Normative Clinical Language Data and Task Specific Effects

Brielle C. Stark, PhD[1] and Charalambos Themistocleous, PhD[2]

1 Indiana University Bloomington, Department of Speech, Language and Hearing Sciences

2 University of Oslo, Department of Special Needs Education

Corresponding author: BC Stark, bcstark@iu.edu, +1 812-855-7760, 2631 E Discovery Parkway Bloomington IN 47408

Highlights

- Automated analysis of 9,955 discourse samples across LHD, RHD, TBI, MCI, dementia
- Task type strongly shapes linguistic output in all clinical groups studied
- Task × diagnosis interactions reveal unique linguistic “fingerprints”
- Nearly 300 phonological, lexical, syntactic, semantic features analyzed
- Framework guides task-specific assessment and therapy in neurogenic disorders

Normative Clinical Language Data and Task Specific Effects

Abstract

Language production in clinical populations varies not only by neurological condition but also by the type of discourse task used during assessment. This study introduces a detailed analysis of connected speech production tasks—Narrative, Procedural, and Picture-based—conducted in etiologically heterogeneous patient groups with neurological damage, namely, 1,394 individuals with Left and Right Hemisphere Damage (LHD, RHD), Traumatic Brain Injury (TBI), Mild Cognitive Impairment (MCI), dementia, and healthy controls. Drawing on over 290 linguistic features spanning phonology, morphology, syntax, semantics, lexicon, and readability, we conducted a large-scale analysis of spoken texts using mixed-effects models to isolate the effects of diagnosis, task, and their interaction. Results revealed that task type exerts a pervasive influence on linguistic output, significantly interacting with diagnostic group across nearly all linguistic domains. Narrative tasks elicited more complex syntax and aspectual morphology, while procedural and descriptive tasks prompted simpler grammatical structures and greater lexical diversity, respectively. To explore high-dimensional linguistic patterns visually, we applied Uniform Manifold Approximation and Projection (UMAP), which revealed clear clustering of observations according to task type. The spatial arrangement of clusters reflected a continuum of contextual constraint: from highly structured picture description tasks (*e.g.*, Cookie Theft), through procedural instruction tasks (*e.g.*, sandwich-making), to conversational speech and open-ended narratives (*e.g.*, accounts of brain injury, recovery, or illness). This gradient likely corresponds to increasing cognitive–linguistic demands, with unstructured tasks requiring greater topic generation, discourse planning, and memory retrieval. Crucially, the interaction between task and diagnosis modulates how underlying impairments manifest, indicating that linguistic deficits are not uniformly expressed across tasks. This challenges the validity of task-agnostic assessment and analysis practices, as well as normative comparisons. We propose a task-specific framework for clinical language assessment, enabling more precise identification of linguistic impairments and informing targeted therapeutic interventions. Our findings underscore the necessity of accounting for task effects in both clinical evaluation and the development of computational tools for neurogenic language disorders.

1 Introduction

Language production in neurogenic communication disorders is shaped not only by lesion location or disease process but also by the discourse task used to elicit speech. Research in aphasia, traumatic brain injury (TBI), dementia, and mild cognitive impairment (MCI) has consistently demonstrated that task structure drives systematic variation in connected speech, often interacting with diagnostic profile to reveal or obscure impairments (Fridriksson et al., 2018; Stark & Fukuyama, 2021). For example, narrative tasks (e.g., story retell) typically elicit greater syntactic complexity, denser propositional content, and increased use of story grammar elements (Richardson et al., 2021; Stark & Fukuyama, 2021). Procedural tasks (e.g., explaining how to make a sandwich) yield simplified syntax, frequent imperative constructions, and a higher noun-to-verb ratio, alongside heavy use of prepositions and "light" verbs (Stark & Fukuyama, 2021; Ulatowska et al., 1981). Picture descriptions emphasize lexical retrieval of core scene items and promote lexical variety, but they often elicit fewer words overall, rely on simpler clause structures, and show a high proportion of nouns relative to verbs, with limited cohesion (Fergadiotis & and Wright, 2011; Stark Brielle, 2019). Picture sequences add demands for temporal sequencing and causal reasoning, eliciting more global coherence markers and causal connectives than single pictures, but with relatively simple syntax (Stark & Fukuyama, 2021). These genre-specific profiles are reproducible across samples and languages, underscoring that discourse tasks are not interchangeable.

Critically, the expression of impairment depends on task-by-diagnosis interactions. In comparative studies of aphasia and age-matched controls, group differences co-occurred with robust genre effects: narrative retells were richest in content but slowest in rate, procedural discourse was simplest syntactically, and even structurally similar picture description tasks produced significantly different linguistic profiles (Stark & Fukuyama, 2021). Moreover, linguistic markers themselves differ by task: people with aphasia show reduced lexical diversity, informativeness, and efficiency, but the extent of these deficits varies depending on elicitation context (Fergadiotis & and Wright, 2011; Stark et al., 2023; Stark & Fukuyama, 2021; Stark Brielle, 2019). Altogether, this evidence demonstrates that the same individual may present with variable linguistic characteristics depending on whether the elicitation method is descriptive, procedural, or narrative, complicating diagnosis and longitudinal tracking when tasks are treated as equivalent

More recent evidence shows that task effects extend beyond overt aphasia. Speakers with latent aphasia and MCI were differentiated from cognitively healthy adults on core lexicon measures (Kim et al., 2024), but the sensitivity of those measures depended on task type (Stark et al., 2025). For example, Cinderella retellings revealed between-group differences most reliably, whereas picture descriptions sometimes masked subtle impairments. This pattern aligns with theoretical accounts of discourse processing: highly structured tasks like the Cookie Theft scene provide strong contextual scaffolding, while open-ended narratives demand greater topic generation, sequencing, and memory retrieval, thereby amplifying cognitive-linguistic vulnerabilities (Stark & Fukuyama, 2021). Similarly, patients with TBI, whilst demonstrating some impaired microlinguistic features, typically associate with reduced macrostructural organization and pragmatic appropriateness (Coelho, 2007), while those with right hemisphere damage demonstrate preserved sentence-level grammar but impaired inferencing, cohesion, and global coherence (Minga et al., 2022).

The clinical implications are significant. Despite multiple elicitation contexts improving the validity of a comprehensive linguistic assessment, most protocols rely on a single task (Bryant et al., 2016; Cruice et al., 2020; Stark & Fukuyama, 2021), often the Cookie Theft picture description, because transcription and coding remain time-consuming. Reliability studies likewise show that test–retest stability of discourse measures is task- and sample-specific, rather than a fixed property of a measure (Stark et al., 2023). Thus, relying on one probe both obscures diagnosis and undermines sensitivity to treatment-related change

Taken together, existing work supports three conclusions. First, discourse tasks are not interchangeable; genre and stimulus constrain lexical, syntactic, and discourse patterns in predictable ways. Second, diagnosis interacts with task such that impairment profiles emerge differently depending on discourse context. Third, clinical validity and reliability depend on aligning tasks with the linguistic construct of interest, rather than assuming generalization across elicitation methods.

2 Lexical Production, Grammar, and Readability

Computational linguistic analysis has the potential to automate speech and language analysis and offer quick and easy analysis of speech productions (Callegari & et al., 2024; Charniak & Johnson, 2005; Devlin & et al., 2018; Fraser & et al., 2015; Goldstein & et al., 2022). We have

developed Open Brain AI, a computational platform to automate language (Themistocleous, 2024). Open Brain AI combines computational methods like advanced AI methodologies, including machine learning, natural language processing (NLP), large language models (LLMs), and automated speech-to-text transcription. Open Brain AI combines various computational methods for in-depth linguistic assessment, such as morphosyntactic analysis, named entity recognition, phonology, morphology, syntax, semantics, and lexical measures. It offers features like speech transcription, grammar error detection, and comprehensive linguistic assessments across multiple languages.

Lexical production includes measures about the distribution of words and relationships between types and tokens that can quantify how words are used in different contexts and how they contribute to the overall meaning of a text such as lexical diversity measures (Fergadiotis & and Wright, 2011; Kim et al., 2023). Lexical features developed by Open Brain AI Themistocleous (2024) are employed for Profiling Lexical Retrieval, these involve, Content Words (total & unique) determine noun/verb retrieval deficits and low unique content-word count highlights word-finding difficulty (Efstratiadou & et al., 2018); Assessing Fluency vs. Density, these are measures of Lexical Density balances fluency (more tokens) against informativeness (content words). For example, in non-fluent aphasia, patients often produce more content words than function words; in fluent aphasia, the reverse (Themistocleous, Ficek, et al., 2020; Themistocleous, Webster, et al., 2020); Measuring Lexical Diversity, these are well known measures like the type-token ratio (TTR), corrected TTR (CTTR), and Maas's A^2 help distinguish true lexical richness from mere verbosity or repetition and indicate track progress in therapy or compare severity across patients; Detecting Syntactic Simplification, which includes the Average Word Length and function-word ratios, the latter can reflect trade-offs between lexical complexity (as measured by content words) and grammatical structure. Total words may differ between different types of tasks as produced by different groups of patients. Crucially, in (Stark et al., 2023) total word count proved an unreliable measure of linguistic ability in the aphasia group, reflecting overall severity more than task performance.

Phonology measures quantify how users employ speech sounds, the sound combinations, and the complexity of syllables. Comparing these measures across patients with different language impairments can reveal characteristics that pertain to the effects of impairment on the cognitive representation of sounds and speech production (Barbieri et al., 2018; Croot et al.,

2000). The phonological measures in Open Brain AI Themistocleous (2024) process the text and provide information about Syllable-Shape Frequencies count how many syllables in the sample conform to each consonant-vowel (CV) templates (e.g., CV, CVC, CCCV, V, VC, VCC). These aim to function as a measure of Articulatory/Phonological Complexity. Multiconsonantal onsets or codas (e.g. CCCV, CCVCC) demand more precise phonological planning and motor control than simple CV or CVC shapes. Speakers with apraxia or conduction aphasia often simplify clusters (avoiding CCCV or CCVCC) or reduce complex codas (CVCC → CVC). Tracking which shapes drop out—or are disproportionately rare—can pinpoint a phonological/articulatory bottleneck. Additionally, it provides calculations of the total number of syllables produced, providing it a measure of Word-Level Complexity, namely whether polysyllable (long) words are under-represented in patients with different symptomatology. For example, patients with non-fluent aphasia may have a significantly fewer syllables whereas fluent but empty speech may have many syllables but low content density.

Morphological measures quantify the structure and form of words, the distribution of parts of speech, and inflectional categories, such as tense, Number, Gender, and Case. Comparing patients with morphology impairments can reveal pathologies, like agrammatism and anomia (Badecker et al., 1990; Caramazza & Hillis, 1991; Fridriksson et al., 2018; Hillis, 1989; Hillis et al., 2018; Stockbridge et al., 2021). Table 1 presents six key linguistic measure groups developed in Open Brain AI Themistocleous (2024) for use clinical language assessment, detailing specific grammatical categories within each group and their associated clinical patterns. The measures range from basic part-of-speech classifications to complex morphological and syntactic features, with clinical insights highlighting how different language impairments (such as non-fluent/agrammatic and fluent/Wernicke's aphasia) manifest as specific patterns of linguistic breakdown or compensatory strategies.

Table 1 Morphology and Inflectional Morphology Measures for Clinical Assessment.

Measure Group	Examples (Count & Ratio)	Clinical Insight
POS Classes	Noun, Verb, Adjective, Adverb, Function words	Non-fluent/agrammatic: ↓nouns & verbs, ↑function words; fluent/Wernicke's: errors in class labels
Verb Inflectional Morphology	Tense (Past/Present), Aspect (Perfective/Progressive), Mood (Indicative), Voice (passive auxiliaries)	Loss of tense/aspect or passive marking → breakdown in grammatical morphology

Noun Inflectional Morphology	Number (Singular/Plural), Gender (Masc/Fem/Neut), Case (Nom/Acc/Dative), Possessive	Avoidance of plurals, gender mismatches, dropped case or possessives → reduced inflectional control
Pronouns & Determiners	Personal/Demonstrative/Relative Pronouns; Determiner; Definite/Indefinite articles	Over-use of "it" or "the" and omission of "a"/"this" → reliance on "easy" words, retrieval failure
Connectives	Coordinating (and/but/or), Subordinating (because/when), Degree (comparative/positive)	Few conjunctions or comparatives → simplified clause structure and morphology avoidance
Polarity & Particles	Negation (not/never), Particles (out, up), Interjections (um, well), Modal verbs	Low negation or modal use; high filler/interjection rates → syntactic simplification or planning gap

Note. Count: raw token count for that category; Ratio: (category tokens)/(total words) or (total class tokens)

Syntactic measures quantify impairments of sentence structure (e.g., subject-verb-object order), grammatical rules (e.g., agreement between subject and verb), and phrase structure (e.g., noun phrases, verb phrases) (Bastiaanse, 2013; Caramazza & Hillis, 1989; Mack et al., 2021; Thompson & Mack, 2014; Wilson et al., 2016). By mapping which phrase types, modifiers, clauses, and tree-depth metrics drop out or simplify, clinicians can pinpoint whether breakdowns lie in phrase building, clause embedding, sentence planning, or hierarchical complexity. Table 2 presents five linguistic measure groups developed in Open Brain AI Themistocleous (2024) for use clinical language assessment, detailing specific grammatical categories within each group and their associated clinical patterns. The mapping of phrase types, modifiers, clauses, and tree-depth metrics allows clinicians to detect syntax related impairments, as in phrase building, clause embedding, sentence planning, and hierarchical complexity.

Table 2 Syntax and Inflectional Syntax Measures for Clinical Assessment.

Measure Group	Examples (Count & Ratio)	Clinical Insight
Phrase Structures	Noun Phrases, Verb Phrases, Adjective Phrases, Adverbial Phrases, Prepositional Phrases	Fewer or truncated phrases indicate limited hierarchical embedding (non-fluent output).

Modifiers & Complements	Adjectival modifier, Adjectival complement, Adverbial modifier, Appositional modifier, Relative clause modifier	Reduced modifiers/complements signal simplified noun and clause expansions.
Clause Constructions	Direct object, Object of preposition, Clausal complement, Open clausal complement, Nominal subject passive, Passive Sentences Percent	Low counts of objects/complements or passives reveal avoidance of complex clause roles and voice alternations.
Sentence & Clause Counts	Total Sentences, Clauses Total Clauses in Text, Total Dependent Clauses, Total Coordinate Phrases, Avg /Min/Max Sentence Length	Short sentences with few clauses or coordination reflects non-fluent, agrammatic patterns.
Tree & Branch Metrics	Mean/Max Classical Yngve Load, Mean/Max Left Branching Depth, Average Tree Height, Total Tree Height	Measures of Syntactic Complexity. Lower tree heights and branching loads indicate shallow syntactic structure and reduced processing demands.

