Age-Related Decline in Syntactic Complexity Reflects Local, Not Global Planning Processes

Albrecht M.A.1, Zandhuis, J.1, Wetterlin, A.1, Khoe, Y.H.2, Fernandes, E. G. 3, Segaert, K.*2, Wheeldon, L.*1

1 Institute of Foreign Languages and Translation, University of Agder, Norway

2 School of Psychology and Centre for Human Brain Health, University of Birmingham, UK

3 School of Psychology, University of Minho, Portugal

ORCID (Mikael): https://orcid.org/0000-0002-9822-9510

ORCID (Eunice): https://orcid.org/0000-0002-1448-3256

*KS and LW are shared senior authors

Corresponding author: Mikael André Albrecht (mikael.a.albrecht@uia.no)

Acknowledgements

KS, LW and data collection for this project were funded by (NFR, FRIPRO 300030). We are indebted to the wider FAB research team for their contributions throughout the project, to Dr. Evelien Heyselaar for drawing the scenes, to Emma Trapper Pedersen and Eliza Mercuri for their work on the transcripts, and to Ben Elswood for his help in data preparation and pre-processing. YHK, KS and LW were funded by The Leverhulme Trust (RPG-2021-360). We are very grateful to Dr. Galit Agmon for sharing the scripts to analyse syntactic complexity.

Declaration of Competing Interests

The authors report no competing interests.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Age-Related Decline in Syntactic Complexity Reflects Local, Not Global Planning Processes

Abstract

Healthy ageing is associated with reduced syntactic complexity in spoken language, yet the mechanisms underlying this decline remain unclear. In particular, it is unknown whether age-related simplification reflects impairments in global syntactic planning scope or arises from more local planning difficulties. In the present study, we elicited natural speech from younger (18 – 35 years) and older (65 – 81 years) monolingual English speakers who described pictured scenes. We then combined transcript-based measures of lexical and syntactic complexity with fine-grained speech analyses of articulation and pausing behaviour to address this question. Our findings show that older adults produced longer and more lexically diverse language overall, but their sentences were syntactically less complex, containing fewer clauses and noun phrases than those of younger adults. Crucially, while longer planning pauses (≥250 ms) strongly predicted syntactic complexity in both age groups, these effects did not differ by age, suggesting that global syntactic planning mechanisms remain largely intact in healthy ageing. In contrast, the shorter medial pauses (100–249 ms) occurring between planning pauses, predicted both the number of clauses and number of words in a sentence more strongly in younger than older adults, suggesting that local planning was more successful for younger than older adults. Together, these findings indicate that age-related reductions in syntactic complexity are best explained as downstream consequences of less efficient local planning rather than impaired syntactic knowledge or global structure-building processes.

Data availability statement

The data for the present study were collected as part of a larger study, The **REDACTED** Project: preregistration: **[REDACTED URL]**; materials and data: https://osf.io/5kg9q/overview?view_only=51cd82da54f54a46814f2beb60f4bf6c

Keywords

Language Production, Ageing, Syntactic Complexity, Lexical Complexity, Pauses

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Introduction

In this paper, we focus on language production abilities in healthy older adults. Language production is a core cognitive function that enables communication of complex ideas, a skill that is essential for forming and sustaining relationships, accessing support, and participating meaningfully in society. Language production is known to decline with ageing even when individuals are ageing healthily. Previous research has demonstrated reductions in syntactic complexity in older adults (e.g., Kemper et al., 2003; Marini et al., 2005) but has not yet elucidated the mechanisms through which these difficulties arise. Current and influential theories of language production agree that sentence production is an incremental process in which word retrieval and the building of grammatical structure occurs rapidly over time such that syntactic complexity and fluency are interrelated (e.g., Bock & Levelt, 1994; Dell & Chang, 2014; Levelt, 1989; Momma & Ferreira, 2021). We advance theory by demonstrating how these processes are differentially interrelated across the lifespan. Through doing this, we elucidate which aspect of language planning is the main culprit underpinning age-related syntactic complexity decline: a breakdown in the processes of planning complex syntactic structures, or a consequence of problems in local planning such as the retrieval and integration of lexical items.

Theories of spoken language production

Language production is a complex process requiring both language specific representations and processes (see Levelt, 1989; Levelt et al., 1999), general cognitive mechanisms such as working memory (e.g., Marlin & Slevc, 2014; Swets et al., 2014), and attention (e.g., Jongman et al., 2015). Current theories agree that sentences are produced in successive planning stages during which conceptual, grammatical, and phonological representations are generated prior to articulation. There is also broad agreement that these processes run in an incremental fashion such that initial parts of an utterance can be articulated while the remaining parts are being planned (e.g., Kempen & Hoenkamp, 1987; Levelt et al., 1999). The linearisation of conceptual information occurs during grammatical encoding, which involves the generation of syntactic structure and the retrieval of lexical content. Theories diverge in the mechanisms that they claim drive the size of the planning increments during grammatical encoding. Linear models propose that lexical retrieval precedes and determines structure building (e.g., Bock & Levelt, 1994; Griffin, 2001; Pickering & Branigan, 1998). In contrast, hierarchical models propose that syntactic structure building precedes and guides lexical access (e.g., Chang et al., 2006; Dell et al., 2008). Evidence from experimentally elicited fluent sentences demonstrates clear effects of syntactic structure on sentence production latencies (e.g., Allum & Wheeldon, 2007; Smith & Wheeldon, 1999) and on both the extent and order of word retrieval (e.g., Allum & Wheeldon, 2009; Konopka & Meyer,

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

2014; Momma & Ferreira, 2021; Schriefers et al., 1990). These findings, on young adults, are consistent with hierarchical planning of global syntax and the incremental output of phrase structure and lexical content.

