Cross-Lingual Examination of Language Features and Cognitive Scores From Free Speech

Hali Lindsay[1], **Giorgia Albertin**[2], **Louisa Schwed**[1], **Nicklas Linz**[1], **Johannes Tröger**[1]

[1]ki-elements, Bleichstraße 27, 66111 Saarbrücken, DE
[2]FICLIT-University of Bologna, Via Zamboni, 32, 40126 Bologna BO, Italy
hali.lindsay@ki-elements.de

Abstract

Speech analysis is gaining significance for monitoring neurodegenerative disorders, but with a view of application in clinical practice, solid evidence of the association of language features with cognitive scores is still needed. A cross-linguistic investigation has been pursued to examine whether language features show significance correlation with two cognitive scores, i.e. Mini-Mental State Examination and ki:e SB-C scores, on Alzheimer's Disease patients. We explore 23 language features, representative of syntactic complexity and semantic richness, extracted on a dataset of free speech recordings of 138 participants distributed in four languages (Spanish, Catalan, German, Dutch). Data was analyzed using the speech library SIGMA; Pearson's correlation was computed with Bonferroni correction, and a mixed effects linear regression analysis is done on the significant correlated results. MMSE and the SB-C are found to be correlated with no significant differences across languages. Three features were found to be significantly correlated with the SB-C scores. Among these, two features of lexical richness show consistent patterns across languages, while determiner rate showed language-specific patterns.

Keywords: Language features, Cross-linguistic analyses, Alzheimer's Disease

1. Introduction

Speech analysis for Alzheimer's Disease (AD) diagnosis holds promise for facilitating timely interventions and improving patient outcomes through early detection and personalized care strategies (Vigo et al., 2022). Language deficits, alongside episodic memory impairment, are hallmark symptoms of AD even in its early stages (Drummond et al., 2015; Szatloczki et al., 2015). The process of using speech to enhance screening and provide support for AD diagnosis has been a popular research topic in recent years, also enhanced by the increasing application of Natural Language Processing (NLP) and Machine Learning (ML) technologies in this domain (De la Fuente Garcia et al., 2020). Despite the growing research of NLP and ML technologies in analyzing speech and language features, particularly in Alzheimer's Disease (AD) diagnosis, challenges such as small datasets and low repeatability (Stegmann et al., 2020) and susceptibility to overfitting (Berisha et al., 2022) hinder the generalizability of results. While leveraging NLP and ML methodologies provides expedited and cost-effective means of assessing cognitive decline through spontaneous speech analysis, it is imperative to establish robust associations between linguistic features and cognitive decline to ensure their clinical utility (De la Fuente Garcia et al., 2020).

Exploring linguistic markers of cognition across languages offers a valuable avenue for research, emphasizing the profound insights it provides into cognitive processes across diverse linguistic backgrounds. The early detection of AD through linguistic analysis faces challenges in translating research findings into clinical practice (Berisha et al., 2022). Small datasets and a plethora of potential features hinder generalizability, while the lack of clinical context further complicates matters. To address this, exploring the consistency of discriminative features across different languages offers a novel approach. By examining linguistic patterns, researchers gain a deeper understanding of cognition and language-specific influences. Comparative analysis facilitates the identification of commonalities and differences in linguistic markers associated with cognition, contributing to theoretical advancements in cognitive science and linguistics. Ultimately, studying linguistic markers of cognition across languages adds generalizability through multilingual feature statistics to computational approaches for the detection of language impairment in AD. If these language features demonstrate consistent patterns of cognitive performance across multiple languages, it suggests they capture relevant cognitive aspects, enhancing their potential for clinical use (Lindsay et al., 2021).

In this study, the investigation focuses on understanding cognitive decline across four different indo-european languages (Catalan, Spanish, German, and Dutch) by analyzing specific language features. The goal is to determine whether these language features can provide insights into cognitive decline, regardless of the language spoken. Two clinical score are considered: the Mini Mental State Ex-

Proceedings of the RaPID-5 @LREC-COLING 2024, pages 16–25
21 May, 2024.

Language	N	Age	MMSE	SB-C
Spanish	18	65.46(7.40)	29.33(1.08)	0.42(0.11)
Catalan	16	67.14(6.73)	28.56(1.46)	0.39(0.08)
German	43	68.57(5.69)	28.88(1.16)	0.46(0.11)
Dutch	61	64.02(10.76)	28.11(1.73)	0.33(0.11)

Table 1: Demographic information for the participants. The Mini-Mental State Exam (MMSE) is a test to measure cognitive function (Max score 30) The SB-C is a composite score of automatically extracted speech features. Means are given with standard deviation in parentheses.

amination (MMSE) (Folstein et al., 1975) and the ki-element's SB-C (Speech Biomarker-Cognition) (Tröger et al., 2022), are used to measure cognitive function. The MMSE is a traditional cognitive screening tool administered by a clinician in the clinic, where as the SB-C is an automatically extracted marker that can be administered in either the clinic or remotely over the phone. By comparing the results of these tests with features extracted from individuals' speech, the study aims to identify if language can serve as an indicator of cognitive health across different languages. Additionally, the study explores whether the SB-C test yields results similar to the MMSE in various linguistic contexts.