Note. Count: raw tokens; *Ratio*: tokens in category / total words (or per relevant class).

Semantics quantify predefined semantic categories such as the names of persons, organizations, locations, expressions of times, quantities, monetary values, and percentages. Semantics helps in understanding how meaning is constructed and interpreted in language. raw count (how many times that type of entity appears) and a ratio (that count normalized by total words or total content words) of semantic measures using a named-entity recognition method but in case of impairment are critical for setting therapeutic targets, in assessment, and diagnostis (Antonucci, 2009; Boyle, 2010; Boyle & Coelho, 1995; Coelho & Boyle, 2000; Efstratiadou & et al., 2018; Gravier & et al., 2018; Hashimoto & Frome, 2011; Kiran & Thompson, 2003; Wallace & Wood, 2013; Wambaugh & et al., 2013). Open Brain AI Themistocleous (2024) computes for each category both a to control for overall speech output. Clinically, a low cardinal number ratio may reveal difficulty retrieving numeric concepts, whereas reduced date mentions can signal impaired temporal orientation or event-memory recall (common in amnestic syndromes); likewise, fewer "person" or "organization" entities often marks classic anomia or semantic-dementia profiles, where proper-name retrieval and real-world schema access break down—even if fluency and basic grammar remain relatively intact.

Readability metrics quantify how easy a text can be to be read and understood by a reader. (Dale & Chall, 1948; Fitzsimmons et al., 2010; Klare, 1974; Themistocleous, 2024). Open Brain AI Themistocleous (2024) provides Readability indices all mine simple text features—word length, sentence length, syllable counts, and word familiarity—to estimate how hard a passage is and what grade level or time it requires. The Automated Readability Index (ARI) uses the ratio of characters per word and words per sentence to output a U.S. grade level, while the Coleman–Liau Indexswaps syllable counts for average letters per 100 words and sentences per 100 words to do the same. The Dale–Chall Readability Score compares each word against a 3,000-word "easy" list, computes the percentage of "difficult words," and then applies an adjustment to yield a grade; Difficult Words simply counts how many words fall outside that list. Estimated Reading Time divides total words by a standard reading speed (e.g., 200 wpm) to give seconds. The Flesch–Kincaid Grade Level and its cousin Flesch Reading Ease both combine average sentence length with average syllables per word—the former spits out a school-grade level, the latter a 0–100 ease score. The Gunning Fog Index tallies words of three or more syllables plus average sentence length to estimate required years of education, while the Linsear Write Formula scores easy (≤2 syllables) versus hard (>2) words in a 100-word sample to derive a grade level. Finally, the SMOG Index samples thirty sentences, counts all polysyllabic words, and applies a square-root formula to predict the necessary education level. By triangulating these scores, you get complementary views on vocabulary difficulty, morphological complexity, sentence structure, and overall text accessibility. It is typically influenced by factors such as sentence length, word complexity, and the overall structure of the text. Text produced by healthy controls should be readable but in patients with neurological conditions readability might be easier because the produce simpler texts or might be more complex, for example by containing hapax legomena (i.e., novel and unknown words).

3 This study

In clinical practice, we know that the tasks we use to assess language matter. However, it is often unclear precisely *how* a patient's performance on a picture description differs from their ability to tell a story, making it challenging to select the best tasks to reveal specific deficits or to set targeted treatment goals. Furthermore, the manual transcription and analysis of these language samples is an incredibly laborious and time-consuming process. This study aims to bridge this

gap by using a powerful and automated computational platform, Open Brain AI, to analyze a massive dataset of clinical language samples and by providing a comprehensive set of features that cover the linguistic profile of speakers unlike earlier studies that focused on a few mostly morphological and lexical measures. This aim is critical because it has the potential to explaining the feasibility of comparing tasks for domains of language such as phonology (e.g., phonological paraphasia and phonological errors), morphology (agrammatism), and semantics (semantic paraphasia and naming). Our goals are to:

1) Provide a clear, evidence-based guide on how different assessment tasks (e.g., picture description, storytelling) influence the speech and language production of patients with various neurological conditions, LHD, RHD, TBI, MCI, and dementia.
2) Identify the unique "linguistic fingerprints" that distinguish conditions like LHD, RHD, TBI, MCI, and dementia within specific tasks, leading to more nuanced assessment.
3) Help clinicians select the most effective tasks strategically to assess or treat specific linguistic domains (e.g., vocabulary, sentence structure, narrative coherence) and to measure therapeutic progress more efficiently.

To achieve these aims, our research is guided by the following core questions:

1) Do tasks vary in each measure and which linguistic features are more sensitive to task variation?
2) How do different neurological conditions reveal provide unique linguistic patterns within each task? For example, when describing the same picture, do individuals with TBI simplify their sentences differently than individuals with dementia?

Based on prior research and clinical theory, we predict the following outcomes. First, we hypothesized that different tasks would elicit predictably different language, regardless of the patient's diagnosis. Specifically, *Narrative Tasks* (e.g., retelling "Cinderella," sharing a personal story) will produce speech that is rich in content and coherence but may feature slower speech rates and more common, less diverse vocabulary. *Procedural Tasks (*e.g., "how to make a sandwich") will yield the simplest grammar, characterized by shorter sentences and a more direct, noun-heavy style; and *Picture Description Tasks* (e.g., "Cookie Theft") will prompt the

greatest vocabulary variety (lexical diversity) as speakers access specific labels for objects and actions. Second, we hypothesize that while the task effect is universal, it will be magnified or altered by a specific neurological condition. For instance, the demand for complex sentence structure in a narrative task will likely reveal more severe syntactic errors in a patient with agrammatic aphasia compared to a patient with dementia, who might instead show more prominent breakdowns in semantic coherence and topic maintenance in the same task. Third, we predict that we can create a "clinical roadmap" for task selection. Picture descriptions will prove superior for quickly assessing word-finding and vocabulary, while personal narratives will be most sensitive for evaluating higher-level discourse organization and pragmatic skills. This will empower clinicians to choose the right tool for the right therapeutic job.

This study assesses the distribution of lexicogrammatical features across various clinical discourse tasks to quantify neurolinguistic differences by profiling phonology, morphology, syntax, semantics, lexicon, and readability. To achieve this, we have constructed a substantial and well-annotated corpus of 299 measures we implemented in the online platform Open Brain AI Themistocleous (2024) leveraging an exceptionally large dataset of 9,955 speech samples. This extensive compilation is drawn from diverse and established TalkBank resources (MacWhinney, 2025), including AphasiaBank, DementiaBank, TBI Bank, and RHD Bank. Such a varied collection allows us to capture the inherent heterogeneity within clinical populations, enabling a more robust and generalizable analysis of linguistic signatures across conditions such as LHD, RHD, dementia, MCI, TBI, and healthy controls. In total, the dataset contains measures from fifteen distinct tasks, which can be categorized into four broad elicitation types, each designed to portray various aspects of discourse. The first category, picture description and connected speech, includes tasks such as describing the classic "Cookie Theft" scene, a single-image cartoon of a cat rescue, an image of a person with an umbrella, the "Broken Window" picture-sequence, a multi-panel scene of a flood, a Norman Rockwell illustration, and images depicting illness or someone giving a speech. A second category involves narrative retelling, where patients are asked to recount the "Cinderella" fairy tale, a task that taps into memory, sequencing, and cohesive devices. The third type, procedural explanation, elicits step-by-step instructions and temporal language through prompts like explaining how to make a peanut butter and jelly sandwich. Finally, personal event narratives are prompted by asking patients to describe

an important life event, a time they were sick, or to recount their own neurological event and subsequent recovery. By combining these task types, clinicians can effectively elicit and evaluate lexical richness, syntactic complexity, discourse cohesion, and pragmatic structure.

The present study extends this literature by leveraging a large, heterogeneous corpus of discourse samples across left and right hemisphere damage, TBI, dementia, MCI, and healthy controls. Using automated, comprehensive linguistic computational profiling across nearly 300 linguistic features, we systematically test how task-by-diagnosis interactions modulate phonological, morphological, syntactic, semantic, lexical, and readability measures. Prior work has evaluated only a few constructs per linguistic profile (e.g., using "MLU" to evaluate syntax; (Bryant et al., 2016)) despite there being robust parameters to evaluate within each linguistic construct. Some metrics are presented for the first time in this study, like measures about phonological information across tasks, readability measures, which are critical for understanding how easily a text can be understood by a reader, and measure of morphological inflection and syntactic measures, which we designed by this team to evaluate the ease of understanding text produced by patients with a range of neurological conditions. Our goals are therefore to delineate the linguistic fingerprints of different task types across diagnostic groups, identify which measures are most sensitive to task-driven variation, and provide an evidence-based framework for task selection in clinical and research contexts. By doing so, we address a critical gap in current practice: the need to move beyond task-agnostic assessment toward a task-specific, diagnosis-informed model of discourse evaluation.

4 Methodology

4.1 Participants and Data

The individuals for this study were drawn from the TalkBank database (MacWhinney, 2025), a comprehensive resource featuring language samples from people with various neurological conditions. The inclusion of data from individuals with no known neurological or language disorders serves as a critical baseline for comparison with these clinical groups, allowing for a thorough analysis of language impairments across different conditions. The databases within the TalkBank consortium, such as AphasiaBank (MacWhinney et al., 2011), RHD Bank (Minga et al., 2022), TBI Bank (Elbourn et al., 2019; Steel et al., 2017), and DementiaBank (Lanzi Alyssa et al., 2023), adhere to shared, standardized protocols for assessment and data collection.

However, it is important to note that each clinical bank is a compilation of contributions from numerous sites, which can result in slight variations in how data is collected. A common thread across these banks is an established discourse protocol that elicits a variety of genres, including picture descriptions, story narratives, procedural explanations, and personal narratives. The present analysis is based on a comprehensive corpus of 9,900 language samples produced by the individuals whose demographic characteristics are detailed in Table 2 (with further details in Table 1 and Supplementary Data). As many participants completed several different tasks, they often contributed multiple samples to this dataset. A defining feature of these source databases is their significant clinical heterogeneity. The LHD cohort, for instance, encompasses various clinical subtypes, including anomic, Wernicke's, and Broca's aphasia. This diversity is further illustrated by the dementia subgroup from the Pitt corpus (N=193), which consists primarily of patients diagnosed with dementia (91%), who have lower average Mini-Mental State Examination (MMSE) scores of 17–18, but also includes individuals with Mild Cognitive Impairment (MCI).

Specifically, AphasiaBank provides spoken discourse samples from 536 individuals with Left Hemisphere Damage (LHD) and 359 healthy controls, emphasizing connected speech to study language production and its neural foundations. The Right Hemisphere Damage Bank (RHDBank) focuses on communication in 38 individuals with RHD and 40 healthy controls, targeting pragmatic language abilities and discourse coherence. TBIBank offers a multimedia database for studying communication disorders in 58 individuals with Traumatic Brain Injury, utilizing tasks like story retelling and personal narratives; while some datasets within TBIBank are longitudinal to identify recovery patterns, others are not. Finally, DementiaBank includes several valuable corpora. The Delaware MCI dataset contains language productions from 71 adults with Mild Cognitive Impairment, aiding in the early detection of subtle language changes. Similarly, the Pitt Study dataset, also part of DementiaBank, includes “Cookie Theft” picture descriptions from 193 individuals with dementia and 99 healthy controls, providing a basis for detecting language abnormalities. Together, these resources offer a rich foundation for investigating neurolinguistic profiles. A key methodological decision was to incorporate these databases in their entirety. This approach preserves the ecological validity of the data, ensuring our findings reflect the natural heterogeneity inherent in clinical populations. Moreover, using

these standard corpora without modification maintains their integrity, a crucial factor for ensuring the reproducibility and comparability of our results within the wider research community.

Table 3 Demographic Characteristics of Participant Groups by Diagnosis and Data Source

Diagnostic Group	**Data Source**	**N (Speakers)**	**Age in Years (M, SD)**	**Education in Years (M, SD)**
Healthy Control (HC)	AphasiaBank	359	56.89 (15.91)	15.91 (2.64)
	RHD Bank	40	47.95 (13.54)	17.09 (2.93)
	DementiaBank (Pitt)	99	63.70 (7.90)	13.90 (2.50)
Left Hemisphere Damage (LHD)	AphasiaBank	536	61.04 (12.40)	15.70 (2.91)
Dementia	DementiaBank (Pitt)	193	71.00 (8.60)	12.20 (2.90)
Mild Cognitive Impairment (MCI)	DementiaBank (Delaware)	71	73.50 (8.03)	*
Right Hemisphere Damage (RHD)	RHD Bank	38	57.40 (12.33)	17.10 (3.99)
Traumatic Brain Injury (TBI)	TBI Bank	58	36.25 (13.47)	13.91 (3.05)

Note. Values for Age and Education are presented as mean (M) and standard deviation (SD). * Education for the MCI group was reported as percentages of highest attainment: PhD (10.81%), Bachelor's/Master's (67.57%), and Vocational Training (21.62%).