Because most of the evidence for hierarchical planning comes from highly constrained sentence production experiments, the data do not fully represent the complexity of daily language production processes which operate in an environment of fluctuating cognitive demands. In addition to linguistic factors, there is evidence that cognitive load can affect planning scope during speech production. For example, increased complexity at the conceptual level can decrease planning scope (e.g., Konopka & Meyer, 2014; Kuchinsky et al., 2011; van de Velde et al., 2014). Conversely, lexical planning scope can increase when syntactic processing is facilitated (e.g., Konopka, 2012; Konopka & Kuchinsky, 2015). There is also evidence that individual differences in cognitive abilities can influence planning: better working memory is associated with increased planning scope (e.g., Martin et al., 2014; Swets et al., 2007, 2014, 2021) and individual differences in some components of attention influence lexical retrieval in picture-naming (e.g., Piai & Roelofs, 2013; Shao et al., 2012) as well as phrase production latencies (e.g., Jongman, Meyer, et al., 2015; Jongman, Roelofs, et al., 2015). Some theories therefore do not assume a default form of incrementality but allow for flexibility such that planning may be more linear or more hierarchical under different planning conditions (e.g., Konopka et al., 2018). Finally, it is also possible that planning scope may, to some extent, be under a speaker's strategic control such that smaller or less complex structures may be prioritised in order to maintain fluency in the face of cognitive challenges (e.g., Ferreira & Swets, 2002). A complex set of factors therefore combine to characterise the relationship between fluency and complexity during spoken language production. We turn now to the available evidence relating to the impact of ageing on these processes.

Syntactic and lexical production in healthy ageing

Key components of language production are adversely affected in healthy ageing. Older adults have been found to produce sentences which are less syntactically complex than those of younger adults (e.g., Kemper et al., 2003; Marini et al., 2005; Sung et al., 2024) - though note that others have not observed any age-related differences in syntactic complexity (Caplan & Waters, 1999; Horton et al., 2010; Payne & Stine-Morrow, 2014). Syntactic complexity itself has been defined by various measures such as number of words in a sentence, verb voice, and number of clauses. Kemper et al. (2003) observed that younger and older adults produced similar sentences when using intransitive or monotransitive verbs, while sentences using verbs that take further complements were less complex in older adults. Agmon et al. (2024) confirmed an adverse effect of ageing on syntactic complexity using natural continuous speech elicited as descriptions of a pictured scene (the cookie theft picture task, Goodglass & Kaplan, 1983). They used a multidimensional metric of syntactic complexity, which automatically extracted syntactic

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

structures as its features. We will be using these same methods in the present study. At present, no evidence is available on which aspects of speech production planning drive age-related declines in syntactic complexity. However, there is a wealth of evidence that the speed and accuracy of lexical retrieval decrease with age (e.g., Feyereisen, 1997; Kavé et al., 2010; Wei et al., 2024). A commonly cited example of word-finding failures are tip-of-the-tongue states in which speakers experience a temporary, partial failure of word retrieval, more common in older age-groups (e.g., Burke et al., 1991). Schmitter-Edgecombe et al. (2000) showed that older participants have significantly more word finding difficulties than younger participants when producing descriptive narrations of pictures. In a review of studies examining word retrieval in connected speech in healthy ageing, Kavé and Goral (2017) showed that word-retrieval difficulties in older adults can lead to an overall reduction in productivity, reduce lexical diversity, increase speech errors and decrease verbal fluency.

Language planning, speech fluency and healthy ageing

The research reviewed above demonstrates that healthy ageing can be associated with decreases in spoken language complexity. However, the available data do not allow us to determine the source of this reduction; namely, whether it is driven by age-related changes in more global syntactic planning processes or in more local planning involving the integration of words and structure. As reviewed above, current theories of language production agree that sentence production is an incremental process in which word retrieval and the building of grammatical structure occurs rapidly over time, and that syntactic complexity, and the speed and fluency of speech production are interrelated. In order to address questions about the effects of ageing on planning processes, it is therefore necessary to relate measures of linguistic complexity to measures of spoken fluency across the age-span.

