Universal and language-specific connected speech characteristics of bilingual speakers with Alzheimer's disease: insights from case studies of structurally distinct languages

Article

Accepted Version

Dutta, M., Mello, T. M., Cheng, Y., Dash, N. S., Nandi, R., Dutt, A., & **Bose, A.** (2024). Universal and Language-Specific Connected Speech Characteristics of Bilingual Speakers With Alzheimer's Disease: Insights From Case Studies of Structurally Distinct Languages. *Journal of Speech, Language, and Hearing Research*, 1-22.

https://pubs.asha.org/doi/abs/10.1044/2024_JSLHR-23-00254

Universal and Language-Specific Connected Speech Characteristics of Bilingual Speakers with Alzheimer's Disease: Insights from Case Studies of Structurally Distinct Languages

Manaswita Dutta[1], Tina M. D Mello[2], Yesi Cheng[2], Niladri Sekhar Dash[3], Ranita Nandi[4], Aparna Dutt[4], and Arpita Bose[2]*

[1]Department of Speech and Hearing Sciences, Portland State University, Portland, OR, USA

[2]School of Psychology and Clinical Language Sciences, University of Reading, Reading, UK

[3]Linguistic Research Unit, Indian Statistical Institute, Kolkata, India

[4]Neuropsychology & Clinical Psychology Unit, Duttanagar Mental Health Center, Kolkata, India

Keywords: Connected speech, Spoken Discourse, Bilinguals, Alzheimer's Disease, Linguistic Analysis, Bengali, English

*** Corresponding author:**

Arpita Bose, PhD.

School of Psychology and Clinical Language Sciences,

University of Reading, Reading, UK.

Email: a.bose@reading.ac.uk

Conflict of Interest: No authors have a conflict of interest

Funding: The Center of Literacy and Multilingualism (CELM) Pump Priming Grant, University of Reading (Arpita Bose), the British Academy International Visiting Fellowship Grant –

VF1/103620 (Niladri S. Dash), and the University of Reading's Undergraduate Research Opportunities Program (UROP, Tina M. D Mello).

Acknowledgments: We are grateful to all participants for their enthusiasm and time to participate in this research.

Abstract

Purpose: Connected speech analysis has been effectively utilized for the diagnosis and disease monitoring of individuals with Alzheimer's Disease (AD). Existing research has been conducted mostly in monolingual English speakers with a noticeable lack of evidence from bilinguals and non-English speakers, particularly in non-European languages. Using a case study approach, we characterized connected speech profiles of two Bengali-English bilingual speakers with AD to: determine the universal features of language impairments in both languages, identify language-specific differences between the languages, and explore language impairment characteristics of the participants with AD in relation to their bilingual language experience.

Method: Participants included two Bengali-English bilingual speakers with AD and a group of age-, gender-, education- and language-matched neurologically healthy controls. Connected speech samples were collected in L1 (Bengali) and L2 (English) using a novel storytelling task (i.e., "Frog, where are you?"). These samples were analyzed using an augmented Quantitative Production Analysis and Correct Information Unit analyses for productivity, fluency, syntactic and morpho-syntactic features, lexical and semantic characteristics.

Results: Irrespective of the language, AD impacted speech productivity (speech rate and fluency) and semantic characteristics in both languages. Unique language-specific differences were noted on syntactic measures (reduced sentence length in Bengali), lexical distribution (fewer pronouns and absence of reduplication in Bengali) and inflectional properties (no difficulties with noun or verb inflections in Bengali). Among the two participants with AD, the individual who showed lower proficiency and usage in L2 (English) demonstrated reduced syntactic complexity and morpho-syntactic richness in English.

Conclusion: Evidence from these case studies suggest that language impairment features in AD are not universal across languages, particularly in comparison to impairments typically associated with language breakdowns in English. This study underscores the importance of establishing connected speech profiles in AD for non-English speaking populations, especially for structurally different languages. This would in turn lead to the development of language-specific markers that can facilitate early detection of language deterioration and aid in improving diagnosis of AD in individuals belonging to underserved linguistically diverse populations.

Introduction

Alzheimer's Disease (AD) is a progressive neurodegenerative disease that can impair a range of cognitive skills including memory, language, attention, processing speed, and executive functioning (McKhann et al., 2011; Weiner et al., 2008). The assessment of connected speech (also referred to as *spoken discourse*) or the use of language beyond individual words and sentences, has sparked significant interest in AD research and clinical practice. This is due to its ease of administration, minimal burden on clients, and usefulness in diagnosing and monitoring disease progression (Ahmed et al., 2013; Boschi et al., 2017; Filiou et al., 2020; Forbes-McKay et al., 2013; Mueller et al., 2018; Slegers et al., 2018). Existing empirical research in connected speech in AD has focused primarily on studying monolingual English speakers. Consequently, very little is known regarding the connected speech profiles of non-English speakers and/or bilingual individuals (Bose et al., 2021; Calabria et al., 2017; Stilwell et al., 2016). With over half of the world's population speaking more than one language and the incidence of older bilingual speakers with AD rising globally, it is important to expand research beyond English and identify language impairment patterns of persons with AD who speak different languages, particularly non-European languages (Petti et al., 2020). These lines of research would lead to

the development of language-specific impairment markers and facilitate early identification of language decline and improve language assessment in these underserved populations.

Bilingualism and AD

Bilingualism and AD research has been undertaken from the perspective of establishing links between the benefits of bilingualism and cognitive reserve and its consequences on the onset of dementia (Alladi et al., 2013; Bak & Robertson, 2017; Bialystok et al., 2016). However, limited research exists in how dementia impacts languages spoken by bilingual speakers, and particularly how AD impacts linguistic features across different languages (see review by Stilwell et al., 2016). Untangling the effects of AD on both languages spoken by bilinguals is a critical line of research for clinical and theoretical reasons. First, comprehensive linguistic assessments such as connected speech analysis can aid in identifying language-universal and language-specific impairments in structurally different languages spoken by bilingual speakers with AD. Second, profiling linguistic impairments in both languages of bilingual AD can help determine the progression (deterioration) of each language through the course of the disease, and their relationship to bilingualism variables and typological factors. Third, detailed characterization can play an important role in facilitating early diagnosis and effective intervention. Collectively, this information is clinically useful for early identification of language decline, developing appropriate assessments and generating functional communication strategies for individuals with AD.

Connected speech characteristics in AD across languages

Persons with AD frequently demonstrate impairments in connected speech which can affect both the structural (i.e., microlinguistic) and global (i.e., macrolinguistic) levels of language (Carlomagno et al., 2005; Mueller et al., 2018; Slegers et al. 2018). Research primarily

from English speaking populations has shown that specific connected speech features distinguish AD from neurologically healthy adults. These include reduced speech rate and spontaneity characterized by increased repetitions and revisions (Ahmed et al., 2013; Sajjadi et al., 2012), simplified syntax and sentence structures including shorter sentences with reduced grammatical complexity (Fraser et al., 2015), word finding difficulties, increased use of pronouns (Forbes-McKay et al., 2013; Gayraud et al., 2010), inflectional errors in nouns and verbs (Ahmed et al., 2013; Sajjadi et al., 2012), and reduced semantic content and less informative language output (Cuetos et al., 2007). Although language-universals exist in impairments (i.e., overlap in linguistic features that would be similarly impaired across languages), prior research indicates that features of language impairments and specific linguistic markers in AD vary with the structure of the language being studied (Bose et al., 2021, 2022; Kavé & Levy, 2003; Paradis, 1998). To illustrate, whereas increased production of pronouns is a characteristic feature in English speakers with AD (Ahmed et al., 2013), decreased pronoun production was identified to be a feature of pro-drop languages such as Bengali (Bose et al., 2021) and Hebrew (Kavé & Goral, 2016; Kavé & Levy, 2003). These differences are not idiosyncratic findings across studies but represent the linguistic characteristics of a language. As Bengali is an inflectionally rich pro-drop language and allows dropping of the subject, the subject can be inferred from the other inflected parts of speech. In languages where subjects are obligatorily spelled out, such as in English, dropping the subject is not an option. A simple deduction from this cross-linguistic observation is that when a language allows the avoidance of a linguistic feature or structure, such as subject drop in Bengali, individuals with AD will avoid retrieving and producing the subject, as it may be more cognitively demanding. This finding implies that over-production of pronouns in AD, which is a characteristic feature in English, is not a relevant linguistic marker for a pro-

drop language, such as Bengali. Importantly, linguistic impairments in AD are not comparative across languages and differ from the language breakdowns typically observed in English (Bose et al., 2021, 2022; Kavé & Levy, 2003). With the majority of the research focusing on English-speaking monolingual individuals (refer to Supplementary Table 1 for a summary of relevant experimental speech and language studies in bilingual AD), we lack knowledge about the linguistic impairment profiles of bilinguals and native speakers of languages other than English, particularly in relation to language-specific features.

Factors influencing the presentation of language deficits in bilingual AD

Several factors such as the type of linguistic task, bilingualism variables, language combinations studied, all contribute to the heterogeneity of findings across bilingual AD studies (see Supplementary Table 1). First, a wide range of *language tasks* have been used to explore the linguistic deterioration in bilingual speakers with AD. Most studies have predominantly used single word level tasks focusing on lexical-semantic processing (e.g., confrontational naming, word repetition, object recognition, lexical decision, oral reading, verbal fluency (e.g., Calabria et al., 2017; De Picciotto et al., 2001; Ivanova et al., 2014; Meguro et al., 2003; Nanchen et al., 2017). While these tasks are sensitive in capturing lexical-semantic deficits in AD, they do not provide information regarding individuals' abilities to communicate in everyday conversational situations. Very few studies have examined discourse or conversational-level language in bilingual speakers with AD, and the existing ones have mainly evaluated code switching and language mixing in AD (e.g., Friedland & Miller, 2010; Hyltenstam & Obler, 1989; Hyltensam & Stroud,1993). To the best of our knowledge, no study has conducted a comprehensive analysis of linguistic features of connected speech across two languages in bilingual speakers with AD.