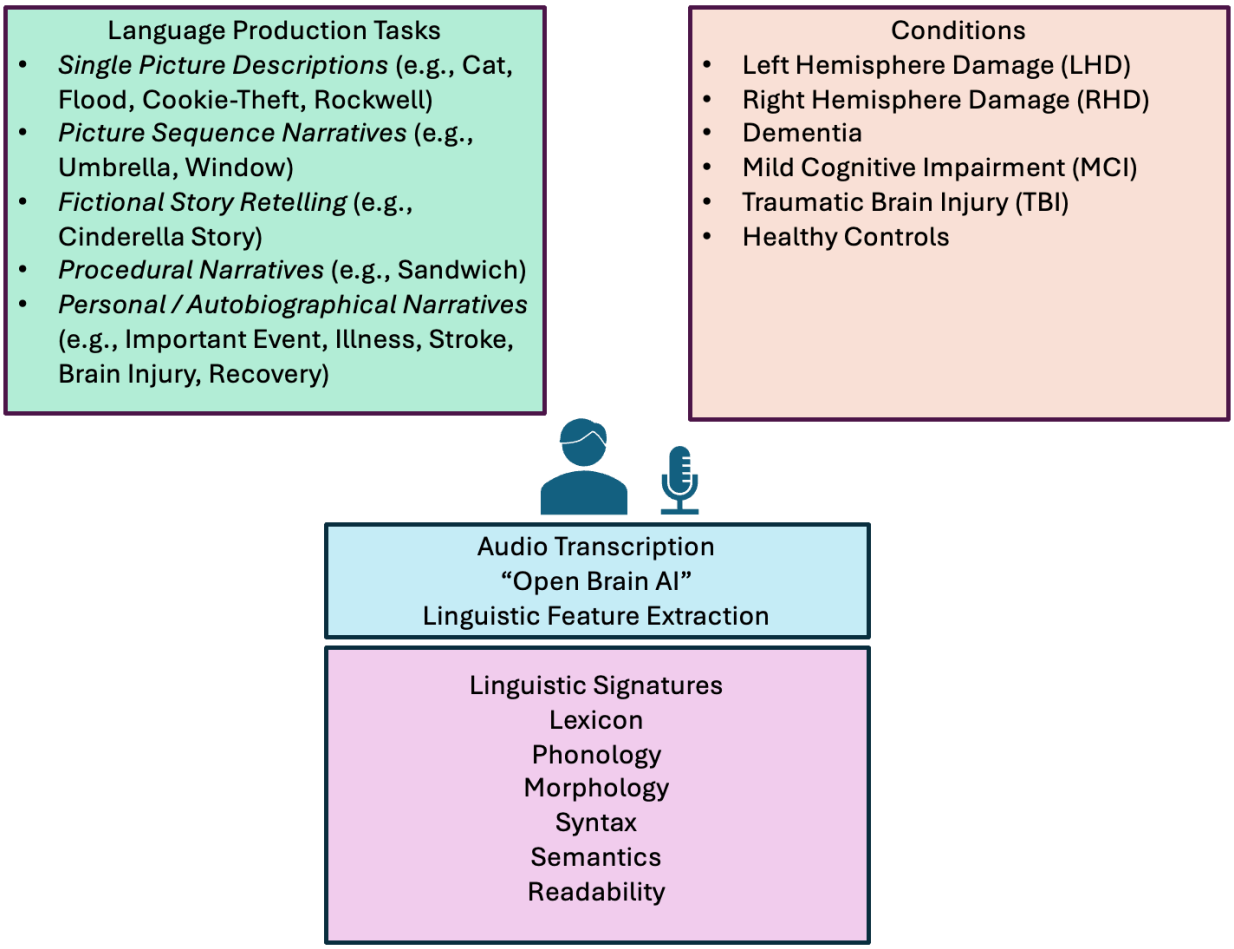


Figure 1 Diagram illustrating a research methodology for studying language production in neurological conditions. The study involves participants from multiple groups (Left Hemisphere Damage, Right Hemisphere Damage, Dementia, Mild Cognitive Impairment, Traumatic Brain Injury, and Healthy Controls) performing various language production tasks including narrative tasks like retelling "Cinderella," procedural tasks such as explaining "how to make a sandwich," and picture description tasks like describing "Cookie Theft." The participants' speech is recorded and the transcripts processed through "Open Brain AI" for linguistic feature extraction, which analyzes multiple linguistic signatures including lexicon, phonology, morphology, syntax, semantics, and readability measures.

Data from the TalkBank consortium clinical banks is available by becoming a member to the TalkBank project.

4.2 Measures

The speech samples were subjected to automated linguistic analysis using OpenBrainAI (http://openbrainai.com; Figure 1), a custom clinical linguistics platform we developed to address the limitations of standard computational tools (Themistocleous, 2024). Unlike generic models, OpenBrainAI is specifically designed for the granular phenotyping of language features, enabling hypothesis-driven research into speech pathology and neurogenic communication disorders. The platform executes a cascade of Natural Language Processing (NLP) techniques, beginning with core steps such as tokenization, part-of-speech tagging, and dependency parsing. For each feature extracted from this process, both raw counts and normalized ratios are computed to account for variations in text length. These quantitative linguistic data were then automatically exported as spreadsheet files, facilitating subsequent statistical analysis. A complete list of all measures and their operational definitions is provided in Appendix 1. Given this large feature set, the analyses presented in this paper prioritize a subset of measures selected for their high sensitivity and specificity in distinguishing between the diagnostic groups and healthy controls. An exhaustive output of all statistical results for every measure is available in the Supplementary Materials.

From these foundational analyses, a comprehensive suite of linguistic measures was automatically extracted to quantify multiple dimensions of language production. These included readability metrics, such as the Flesch Reading Ease and Gunning Fog Index, which assess textual complexity. The lexicon was analyzed through features of vocabulary richness and diversity, including measures like Type-Token Ratio and counts of content versus function words. Phonological characteristics were quantified by tabulating word counts by syllable number and the distribution of various consonant-vowel syllable structures. Morphological analysis encompassed the distribution of parts of speech and inflectional categories. Syntactic complexity was measured through the quantification of phrase types, an analysis of core grammatical dependencies, and metrics such as average sentence length and T-units. Finally, semantic analysis focused on Named Entity Recognition (NER) to identify and categorize entities like persons, organizations, and locations.

These grammatical analyses were grounded in the Universal Dependencies framework for standardized annotation (Nivre et al., 2020). The measures were systematically selected to include both established metrics with demonstrated sensitivity to pathological language changes—such as noun and verb counts—and novel measures targeting microstructural (phonology, morphology), macrostructural (syntax, semantics), and pragmatic dimensions. This comprehensive approach aims to characterize language impairments in a way that aligns with current models of linguistic breakdown, enabling the detection of subtle but clinically significant changes that might otherwise be overlooked.

4.3 Statistical Analysis

We performed two types of analysis. An unsupervised ML approach and a generalized linear mixed effects model. These are presented in the following sections.

4.3.1 Unsupervised Machine Learning

The number of measures resulted in a very high-dimensional dataset, with each row representing a speech-language sample tied to a specific Task (like Cookie, Cat, Cinderella, etc.) and containing hundreds of linguistic features—ratios, counts, complexity metrics, and phonetic patterns. This richness enables us to provide a detailed analysis but makes direct representation of the data structure and visualization impossible without dimensionality reduction. This is a critical step for us to answer the first question on whether the measures differ based on the tasks.

We evaluated two main Machine Learning Models for unsupervised analysis: a PCA (Principal Component Analysis), which is a linear dimensionality reduction method and has been universally used for analyzing neurolinguistic data in several studies (Fridriksson et al., 2018; Ingram et al., 2020; Lacey et al., 2017; Marcotte et al., 2017). The PCA models detect new orthogonal axes (“principal components”) that capture the greatest variance in the data: PC1 (the first PCA component) explains the most variance, followed by PC2, and so on. PCA has the advantage that it preserves global structure, namely the large-scale relationships between samples. However, it assumes linear relationships; can miss non-linear patterns. To address this point we conducted a second ML approach, the Uniform Manifold Approximation and Projection (UMAP), which is a non-linear dimensionality reduction method, as we hypothesize

that non-linearities will be prevalent in the dataset given its variation (Ghojogh et al., 2023). The UMAP constructs a high-dimensional graph of data points based on nearest neighbor and optimizes a low-dimensional embedding to preserve local neighborhood structure.

4.3.2 Mixed effects Models

To assess the influence of clinical diagnosis on each linguistic outcome variable, we utilized an automated mixed-effects modeling pipeline. This analysis included participants from the five diagnostic groups (LHD, Dementia, MCI, RHS, TBI) and the Healthy Control (HC) group. The pipeline, developed in R (R Core Team, 2025) was designed to be flexible, data-driven, and robust to violations of statistical assumptions common in linguistic data.

For each linguistic variable, a linear mixed-effects model was implemented. The core of the analysis was specifying Task, Diagnosis, and their interaction (Task * Diagnosis) as fixed effects. This allows the model to determine not only the main effect of the task and the diagnosis but, more importantly, whether the linguistic differences between tasks are dependent on the patient's diagnosis.

To appropriately account for the non-independence of data arising from the study design, a random intercept was included for each participant via the (1 | SpeakerID) term. This term addresses that multiple observations (i.e., linguistic measures from one or more tasks) originate from the same individual. By including this random intercept, the model accounts for individual-specific baseline differences in linguistic performance, thereby modeling the repeated measures dimension of the data. This structure is robust to the unbalanced nature of task administration (i.e., not all participants completed all tasks).

The general model structure was:

$$Outcome \sim Diagnosis \times \text{Task} + (1|\text{Speaker}) \quad (1)$$

The analytical pipeline systematically selected the most appropriate statistical model based on the distribution of each dependent variable. This adaptive process involved fitting Gaussian Linear Mixed-Effects Models (LMMs) for continuous variables, using robust LMMs if residual diagnostics (via the DHARMa package (Hartig, 2016)) indicated violations of model assumptions, and employing Generalized Linear Mixed-Effects Models (GLMMs) with appropriate distributions (e.g., binomial, Poisson, or negative binomial) for binary or count data. If a suitable model could not be fitted through these steps, a rank-based LMM was applied as a

robust fallback. (Further details on the specific model selection criteria and R packages, such as lmerTest (Kuznetsova et al., 2016) and robustlmm (Koller, 2016) are available in the script).

When a significant Task: Diagnosis interaction was found ($p < .05$) from a Type III ANOVA, post-hoc pairwise comparisons were conducted to understand the nature of the interaction. These comparisons were made between tasks within each diagnostic group using estimated marginal means (via the emmeans package (Russell, 2020)). Tukey's method was applied to adjust for multiple comparisons. This approach allows for the identification of specific task contrasts that are significantly different for one diagnostic group but not necessarily for others.

5 Results

The first research question aims to determine whether linguistic measures differ significantly between tasks and which linguistic features are more sensitive to task variation. To answer this question, we performed two types of analysis: an unsupervised machine learning and a post-hoc analysis of glmer results for the effects of Task and Diagnosis on each linguistic measure. We then provide the ten most significant ones that discriminate the tasks in terms of absolute t-ratio.

The unsupervised ML, using information from all linguistic levels, revealed clear clusters, which show that the type of task a clinician selects for the analysis of connected speech determines the distribution. PCA components explain large parts of the variance (PC1 = 28%, PC2 = 5%), in other words the first two PCA dimensions captured ~33% of the total variance (PC1 = 28%, PC2 = 5%), which is substantial given the high dimensionality of the linguistic dataset. PCA emphasizes global variance, whereas UMAP preserves local and nonlinear structure, making the latter more suitable for visualizing clinically relevant clusters. Overall, the UMAP emphasizes local neighborhood preservation and non-linearity. Trustworthiness on the full dataset was for UMAP 0.868 (closer to 1 indicates better local structure preservation). The displayed UMAP plot removed outliers (IQR; k=1.5) to enhance interpretability (kept 9791/9996 points). These 2D maps reveal clusters/subgroups of speech-language patterns related to diagnostic categories or therapy response, flag potential outliers for review, and support communication across clinical teams (Figure 2).

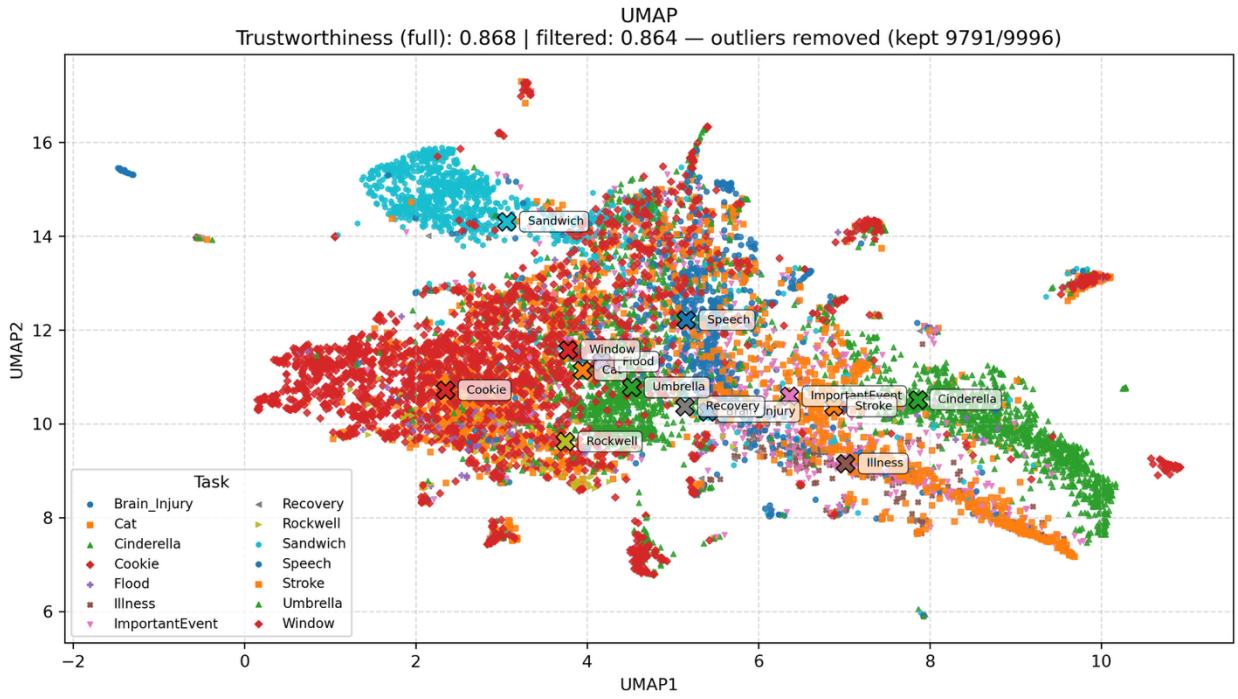


Figure 2 *UMAP projection of linguistic feature embeddings across tasks.* Each point represents one sample, colored and shaped by task. The 2D layout emphasizes local neighborhood structure (trustworthiness = 0.868 full, 0.864 filtered). Outliers were excluded using IQR filtering (kept 9791/9996 points). The x symbol and the corresponding label indicate the centroid of each group.

As a reminder, the tasks cluster, broadly into these genres: Single Picture Descriptions (Cat, Flood, Cookie-Theft, Rockwell), Picture Sequence Narratives (Umbrella, Window), Fictional Story Retelling (e.g., Cinderella, Fictional Retelling), Procedural Narratives (Sandwich), and Personal / Autobiographical Narratives (Important Event, Illness, Stroke, Brain Injury, Recovery).

