Clinical Linguistics & Phonetics

ISSN: 0269-9206 (Print) 1464-5076 (Online) Journal homepage: www.tandfonline.com/journals/iclp20

# Identifying novel linguistic biomarkers of mild cognitive impairment in Mandarin-speaking older adults: A quantitative syntactic approach

Tsy Yih, Yiran Yang, Mu Yang, Haitao Liu & Lihe Huang

**To cite this article:** Tsy Yih, Yiran Yang, Mu Yang, Haitao Liu & Lihe Huang (16 Oct 2025): Identifying novel linguistic biomarkers of mild cognitive impairment in Mandarin-speaking older adults: A quantitative syntactic approach, Clinical Linguistics & Phonetics, DOI: 10.1080/02699206.2025.2571660

**To link to this article:** https://doi.org/10.1080/02699206.2025.2571660

Published online: 16 Oct 2025.

CLINICAL LINGUISTICS & PHONETICS
https://doi.org/10.1080/02699206.2025.2571660

Check for updates

# Identifying novel linguistic biomarkers of mild cognitive impairment in Mandarin-speaking older adults: A quantitative syntactic approach

Tsy Yih [a*], Yiran Yang [a], Mu Yang [b], Haitao Liu [c], and Lihe Huang [a]

[a]Research Center for Ageing, Language and Care, School of Foreign Studies, Tongji University, Shanghai, China; [b]School of International Studies, Zhejiang University, Hangzhou, China; [c]College of Foreign Languages and Literature, Fudan University, Shanghai, China

**ABSTRACT**

Mild Cognitive Impairment (MCI) is an early symptom of Alzheimer's disease, commonly observed in older adults. The use of low-cost language biomarkers is becoming an emerging trend. This study aims to investigate whether recently proposed measures in the field of Quantitative Syntax and their combinations have the potential to serve as biomarkers for distinguishing between the MCI and cognitively normal (CN) groups. A portion of the Chinese corpus MCGD (CN = 25, MCI = 16) was used. The elderly participants performed a sequential picture description task to produce connected speech. The transcription and annotation were semi-automatically conducted and manually checked. Eleven dependency-based syntactic features were calculated. We assessed the discriminability of both univariate features and multivariate feature combinations using support vector machine. Results show that all features can be grouped into four clusters. Most measures within the largest cluster demonstrate high intercorrelations and are statistically significant in distinguishing between the MCI and CN groups. Among these, mean dependency distance (MDD) exhibits the strongest discriminative ability (AUC = 0.791 [0.610, 0.944]). Two hierarchical features have relatively weaker performance, while dependency direction indicators show almost no group differentiability. Several feature combinations identified slightly improved performance, but the difference was not statistically significant. Our findings suggest that the classic syntactic biomarker MDD remains the best-performing measure for distinguishing between MCI and CN for Mandarin-speaking older adults, while most dependency-based syntactic measures can serve as alternative markers. In the future, combining MDD with features in other domains holds promising potential for early diagnosis.

**ARTICLE HISTORY**
Received 10 February 2025
Revised 2 July 2025
Accepted 26 September 2025

**KEYWORDS**
Mild cognitive impairment; quantitative syntax; linguistic biomarker; machine learning; gerontolinguistics

## Introduction

Mild Cognitive Impairment (MCI) is a common condition in older adults, which can later develop into neurodegenerative diseases such as Alzheimer's disease (AD) and other types of dementia (Cummings, 2020: Ch. 4). According to epidemiological predictions, the

**CONTACT** Lihe Huang cranehlh@tongji.edu.cn Research Center for Ageing, Language and Care, School of Foreign Studies, Tongji University, No. 1239 Siping Road, Yangpu District, Shanghai 200092, China

*Tsy Yih is the transliteration of the name of the first author in his mother tongue, Shanghai Wu Chinese. He is also known as ZI YE in Mandarin pinyin.

overall prevalence of MCI in the population aged 60 and above in China was estimated to be about 15.5% (Jia et al., 2020) to 19.6% (Wu et al., 2024). Early detection can provide strong support for early intervention and treatment.

Traditionally, fluids (Blanco et al., 2023; Gaur et al., 2023; Qu et al., 2021) and neuroimaging (Li et al., 2022; Ruan et al., 2016) biomarkers were commonly used as the basis for early MCI diagnosis. However, these methods are costly, invasive, and difficult to implement in community-based early screening. Clinical linguistic and phonetic studies have suggested that subtle changes in spoken language can be detected in the early phases of cognitive decline, particularly in terms of both acoustic properties and textual structure (Duboisdindien, 2024; Huang et al., 2023; Walker et al., 2021). Consequently, in recent years, low-cost detection methods based on connected speech production have gradually become an emerging research topic (Filiou et al., 2020; Huang et al., 2024; Mueller et al., 2018). Several diagnostic criteria from relevant organisations also highlight the critical role of linguistic features as reliable biomarkers for dementia and cognitive impairment in clinical practice (DSM-5, American Psychiatric Association, 2013; IWG-2; Dubois et al., 2014; Chinese Diagnosis and Treatment of Dementia and Cognitive Impairment Guide; Jia, 2016). In particular, with the rapid development of AI technologies such as natural language processing and machine learning, it is easy to perform transcription, automatic extraction of measures, and construct models that integrate multiple features to predict the cognitive status of older adults.

Previous studies have found that as cognitive decline and dementia symptoms worsen, older adults' language abilities deteriorate in multiple aspects. In the severe stages of AD, this deterioration is very evident (Huang, 2024). In contrast, in the early stages of MCI, identifying subtle changes in language features remains a significant research challenge, and several studies indicate that language decline usually occurs at the phonetic (Kleiman & Galvin, 2024; Pistono et al., 2016; Qiao et al., 2020; Xu et al., 2025) and lexical-semantic levels (Che & Huang, 2025; Stark et al., 2025). However, dedicated research specifically addressing how syntactic abilities change with cognitive decline in older adults remains relatively scarce (Yih et al., 2025). Therefore, in this study, we attempt to focus on the syntactic abilities of older adults with MCI, using novel quantitative measures combined with machine learning methods to differentiate between the MCI and cognitively normal (CN) populations in the syntactic layer, further laying a technological foundation for large-scale, language-based automatic cognitive screening.

## Syntactic complexity measures as MCI biomarkers

It has long been observed that syntactic ability tends to deteriorate alongside cognitive decline in older adults (Snowdon et al., 1996). As a key component of the multidimensional construct of language ability, syntactic ability is typically manifested in a speaker's capacity to organise sentences and accurately produce complex syntactic structures, such as complex noun phrases (often with multiple modifiers) and complex sentences (containing embedded or coordinated clauses). These structures are generally characterised by greater length, a higher number of linguistic elements, more intricate structural relationships, and increased cognitive demands during processing. A common way to assess speakers' syntactic ability is based on the textual syntactic complexity features of their connected speech production, which can be categorised into two types.

The first approach involves using micro-linguistic features, such as the frequency or length of specific linguistic phenomena or constructions. For instance, some studies have examined the differences between cognitively impaired older adults and cognitively normal controls in their use of parts of speech (Lyons et al., 1994), phrases (Lundholm Fors et al., 2018), certain construction such as relative clauses, negative constructions, etc. (Kemper et al., 1993; Lai & Pai, 2009; Lyons et al., 1994), and clauses types (de Lira et al., 2011; Kempler et al., 1987; Orimaye et al., 2017). Chapin et al. (2022) further compared fine-grained features across different syntactic levels, including noun phrasal, verb phrasal, and clausal constructions. However, the problem with such features is that, in short, speech samples elicited through specific tasks, not every speaker necessarily produces the syntactic structure of interest. Thus, the feature value for many individuals is zero, not because they are unable to produce the structure, but because it simply did not occur in the limited sample. This leads to a flooring effect: the distribution of the feature is heavily skewed towards zero across participants, making it difficult to capture meaningful group differences, and reducing its discriminative power in distinguishing between older adults with and without cognitive impairments.

Another approach is to use macro-level syntactic measures, which refer to features that characterise the overall syntactic properties of a text rather than focusing on specific linguistic phenomena or syntactic structures. These measures are typically applied to any given sentence or text and generally yield non-zero values, making them more stable for comparison across samples. Among the various macro-level indices, the mean sentence length (MSL)[1] is the most common one, defined as the average length of all sentences in a given text, typically calculated based on the number of words or morphemes. MSL is a surface-level sentence measure, and does not take into account the internal syntactic structure of sentences, which makes it easy to calculate. This measure was first introduced by Brown (1973) in the field of child language development and has since been widely adopted, and borrowed to clinical linguistic research. Some researchers have identified significant differences in MSL between the MCI and healthy groups (Calzà et al., 2021; Wang et al., 2019), while other studies have not observed such differences (Kemper et al., 1993; Lundholm Fors et al., 2018). Sometimes, even within the same study, the same batch of participants may show different performances across different tasks (Beltrami et al., 2018). More importantly, MSL presents several issues. First, it disregards the influence of internal sentence structure. In other words, sentences of the same length but differing syntactic structures cannot be distinguished in terms of complexity by MSL. Second, the value of MSL is influenced by various qualitative factors in preprocessing steps (Brown, 1973; Leadholm & Miller, 1994; Parker & Brorson, 2005), which are often difficult to report in detail within a single journal article, and these processes are typically opaque. For instance, in utterance segmentation, the determination of whether two clauses belong to the same sentence or should be treated as separate sentences may lead to substantial differences in the values of sentence length and, therefore, MSL. In sum, MSL is, in fact, the result of a combination of various factors.

[1]This measure is alternatively known as mean length of unit (MLU) or mean length of sentence (MLS). There are various perspectives on the definitions and distinctions between 'utterance' and 'sentence'. This study does not intend to engage in a discussion of the differences between these two terms. For the sake of consistency, we will refer to the measure as MSL throughout the paper, following https://github.com/YuhuYang/QuanSyn.

To address the limitations of surface-level sentence measures, it is necessary to use syntactic trees with richer structural information to extract more refined syntactic complexity measures. Generally, syntactic frameworks are categorised into constituency (or phrase structure) grammar and dependency grammar. Yet the number and scope of syntactic measures based on these two frameworks are not well-matched. Common examples of syntactic complexity features based on constituent trees include Yngve and Frazier scores (Frazier, 1985; Yngve, 1960). The application of these measures in clinical linguistics can be traced back to early works, such as Cheung and Kemper (1992). However, on the one hand, the complex weighting components in these measures originate from studies of child language development, and such weightings may not necessarily be applicable to the discourse of older adults. On the other hand, due to the numerous varieties and versions in phrase structure grammar, any discrepancy in the qualitative syntactic analysis of linguistic phenomena or in the outputs of automatic syntactic parsers can lead to significant differences in the measures. In other words, the robustness of these measures is relatively weak.