Individual-specific *bilingualism profiles* also influence the presentation of linguistic deficits in bilingual individuals with AD. For instance, Gollan et al. (2010), Ivanova et al. (2014), and Salvatierra et al. (2007) studied non-balanced Spanish-English bilingual speakers with AD. While Salvatierra et. (2007) reported equal deterioration of both languages in their participants; both Gollan et al. (2010) and Ivanova et al. (2014) reported differential impairments between dominant and non-dominant languages. That is, Gollan et al. (2010) found that the dominant language was more impaired; in contrast, Ivanova et al.'s longitudinal analysis revealed that the non-dominant language deteriorated faster than the dominant language. Although participants of all three studies were sequential Spanish-English speakers, participants in Gollan et al. (2010) and Ivanova et al. (2014) were English-dominant and Spanish-dominant bilinguals respectively whereas, participants in Salvatierra et al. (2007) were equally proficient in both English and Spanish. Therefore, this highlights the differential impact that bilingualism profiles can have on linguistic impairments in bilingual speakers with AD. In addition to bilingualism profiles of the participants, the varied findings across these three studies could be due to different tasks and designs used (i.e., picture naming in Gollan et al. [2010] and Ivanova et al. [2014]; phonemic and semantic verbal fluency in Salvatierra et al. [2007]; see Supplementary Table 1). Another critical issue regarding the bilingual profile characterization pertains to the various definitions used for L1 (first language) and L2 (second language) across different studies. It is important to note that individuals may not necessarily be proficient and/or frequently use their L1, as highlighted in a study by Ellajosyula et al. (2020) on bilingual individuals with primary progressive aphasia. Nevertheless, it is still essential to account for these variables when interpreting bilingual language characteristics in neurologically impaired populations (Costa et al., 2012; Lerman et al., 2019).

Specific linguistic impairments in bilingual speakers with AD also vary based on the *language combinations* studied. Costa et al. (2012) studied two groups of early and highly proficient Catalan-Spanish speakers with mild (n = 23) and moderate (n = 24) AD in comparison to a control group of participants with mild cognitive impairment (MCI; n = 24) on lexical-semantic tasks. Results revealed that both languages were similarly affected in participants with AD indicating a parallel decline in the underlying lexical-semantic system and shared neural network supporting both languages. These findings from Costa et al.'s study was not surprising as Catalan and Spanish are structurally similar languages. In contrast, Meguro et al. (2003) studied four Japanese-Portuguese bilingual speakers using several single word lexical-semantic tasks; participants demonstrated differential patterns of impairments in single word reading tasks across Japanese and Portuguese, which are two structurally distinct languages.

As can be seen from the ongoing evaluation of the research evidence of linguistic impairments in the two languages spoken by bilinguals with AD, impairment profiles are determined by a complex interaction of methodological variables (e.g., task used, linguistic level analyzed), bilingualism profiles of the AD participants, and linguistic characteristics of the languages studied. At present, the extant literature is limited to make strong claims regarding how the two languages will deteriorate with the progression of AD. Contemporary bilingual research has primarily focused on specific languages such as Spanish. Aside from the Spanish-speaking community, South Asians represent one of the most rapidly expanding demographic groups in the United States (Faroqi-Shah, 2012; Mahendra, 2012). Moreover, as AD is becoming more prevalent in South Asian and Western Pacific regions, including countries like China and India (Li et al., 2022), it is imperative to broaden our empirical knowledge regarding how AD affects other languages such as Bengali, Hindi, and Urdu. This expansion of research is crucial

for discerning early identification of language decline which could facilitate prompt and accurate dementia diagnosis as well as rehabilitation of individuals from these linguistic backgrounds. The following section highlights the key linguistic characteristics of Bengali and English, as this knowledge is pertinent to understanding the characterization of connected speech profiles of the two case studies we will be discussing in this paper.

Bengali and English: Cross-Linguistic Differences

Bengali (*Bangla*) is an Indo-Aryan language and is spoken by an estimated 272 million people as a first or second language globally and it is the sixth most commonly spoken language in the world (Noack & Gamio, 2015; Census of India, 2011; U.S. Census Bureau., 2017). Bengali and English are syntactically, lexically, and morphologically distinct languages. Table 1 provides a summary of the cross-linguistic differences between Bengali and English that are relevant for characterizing language features in AD. Key differences between English and Bengali include the word order: English adopts a rigid SVO (Subject-Verb-Object) order, whereas Bengali adopts a more fluid word order with SOV (Subject-Object-Verb), as it's canonical form. Further, Bengali is an agglutinative language with extensive, complex and systematic inflectional morphology (Thompson, 2010). In terms of lexical distribution, in contrast to English, Bengali has fewer closed-class words (Bengali: pronouns, postpositions, indeclinables vs. English: prepositions, determiners, pronouns, conjunctions, modals, auxiliaries). In Bengali, extensive inflectional properties of open-class words serve the role of closed-class words that are utilized in English. As a pro-drop language, Bengali allows for context-based omission of pronouns in the subject position, whereas English does not allow for omission of obligatory pronouns. Reduplication is a prevalent linguistic feature found in several Indian languages, including Bengali. It involves the duplication of a word, either wholly or

partially, to create a new word that is different in form and adds a new sense in meaning. Reduplication serves multiple semantic functions, such as sense of multiplicity, continuation of action, recurrence of an event or emotional state, or emphasis (e.g., [d̪in] "day" à [d̪ind̪in] "day-by-day"; [ʤɔl] "water" à [ʤɔlʤɔle] "watery"; [gʰɔr-e] "house-$_{\text{locative}}$"à[gʰɔre gʰɔre]"in every house").

These linguistic differences between the languages are likely to manifest distinctively in connected speech characteristics – *language-specific features*. It is important to reiterate that certain features will be affected in both languages - *universal features* - among bilingual individuals with AD. Thus, considering the variations between languages, detailed examination of language impairment and decline in both languages is warranted.

Insert Table 1 about here.

Research Gaps and the Current Study

Despite increased recognition that linguistic impairments are important markers for AD (e.g., Cuetos et al., 2007; Fraser et al., 2015), very little is known regarding patterns of linguistic deficits in speakers of languages other than English and in bilingual speakers. Given the lack of research on detailed characterization of connected speech in bilingual speakers with AD, these case studies are the first ever attempt to document and comprehensively analyze the language profiles of Bengali-English bilingual speakers with AD. Additionally, the inclusion of Bengali-English speakers serves to extend the diversity of published research in acquired neurogenic communication disorders (Beveridge & Bak, 2011).

Research Questions and Predictions

In this study, we aimed to characterize connected speech profiles of two Bengali-English bilingual speakers with AD in Bengali (L1) and English (L2) across six linguistic levels (speech

rate, syntactic, lexical, morphological, semantic and speech fluency). The specific aims of this research were to:

1. Determine the universal features of language impairment in the two bilingual individuals with AD, that is, linguistic features which are affected in both languages.
2. Identify the language-specific impairments between Bengali and English, that is, linguistic features which are only affected in the specific language.
3. Explore connected speech features of individual bilingual AD participants in relation to their language experiences and background.

Based on the distinct linguistic properties of Bengali and English, we expected to find differential impairment patterns in L1 (Bengali) and L2 (English) across some of the linguistic levels but not all. Specifically, we predicted that individuals with AD would present similar patterns of impairments in both languages in terms of speech productivity, semantic content, and fluency. However, we expected to observe language-specific variations in syntax, morphology, and lexical distribution between Bengali and English (i.e., proportion of pronouns, postpositions, reduplications, word order, nouns and verb inflections), given the structural differences between both languages. Additionally, we hypothesized that cross-linguistic patterns of deficits would vary with the participants' bilingualism profiles.

Methods

Participants

Two Bengali-English bilingual adults (AD03; AD09) with a clinical diagnosis of probable AD based on the NINCDS/ADRAA criteria (McKhann et al., 2011) and eight age-, gender-, education- and language-matched neurologically healthy controls (HC) participated in the study. AD03 and AD09 were part of a larger research project investigating language

production in bilingual AD involving six participants. Of these six individuals, only two of them (AD03; AD09) were able to complete the connected speech tasks in both languages. All participants were right-handed urban community dwelling sequential Bengali-English bilingual speakers with better proficiency and higher usage in Bengali (L1) than English (L2). AD03 and AD09 were recruited from the Duttanagar Mental Health Centre, Kolkata, eastern India, and were living with their families at the time of the study.

HC participants were recruited from a volunteer participant pool. Exclusion criteria for both groups included: (1) a known history of alcohol or drug abuse, (2) a history of other neurological or psychiatric illness, or (3) less than ten years of education. Ethical approval for this study was obtained from the University of Reading (2017-035-AB). The demographic details for all participants are provided in Table 2.

Participant AD03 was a retired electrical supervisor. As per his wife, his symptoms began 36 months prior to receiving his AD diagnosis. His initial presenting symptom was impaired memory, evidenced by forgetting recent conversations and meals. He was reported to have difficulty recognizing his house, among other visuospatial difficulties. Moreover, he was reported to present with neuropsychiatric symptoms such as increased aggression. Participant AD09 used to work as an English teacher at a high school. However, as per her husband, she retired as she was no longer able to cope with the job's demands. Her symptoms began four years prior to receiving her AD diagnosis and initially presented with progressive loss of memory evidenced by difficulties with recalling the names of places, remembering recent conversations and books she had read, and remembering meals she had eaten. She exhibited behavioral changes such as withdrawal from social groups, confabulations, and difficulties with recognizing familiar places. Additionally, she demonstrated difficulty with comprehension and an increased reliance

on her spouse for activities of daily living. AD09 was noted to need more time understanding instructions while completing the assessments in this study.

Insert Table 2 about here.

Background Assessments

Neuropsychological assessments. General cognitive functioning was measured using the *Kolkata Cognitive Screening Battery* (KCSB), an adapted Bengali version of *Mini-Mental State Examination* (BMSE; Das et al., 2006), the Bengali adapted version of *Addenbrooke's Cognitive Examination* (ACE-III; Hsieh et al., 2013) and the *Clinical Dementia Rating Scale* (CDR, Morris, 1993). The CDR is a measure of dementia severity based on the individual's cognitive and daily functions across several domains including memory, orientation, judgement, problem solving, community affairs, home and hobbies, and personal care. The study participants as well as their family members completed the CDR. In addition, the *Instrumental Activities of Daily Living Scale for Elderly* (IADL-EDR; Mathuranath et al., 2005) was completed by the participants' family members as a proxy measure of each participant's ability to undertake day-to-day activities; this included cognitive activities, social and recreational activities, community activities, household activities and self-care activities. A composite percentage score derived from the questionnaire indicated overall physical and cognitive disability.

None of the HC participants exhibited any cognitive symptoms or neurological disorders and performed within the normal range on the KCSB, ACE-III, CDR and IALD-EDR assessments. Compared to HC, both AD03 and AD09 demonstrated significantly lower ACE-III and KCSB scores, marked difficulties in IADLs, and an overall severity of mild dementia (i.e., CDR global score of 1). All clinical interviews and assessments were performed by the fifth and

sixth authors (RN, AD) who are clinical neuropsychologists with extensive research experience in cross-cultural neuropsychology. See Table 2 for neuropsychological test results.