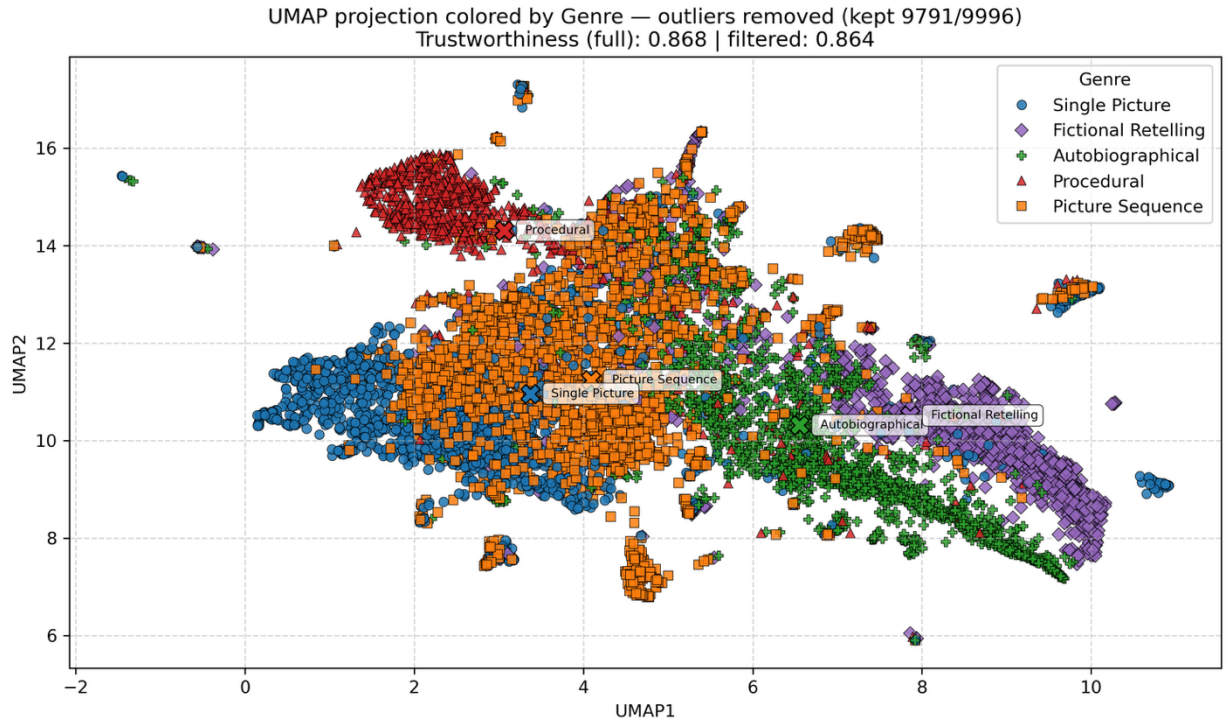


Figure 3*UMAP projection of linguistic feature embeddings across tasks (colors indicate genre categories).* Each point represents one sample, colored and shaped by task. The 2D layout emphasizes local neighborhood structure (trustworthiness = 0.868 full, 0.864 filtered). Outliers were excluded using IQR filtering (kept 9791/9996 points). The x symbol and the corresponding label indicate the centroid of each group.

The unsupervised ML highlights the global data structure. Each point represents an individual observation projected from high-dimensional linguistic feature space into two dimensions using Uniform Manifold Approximation and Projection (UMAP). Colors and shapes indicate the task associated with each observation. The algorithm emphasizes preservation of local neighborhood structure, allowing similar linguistic profiles to appear close together. Outliers, identified using the interquartile range (IQR) method, were removed to improve visualization clarity. Clusters or separations in the plot may reflect underlying similarities or differences in speech–language patterns relevant to diagnosis or treatment monitoring. Specifically, the clusters in Figure 3, almost read from left to right: close context tasks (Single Picture, Procedural) and autobiographical and fictional retelling to right.

Table 4. Top 10 Linguistic Features Discriminating Tasks and Diagnostic Groups.

Rank	Diagnosis	Variable	Contrast	|t-ratio|	p.value
1.00	LHD	Pronoun Type: Relative Ratio	Speech - Stroke	12.42	0.00
2.00	LHD	Pronoun Type: Relative Ratio	Cinderella - Speech	11.11	0.00
3.00	LHD	Pronoun Type: Relative Ratio	Stroke - Window	10.32	0.00
4.00	LHD	Pronoun Type: Relative Ratio	Cinderella - Window	9.31	0.00
5.00	LHD	Pronoun Type: Relative Ratio	Important Event - Speech	8.26	0.00
6.00	LHD	Pronoun Type: Relative Ratio	Sandwich - Stroke	7.83	0.00
7.00	LHD	Pronoun Type: Relative Ratio	Cat - Stroke	7.68	0.00
8.00	LHD	Pronoun Type: Relative Ratio	Cinderella - Sandwich	6.62	0.00
9.00	LHD	5 syllables word	Important Event - Stroke	6.35	0.00
10.00	LHD	Pronoun Type: Relative Ratio	Important Event - Window	6.31	0.00
1.00	HC	Case marker Count	Cat - Cinderella	10.71	0.00
2.00	HC	Case marker Count	Cinderella - Window	9.91	0.00
3.00	HC	Pronoun Type: Relative Ratio	Cookie - Cinderella	9.60	0.00
4.00	HC	Case marker Count	Cinderella - Umbrella	9.15	0.00
5.00	HC	Pronoun Type: Relative Ratio	Cat - Cinderella	7.05	0.00
6.00	HC	Pronoun Type: Relative Ratio	Cookie - Illness	6.88	0.00
7.00	HC	Pronoun Type: Relative Ratio	Cinderella - Flood	6.66	0.00
8.00	HC	5 syllables word	ImportantEvent - Window	6.50	0.00
9.00	HC	Clausal modifier of noun Count	Cinderella - Window	6.49	0.00
10.00	HC	Case marker Count	Cinderella - Sandwich	6.48	0.00
1.00	MCI	Pronoun Type: Relative Ratio	Cat - Cinderella	3.08	0.03
2.00	MCI	Pronoun Type: Relative Ratio	Cookie - Cinderella	2.76	0.08
3.00	MCI	Pronoun Type: Relative Ratio	Cat - Rockwell	2.53	0.15
4.00	MCI	Clausal modifier of noun Count	Cat - Rockwell	2.48	0.10
5.00	MCI	Case marker Count	Cat - Rockwell	2.42	0.15
6.00	MCI	Pronoun Type: Relative Ratio	Cookie - Rockwell	2.23	0.28
7.00	MCI	Clausal modifier of noun Count	Cookie - Rockwell	2.15	0.20
8.00	MCI	Pronoun Type: Relative Ratio	Cinderella - Sandwich	1.81	0.54
9.00	MCI	Clausal modifier of noun Count	Cinderella - Rockwell	1.79	0.38
10.00	MCI	Case marker Count	Cookie - Cat	1.76	0.49
1.00	RHD	Pronoun Type: Relative Ratio	Cookie - Stroke	3.32	0.01
2.00	RHD	5 syllables word	Speech - Stroke	2.61	0.09
3.00	RHD	Pronoun Type: Relative Ratio	Cookie - Cinderella	2.61	0.10
4.00	RHD	Pronoun Type: Relative Ratio	Cat - Stroke	2.60	0.10
5.00	RHD	5 syllables word	Cat - Stroke	2.43	0.15
6.00	RHD	Case marker Count	Cinderella - Stroke	2.19	0.24
7.00	RHD	5 syllables word	Sandwich - Stroke	2.15	0.26
8.00	RHD	Pronoun Type: Relative Ratio	Speech - Stroke	1.98	0.36
9.00	RHD	Clausal modifier of noun Count	Speech - Stroke	1.93	0.38
10.00	RHD	Clausal modifier of noun Count	Sandwich - Stroke	1.88	0.42
1.00	TBI	Pronoun Type: Relative Ratio	Cinderella - Speech	5.50	0.00
2.00	TBI	Pronoun Type: Relative Ratio	Cinderella - Window	5.42	0.00
3.00	TBI	Pronoun Type: Relative Ratio	ImportantEvent - Window	4.49	0.00
4.00	TBI	Pronoun Type: Relative Ratio	Important Event - Speech	4.14	0.00

5.00	TBI	Pronoun Type: Relative Ratio	Cat - Window	3.62	0.01
6.00	TBI	Pronoun Type: Relative Ratio	Umbrella - Window	3.60	0.01
7.00	TBI	Pronoun Type: Relative Ratio	Brain_Injury - Window	3.58	0.01
8.00	TBI	Pronoun Type: Relative Ratio	Recovery - Window	3.54	0.01
9.00	TBI	Pronoun Type: Relative Ratio	Brain_Injury - Cinderella	3.01	0.07
10.00	TBI	Pronoun Type: Relative Ratio	Cinderella - Recovery	3.00	0.07

close context tasks (Single Picture, Procedural) and autobiographical and fictional retelling to right.

Table 4 shows the local relationship between the task-diagnosis and linguistic measures. Specifically, it presents the ten most significant post-hoc comparisons that differentiate between pairs of discourse tasks for each diagnostic group (LHD, HC, MCI, RHD, and TBI). The rankings are determined by the absolute t-ratio (|t-ratio|), where a higher value indicates a more powerful distinction between the two tasks for a specific linguistic variable. These results show that the choice of assessment task has a profound and measurable impact on a speaker's linguistic output. The data clearly shows that tasks are not interchangeable. Features that measure syntactic complexity are the most sensitive indicators of this variation, providing a powerful tool for understanding how different discourse demands reveal distinct linguistic abilities in both healthy individuals and those with neurological conditions. A detailed analysis of the most significant task contrasts reveals several key patterns.

Specifically, syntactic complexity is the primary differentiator. Across all five patient groups and the healthy controls, the single most sensitive linguistic feature was the Pronoun Type: Relative Ratio. This measure, which reflects the use of relative clauses (e.g., “the boy who is on the stool”), consistently showed the largest differences when comparing narrative tasks to descriptive tasks. This indicates that the demand to produce complex sentences is a primary factor that distinguishes one task from another. Other measures of syntactic complexity, such as Clausal modifier of a noun Count and Case marker Count, were also sensitive, particularly in the Healthy Control group.

Narrative and descriptive tasks elicit different structures. Narrative tasks inherently require speakers to connect events, characters, and ideas, leading to more complex sentence structures. Descriptive tasks, on the other hand, can often be completed with simpler, more direct

statements. Also, we see that Lexical Complexity Varies by Task. Beyond grammar, tasks also differed in the complexity of the vocabulary they elicited. The use of 5-syllable words was a significant differentiator between tasks for the LHD, RHD, and Healthy Control groups, suggesting that certain tasks prompt more sophisticated word choices.

An especially critical finding is that the strength of task-related distinctions depends on diagnostic group. The same types of linguistic features tended to be sensitive across groups, but their magnitude of separation varied. In the LHD and Healthy Control groups, the contrasts between tasks were especially sharp, reflected in very high t-ratios (many > 8.0) — i.e., tasks are more cleanly separated in feature space. In contrast, the MCI and RHD groups still showed task-related differences, but the separations were less pronounced, with lower t-ratios. This indicates that it's not just the features themselves, but the interaction between diagnosis and task that determines how strongly the clusters emerge.

5.1 Task and Diagnosis on Language measures

Our second question was how different neurological conditions reveal unique linguistic patterns within each task. To answer it, we conducted generalized mixed effects models. The complete analysis is provided in the supplementary tables. Here, we focus on the effects of the interaction between task and diagnosis on phonology, morphology, syntax, semantics, lexicon.

5.1.1 Phonology

Table 5 Phonological analysis results by syllable count and syllable structure patterns: results from the ANOVA output on the glmer models.

Variable	Task Effect	Diagnosis Effect	Interaction Effect
1 syllable word	***	n.s.	***
2 syllables word	***	***	***
3 syllables word	***	**	***
4 syllables word	***	n.s.	***
5 syllables word	***	n.s.	**
CCCV	***	n.s.	n.s.
CCCVC	***	*	n.s.
CCV	***	n.s.	***
CCVC	***	n.s.	***

CCVCC	***	n.s.	***
CCVCCC	***	n.s.	n.s.
CV	***	***	***
CVC	***	***	***
CVCC	***	***	***
CVCCC	***	***	***
V	***	n.s.	***
VC	***	n.s.	***
VCC	***	n.s.	***
Syllables	***	**	***
Total Characters in Text All symbols	***	*	***
Total Characters in Text Letters Only	***	**	***

*Note: Asterisks (**) indicate statistical significance levels, with more asterisks representing stronger significance (e.g., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). "n.s." indicates non-significant results. C = consonant, V = vowel in syllable structure notation.

The table presents statistical overall ANOVA output of the glmer models examining the effects of task (e.g., different language production or processing tasks like naming, repetition, or reading) and diagnosis on word lengths by syllable count, syllable structures (where C = consonant, V = vowel), total syllables, and total characters in produced text measures. The results highlight a consistent main effect of task across all variables, indicating that the nature of the task strongly influences phonological output regardless of diagnosis; a more variable main effect of diagnosis, suggesting that these neurological conditions differentially impact certain aspects of phonology but not others; and widespread interaction effects between task and diagnosis, implying that the influence of task on phonological measures varies depending on the specific diagnosis, e.g., one task might exacerbate phonological deficits in one group but not others. Interactions suggest non-additive effects, where certain diagnosis-task combinations amplify phonological vulnerabilities. This informs models of language breakdown, emphasizing how neurological damage interacts with contextual demands.

5.1.2 Morphology – Part of Speech Measures

The full results are shown in Appendix 1. Task was significant in all measures ($p < 0.5$) and the diagnosis was significant for most measures ($p < 0.5$), except for Conjunct Count, Coordinating conjunction Count, Expletive Count, Noun Ratio, Numeral Count, Numeral Ratio, Pronoun

Count, Pronoun Ratio, Proper noun Count (n.s.). The effect of the task x diagnosis interaction on Part of Speech Measures was significant in most distributions ($p < 0.5$), except for Expletive Counts.

5.1.3 Morphology – Inflectional Morphology

Across nearly all categories of inflectional morphology, task effects were highly significant: verb aspect, tense, mood, voice, form, modality, number, gender, definiteness, possession, case, comparison, degree, pronoun type, and polarity all varied systematically by the discourse task. This means that the linguistic context provided by the task strongly shapes how morphological features appear, even when diagnosis alone was not significant.

At the same time, diagnosis and its interaction with task modulated these effects. For example, groups differed in how they marked tense and aspect, produced passive forms, used complex verb forms, and handled pronouns, gender, and possession. These differences often appeared in the ratios (relative frequency of forms), highlighting that diagnostic groups vary not only in whether features are present but in how heavily they are relied upon relative to other structures. Taken together, this shows that task type is a powerful driver of inflectional morphology, while diagnosis shapes the strength and distribution of these effects.

5.1.4 Syntax

Syntactic measures represent the constructions of sentences and phrases providing information about their role in the syntactic hierarchy. For presentation purposes we break these measures down into Sentence Complexity and Structural Measures, Clauses, Sentences, Phrase Types and Units Counts and Ratios, and Grammatical Dependency Types (Modifiers and Complements) (Appendix 3).

Table 6 Sentence Complexity and Structural Measures results from the Anova output on the glmer models.

