Can we track the progression of Alzheimer's Disease via lexical-semantic variables in connected speech?

Marte Mestach[1], Robert J. Hartsuiker[1] and Aurélie Pistono[1]

[1]Department of Experimental Psychology, Ghent University, Belgium

Correspondence

Aurélie Pistono

Ghent University, Department of Experimental Psychology

Henri Dunantlaan 2

9000 Gent (Belgium)

aurelie.pistono@ugent.be

Abstract

Background: Alzheimer's disease (AD) is one of the most common neurodegenerative disorders worldwide and is characterized by problems with cognition and language, especially word-finding difficulties. The present study focuses on lexical-semantic features via five discourse variables reflecting word-finding difficulties, namely indefinite terms, lexical frequency, repetitions, semantic paraphasias, and use of pronouns. Our aim is twofold: testing whether these variables can discriminate healthy aging from AD, but also mild from moderate AD.

Method: 105 participants were examined from the existing Pitt corpus (available on DementiaBank), which includes the Cookie Theft Picture Description task. 40 participants were healthy controls, 25 were mild AD participants, and 40 moderate AD participants.

Results: The moderate AD group differed significantly from healthy controls in terms of indefinite terms, repetitions, semantic paraphasias, and pronouns. For the latter variable, mild AD patients also differed significantly from healthy controls. However, none of the variables could differentiate mild from moderate AD.

Conclusion: Four out of five discourse variables could discriminate healthy aging from moderate AD, while only one could discriminate mild AD patients. This is therefore questioning current literature on connected-speech measures in AD and calling for further research on the variables that could better distinguish mild to moderate AD.

Keywords: Alzheimer's disease, lexical-semantic impairment, language, word-finding difficulties, connected speech, discourse

1. Introduction

1.1 Alzheimer's Disease

Alzheimer's disease (AD) is one of the most common neurodegenerative disorders worldwide (Checkoway et al., 2011; Gezondheid En Wetenschap, 2021; Mucke, 2009) and is the most frequent cause of dementia. AD is caused by the accumulation of certain proteins in the brain, namely a combination of amyloid-β (Aβ) and Tau peptide (Bali et al., 2010). This accumulation results in neural loss and deterioration of cognitive functions, such as memory, language use, and performing and understanding complex actions (Gezondheid En Wetenschap, 2021).

Memory loss is the main and most well-known symptom of AD (McKhann et al., 2011). In addition, the presence of lexical-semantic impairment is another influential symptom of AD. This impairment occurs early in the process of the disease. Amieva et al. (2008) investigated whether cognitive impairment -including language-related difficulties- are detectable years before typical AD symptoms are apparent. In particular, they examined participants' semantic verbal fluency and conceptual knowledge. Semantic verbal fluency was measured using the Isaacs Set Test (IST), in which words belonging to a particular semantic category must be generated within 15 seconds. Conceptual knowledge was measured by the Wechsler Similarities test, which required saying in what ways two things are alike. They found, as measured by the IST, that cognitive performance already exhibits the first decline 12 years before the onset of dementia. This decline is observed in semantic memory and possibly results in a weaker ability to access information in long-term memory. The authors also found a decline in conceptual formation about 10 years before dementia, as measured by the Wechsler Similarities test. These findings indicate that AD is characterized by a long and progressive prodromal phase, which can be identified by the appearance of certain symptoms. Thus, language-related tasks may be more discriminant than other cognitive tasks and contribute to the detection of AD. Moreover, several authors (e.g., Pistono et al., 2019; Salehi et al., 2017) indicated that lexical-semantic impairments (i.e., a mix of semantic memory loss and lexical access difficulties) are the first and most important complaints in early AD. Indeed, AD is characterized by a breakdown of semantic knowledge together with difficulties in accessing semantic information. As a result, patients experience word-finding difficulties that are more frequent and more persistent than the ones observed in healthy aging.

1.2 Language in Healthy Aging and Alzheimer's Disease

As stated by Ansado et al. (2013), it is essential to know how the brain copes with typical, healthy aging in order to uncover how the brain changes by, responds to, and compensates for cognitive disorders due to aging. Over the years, structural changes occur in the brain such as the loss of brain mass, which is not evenly distributed. As a consequence, some cognitive abilities like numerical abilities and spatial orientation (Schaie & Willis, 1993) decline with typical aging, while others do not. In particular, language abilities are relatively well preserved although they depend on large sets of neural networks distributed throughout the brain. One preserved language ability is vocabulary, which does not decrease with age (Park et al., 2002). These nearly unaffected language abilities indicate the occurrence of neural compensatory mechanisms during aging (Ansado et al., 2013). However, a recent study by Biran et al. (2023) showed that some language domains are more vulnerable to aging, with lexical retrieval being the most vulnerable domain. In contrast, the syntactic domain appeared to be more resistant, even though a large variance exists across participants in terms of scope of the decline.

While extensive research shows that language is relatively preserved, older people do report more word-finding difficulties, such as tip-of-the-tongue states. This tip-of-the-tongue state arises when one has a sense of knowing the word but cannot retrieve it. The transmission deficit hypothesis (MacKay & Burke, 1990) states that healthy aging weakens the connections between the semantic and phonological representation of a word. These weak connections result in a production failure because they induce too little excitation, which prevents the representation from reaching the activation threshold. Unlike AD, healthy aging does not affect semantic and lexical representations, but it affects the phonological level (i.e., the person will have a strong feeling of knowing the target word, but the phonological information about this word will be incomplete). In addition, these transmission deficits may have more detrimental effects on single-word production than on connected speech (i.e., discourse). This is not unexpected since retrieval in isolation requires a one-to-one mapping between semantic and phonological levels, whereas retrieval in context is more flexible (Kavé & Goral, 2016).