Bilingualism measures. All participants completed a set of subjective language background questionnaires (language acquisition history, language of instruction, self-rated language proficiency, language usage, and language dominance). For individuals with AD, information was corroborated by their family members. To measure language acquisition history, self-rated language proficiency (in speaking, comprehension, reading, and writing), and the current language usage pattern, we adapted and modified the questionnaire developed by Muñoz et al. (1999). Language dominance was measured using the language dominance questionnaire (Dunn & Fox Tree, 2009). See Patra et al. (2020) for the adapted versions of these questionnaires. In addition, participants completed a questionnaire regarding their medium (i.e., language) of instruction during different levels of education. Bilingualism profiles generated from the bilingualism measures are presented in Table 3.

Insert Table 3 about here.

The onset of exposure to Bengali for all participants began since birth as they all grew up in Bengali households and were educated in Kolkata. The language acquisition questionnaire indicated that AD03, AD09 and the HC participants all had greater immersion in Bengali compared to English. AD09 reported an early onset of exposure to English (at around 5 years), both formally and informally in contrast to AD03 who was formally exposed to English in primary school at the age of six years and informally with friends and neighbors at around 11 years. All participants were exposed to English after the age of three years.

In terms of formal education, AD03's primary medium of instruction until high school was Bengali, and only in his undergraduate degree the medium of instruction included both Bengali

and English. In contrast, AD09 had mixed medium of instruction during schooling (i.e., both Bengali and English) from primary school to her postgraduate master's degree.

Data from the usage and proficiency questionnaires indicated that both AD03 and AD09 had greater current language proficiency in L1 compared to L2. Following the guidelines provided by Dunn and Fox Tree (2009), scores from the *Bilingual Dominance Scale* showed similar results for both AD participants, with L1 (Bengali) being their dominant language (AD03 dominance scores for L1 = 26, L2 = -4; AD09 dominance scores for L1 = 23, L2 = 7; total possible score = 31). For purposes of interpretation, higher scores showed higher dominance in one language (i.e., Bengali) compared to the other language (i.e., English). Furthermore, both participants indicated higher frequency and greater use of L1 compared to L2. An important difference of note between AD03 and AD09 was that for the measures of proficiency, dominance and current use, the difference in measures between L1 and L2 was smaller in AD09 (i.e., proficiency rating L1 = 7, L2 = 5.88; usage L1 =5, L2 = 3.75); in contrast to AD03 whose difference between L1 and L2 on these parameters were greater (i.e., proficiency rating L1 = 7, L2 = 2.38; usage L1 = 5, L2 = 2). It is worthy to reiterate that AD09 had a master's degree in English and was working as a high school English teacher until her dementia onset, likely indicating a greater proficiency in English compared to AD03.

In summary, based on the bilingualism profiles, AD03 and AD09 in addition to all HC participants were considered sequential bilinguals, with Bengali as their L1 and English as their L2. Both AD03 and AD09 were dominant in L1 with greater proficiency and current use in L1. However, compared to AD03, AD09 was more "balanced" with lesser difference in L1 and L2 proficiency and usage.

Connected speech task

Data collection and recording. Narrative samples using the wordless picture book, "Frog, where are you?" (Mayer, 1969) were elicited by author RN in Bengali and English on separate testing days. Storytelling was chosen as it has shown to elicit richer and more descriptive language samples with the greater sensitivity to capture language-specific distinctions when compared to single-picture descriptions (Bose et al., 2022). This story was selected to ensure that the participants found the story relevant and culturally appropriate. While most studies in English speakers with AD and other dementias have used the Cinderella Story retelling narrative task (e.g., Fraser et al., 2015; Kavé et al., 2007), Bengali speakers residing in Kolkata, India, are less likely to be familiar with all the specifics of Cinderella, even if they are acquainted with the general concept of the story. The "Frog, where are you?" story has been used dementia literature (e.g., Ash et al., 2007; Clark et al., 2021) and does not rely on specific knowledge of specific story concepts (such as those in Cinderella), therefore making the story more appropriate for the participants.

Prior to administering the narrative task, participants were given a brief background about the story and were told that the main characters of the story are a boy, his dog, and a frog. The story is about a boy who is searching for his missing frog along with his dog. Participants were instructed to look through the wordless picture book and then asked to narrate the story using complete sentences. Instructions for eliciting connected speech tasks in Bengali were delivered in Bengali, while instructions for the English samples were presented in English. Participants could keep the book with them while narrating the story. Other than occasional prompts and generic encouragement, tester interruptions were kept to a minimum. No feedback was provided during narrative elicitation. Instructions for testing and feedback were written

down for the tester to ensure consistency in instruction across participants. Testing was conducted in a quiet room. The language samples were recorded using the digital audio recorder Olympus voice recorder WS-833 for subsequent orthographic transcription.

Transcription and data extraction. Each narrative sample was transcribed manually verbatim, segmented, and analyzed in accordance with the procedures outlined in the Quantitative Production Analysis (QPA; Berndt et al., 2000). All raters involved in the scoring process listened the full audio samples for each participant, transcribing the entirety of the samples. Author YC completed the transcriptions in English and author ND performed the transcriptions in Bengali. As in the original QPA, utterances were defined as segments of running speech that were coherent in terms of syntax and/or prosody. Placement of sentence boundaries was guided by semantic, syntactic, and prosodic features. An utterance did not have to constitute a fully grammatical sentence. Using the QPA rules of extracting the narrative core, words that did not contribute to the narrative (i.e., repetitions, repairs, examiner's prompts, discourse markers, nonwords) were removed (see Berndt et al., 2000; Rochon et al., 2000 for specific steps in extracting the narrative words). Authors AB and NSD performed the narrative core extraction individually for all the 10 speech samples (AD n = 2; HC n = 8) in Bengali, and authors YC and AB performed the narrative core extraction individually for all the English (AD n = 2; HC n = 8) samples.

The minimum length of the language sample for obtaining meaningful results from a narrative production has been widely debated (e.g., Berndt et al., 2000; Sajjadi et al., 2012). The QPA analysis protocol recommends a corpus of 150 words for narrative analysis (Berndt et al., 2000). Moreover, previous research with different sample lengths have shown that a 150 narrative word corpus produced an adequate and reliable analysis (Sajjadi et al., 2012).

Following the abovementioned recommendations and to keep the sample length consistent across participants, we derived the measures after extracting 150±10 narrative words. The total narrative duration and total number of words produced by each participant were also recorded. To ensure reliability, transcriptions and coding were further reviewed and agreed upon by multiple authors (AB, MD, NSD, YC, and TMD). Authors AB and NSD (a linguist with extensive expertise in Bengali linguistics) jointly scored all language samples and coded for the variables in Bengali. Given that this was the first time QPA was being augmented for Bengali, the scoring process was conducted collaboratively to ensure the analysis was linguistically robust. Authors YC and TDM coded and analyzed the English samples. For CIUs, authors MD, AB, and YC jointly conducted the coding for variables. Any disagreements in narrative core extraction or utterance segmentation were resolved through review of the QPA rules, and re-listening of the audio samples. Variable scores for each sample were finalized only after unanimous agreement among all scorers. The coding agreement between scorers was 100%. Excerpts of transcripts in Bengali and English for the two participants with AD (AD03 and AD09) and one HC participant are provided in the Table 4. Best Practice Guidelines for Reporting Spoken Discourse in Aphasia and Neurogenic Communication Disorders by Stark et al. (2022) were followed (see Supplementary Table 2 for completed checklist).

Insert Table 4 about here.

Quantitative Analysis of Narratives. The QPA and the Correct Information Unit (CIU; Nicholas & Brookshire, 1993) analyses were implemented to calculate a set of count and proportional measures for each connected speech sample. The QPA scheme was augmented, and additional variables were included in the analysis scheme to capture specific linguistic features of Bengali (e.g., verbal and nominal morphology, proportion of postpositions, number of

reduplications; see Table 5). CIUs are a widely used metric in narrative analysis that assess the informativeness and efficiency of information conveyed through connected speech (Ahmed et al., 2013). The multidimensional nature of connected speech analysis and the large number of different variables used by researchers makes the choice of reporting appropriate variables a challenging task, especially for cross-linguistic comparisons. To keep the comparisons succinct for facilitating cross-linguistic comparisons, we focused on the proportional measures that provide insights into cross-linguistic differences and also capture language-specific features of Bengali (e.g., proportion of postpositions) and English (e.g., proportion of prepositions). The measures reported in the current study are aligned with the recommendations from recent reviews regarding the linguistic levels that are essential for characterizing language in AD (Slegers et al., 2018; Filiou et al., 2020). Previous studies examining connected speech in AD (e.g., Ahmed et al., 2013; Boschi et al., 2017; Frazer et al., 2016) have quantified six different aspects of language production: (1) speech rate; (2) structural and syntactic measures; (3) lexical measures; (4) morphological and inflectional measures; (5) semantic measures; and (6) measures of spontaneity and fluency disruptions.

The following section provides a brief description of the six linguistic levels and specific variables used for characterizing the samples and to capture the cross-linguistic differences between Bengali and English. See Table 5 for the definition of these variables and features relevant for Bengali versus English (see Supplementary Table 3 for a complete list of variables derived from the QPA and CIU analyses).

- **Speech rate:** defined as the number of words produced per minute.

- **Structural and syntactic measures**: this level measured the mean length, complexity, and grammaticality of sentences to capture the structural and syntactic aspects of language production.
- **Lexical measures**: these measures captured participant's production of various types of lexical items (e.g., nouns, verbs, pronouns, postpositions, prepositions) across all extracted narrative words, independent of utterance type. For this study, the following set of variables were calculated: proportion of open class words, closed class words, nouns, personal pronouns, verbs, postpositions (Bengali only), prepositions (English only), and number of reduplications (Bengali only).
- **Morphological and inflectional measures**: in QPA, morphosyntactic properties of nouns and verbs in English are indexed by two metrices “noun determiner index” and “verb inflection score”. To capture the morphological richness and intricacies of Bengali noun and verb inflectional system, we generated the following measures (see Supplementary Table 3 for definitions): For nominal inflections, we determined the total number of: (1) nouns, (2) nouns in their base form (i.e., uninflected forms), (3) nouns that are possible to be inflected, (4) nouns with appropriate inflections, and (5) inflections on each noun (i.e., one, two, > two inflections). From these count measures, we derived the noun inflection index, which indicated if inflected nouns are correctly inflected or not. For verbs, we determined the total number of: (1) verbs, (2) inflectable verbs, and (3) inflected verbs with appropriate inflections. From these count measures, the verb inflection index was calculated to indicate if inflected verbs are inflected correctly or not. By doing this, we aimed to arrive at comparable metrices of noun and verb inflections between the languages (i.e., noun determiner index and verb inflection score in English; noun inflection index and verb inflection index in Bengali).