Variable	Task	Diagnosis	Interaction
Average Sentence Length	***	***	***
Mean Sentence Length in Words	***	***	***
Max Sentence Length	***	***	***
Min Sentence Length	***	***	***

Average Tree Height	***	***	***
Total Tree Height	***	n.s.	***
Max Classical Yngve Load	***	*	***
Mean Classical Yngve Load	***	***	***
Total Classical Yngve Load	***	n.s.	***
Max Left Branching Depth	***	n.s.	***
Mean Left Branching Depth	***	***	***
Passive Sentences Percent	***	***	***

*Note: Asterisks (**) indicate statistical significance levels, with more asterisks representing stronger significance (e.g., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

These features indicate robust task effects for structural complexity, with more complex syntax in certain task types. Diagnostic effects are subtler but emerge in measures related to syntactic embedding depth (Yngve Load, tree height) and passive constructions, suggesting some diagnostic groups may produce simpler or flatter syntactic trees or struggle with complex sentential structures.

Table 7 Clauses, Sentences, Phrase Types and Units Counts and Ratios results from the Anova output on the glmer models.

Variable	**Task**	**Diagnosis**	**Interaction**
Total Sentences	***	n.s.	***
Sentences Alphabetic Only	***	n.s.	***
Total T units	***	*	***
Total Complex T units	***	***	***
Total Clauses	***	*	***
Clauses Total Clauses in Text	***	n.s.	***
Total Dependent Clauses	***	***	***
Total Coordinate Phrases	***	n.s.	***
Total Complex Nominals	***	***	***
Adjective Phrases	***	***	***
Adverbial Phrases	***	*	***
Noun Phrases	***	*	***
Verb Phrases	***	n.s.	***
Prepositional Phrases	***	***	***

*Note: Asterisks (*) indicate statistical significance levels, with more asterisks representing stronger significance (e.g., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

There is strong evidence that syntactic packaging—how clauses and phrases are organized into T-units, sentences, and phrase structures—is shaped by both task demands and diagnostic group. Diagnostic effects emerge in clausal complexity (e.g., dependent clauses) and phrasal expansion (complex nominals), indicating linguistic planning and structural embedding may be constrained in some clinical groups. This set captures the microstructure of clause syntax—how dependents modify or complement heads. Diagnostic group differences are seen in ratio-based measures, which normalize for production length, revealing distinctions in how subjects, modifiers, and complements are structured. These patterns suggest certain diagnostic populations may underproduce or simplify syntactic dependencies, especially in relative and clausal constructions.

5.1.5 Semantics

Table 8 Semantic results from the Anova output on the glmer models.

Variable	Task	Diagnosis	Interaction
Cardinal Number Count	***	n.s.	***
Cardinal Number Ratio	***	n.s.	***
Date Count	***	n.s.	***
Date Ratio	***	***	***
Organization Count	***	***	***
Organization Ratio	***	***	***
Person Count	***	n.s.	***
Person Ratio	***	***	***

*Note: Asterisks (*) indicate statistical significance levels, with more asterisks representing stronger significance (e.g., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Like in the previous measures, the interaction between task and diagnosis results in significant effects for Cardinal Numbers, Dates, Organizations, and Persons, suggesting that semantics are affected by the interaction between Diagnosis and Task, even when the Diagnosis does not influence them independently.

5.1.6 Readability

Table 9 Readability results from the Anova output on the glmer models.

Variable	Task	Diagnosis	Interaction
Automated Readability Index	***	***	***
Coleman Liau Index	***	***	***
Dale Chall Readability Score	***	***	***
Difficult Words	***	n.s.	***
Estimated Reading Time sec	***	*	***
Flesch Kincaid Grade Level	***	***	***
Flesch Reading Ease	***	***	***
Gunning Fog Index	***	***	***
Linsear Write Formula	***	***	***
Smog Index	***	***	***

*Note: Asterisks (**) indicate statistical significance levels, with more asterisks representing stronger significance (e.g., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

The interaction between task and diagnosis resulted in significant effects for all Readability, suggesting that how difficult or easy a test depends on the task and the diagnosis.

6 Discussion

This research examines the intelligibility of patient-produced text and the influence of discourse task on its comprehensibility. A central question we address is the validity of comparing speech productions across different tasks to draw conclusions about underlying pathology. Consequently, we question the customary practice of using normative corpora without controlling for the confounding effect of the task itself. Currently, different tasks like the 'Cookie Theft' picture description task, other descriptive tasks, storytelling, and story retelling are often used interchangeably. However, it is of utmost importance to model how the different tasks affect language production before drawing results from patient and neurological conditions comparisons. This is because as it became evident by the study the comparison of the linguistic output from an individual to a group, or from one individual to another, without accounting for the specific task and even the specific stimulus used—for instance, the 'Cookie Theft' image versus a different narrative picture—creates a significant measurement problem. This is not only a methodological issue; it is a fundamental barrier to achieving comparable measures across diagnostic groups or measurement points. The typological differences between languages, such as the reliance on word order in English versus rich morphological systems in other languages,

mean that the grammatical features elicited by the same conceptual task can be fundamentally different. Furthermore, for the growing field of computational analysis, failing to control for task-induced variability can lead to the development of flawed algorithms and unreliable diagnostic markers.

A systematic investigation of these task effects is therefore essential for advancing the accuracy and validity of language assessment in neurology. This study advances beyond prior work that has typically relied on one or two measures per domain (e.g., mean length of utterance for syntax, type–token ratio for lexicon) by demonstrating that task–diagnosis interactions permeate a much wider set of linguistic features. By applying a computational pipeline to 290 features spanning phonology, morphology, syntax, semantics, lexicon, and readability, we show for the first time that discourse tasks generate distinct, reproducible "fingerprints" across multiple domains. This directly challenges the longstanding assumption that tasks can be collapsed into a single discourse category for analysis or clinical interpretation.

6.1 The Interaction of Task and Diagnosis in Linguistic Production

Our findings show significant interaction effects between task and diagnosis across all major linguistic domains. This indicates that the linguistic and cognitive profile of a neurological condition is not static (Aamodt et al., 2021; Cousins et al., 2021; Delaby et al., 2022; Khalil et al., 2024; Ljubenkov & et al., 2018; Mollenhaue & et al., 2020; Rohrer & et al., 2016); it is revealed differently depending on the cognitive and linguistic demands of the task at hand (Ash & et al., 2013; Bose & et al., 2021; Chapin & et al., 2022; Charles & et al., 2014; Cho & et al., 2021; Giannini & et al., 2017; Gorno-Tempini & et al., 2011).

First, one of the most critical domains of language impairment, phonology is typically related to left hemisphere frontotemporal damage (Best & et al., 2002; Goodglass & et al., 1997; Hickin & et al., 2002; Lorenz & Nickels, 2007; Meteyard & Bose, 2018; Van Hees & et al., 2013; Wambaugh, 2003; Wambaugh & et al., 2001). We found that both word length (by syllable count) and the complexity of syllable structures are significantly influenced by the interplay of task and diagnosis. More demanding tasks appear to amplify phonological deficits, revealing vulnerabilities in motor planning and articulatory control that might be less apparent in simpler contexts. For example, a patient with apraxia of speech might produce simple CV structures adequately in a straightforward task but show marked simplification of complex consonant

clusters (e.g., CCCV) in a more demanding narrative task. This suggests that a patient's phonological system can be selectively stressed by certain tasks, providing a more nuanced diagnostic picture.

Core morphology measures are critical for detecting conditions like agrammatism and anomia (Ahlsén & et al., 1996; Ballard & Thompson, 1999; Berndt & et al., 1996; Caramazza & Zurif, 1976; Duman & et al., 2011; Friedmann, 2002). Nevertheless, our findings show that the task and consequently the studies employing different tasks should not be directly compared as the distribution of parts-of-speech and the use of inflectional morphology were both extremely sensitive to the task-diagnosis interaction. Clusters or separations as was demonstrated in the unsupervised ML may indicate underlying similarities or differences in speech–language patterns that are relevant to clinical diagnosis or the monitoring of treatment progress. In Figure 2, the arrangement of points—from left to right—appears to follow the degree of contextual constraint in the tasks: highly structured activities such as picture description (e.g., Cookie Theft) are followed by instruction-based tasks (e.g., sandwich-making), then by conversational exchanges (e.g., spontaneous speech), and finally by open-ended narratives (e.g., accounts of brain injury, recovery, or illness). From a clinical perspective, this left-to-right progression may reflect increasing cognitive–linguistic demands: structured tasks provide clear visual or procedural cues, reducing planning and lexical retrieval load, whereas unstructured narratives require greater topic generation, organization, and memory retrieval (Afthinos et al., 2022).

These findings can facilitate the interpretation of patient performance patterns and identify task types that are most sensitive to subtle impairments. Specifically, the significant interaction effects on inflectional categories—such as verb tense and aspect, noun number, and case marking—are particularly revealing. These features are the building blocks of grammatical cohesion and narrative structure. A task requiring storytelling (e.g., “Cinderella” retelling) will inherently demand more complex temporal marking (past tense, perfective aspect) than a static picture description. The failure to deploy these features under such task demands can signal a core deficit in grammatical morphology, a hallmark of conditions like agrammatic aphasia.

Syntactic complexity indicates language processing and consequently the effects of neurological conditions on language processing (Agmon & et al., 2024; Cheung & Kemper, 1992; Kemper,

1987; Kyle & Crossley, 2017; Lan & et al., 2022). Our findings prove that syntactic complexity is a powerful differentiator, with nearly all measures showing significant interaction effects. Measures of sentence length, clausal embedding (e.g., dependent clauses), and hierarchical structure (e.g., tree height) all varied based on the specific pairing of task and diagnosis. This supports the hypothesis that tasks with higher narrative or procedural demands force the production of more complex syntax, thereby revealing latent deficits in sentence planning and construction. For instance, the reduced production of complex nominals or dependent clauses in a narrative task by a patient with TBI may reflect impaired executive functions impacting linguistic planning, a different underlying cause than the syntactic simplification seen in aphasia.

Specifically, to semantics, the ability to retrieve and use specific semantic content, as measured by named-entity recognition, was also significantly affected by the task-diagnosis interaction (Antonucci, 2009; Boyle, 2010; Boyle & Coelho, 1995; Coelho & Boyle, 2000; Efstratiadou & et al., 2018; Gravier & et al., 2018; Hashimoto & Frome, 2011; Kiran & Thompson, 2003; Wallace & Wood, 2013; Wambaugh & et al., 2013). Even when a diagnosis alone did not have a main effect, its combination with a specific task often did. This suggests that a patient's ability to access semantic categories like persons, organizations, or dates is not static but is modulated by the task's contextual demands. A personal event narrative, for example, directly probes the recall of temporal and personal information, making it a more sensitive tool for detecting impaired temporal orientation or anomia for proper names than a generic picture description.

Readability metrics, which synthesize features like word choice, sentence length, and syntactic complexity, offer a holistic measure of a text's accessibility. The analysis showed that the ease of understanding a patient's language is highly dependent on the task they are performing. A patient might produce simpler, more "readable" text in one task due to syntactic simplification, while producing more complex but disorganized and error-prone (and thus less readable) text in another. This finding is critical, as it demonstrates that a single readability score is insufficient for characterizing a patient's communicative effectiveness; the context of the elicitation task is paramount.

In sum, from a clinical perspective, these findings have direct implications for assessment and treatment planning. Narrative tasks were most sensitive to deficits in syntactic complexity,

picture descriptions highlighted lexical retrieval demands, and procedural tasks emphasized morphosyntactic simplification. Thus, each genre provides complementary information: a picture description may be optimal for rapidly probing word-finding, whereas a personal narrative better captures global coherence and pragmatic breakdowns. Developing task-specific normative data will allow clinicians to select the probe that best aligns with their diagnostic question and to track treatment-related changes with greater precision. In practice, this moves assessment toward a modular, task-matched model rather than relying on a single "one-size-fits-all" probe. Importantly, the use of an automated computational pipeline makes this level of task-sensitive profiling feasible at scale. Manual transcription and coding previously restricted clinical protocols to one or two elicitation tasks, such as Cookie Theft. Automation reduces the burden of time and training, enabling clinicians and researchers to incorporate multiple tasks into routine assessment without sacrificing efficiency. In this way, computational methods transform the theoretical insight that "tasks are not interchangeable" into a practical, implementable clinical workflow.

Although the present analyses grouped participants by broad diagnostic categories (LHD, RHD, TBI, MCI, dementia), it is important to note that substantial heterogeneity exists within each group. For example, Broca's versus Wernicke's aphasia, amnestic versus non-amnestic MCI, and focal versus diffuse TBI all yield different discourse phenotypes. Our framework provides a foundation that can be extended to more narrowly defined subgroups, allowing for identification of subtype-specific task effects. This is a critical step toward precision assessment, where both the diagnosis and its subtype inform the choice of discourse probe.

6.2 Generalizing beyond the tasks studied

These findings and earlier research lead to the assumption that the results have consequences that extend beyond the tasks and potentially the populations studied in this work including affecting the acoustic and linguistic measures (Themistocleous, 2016, 2017, 2019). For example, earlier research has examined written and spoken language production in various populations, including individuals with post-stroke aphasia (e.g., writing; (Mortensen, 2005)), traumatic brain injury (TBI; e.g., writing; (Wilson & Proctor, 2002)), Alzheimer's disease (writing and speech; (Croisile et al., 1996)), non-fluent progressive aphasia (writing and speech; Graham et al., 2004),

and even healthy young adults (writing and speech; (Behrns et al., 2009; Croisile et al., 1996; Ulatowska et al., 1983)).

6.3 Limitations and Future Directions

While this study's scale and computational depth provide novel insights, several limitations must be acknowledged, which in turn pave the way for future research.

First, the analysis relies on aggregating large, publicly available clinical corpora. A significant strength of this approach is its ecological validity and the large sample size; however, it also introduces limitations. The dataset is inherently unbalanced, as not all participants completed every task. Although our use of mixed-effects models is statistically robust for managing such missing data, the statistical power to detect interaction effects may be lower for specific task-and-diagnosis combinations where data is sparse. Furthermore, combining data from various sources (e.g., AphasiaBank, DementiaBank) means there could be subtle, unmeasured variations in data collection protocols, recording quality, or examiner instructions across the original study sites that contribute to noise in the data.

Second, the diagnostic categories used in this study (e.g., LHD, TBI, Dementia) are broad and encompass significant internal heterogeneity. For instance, the LHD group includes individuals with varying aphasia subtypes (e.g., Broca's, Wernicke's, anomic), each with a distinct linguistic profile. Our analysis successfully identifies overarching patterns at the group level, but it may mask more granular, subtype-specific interactions between task and diagnosis. Future research could apply this framework to more narrowly defined clinical subgroups to uncover finer-grained distinctions.