Regarding AD, there is a growing consensus that, in addition to episodic memory deficit, language impairments can also serve as a manifestation of AD or as a tool to detect AD (Filiou et al., 2020). This is in line with Szatloczki et al.'s (2015) claim that memory functions are necessary for linguistic functioning, making both functions strongly correlated in AD. Correspondingly, Kempler (2004) previously found a continuous decline in language for AD patients, implying that the severity of language impairment can potentially be linked to the different stages of AD. For each stage, specific patterns of language difficulties arise in one or more of the five language domains, namely phonetics and phonology, morphology, lexicon and semantics, syntax, and pragmatics (Szatloczki et al., 2015). In particular, lexical-semantic impairments occur very early in the time course of AD and are therefore the most obvious and important hallmark of AD (Pistono et al., 2019; Salehi et al., 2017). AD patients therefore also experience word-finding difficulties, but contrary to healthy aging, these difficulties are rather due to lexical access difficulties and/or semantic deterioration. The assessment of word-finding difficulties is usually done using standardized single-word tasks, such as naming or fluency tasks. These single-word production tasks yielded repeated evidence of AD patients scoring lower on naming and fluency tasks than healthy older adults (Slegers et al., 2018; Weiner et al., 2008). Yet, these tasks have rather low ecological validity meaning that their results are a rather bad prediction of the real-world situation (Kavé & Goral, 2016).

1.3 Connected-speech production in Alzheimer's Disease

Connected speech refers to the natural language used to communicate in daily life (Filiou et al., 2020). In recent years there has been an increase in studies that analyze, among other things, word-finding difficulties in a real-life context by focusing on discourse-level performance. These studies indicate that connected-speech analysis is the most informative part of language assessment (Boschi et al., 2017) and as such, this form of assessment contributes to the diagnosis of AD. However, the different findings on connected-speech production in AD are rather difficult to reconcile. First, there is a lack of consensus on the tasks used (e.g., personal narrative (Gayraud et al., 2011), picture description tasks (Croisile et al., 1996), storytelling task (Ash et al., 2007), etc.), which may influence participants' lexical-semantic difficulties, and participants' cognitive load more generally. Second, the mapping between discourse features and different stages of AD is still unclear. Regarding the first point, several reviews aimed at summarizing current findings about connected-speech production in AD. However, some of them (e.g., Mueller et al. (2018), Slegers et al. (2018)) only considered studies using picture description tasks in their literature review. On the contrary, Boschi et al. (2017) included studies using picture description tasks, story telling, and interviews. Although these three reviews present some consistency in their conclusions (e.g., they all indicate that participants with AD produce more repetitions, or that discourse coherence declines in AD), they also stress an influence of the task under study on language processes. In particular, Boschi and colleagues suggest that pragmatics is better assessed during interviews, whereas lexical-semantic impairments are best assessed by using a picture description task. Varlokosta et al. (2023) drew similar conclusions regarding picture description tasks, but they also indicated that discourse elicitation tasks are more constrained in assessing morphosyntactic production than structured experimental tasks. To get a more complete understanding of language preservations and impairments in AD patients, they therefore suggest using a combination of narrative discourse elicitation tasks and structured experimental tasks.

Concerning the second point, it remains difficult to comprehend discourse phenomena that characterize different stages of AD. While more and more studies are focusing on the earliest stages, in order to improve the detection of AD (e.g., Mueller et al., 2016), it is less clear how connected-speech production evolves from early to moderate AD. In addition, some studies do not mention the stages at which their patients were tested. For instance, Boschi et al. (2017) reported that some of their included studies, such as Altmann et al., (2001), Nicholas et al., (1985), Orimaye et al., (2014), did not report the results separately per AD stage. Last, the definition of the variables reflecting word-finding difficulties is not always clear, nor consistent across studies. For example, the umbrella term "word-finding difficulties" often groups together different language phenomena depending on the authors (e.g., it includes pauses and indefinite terms in Croisile et al., 1996, but pauses, repetitions and indefinite terms in Forbes-McKay & Venneri, 2005). On the contrary, substitutions (Kavé & Goral, 2017), self-corrections (Pistono et al., 2021) or revisions (de Lira et al., 2011) often refer to the same type of lexical errors. In other words, although a fair amount of research on language assessment in AD has already been done, few consistent assessment techniques, results, and definitions exist within this field of research (i.e., different types of tasks and phenomena, different stages of AD). Nonetheless, based on existing literature reviews on language in AD, it seems that some variables show more consistency, both in their definition and their results. More precisely, it seems that AD patients tend to use more high-frequency words than healthy older adults (Boschi et al., 2017; Kavé & Goral, 2017; Slegers et al., 2018), more repetitions (Boschi et al., 2017; Mueller et al., 2018; Slegers et al., 2018; Szatloczki et al., 2015), more pronouns (Almor et al., 1999; Boschi et al., 2017), more indefinite terms (Boschi et al., 2017) and more semantic paraphasias (Mueller et al., 2018; Szatloczki et al., 2015). However, there is no consensus in the literature regarding the presence or absence of these variables in early or later stages of AD, from which stage these variables can be detected, and how they evolve over the time-course of AD. As a result, the current study will focus on these five variables (i.e., lexical frequency, repetitions, pronouns, indefinite terms and paraphasias) in order to test whether they can discriminate (i) healthy aging from (mild) AD; (ii) mild AD from moderate AD. This will allow us to uncover the characteristics of word-finding difficulties in different stages of AD compared to healthy older adults by using a set of variables that have a clear definition but show inconsistency across different studies, due to different stages of patients or different tasks. As a result, the current study will delve deeper into the detection of Alzheimer's disease by using connected-speech measures, and improve the understanding of the different stages of AD based on discourse measures.

2. Method

2.1 Corpus

Data was collected from TalkBank (MacWhinney, 2019), more precisely from DementiaBank. The DementiaBank contains multiple databases for the study of language in dementia and the original data acquisition of the DementiaBank was supported by NIH grants AG005133 and AG003705 to the University of Pittsburgh. The present study uses the Pitt corpus (Becker et al., 1994) This project was part of a larger protocol administered by Alzheimer and Related Dementias Study at the University of Pittsburgh School of Medicine. In total, the study included 104 older adults as controls, 208 people with probable and possible Alzheimer's disease, and 85 people with other diagnoses of dementia. The Pitt corpus comprises a picture description task, the Cookie Theft picture (Goodglass & Kaplan, 1972). An advantage of the common use of this task is that it facilitates comparisons across the large existing literature on language features. This corpus, like any other corpus available on the TalkBank was orthographically transcribed with the Child Language Data Exchange System (CHILDES; MacWhinney, 2011), using the embedded Computerized Language Analysis (CLAN) software program and its CHAT transcription norms. This system allows a semi-automatic analysis of discourse variables. The command lines used in the CLAN program to analyze lexical-semantic variables can be found as Supplementary Material, to allow future replications of the current study.