- **Semantic measures**: semantic content of the narrative samples was quantified separately using the CIU measures. Three measures were derived from the CIU analysis: number of CIUs, idea density and idea efficiency.
- **Measures of spontaneity and fluency disruptions**: the total counts of disruption in spontaneity and fluency were included to capture the number of repetitions, revisions, and reformulations in the narrative samples.

Insert Table 5 about here.

Statistical analysis

Given that we report two case studies, we implemented the well-established single-subject statistical method by Crawford and colleagues for comparing a single case to a small control group (at least five) to identify differences between each AD participant and controls (Crawford & Garthwaite, 2002; 2006; Crawford et al., 2010). The score for each variable was extracted and compared to the HC group's average scores for the variable in the same language (i.e., a within-language comparison) with the significance value set at 0.05 ($p < 0.05$, one-tailed).

Results

Table 6 presents the data from participants AD03 and AD09, and the means and standard deviations from HC participants along with the results of the single-subject statistics. Figure 1 illustrates the variables that were affected in both languages for AD03 and AD09 (i.e., language-universal variables) and those that were affected only in one language for both participants (i.e., language-specific variables). The readers are encouraged to review Table 4 of illustrative examples of narrative productions of AD03 and AD09 and their comparison with a control participant.

Insert Table 6 and Figure 1 about here.

Universal features of connected speech impairments in bilingual speakers with AD

For both Bengali and English, participants AD03 and AD09 showed reduced speech rate, increased number of dysfluencies, and reduced CIU% (idea density) and CIUs per minute (idea efficiency) in their narrative productions compared to the control group.

Language-specific differences between Bengali and English in bilingual speakers with AD

Prominent differences in Bengali and English emerged on the structural and syntactic, lexical, and inflectional measures (see Table 6). Compared to HC, participants AD03 and AD09 both produced shorter mean length of utterance (MLU) with fewer embeddings in Bengali. In terms of lexical distribution, AD03 and AD09 showed reduced proportion of pronouns in Bengali, with no such difference in English. Both participants showed an absence of reduplications in Bengali, which is a characteristic feature of Bengali. Neither individual with AD showed any difficulty with noun and verb inflectional indices in Bengali.

Connected speech characteristics in AD in relation to language experience and background

As a reminder to our readers, AD03 and AD09 were sequential bilinguals, with Bengali as their L1 and English as their L2, with greater dominance, proficiency, and usage in L1. However, compared to AD03, AD09 was a more "balanced bilingual" with lesser difference in L1 and L2 proficiency and usage, and her medium of instruction during all school years had been both in Bengali and English.

In English, participant AD03 showed reduced well-formedness of sentences, whereas AD09 demonstrated no impairments on syntactic and structural measures. Additionally, participant AD09 showed significant differences in the proportion of open class and closed class words for both Bengali and English; however, the direction of the difference was opposite in two languages: The proportion of open class words in Bengali increased, however in English, there

was a notable decrease in their proportion. AD03's difficulty with noun inflections was indicated by reduced noun determiner index in English. Although, AD03's verb inflection index score was lower than the HC mean, the statistical analysis did not reveal a significant difference, possibly due to higher variability in the control group.

In summary, for both participants AD03 and AD09, speech productivity (i.e., speech rate, fluency) and semantic measures were affected in both languages; cross-linguistic differences appeared in syntax, lexical distribution, and inflectional properties. Bengali-specific features that were consistently observed in both participants included reduced well-formedness of sentences, decrease in pronouns, absence of reduplications, and no difficulty with noun or verb inflections in Bengali. In terms of English variables, AD09 showed fewer differences from HC on the syntactic/structural and morphological variables than AD03 who demonstrated less well-formed sentences and difficulty with noun inflections in English.

Discussion

These case studies present a comprehensive analysis of connected speech to evaluate cross-linguistic profiles of two sequential Bengali-English bilingual individuals diagnosed with AD. In line with the existing monolingual AD literature (e.g., Ahmed et al., 2013; Ash et al., 2007; Forbes-McKay et al., 2013; Sajjadi et al., 2012), our findings revealed that both bilingual participants with AD performed significantly differently compared to age-, education-, and gender-matched HC on the narrative language task, affecting all linguistic levels. Impairments in speech productivity, fluency, and semantics were universally affected in both languages; while language-specific differences were noted in the syntactic, lexical, and morphological domains. Additionally, the two bilingual speakers with AD showed differential patterns of impairments in their L1 (i.e., Bengali) and L2 (i.e., English) on select linguistic measures and these patterns

varied with the participants' language experiences and background. Our results add to the extant literature and demonstrate that not all linguistic levels in connected speech are equally affected as a result of AD in bilingual individuals.

Universal features of connected speech impairments in bilingual speakers with AD

We found that both AD03 and AD09 demonstrated slower speech rate, more frequent fluency disruptions, and reduced informativeness and efficiency of ideas in their narrative productions in both Bengali and English. These features are typically known to be affected in AD (e.g., Ahmed et al., 2013; Fraser et al., 2016; Sajjadi et al., 2012). However, in the context of bilingualism, similar deficits in both languages as evidenced by the two case studies highlight that the degenerative changes associated with AD universally impact certain domains irrespective of the language. These deficits reflect the ongoing deterioration in the underlying semantic networks causing slowed retrieval of lexical-semantic content, and changes in cognitive skills including processing speed, working memory, planning, and monitoring of information (Gayraud et al., 2010; Grossman & Rhee, 2001). Furthermore, semantic impairment in both languages among AD03 and AD09 are consistent with the notion that the semantic system and conceptual representations involved during a narrative task are perhaps common across both languages (Costa et al., 2012). Our results are consistent with Costa et al. (2012) and Gollan et al. (2010) and indicate that these linguistic features are similarly affected due to cognitive decline and can be universally impacted irrespective of the language used.

Language-specific differences between Bengali and English in bilingual speakers with AD

Cross-linguistic differences were observed on the syntactic, lexical, morphological, and inflectional measures. Several linguistic features and impairment patterns were unique to Bengali and were consistently observed for both AD participants. In terms of structural and syntactic

variables, both individuals with AD produced syntactically simpler and shorter sentences with lower embeddings in their narrative output. It is important to note that these differences were more prominent in the participants' L1 (i.e., Bengali) and were not manifested in their less proficient language, English (L2). It could be that the L2 of HC participants was not as proficient as their L1 as they were sequential bilinguals themselves, and thus the magnitude of the between group difference was smaller in L2 than in L1. Our results generally align with previous studies that have reported reduced syntactic complexity of productions in AD (e.g., De Lira et al., 2011; Forbes-McKay et al., 2013; Sajjadi et al., 2012), but differ from a bilingual study by Gómez-Ruiz et al. (2012) who noted preservation of syntactic skills in spontaneous speech productions of Catalan-Spanish bilingual speakers with AD cross-linguistically. The key finding is that at the syntactic level, unique profiles are observed across languages, which relate to the linguistic properties of the languages being compared.

In the current study, we evaluated Bengali and English which are structurally distinct from each other (with differences in word order, fluidity of word order, and morphosyntax). As noted previously, Bengali is a highly inflectional language with a fluid word order that allows dropping of the subject (i.e., pro-drop, discussed below), while English has less flexibility in terms of the structure. This is in stark contrast to a language combination such as Catalan and Spanish, which are structurally and morpho-syntactically more similar languages, as both are pro-subject drop languages sharing comparable word order and morpho-syntactic properties. Accordingly, patterns of linguistic impairments vary based on the structural differences across the languages.

Differences in Bengali language-specific markers were consistently observed for both AD participants in the lexical, morphological and inflectional domains. In terms of the lexical

features, neither person with AD showed significant differences when compared to HC in the proportion of nouns, verbs, or prepositions/postpositions produced in either in L1 or L2. However, compared to controls, participants AD03 and AD09 demonstrated the use of fewer pronouns in Bengali. As mentioned previously, Bengali is an inflectionally rich pro-drop language and allows the dropping of the subject; the subject can be inferred from the other inflected parts of speech. Contrastingly, in languages where subjects are obligatorily spelled out, such as in English, dropping the subject is not an option. Therefore, speakers with AD of those languages such as English, will opt for a cognitively less demanding option and prefer overproducing pronouns over nouns as the former is semantically vague and more frequently used, and thus might be easier to retrieve (Almor et al., 1999; Bloom, 1990). One simple deduction from this cross-linguistic observation is that when a language allows the avoidance of a linguistic feature or structure, such as subject drop in Bengali, participants with AD will avoid retrieving and producing the subject, as it is more cognitively demanding. Therefore, this finding implies that over-production of pronouns in AD, which is a characteristic feature in English, is not a relevant linguistic marker for a pro-drop language, such as Bengali.

Interestingly, we did not observe an overuse of pronouns in either AD participant in English, a feature that is commonly documented in English speakers with AD (e.g., Ahmed et al., 2013; Fraser et al., 2015). This finding may be attributed to the lower proficiency of both AD participants in their second language. Speakers who are predominantly Bengali-dominant with limited English proficiency may replicate the pronoun usage patterns of Bengali. Currently, this remains a speculation, which can be tested empirically in future with greater number of participants. Furthermore, compared to HC, both AD participants demonstrated an absence of reduplications, a feature unique to Bengali and many other Indian languages (Bose et al., 2021).

We posit that the reduced quantity of reduplications in AD could signify a potential underlying impairment with semantic processing that results in difficulties retrieving and producing reduplications. The speculation for underlying cause of the lack of reduplication in AD would need to be experimentally tested in future.

The morphological and inflectional measures indicated that, contrary to some studies involving English speakers with AD (e.g., Sajjadi et al., 2012; Ahmed et al., 2013), both AD03 and A09 performed similarly to HC in terms of the noun and verb inflections produced in Bengali revealing no inflectional impairments. They produced similar amounts of correctly inflected nouns and verbs to the bilingual control group. This is consistent with previous studies in languages with high inflectional morphology, in that, the inflectional morphology is spared in cases of language impairments (Auclair-Ouellet et al., 2019). In contrast to English where overproduction of pronouns and inflectional impairments are common, bilingual speakers with AD in the current study tended to omit the pronouns while still maintaining the correct syntax (albeit simpler) and demonstrated no difficulties with verb and noun inflections.