2. Background

2.1. Cognitive Scores

The Mini-Mental State Examination (MMSE) is a widely-used cognitive screening tool comprised of several tasks assessing various cognitive domains, including orientation, memory, attention, language, and visuospatial abilities (Folstein et al., 1975). With a total score ranging from 0 to 30, the MMSE provides a quantitative measure of cognitive function, with higher scores indicative of better cognitive performance. Tasks within the MMSE include orientation to time and place, immediate and delayed recall of words or phrases, serial subtraction, naming of objects, repetition of sentences, and copying a complex figure. Administration of the MMSE typically takes around 10 minutes and can be easily conducted by healthcare professionals or trained administrators. Due to its brevity and simplicity, the MMSE is commonly used in clinical settings to screen for cognitive impairment, monitor cognitive changes over time, and inform treatment planning.

The ki:e SB-C (Tröger et al., 2022) is a composite score comprised of over 50 automatically extracted speech features, which are organized into three distinct neurocognitive domains: learning and memory, executive function, and processing speed. These domains are utilized to generate a single aggregated global cognition score. The ki:e SB-C utilizes speech recordings from two standard neuropsychological assessments, the Rey Auditory Verbal Learning Test (RAVLT) and the Semantic Verbal Fluency task (SVF). These speech recordings undergo automatic processing via the proprietary speech analysis pipeline from ki:elements, which includes automatic speech recognition and feature extraction. Following this processing, domain scores and the global cognition score are calculated. The ki:e SB-C can be collected automatically via traditional landline phone infrastructure or in face-to-face on-site settings using mobile front ends (Konig et al., 2018). The SB-C does not currently make use of pure language features from free speech that are described in the following section.

2.2. Language Features

Cognitive decline profoundly impacts language abilities, as evidenced by changes observed in free speech tasks among individuals with neurodegenerative disorders such as Alzheimer's Disease (Slegers et al., 2018; Deters et al., 2017). As cognitive functions deteriorate, language skills deteriorate, manifesting in various linguistic deficits. These deficits may include reductions in vocabulary richness, syntactic complexity, and semantic coherence, as well as increased hesitations, pauses, and speech errors (Fraser et al., 2016; Ammar and Ayed, 2020; Mueller et al., 2018). Individuals experiencing cognitive decline often exhibit difficulties in generating coherent narratives, organizing thoughts logically, and maintaining topic coherence during free speech tasks (Slegers et al., 2018). Moreover, declines in executive functions, such as attention, planning, and inhibition, further exacerbate language impairments by impairing the individual's ability to monitor and regulate speech production (Gonçalves et al., 2018). Consequently, changes in language abilities observed in free speech tasks serve as valuable markers of cognitive decline and are instrumental in assessing the progression of neurodegenerative diseases. Understanding the intricate relationship between cognitive decline and language abilities in free speech tasks is essential for developing effective diagnostic and intervention strategies for individuals affected by neurodegenerative disorders.

The linguistic features selected for extraction in this study predominantly encompass morpho-