2.2 Sample Size

Before selecting the participants, a sample size estimation has been performed to detect moderate effect sizes ($d > 0.52$, with power $(1 - \beta) = 0.80$, alpha = 0.05, number of groups = 3). This moderate effect size can be achieved with 39 participants per group. This power analysis for a one-way ANOVA was performed by using the G*Power 3.1 program (Faul et al., 2009).

2.3 Selection Criteria

In order to have three groups of participants that are comparable, we applied strict criteria for each group (see Table 1). The subject selection started from a database consisting of 510 subjects in total, first excluding all the subjects with missing data like missing audio files or transcripts. Subjects were then selected and placed into one of the three groups based on their diagnosis of either no dementia, mild dementia, or moderate dementia, based on the previously mentioned definition of these three groups. People who received other diagnoses were excluded.

Then, we used a cut-off Mini-Mental State Examination (MMSE, Folstein et al., 1975) score for each group, based on the study of Perneczky et al. (2006). These authors examined how valid the MMSE assessment is as a substitute for the Clinical Dementia Rating (CDR), by mapping the MMSE scores onto the CDR categories. From this analysis the following MMSE ranges were obtained: a score of 30 out of 30 indicating no dementia, a score between 26 and 28 was seen as questionable, 21-25 for mild dementia, 11-20 out of 30 for moderate dementia and 0-10 for severe dementia. Thus, the present study focuses on mild AD (MMSE = 21-25) and moderate AD (MMSE = 11-20), besides a group of healthy aging subjects (MMSE > 27). The severe AD stage will not be considered since it is quite difficult to analyze connected-speech production in this group.

Table 1

Selection Criteria of Participants

Criteria	Healthy Aging	Mild AD	Moderate AD
MMSE Score	> 27	21-25	11-20
Age	60-80 years	60-80 years	60-80 years
Level of Education	6-18 years	6-18 years	6-18 years

2.2 Connected-speech measures

An example of a transcript as well as an example of the output per variable is depicted in Supplementary Material. Discourse variables are summarized in Table 2.

2.2.1 Repetitions

The current study considers repetitions as the immediate repetition of words, and parts of sentences such as “the the boat”. The proportion of repetitions therefore represents the number of immediate repetitions of words and parts of sentences in relation to the total number of morphemes produced.

2.2.2 Indefinite Terms

Indefinite terms can be defined as empty words without specific meaning and nonspecific nouns that make ambiguous or general reference (Boschi et al., 2017; Lai, 2014). The indefinite terms that were of interest in the current study were: “something”, “this”, “that”, “what”, “stuff”, “whatever” and “thing”. After adding up the amount of all indefinite terms produced, a percentage, based on the total number of morphemes produced, was computed for each subject.

2.2.3 Pronouns

Pronouns are words that substitute for nouns, noun phrases, or larger nominal units, usually to avoid repetition. In the current study, this category only comprises the personal pronouns (i.e., I, you, she, he, it, we, and they) because “this” and “that” are analyzed as indefinite terms and “which” was not considered as a proper pronoun as it cannot be replaced by a different term in an utterance. As for previous variables, a percentage, based on the total number of morphemes produced, was computed for each subject.

2.2.4 Semantic Paraphasias

Semantic paraphasias involve superordinate terms, where patients use the bigger category (e.g., animal) instead of a specific item of this category (e.g., cat), or coordinate terms, where patients name the wrong item of a category that possibly has some overlapping features (e.g., say “dog” instead of “cat”) (Boschi et al., 2017). The proportion of paraphasias was also based on the total number of morphemes produced.

2.2.5 Lexical Frequency

The current study measured word frequency via lexical frequency of nouns. All nouns and their corresponding frequencies were selected from each subject’s transcript. Lexical frequency was measured using SUBLTEX$_{US}$, a database containing word frequencies based on subtitles of movies and TV series in American English (Brysbaert & New, 2009 as cited in New

& Pallier, 2022). The $SUBLTEX_{US}$ was expanded a few years ago with the Zipf scale, which is a recently developed standardized frequency measure, that is an index with the same interpretation in all collected corpora. This logarithmic scale roughly goes from 1 to 6 or 7 and its biggest advantage is that this measure does not depend on the size of the corpus. For this measure, we averaged nouns' lexical frequency for each individual.

Table 2

Terms Used in Analysis of the Five Discourse Variables

Discourse variables	Definitions or terms used for measurement
Repetitions	The immediate repetition of syllables, words, and parts of sentences
Indefinite terms	This, that, what, stuff, whatever, thing, something
Use of pronouns	I, you, she, he, it, we, they
Semantic paraphasias	Erroneous word with a narrow or distant semantic relation to the target word, substitution of another word that is semantically related to the target word
Lexical frequency	Measured with Zipf values. The calculation of Zipf values equals log10(frequency per billion words) or log10(frequency per million words) + 3. The formula to calculate the Zipf values based on the frequency counts of the total corpus is $Zipf = \log 10\left(\frac{frequency\_count+1}{201.336+.159}\right)+3.0$.

2.2.6 Inter-rater Agreement

The inter-rater agreement for the variables "indefinite terms", "use of pronouns" and "semantic paraphasias" was examined in 30% of the data which was selected randomly across the three participant groups. Ratings by an independent researcher were compared with those of the author. This resulted in a complete agreement on the use of pronouns and on semantic paraphasias, for the indefinite terms an agreement of 87.5% was obtained.

2.2.5 Statistical Analysis

A Kruskal-Wallis test was performed for each variable. Post-hoc comparisons were then based on the Tukey's HSD Test which compares the means of all conditions with each other. This test is recommended for samples of different sizes. The analyses were done with the open-source statistics program JASP (*JASP Team*, 2022).

3. Results

3.1 Population

Based on the selection criteria, the population consists of three groups: a healthy aging group (N = 40, mean age = 68.95 $\pm$ 5.048), a mild AD group with twenty-five subjects (mean age = 69.56 $\pm$ 5.43), and the moderate AD group with forty subjects (mean age = 69.33 $\pm$ 5.34). This results in a total sample size of 105 subjects (72 women, mean age = 69.24 $\pm$ 5.21). The mild AD group consists of only twenty-five participants due to missing audio files and participants failing to meet the selection criteria. For this reason, a non-parametric ANOVA was preferred.