Collectively, these cross-linguistic differences highlight the importance of developing language-specific connected speech markers for early identification of language decline across structurally distinct languages. Furthermore, these language-specific markers of Bengali identified in the current case studies (i.e., pronouns, reduplications, noun/verb inflections) must be taken into account in evaluation and language therapy.

Connected speech characteristics in AD in relation to language experience and background

The current findings also highlight that extent to which linguistic impairments in L2 (English) vary with the participants' bilingualism profiles such as usage, proficiency, and educational background. The manifestation of cross-linguistic patterns varied between AD03 and

AD09 in certain linguistic domains. Participant AD09, in comparison to AD03 demonstrated fewer significant differences compared to the bilingual control speakers on syntactic and morphological levels in her English. Further, AD03 produced significantly fewer well-formed sentences in English while AD09 did not show any significant differences on this measure compared to controls. Moreover, AD03 produced fewer determiners relative to nouns but only in English, suggesting reduced grammaticality in his connected speech. Reduced grammatical output is a common observation in less-proficient bilingual speakers. This could potentially be related to AD03's later exposure to English informally and in schooling compared to AD09 who had greater exposure to English through her education, having earned a master's degree in English, and through her occupation as an English school teacher. As a result, she was considered a more proficient bilingual than AD03 with relatively similar proficiency in both languages.

Limitations and Future Directions

The present study acknowledges some limitations that warrant consideration. First, the use of a case study methodology was necessitated by the relatively small sample size, comprising only two bilingual individuals with AD and a matched control group. Recruitment of clinical participants is a universal challenge for any research study, this problem is exacerbated when working in locations with limited research infrastructure, lack of awareness for the importance of research and when study goal targets underrepresented populations, such as bilingual speakers who have been underexplored in previous studies. Although, this research provides a detailed profile of connected speech for Bengali and English in bilingual participants, the results must be interpreted with caution as small sample size limits the generalizability of findings.

Second, the current study focused on a novel storytelling task. Future studies must extend this research by incorporating more ecologically valid language tasks (e.g., free speech, conversations with familiar and unfamiliar partners). Furthermore, the efficient production of meaningful narratives involves interaction of both micro- and macro-linguistic levels and impairments in the structural elements can impact quality of narratives at the global level (Farivar et al., 2019). Needless to say, future research should prioritize multi-level language analysis (i.e., examination of both micro- and macro-linguistic variables and the interactions between them). These limitations collectively suggest the need for larger-scale studies with diverse samples (e.g., varying AD profiles, monolingual populations from structurally distinct languages, variables related to bilingualism and demographics, different levels of education, race, ethnicities), other language combinations (including both similar and dissimilar language structures), and multiple language tasks to corroborate and expand upon the current findings. This would allow more systematic investigations of cross-linguistic effects in bilingual speakers and establish connected speech impairment profiles in the AD population and other types of dementia pathologies.

Conclusion and Empirical Contributions

To the best of our knowledge, this research is the first ever attempt to provide detailed microlinguistic analysis of connected speech in two structurally different languages – Bengali and English – spoken by bilingual individuals with AD. By researching English and a non-European language combination, the findings add to and improve the diversity of research in the field and languages studied in bilingual speakers with AD. The findings from the two case studies revealed that domains of speech productivity, fluency, and semantic content were universally impacted, irrespective of the language used. Also, as Bengali and English differ in

syntactic, morpho-syntactic and lexical distributions, unique language-specific profiles were noted on syntactic measures, lexical distribution and inflectional properties. Thus, in contrast to English where overproduction of pronouns and inflectional impairments are common, individuals with AD in Bengali tended to omit the pronouns while still maintaining the correct yet simpler syntax and demonstrated no difficulties with verb and noun inflections. In addition, the participant with lower proficiency and usage in L2 (English) showed impoverished syntactic and morpho-syntactic features in their English output.

This research also makes important methodological contributions. First, with bilingual AD becoming increasingly relevant in clinics, research such as this should strive to move beyond restricted language tasks and provide an exhaustive analysis of connected speech to facilitate cross-linguistic comparisons and establish unique linguistic markers across various languages. As observed in the findings from this study, the patterns of linguistic changes observed in bilingual speakers with AD would depend on the linguistic level investigated, specific cross-linguistic differences between languages and bilingualism variables. Second, we provide an analysis framework for studying different language combinations, more so, for structurally different languages. For instance, the current study employed a story retelling task, which is a more natural and ecologically valid method to elicit connected speech samples in AD. Compared to structured language tasks, narrative samples are relatively less restricted, and they capture a more extensive range of language features, making them a valuable tool for language assessment in both clinical and research settings (Bose et al., 2022). Although we followed the general QPA framework to assess connected speech, we were able to adapt and incorporate Bengali-specific features into the analysis. These methodological contributions can encourage researchers to

pursue future studies by thinking about the different linguistic levels to target and tap into cross-linguistic variations.

Comprehensive connected speech analysis is receiving increasing empirical attention due to its importance in diagnosing and monitoring AD, as well as in determining therapy approaches. Given its theoretical and clinical importance, our findings emphasize the need to incorporate the evaluation of connected speech more frequently in research and focus on examining AD-related changes in languages that continue to remain under-researched. To enhance communication efficiency of bilingual speakers with AD, it is critical to first identify and establish patterns of impairments and language-specific markers of language decline, which will in turn inform clinical decision making for these individuals.

Data Availability Statement: Data collected for this study (i.e. deidentified participant data) will be made available under a signed data access agreement, after the online publication date, in response to reasonable requests from academic researchers emailed to the corresponding author.

References

Ahmed, S., Haigh, A. M. F., de Jager, C. A., & Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*, *136*(12), 3727–3737.

Alladi, S., Bak, T. H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A. K., Chaudhuri, J. R., & Kaul, S. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology*, *81*(22), 1938–1944.

Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S., & Tyler, L. K. (1999). Why Do Alzheimer Patients Have Difficulty with Pronouns? Working Memory, Semantics, and

Reference in Comprehension and Production in Alzheimer's Disease. *Brain and Language*, *67*(3), 202–227.

Ash, S., Moore, P., Vesely, L., & Grossman, M. (2007). The decline of narrative discourse in Alzheimer's disease. *Brain and Language*, *1–2*(103), 181–182.

Auclair-Ouellet, N., Pythoud, P., Koenig-Bruhin, M., & Fossard, M. (2019). Inflectional Morphology in Fluent Aphasia: A Case Study in a Highly Inflected Language. *Language and Speech*, *62*(2), 250–259.

Bak, T. H., & Robertson, I. (2017). Biology enters the scene—A new perspective on bilingualism, cognition, and dementia. *Neurobiology of Aging*, *50*, iii–iv.

Berndt, R. S., Wayland, S., Rochon, E., Saffran, E., and Schwartz, M. (2000). *Quantitative Production Analysis: A Training Manual for the Analysis of Aphasic Sentence Production.* Psychology Press.

Beveridge, M. E. L., & Bak, T. H. (2011). The languages of aphasia research: Bias and diversity. *Aphasiology*, *25*(12), 1451-1468.

Bialystok, E., Abutalebi, J., Bak, T. H., Burke, D. M., & Kroll, J. F. (2016). Aging in two languages: Implications for public health. *Ageing Research Reviews*, *27*, 56–60.

Bloom, P. (1990). Subjectless Sentences in Child Language. *Linguistic Inquiry*, *21*(4), 491–504.

Boschi, V., Catricalà, E., Consonni, M., Chesi, C., Moro, A., & Cappa, S. F. (2017). Connected Speech in Neurodegenerative Language Disorders: A Review. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00269

Bose, A., Dash, N. S., Ahmed, S., Dutta, M., Dutt, A., Nandi, R., Cheng, Y., & Tina, T. M. (2021). Connected Speech Characteristics of Bengali Speakers With Alzheimer's

Disease: Evidence for Language-Specific Diagnostic Markers. *Frontiers in Aging Neuroscience*, *13*. https://doi.org/10.3389/FNAGI.2021.707628/BIBTEX

Bose, A., Dutta, M., Dash, N. S., Nandi, R., Dutt, A., & Ahmed, S. (2022). Importance of Task Selection for Connected Speech Analysis in Patients with Alzheimer's Disease from an Ethnically Diverse Sample. *Journal of Alzheimer's Disease*, *87*(4), 1475–1481.

Calabria, M., Cattaneo, G., Marne, P., Hernández, M., Juncadella, M., Gascón-Bayarri, J., Sala, I., Lleó, A., Ortiz-Gil, J., Ugas, L., Blesa, R., Reñé, R., & Costa, A. (2017). Language deterioration in bilingual Alzheimer's disease patients: A longitudinal study. *Journal of Neurolinguistics*, *43*, 59–74.

Census of India. Language Tables from 1981 to 2011. Retrieved from https://censusindia.gov.in/census.website/data/LSI

Clarke, N., Barrick, T. R., & Garrard, P. (2021). A comparison of connected speech tasks for detecting early Alzheimer's disease and mild cognitive impairment using natural language processing and machine learning. *Frontiers in Computer Science*, *3*, 634360.

Costa, A., Calabria, M., Marne, P., Hernández, M., Juncadella, M., Gascón-Bayarri, J., Lleó, A., Ortiz-Gil, J., Ugas, L., Blesa, R., Reñé, R., & Reñé, R. (2012). On the parallel deterioration of lexico-semantic processes in the bilinguals' two languages: Evidence from Alzheimer's disease. *Neuropsychologia*, *50*, 740–753.

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia*, *40*(8), 1196–1208.

Crawford, J. R., & Garthwaite, P. H. (2006). Comparison of a single case to a control or normative sample in neuropsychology: Development of a Bayesian approach. *Cognitive Neuropsychology*, 24, 343-372.

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case-controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology*, *27*(3), 245–260.

Cuetos, F., Arango-Lasprilla, J. C., Uribe, C., Valencia, C., & Lopera, F. (2007). Linguistic changes in verbal expression: a preclinical marker of Alzheimer's disease. *Journal of the International Neuropsychological Society*, *13*(3), 433-439.

Das, S. K., Banerjee, T. K., Mukherjee, C. S., Bose, P., Biswas, A., Hazra, A., ... & Roy, T. (2006). An urban community-based study of cognitive function among non-demented elderly population in India. *Neurol Asia*, 11, 37-48.