In contrast, due to the simplicity of the representation, the relatively fewer analytical controversies, and the widespread use of large-scale publicly annotated treebanks such as Universal Dependencies (Nivre et al., 2016, 2020), dependency grammar has played an increasingly important role in the field of natural language processing in recent years. Among the dependency-based syntactic biomarkers for Alzheimer's disease and mild cognitive impairment, mean dependency distance (MDD) is the most commonly used measure. These studies cover various languages, including English (Gao & He, 2024; Orimaye et al., 2017; Roark et al., 2011), Swedish (Lundholm Fors et al., 2018; Sand Aronsson et al., 2021), Italian (Calzà et al., 2021). As for Mandarin-speaking older adults, Liu et al. (2021) found a slightly significant decrease in MDD in the discourse production of the AD group, while Yih et al. (2025) employed a wide range of syntactic measures, including MDD, and applied machine learning algorithms across three speech tasks to examine the discriminative power of feature combinations between groups. However, similar to the case for MSL, the significance levels found in different studies may vary, which could be attributed to factors such as language, the type of speech elicitation tasks, the preprocessing of data, and the computational methods used, or even some unknown random factors. For instance, it is generally expected that the syntactic complexity of individuals with cognitive impairment would decline, and consequently exemplified by the MDD measures. This has indeed been observed in studies by Roark et al. (2011), Sand Aronsson et al. (2021), Calzà et al. (2021), Gao and He (2024). However, Orimaye et al. (2017) found that the total DD score decreased in the probable AD group, but the dropping was not statistically significant, whereas Lundholm Fors et al. (2018) even obtained an unusual result, where the MDD in the cognitive impairment group did not decrease but rather increased, though this change was not statistically significant. Given the conflicting conclusions in previous studies, whether MDD is an effective indicator of cognitive impairment requires further validation through more cohorts and datasets.

### *Recent developments in quantitative syntax*

Quantitative Linguistics (QL) is a methodological branch of linguistics with a long history, concerned with quantifying textual notions and constructs. Early research in this field primarily focused on measuring lexical richness and the distribution of vocabulary within texts. Early studies, such as Miller and Chomsky (1963), Yngve (1960) and Frazier (1985), proposed scattered measures of grammatical complexity based on manually analysed phrase-structure syntactic trees. More recently, with the advancement of natural language processing and syntactic parsers, it has become possible to extract a wider range of text-based syntactic measures from syntactic trees. Consequently, a dedicated subfield of Quantitative Syntax has gradually taken shape (Köhler, 2012). In recent years, in the international academia of QL, numerous quantitative measures based on dependency syntax have been proposed. For example, Liu (2010) proposed the measure of dependency direction, which was later referred to as Liu-directionality, for assessing word order typology, and has been applied in cross-linguistic classification. In addition, in contrast to features that measure texts horizontally, such as MSL and MDD, mean hierarchical distance (Jing & Liu, 2015) captures the depth of syntactic trees, offering an alternative perspective on syntactic complexity. More have been put forward and widely applied in fields such as linguistic typology (Liu, 2010; Yan & Liu, 2021a), translation and interpreting studies (Fan & Jiang, 2019; Liang et al., 2017; Xu & Liu, 2023), second language acquisition (Li & Yan, 2021; Ouyang et al., 2022; Yang & Li, 2024), academic writing (Bi & Tan, 2024; Gao & He, 2023), etc.

However, based on our preliminary research, we are aware of only a few studies that utilise dependency-based measures other than mean dependency distance to investigate the syntactic abilities of cognitively healthy or impaired older adults. Liu et al. (2021) adopted a complex network approach, where they converted dependency treebanks into syntactic networks and used network parameters to examine the linguistic characteristics of older adults who are native speakers of Mandarin Chinese, comparing those with Alzheimer's disease to healthy individuals. They found that the vertices representing function words in the dependency networks of older adults with AD exhibit distinct network parameters, such as higher betweenness centrality, closeness centrality, and clustering coefficients, suggesting that the syntax in AD patients is impaired and tends to show more simplified patterns. Gao and He (2024) used the English data in the classic DementiaBank to investigate measures such as the mean dependency distance and dependency direction of the whole corpus and of certain dependency types, as well as the proportion of adjacent dependencies, and mean sentence length globally. Except for the proportion of adjacent dependencies, all other measures showed significant differences between the AD group and the healthy control group.

In sum, there is still a considerable amount of unexplored space that remains to be investigated. With the emergence of a large number of new dependency-related syntactic measures in recent years, it is necessary to investigate whether these measures can capture language changes in older adults, and assess their potential as linguistic biomarkers of cognitive impairment for early detection.

### *The current study*

Given the gaps in the aforementioned aspects, as well as the recent developments in the field of Quantitative Syntax, the aim of this study is to apply several newly proposed measures within this field to examine the cognitive state of Mandarin-speaking older adults based on connected speech, and to identify syntactic features and their combinations that show high discriminative ability. In particular, this study emphasises the automated computation of dependency-based textual measures and the application of machine learning techniques to uncover composite biomarkers. By leveraging automated methods, we aim to streamline the process of extracting and analysing complex linguistic features, facilitating more efficient and scalable detection of cognitive impairments. Machine learning algorithms, with their ability to identify the optimal combination of basic features for classification tasks, are employed to identify patterns and interactions among various linguistic measures, ultimately improving the accuracy and reliability of early diagnostic models for Alzheimer's disease and cognitive decline. Specifically, we seek to address the following three research questions:

(1) What is the correlation among these new QuanSyn measure? Do they reflect the same construct or different constructs?
(2) In comparison with MDD, which of the new measures show highly discriminative ability between the MCI and CN groups?
(3) What combinations of measures can better discriminate between the cognitively impaired and cognitively normal groups?

## Methods

### Participants

The cohort utilised in this study was selected from the Multimodal Corpus Gerontic Discourse (MCGD) (Zhou, 2024). The cohort was collected in two batches, with the second batch being the focus of the present study. This batch consists of 52 Chinese older adults, recruited from various communities in Shanghai during April to July 2023. These participants underwent the Chinese version of MoCA-B (Montreal Cognitive Assessment – Basic) and participated in a series of speech elicitation tasks, including single and sequential picture description, story recall, lifespan interview, and so on.

We excluded cases with unclear recordings, participants who spoke only Shanghainese, and those whose speech productions were too brief to be used for subsequent differentiation. Ultimately, for data compatibility, 41 participants' data were selected for analysis. The demographic information of these individuals is presented in Table 1. The MoCA-B cut-off scores for cognitive classification are presented in Table A1 and are adjusted based on years of education (Chen et al., 2016).

**Table 1.** Demographic information and cognitive scores of the participants.

| | CN ($N$ = 25) | MCI ($N$ = 16) | $p$-value |
|---|---|---|---|
| Age, years | 71.0 (6.0) | 71.5 (9.3) | 0.658 |
| Sex | | | 0.434 |
| Female, n (%) | 18 (72%) | 14 (87.5%) | |
| Male, n (%) | 7 (28%) | 2 (12.5%) | |
| Education, years | 11.0 (3.0) | 9.0 (2.5) | 0.143 |
| MoCA-B | 25.5 (4.0) | 18.0 (4.3) | 2.854e-07 |

CN and MCI stand respectively for cognitively normal and mild cognitively impaired. The age, years of education, MoCA-B scores are shown in the form of median (IQR), with the $p$-values assessed with the Mann-Whitney U-tests. The sex is presented as the number of each gender (percentage), with the p-value for sex calculated using a chi-square test.

All elderly participants were bilingual in Shanghainese and Mandarin, and the discourse elicitation was conducted in Mandarin. Ethical approval was granted by the Ethics Committee of Tongji University (tjsflrec202306). All participants provided written informed consent to participate in the study.

### Speech task

In the present study, we selected a sequential picture description (SeqPD) task. This paradigm has been used in a series of studies on the language-based detection of AD and MCI (Duong et al., 2005; Kong et al., 2016; Lima et al., 2014; Malcorra et al., 2024; March et al., 2006; Pistono et al., 2019). The participants were asked to narrate a story based on a stimulus consisting of several related sequential line drawings. The SeqPD task placed higher organisational demands on the produced story because the sequence of pictures visually provided the structure of the narrative, eliminating the need for participants to organise their ideas before expression. This task was particularly challenging for AD and MCI patients from a visuospatial perspective (Duong et al., 2005), as the repetitive characters often shifted through similar contexts while engaging in various behaviours.

Compared with other connected speech tasks, on the one hand, the SeqPD task is more story-driven than the classic single picture description tasks (SinPD, e.g. the Cookie Theft task), as participants needed to connect each picture to the preceding and subsequent ones (March et al., 2006). Our preliminary experiments found that in single picture descriptions with static content, patients tend to use more deictic expressions such as 'This is/These are (a boy)/ … ', which create an interactive relationship with the investigator or with the picture. Although such expressions also appear in multi-picture descriptions, the stronger narrative nature of the latter allows participants to immerse themselves better, resulting in the use of more predicate-headed sentences to describe the content of the pictures directly. On the other hand, in contrast to memory-based story narration tasks (e.g. autobiography, familiar or unfamiliar story retelling), the SeqPD task reduces the demands on memory as the story is visually presented to the participants (Duong et al., 2003). It does not depend on the participants' delayed memory that lasts from several tens of seconds to a few minutes, or episodic memory that lasts longer. Instead, it focuses on the working memory within seconds, which is used for syntactic construction during the linguistic organisation process.

Specifically, in our task, the participants were asked to describe *the Wanderings of Sanmao* comic strips, drawn by the Chinese artist Zhang Leping in the 1940s (Figure 1). The investigator gave the following instructions in Chinese, 'These are several pictures about Sanmao, and you need to tell a story. Please look at the pictures in this order (pointing to the sequence of pictures), and then tell me a story with a beginning, middle, and end'. If a participant did not provide a verbal response within 10 seconds, or if their response consisted of fewer than two utterances, an additional prompt was given, 'Anything else you can tell me about the picture?' No time limit was imposed for this task.

### Preprocessing procedure

The corpus used in this study has undergone careful manual cleaning and proofreading, with repetitions and self-repairs removed. The transcription work was completed in two

**Figure 1.** The wanderings of Sanmao comic strips (Zhang, 2017).

rounds, based on the automatic speech recognition transcription provided by Feishu,[2] and the manual checking of a postdoctoral researcher with a PhD degree in linguistics with a master student in linguistics. Any disputes in transcription were thoroughly discussed and resolved. Prior to sending the sentences into the syntactic parser, every effort was made to ensure that the sentences were as grammatically clear and unambiguous as possible, which is likely to improve the accuracy of the syntactic analysis. We used SpaCy[3] for automatic dependency syntactic analysis, and saved the files into the CoNLL-U format. Subsequently, the syntactically analysed files were manually checked a second time. The syntactic features were calculated using the QuanSyn package (version 0.0.9) in Python (Yang & Liu, 2025), by importing the analysed CoNLL-U file for further analysis.

## *Measures*

Table 2 lists the definitions, algorithms, and sources of the measures used in this study, which fall into four categories.

The first category indicates length or distance in the horizontal direction. In addition to our baseline mean dependency distance and mean sentence length, it also includes the mean total dependency length (MTDL) and mean tree width (MTW). The typical characteristic of horizontal syntactic measures is that they tend to increase the burden on working memory (Gibson, 1998).