Table 3 summarizes all descriptive results related to the variables age, gender, education level, and MMSE score across the three groups. As gender is a binomial variable, a Chi-squared test was performed showing that the groups are matched for gender ($\chi^2(2, 105) = 4.70$, p = .1). The Kruskal-Wallis test for age also revealed a non-significant p-value (H(2) = 0.22, p = .89). Since the Kruskal-Wallis test showed a statistically significant difference for level of education between the three groups (H(2) = 8.31, p = .02), post-hoc comparisons had to be computed. A non-significant p-value was found for the comparison between the healthy and mild AD group, and between the mild and moderate AD group. However, the p-value for the healthy–moderate AD group comparison was at p=0.05. For the MMSE scores, the Kruskal-Wallis test (H(2) = 91.95, p < .001) showed a main effect of group.

Table 3

Descriptive of the Population

Population	*N*	Age	Level of education	Gender (males:females)	MMSE
Healthy	40	68.95 (5.05)	13.43 (1.84)	17:23	29.15 (0.89)
Mild	25	69.56 (5.43)	12.72 (3.01)	8:17	23.40 (1.32)
Moderate	40	69.33 (5.34)	12.30 (1.51)	8:32	17.03 (2.82)

p	-	.89	.02	.1	< .001
p_{tukey} (HC – mild)	-	-	.39	-	< .001
p_{tukey} (HC – moderate)	-	-	.05	-	< .001
p_{tukey} (mild – moderate)	-	-	.71	-	< .001

Note. HC = healthy controls; MMSE = Mini-Mental State Examination; *N* = sample size. Reporting means of age, level of education, and MMSE score. Standard deviations are between brackets. *p*-value of the Tukey method is adjusted for comparing a family of 3. The first Tukey *p*-value is for healthy-mild AD comparison. The second Tukey *p*-value is for healthy-moderate AD comparison. $p < 0.05$ = significant.

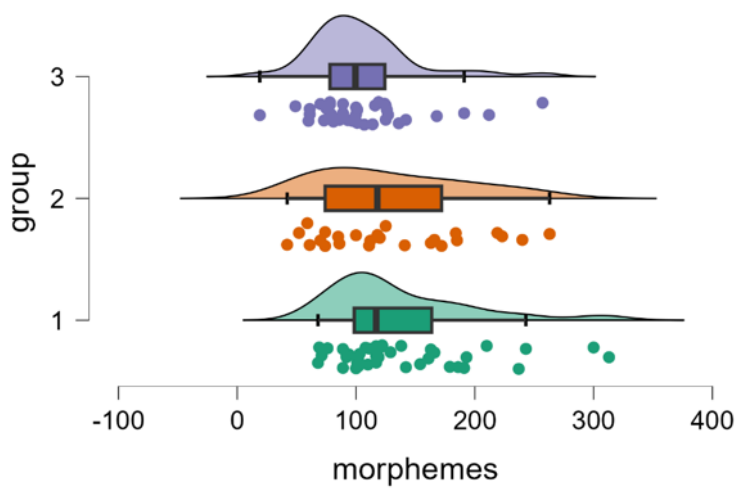
3.2 Connected-speech Measures

3.2.1 Morphemes

A statistically significant difference was found for the number of morphemes across the three groups (H(2) = 6.49, $p = 0.04$), with a mean value of 136.68 $\pm$ 59.81 (SD) for healthy controls, 129.8 $\pm$ 63.21 (SD) for the mild AD group and 104.58 $\pm$ 44.09 (SD) for the moderate AD group. The standard post-hoc comparisons using the Tukey's HSD test indicated that only the moderate AD group differed significantly from healthy controls (Table 4 and Figure 1). These results confirm that it was preferable to normalize the discourse variables based on the total number of morphemes rather than using the absolute values of the variables.

Figure 1

Raincloud Plot of Morphemes



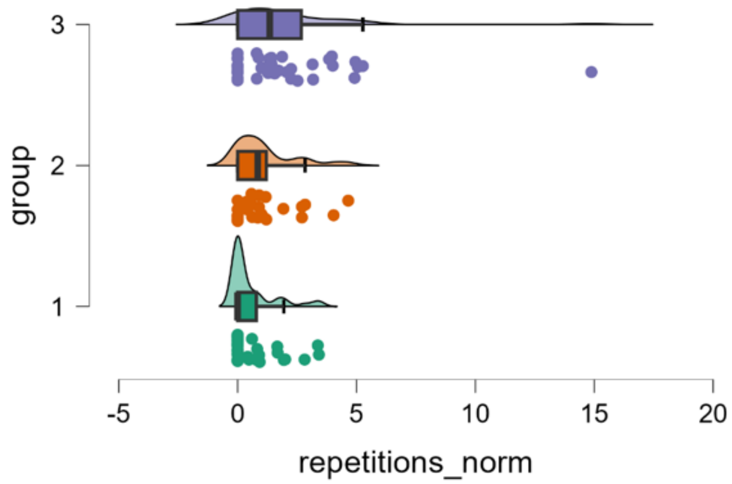
Note. Group 1 = healthy controls, 2 = mild AD, 3 = moderate AD.

3.2.2 Repetitions

The repetitions participants produced also showed a main effect of group (H(2) = 16.39, $p < .001$), with a mean value of 0.55 ± 0.96 (SD) for healthy controls, 1.14 ± 1.3 (SD) for the mild AD group and 2.03 ± 2.64 (SD) for the moderate AD group. The Tukey's HSD test showed that only the moderate AD group differed statistically significantly from healthy controls (Table 4 and Figure 2).

Figure 2

Raincloud Plot of Repetitions (normalized)



Note. Group 1 = healthy controls, 2 = mild AD, 3 = moderate AD.