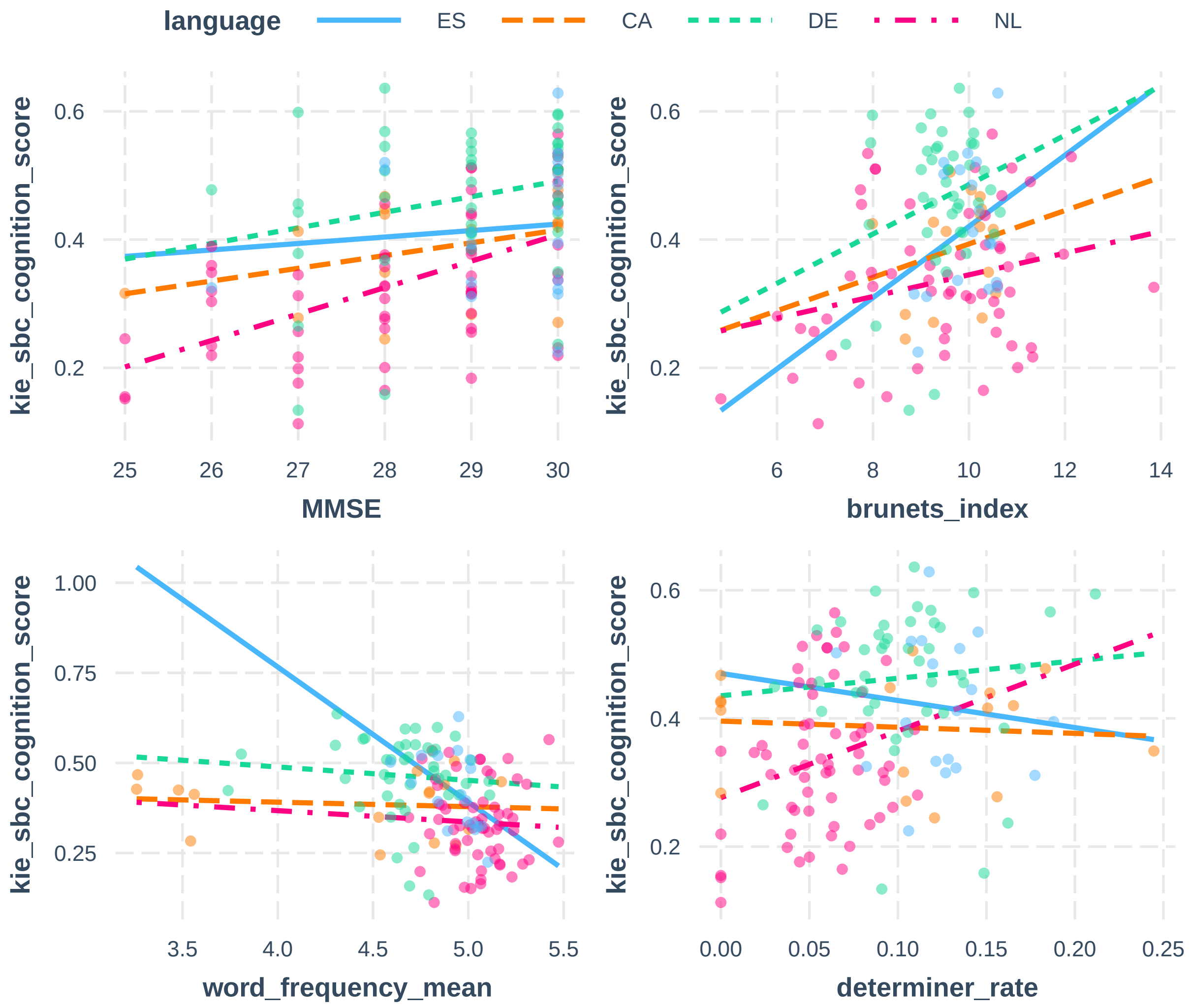

Figure 1: Interaction plots from the mixed effects linear regressions for significantly correlated language features and MMSE with the SB-C. Points represent individual scores where as the lines denote the overall trend from the linear model. (ES)Spanish, (CA)Catalan, (DE)German, (NL)Dutch.

syntactic aspects. These features include the rates of various part-of-speech categories such as adjectives, adpositions, adverbs, conjunctions, determiners, inflected and total verbs, nouns, pronouns, and proper nouns. Additionally, indices of lexical richness, including Brunet's Index, Honoré's Statistic, and the Type-Tokens ratio, were calculated (Hernández-Domínguez et al., 2018).

Furthermore, features were chosen to explore syntactic structures, such as the mean number of subordinate clauses in a sentence, the proportion of verb phrases with objects and subjects, and the number of verb phrases with auxiliaries. General aspects of language, such as word count, word frequency (mean, standard deviation, and range), and the number of consecutive repetitions, were also included.

The word count and number of consecutive repetitions serve as indicators of response amount and fluency, respectively. Semantic richness is assessed through features like adjective rate, Brunet's Index, Honoré's Statistic, noun rate, proper noun rate, type-token ratio, and word frequency, which tap into semantic memory and lexical retrieval abilities (Hernández-Domínguez et al., 2018).

Higher rates of morpho-syntactic features are anticipated to correlate positively with the MMSE and SB-C, reflecting stronger cognitive abilities. Lower Honoré's statistic and Larger Brunet's Index values may indicate efficient word retrieval processes and a larger mental lexicon, while word frequency can reveal vocabulary knowledge and lexical access abilities (Deepa and Shyamala, 2010).

Syntactic complexity is monitored by adposition and adverb rates, reflecting grammatical proficiency and syntactic processing abilities. Features like subordinate clauses and conjunction rate introduce additional information or qualifications to main clauses, allowing for the expression of complex relationships and ideas (Lindsay et al., 2021).

Determiners provide insights into the specificity, definiteness, or quantity of nouns, suggesting extensive semantic processing and comprehension abilities with higher determiner rates. Pronoun rates may indicate stronger theory of mind abilities, contributing to narrative coherence and discourse cohesion through referential continuity.

Moreover, higher verb rates suggest faster cognitive processing speed and play a crucial role in establishing narrative structure and discourse coherence.

3. Data

This study considered a total of 138 participant who completed a one minute free speech task (e.g. tell me about your last vacation) in one of 4 languages; Spanish, Catalan, German and Dutch. The German, Spanish, and Catalan data was collected as part of the Prospect AD study (König et al., 2023). In this clinical study, speech protocol of neurocognitive tests—including the a word list test, verbal fluency task, and spontaneous free speech to assess psychological and/or behavioral symptoms—is administer remotely, via a phone call.