The second category is related to the vertical dimension and includes three measures (MHD, MHDD, MTH). These measures can be viewed as the embedding depth at the word level. Early syntactic studies focused on phrase structure and proposed complexity

[2]https://www.feishu.cn/
[3]https://spacy.io/

**Table 2.** Measures of quantitative syntax.

| Abbreviation | Full name | Calculation | Original literature |
|---|---|---|---|
| Horizontal measures | | | |
| MSL* | mean sentence length | the average of all sentence lengths across the text | Brown (1973) |
| MDD* | mean dependency distance | the average of all dependency distances across all sentences in the text | Liu (2008) |
| MTDL | mean total dependency distance | the average of all sentence lengths across the text | Futrell et al. (2015) |
| MTW | mean tree width | the average of all tree widths across all sentences in the text | Zhang and Liu (2018) |
| Vertical measures | | | |
| MHD | mean hierarchical distance | the average of all hierarchy distances across all sentences in the text | Jiang and Liu (2015) |
| MTH | mean tree height | the average of all tree heights across all sentences in the text | Zhang and Liu (2018) |
| MHDD | mean hierarchical dependency distance | the average of all tdl values across all sentences in the text | Chen et al. (2022) |
| Valency-related measures | | | |
| MV | mean valency | the average of all valency values across all sentences in the text | Yan and Liu (2021b) |
| VK | variance of dynamic valencies | the average of all VK values across all sentences in the text | Lu et al. (2018) |
| Dependency direction | | | |
| HI | the proportion of head-initial dependencies | proportion of head-initial dependencies in the text | Liu (2010) |
| HF | the proportion of head-final dependencies | proportion of head-final dependencies in the text | Liu (2010) |

*Baseline measures.

measures such as Yngve's score, Frazier' score, and Miller and Chomsky (1963)'s structural complexity index, which are related to hierarchical, embedded structures. In contrast, the measures used in this study were derived from the more recent approach based on dependency grammar (DG). Due to the near-isomorphism between DG and phrase structure grammar (PSG) (Gaifman, 1965), there are commonalities between the two.

The third category consists of valency-related measures. Valency is the ability of a word to attract, or 'hook', other words (Tesnière, 1959). In its narrow sense, valency in linguistics originally referred to the number of arguments (e.g. subject, object, etc.) that a verb can take. Later, notions such as the valency of nouns and adjectives were also developed. In the broadest sense, valency now refers to the number of words governed by a particular word of any word class in a syntactic tree, plus the head word that governs it. The value of the latter is often 1 in non-directed acyclic graph tree structures. In other words, generalised valency refers to the sum of the in-degree and out-degree of a specific node in a syntactic tree, or the total number of words that have a dependency relationship with a particular word. Both the mean and variance of valencies at the textual level (Lu et al., 2018) were investigated.

The fourth category is related to the direction of dependency relations (Liu, 2010), and is considered to reflect the feature of language typology. It contains two measures, the proportion of head-final (HF) or head-initial (HI) dependency relations among all dependency relations. The former is also known as Liu-directionality (Fisch et al., 2019) or simple dependency direction (DDir).

The calculation methods for all the measures used in this paper, along with concrete examples, can be found at https://github.com/YuhuYang/QuanSyn.

### Statistical analysis and machine learning

To address the first research question, we conducted correlation tests to examine whether these measures reflect the same construct and to minimise collinearity in subsequent analyses. Specifically, we calculated the pairwise Pearson correlation coefficients ($r$) between the metrics and visualised the results using a heatmap. The range is as follows: $|r| \leq 0.3$: weak correlation; $0.3 < |r| \leq 0.7$: moderate correlation; $0.7 < |r| \leq 1$: strong correlation (Levshina, 2015). In addition, we employed the correlation coefficient as a measure of similarity and applied the Ward method to construct a hierarchical clustering dendrogram for the measures.

As for the second research question, given that each group had fewer than 30 samples, which violates the assumptions required for the use of the t-test, we applied the Mann-Whitney U test for each measure on a univariate basis. A box plot with jitter plot was used to display the distribution between different groups of measures. Then, we used the area under the Receiver Operating Characteristic curve (AUC) as the corresponding effect size (Bamber, 1975). In addition, the DeLong test was utilised to compare whether the differences in AUCs between the measures were statistically significant (DeLong et al., 1988).

Finally, to address the third research question, we used a support vector machine (SVM), a commonly used linear machine learning model, to identify feature pairs that would enhance the distinction between the CN and MCI groups compared to individual features. This model can be understood as a simple weighted combination of features and is well-known for its advantage in interpretability. The candidate features included all the syntactic measures as well as demographic information, except for HI and MV for the reasons mentioned above. We first applied logistic regression to transform and standardise each individual feature, scaling the values to the range between 0 and 1. It is equivalent to adding a simple classifier in front of each feature, and then applying ensemble learning to the overall model. These transformed values were then fed into the support vector machine for prediction. We then used cross-validation to select hyper-parameters (C in the SVM) and perform feature selection. Given that the sample size in this study is not large, we set the maximum number of features to 5. The average AUC from cross-validation was used as the evaluation metric. For cross-validation, the Stratified Leave-Pair-Out (SLPO) method was employed to assess the model's generalisation ability and prevent overfitting. In SLPO, each time one sample from the positive class and one from the negative class were selected as the test set, while the rest served as the training set. The process was repeated until the entire dataset was traversed. This method has been utilised in a number of previous studies (Fraser et al., 2019; Orimaye et al., 2017; Roark et al., 2011). It was considered the cross-validation approach most consistent with the intrinsic nature of the AUC metric, that is, the proportion of sample pairs where the positive sample is always situated on the same side with respect to the negative sample along a feature dimension. The performance of the final model was evaluated by ranking the models based on the average AUC from SLPOCV, from highest to lowest. Bootstrap resampling with 1000 iterations was then used to calculate the median AUC and the 95% confidence intervals for these combined features. In addition, ROC curves were plotted, and the DeLong test was performed to compare the models with the baseline features.

The statistical tests in this study were conducted using the SciPy[4] package in Python, the machine learning tasks were primarily carried out using the Scikit-learn[5] package, and all visualisations were generated with Matplotlib[6] and Seaborn.[7]

## Results

Figure 2 displays the correlation heatmap between various quantitative syntactic measures derived from dependency treebanks. To provide a clearer and more direct representation of the correlations between these measures, Figure 3 presents a hierarchical clustering

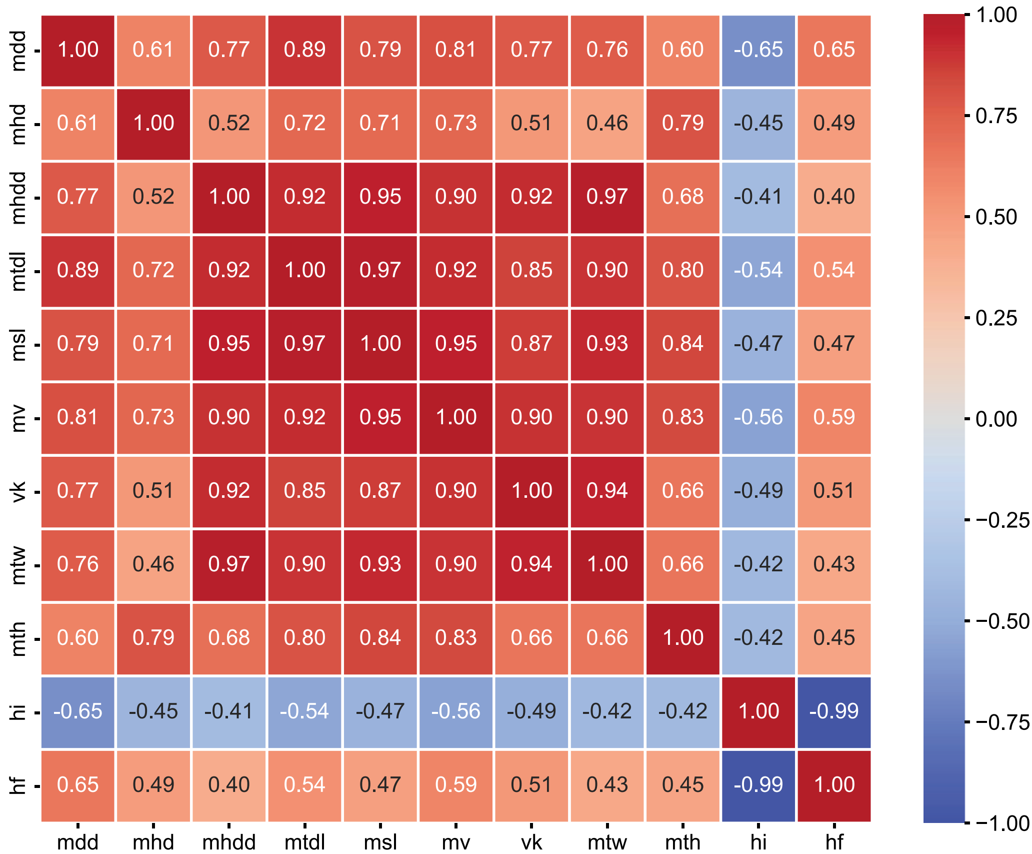


**Figure 2.** Correlation heatmap of QuanSyn measures.

[4]https://scipy.org/
[5]https://scikit-learn.org/
[6]https://matplotlib.org/
[7]https://seaborn.pydata.org/

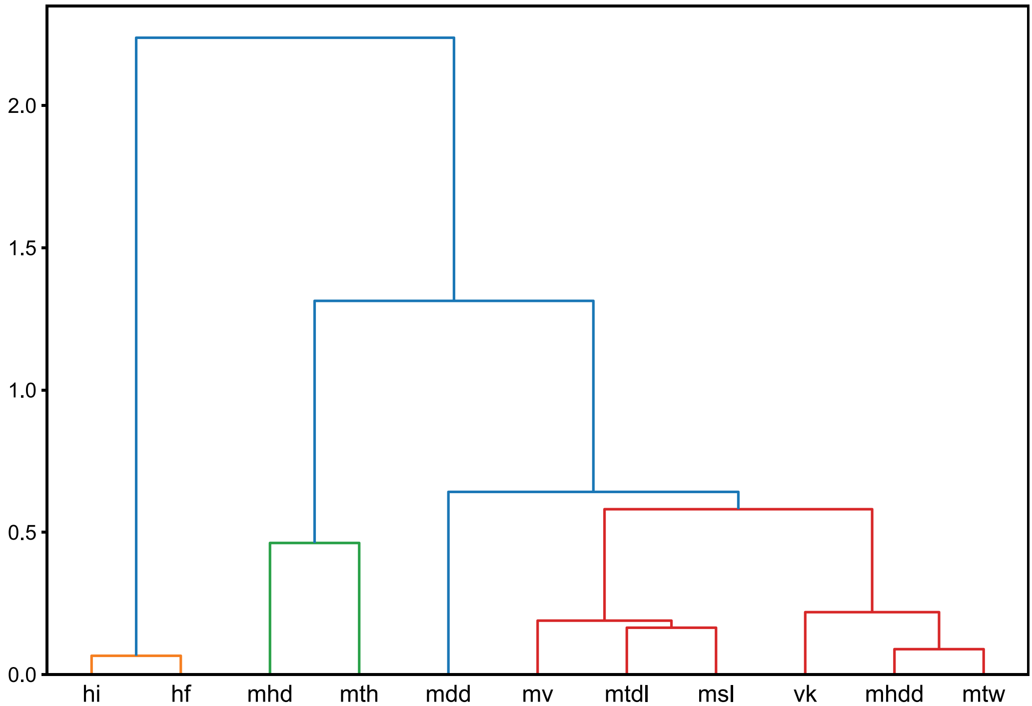


**Figure 3.** Hierarchical clustering dendrogram of the correlations between measures.

dendrogram. Both Figures 2 and 3 highlight the following points. First, by setting the appropriate threshold, the dendrogram can be roughly divided into three to four clusters. The majority of the horizontal measures (MTDL, MSL, MTW) and the valency-related measures (MV, VK) exhibit high correlations, and therefore, are grouped into the same cluster (the one on the far right), with the exception of a horizontal measure MHDD. The Pearson coefficients between the pair of these measures are greater than 0.85, indicating strong correlations. MDD, though has relatively lower $r$, is closely associated with this cluster. Second, MTH and MHD, both as measures reflecting the depth of syntactic trees, form the second cluster. Their correlations with MDD are around 0.6, while those with other measures in the rightmost cluster ranges from 0.7 to 0.9. Finally, the correlation between HI and HF is strictly −1, as determined by their definition, indicating that these two features in fact measure the construct on the same dimension, albeit in opposite directions. Furthermore, these two measures show relatively low correlations with the other metrics.