3.2.3 Indefinite Terms

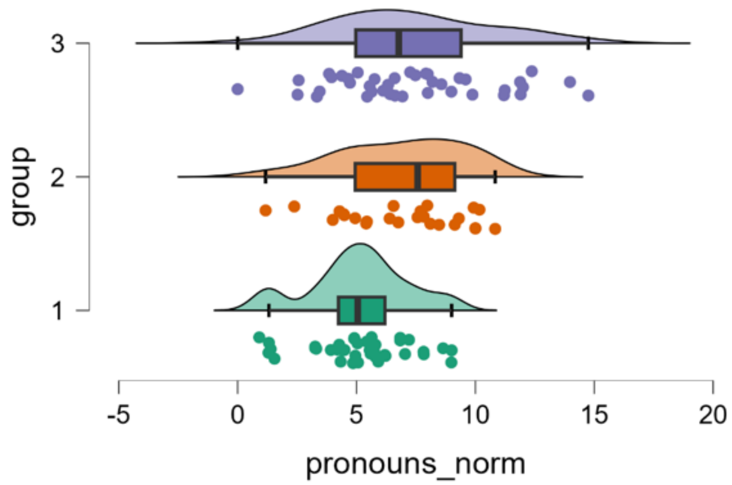
The analysis of the indefinite terms resulted in a non-significant Kruskal-Wallis test (H(2) = 5.15, $p = .08$)), with a mean value of 2 ± 1.44 (SD) for healthy controls, 2.20 ± 1.61 (SD) for the mild AD group and 3.13 ± 2.34 (SD) for the moderate AD group. As a result, no pairwise comparisons were performed.

3.2.4 Use of Pronouns

The use of pronouns resulted in a main effect of group (H(2) = 12.11, $p = .002$), , with a mean value of 5.10 ± 2.06 (SD) for healthy controls, 6.93 ± 2.59 (SD) for the mild AD group and 7.31 ± 3.36 (SD) for the moderate AD group. The Tukey's HSD test showed that both the mild AD group and the moderate AD group significantly use more pronouns than healthy controls (Table 4 and Figure 3).

Figure 3

Raincloud Plot of the Use of Pronouns (normalized)



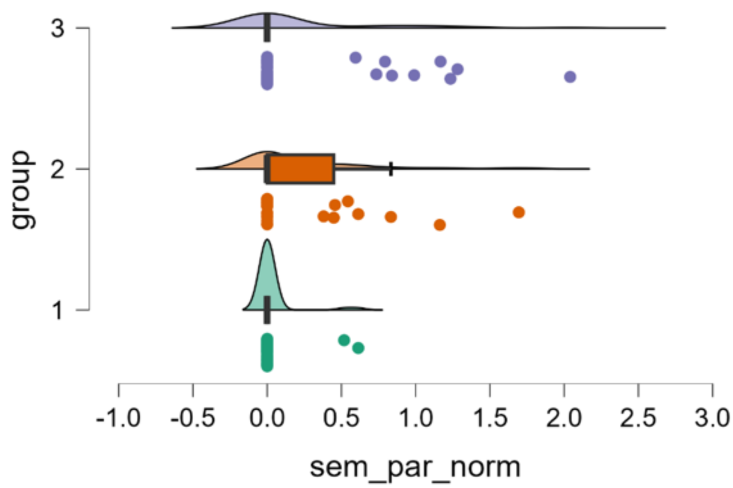
Note. Group 1 = healthy controls, 2 = mild AD, 3 = moderate AD.

3.2.5 Semantic Paraphasias

The production of semantic paraphasias resulted in a main effect of group (H(2) = 8.24, p = .02), with a mean value of 0.03 ± 0.13 (SD) for healthy controls, 0.25 ± 0.44 (SD) for the mild AD group and 0.24 ± 0.50 (SD) for the moderate AD group. Post-hoc comparisons using the Tukey's HSD test showed that the difference between the moderate AD group and the healthy control group was the only statistically significant difference (Table 4 and Figure 4).

Figure 4

Raincloud Plot of Semantic Paraphasias (normalized)



Note. Group 1 = healthy controls, 2 = mild AD, 3 = moderate AD.

3.2.6 Lexical Frequency

To analyze the effect of lexical frequencies, 2,200 words over the three groups were manually analyzed to compute the lexical frequencies. This resulted in a non-significant effect (H(2) = 2.87, p = .24), with a mean value of 6.06 $\pm$ 0.93 (SD) for healthy controls, 6 $\pm$ 1.19 (SD) for the mild AD group and 6.46 $\pm$ 1.32 (SD) for the moderate AD group, revealed no significant effect of lexical frequencies. Therefore, no pairwise comparisons were performed.

Table 6

Standard Post Hoc Tests for Discourse Variables that showed a main effect of Group

Discourse variable			Mean Difference	SE	t	p_{tukey}
Morphemes	Healthy	Mild	6.88	14.08	0.49	.88
		Moderate	32.1	12.35	2.6	**.03***
	Mild	Moderate	25.23	14.08	1.79	.18
Repetitions	Healthy	Mild	-0.59	0.47	-1.26	.42
		Moderate	-1.49	0.41	-3.61	**.001****
	Mild	Moderate	-0.89	0.47	-1.9	.14
Use of Pronouns	Healthy	Mild	-1.83	0.70	-2.63	**.03***
		Moderate	-2.21	0.61	-3.61	**.001****
	Mild	Moderate	-0.38	0.70	-0.54	.85
Semantic	Healthy	Mild	-0.22	0.10	-2.23	.07
Paraphasias		Moderate	-0.21	0.09	-2.51	**.04 ***
	Mild	Moderate	0.003	0.10	0.03	1

Note. p-value adjusted for comparing family of 3. * $p < .05$, ** $p < .01$, *** $p < .001$.

4. Discussion

The goal of the present study was to delve deeper into the variables that could track the progression of Alzheimer's disease, and to help improve the classification of the different stages of AD based on discourse features. The first goal was to examine whether AD patients can be discriminated from healthy older adults by discourse variables that reflect word-finding difficulties. The second goal was to investigate which variables are able to discriminate early from moderate AD.