Third, our findings are based on data from English-speaking participants. The specific linguistic features that are most sensitive to the task-diagnosis interaction (e.g., reliance on word order, specific inflectional morphemes) are tied to the typological structure of English. The broader principle—that task demands interact with underlying deficits—is likely universal, but the specific "linguistic fingerprints" will almost certainly differ in other languages, particularly those with richer morphological systems or different syntactic rules. Cross-linguistic replication

of this study is a critical next step to establishing a more universal model of task effects in clinical assessment.

Fourth, our analysis is contingent on an automated computational pipeline. While OpenBrainAI enables a large-scale analysis that would be impossible to perform manually, no NLP tool is perfect. The accuracy of part-of-speech tagging, dependency parsing, and semantic entity recognition can be lower for atypical, error-prone, or disfluent speech common in neurogenic disorders. These potential inaccuracies could influence the precise values of the linguistic measures, though the large-scale patterns observed are likely robust. Future work should include targeted validation of the platform's performance on diverse clinical speech samples.

Looking ahead, these limitations highlight clear future directions. The clinical takeaway that assessment tasks are not interchangeable is paramount. Future research should focus on developing task-specific normative databases, allowing for more precise and valid comparisons. Clinicians can use the framework presented here to make hypothesis-driven choices about assessment, selecting tasks strategically to probe specific linguistic abilities. For example, a picture description is ideal for a quick lexical assessment, whereas a personal narrative is a superior tool for evaluating high-level discourse organization and syntactic complexity. By embracing the interaction between task and diagnosis, we can refine our diagnostic tools, set more targeted therapeutic goals, and ultimately deepen our understanding of the multifaceted nature of neurogenic language disorders.

Building on these principles, three translational priorities emerge. First, cross-linguistic validation is essential to determine whether the task–diagnosis “fingerprints” identified here generalize to languages with richer morphology or different syntactic structures. Second, linking linguistic profiling with neuroimaging (e.g., lesion–symptom mapping, functional connectivity) will help clarify mechanistic underpinnings of task effects and support biomarker development. Third, there is a pressing need to translate these findings into implementable clinical tools. This includes developing a modernized library of discourse tasks that are co-designed with patients and clinicians, systematically tiered by genre and complexity, and validated for psychometric rigor. Parallel efforts should create streamlined, clinician-informed scoring frameworks that

capture both linguistic and pragmatic features but can be completed efficiently in routine practice.

Appendix 1. Morphological analysis results from the Anova output on the glmer models.

Variable	Task	Diagnosis	Interaction
Adjective Count	***	***	***
Adjective Ratio	***	***	*
Adposition Count	***	***	***
Adposition Ratio	***	***	***
Adverb Count	***	*	***
Adverb Ratio	***	***	***
Auxiliary Count	***	**	***
Auxiliary Ratio	***	***	***
Conjunct Count	***	n.s.	***
Conjunct Ratio	***	***	***
Coordinating conjunction Count	***	n.s.	***
Coordinating conjunction Ratio	***	***	***
Determiner Count	***	**	***
Determiner Ratio	***	***	***
Expletive Count	***	n.s.	n.s.
Expletive Ratio	***	***	***
Interjection Count	***	***	***
Interjection Ratio	***	***	***
Noun Count	***	**	***
Noun Ratio	***	n.s.	***
Numeral Count	***	n.s.	**
Numeral Ratio	***	n.s.	***
Other Count	***	***	***
Other Ratio	***	***	***
Particle Count	***	***	***
Particle Ratio	***	***	***
Pronoun Count	***	n.s.	***
Pronoun Ratio	***	n.s.	***
Proper noun Count	***	n.s.	***
Proper noun Ratio	***	***	***
Subordinating conjunction Count	***	***	***
Subordinating conjunction Ratio	***	***	***
Verb Count	***	***	***
Verb Ratio	***	***	***

*Note: Asterisks (*) indicate statistical significance levels, with more asterisks representing stronger significance (e.g., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Appendix 2. Inflectional morphology results from the Anova output on the glmer models

Variable	Task	Diagn.	Inter.	Variable	Task	Diagn.	Inter
Aspect Perfective Count	***	*	***	Number Singular Ratio	***	***	***
Aspect Perfective Ratio	***	***	***	NumType Cardinal Count	***	n.s.	n.s.
Aspect Progressive Count	***	***	***	NumType Cardinal Ratio	***	**	n.s.
Aspect Progressive Ratio	***	n.s.	***	Person First Count	***	**	***
Auxiliary passive Count	***	n.s.	***	Person First Ratio	***	***	***
Auxiliary passive Ratio	***	***	***	Person Second Count	***	n.s.	**
Case Accusative Count	***	n.s.	***	Person Second Ratio	***	n.s.	***
Case Accusative Ratio	***	*	***	Person Third Count	***	n.s.	***
Case marker Count	***	n.s.	***	Person Third Ratio	***	***	***
Case marker Ratio	***	***	***	Polarity Negative Count	***	n.s.	n.s.
Case Nominative Count	***	n.s.	***	Polarity Negative Ratio	***	**	n.s.
Case Nominative Ratio	***	***	***	Poss Possessive Count	***	***	***
ConjType Cmp Count	***	n.s.	***	Poss Possessive Ratio	***	***	***
ConjType Cmp Ratio	***	*	***	PronType Article Count	***	n.s.	***
Dative Count	***	n.s.	n.s.	PronType Article Ratio	***	n.s.	***
Dative Ratio	***	n.s.	***	PronType Demonstrative Count	***	n.s.	***
Definite Definite Count	***	n.s.	***	PronType Demonstrative Ratio	***	***	***
Definite Definite Ratio	***	***	***	PronType Indefinite Count	***	n.s.	n.s.
Definite Indefinite Count	***	*	***	PronType Indefinite Ratio	***	n.s.	n.s.
Definite Indefinite Ratio	***	n.s.	***	PronType Personal Count	***	*	***
Degree Cmp Count	***	n.s.	n.s.	PronType Personal Ratio	***	n.s.	***
Degree Cmp Ratio	***	n.s.	n.s.	PronType Relative Count	***	n.s.	n.s.
Degree Positive Count	***	**	***	PronType Relative Ratio	***	*	*
Degree Positive Ratio	***	n.s.	n.s.	Tense Past Count	***	**	***
Gender Feminine Count	***	n.s.	***	Tense Past Ratio	***	***	***
Gender Feminine Ratio	***	n.s.	***	Tense Present Count	***	n.s.	***
Gender Masculine Count	***	**	***	Tense Present Ratio	***	**	***
Gender Masculine Ratio	***	**	***	VerbForm Finite Count	***	n.s.	***

Gender Neuter Count	***	n.s.	***	VerbForm Finite Ratio	***	n.s.	***
Gender Neuter Ratio	***	***	***	VerbForm Infinitive Count	***	*	***
Mood Indicative Count	***	n.s.	***	VerbForm Infinitive Ratio	***	**	***
Mood Indicative Ratio	***	***	***	VerbForm Participle Count	***	***	***
Number Plural Count	***	n.s.	***	VerbForm Participle Ratio	***	***	***
Number Plural Ratio	***	***	***	VerbType Mod Count	***	n.s.	n.s.
Number Singular Count	***	*	***	VerbType Mod Ratio	***	n.s.	n.s.

*Note: Asterisks (**) indicate statistical significance levels, with more asterisks representing stronger significance (e.g., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Appendix 3. Grammatical Dependency Types (Modifiers and Complements) results from the Anova output on the glmer models.

Variable	Task	Diagnosis	Interaction
Adjectival complement Count	***	n.s.	***
Adjectival complement Ratio	***	***	***
Adjectival modifier Count	***	***	***
Adjectival modifier Ratio	***	***	***
Adverbial clause modifier Count	***	**	***
Adverbial clause modifier Ratio	***	***	***
Adverbial modifier Count	***	*	***
Adverbial modifier Ratio	***	***	***
Appositional modifier Count	***	n.s.	***
Appositional modifier Ratio	***	***	n.s.
Attribute Count	***	n.s.	***
Attribute Ratio	***	***	***
Clausal complement Count	***	n.s.	***
Clausal complement Ratio	***	*	***
Clausal modifier of noun Count	***	n.s.	*
Clausal modifier of noun Ratio	***	***	***
Complement of preposition Count	***	n.s.	n.s.
Complement of preposition Ratio	***	***	***
Compound modifier Count	***	n.s.	***
Compound modifier Ratio	***	***	***

Direct object Count	***	***	***
Direct object Ratio	***	***	***
Marker Count	***	n.s.	***
Marker Ratio	***	***	***
Negation modifier Count	***	n.s.	n.s.
Negation modifier Ratio	***	n.s.	n.s.
Nominal subject Count	***	n.s.	***
Nominal subject Ratio	***	**	***
Nominal subject passive Count	***	n.s.	***
Nominal subject passive Ratio	***	***	***
Number modifier Count	***	*	n.s.
Number modifier Ratio	***	n.s.	***
Object of preposition Count	***	***	***
Object of preposition Ratio	***	***	***
Open clausal complement Count	***	***	***
Open clausal complement Ratio	***	***	***
Possession modifier Count	***	***	***
Possession modifier Ratio	***	***	***
Prepositional modifier Count	***	***	***
Prepositional modifier Ratio	***	***	***
Relative clause modifier Count	***	n.s.	***
Relative clause modifier Ratio	***	*	***
Root Count	***	n.s.	***
Root Ratio	***	***	***
Unclassified dependent Count	***	n.s.	**
Unclassified dependent Ratio	***	***	n.s.

*Note: Asterisks (*) indicate statistical significance levels, with more asterisks representing stronger significance (e.g., * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

7 References

Aamodt, W. W., Waligorska, T., Shen, J., Tropea, T. F., Siderowf, A., Weintraub, D., Grossman, M., Irwin, D., Wolk, D. A., Xie, S. X., Trojanowski, J. Q., Shaw, L. M., & Chen-Plotkin, A. S. (2021). Neurofilament light chain as a biomarker for cognitive decline in Parkinson Disease. *Movement Disorders*, *36*(12), 2945-2950. https://doi.org/10.1002/mds.28779

Afthinos, A., Themistocleous, C., Herrmann, O., Fan, H., Lu, H., & Tsapkini, K. (2022). The Contribution of Working Memory Areas to Verbal Learning and Recall in Primary Progressive Aphasia. *Frontiers in Neurology*, *13*, 1-11. https://doi.org/10.3389/fneur.2022.698200

Agmon, G., & et al. (2024). Automated measures of syntactic complexity in natural speech production: Older and younger adults as a case study. *Journal of Speech, Language, & Hearing Research*, *67*(2), 545-561. https://doi.org/10.1044/2023_JSLHR-23-00009

Ahlsén, E., & et al. (1996). Noun phrase production by agrammatic patients: A cross-linguistic approach. *Aphasiology*, *10*(6), 543-559. https://doi.org/10.1080/02687039608248436

Antonucci, S. M. (2009). Use of semantic feature analysis in group aphasia treatment. *Aphasiology*, *23*(7-8), 854-866. https://doi.org/10.1080/02687030802634405

Ash, S., & et al. (2013). Differentiating primary progressive aphasias in a brief sample of connected speech. *Neurology*, *81*(4), 329-336. https://doi.org/10.1212/WNL.0b013e31829c5d0e

Badecker, W., Hillis, A., & Caramazza, A. (1990). Lexical morphology and its role in the writing process: evidence from a case of acquired dysgraphia. *Cognition*, *35*(3), 205--243.

Ballard, K. J., & Thompson, C. K. (1999). Treatment and generalization of complex sentence production in agrammatism. *Journal of Speech, Language, & Hearing Research*, *42*(3), 690-707. https://doi.org/10.1044/jslhr.4203.690

Barbieri, Z., Fernández, M. A., Newbury, D. F., & Villanueva, P. (2018). Family aggregation of language impairment in an isolated Chilean population from Robinson Crusoe Island. *Int J Lang Commun Disord*, *53*(3), 643-655. https://doi.org/10.1111/1460-6984.12377

Bastiaanse, R. (2013). Why reference to the past is difficult for agrammatic speakers [Conference Paper]. *Clinical Linguistics & Phonetics*, *27*(4), 244-263. https://doi.org/10.3109/02699206.2012.751626

Behrns, I., Wengelin, Å., Broberg, M., & Hartelius, L. (2009). A comparison between written and spoken narratives in aphasia. *Clinical Linguistics & Phonetics*, *23*(7), 507-528.

Berndt, R. S., & et al. (1996). Comprehension of reversible sentences in 'agrammatism': A meta-analysis. *Cognition*, *58*(3), 289-308. https://doi.org/10.1016/0010-0277(95)00682-6

Best, W., & et al. (2002). Phonological and orthographic facilitation of word-retrieval in aphasia: Immediate and delayed effects. *Aphasiology*, *16*(1–2), 151–168. https://doi.org/10.1080/02687040143000483

Bose, A., & et al. (2021). Connected speech characteristics of Bengali speakers with Alzheimer's disease: Evidence for language-specific diagnostic markers. *Frontiers in Aging Neuroscience*, *13*, 707628. https://doi.org/10.3389/fnagi.2021.707628

Boyle, M. (2010). Semantic feature analysis treatment for aphasic word retrieval impairments: What's in a name? *Topics in Stroke Rehabilitation*, *17*(6), 411-422. https://doi.org/10.1310/tsr1706-411

Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology*, *4*(4), 94-98. https://doi.org/10.1044/1058-0360.0404.94

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

Callegari, E., & et al. (2024). Automatic extraction of language-specific biomarkers of healthy aging in Icelandic.

Caramazza, A., & Hillis, A. E. (1989). The disruption of sentence production: some dissociations. *Brain and Language*, *36*(4), 625--650.

Caramazza, A., & Hillis, A. E. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, *349*(6312), 788--790.

Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, *3*(4), 572-582. https://doi.org/10.1016/0093-934X(76)90048-1

Chapin, K., & et al. (2022). A finer-grained linguistic profile of Alzheimer's disease and Mild Cognitive Impairment. *Journal of Neurolinguistics*, *63*, 101069. https://doi.org/10.1016/j.jneuroling.2022.101069

Charles, D., & et al. (2014). Grammatical comprehension deficits in non-fluent/agrammatic primary progressive aphasia. *Journal of Neurology, Neurosurgery & Psychiatry*, *85*(3), 249-256. https://doi.org/10.1136/jnnp-2013-305749

Charniak, E., & Johnson, M. (2005). Coarse-to-fine n-best parsing and MaxEnt discriminative reranking.