For the Dutch participants, the study recruited participants from the memory clinic of the Maastricht University Medical Center+, where a test leader facilitated a semi-automated phone assessment. The test battery included a verbal learning test (VLT), semantic verbal fluency (SVF), and free speech assessment were administered as part of this comprehensive evaluation (Ter Huurne et al., 2023). Part of this study completed an analysis comparing ASR and manual transcripts for the SVF and VLT and found a high agreement between the ASR and manual scores.

The demographic data for the sample population is given in Table 1.

4. Methods

Linguistic features were extracted using SIGMA, a proprietary pipeline for speech and language feature extraction. SIGMA incorporates a comprehensive suite of linguistic analysis tools, providing insights into various language dimensions such as lexical richness, syntactic complexity, and discourse coherence. The transcription of data was automated through Google Automatic Speech Recognition (ASR)[1], ensuring consistency and efficiency in data processing. Additionally, part-of-speech tagging was performed using the python library Stanza, a natural language processing toolkit, to identify and label the grammatical categories of words within the transcribed text (Qi et al., 2020).

Once transcribed, 23 language features were extracted from each transcript. These features included various linguistic aspects, including the rates of adjectives, adpositions, adverbs, conjunctions, determiners, inflected verbs, nouns, and pronouns, as well as verbs. Additionally, type token ratio (TTR), Brunet's index (Brunet et al., 1978) and Honore's Statistic (Honoré et al., 1979) were calculated as measures of vocabulary richness (Ntracha et al., 2020). Furthermore, word count, number of consecutive repetitions, and descriptive statistics such as word frequency mean, standard deviation, and range were extracted. Finally, syntactic features were considered, including the mean number of subordinate clauses and various measures related to the complexity of verb phrases.

A full list of extracted features is given in Table 2 and feature descriptions are given in Section 2.2.

4.1. Correlation Analysis

To explore the relationship between MMSE and SB-C scores and various language features, we calculated Pearson's correlation coefficient (r). This statistic helps us understand the strength and direction of the linear connection between the two continuous variables (cognitive score and language feature), ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no linear relationship. A significant correlation suggests that the observed association is unlikely due to chance alone, indicating a meaningful connection in the population.

Considering the multiple comparisons made, we applied the Bonferroni correction to control for Type I error. This method adjusts the significance threshold by dividing the standard alpha level (0.05) by the number of comparisons conducted.

We report Pearson's correlation coefficients and their corresponding p-values after Bonferroni correction. Statistical significance was determined with a threshold of $p < 0.05$, adjusted for multiple comparisons.

All analyses, including correlations, significance testing, and Bonferroni correction, were performed in Python 3.9 using the scipy library (Virtanen et al., 2020).

4.2. Linear Mixed-Effects Modeling

To investigate the effects of cognition and language, a post-hoc linear regression mixed-effects model was used to explore the relationship between cognitive scores (MMSE or SB-C) and each significantly correlated language feature, while considering potential variations across languages.

[1]Google. Google Speech API, Available from: https://cloud.google.com/speech-to-text/

Table 2: Correlation coefficients (r) and statistical significance (p) for the relationships between cognitive scores (MMSE and SB-C) and linguistic features, along with mean (μ) and standard deviation (σ) values for each feature across different languages (Spanish, Catalan, German, and Dutch).