Most of the discourse variables, namely repetitions, semantic paraphasias, and use of pronouns, showed a difference between the moderate AD group and healthy control group. Regarding the use of pronouns, current results align with previous reviews on connected-speech in individuals with AD. These reviews demonstrate a high proportion of pronouns in patients' speech (e.g., Kavé and Goral, 2017; Slegers et al., 2018). However, these reviews did not consider the stage of AD when interpreting their results, and it is therefore unclear whether mild AD and moderate AD both differed from healthy controls. In the current study, it is the only discourse variable showing differences between both the mild AD group and the healthy control group and the moderate AD group and the healthy control group. This suggests that the use of pronouns already changes during the mild stage. This is in line with the statement of Slegers et al. (2018) that AD patients are able to preserve relatively fluent speech when confronted with lexical-semantic difficulties by replacing an irretrievable target noun with a pronoun. Almor et al. (1999) also found an increased production of pronouns in spontaneous speech in the AD group compared to the healthy control group. According to these authors, these results can be explained by working memory problems, which prevent patients from keeping a semantic representation in mind. However, this interpretation is based on correlations between pronoun rate and working memory performance, which does not necessarily mean that there is a causal relationship between working memory and pronouns. Alternatively, and as mentioned by Slegers and colleagues, pronouns may reflect lexical access difficulties and situations in which participants replace a word they are unable to retrieve. An advantage of the current study over that of Almor et al. (1999) is that it compared mild and moderate AD stages on top of the comparison between healthy subjects and mild AD patients, as opposed to the comparison between healthy subjects and a mixed group of AD patients (MMSE range: 16-24) by the latter study. According to the results of the current study, the use of pronouns seems to be the only discourse variable that can distinguish mild AD from healthy aging. It is therefore possible that the use of pronouns is an initial strategy people use to hide their word-finding difficulties. This strategy disappears as word-finding difficulties and metacognitive difficulties increase, which then leads to an increase of other phenomena, such as semantic paraphasias. Regarding repetitions, many reviews also found an increase in repetitions in AD patients compared to healthy older adults (e.g., Boschi et al., 2017: Kavé and Goral, 2017; Slegers et al., 2018). Again, whether it was the case for each stage of AD was not clear from these studies. In the current work, we only found differences between healthy controls and moderate AD. This is coherent with Mueller et al., 2018 and Pistono et al., 2021, who did not find differences in the use of repetitions between healthy older adults and MCI participants. It therefore seems that this variable cannot contribute to the early detection of AD. In the same way, healthy controls differed from the moderate AD group only in terms of semantic paraphasias. This is in contrast with the review conducted by Mueller et al. (2018), which showed an increased production of semantic paraphasias from the earliest stage of AD. Contrary to the review conducted by Mueller et al. (2018), the current study used a database which used the standard Cookie Theft Picture. Conversely, the studies by Forbes et al. (2002) and Forbes-McKay et al. (2005), as reported in the review paper, employed both a simple and a complex version of the Cookie Theft picture description task. Both studies found an increased production of semantic paraphasias at earlier stages of AD when using the complex version of the Cookie Theft Picture. A possible explanation based on the review paper by Mueller et al. (2018) is that an increased difficulty of a picture description task may allow for more precise detection of patients' AD stages. However, one striking finding over the three groups is that some participants did not produce any semantic paraphasias, while others did. Future work is therefore required to analyze this type of inter-individual differences more thoroughly.

Finally, two variables did not show any difference between the three groups. First, no significant difference was found regarding the proportion of indefinite terms. This is contrary to Boschi et al.'s (2017) findings. Indeed, Boschi et al.'s (2017) review included two studies using picture description tasks and three studies using an interview to assess indefinite terms. Four of these studies showed that mild AD patients use significantly more indefinite terms than healthy controls, and only one (i.e., using interview) did not result in significant differences. However, indefinite terms were defined as "*total number of empty words without specific meaning; nonspecific nouns or pronouns*" (p. 8), meaning that some pronouns were also included in this category. More generally, indefinite terms sometimes include pronouns without proper noun antecedents such that the significant effect of indefinite terms in previous studies may have been driven by the effect of pronouns (e.g., Nicholas et al., 1985). Second, the mean lexical frequency of nouns was similar across the three groups. Several researchers found an increase in lexical frequency for AD patients compared to healthy controls (e.g., Slegers et al., 2018), however in the present study no significant effects have been retrieved. This finding can be due to the discourses possibly being too short, as the discourses that participants produced consisted on average of 123.69 morphemes, meaning that the sample of nouns may be too small to get a robust result. This lack of differences could also be due to an intertwining relationship with the use of pronouns, namely that the increased use of pronouns in AD patients, due to difficulties in coming up with specific names, may hide the effect of lexical frequency on nouns. It may also be due to focusing only on lexical frequencies of nouns, in contrast to other studies (e.g., Slegers et al., 2018), which included verb frequency, mean frequency of all words, and mean frequency of nouns. Indeed, the lexical frequency of adjectives and adverbs was not considered since AD patients tend to use less of these parts of speech. Moreover, since picture description tasks constrain the use of some specific nouns, it seemed interesting to see whether AD patients used target words that are similar to their control group, or more frequent ones. A final reason why this study opted to analyze the lexical frequency of nouns is because nouns offer more choices than verbs. For example, Buckingham (2011) argues that certain verbs can sometimes evoke multiple nouns. Nevertheless, longer speech samples would be required to further analyze lexical frequency in AD.

The current study contains some considerable strengths as well as some important limitations. As first strength, the current investigation compared two AD groups, namely mild AD and moderate AD. This comparison was rarely examined in previous studies, as they most often used a mixed AD group (e.g., Kavé and Dassa, 2018; Ripich et al., 2000; Visch-Brink et al., 2009). We indeed consider that a division of patients according to AD stages may contribute to improve the description of AD stages. Second, this study used a larger sample size which was often not done in previous studies (e.g., Ahmed et al., 2013: 15 AD, 15 controls; Altmann et al., 2001: 10 AD, 15 controls; Duong et al., 2003: 5 AD, 27 controls; Nicholas et al., 1985: 19 AD, 30 controls; Tomoeda & Bayles, 1993: 3 AD, 3 controls, etc.). Still, the mild AD group only had 25 participants, which limits the generalizability of the current results and increases type II errors, i.e., false-negative results. Nonetheless, we showed that many lexical-semantic variables only serve as good evidence for word-finding difficulties in moderate AD patients. Indeed, mild AD patients seem to be in the middle of a continuum between healthy aging and moderate AD, making this group not significantly different from the two others. This lack of differences may also be due to the fact that the selected discourse variables are not sensitive enough to differentiate mild AD patients from healthy participants. As a result, future work is required to uncover discourse variables that will qualitatively differ between groups (i.e., that already appear in early stages of AD). Based on the current study, it can be suggested that qualitative differences between healthy subjects and mild AD patients could be yielded by analyzing the location of the phenomena (e.g., whether repetitions are followed by high-frequency or low-frequency words), patients' reactions to difficulties (e.g., response to word finding delay, as described in Boschi et al., 2017) or by investigating whether different phenomena are mixed together (e.g., whether there is an intertwined relationship between the use of pronouns and lexical frequency of nouns). Future research could also use other discourse tasks, in particular connected-speech tasks that can elicit longer speech samples, which may lead to more differences between groups.