Figure 4 illustrates the discriminative ability of each measure for the MCI and CN groups as well as the corresponding degrees of significance. The Mann–Whitney U tests show that all measures, except for HF, exhibit statistical significance. Among these, only MHD is significant ($p < 0.05$), while all others show moderate significance ($p < 0.01$). After Bonferroni corrections, other measures still show significance, while MHD is not significant. This suggests that most features reflect a decrease in syntactic complexity in the MCI group compared to the CN group. As can be seen from Figure 4, for the HF dimension, the distributions of the two groups overlap to

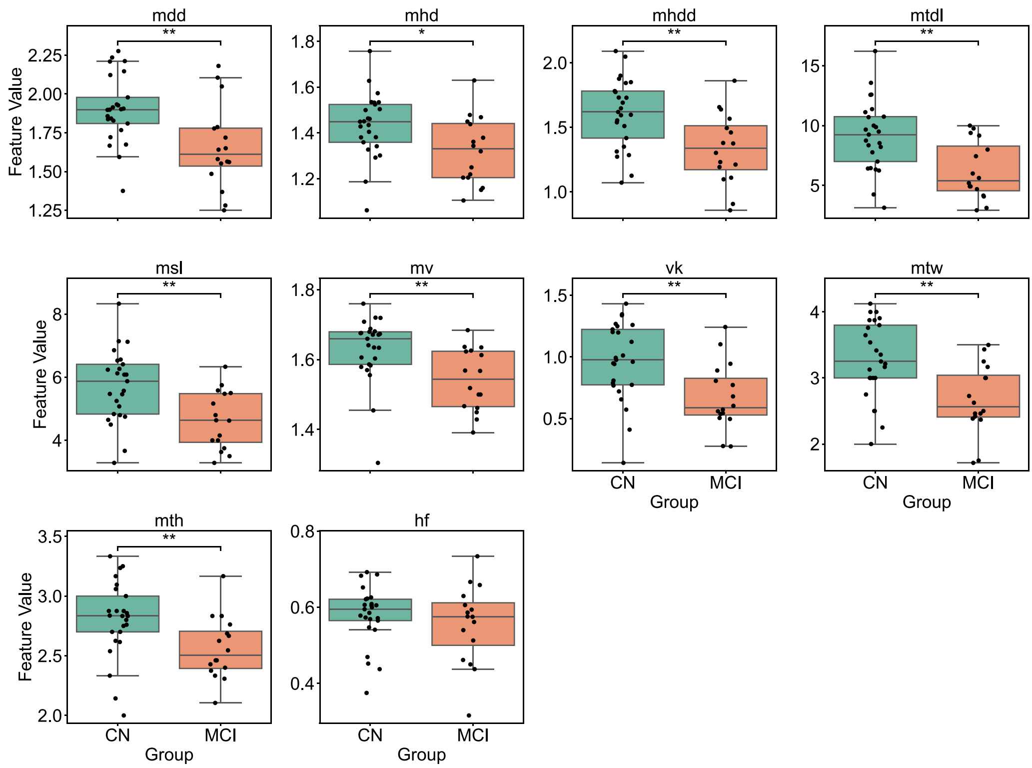


**Figure 4.** The boxplots and jitter plots of syntactic measures.

a large extent. It is indicated that this measure is nearly unable to differentiate between the two groups, performing no better than random guessing. In terms of effect size, Table 3 shows the median AUC values and the 95% confidence intervals for each measure after bootstrapping 1000 times, ordered by AUC in descending order. It is also shown that the AUC values are generally close, ranging from 0.746 to 0.805. Compared to the two baseline measures, MTDL, MV, VK, and MTW all slightly outperform MSL, but none exceed MDD. This suggests that, based on the cohort used in this study, MDD remains the most discriminative measure among all the measures we examined. However, as shown by the DeLong test, none of these differences between AUCs are statistically significant (Figure A1). Thus, with the exception of HF, most syntactic measures demonstrate a strong discriminative ability to distinguish between cognitive impairment and non-impairment.

**Table 3.** The AUC values of syntactic measures.

| Features | Median [95% CI] | *p*-value | Features | Median [95% CI] | *p*-value |
|---|---|---|---|---|---|
| MDD | 0.791 [0.610, 0.944] | 0.0010 | MSL | 0.765 [0.603, 0.897] | 0.0024 |
| MTDL | 0.788 [0.618, 0.918] | 0.0012 | MHDD | 0.760 [0.592, 0.897] | 0.0028 |
| VK | 0.782 [0.613, 0.908] | 0.0014 | MTH | 0.755 [0.594, 0.902] | 0.0032 |
| MV | 0.781 [0.622, 0.913] | 0.0014 | MHD | 0.710 [0.538, 0.864] | 0.0128 |
| MTW | 0.781 [0.610, 0.904] | 0.0014 | HF | 0.577 [0.371, 0.758] | 0.2191 |

Subsequently, we attempted to find feature combinations with better performance with SVM. Among all models, we opted to select feature pairs consisting of only two features to simplify the model as much as possible. Achieving a high level of distinction with just two features would, therefore, be the most efficient and economical solution. We identified several models that outperformed or behaved equivalently as the best single feature, MDD, in terms of the mean AUC in SLPO cross-validation. These models all utilised two-feature panels with SVM parameter C set to 0.1, including MTDL+MSL (mean $AUC_{SLPOCV} = 0.795$), MTW+VK (mean $AUC_{SLPOCV} = 0.792$), MDD+MTW (mean $AUC_{SLPOCV} = 0.790$).[8] Figure 5 shows the ROC curves for these models against the baseline features, and the median AUC and 95% CI through bootstrapping 1000 times. The DeLong test indicates that, compared to the baseline model, these feature combinations are not statistically significant ($p > 0.05$). The inclusion of the demographic information, in our case, did not improve the performance of the models.

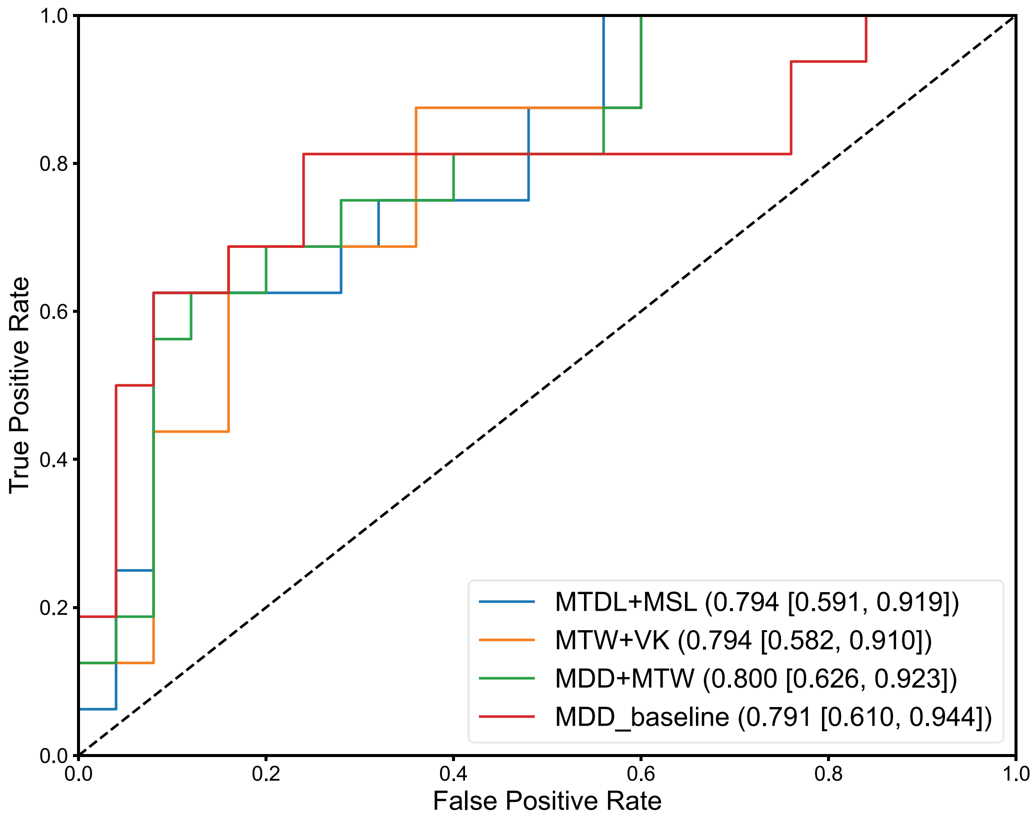


**Figure 5.** The ROC curve of the machine learning model.

[8]It is important to note that the average AUC from cross-validation is generally lower than the AUC of the final combined feature set on the entire dataset, as the latter in fact involves data leakage. In contrast, cross-validation separates the training and testing sets, avoiding this issue.

## Discussion

The aim of this study is to take a quantitative syntactic approach, based on recently proposed dependency-based measures in the academic literature, to identify new linguistic biomarkers on the syntactic level for cognitive decline of Mandarin-speaking older adults.

The findings of this study suggest that, apart from dependency direction (i.e. the proportion of head-final or head-initial dependency relations), most of the recently proposed syntactic measures can to some extent differentiate between cognitively impaired older adults and their healthy peers. Furthermore, several of the new measures are not entirely correlated with MDD, indicating that they capture different or extra information.

### *Correlations between dependency-based measures*

Regarding the first research question on the interrelationships between the measures under investigation, we will go through the four clusters of measures as shown in Figure 3 based on their correlations, and discuss each cluster in turn.