Cheung, H., & Kemper, S. (1992). Competing complexity metrics and adults' production of complex sentences. *Applied Psycholinguistics*, *13*(1), 53-76. https://doi.org/10.1017/S0142716400005427

Cho, S., & et al. (2021). Automated analysis of lexical features in frontotemporal degeneration. *Cortex*, *137*, 215-231. https://doi.org/10.1016/j.cortex.2021.01.012

Coelho, C. A. (2007). Management of discourse deficits following traumatic brain injury: Progress, caveats, and needs. Seminars in Speech and Language,

Coelho, C. A. M. R. E., & Boyle, M. (2000). Semantic feature analysis as a treatment for aphasic dysnomia: A replication. *Aphasiology*, *14*(2), 133-142. https://doi.org/10.1080/026870300401513

Cousins, K. A. Q., Phillips, J. S., Irwin, D. J., Lee, E. B., Wolk, D. A., Shaw, L. M., Zetterberg, H., Blennow, K., Burke, S. E., Kinney, N. G., Gibbons, G. S., McMillan, C. T., Trojanowski, J. Q., & Grossman, M. (2021). ATN incorporating cerebrospinal fluid neurofilament light chain detects frontotemporal lobar degeneration. *Alzheimer's & Dementia*, *17*(5), 822-830. https://doi.org/10.1002/alz.12233

Croisile, B., Ska, B., Brabant, M.-J., Duchene, A., Lepage, Y., Aimard, G., & Trillet, M. (1996). Comparative study of oral and written picture description in patients with Alzheimer's disease. *Brain and Language*, *53*(1), 1-19. https://www.sciencedirect.com/science/article/abs/pii/S0093934X96900334?via%3Dihub

Croot, K., Hodges, J. R., Xuereb, J., & Patterson, K. (2000). Phonological and Articulatory Impairment in Alzheimer's Disease: A Case Series. *Brain and Language*, *75*(2), 277-309. https://doi.org/https://doi.org/10.1006/brln.2000.2357

Cruice, M., Botting, N., Marshall, J., Boyle, M., Hersh, D., Pritchard, M., & Dipper, L. (2020). UK speech and language therapists' views and reported practices of discourse analysis in aphasia rehabilitation. *International journal of language & communication disorders*, *55*(3), 417-442.

Dale, E., & Chall, J. S. (1948). A formula for predicting readability: Instructions. *Educational research bulletin*, 37-54.

Delaby, C., Bousiges, O., Bouvier, D., Fillée, C., Fourier, A., Mondésert, E., Nezry, N., Omar, S., Quadrio, I., Rucheton, B., Schraen-Maschke, S., van Pesch, V., Vicca, S., Lehmann, S., & Bedel, A. (2022). Neurofilaments contribution in clinic: State of the art. *Frontiers in Aging Neuroscience*, *14*, 1265. https://doi.org/10.3389/fnagi.2022.1034684

Devlin, J., & et al. (2018). Pre-training of deep bidirectional transformers for language understanding. *arXiv*. https://doi.org/10.48550/arXiv.1810.04805

Duman, T. Y., & et al. (2011). Sentence comprehension in Turkish Broca's aphasia: An integration problem. *Aphasiology*, *25*(8), 908-926. https://doi.org/10.1080/02687038.2010.550629

Efstratiadou, E. A., & et al. (2018). A systematic review of semantic feature analysis therapy studies for aphasia. *Journal of Speech, Language, & Hearing Research*, *61*(5), 1261-1278. https://doi.org/10.1044/2018_JSLHR-L-16-0330

Elbourn, E., Kenny, B., Power, E., & Togher, L. (2019). Psychosocial Outcomes of Severe Traumatic Brain Injury in Relation to Discourse Recovery: A Longitudinal Study up to 1 Year Post-Injury. *American Journal of Speech-Language Pathology*, *28*(4), 1463-1478. https://doi.org/doi:10.1044/2019_AJSLP-18-0204

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

Fitzsimmons, P. R., Michael, B. D., Hulley, J. L., & Scott, G. O. (2010). A readability assessment of online Parkinson's disease information. *J R Coll Physicians Edinb*, *40*(4), 292-296. https://doi.org/10.4997/JRCPE.2010.401

Fraser, K. C., & et al. (2015). Sentence segmentation of aphasic speech.

Fridriksson, J., den Ouden, D. B., Hillis, A. E., Hickok, G., Rorden, C., Basilakos, A., & Bonilha, L. (2018). Anatomy of aphasia revisited. *Brain*, *141*(3), 848-862. https://doi.org/10.1093/brain/awx363

Friedmann, N. (2002). Question production in agrammatism: The tree pruning hypothesis. *Brain and Language*, *80*(2), 160-187. https://doi.org/10.1006/brln.2001.2587

Ghojogh, B., Crowley, M., Karray, F., & Ghodsi, A. (2023). Uniform Manifold Approximation and Projection (UMAP). In B. Ghojogh, M. Crowley, F. Karray, & A. Ghodsi (Eds.), *Elements of Dimensionality Reduction and Manifold Learning* (pp. 479-497). Springer International Publishing. https://doi.org/10.1007/978-3-031-10602-6_17

Giannini, L. A. A., & et al. (2017). Clinical marker for Alzheimer disease pathology in logopenic primary progressive aphasia. *Neurology*, *88*(24), 2276-2284. https://doi.org/10.1212/wnl.0000000000004034

Goldstein, A., & et al. (2022). Shared computational principles for language processing in humans and deep language models. *Nature Neuroscience*, *25*(3), 369-380. https://doi.org/10.1038/s41593-022-01026-4

Goodglass, H., & et al. (1997). The importance of word-initial phonology: Error patterns in prolonged naming efforts by aphasic patients. *Journal of the International Neuropsychological Society*, *3*(2), 128–138. https://doi.org/10.1017/S1355617797001288

Gorno-Tempini, M. L., & et al. (2011). Classification of primary progressive aphasia and its variants. *Neurology*, *76*(11), 1006-1014. https://doi.org/10.1212/wnl.0b013e31821103e6

Gravier, M. L., & et al. (2018). What matters in semantic feature analysis: Practice-related predictors of treatment response in aphasia. *American Journal of Speech-Language Pathology*, *27*(1S), 438-453. https://doi.org/10.1044/2017_AJSLP-16-0196

Hartig, F. (2016). DHARMa: residual diagnostics for hierarchical (multi-level/mixed) regression models. *CRAN: Contributed Packages*.

Hashimoto, N., & Frome, A. (2011). The use of a modified semantic features analysis approach in aphasia. *Journal of Communication Disorders*, *44*(4), 459-469. https://doi.org/10.1016/j.jcomdis.2011.02.004

Hickin, J., & et al. (2002). Phonological therapy for word-finding difficulties: A re-evaluation. *Aphasiology*, *16*(10–11), 981–999. https://doi.org/10.1080/02687030244000509

Hillis, A. E. (1989). Efficacy and generalization of treatment for aphasic naming errors. *Archives of Physical Medicine and Rehabilitation*, *70*(8), 632--636.

Hillis, A. E., Beh, Y. Y., Sebastian, R., Breining, B., Tippett, D. C., Wright, A., Saxena, S., Rorden, C., Bonilha, L., & Basilakos, A. (2018). Predicting Recovery in Acute Post-stroke Aphasia. *Annals of Neurology*.

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.

Kemper, S. (1987). Life-span changes in syntactic complexity. *Journal of Gerontology*, *42*(3), 323-328. https://doi.org/10.1093/geronj/42.3.323

Khalil, M., Teunissen, C. E., Lehmann, S., Otto, M., Piehl, F., Ziemssen, T., Bittner, S., Sormani, M. P., Gattringer, T., Abu-Rumeileh, S., Thebault, S., Abdelhak, A., Green, A., Benkert, P., Kappos, L., Comabella, M., Tumani, H., Freedman, M. S., & Leppert, D. (2024). Neurofilaments as biomarkers in neurological disorders — towards clinical application. *Nature Reviews Neurology*, *20*(5), 269-287. https://doi.org/10.1038/s41582-024-00955-x

Kim, H., Berube, S., & Hillis, A. E. (2023). Core lexicon in aphasia: A longitudinal study. *Aphasiology*, *37*(10), 1679-1691. https://doi.org/10.1080/02687038.2022.2121598

Kim, H., Hillis, A. E., & Themistocleous, C. (2024). Machine Learning Classification of Patients with Amnestic Mild Cognitive Impairment and Non-Amnestic Mild Cognitive Impairment from Written Picture Description Tasks. *Brain Sciences*, *14*(7).

Kiran, S., & Thompson, C. K. (2003). The role of semantic complexity in treatment of naming deficits. *Journal of Speech, Language, & Hearing Research*, *46*(4), 773-787. https://doi.org/10.1044/1092-4388(2003/061)

Klare, G. R. (1974). Assessing Readability. *Reading Research Quarterly*, *10*(1), 62-102. https://doi.org/10.2307/747086

Koller, M. (2016). robustlmm: an R package for robust estimation of linear mixed-effects models. *Journal of Statistical Software*, *75*, 1-24.

Kuznetsova, A., Bruun Brockhoff, P., & Haubo Bojesen Christensen, R. (2016). *lmerTest: Tests in Linear Mixed Effects Models*. In R Foundation for Statistical Computing. https://CRAN.R-project.org/package=lmerTest

Kyle, K., & Crossley, S. (2017). Assessing syntactic sophistication in L2 writing: A usage-based approach. *Language Testing*, *34*(4), 513-535. https://doi.org/10.1177/0265532217712554

Lacey, E. H., Skipper-Kallal, L. M., Xing, S., Fama, M. E., & Turkeltaub, P. E. (2017). Mapping common aphasia assessments to underlying cognitive processes and their neural substrates. *Neurorehabilitation and Neural Repair*, *31*(5), 442-450. https://pmc.ncbi.nlm.nih.gov/articles/PMC5393922/

Lan, G., & et al. (2022). A corpus-based investigation on noun phrase complexity in L1 and L2 English writing. *English for Specific Purposes*, *67*, 4-17. https://doi.org/10.1016/j.esp.2022.02.002

Lanzi Alyssa, M., Saylor Anna, K., Fromm, D., Liu, H., MacWhinney, B., & Cohen Matthew, L. (2023). DementiaBank: Theoretical Rationale, Protocol, and Illustrative Analyses. *American Journal of Speech-Language Pathology*, *32*(2), 426-438. https://doi.org/10.1044/2022_AJSLP-22-00281

Ljubenkov, P. A., & et al. (2018). Cerebrospinal fluid biomarkers predict frontotemporal dementia trajectory. *Annals of Clinical & Translational Neurology*, *5*(10), 1250-1263. https://doi.org/10.1002/acn3.643

Lorenz, A., & Nickels, L. (2007). Orthographic cueing in anomic aphasia: How does it work? *Aphasiology*, *21*(6–8), 670–686. https://doi.org/10.1080/02687030701192182

Mack, J. E., Barbieri, E., Weintraub, S., Mesulam, M. M., & Thompson, C. K. (2021). Quantifying grammatical impairments in primary progressive aphasia: Structured language tests and narrative language production. *Neuropsychologia*, *151*, 107713. https://doi.org/10.1016/j.neuropsychologia.2020.107713

MacWhinney, B. (2025). Understanding Language Through TalkBank. *Current Directions in Psychological Science*, 09637214241304345.

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

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. *Dement Geriatr Cogn Dis Extra*, *7*(1), 52-73. https://doi.org/10.1159/000456710

Meteyard, L., & Bose, A. (2018). What does a cue do? Comparing phonological and semantic cues for picture naming in aphasia. *Journal of Speech, Language, & Hearing Research*, *61*(3), 658–674. https://doi.org/10.1044/2017_JSLHR-L-17-0214

Minga, J., Stockbridge Melissa, D., Durfee, A., & Johnson, M. (2022). Clinical Guidelines for Eliciting Discourse Using the RHDBank Protocol. *American Journal of Speech-Language Pathology*, *31*(5), 1949-1962. https://doi.org/10.1044/2022_AJSLP-22-00097

Mollenhaue, B., & et al. (2020). Validation of serum neurofilament light chain as a biomarker of Parkinson's disease progression. *Movement Disorders*, *35*(11), 1999-2008. https://doi.org/10.1002/mds.28206

Mortensen, L. (2005). Written discourse and acquired brain impairment: Evaluation of structural and semantic features of personal letters from a Systemic Functional Linguistic perspective. *Clinical Linguistics & Phonetics*, *19*(3), 227-247. https://doi.org/10.1080/02699200410001698652

Nivre, J., de Marneffe, M.-C., Ginter, F., Hajič, J., Manning, C. D., Pyysalo, S., Schuster, S., Tyers, F., & Zeman, D. (2020, May). Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis, *Proceedings of the Twelfth Language Resources and Evaluation Conference* Marseille, France.

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. In R Foundation for Statistical Computing. https://www.R-project.org/

Richardson, J. D., Dalton, S. G., Greenslade, K. J., Jacks, A., Haley, K. L., & Adams, J. (2021). Main Concept, Sequencing, and Story Grammar Analyses of Cinderella Narratives in a Large Sample of Persons with Aphasia. *Brain Sciences*, *11*(1), 110. https://www.mdpi.com/2076-3425/11/1/110

https://mdpi-res.com/d_attachment/brainsci/brainsci-11-00110/article_deploy/brainsci-11-00110.pdf?version=1610711274

Rohrer, J. D., & et al. (2016). Serum neurofilament light chain protein is a measure of disease intensity in frontotemporal dementia. *Neurology*, *87*(13), 1329-1336. https://doi.org/10.1212/WNL.0000000000003154

Russell, L. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. In R Foundation for Statistical Computing. https://CRAN.R-project.org/package=emmeans

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

Stark, B. C., Dalton, S. G., & Lanzi, A. M. (2025). Access to context-specific lexical-semantic information during discourse tasks differentiates speakers with latent aphasia, mild cognitive impairment, and cognitively healthy adults [Original Research]. *Frontiers in Human Neuroscience*, *Volume 18 - 2024*. https://doi.org/10.3389/fnhum.2024.1500735

Stark, B. C., & Fukuyama, J. (2021). Leveraging big data to understand the interaction of task and language during monologic spoken discourse in speakers with and without aphasia. *Language, Cognition and Neuroscience*, *36*(5), 562-585. https://doi.org/10.1080/23273798.2020.1862258

Stark Brielle, 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

Steel, J., Ferguson, A., Spencer, E., & Togher, L. (2017). Language and cognitive communication disorder during post-traumatic amnesia: Profiles of recovery after TBI from three cases. *Brain Injury*, *31*(13-14), 1889-1902. https://doi.org/10.1080/02699052.2017.1373200

Stockbridge, M. D., Matchin, W., Walker, A., Breining, B., Fridriksson, J., Hickok, G., & Hillis, A. E. (2021). One cat, Two cats, Red cat, Blue cats: Eliciting morphemes from individuals with primary progressive aphasia. *Aphasiology*, *35*(12), 1-12. https://doi.org/10.1080/02687038.2020.1852167

Themistocleous, C. (2016). The bursts of stops can convey dialectal information. *The Journal of the Acoustical Society of America*, *140*(4), EL334-EL339. https://doi.org/doi:http://dx.doi.org/10.1121/1.4964818