De Lira, J. O., Ortiz, K. Z., Campanha, A. C., Bertolucci, P. H. F., & Minett, T. S. C. (2011). Microlinguistic aspects of the oral narrative in patients with Alzheimer's disease. *International Psychogeriatrics*, *23*(3), 404–412.

De Picciotto, J., Friedland, D., & De Picciotto, J. (2001). Verbal Fluency in Elderly Bilingual Speakers: Normative Data and Preliminary Application to Alzheimer's Disease. *Folia Phoniatrica et Logopaedica*, *53*, 145–152.

Dunn, A. L., & Fox Tree, J. E. (2009). A quick, gradient Bilingual Dominance Scale*. *Bilingualism: Language and Cognition*, *12*(3), 273–289.

Ellajosyula, R., Narayanan, J., & Patterson, K. (2020). Striking loss of second language in bilingual patients with semantic dementia. *Journal of Neurology*, *267*(2), 551–560.

Farivar, M., Anaraki, Z. G., Derakhshandeh, F., Baharloei, N., & Poorjavad, M. (2019). Narrative discourse in Persian-speaking patients with mild Alzheimer's disease. *Dementia & Neuropsychologia*, *13*(2), 225–231.

Filiou, R.P., Bier, N., Slegers, A., Houzé, B., Belchior, P., & Brambati, S. M. (2020). Connected speech assessment in the early detection of Alzheimer's disease and mild cognitive impairment: A scoping review. *Aphasiology*, *34*(6), 723–755.

Forbes-McKay, K., Shanks, M. F., & Venneri, A. (2013). Profiling spontaneous speech decline in Alzheimer's disease: A longitudinal study. *Acta Neuropsychiatrica*, *25*(6), 320–327.

Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2015). Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, *49*(2), 407–422.

Friedland, D., & Miller, N. (2010). Language mixing in bilingual speakers with Alzheimer's dementia: A conversation analysis approach. *Aphasiology,13*(4–5), 427–444.

Gayraud, F., Lee, H.-R., & Barkat-Defradas, M. (2010). Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects. *Clinical Linguistics & Phonetics*, *25*(3), 198-209.

Gollan, T. H., Salmon, D. P., Montoya, R. I., & da Pena, E. (2010). Accessibility of the nondominant language in picture naming: A counterintuitive effect of dementia on bilingual language production. *Neuropsychologia*, *48*(5), 1356–1366.

Gómez-Ruiz, I., Aguilar-Alonso, Á., & Espasa, M. A. (2012). Language impairment in Catalan-Spanish bilinguals with Alzheimer's disease. *Journal of Neurolinguistics*, *25*(6), 552–566.

Grossman, M., & Rhee, J. (2001). Cognitive resources during sentence processing in Alzheimer's disease. *Neuropsychologia*, *39*(13), 1419–1431.

Hsieh, S., Schubert, S., Hoon, C., Mioshi, E., & Hodges, J. R. (2013). Validation of the Addenbrooke's Cognitive Examination III in Frontotemporal Dementia and Alzheimer's Disease. *Dementia and Geriatric Cognitive Disorders*, *36*(3–4), 242–250.

Hyltenstam, K., & Obler, L. K. (1989). *Bilingualism Across the Lifespan: Aspects of Acquisition, Maturity and Loss*. Cambridge University Press.

Hyltenstam, K. & Stroud, C. (1993). Second language regression in Alzheimer's dementia. In K. Hyltenstam, and A. Viberg (Eds.), *Progression and Regression in Language. Socio-cultural, Neuropsychological and Linguistic Perspectives* (pp. 222-242). Cambridge University Press.

Ivanova, I., Salmon, D. P., & Gollan, T. H. (2014). Which Language Declines More? Longitudinal versus Cross-sectional Decline of Picture Naming in Bilinguals with Alzheimer's Disease. *Journal of the International Neuropsychological Society*, *20*(5), 534–546.

Kavé, G., & Goral, M. (2016). Word retrieval in picture descriptions produced by individuals with Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology*, *38*(9), 958–966.

Kavé, G., Leonard, C., Cupit, J., & Rochon, E. (2007). Structurally well-formed narrative production in the face of severe conceptual deterioration: A longitudinal case study of a woman with semantic dementia. *Journal of Neurolinguistics*, *20*(2), 161–177.

Kavé, G., & Levy, Y. (2003). Morphology in Picture Descriptions Provided by Persons With Alzheimer's Disease. *Journal of Speech Language and Hearing Research, 46* (2), 341-352.

Lehmann, C. (1982). Directions for interlinear morphemic translations. *Folia Linguistica, 16,* 199-224.

Lerman, A., Goral, M., & Obler, L. K. (2019). The complex relationship between pre-stroke and post-stroke language abilities in multilingual individuals with aphasia. *Aphasiology*, *34*(11), 1319–1340.

Mathuranath, P. S., George, A., Cherian, P. J., Mathew, R., & Sarma, P. S. (2005). Instrumental activities of daily living scale for dementia screening in elderly people. *International Psychogeriatrics*, *17*(3), 461–474.

Mayer, M. (1969). *Frog, where are you?.* Dial Press.

McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia*, *7*(3), 263–269.

Meguro, K., Senaha, M. L., Caramelli, P., Ishizaki, J., Chubacci, R. Y., Meguro, M., Ambo, H., Nitrini, R., & Yamadori, A. (2003). Language deterioration in four Japanese–Portuguese bilingual patients with Alzheimer's disease: A trans-cultural study of Japanese elderly immigrants in Brazil. *Psychogeriatrics*, *3*(2), 63–68.

Morris, J. C. (1993). The Clinical Dementia Rating (CDR): Current version and scoring rules. *Neurology, 43,* 2412–2414.

Mueller, K. D., Hermann, B., Mecollari, J., & Turkstra, L. S. (2018). Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture

description tasks. *Journal of Clinical and Experimental Neuropsychology*, *40*(9), 917-939.

Muñoz, M. L., Marquardt, T. P., & Copeland, G. (1999). A comparison of the codeswitching patterns of aphasic and neurologically normal bilingual speakers of English and Spanish. *Brain and Language*, 66(2), 249-274.

Nanchen, G., Abutalebi, J., Assal, F., Manchon, M., Démonet, J. F., & Annoni, J. M. (2017). Second language performances in elderly bilinguals and individuals with dementia: The role of L2 immersion. *Journal of Neurolinguistics*, *43*, 49–58.

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

Noack, R., & Gamio, L. (2015). The world's languages, in 7 maps and charts. *The Washington Post*, *4*(23), 65-70.

Paradis, M. (1998). The other side of language: Pragmatic competence. *Journal of Neurolinguistics*, *11*(1), 1–10.

Patra, A., Bose, A., & Marinis, T. (2020). Lexical and Cognitive Underpinnings of Verbal Fluency: Evidence from Bengali-English Bilingual Aphasia. *Behavioral Sciences*, *10*(10), 155.

Petti, U., Baker, S., & Korhonen, A. (2020). A systematic literature review of automatic Alzheimer's disease detection from speech and language. *Journal of the American Medical Informatics Association*, *27*(11), 1784–1797.

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

Sajjadi, S. A., Patterson, K., Tomek, M., & Nestor, P. J. (2012). *Abnormalities of connected speech in semantic dementia vs Alzheimer's disease*. *Aphasiology*, *26*(6), 847-866.

Salvatierra, J., Rosselli, M., Acevedo, A., & Duara, R. (2007). Verbal Fluency in Bilingual Spanish/English Alzheimer's Disease Patients. *American Journal of Alzheimer's Disease & Other Dementias, 22*(3), 190-201.

Slegers, A., Filiou, R.-P., Montembeault, M., & Brambati, S. M. (2018). Connected Speech Features from Picture Description in Alzheimer's Disease: A Systematic Review. *Journal of Alzheimer's Disease*, *65*(2), 519-542.

Stark, B. C., Bryant, L., Themistocleous, C., den Ouden, D.-B., & Roberts, A. C. (2022). Best practice guidelines for reporting spoken discourse in aphasia and neurogenic communication disorders. *Aphasiology,* 1-24.

Stilwell, B. L., Dow, R. M., Lamers, C., & Woods, R. T. (2016). Language changes in bilingual individuals with Alzheimer's disease. *International Journal of Language & Communication Disorders*, *51*(2), 113–127.

Thompson, H. R. (2020). *Bengali: A Comprehensive Grammar*. Routledge.

U.K. Census Bureau. (2011). Language in England and Wales: 2011, Office for National Statistics, 2011 Census. https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/language/articles/language in england and wales/2013-03-04

U.S. Census Bureau. (2017). Bengali speakers, 2012–2016 American Community Survey 5-year estimate. https://www.census. gov/library/visualizations/2017/comm/bengali.html

Weiner, M. F., Neubecker, K. E., Mary, B. A., Bret, E., & Hynan, L. S. (2008). Language in Alzheimer's Disease. *Journal of Clinical Psychiatry*, *69*(8), 1223–1227.

FIGURES

Figure 1. The top two panels illustrate the variables (speech rate, fluency disruptions, idea density and idea efficiency) that were affected both in Bengali and English for AD03 and AD09 in comparison to the neurologically healthy controls (HC). The lower bottom panel illustrates variables (mean sentence length and proportion of pronouns) that were only affected in Bengali for both participants. An asterix indicates significant difference from control group.

SUPPLEMENTARY MATERIALS

Supplementary Table 1. A summary of relevant experimental speech and language studies in bilingual AD (adapted from Stilwell et al., 2016 and Calabria et al., 2017).

Supplementary Table 2. Best Practice Guidelines for Reporting Spoken Discourse in Aphasia and Neurogenic Communication Disorders (Stark et al., 2021).

Supplementary Table 3. A complete list and definitions of variables derived from the Quantitative Production Analysis (QPA) and Correct Information Unit (CIU) analyses.

Table 1. Cross linguistic differences in Bengali and English.