	MMSE		SB-C		$\mu(\sigma)$			
Feature	r	p	r	p	Spanish	Catalan	German	Dutch
MMSE	-	-	0.478	0.00	29.33(1.09)	28.56(1.46)	28.88(1.16)	28.12(1.73)
adjective rate	0.11	1.0	0.05	1.0	0.07(0.03)	0.04(0.04)	0.05(0.03)	0.07(0.04)
adposition rate	-0.21	0.40	-0.11	1.0	0.10(0.04)	0.07(0.06)	0.08(0.03)	0.10(0.03)
adverb rate	-0.11	1.0	-0.14	1.0	0.08(0.03)	0.06(0.06)	0.12(0.05)	0.12(0.05)
Brunets Index	0.23	0.20	0.27	0.04	9.94(0.57)	9.73(0.77)	9.44(0.72)	9.33(1.79)
conjunction rate	0.09	1.0	0.06	1.0	0.10(0.03)	0.07(0.05)	0.06(0.03)	0.07(0.03)
determiner rate	0.25	0.07	0.29	0.01	0.13(0.03)	0.10(0.08)	0.11(0.04)	0.06(0.03)
honore stat	-0.01	1.0	-0.08	1.0	1928.2(348.5)	2189.2(612.7)	2418.6(1113.0)	2538.0(1590.3)
inflected verb rate	0.12	1.0	0.21	0.32	0.73(0.16)	0.38(0.29)	0.73(0.15)	0.62(0.27)
mean number subordinate clauses	0.13	1.0	0.12	1.0	5.36(5.55)	8.75(6.99)	1.08(1.63)	0.07(0.16)
noun rate	0.08	1.0	0.07	1.0	0.14(0.03)	0.10(0.08)	0.14(0.04)	0.14(0.05)
number consecutive repetitions	0.09	1.0	0.08	1.0	0.50(0.86)	0.31(0.70)	0.23(0.53)	0.54(0.91)
pronoun rate	0.04	1.0	-0.04	1.0	0.12(0.04)	0.06(0.05)	0.13(0.04)	0.12(0.03)
proper noun rate	-0.02	1.0	-0.04	1.0	0.01(0.01)	0.32(0.47)	0.02(0.03)	0.04(0.04)
proportion verb phrase with objects	0.20	0.55	0.16	1.0	0.37(0.08)	0.26(0.19)	0.30(0.13)	0.18(0.14)
proportion verb phrase with subjects	0.15	1.0	0.06	1.0	0.62(0.20)	0.55(0.44)	0.79(0.16)	0.72(0.19)
type token ratio	-0.14	1.0	-0.19	0.70	0.65(0.06)	0.72(0.08)	0.70(0.07)	0.71(0.14)
verb phrase with aux and vp rate	0.11	1.0	0.08	1.0	0.01(0.02)	0.06(0.09)	0.04(0.06)	0.02(0.04)
verb phrase with aux rate	0.07	1.0	-0.08	1.0	0.27(0.24)	0.35(0.36)	0.45(0.22)	0.40(0.49)
verb rate	0.01	1.0	-0.06	1.0	0.12(0.02)	0.09(0.07)	0.11(0.03)	0.10(0.04)
word count	0.13	1.0	0.15	1.0	102.7(25.4)	101.9(30.9)	92.2(26.9)	108.2(74.3)
word frequency mean	-0.15	1.0	-0.27	0.04	4.93(0.14)	4.39(0.70)	4.66(0.28)	5.05(0.17)
word frequency sd	0.14	1.0	0.16	1.0	0.82(0.11)	1.12(0.24)	0.92(0.15)	0.91(0.18)
word frequency range	0.18	1.0	0.19	0.60	3.53(0.73)	4.66(0.58)	3.82(0.75)	3.97(1.00)

A linear regression model was used to investigate the relationship between cognition scores and language features, while also considering the interaction between language features and language. The dependent variable is represented by $CogScore$. The fixed effects of the model were defined as the language feature and language $(Feature \times Language)$, as well as their interaction, which allows for the assessment of how language features influence cognition scores across different languages.

$$CogScore \sim Feature \times Language + (1 \mid Language) \quad (1)$$

The models consider potential correlation among observations from the same language group by incorporating a random intercept for language, $(1 \mid language)$.

5. Results

5.1. What is the relationship between the MMSE and SB-C?

The MMSE and SB-C showed significant (p=0.00) positive correlations across all for languages (r=0.478). As the MMSE increases the SB-C also increases. The MMSE did not show significant correlations with any language features in this analysis. However, the SB-C showed significant correlations with three features: Brunet's Index, determiner rate, and mean word frequency.

In addition, to the feature models, we also examined the relationship between the MMSE and SB-C across the four languages using a mixed effects linear regression model. Results for the linear model are visualized in the top left corner of Figure 1. Our analysis of fixed effects revealed that neither MMSE nor language had a statistically significant difference with SB-C scores. In addition, interaction terms between MMSE and language also failed to show significant effects on SB-C scores. Examining the random effect of language showed minimal variability between language groups, with a low Intraclass Correlation Coefficient (ICC=0.016), suggesting negligible group-level differences relative to total variability. Overall, our findings indicate that there were no significant differences in cognitive abilities measured by SB-C in relation to MMSE or across the languages studied.

5.2. Do language features generalize across languages?

In our study, we analyzed 23 linguistic features extracted from the free speech task conducted in four different languages. Surprisingly, none of these features showed significant correlations with the Mini-Mental State Examination (MMSE). However, when examining the Subjective Cognitive Decline (SB-C), three linguistic features stood out: Brunet's Index, determiner rate, and mean word frequency.

Brunet's Index, a measure of lexical richness, revealed a consistent positive correlation with SB-C scores across all four languages. This suggests that individuals with higher cognitive function tended to produce speech that was more diverse and varied in vocabulary.

Similarly, we found a negative correlation between average word frequency and SB-C scores. This implies that individuals with lower cognitive scores tended to use more common words, while those with higher cognitive scores used less common words, indicating a greater lexical sophistication.

Interestingly, determiner rate exhibited distinct patterns of correlation based on language family. In Germanic languages such as German and Dutch, we observed an increase in determiner rate with higher cognitive performance. Conversely, Romance languages like Spanish and Catalan showed a mild negative trend, where lower cognitive scores were associated with higher determiner rates. These findings underscore the complexity of linguistic patterns in relation to cognitive function across different language groups.