The leftmost cluster (including HF and HI), when examined alongside the heatmap, suggests that these two measures are perfectly identical, despite their opposite trends and interpretations. Although several previous studies have investigated both two measures simultaneously (Gao & He, 2024; Jiang et al., 2019; Liu, 2010), we assume that it may not be necessary to use them both at the same time in future studies. The second cluster includes two measures related to the vertical dimension, MHD and MTH. It seems that they both roughly point to the same construct, namely embedding depth. However, another measure, MHDD, shows a low correlation with them. This discrepancy may be attributed to the definition of MHDD (Chen et al., 2022). Both the numerator and denominator of MHDD essentially capture syntactic complexity, and their ratio determines whether it reflects syntactic complexity or incomplexity, depending on which component increases more rapidly. As a result, despite its definition being roughly the inverse of MHD, MHDD still exhibits a certain degree of positive correlation. This suggests that the measure may not be suitable. A more appropriate definition could be to define it as the average of MaxHL over the number of sentences. The current definition, unlike measures such as MDD and MHD, fails to adequately capture the concept of averaging.

The fourth and the largest cluster on the right in Figure 3 show strong correlation among the measures within this cluster. It is also indicated that these measures are significantly related to mean sentence length, or that the factor sentence length predominates. This is consistent with a number of previous findings (see Zhang & Liu, 2018 for the relationship of sentence length with tree width and tree height, Lu et al., 2018 for that with variance of dynamic valencies, Futrell et al., 2015; Niu & Liu, 2022 for that with total dependency length/distance, and Chen et al., 2022 for that with mean hierarchical dependency distance). However, it is noteworthy that some relationships established in previous studies are in fact between sentence-level measures. For example, Zhang and Liu (2018) identified a statistical power-law relationship between sentence-level tree width, tree height, and sentence length. In contrast, the present study examines the relationships between their corresponding counterparts at the text level. Therefore, our research further complements the network of the relationships between quantitative measures at different levels from a perspective of synergetic linguistic (Köhler, 1993).

However, it should be noted that although most of the measures in this cluster exhibit quite high intercorrelation coefficients (mostly greater than 0.85), they do not reach 100%, and therefore cannot be considered as completely identical indicators of the same construct, unlike the relationship between HI and HF.[9] Correlation coefficients are influenced by the authentic distribution of the samples. In other words, this imperfect match suggests the presence of different or additional information. For example, all the above measures can be divided into two main groups: one group directly averages across the entire corpus (MDD, MHD, MV), while the other group first averages within the sentence and then averages across the entire sentence (MTDL, MTW, MTH). The latter is influenced by and will carry the information of the sentence length distribution, but such information differences contained in the indicators masked by the strong correlation may be related to the relatively small sample size, or the single text type.

Based on the threshold we selected, MDD forms a separate third cluster. However, from the correlation coefficients, there is also a moderate correlation between MDD and MSL, although weaker than the correlations within the other members of this cluster. Given that we employed a linear correlation test, the correlation appears weaker at first sight. However, Jiang and Liu (2015) distinguished the relationship between the average MDD and sentence length across sentences of varying lengths, while Niu and Liu (2022) examined the relationship between sentence-level MDD and sentence length, both finding a power-law relationship. This may explain the relatively modest correlation observed. Therefore, it can be concluded that MDD encompasses factors beyond sentence length, as indicated earlier in this paper, including internal syntactic structure information, which plays a critical role in distinguishing MCI from healthy age-matched controls.

Overall, this part of our work aligns with previous attempts (Agmon et al., 2024; Cheung & Kemper, 1992; Roark et al., 2011) which endeavoured to investigate the relationships between several syntactic measures. However, due to the limitations of their time, they only examined some of the measures proposed in the literature at that point, such as those based on phrase structure trees (e.g. Frazier and Yngve scores) and those that involve scoring specific syntactic phenomena (e.g. DLevel, IPSyn, DSS). Although the features used in this study only partially overlap with those in previous research (e.g. MDD, MSL), we are also contributing to the establishment of a more comprehensive and complete network of correlations among syntactic measures, and to the synergetic linguistics endeavour aimed at building a system of linguistic laws and the functional and statistical relationships between quantitative measures (Köhler, 1993). Moreover, since as mentioned above the relationships between these indicators are influenced by the empirical distribution, multiple studies such as Wang and Liu (2017), Chen et al. (2017, 2022) have investigated how the relationships among these quantitative measures are affected by text type or genre. In this study, we also explored the more relationships using clinical and elderly speech corpora, expanding our understanding of how these they manifest in new domains.

[9]Note that one exception is the pair of MV and MSL. In fact, there is a functional relationship between these two metrics: MV = 2 – 2/MSL, which can be proven mathematically, although the linear correlation coefficient does not capture this nonlinear relationship. Therefore, we suggest that, similar to the relationship between HI and HF, only one of these two measures should be retained in future studies. In fact, MV is more applicable to specific parts of speech rather than to words across the entire text (Yan & Liu, 2021b).

### ***Discriminative power of unidimensional features***

The second research question aims to examine the discriminative power of our novel syntactic features for distinguishing between MCI and CN elderly populations. We found that using MDD and MSL as baselines, three measures (MTDL, MV, and MTW) slightly outperformed MSL in terms of discriminative ability, but MDD proved to be the best-performing indicator among all the candidate features in our dataset. Among the remaining features, MTH and MHD are somewhat weaker, but still demonstrate strong diagnostic potential. In contrast, HF (and HI) performs poorly, offering little to no value in determining whether the speaker exhibits cognitive impairment. These results are largely consistent with the correlations between the measures, as well as the degree of similarity between them, respectively, shown in Figures 2 and 3. Overall, our findings indicate that compared to the normal group, the MCI patients show significant decline in sentence and dependency length, embedding depth, as well as in the words' ability in attracting other words or taking arguments or modifiers, while no difference is observed in the dependency direction of overall speech production between the MCI and normal groups. Therefore, most measures in our study, except HF, can all serve as substitutes for MDD and as potential biomarkers for MCI.

Specifically, for MDD and MSL, our study offers a new case addressing the contradictory conclusions in previous research regarding whether there are significant differences between cognitively impaired and cognitively healthy older adults. In our case, both demonstrate strong discriminative ability between the MCI and CN boundaries, reaching a high level of statistical significance. Such findings align with those of Roark et al. (2011), Sand Aronsson et al. (2021), Calzà et al. (2021) for MDD, and those of Wang et al. (2019), Calzà et al. (2021) for MSL. These suggest that older adults with cognitive impairment do tend to use shorter sentences, and the dependencies within these sentences are relatively simple, likely consisting mostly of adjacent dependencies (Jiang & Liu, 2015). We point out that the inconsistency in the significance or the changing tendency of a measure found in previous literature may also be attributed to differences in the definition of indicators. For instance, apart from MDD, there are several variants of dependency distance-related metrics, such as Dependency Length (also known as Total Dependency Distance, or TotalDD), Maximum Dependency Distance (MaxDD), and Averaged MDD (AMDD) (Lundholm Fors et al., 2018), the number of unique dependencies, and the average number of unique dependencies per sentence (Orimaye et al., 2017). Taking the distinction between AMDD and MDD as an example. Similar to what was mentioned in the previous section, their difference lies in whether the averaging is done over the entire sentence or within the sentence first, and then across all sentences. The former is even referred to as MDD in some studies, which makes it particularly confusing. In fact, these slightest qualitative differences in definitions can essentially lead to viewing them as different measures (Yih & Liu, 2025). Therefore, to enhance the comparability across studies, it is necessary in the future to either standardise the measurement of a feature, or to specifically investigate the relationships between these superficially identical but substantially different index variants.

MDD has previously been shown to reflect cognitive abilities in the human brain, particularly working memory (Gibson, 1998, 2000; Hudson, 1995; Xu & Liu, 2022). Our study shows that the decline in MDD, an indicator claimed to reflect working memory, is consistent with memory decline, a common symptom of dementia. However, since MCI can be classified into the amnestic and non-amnestic types, and we did not make such distinction in the current study, this issue remains unresolved. In addition, previous studies have mostly related MDD to reaction times in sentence comprehension (Niu & Liu, 2022), while in the field of clinical linguistics, the digit span test (DST) is commonly used as a neuropsychological test specifically designed to measure working memory (Cheung & Kemper, 1992; Roark et al., 2011). Therefore, it is necessary in the future to correlate the results of MDD and DST, specifically examining whether MDD truly reflects working memory.

It is worth noting that, although dependency direction has been found to be a strong indicator of second language learners' linguistic abilities in some second language acquisition (Ouyang & Jiang, 2023) and translation/interpreting literature (Fan & Jiang, 2019), in our study it was found to be nearly equivalent to random guessing. This finding suggests that, regardless of whether they have cognitive impairment, older adults may use similar proportions of head-final or head-initial dependency relations. It may be due to the fact that second language acquisition, as well as translation/interpretation, involves two languages, and the languages used in these studies have different proportions of head-final dependencies. This leads to changes in learners' linguistic abilities when transitioning from one language to another. Liu et al. (2024) also compared the dependency directionality between Mandarin and English native older adults and found significant differences. Yet all these in fact can be ultimately attributed to the underlying differences in language types (Liu, 2010). In contrast, in our study, speakers only used one language, which explains why no variation in dependency direction was observed.

Nevertheless, our result and explanation contrasts with the findings of Gao and He (2024), who reported significant differences in DDir between the AD and healthy control groups based on the data from the DementiaBank corpus in a single language, English. We conjecture the first potential reason for this discrepancy is that they did not use all the data from DementiaBank. Thus, the results may be related to the specific subset they selected. Another explanation could be that the language differences in the AD group were more apparent than in the MCI group, which might explain why no such differences were observed in our study. In sum, this point requires further validation in the future.

### *Discriminative power of feature composites*

Finally, to address the third research question, we attempted to identify feature combinations and models that would yield better performance. We indeed obtained several models, improved the differentiation of the MCI group compared to using MDD only, but this improvement was quite minimal. However, it is beyond expectation that despite the moderate correlation between MHD/MTH and MDD, the inclusion of the hierarchical information did not help improve the discriminative power.

There is already a substantial body of literature based on linguistic biomarkers and machine learning as well as deep learning methods (de la Fuente Garcia et al., 2020; Parsapoor, 2023). Specifically for Mandarin, compared to several previous studies on AD detection using Chinese older adults' speech corpora, our performance may not be as high, with the state of the art being AUC generally above 80% (Chien et al., 2019; Wang et al., 2023). However, some studies do not approach the problem from the perspective of identifying biomarkers, but instead focus solely on classification accuracy within a specific dataset, without reporting the AUC values (Chou et al., 2024; Wang et al., 2021). Others employed pre-trained embedding vectors or focus on developing new deep learning models (Chien et al., 2019; Tsai et al., 2021), whose features lack interpretability, in terms of which our approach has an advantage. In addition, the relatively low discriminative ability observed in this study may be due to the fact that we are distinguishing the boundaries between MCI and CN, which is the most challenging task, more difficult than distinguishing between pure CN vs. AD and even CN vs. MCI + AD. Since the latter involves a mix of MCI and AD samples, the AD samples, which overlap less with the CN group, may artificially cause the increase of the overall AUC value, leading to an overestimation (e.g. Vincze et al., 2022). Overall, our results are closest to those of Wang et al. (2023), who examined the binary classification of MCI and CN, achieving an ACC of 0.81 and an AUC of 0.81. However, they used 192 speech features across three major categories, whereas we only used 1 to 2 feature(s), making our approach clearly more economical.