Linguistic features	**Bengali**	**English**
Syntactic and structural characteristics		
Canonical word order	SOV (Subject-Object-Verb)	SVO (Subject-Verb-Object)
Flexibility of word order	Fluid, at least for canonical forms; can be extended to other forms	Rigid
Branching	Left branching	Right branching
Passive constructions	Rare	Common
Pro-drop	Yes (subject); Object (rarely dropped)	No
Negation	Postverbal (preverbal in conditional constructions)	Preverbal (with added auxiliary)
Complementizer	Optional (sense of the complementizer is implicit)	Optional (that)
Lexical characteristics		
Open class words		
Nouns	Present	Present
Verbs	Present	Present
Copular verbs	Not used in simple declarative present tense, available in other tenses.	One in number (to be)
Adjectives	Present, inflected for comparative and superlative	Present, uninflected
Adverbs	Present	Present
Reduplication	Pervasive use	Rare use
Closed class words		
Pronouns	Present: overt and null	Present: overt in most contexts
Prepositions	Absent	Present
Postpositions	Present	Absent
Auxiliaries	Absent as independent form, but marked in the main inflected finite verb	Present
Clitics	Present (mostly emphatic in sense)	Absent (but debated)
Modals	Absent as independent form but marked by inflections on verbs.	Present
Articles	Absent as independent form, but marked by inflections on nouns.	Present
Morphological characteristics		
Nominal morphology	Highly inflected morphology	Limited inflectional morphology

Base form (uninflected)	Yes	Yes
Nouns can be inflected for:		
Number	Marked with suffix, and words of multitude	Marked with suffix
Definiteness	Marked with suffix	Use of determiner
Case	Marked with suffix and postposition	Not marked
Particles	Marked with suffix	
Pronoun morphology	Some inflected morphology	Limited inflectional morphology
Pronouns can be inflected for:		
Number	No	Yes
Person	Yes	Yes
Case	Yes	Yes but limited
Verbal morphology	Highly inflected morphology	Limited inflectional morphology
Base form (uninflected)	No (with the exception of unmarked 2nd person, imperative)	Yes
Verbs can be inflected for:		
Mood	Marked with suffix	Marked with auxiliary and suffix
Tense	Marked with suffix	Marked with suffix
Aspect	Marked with suffix	Marked with auxiliary and suffix (-ing)
Person	Marked with suffix	Marked with suffix but limited
Number	Not marked	Marked with suffix, limited to third person singular
Gender	Not marked	
Honorification	Marked with suffix	Not marked
Particles (emphatic & negative)	Marked with suffix	Expressed analytically
Auxiliaries can be inflected for:		
Mood	Not applicable	
Tense	Not applicable	Marked with suffix
Number	Not applicable	Marked with suffix but limited
Person	Not applicable	Marked with suffix but limited

Table 2. Demographics and neuropsychological assessment data for participants with Alzheimer's Disease (AD03 and AD09) and neurologically healthy controls (HC).

Demographic	AD03	AD09	HC (n = 8) Mean (SD)
Age at the time of study (years)	76	56	71.4 (4.2)
Education (years)	14	17	16 (1.2)
Age at the onset of symptoms (years)	73	52	-
Sex	M	F	M=4 F=4
Handedness	R	R	R
Occupation	Retired Electrical Supervisor	Retired English Teacher	
Neuropsychological measures			
Bengali version of Addenbrooke's Cognitive Examination-III[a]			
Total (/100)	**40**	**31**	92.6 (2.5)
Attention (/18)	**10**	**8**	17.6 (0.7)
Memory (/26)	**9**	**4**	25.4 (0.7)
Fluency (/14)	**1**	**1**	8.0 (1.1)
Language (/26)	**12**	**15**	25.9 (0.4)
Visuoconstructional (/16)	**8**	**3**	15.8 (0.5)
Kolkata Cognitive Screening Battery (KCSB)[b]			
Total (/30)	20*	16*	30 (0)
Clinical Dementia Rating Scale[c]			
Rating	1*	1*	0
Instrumental Activities of Daily Living (IADL-EDR)[d]			
% Impairment	50*	36*	0
Verbal Fluency Scores			
Semantic (Animals)			
Bengali	8	**5**	14.4 (3.6)
English	6	7	12.3 (4.2)
Letter			
Bengali (Average)	15	6.6	12.7 (3.2)
K	17	6	12.6 (3.7)
P	15	**4**	12.3 (3.2)
M	13	10	13.1 (2.8)
English (Average)	11.3	5	12.3 (4.5)
F	9	5	12.4 (4)
A	18	**	13 (4.8)
S	7	**	11.6 (4.7)

a- Hsieh et al. (2013)
b- Adapted Bengali version of the Mini-Mental State Examination, BMSE (Das et al., 2006)
c- Morris (1993); CDR score 0 = no dementia; 0.5 = questionable dementia; 1 = mild dementia; 2 = moderate dementia; 3 = severe dementia
d- Mathuranath et al. (2005); IADL score of > 16 indicate impairment, with higher scores indicating greater impairment.
Values in bold font = Score for each participant with AD was significantly lower compared to the control group mean ($p < 0.054$ on SingleBayes_ES; Crawford et al., 2010)
*Statistical difference unable to be computed due to a standard deviation of 0
**Participant was unable to complete these tests.

Table 3. Bilingualism Profiles of Participants with Alzheimer's Disease (AD03 and AD09) and neurologically healthy controls (HC).

Bilingualism measures	AD03		AD09		HC (n=8)			
	Bengali (L1)	English (L2)	Bengali (L1)	English (L2)	Bengali (L1)		English (L2)	
					Mean	SD	Mean	SD
Language Acquisition History								
Language Acquisition[a] (/7)	6	1	6.5	0.5	7	0	0	0
Onset of exposure (age in years)								
Formal (in classrooms, schooling)	5	6	3	5	3.9	1.4	6.9	1.6
Informal (interactions with friends, neighbors)	Since birth	11	Since birth	5	Since birth	0	10.9	3.2
Medium of instruction in education ('✓' indicates language(s) of instruction in school)								
Primary school	✓		✓	✓				
Secondary school	✓		✓	✓				
High school	✓		✓	✓				
Undergraduate	✓	✓	✓	✓				
Postgraduate	n/a	n/a	✓	✓				
Current Language Proficiency Rating[b] (/7)								
Average	**7**	**2.38**	**7**	**5.88**	**7**	**0**	**5.08**	**1.58**
Speaking	7	2	7	5.5	7	0	4.44	1.92
Comprehension	7	2.5	7	6	7	0	4.75	1.6
Reading	7	3	7	6	7	0	5.75	1.49
Writing	7	2	7	6	7	0	5.38	1.85
Bilingual Language Dominance[c] (/31)								
Total score	**26**	**-4**	**23**	**7**	**24**	**1.93**	**4.75**	**4.71**
Current language usage and frequency of usage[d]								
Average scores	**5**	**2**	**5**	**3.75**	**4.83**	**0.24**	**2.86**	**0.49**
At home	5	1	5	3	5	0	1.63	0.74
At community gatherings	5	1	5	3	5	0	2.25	0.71
At social gatherings	5	2	5	4	4.75	0.71	2.75	0.89
At formal situations	5	4	5	5	4.75	0.71	3.14	1.07
With Friends	5	2	5	4	5	0	2.13	0.64
Telecommunication	5	1	5	3	5	0	2.75	0.89
Reading	5	3	5	4	5	0	3.88	0.64
Writing	5	2	5	4	4.13	0.64	4.5	0.76

a- Munoz et al. (1999), adapted by Patra et al. (2020); Maximum score of 7, with higher scores indicating greater immersion in that language in childhood.

b- Munoz et al. (1999), adapted by Patra et al. (2020); Maximum score of 7, with 0 indicating no proficiency and 7 indicating native-like proficiency.

c- Dunn & Fox Tree (2009), adapted by Patra et al. (2020); Maximum score of 31, with a higher score indicating greater dominance in that language compared to the other language.

d- Rating of 1 indicates 'not at all' and 5 indicates 'very often.'

Table 4. Illustrative samples of the "Frog, where are you?" story narrations in English and Bengali by AD03 and AD09, and one control participant (HC05). The excerpts are the first five utterances from their transcripts. English narration is transcribed orthographically. Bengali transcription follows five tier system: Tier 1 (Bengali orthographic transcription); Tier 2 (transliteration with Indic Roman script); Tier 3 (Leipzig interlinear glossing*); Tier 4 (Word-by-word translation in English); Tier 5 (possible meaning in English).

AD03 (English)

1) Boy is with the frog
2) Dog
3) Boy sleeping
4) And the dog trying to go away trying to go out trying to go out
5) The boy get up not seeing the dog

AD09 (English)

1) The moon can be seen from the window
2) And there is a there is a window below it
3) And the child was there with um his um dog
4) And and on the back the oh on the back there was ka ka cushion and and which the um child was sleeping
5) And then the child come to the come to the and come to the went to the and started talking to him

HC05 (English)

1) A child had two pets a frog and a dog
2) While the dog roamed all over the house, the frog remained confined in a jar
3) One day it so happened the frog went out of the jar
4) The child was very anxious not to find his dear frog
5) He looked for it within the house he turned many things

AD03 (Bengali)

(1)	একটা	ছেলে
	ekṭā	chele
	ADJ-DET	NN-3.M.SG
	one	boy
	'A boy'	

(2)	এটা	একটা	কুকুর ...	কুকুর	ও	বা	সামথিং	এলস্...	বেড়াল
	eṭā	ekṭā	kukur...	kukur	o	bā	something	else...	beṛāl
	PN-DEM	ADJ-DET	NN-3.N.SG	NN-3.N.SG	CONJ	COMP	PN-INDF.N	ADV	NN-3.N.SG
	this	one	dog...	dog	and	or	something	else...	cat
	This is a dog and a cat or something else'								

(3)	এটা	ফ্রগ ...	হ্যাঁ	ব্যাঙ
	eṭā	frog ...	hyã	byāṅ
	PN-DEM	NN-3.N.SG...	FILL	NN-3.N.SG
	this	frog ...	yes	frog
	'This is a frog'			

(4)	কুকুর	ছেলে	ডগ ...	এই	তিন	জন
	kukur	chele	ḍog...	ei	tin	jan
	NN-3.N.SG	NN-3.M.SG	NN-3.N.SG	PN-DEM	ADJ	QNTF

	dog	boy	dog	this	three	persons
	‘Dog, boy and dog these three people’					

(5)	আরেকটা	ঘর
	ārekṭā	ghar
	ADJ-CONJ.DET	NN-N.SG
	another	room
	‘Another room’	

AD09 (Bengali)

(1)	ঘরে	জানলা
	ghare	jānlā
	NN-N.SG-LOC	NN-N.SG
	in room	window
	‘Window in the room’	

(2)	বাইরে	উমম	চাঁদ	দেখা যাচ্ছে
	bāire	umm	cãd	dekhā_ẏācche
	ADV-LOC	FILL	NN-M.SG	FV-PRS-PROG-PASS
	at outside	umm	moon	is being seen
	‘The moon is being seen at outside’			