5.3. How do cognition and language influence language features?

In our study, we employed linear mixed effects models to investigate the factors influencing cognition scores using data from 137 observations. The cognition score (SB-C) served as the dependent variable. The models demonstrated good fit, with AIC values ranging from -169.098 to -189.176 and BIC values from -139.898 to -159.976. The pseudo-R^2 values indicated that the fixed effects accounted for 14.3% (determiner rate), 21.9% (Brunet's Index), and 23.3% (mean word frequency) of the variance in cognition scores, while the total model explained 24.7% (mean word frequency), 38.8% (Burnet's Index), 53.9% (determiner rate) of the variance.

Across the models, no significant main effects of language features, such as Brunet's Index, word frequency mean, or determiner rate, were found on cognition scores. Additionally, the language did not exhibit significant main effects on cognition scores.

Interaction effects between language features and language variables were explored but did not reach statistical significance, suggesting that the relationship between these language features and cognition scores did not significantly vary across different languages.

Analysis of random effects revealed variability between language groups, with moderate to high

Intraclass Correlation Coefficients (ICCs) of 0.018 (mean word frequency), 0.217 (Burnet's Index), and 0.462 (determiner rate). This suggests that differences between language groups accounted for a portion of the total variance in cognition scores.

Overall, our findings suggest that while certain language features may play a role in predicting cognition scores, their effects were not statistically significant in our study. Further research is needed to explore other factors that may contribute to variability in cognition scores across different language groups.

6. Discussion

The results of this study demonstrate a significant correlation between both the SB-C and MMSE scores and language features across four distinct languages. Notably, these languages represent different linguistic families, with Spanish and Catalan belonging to the Romanic group, while German and Dutch fall under the Germanic category. This cross-linguistic correlation of cognitive scores suggests that certain speech-derived features for lexical richness may exhibit a consistent relationship with cognition that can be generalized across languages. Although variations in the overall means of the features are observed, the patterns of correlation with cognition remain consistent across languages, as depicted in Table 2.

The positive correlation observed between SB-C scores and language features associated with lexical richness, such as Brunet Index and average word frequency, indicates an association between a richer vocabulary and higher cognitive function. This finding aligns with existing literature suggesting a link between mental lexicon and cognition, although this relationship becomes more complex with age due to various factors beyond cognitive decline. These factors include alterations in the ability to learn new word-concept associations, influenced by prior knowledge(Wulff et al., 2019). Additionally, compromised word retrieval and verbal fluency, observed in language disruptions in Alzheimer's Disease (AD), may affect the richness of vocabulary(Taler and Phillips, 2008).

The significant relationship between vocabulary richness features (Brunet's Index and word frequency mean) and both the MMSE and SB-C suggests that these linguistic measures may serve as indicators of cognitive ability. This implies that individuals with higher cognitive function, as measured by MMSE and SB-C, tend to exhibit richer and more diverse vocabularies. This outcome can be anticipated, considering that the MMSE and SB-C primarily evaluate cognitive ability, which is likely being assessed by the vocabulary richness features.

In addition to Brunet's Index and mean word frequency, another linguistic feature, determiner rate, showed a significant positive correlation (r=0.29) with cognitive score. However, this correlation revealed a more nuanced relationship across languages, as illustrated in Figure 1. While there was an overall positive correlation between determiner rate and cognitive score, distinct language-specific patterns emerged between the Germanic (Dutch and German) and Romance languages (Catalan and Spanish). Specifically, the trend indicated a positive relationship between determiner rate and cognitive score in Germanic languages, suggesting that higher cognitive function was associated with a greater use of determiners. In contrast, the inverse relationship was observed in Romance languages, where lower cognitive scores were associated with higher determiner rates. Determiners, including articles like "the" and "a/an," as well as demonstratives such as "this" and "that," are essential for shaping sentences and communicating meaning in Romance and Germanic languages. However, their impact on cognitive load might vary across these language groups. This variability could stem from differences in morphological complexity, inflectional patterns, and agreement rules inherent in these languages (Foucart et al., 2010). These findings highlight the complexity of linguistic patterns and their associations with cognitive function, emphasizing the need for language-specific analyses in cognitive research

In our study, we observed that the Speech-Based Cognition Score (SB-C) correlated with language features, while the Mini-Mental State Examination (MMSE) did not. One speculative insight into this discrepancy is the difference in the spread of scores represented in the data. The MMSE scores in our relatively healthy population were consistently above 25, indicating a ceiling effect and limited variability. In contrast, the SB-C exhibited a more continuous distribution with greater spread.

A cognitive score with a broader spread of values provides more variability in the data, enhancing its sensitivity to changes and differences. This increased variability allows for the capture of more nuanced relationships and may lead to stronger correlations with other variables, such as language features. Therefore, the broader spread of scores in the SB-C may explain why it exhibited stronger correlations with language features compared to the MMSE. This speculation suggests that the nature of the cognitive score, particularly its variability, influences its ability to capture associations with language features.