References

Ahmed, S., Haigh, A.-M. F., de Jager, C. A., & Garrard, P. (2013). Connected speech as a marker of disease progression in autopsy-proven Alzheimer's disease. *Brain*, *136*(12), 3727–3737. https://doi.org/10.1093/brain/awt269

Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S., & Tyler, L. K. (1999). Why do Alzheimer patients have difficulty with pronouns? working memory, semantics, and reference in comprehension and production in Alzheimer's disease. *Brain and Language, 67*(3), 202–227. https://doi.org/10.1006/brln.1999.2055

Altmann, L. J., Kempler, D., & Andersen, E. S. (2001). Speech errors in Alzheimer's disease: Reevaluating morphosyntactic preservation. *Journal of Speech, Language, and Hearing Research: JSLHR*, *44*(5), 1069–1082. https://doi.org/10.1044/1092-4388(2001/085)

Amieva, H., Le Goff, M., Millet, X., Orgogozo, J. M., Pérès, K., Barberger-Gateau, P., Jacqmin-Gadda, H., & Dartigues, J. F. (2008). Prodromal Alzheimer's disease: Successive emergence of the clinical symptoms. *Annals of Neurology*, *64*(5), 492–498. https://doi.org/10.1002/ana.21509

Ansado, J., Marsolais, Y., Methqal, I., Alary, F., & Joanette, Y. (2013). The adaptive aging brain: Evidence from the preservation of communication abilities with age. *The European Journal of Neuroscience*, *37*(12), 1887–1895. https://doi.org/10.1111/ejn.12252

Ash, S., Moore, P., Vesely, L., & Grossman, M. (2007). The decline of narrative discourse in Alzheimer's disease. *Brain and Language*, *103*(1–2), 181–182. https://doi.org/10.1016/j.bandl.2007.07.105

Bali, J., Halima, S. B., Felmy, B., Goodger, Z., Zurbriggen, S., & Rajendran, L. (2010). Cellular basis of Alzheimer's disease. *Annals of Indian Academy of Neurology*, *13*(Suppl2), S89–S93. https://doi.org/10.4103/0972-2327.74251

Becker, J. T., Boller, F., Lopez, O. L., Saxton, J., & McGonigle, K. L. (1994). The natural history of Alzheimer's disease. Description of study cohort and accuracy of diagnosis. *Archives of Neurology*, *51*(6), 585–594. https://doi.org/10.1001/archneur.1994.00540180063015

Buckingham, H. W. (2011). Semantic paraphasia. In *Springer eBooks* (pp. 2248–2250). https://doi.org/10.1007/978-0-387-79948-3_922

Biran, M., Gvion, A., & Shmuely-Samuel, S. (2022). Language in healthy ageing: A comparison across language domains. *Folia Phoniatrica Et Logopaedica*, *75*(2), 90–103. https://doi.org/10.1159/000527005

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

Boyé, M., Grabar, N., & Thi Tran, M. (2014). Contrastive conversational analysis of language production by Alzheimer's and control people. *Studies in Health Technology and Informatics*, *205*, 682–686.

Checkoway, H., Lundin, J. I., & Kelada, S. N. (2011). Neurodegenerative diseases. *IARC Scientific Publications*, *163*, 407–419.

Croisile, B., Ska, B., Brabant, M.-J., Duchene, A., Lepage, Y., Aimard, G., & Trillet, M. (1996). Comparative Study of Oral and Written Picture Description in Patients with Alzheimer ' s Disease. *Brain and Language*, *19*(53), 1–19.

Duong, A., Tardif, A., & Ska, B. (2003). Discourse about discourse: What is it and how does it progress in Alzheimer's disease? *Brain and Cognition*, *53*(2), 177–180. https://doi.org/10.1016/S0278-2626(03)00104-0

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Filiou, R.-P., Bier, N., Slegers, A., Houzé, B., Belchior, P., & Brambati, S. M. (2020). Connected speech assessment in the early detection of Alzheimer's disease and mild cognitive impairment: A scoping review. *Aphasiology*, *34*(6), 723–755. https://doi.org/10.1080/02687038.2019.1608502

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (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. https://doi.org/10.1016/0022-3956(75)90026-6

Forbes-McKay, K. E., & Venneri, a. (2005). Detecting subtle spontaneous language decline in early Alzheimer's disease with a picture description task. *Neurological Sciences : Official Journal of the*