(3)	নিচে	এখানে	একটা	কুকুর	অঅঅ	বসে_আছে
	niche	ekhāne	ekṭā	kukur	aaa	base_āche
	POST-LOC	ADV-LOC	ADJ-DET	NN-3.N.SG	FILL	FV-PRS-3
	At underneath	here	one	dog	aaa	is sitting
	‘A dog is sitting here at underneath’					

(4)	তার	নিচে	এখানে	একটা	ব্যাঙের	মতো	ব্যাঙ	বসে আছে
	tār	nice	ekhāne	ekṭā	byāṅer	mato…	byāṅ	base_āche
	PN-3.N.SG	POST	ADV-LOC	ADJ-DET	NN-3.N.SG-GEN	ADJ	NN-3.N.SG	FV-PRS-3
	of it	under	here	one	frog	like…	frog	sits
	‘A frog is sitting under it’							

(5)	পেছনে	খাটটা	রয়েছে
	pechane	khāṭṭā	rayeche
	POST	NN-3.N.SG-DET	FV-PRS-PRF-3
	behind	the cot	is present
	‘The cot is behind’		

HC05 (Bengali)

(1)	একটি	বাচ্চা
	ekṭi	bācchā
	ADJ-DET	NN-3.N.SG
	One	child
	‘One child’	

(2)	তার	ছিল	এক	কুকুর	আর	একটি	ব্যাঙ
	tār	chhila	ek	kukur	ār	ekṭi	byāṅ
	PN-3.N.SG	FV-PST	ADJ	NN-3.N.SG	CONJ	ADJ-DET	NN-3.N.SG

	His	had	one	dog	and	one	frog
	‘He had one dog and one frog’						

(3)	দুটোকেই	সে	খুব	ভালোবাসতো
	duṭokei	se	khub	bhālobāsto
	NN-3.N.PL.DET.EMPH	PN-3.M.SG	ADV	FV-PST.HAB.3
	to both of them	he	deeply	love used to
	‘He used to love both of them deeply’			

(4)	কুকুর	বাইরেই	ঘুরে_ বেড়াত	আর	ব্যাঙটা	থাকত	একটা	জারের	মধ্যে
	kukur	bāirei	ghure_beṛāta	ār	byāṅṭā	thākta	ekṭā	jārer	madhye
	NN-3.N.SG	ADV-LOC-EMPH	FV-PST.HAB.3	CONJ	NN-3.N.SG.DET	FV-PST.HAB.3	ADJ-DET	NN-3.N.SG	POST
	Dog	outside only	used to roam	and	the frog	used to live	one	of jar	inside
	‘Dog used to roam only outside and the frog used to live inside a jar’								

(5)	যখন	ছেলেটি	ঘুমোচ্ছিল	ব্যাঙটি	সেই	জার	থেকে	বেরিয়ে পড়ল
	ẏakhan	chheleṭi	ghumochchhila	byāṅṭi	sei	jār	theke	beriye_paṛla
	ADV	NN-3.M.SG.DET	FV-PST.PROG.3	NN-3.N.SG.DET	PN-DEM	NN-3.N.SG	POST	FV-PST.3
	When	the boy	was sleeping	the frog	that	jar	from	went out
	‘When the boy was sleeping the frog got out of the jar’							

*Leipzig interlinear glossing codes: ADJ=Adjective; ADV=Adverb; COMP=Complementizer; CONJ=Conjunction; DAT=Dative; DEM=Demonstrative; DET=Determiner; EMPH=Emphatic; F=Feminine; FILL=Filler; FV=Finite verb; GEN=Genitive; INDF=Indefinite; INF=Indefinite; LOC=Locative; M=Masculine; N=Neuter; NN=Noun; PART=Particle; PASS=Passive; PN=Pronoun; POST= Postposition; PRF=Perfect; PROG=Progressive; PRS=Present; PST=Past; QNTF= Quantifier; SG=Singular; 1, 2, 3=1st, 2nd and 3rd person, respectively.

Table 5. Definition of connected speech variables extracted. Checkmarks with an * (✓*) indicate cross-linguistic differences expected in these variables due to the differences in linguistic properties between Bengali and English.

Variable Name	Variable Definition	Relevant for Bengali	Relevant for English
Speech rate			
Words per minute	Speech rate was defined as the number of words per minute, including repetitions, corrections, restarts, and paraphasias as well as patients' direct responses to the questions, but excluding indistinct strings of phonemes and discourse markers.	✓	✓
Structural and syntactic measures			
Mean sentence length	The average number of words produced per sentence.	✓*	✓*
Proportion of well-formed sentences	Total number of well-formed sentences divided by the total number of sentences.	✓*	✓*
Embedding index	Total number of embeddings divided by the total number of sentences. This measure provides a quantification for utterance complexity. Fewer embedding would imply less complex utterances.	✓	✓
Lexical measures			
Proportion of open class words	Total number of open class words divided by total number of narrative words.	✓	✓
Proportion of closed class words	Total number of closed class words divided by total number of narrative words.	✓*	✓*
Proportion of noun (N/all NW)	Total number of nouns divided by total number of narrative words.	✓	✓
Proportion of pronoun (P/all NW)	Total number of personal pronouns divided by total number of narrative words.	✓	✓
Proportion of verb (V/all NW)	Total number of verbs divided by total number of narrative words.	✓	✓
Proportion of postposition (Po/all NW)	Total number of postposition divided by total number of narrative words.	✓*	NR
Proportion of preposition (Pr/all NW)	Total number of preposition divided by total number of narrative words.	NR	✓*
Number of reduplication	Total sum of all reduplications.	✓	
Morphological and inflectional measures			
Nouns inflections			
Noun inflection index	Total number of appropriately inflected nouns to the number of nouns that are possible to be inflected.	✓*	NR
Determiner index (determiner/noun)	Proportion of determiners produced in obligatory contexts.	NR	✓*
Verb inflections			
Verb inflection index	Total number of appropriately inflected verbs to the number of verbs that are possible to be inflected. This is conceptually similar to the verb inflection index of the QPA in English.	✓	✓
Auxiliary Index (Aux/MV-1)	It is an index of elaboration of the auxiliary in matrix verbs and derived by taking the ratio of total auxiliary score to total number of matrix verb minus 1.	NR	✓*
Semantic measures			
Number of CIU	The total number of intelligible, accurate and informative words that were relevant to the "Frog, where are you?" story (Nicholas & Brookshire, 1993)	✓	✓
CIU% (Idea density)	Total number of CIUs (i.e., semantic units) divided by the total number of words used in the sample.	✓	✓
CIUs per minute (Idea efficiency)	Total number of CIUs (i.e., semantic units) divided by the duration of the sample used for calculation of the CIUs.	✓	✓
Measures of spontaneity and fluency disruptions			
Total count of disruption of spontaneity and fluency	Sum of count of repetitions, revisions and reformulations.	✓	✓
NR = Not Relevant			

Table 6. Results from the single-subject statistical analysis of connected speech variables for participants with Alzheimer's Disease (AD03 and AD09) in relation to neurologically healthy controls (HC).

	AD03		AD09		HC	
Variable Name	**Bengali**	**English**	**Bengali**	**English**	**Bengali M (SD)**	**English M (SD)**
Speech rate						
Words per minute	**48.21***	**44.04***	**42.17***	**45.94***	135.92 (31.89)	107.75 (10.74)
Structural and syntactic measures						
Mean sentence length	**4.31***	5.77	**3.34***	6.70	7.59 (0.73)	8.95 (2.16)
Proportion of well-formed sentences	0.88	**0.46***	0.87	0.78	0.92 (0.07)	0.89 (0.12)
Embedding index	**0.00***	0.04	**0.00***	0.09	0.50 (0.18)	0.21 (0.14)
Lexical measures						
Proportion of open class words	0.78	0.47	**0.86***	**0.39***	0.76 (0.04)	0.47 (0.03)
Proportion of closed class words	0.22	0.53	**0.14***	**0.61***	0.24 (0.04)	0.53 (0.03)
Proportion of noun (N/all NW)	0.31	0.23	0.32	0.19	0.33 (0.03)	0.22 (0.03)
Proportion of pronoun (P/all NW)	**0.02***	0.10	**0.04***	0.15	0.10 (0.03)	0.14 (0.03)
Proportion of verb (V/all NW)	0.25	0.24	0.28	0.18	0.24 (0.04)	0.18 (0.03)
Proportion of postposition (Po/all NW)	0.06	NR	0.12	NR	0.08 (0.02)	NR
Proportion of preposition (Pr/all NW)	NR	0.10	NR	0.11	NR	0.11 (0.02)
Number of reduplications	**0***	NR	**0***	NR	3 (2.78)	NR
Morphological and inflectional measures						
Nouns inflections						
Noun inflection index	1.00	NR	0.92	NR	1.00 (0.00)	NR
Determiner index	NR	**0.82***	NR	0.96	NR	0.98 (0.03)
Verb inflections						
Verb inflection index	1	0.89	1	1.00	1.00 (0.00)	0.94 (0.07)
Auxiliary Index	NR	1.07	NR	1.23	NR	1.00 (0.12)
Semantic measures						
Number of CIU	154	151.00	**143***	**136.00***	161.63 (5.71)	149.43 (3.74)
CIU% (Idea density)	**63.90***	**59.92***	**49.48***	**51.71***	90.87 (5.54)	89.72 (4.86)
CIUs per minute (Idea efficiency)	**47.87***	**29.04***	**33.78***	**24.21***	98.24 (15.93)	97.91 (11.16)
Measures of spontaneity and fluency disruptions						
Total count of disruption of spontaneity and fluency	**12***	**11***	**21***	**35***	3.13 (2.90)	2.71 (2.43)

Values in bold with asterisk = Score for each participant with AD was significantly lower compared to the control group mean ($p < 0.054$ on SingleBayes_ES; Crawford et al., 2010); NR = Not Relevant; N= Number of nouns; NW=Number of narrative words; P=Number of pronouns; V=Number of verbs; Po=Number of postpositions; Pr=Number of prepositions; CIU=Correct Information Units.

Figure 1. The top two panels illustrate the variables (speech rate, fluency disruptions, idea density and idea efficiency) that were affected both in Bengali and English for AD03 and AD09 in comparison of the healthy control group (HC). The lower bottom panel illustrates variables (mean sentence length and proportion of pronouns) that were only affected in Bengali for both participants. Asterix indicates significant difference from control group.