To sum up, our results suggest that, among the measures selected in this study, the classical MDD may still outperform all the rest. In addition, as previously noted, due to the strong collinearity between different measures, adding additional features to MDD provides little contribution to distinguishing between the MCI and CN groups. Since MDD is a single measure, it is more economical than two-feature pairs. From a clinical perspective, MDD may be the preferable choice, as using a single measure reduces measurement error compared to feature combinations.

## Conclusions

This study follows the approach of Quantitative Syntax and examines the effectiveness of multiple new measures and their combinations for identifying mild cognitive impairment in Mandarin-speaking older adults. The results indicate that, at a unidimensional level, most novel measure shows significant discriminative power for distinguishing between MCI and NC, except for the dependency direction (HF). Meanwhile, the classic measure MDD shows highest discriminative power, outperforming all others measures. However, based on the interpretable linear machine learning model SVM, it was found that although two feature combinations yield better discriminative performance than MDD alone, the difference is not significant. This suggests that MDD remains one of the most ideal syntactic measures currently available. In sum, this study contributes additional alternative measures to the repertoire of linguistic biomarkers for identifying older adults with mild cognitive impairment, Alzheimer's disease and related cognitive decline based on speech production. It significantly advances early diagnosis and large-scale community screening, enhances clinical decision-making, and contributes to more personalised interventions for cognitive impairment in ageing populations. The findings highlight the importance of integrating advanced computational techniques with linguistic analysis, offering new insights into the complex relationship between language and cognition in neurodegenerative conditions.

There are no doubt certain limitations to this study. First, the sample size in our study was relatively small, and some participants produced very brief speech outputs. This was due to the experimenter not adequately controlling the reminders during the induction process to ensure that the speakers' outputs met a minimum length. Furthermore, although we employed cross-validation in our study, we relied on a single dataset, without performing external validation. This limits the generalisability of the findings to other populations or settings, as the good discriminatory performance achieved in this paper may be due to chance and the specific dataset used. Second, the discourse types in our study were restricted to a particular task, which may not capture the full range of linguistic features that could be relevant for differentiating cognitive impairments. In future studies, it is important to consider a wider variety of speech tasks to identify metrics that are robust across different linguistic contexts. This approach could help to uncover indicators that hold across various types of discourse production, thus enhancing the reliability and applicability of the findings. Finally, some of the metrics used in our study were correlated, resulting in a smaller number of valid contributors in a multivariate predictive model. In future research, it will be necessary to explore and incorporate highly discriminative measures in other domains to train models that achieve enhanced detection performance.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This research was supported by the National Social Science Fund of China [Grant: 24BYY120] awarded to Lihe Huang, and by the Fundamental Research Funds for the Central Universities [Grant: 22120250439] awarded to Tsy Yih.

## ORCID

Tsy Yih http://orcid.org/0000-0003-2401-8486
Yiran Yang http://orcid.org/0009-0006-8546-5668
Mu Yang http://orcid.org/0009-0009-0883-1910
Haitao Liu http://orcid.org/0000-0003-1724-4418
Lihe Huang http://orcid.org/0000-0002-6079-4602

## Data availability statement

The original speech data used in this study are not openly available due to project confidentiality restrictions. However, the data that support the findings of this study are available from the first author upon reasonable request.

## References

Agmon, G., Pradhan, S., Ash, S., Nevler, N., & Liberman, M., Grossman, M., Cho, S. (2024). Automated measures of syntactic complexity in natural speech production: Older and younger adults as a case study. *Journal of Speech, Language, & Hearing Research*, *67*(2), 545–561. https://doi.org/10.1044/2023_JSLHR-23-00009

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425787

Bamber, D. (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology*, *12*(4), 387–415. https://doi.org/10.1016/0022-2496(75)90001-2

Beltrami, D., Gagliardi, G., Rossini Favretti, R., Ghidoni, E., Tamburini, F., & Calzà, L. (2018). Speech analysis by natural language processing techniques: A possible tool for very early detection of cognitive decline? *Frontiers in Aging Neuroscience*, *10*, 369. https://doi.org/10.3389/fnagi.2018.00369

Bi, Y., & Tan, H. (2024). Language transfer in L2 academic writings: A dependency grammar approach. *Frontiers in Psychology*, *15*, 1384629. https://doi.org/10.3389/fpsyg.2024.1384629

Blanco, K., Salcidua, S., Orellana, P., Sauma-Pérez, T., León, T., Steinmetz, L. C. L., de la Cruz, R., Duran-Aniotz, C., & de la Cruz, R. (2023). Systematic review: Fluid biomarkers and machine learning methods to improve the diagnosis from mild cognitive impairment to Alzheimer's disease. *Alzheimer's Research and Therapy*, *15*(1), 176. https://doi.org/10.1186/s13195-023-01304-8

Brown, R. (1973). *A first language: The early stages*. Harvard University Press. https://doi.org/10.4159/harvard.9780674732469

Calzà, L., Gagliardi, G., Favretti, R. R., & Tamburini, F. (2021). Linguistic features and automatic classifiers for identifying mild cognitive impairment and dementia. *Computer Speech and Language*, *65*, 101113. https://doi.org/10.1016/j.csl.2020.101113

Chapin, K., Clarke, N., Garrard, P., & Hinzen, W. (2022). A finer-grained linguistic profile of Alzheimer's disease and mild cognitive impairment. *Journal of Neurolinguistics*, *63*, 101069. https://doi.org/10.1016/j.jneuroling.2022.101069

Che, Y., & Huang, L. (2025). Lexical and semantic features of Chinese-speaking older adults with mild cognitive impairment across elicitation methods. *Aphasiology*. https://doi.org/10.1080/02687038.2025.2510323

Chen, K. L., Xu, Y., Chu, A. Q., Ding, D., Liang, X. N., Nasreddine, Z. S., Dong, Q., Hong, Z., Zhao, Q. H., & Guo, Q. H. (2016). Validation of the Chinese version of Montreal cognitive assessment basic for screening mild cognitive impairment. *Journal of the American Geriatrics Society*, *64*(12), e285–e290. https://doi.org/10.1111/jgs.14530

Chen, R., Deng, S., & Liu, H. (2022). Syntactic complexity of different text types: From the perspective of dependency distance both linearly and hierarchically. *Journal of Quantitative Linguistics*, *29*(4), 510–540. https://doi.org/10.1080/09296174.2021.2005960

Chen, R., Liu, H., & Altmann, G. (2017). Entropy in different text types. *Digital Scholarship in the Humanities*, *32*(3), 528–542. https://doi.org/10.1093/llc/fqw008

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

Chien, Y. W., Hong, S. Y., Cheah, W. T., Yao, L. H., Chang, Y. L., & Fu, L. C. (2019). An automatic assessment system for Alzheimer's disease based on speech using feature sequence generator and recurrent neural network. *Scientific Reports*, *9*(1), 19597. https://doi.org/10.1038/s41598-019-56020-x

Chou, C. J., Chang, C. T., Chang, Y. N., Lee, C. Y., Chuang, Y. F., Chiu, Y. L., Liang, W. L., Fan, Y. M., & Liu, Y. C. (2024). Screening for early Alzheimer's disease: Enhancing diagnosis with linguistic features and biomarkers. *Frontiers in Aging Neuroscience*, *16*, 1451326. https://doi.org/10.3389/fnagi.2024.1451326

Cummings, L. (2020). *Language in dementia*. Cambridge University Press. https://doi.org/10.1017/9781108587921

de la Fuente Garcia, S., Ritchie, C. W., & Luz, S. (2020). Artificial intelligence, speech, and language processing approaches to monitoring Alzheimer's disease: A systematic review. *Journal of Alzheimer's Disease*, *78*(4), 1547–1574. https://doi.org/10.3233/JAD-200888

de Lira, J. O., Ortiz, K. Z., Campanha, A. C., Bertolucci, P. H., & Minett, T. S. (2011). Microlinguistic aspects of the oral narrative in patients with Alzheimer's disease. *International Psychogeriatrics*, *23*(3), 404–412. https://doi.org/10.1017/S1041610210001092

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, *44*(3), 837–845. https://doi.org/10.2307/2531595

Dubois, B., Feldman, H. H., Jacova, C., Hampel, H., Molinuevo, J. L., Blennow, K., Cummings, J. L., Gauthier, S., Selkoe, D., Bateman, R., Cappa, S., Crutch, S., Engelborghs, S., Frisoni, G. B., Fox, N. C., Galasko, D., Habert, M.-O., Jicha, G. A.Nordberg, A., . . . Scheltens, P. (2014). Advancing research diagnostic criteria for Alzheimer's disease: The IWG-2 criteria. *Lancet Neurology*, *13*(6), 614–629. https://doi.org/10.1016/S1474-4422(14)70090-0

Duboisdindien, G. (2024). The analysis of gestural and verbal pragmatic markers produced by mild cognitive impaired participants during longitudinal and autobiographical interviews. *Clinical Linguistics and Phonetics*, *38*(2), 116–137. https://doi.org/10.1080/02699206.2023.2174450

Duong, A., Giroux, F., Tardif, A., & Ska, B. (2005). The heterogeneity of picture-supported narratives in Alzheimer's disease. *Brain and Language*, *93*(2), 173–184. https://doi.org/10.1016/j.bandl.2004.10.007

Duong, A., Tardif, A., & Ska, B. (2003). Discourse about discourse: What is it and how does it progress in Alzheimer's disease? *Brain & Cognition*, *53*(2), 177–180. https://doi.org/10.1016/s0278-2626(03)00104-0

Fan, L., & Jiang, Y. (2019). Can dependency distance and direction be used to differentiate translational language from native language? *Lingua*, *224*, 51–59. https://doi.org/10.1016/j.lingua.2019.03.004

Filiou, R. P., Bier, N., Slegers, A., Houze, 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. https://doi.org/10.1080/02687038.2019.1608502

Fisch, A., Guo, J., & Barzilay, R. (2019). Working hard or hardly working: Challenges of integrating typology into neural dependency parsers. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 5714–5720). Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1574

Fraser, K. C., Lundholm Fors, K., Eckerström, M., Öhman, F., & Kokkinakis, D. (2019). Predicting mci status from multimodal language data using cascaded classifiers. *Frontiers in Aging Neuroscience*, *11*, 205. https://doi.org/10.3389/fnagi.2019.00205

Frazier, L. (1985). Syntactic complexity. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 129–189). Cambridge University Press. https://doi.org/10.1017/CBO9780511597855.005

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341. https://doi.org/10.1073/pnas.1502134112

Gaifman, H. (1965). Dependency systems and phrase-structure systems. *Information & Control*, *8*(3), 304–337. https://doi.org/10.1016/S0019-9958(65)90232-9

Gao, N., & He, Q. (2023). A corpus-based study of the dependency distance differences in English academic writing. *Sage Open*, *13*(3), 21582440231198408. https://doi.org/10.1177/21582440231198408

Gao, N., & He, Q. (2024). A dependency distance approach to the syntactic complexity variation in the connected speech of Alzheimer's disease. *Humanities and Social Sciences Communications*, *11*(1), 1–12. https://doi.org/10.1057/s41599-024-03509-0

Gaur, A., Rivet, L., Mah, E., Bawa, K. K., Gallagher, D., Herrmann, N., & Lanctôt, K. L. (2023). Novel fluid biomarkers for mild cognitive impairment: A systematic review and meta-analysis. *Ageing Research Reviews*, *102046*, 102046. https://doi.org/10.1016/j.arr.2023.102046

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 94–126). The MIT Press. https://doi.org/10.7551/mitpress/3654.003.0008

Huang, L. (2024). Language and ageing. In L. W, H. Zhu, & J. Simpson (Eds.), *The Routledge handbook of applied linguistics* (2nd ed., Vol. 2, pp. 267–279). Routledge. https://doi.org/10.4324/9781003082637-24

Huang, L., Liu, Z., & Li, Y. (2023). Incompleteness features in the descriptive discourse of Chinese elders with and without Alzheimer's disease. *Clinical Linguistics and Phonetics*, *37*(12), 1171–1185. https://doi.org/10.1080/02699206.2022.2092423

Huang, L., Yang, H., Che, Y., & Yang, J. (2024). Automatic speech analysis for detecting cognitive decline of older adults. *Frontiers in Public Health*, *12*, 1417966. https://doi.org/10.3389/fpubh.2024.1417966

Hudson, R. (1995). Measuring syntactic difficulty. Manuscript. University College. https://dickhudson.com/wp-content/uploads/2013/07/Difficulty.pdf,

Jia, J. (2016). *Chinese diagnosis and treatment of dementia and cognitive impairment guide (in Chinese)* (2nd ed. People's Medical Publishing House Co. Ltd.