The Mini-Mental State Examination (MMSE) is a widely used tool for screening cognitive impairment and diagnosing cognitive impairment, offering brevity, ease of administration, and assessment across multiple cognitive domains. While it facilitates diagnosis, limitations such as reduced sensitivity to mild cognitive impairment and lack of specificity(de Jager et al., 2009; Shiri-Feshki, 2009; Tombaugh and McIntyre, 1992), especially in diverse populations, warrant consideration for optimizing its utility in diagnosing AD .

An objective marker of cognition based on speech tasks, such as the SB-C, offers a promising avenue to address some MMSE limitations . By providing an objective and quantifiable measure of cognitive function through linguistic features analysis, it offers nuanced insights into cognitive abilities, including executive function. Integrating such speech-based markers into clinical practice could complement traditional assessments like the MMSE, enhancing the comprehensive and objective diagnosis of AD and cognitive decline progression.

Several limitations should be acknowledged when interpreting the findings of this study. Firstly, the lack of control for education level introduces a potential confounding factor that may influence over results pertaining to cognition. Additionally, the variability in the spread of MMSE scores among different language groups, as illustrated in the figure, reveals disparities in cognitive states across participants. Specifically, Spanish, Catalan, and German participants exhibit mild to no signs of cognitive impairment, where all participants have an MMSE score above 25, indicating there is no confirmed clinical impairment at the time of this analysis. These variations highlight the need for caution when generalizing findings across diverse linguistic backgrounds.

Future work should involve the manual annotation of the speech data to compute the Word Error Rate (WER) to examine the reliability of the automatic speech recognition. While ASR is currently used in the field to transcribe speech into text, there remains an important need to assess its accuracy and performance under various linguistic contexts. One direction is to investigate whether there are differing rates of reliability in ASR systems based on the overall popularity of the language being evaluated. Languages with larger speaker populations or more extensive linguistic resources may have better ASR performance due to the availability of training data and language models. Conversely, less widely spoken languages or those with limited resources may present greater challenges for ASR systems, leading to higher error rates. This is also confounded by using ASR on older populations, where a higher error rate may be expected as older speakers are not typically used to train these systems.

7. Conclusion

This paper set out to investigate the potential correlation between language and cognition in a cross-lingual setting. We find a strong correlation the two markers of cognition, MMSE and SB-C scores. In addition, language features indicating lexical richness (Brunet's Index and mean word frequency) were consistent across four languages: Spanish, Catalan, German and Dutch. In addition, we identify determiner rate as a feature that shows an overall significant positive correlation but differs between language groups This indicates that some language features may be indicative of cognition while displaying inverse relationships due to other factors. Future research endeavors may consider mapping language phenomena of cognition with a comprehensive language score, with the aim of captures patterns of generalizability among language-specific properties.

8. Bibliographical References

Randa Ben Ammar and Yassine Ben Ayed. 2020. Language-related features for early detection of alzheimer disease. *Procedia Computer Science*, 176:763–770.

Visar Berisha, Chelsea Krantsevich, Gabriela Stegmann, Shira Hahn, and Julie Liss. 2022. Are reported accuracies in the clinical speech machine learning literature overoptimistic? *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2022-September:2453–2457.

Dagmar Bittner, Claudia Frankenberg, and Johannes Schröder. 2024. Pronoun use in preclinical and early stages of alzheimer's dementia. *Computer Speech & Language*, 84:101573.

Étienne Brunet et al. 1978. *Le vocabulaire de Jean Giraudoux structure et évolution*. Slatkine.

Celeste A de Jager, Anne-Claire MC Schrijnemaekers, Thurza EM Honey, and Marc M Budge. 2009. Detection of mci in the clinic: evaluation of the sensitivity and specificity of a computerised test battery, the hopkins verbal learning test and the mmse. *Age and ageing*, 38(4):455–460.

Sofia De la Fuente Garcia, Craig W Ritchie, and Saturnino Luz. 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.

MS Deepa and KC Shyamala. 2010. Complex discourse production in persons with mild dementia: Measures of richness of vocabulary. *Journal of the All India Institute of Speech & Hearing*, 29(1).

Kacie D Deters, Kwangsik Nho, Shannon L Risacher, Sungeun Kim, Vijay K Ramanan, Paul K Crane, Liana G Apostolova, Andrew J Saykin, Alzheimer's Disease Neuroimaging Initiative, et al. 2017. Genome-wide association study of language performance in alzheimer's disease. *Brain and language*, 172:22–29.

Cláudia Drummond, Gabriel Coutinho, Rochele Paz Fonseca, Naima Assunção, Alina Teldeschi, Ricardo de Oliveira-Souza, Jorge Moll, Fernanda Tovar-Moll, and Paulo Mattos. 2015. Deficits in narrative discourse elicited by visual stimuli are already present in patients with mild cognitive impairment. *Frontiers in aging neuroscience*, 7:96.

Marshal F Folstein, Susan E Folstein, and Paul R McHugh. 1975. "mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*, 12(3):189–198.

Alice Foucart, Holly P Branigan, and Ellen G Bard. 2010. Determiner selection in romance languages: Evidence from french. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6):1414.