Italian Neurological Society and of the Italian Society of Clinical Neurophysiology, *26*(4), 243–254. https://doi.org/10.1007/s10072-005-0467-9
Gayraud, F., Lee, H., & Barkat-Defradas, M. (2011). Syntactic and lexical context of pauses and hesitations in the discourse of Alzheimer patients and healthy elderly subjects. *Clinical Linguistics & Phonetics*, *25*(3), 198–209. https://doi.org/10.3109/02699206.2010.521612
Gezondheid en wetenschap. (2021). [Gezondheid en wetenschap]. *Ziekte van Alzheimer*. https://www.gezondheidenwetenschap.be/richtlijnen/de-ziekte-van-alzheimer
Goodglass, H., & Kaplan, E. (1972). Boston Diagnostic Aphasia Examination (BDAE), traduit et adapté par MAZAUX JM, ORGOGOZO JM (1982). *Échelle d'évaluation de l'aphasie*.
Kavé, G., & Dassa, A. (2017). Severity of Alzheimer's disease and language features in picture descriptions. *Aphasiology, 32,* 1–14. https://doi.org/10.1080/02687038.2017.1303441
Kavé, G., & Goral, M. (2016). Do age-related word retrieval difficulties appear (or disappear) in connected speech? *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition*, *24*, 1–20. https://doi.org/10.1080/13825585.2016.1226249
Kavé, G., & Goral, M. (2017). Word retrieval in connected speech in Alzheimer's disease: A review with meta-analyses. *Aphasiology*, *32*, 1–23. https://doi.org/10.1080/02687038.2017.1338663
Lai, Y. -h. (2014). Discourse Features of Chinese-Speaking Seniors with and without Alzheimer's Disease. *Language and Linguistics*, *15*(3), 411–434. https://doi.org/10.1177/1606822X14520665
MacKay, D. G., & Burke, D. M. (1990). Cognition and Aging: A Theory of New Learning and the Use of Old Connections. *Advances in Psychology*, *71*(C), 213–263. https://doi.org/10.1016/S0166-4115(08)60159-4
MacWhinney, B. (2019). Understanding spoken language through TalkBank. *Behavior Research Methods*, *51*(4), 1919–1927. https://doi.org/10.3758/s13428-018-1174-9
McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R., Mohs, R. C., Morris, J. C., Rossor, M. N., Scheltens, P., Carrillo, M. C., Thies, B., Weintraub, S., & Phelps, C. H. (2011). The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimer's & Dementia : The Journal of the Alzheimer's Association*, *7*(3), 263–269. https://doi.org/10.1016/j.jalz.2011.03.005
Mucke, L. (2009). Alzheimer's disease. *Nature*, *461*(7266), Article 7266. https://doi.org/10.1038/461895a
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
Mueller, K. D., Koscik, R. L., Turkstra, L. S., Riedeman, S. K., LaRue, A., Clark, L. R., Hermann, B., Sager, M. A., & Johnson, S. C. (2016). Connected Language in Late Middle-Aged Adults at Risk for Alzheimer's Disease. *Journal of Alzheimer's Disease : JAD*, *54*(4), 1539–1550. https://doi.org/10.3233/JAD-160252
Nicholas, M., Obler, L. K., Albert, M. L., & Helm-Estabrooks, N. (1985). Empty speech in Alzheimer's disease and fluent aphasia. *Journal of Speech and Hearing Research*, *28*(3), 405–410. https://doi.org/10.1044/jshr.2803.405
Orimaye, S. O., Wong, J. S.-M., & Golden, K. J. (2014). Learning Predictive Linguistic Features for Alzheimer's Disease and related Dementias using Verbal Utterances. *Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, 78–87. https://doi.org/10.3115/v1/W14-3210
Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320.
Perneczky, R., Wagenpfeil, S., Komossa, K., Grimmer, T., Diehl, J., & Kurz, A. (2006). Mapping scores onto stages: Mini-mental state examination and clinical dementia rating. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry*, *14*(2), 139–144. https://doi.org/10.1097/01.JGP.0000192478.82189.a8
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

Pistono, A., Senoussi, M., Guerrier, L., Rafiq, M., Giméno, M., Péran, P., ... & Pariente, J. (2021). Language network connectivity increases in early Alzheimer's disease. *Journal of Alzheimer's disease*, *82*(1), 447-460.

Ripich, D. N., Carpenter, B. D., & Ziol, E. W. (2000). Conversational cohesion patterns in men and women with Alzheimer's disease: A longitudinal study. *International Journal of Language & Communication Disorders*, *35*(1), 49–64. https://doi.org/10.1080/136828200247241

Schaie, K. W., & Willis, S. L. (1993). Age difference patterns of psychometric intelligence in adulthood: Generalizability within and across ability domains. *Psychology and Aging*, *8*(1), 44-55. https://doi.org/10.1037//0882-7974.8.1.44

Salehi, M., Reisi, M., & Ghasisin, L. (2017). Lexical Retrieval or Semantic Knowledge? Which One Causes Naming Errors in Patients with Mild and Moderate Alzheimer's Disease? *Dementia and Geriatric Cognitive Disorders EXTRA*, *7*(3), 419–429. https://doi.org/10.1159/000484137

Slegers, A., Filiou, R.-P., Montembeault, M., & Brambati, S. M. (2018). Connected Speech Features from Picture Description in Alzheimer's Disease: A Systematic Review. *Journal of Alzheimer's Disease: JAD*, *65*(2), 519–542. https://doi.org/10.3233/JAD-170881

Swan, K., Hopper, M., Wenke, R., Jackson, C., Till, T., & Conway, E. (2018). Speech-Language Pathologist Interventions for Communication in Moderate-Severe Dementia: A Systematic Review. *American Journal of Speech-Language Pathology*, *27*(2), 836–852. https://doi.org/10.1044/2017_AJSLP-17-0043

Szatloczki, G., Hoffmann, I., Vincze, V., Kalman, J., & Pakaski, M. (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. https://doi.org/10.3389/fnagi.2015.00195

Tomoeda, C. K., & Bayles, K. A. (1993). Longitudinal effects of Alzheimer disease on discourse production. *Alzheimer Disease and Associated Disorders*, *7*, 223–236.

Varlokosta, S., Fragkopoulou, K., Arfani, D., & Manouilidou, C. (2023). Methodologies for assessing morphosyntactic ability in people with Alzheimer's Disease. *International Journal of Language & Communication Disorders*. https://doi.org/10.1111/1460-6984.12862

Visch-Brink, E. G., Van Rhee-Temme, W., Rietveld, T., Krulder, J. W. M., Van Harskamp, F., & Van Der Cammen, T. J. M. (2009). Improvement of spontaneous speech in early stage Alzheimer's disease with rivastigmine. *JNHA - The Journal of Nutrition, Health and Aging*, *13*(1), 34–38. https://doi.org/10.1007/s12603-009-0007-z

Weiner, M. F., Neubecker, K. E., Bret, M. E., & Hynan, L. S. (2008). Language in Alzheimer's disease. *The Journal of Clinical Psychiatry*, *69*(8), 1223–1227. https://doi.org/10.4088/jcp.v69n0804

Word classes and phrase classes. (2023). *In Cambridge Grammar*. https://dictionary.cambridge.org/grammar/british-grammar/word-classes-and-phrase-classes