Jia, L., Du, Y., Chu, L., Zhang, Z., Li, F., Lyu, D., Qiu, Q., Li, Y., Zhu, M., Jiao, H., Song, Y., Shi, Y., Zhang, H., Gong, M., Wei, C., Tang, Y., Fang, B., Guo, D., Wang, F., . . . Qin, Q. (2020). Prevalence, risk factors, and management of dementia and mild cognitive impairment in adults aged 60 years or older in China: A cross-sectional study. *Lancet Public Health*, *5*(12), e661–e671. https://doi.org/10.1016/s2468-2667(20)30185-7

Jiang, J., & Liu, H. (2015). The effects of sentence length on dependency distance, dependency direction and the implications-based on a parallel English-Chinese dependency treebank. *Language Sciences*, *50*, 93–104. https://doi.org/10.1016/j.langsci.2015.04.002

Jiang, J., Ouyang, J., & Liu, H. (2019). Interlanguage: A perspective of quantitative linguistic typology. *Language Sciences*, *74*, 85–97. https://doi.org/10.1016/j.langsci.2019.04.004

Jing, Y., & Liu, H. (2015). Mean hierarchical distance augmenting mean dependency distance. In *Proceedings of the third international conference on dependency linguistics (Depling 2015)* (pp. 161–170). https://aclanthology.org/W15-2119/

Kemper, S., LaBarge, E., Ferraro, F. R., Cheung, H., Cheung, H., & Storandt, M. (1993). On the preservation of Syntax in Alzheimer's disease: Evidence from written sentences. *Archives of Neurology*, *50*(1), 81–86. https://doi.org/10.1001/archneur.1993.00540010075021

Kempler, D., Curtiss, S., & Jackson, C. (1987). Syntactic preservation in Alzheimer's disease. *Journal of Speech and Hearing Research*, *30*(3), 343–350. https://doi.org/10.1044/jshr.3003.343

Kleiman, M. J., & Galvin, J. E. (2024). High frequency post-pause word choices and task-dependent speech behavior characterize connected speech in individuals with mild cognitive impairment. *Journal of Alzheimer's Disease*, *102*(3), 815–829. https://doi.org/10.1177/13872877241291239

Köhler, R. (1993). Synergetic linguistics. In R. Köhler & B. B. Rieger (Eds.), *Contributions to Quantitative Linguistics: Proceedings of the First International Conference on Quantitative Linguistics, QUALICO, Trier, 1991* (pp. 41–51). Kluwer, Dordrecht. https://doi.org/10.1007/978-94-011-1769-2_4

Köhler, R. (2012). *Quantitative syntax analysis*. Walter de Gruyter. https://doi.org/10.1515/9783110272925

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

Lai, Y. H., & Pai, H. H. (2009). To be semantically-impaired or to be syntactically-impaired: Linguistic patterns in Chinese-speaking persons with or without dementia. *Journal of Neurolinguistics*, *22*(5), 465–475. https://doi.org/10.1016/j.jneuroling.2009.03.004

Leadholm, B. J., & Miller, J. F. (1994). *Language sample analysis: The Wisconsin guide*. Wisconsin Department of Public Instruction.

Levshina, N. (2015). *How to do linguistics with r: Data exploration and statistical analysis*. John Benjamins. https://doi.org/10.1075/z.195

Li, L., Yu, X., Sheng, C., Jiang, X., Zhang, Q., Han, Y., & Jiang, J. (2022). A review of brain imaging biomarker genomics in Alzheimer's disease: Implementation and perspectives. *Translational Neurodegeneration*, *11*(1), 42. https://doi.org/10.1186/s40035-022-00315-z

Li, W., & Yan, J. (2021). Probability distribution of dependency distance based on a treebank of Japanese EFL learners' interlanguage. *Journal of Quantitative Linguistics*, *28*(2), 172–186. https://doi.org/10.1080/09296174.2020.1754611

Liang, J., Fang, Y., Lv, Q., & Liu, H. (2017). Dependency distance differences across interpreting types: Implications for cognitive demand. *Frontiers in Psychology*, *8*, 2132. https://doi.org/10.3389/fpsyg.2017.02132

Lima, T. M., Brandão, L., Parente, M. A. D. M. P., & Peña-Casanova, J. (2014). Alzheimer's disease: Cognition and picture-based narrative discourse. *Revista CEFAC*, *16*(4), 1168–1177. https://doi.org/10.1590/1982-021620147513

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159–191. https://doi.org/10.17791/jcs.2008.9.2.159

Liu, H. (2010). Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua*, *120*, 1567–1578. https://doi.org/10.1016/j.lingua.2009.10.001

Liu, J., Shen, X., Liu, H., & Du, H. (2024). Dependency network approach to the oral production of English- and Chinese-speaking healthy older adults. *Speech, Language and Hearing*, *27*(2), 130–147. https://doi.org/10.1080/2050571X.2023.2241759

Liu, J., Zhao, J., & Bai, X. (2021). Syntactic impairments of Chinese Alzheimer's disease patients from a language dependency network perspective. *Journal of Quantitative Linguistics*, *28*(3), 253–281. https://doi.org/10.1080/09296174.2019.1703485

Lu, Q., Lin, Y., & Liu, H. (2018). Dynamic valency and dependency distance. *Quantitative Analysis of Dependency Structures*, 145–166. https://doi.org/10.1515/9783110573565-008

Lundholm Fors, K., Fraser, K., & Kokkinakis, D. (2018). Automated syntactic analysis of language abilities in persons with mild and subjective cognitive impairment. In A. Ugon, D. Karlsson, G. O. Klein, & A. Moen (Eds.), *Building continents of knowledge in oceans of data: The future of co-created eHealth* (pp. 705–709). IOS Press. https://doi.org/10.3233/978-1-61499-852-5-705

Lyons, K., Kemper, S., LaBarge, E., Ferraro, F. R., Balota, D., & Storandt, M. (1994). Oral language and Alzheimer's disease: A reduction in syntactic complexity. *Aging, Neuropsychology & Cognition*, *1*(4), 271–281. https://doi.org/10.1080/13825589408256581

Malcorra, B. L. C., García, A. O., Marcotte, K., de Paz, H., Schilling, L. P., da Silva Filho, I. G., Soder, R., da Rosa Franco, A., Loureiro, F., & Hübner, L. C. (2024). Exploring spoken discourse and its neural correlates in women with Alzheimer's disease with low levels of education and socio-economic status. *American Journal of Speech-Language Pathology*, *33*(2), 893–911. https://doi.org/10.1044/2023_AJSLP-23-00137

March, E. G., Wales, R., & Pattison, P. (2006). The uses of nouns and deixis in discourse production in Alzheimer's disease. *Journal of Neurolinguistics*, *19*(4), 311–340. https://doi.org/10.1016/j.jneurolinq.2006.01.001

Miller, A., & Chomsky, N. (1963). Finitary models of language users. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 419–491). John Wiley and Sons, Inc.

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. https://doi.org/10.1080/13803395.2018.1446513

Niu, R., & Liu, H. (2022). Effects of syntactic distance and word order on language processing: An investigation based on a psycholinguistic treebank of English. *Journal of Psycholinguistic Research*, *51*(5), 1043–1062. https://doi.org/10.1007/s10936-022-09878-4

Nivre, J., De Marneffe, M. C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., & Zeman, D. (2016, May). Universal dependencies v1: A multilingual treebank collection. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 1659–1666). https://aclanthology.org/L16-1262/

Nivre, J., de Marneffe, M. C., Ginter, F., Hajic, J., Manning, C. D., Pyysalo, S., & Zeman, D. (2020). Universal dependencies v2: An evergrowing multilingual treebank collection. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.), *12th International Conference on Language Resources and Evaluation (LREC), MAY 11-16, 2020*, Marseille, France (pp. 4034–4043). European Language Resources Association. https://aclanthology.org/2020.lrec-1.497/

Orimaye, S. O., Wong, J. S., Golden, K. J., Wong, C. P., & Soyiri, I. N. (2017). Predicting probable Alzheimer's disease using linguistic deficits and biomarkers. *BMC Bioinformatics*, *18*(1), 34. https://doi.org/10.1186/s12859-016-1456-0

Ouyang, J., & Jiang, J. (2023). Typological variation of Chinese-English interlanguage (in Chinese). *Modern Foreign Languages*, *46*(3), 358–370. https://doi.org/10.20071/j.cnki.xdwy.20230221.012

Ouyang, J., Jiang, J., & Liu, H. (2022). Dependency distance measures in assessing L2 writing proficiency. *Assessing Writing*, *51*, 100603. https://doi.org/10.1016/j.asw.2021.100603

Parker, M. D., & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, *25*(3), 365–376. https://doi.org/10.1177/0142723705059114

Parsapoor, M. (2023). AI-based assessments of speech and language impairments in dementia. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *19*(10), 4675–4687. https://doi.org/10.1002/alz.13395

Pistono, A., Jucla, M., Barbeau, E. J., Saint-Aubert, L., Lemesle, B., Calvet, B., Köpke, B., Puel, M., & Pariente, J. (2016). Pauses during autobiographical discourse reflect episodic memory processes in early Alzheimer's disease. *Journal of Alzheimer's Disease*, *50*(3), 687–698. https://doi.org/10.3233/JAD-150408

Pistono, A., Pariente, J., Bézy, C., Lemesle, B., Le Men, J., & Jucla, M. (2019). What happens when nothing happens? An investigation of pauses as a compensatory mechanism in early Alzheimer's disease. *Neuropsychologia*, *124*, 133–143. https://doi.org/10.1016/j.neuropsychologia.2018.12.018

Qiao, Y., Xie, X. Y., Lin, G. Z., Zou, Y., Chen, S. D., Ren, R. J., & Wang, G. (2020). Computer-assisted speech analysis in mild cognitive impairment and Alzheimer's disease: A pilot study from Shanghai, China. *Journal of Alzheimer's Disease*, *75*(1), 211–221. https://doi.org/10.3233/JAD-191056

Qu, Y., Ma, Y. H., Huang, Y. Y., Ou, Y. N., Shen, X. N., Chen, S. D., Yu, J. T., Tan, L., & Yu, J.-T. (2021). Blood biomarkers for the diagnosis of amnestic mild cognitive impairment and Alzheimer's disease: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *128*, 479–486. https://doi.org/10.1016/j.neubiorev.2021.07.007

Roark, B., Mitchell, M., Hosom, J. P., Hollingshead, K., & Kaye, J. (2011). Spoken language derived measures for detecting mild cognitive impairment. *IEEE Transactions on Audio, Speech, and Language Processing*, *19*(7), 2081–2090. https://doi.org/10.1109/TASL.2011.2112351

Ruan, Q., D'Onofrio, G., Sancarlo, D., Bao, Z., Greco, A., & Yu, Z. (2016). Potential neuroimaging biomarkers of pathologic brain changes in mild cognitive impairment and Alzheimer's disease: A systematic review. *BMC Geriatrics*, *16*, 1–9. https://doi.org/10.1186/s12877-016-0281-7

Sand Aronsson, F., Kuhlmann, M., Jelic, V., & Östberg, P. (2021). Is cognitive impairment associated with reduced syntactic complexity in writing? Evidence from automated text analysis. *Aphasiology*, *35*(7), 900–913. https://doi.org/10.1080/02687038.2020.1742282

Snowdon, D. A., Kemper, S. J., Mortimer, J. A., Greiner, L. H., Wekstein, D. R., & Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life: Findings From the Nun Study. *JAMA*, *275*(7), 528–532.