Kathleen C Fraser, Jed A Meltzer, and Frank Rudzicz. 2016. Linguistic features identify alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, 49(2):407–422.

Ana Paula Bresolin Gonçalves, Clarissa Mello, Andressa Hermes Pereira, Perrine Ferré, Rochele Paz Fonseca, and Yves Joanette. 2018. Executive functions assessment in patients with language impairment a systematic review. *Dementia & neuropsychologia*, 12:272–283.

Laura Hernández-Domínguez, Sylvie Ratté, Gerardo Sierra-Martínez, and Andrés Roche-Bergua. 2018. Computer-based evaluation of alzheimer's disease and mild cognitive impairment patients during a picture description task. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, 10:260–268.

Antony Honoré et al. 1979. Some simple measures of richness of vocabulary. *Association for literary and linguistic computing bulletin*, 7(2):172–177.

Alexandra König, N Linz, E Baykara, J Tröger, C Ritchie, S Saunders, S Teipel, S Köhler, G Sánchez-Benavides, O Grau-Rivera, et al. 2023. Screening over speech in unselected populations for clinical trials in ad (prospect-ad): study design and protocol. *The journal of prevention of Alzheimer's disease*, 10(2):314–321.

Alexandra Konig, Aharon Satt, Alex Sorin, Ran Hoory, Alexandre Derreumaux, Renaud David, and Phillippe H Robert. 2018. Use of speech analyses within a mobile application for the assessment of cognitive impairment in elderly people. *Current Alzheimer Research*, 15(2):120–129.

Hali Lindsay, Johannes Tröger, and Alexandra König. 2021. Language impairment in alzheimer's disease—robust and explainable evidence for ad-related deterioration of spontaneous speech through multilingual machine learning. *Frontiers in aging neuroscience*, 13:642033.

Kimberly D Mueller, Bruce Hermann, Jonilda Mecollari, and Lyn S Turkstra. 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.

Anastasia Ntracha, Dimitrios Iakovakis, Stelios Hadjidimitriou, Vasileios S Charisis, Magda Tsolaki, and Leontios J Hadjileontiadis. 2020. Detection of mild cognitive impairment through natural language and touchscreen typing processing. *Frontiers in Digital Health*, 2:567158.

Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082*.

RStudio Team. 2020. *RStudio: Integrated Development Environment for R*. RStudio, PBC., Boston, MA.

Mojtaba Shiri-Feshki. 2009. Rate of progression of mild cognitive impairment to dementia-meta-analysis of 41 robust inception cohort studies.

Antoine Slegers, Renee-Pier Filiou, Maxime Montembeault, and Simona Maria Brambati. 2018. Connected speech features from picture description in alzheimer's disease: A systematic review. *Journal of Alzheimer's disease*, 65(2):519–542.

Gabriela M Stegmann, Shira Hahn, Julie Liss, Jeremy Shefner, Seward B Rutkove, Kan Kawabata, Samarth Bhandari, Kerisa Shelton, Cayla Jessica Duncan, and Visar Berisha. 2020. Repeatability of commonly used speech and language features for clinical applications. *Digital biomarkers*, 4(3):109–122.

Greta Szatloczki, Ildiko Hoffmann, Veronika Vincze, Janos Kalman, and Magdolna Pakaski. 2015. Speaking in alzheimer's disease, is that an early sign? importance of changes in language abilities in alzheimer's disease. *Frontiers in aging neuroscience*, 7:195.

Vanessa Taler and Natalie A Phillips. 2008. Language performance in alzheimer's disease and mild cognitive impairment: a comparative review. *Journal of clinical and experimental neuropsychology*, 30(5):501–556.

Daphne Ter Huurne, Nina Possemis, Leonie Banning, Angélique Gruters, Alexandra König, Nicklas Linz, Johannes Tröger, Kai Langel, Frans Verhey, Marjolein De Vugt, et al. 2023. Validation of an automated speech analysis of cognitive tasks within a semiautomated phone assessment. *Digital biomarkers*, 7(1):115–123.

Tom N Tombaugh and Nancy J McIntyre. 1992. The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society*, 40(9):922–935.

Johannes Tröger, Ebru Baykara, Jian Zhao, Daphne Ter Huurne, Nina Possemis, Elisa Mallick, Simona Schäfer, Louisa Schwed, Mario Mina, Nicklas Linz, et al. 2022. Validation of the remote automated ki: E speech biomarker for cognition in mild cognitive impairment: Verification and validation following dime v3 framework. *Digital biomarkers*, 6(3):107–116.

Ines Vigo, Luis Coelho, and Sara Reis. 2022. Speech-and language-based classification of alzheimer's disease: a systematic review. *Bioengineering*, 9(1):27.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.

Dirk U Wulff, Simon De Deyne, Michael N Jones, and Rui Mata. 2019. New perspectives on the aging lexicon. *Trends in cognitive sciences*, 23(8):686–698.