Themistocleous, C. (2017). The Nature of Phonetic Gradience across a Dialect Continuum: Evidence from Modern Greek Vowels [Journal Article]. *Phonetica*, *74*(3), 157-172. https://doi.org/10.1159/000450554)

Themistocleous, C. (2019). Dialect Classification From a Single Sonorant Sound Using Deep Neural Networks. *Frontiers in Communication*, *4*, 1-12. https://doi.org/10.3389/fcomm.2019.00064

Themistocleous, C. (2024). Open Brain AI and language assessment. *Frontiers in Human Neuroscience*, *18*. https://doi.org/10.3389/fnhum.2024.1421435

Themistocleous, C., Ficek, B., Webster, K., den Ouden, D.-B., Hillis, A. E., & Tsapkini, K. (2020). Automatic subtyping of individuals with Primary Progressive Aphasia. *bioRxiv*, 2020.2004.2004.025593. https://doi.org/10.1101/2020.04.04.025593

Themistocleous, C., Webster, K., Afthinos, A., & Tsapkini, K. (2020). Part of Speech Production in Patients With Primary Progressive Aphasia: An Analysis Based on Natural Language Processing. *American Journal of Speech-Language Pathology*, 1-15. https://doi.org/10.1044/2020_AJSLP-19-00114

Thompson, C. K., & Mack, J. E. (2014). Grammatical Impairments in PPA. *Aphasiology*, *28*(8-9), 1018-1037. https://doi.org/10.1080/02687038.2014.912744

Ulatowska, H. K., Doyel, A. W., Stern, R. F., Haynes, S. M., & North, A. J. (1983). Production of procedural discourse in aphasia. *Brain and Language*, *18*(2), 315-341. https://www.sciencedirect.com/science/article/abs/pii/0093934X83900238?via%3Dihub

Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and Language*, *13*(2), 345-371. https://www.sciencedirect.com/science/article/abs/pii/0093934X81901000?via%3Dihub

Van Hees, S., & et al. (2013). A comparison of semantic feature analysis and phonological components analysis for the treatment of naming impairments in aphasia. *Neuropsychological Rehabilitation*, *23*(1), 102–132. https://doi.org/10.1080/09602011.2012.726201

Wallace, S. E. K. M. D. Z., & Wood, R. L. (2013). Generalization of word retrieval following semantic feature treatment. *NeuroRehabilitation*, *32*(4), 899-913. https://doi.org/10.3233/NRE-130914

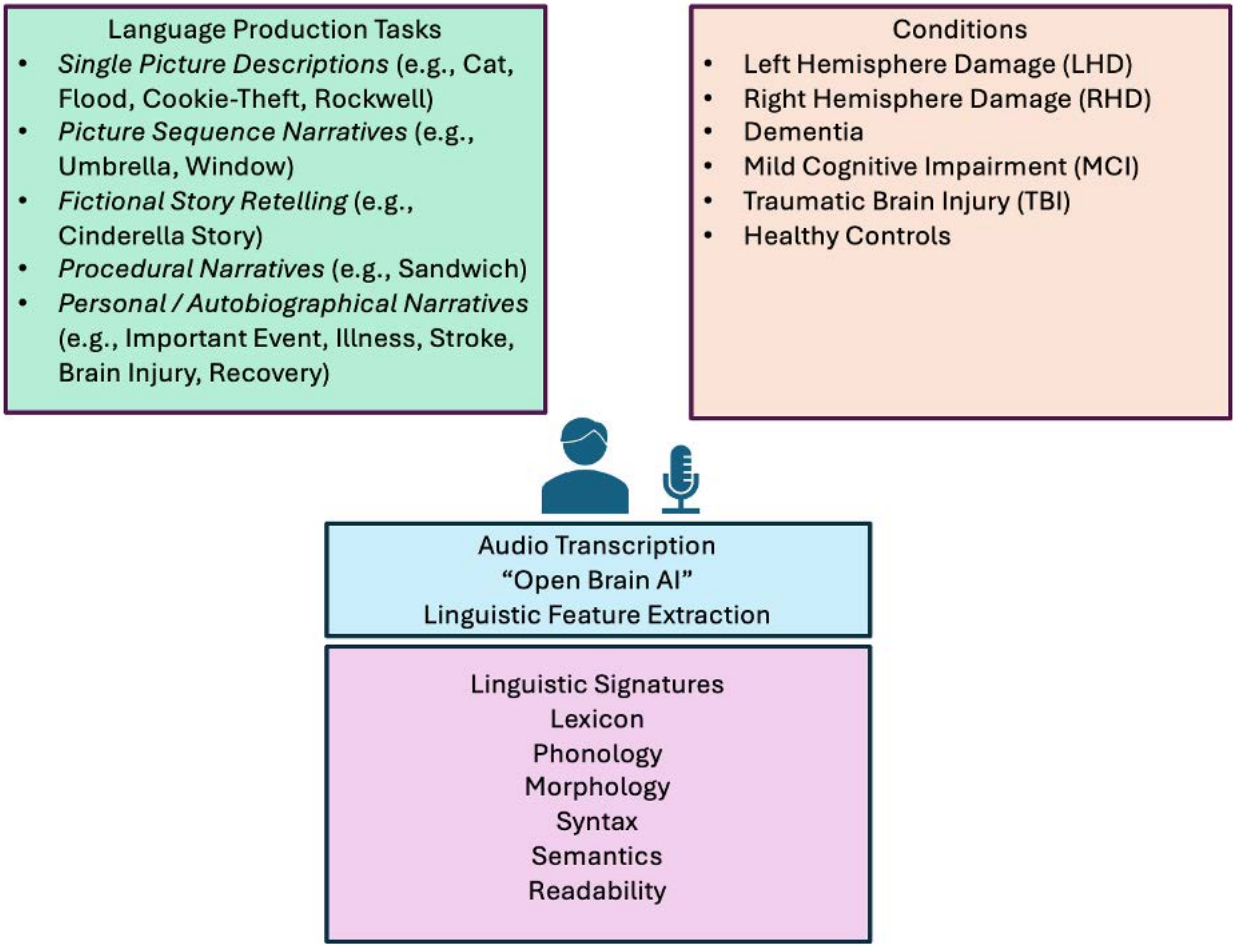
Wambaugh, J. (2003). A comparison of the relative effects of phonologic and semantic cueing treatments. *Aphasiology*, *17*(5), 433–441. https://doi.org/10.1080/02687030344000085

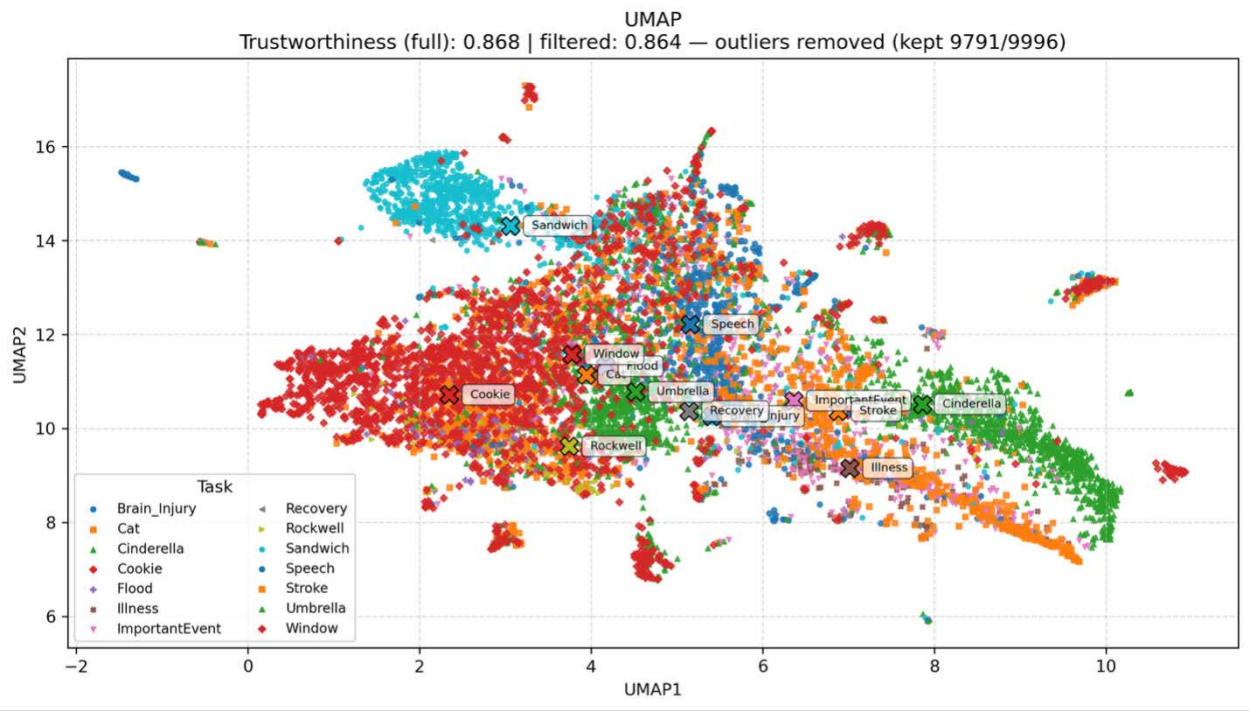
Wambaugh, J. L., & et al. (2001). Effects of two cueing treatments on lexical retrieval in aphasic speakers with different levels of deficit. *Aphasiology*, *15*(10–11), 933–950. https://doi.org/10.1080/02687040143000302

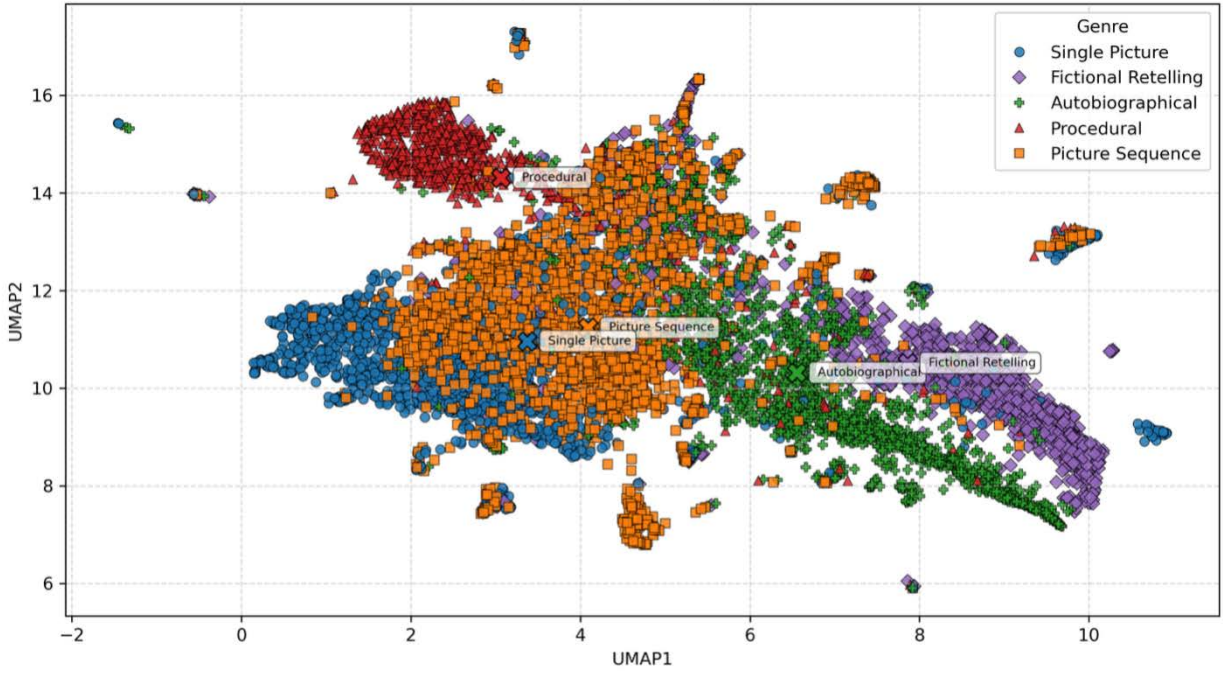
Wambaugh, J. L., & et al. (2013). Semantic feature analysis: Incorporating typicality treatment and mediating strategy training to promote generalization. *American Journal of Speech-Language Pathology*, *22*. https://doi.org/10.1044/1058-0360(2013/12-0070)

Wilson, B. M., & Proctor, A. (2002). Written discourse of adolescents with closed head injury. *Brain Injury*, *16*(11), 1011-1024. https://doi.org/10.1080/02699050210147248

Wilson, S. M., DeMarco, A. T., Henry, M. L., Gesierich, B., Babiak, M., Miller, B. L., & Gorno-Tempini, M. L. (2016). Variable disruption of a syntactic processing network in primary progressive aphasia. *Brain*, *139*(11), 2994-3006. https://doi.org/10.1093/brain/aww218


Language Production Tasks
• Single Picture Descriptions (e.g., Cat, Flood, Cookie-Theft, Rockwell)
• Picture Sequence Narratives (e.g., Umbrella, Window)
• Fictional Story Retelling (e.g., Cinderella Story)
• Procedural Narratives (e.g., Sandwich)
• Personal / Autobiographical Narratives (e.g., Important Event, Illness, Stroke, Brain Injury, Recovery)
Conditions
• Left Hemisphere Damage (LHD)
• Right Hemisphere Damage (RHD)
• Dementia
• Mild Cognitive Impairment (MCI)
• Traumatic Brain Injury (TBI)
• Healthy Controls
Audio Transcription
"Open Brain AI"
Linguistic Feature Extraction
Linguistic Signatures
Lexicon
Phonology
Morphology
Syntax
Semantics
Readability


UMAP
Trustworthiness (full): 0.868 | filtered: 0.864 — outliers removed (kept 9791/9996)
UMAP2
UMAP1
16
14
12
10
8
6
−2
0
2
4
6
8
10
Sandwich
Speech
Window
Flood
Umbrella
Cookie
Recovery
Injury
ImportantEvent
Stroke
Cinderella
Rockwell
Illness
Task
Brain_Injury
Cat
Cinderella
Cookie
Flood
Illness
ImportantEvent
Recovery
Rockwell
Sandwich
Speech
Stroke
Umbrella
Window


UMAP projection colored by Genre — outliers removed (kept 9791/9996)
Trustworthiness (full): 0.868 | filtered: 0.864
Genre
Single Picture
Fictional Retelling
Autobiographical
Procedural
Picture Sequence
Procedural
Picture Sequence
Single Picture
Autobiographical
Fictional Retelling
UMAP2
UMAP1
16
14
12
10
8
6
−2
0
2
4
6
8
10