Stark, B. C., Dalton, S. G., & Lanzi, A. M. (2025). Access to context-specific lexical-semantic information during discourse tasks differentiates speakers with latent aphasia, mild cognitive impairment, and cognitively healthy adults. *Frontiers in Human Neuroscience*, *18*, 1500735. https://doi.org/10.3389/fnhum.2024.1500735

Tesnière, L. (1959). *Éléments de syntaxe structurale.*

Tsai, A. C., Hong, S. Y., Yao, L. H., Chang, W. D., Fu, L. C., & Chang, Y. L. (2021). An efficient context-aware screening system for Alzheimer's disease based on neuropsychology test. *Scientific Reports*, *11*(1), 18570. https://doi.org/10.1038/s41598-021-97642-4

Vincze, V., Szabó, M. K., Hoffmann, I., Tóth, L., Pákáski, M., Kálmán, J., & Gosztolya, G. (2022). Linguistic parameters of spontaneous speech for identifying mild cognitive impairment and Alzheimer disease. *Computational Linguistics*, *48*(1), 119–153. https://doi.org/10.1162/coli_a_00428

Walker, G., Morris, L. A., Christensen, H., Mirheidari, B., Reuber, M., & Blackburn, D. J. (2021). Characterising spoken responses to an intelligent virtual agent by persons with mild cognitive impairment. *Clinical Linguistics and Phonetics*, *35*(3), 237–252. https://doi.org/10.1080/02699206.2020.1777586

Wang, R., Kuang, C., Guo, C., Chen, Y., Li, C., Matsumura, Y., Ishimaru, M., Van Pelt, A. J., & Chen, F. (2023). Automatic detection of putative mild cognitive impairment from speech acoustic features in Mandarin-speaking elders. *Journal of Alzheimer's Disease: JAD*, *95*(3), 901–914. https://doi.org/10.3233/JAD-230373

Wang, T., Hong, Y., Wang, Q., Su, R., Ng, M. L., Xu, J., Wang, L., & Yan, N. (2021). Identification of mild cognitive impairment among Chinese based on multiple spoken tasks. *Journal of Alzheimer's Disease: JAD*, *82*(1), 185–204. https://doi.org/10.3233/JAD-201387

Wang, T., Lian, C., Pan, J., Yan, Q., Zhu, F., Ng, M. L., & Yan, N. (2019). Towards the speech features of mild cognitive impairment: Universal evidence from structured and unstructured connected speech of Chinese. *Interspeech*, 3880–3884. https://doi.org/10.21437/Interspeech.2019-2414

Wang, Y., & Liu, H. (2017). The effects of genre on dependency distance and dependency direction. *Language Sciences*, *59*, 135–147. https://doi.org/10.1016/j.langsci.2016.09.006

Wu, W., Chen, G., Ren, X., Zhao, Y., Yu, Z., Peng, H., Song, W., & Song, W. (2024). The prevalence of mild cognitive impairment in China: Evidence from a meta-analysis and systematic review of 393525 adults. *Neuroepidemiology*, *59*(3), 259–276. https://doi.org/10.1159/000539802

Xu, C., & Liu, H. (2022). The role of working memory in shaping syntactic dependency structures. In J. W. Schwieter & Z. E. Wen (Eds.), *The Cambridge handbook of working memory and language* (pp. 343–367). Cambridge University Press. https://doi.org/10.1017/9781108955638.020

Xu, H., & Liu, K. (2023). Syntactic simplification in interpreted English: Dependency distance and direction measures. *Lingua*, *294*, 103607. https://doi.org/10.1016/j.lingua.2023.103607

Xu, L., Chen, K., Mueller, K. D., Liss, J., & Berisha, V. (2025). Articulatory precision from connected speech as a marker of cognitive decline in Alzheimer's disease risk-enriched cohorts. *Journal of Alzheimer's Disease*, *103*(2), 476–486. https://doi.org/10.1177/13872877241300149

Yan, J., & Liu, H. (2021a). Morphology and word order in Slavic languages: Insights from annotated corpora. *Вопросы языкознания*, *4*(4), 131–159. https://doi.org/10.31857/0373-658X.2021.4.131-159

Yan, J., & Liu, H. (2021b). Quantitative analysis of Chinese and English verb valencies based on probabilistic valency pattern theory. In M. Dong, Y. Gu, & J.-F. Hong (Eds.), *Workshop on Chinese lexical semantics* (pp. 152–162). Springer International Publishing. https://doi.org/10.1007/978-3-031-06547-7_12

Yang, M., & Liu, H. (2025). Quansyn: A package for quantitative syntax analysis. *Journal of Quantitative Linguistics*, *32*(2), 181–198. https://doi.org/10.1080/09296174.2025.2471157

Yang, X., & Li, W. (2024). The development of syntactic complexity of Chinese JFL learners based on mean dependency distance and mean hierarchical distance. *International Review of Applied Linguistics in Language Teaching*, *62*(1), 79–104. https://doi.org/10.1515/iral-2023-0010

Yih, T., & Liu, H. (2025). Decomposing dependency analysis: Revisiting the relation between annotation scheme and structure-based textual measures. *Digital Scholarship in the Humanities*, *40*(1), fqaf003. https://doi.org/10.1093/llc/fqaf003

Yih, T., Yang, H., Huang, L., & Yao, Q. (2025). Identifying syntactic biomarkers of cognitive impairment in Mandarin-speaking older adults by applying machine learning approaches across multiple speech tasks. *Aphasiology*. https://doi.org/10.1080/02687038.2025.2511217

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society*, *104*(5), 444–466. https://www.jstor.org/stable/pdf/985230

Zhang, H., & Liu, H. (2018). Interrelations among dependency tree widths, heights and sentence lengths. *Quantitative Analysis of Dependency Structures*, *72*, 31–52. https://doi.org/10.1515/9783110573565-002

Zhang, L.-P. (2017). *The wanderings of Sanmao (in Chinese)*. China Juvenile & Children's Publishing House.

Zhou, D. (2024). Multimodal corpus of geronto discourse: Construction and reflection (in Chinese). *Linguistic Research*, *36*(1), 20–34. https://kns.cnki.net/kcms2/article/abstract?v=F0lYaGTXfx-GCimCpkbJ_ciL3Omj70eyCKbTP1sKb6Sab-bfsU7Ul9tKB1hSerh3xVXskd-Q6VEpPAht3TWOzbMpPCU8KPWIc3AIMl7fEaiC30fdCe165hO7jHTvqytP-hgTSSFOrrOxqYQ_qpoLBMM4WPok11jOy_l2KyNgfnbbzsNBLoI_fg==&uniplatform=NZKPT&language=CHS

## Appendix

**Table A1.** Criterion for deciding the cognitive status of a participant based on MoCA-B scores.

| | Highest education level (Years of education) | | |
|---|---|---|---|
| Cognitive status | Primary school (≤6) | Middle school (7–12) | University or college (≥13) |
| CN | 20–30 | 23–30 | 25–30 |
| MCI | 14–19 | 16–22 | 17–24 |
| Mild dementia | 11–13 | 12–15 | 14–16 |
| Moderate/severe dementia | ≤10 | ≤11 | ≤13 |

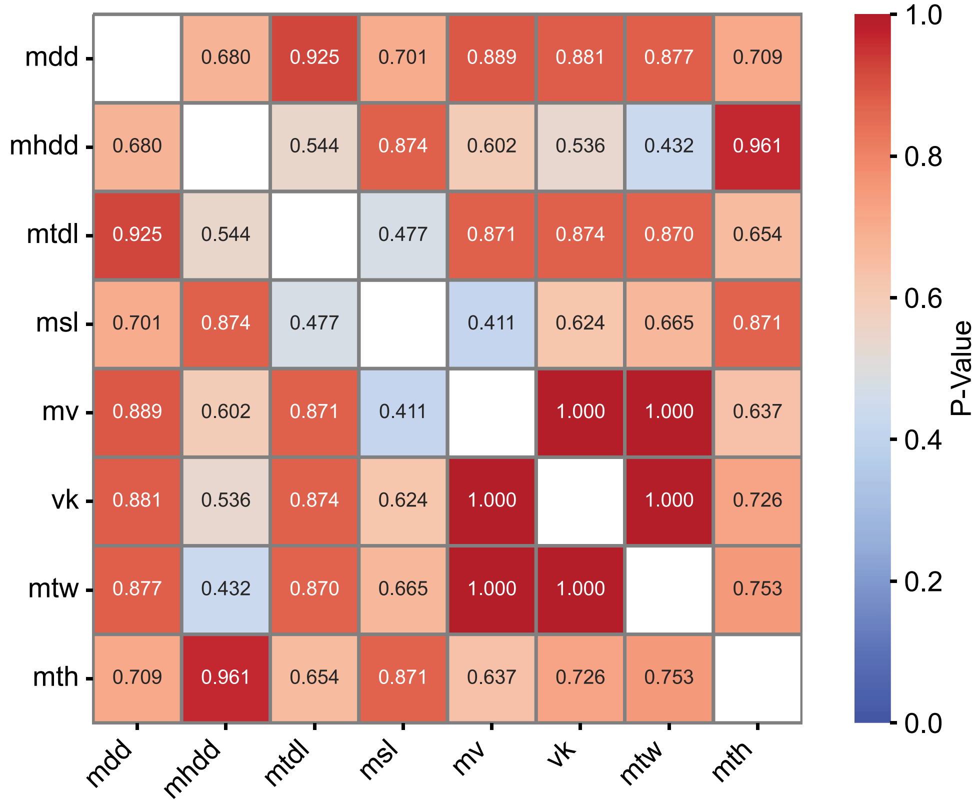


**Figure A1.** The p-values of DeLong tests between similar features.