Analyses of speech in natural dialogues have shown that speech fluency can decrease during healthy ageing. The speed of articulation can be reduced with age (e.g., Beier et al., 2023; Horton et al., 2010) while dysfluencies; such as the number of filled and silent pauses, repetitions and repairs, have been shown to increase (e.g., Horton et al., 2010; Samani et al., 2017, but see Beier et al., 2023). Speech disfluencies have been attributed to increases in planning difficulty, associated with aspects of task complexity (e.g., Bortfeld et al., 2001). However, few studies have related measures of disfluency to measures of language complexity. For example, Horton et al., (2010) examined a recorded conversational speech corpus (Switchboard I, Godfrey et al., 1992) and found that increasing age (speaker range was 20-67 years) was associated with reduced speech rates (measured as number of words by total speaking time) and increased rates of filled pauses. Interestingly, the rate of different fillers (number of “uh” and “um” per 100 words) correlated differently with age: the production of “uh” increased with age while the production of “um” decease with age. In an analysis of a subset of the corpus, Shriberg (1996) found that “uhs” occurred more often within utterances and may signal word finding difficulties. In contrast,

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

"ums" occurred more frequently in utterance initial position and may signal message level planning. However, no relationship between the frequency of these fillers and syntactic complexity was reported.

Wang and Wang (2024) directly investigated the relationship between measures of speech fluency and complexity in spontaneous spoken narratives. Speakers were grouped as early-old (60-69), middle-old (70-79), and later-old (80+). Fluency was operationalised as articulation rate (syllables per second) as well as dysfluency rates (proportion of dysfluent syllables including fillers, repetitions, repairs and restarts). Lexical complexity was measured using a corrected type-token ratio which gives the proportion of unique words (see Carrol, 1964, for the correction used). Syntactic complexity was measured as mean dependency distance (MDD, see Liu et al., 2017) or the linear distance between two syntactically related words. They reported a complex relationship between these fluency and complexity measures across age groups. For example, articulation rate and syntactic complexity interacted in the middle-old group such that faster rates of articulation were associated with sentences that were more complex while in the later-old group the pattern was more competitive with faster articulation rates being associated with reduced syntactic complexity. Faster rates of articulation were also associated with reduced lexical diversity in the middle-old group, whereas both the early-old and the later-old group showed the opposite pattern. However, the overall articulation rates of the later-old group were lower than those of the early-old group indicating that participants' ability to coordinate articulation rate and lexical diversity may decline with age.

These findings suggest that the ability to coordinate linguistic subsystems declines with age and highlight the importance of examining multiple dimensions of language production simultaneously to gain a better picture of how age affects language production. However, the measures in Wang and Wang (2024) did not include the timing information necessary to address the question of planning processes and language complexity in ageing. A key limitation of the research to-date for understanding planning processes is that measures of fluency have been restricted to articulation rate, number of hesitations and filled pauses. However, spontaneous speech is characterised by periods of fluent output interspersed by silent as well as filled pauses both of which show a relationship to speech planning. Filled pauses have been shown to signal different kinds of planning difficulties (e.g., Clark & Fox Tree, 2002). Silent pauses of more than approximately 250 ms tend to occur at prosodic and/or grammatical boundaries (e.g., Butterworth, 1980; Goldman-Eisler, 1968) and the duration of such pauses is positively related to the length and complexity of fluent speech that follows it (e.g., Ferreira, 1991; Grosjean et al., 1979; Henderson et al., 1966; Krivokapić, 2007). These periods have therefore been characterised as planning pauses for the upcoming fluent speech and may provide critical information about language production processes in ageing. Therefore, examining the relationship between

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

pausing, complexity and fluency is essential for understanding the effect of healthy ageing on language production processes.

The Present Study

In the present study, we combined, for the first time, transcript-based analyses of syntactic complexity with detailed speech analysis data of pausing and articulation patterns. We employ a scene description task that allows for free-speech production while constraining the narrative and word choice. Our chosen task addresses limitations of previous research. Experimental studies are limited in their focus on the controlled production of single sentences which may be too simple a task for age effects in planning scope to be observed (e.g., Hardy et al., 2020). On the other hand, the analysis of more complex spontaneous speech has tended to use corpora of unconstrained conversational or narrative speech that may allow older adults to employ their accrued linguistic knowledge to mask evidence of any decline in language planning processing: it is well established that vocabulary size increases with age (Kavé, 2022; Verhaeghen, 2003) and that narrative quality can markedly improve (e.g., Hilviu et al., 2025).

We developed 4 picture scenes based on the classic cookie theft picture task (Goodglass & Kaplan, 1983; **[REDACTED URL]**, see Figure 1 below) that our participants were asked to describe. Speech elicited in this way is regularly used in the study of aphasia (e.g., Andreetta & Marini, 2015; Marini et al., 2026) but less frequently in healthy ageing research. This semi-constrained methodology ensured that both older and younger adults performed the same task, allowing for the direct and controlled comparison of naturalistic speech production across age groups, while minimising the opportunity for older adults to use different strategies that could mask problems. Picture descriptions were collected from a large sample of monolingual English-speaking older adults (between 60-81 years of age) and compared to that elicited from younger adults (18-35 years of age).

Transcripts were parsed into sentences and analysed for several measures of lexical and syntactic complexity (following Agmon et al. 2024). In addition, the spoken language was forced aligned with the transcripts to elicit detailed measurements of speech-rate and pausing patterns. This data set provided a detailed picture of age-related changes in language production. Critically it also allowed us to investigate the effect of healthy ageing on the relationship between pausing and the syntactic complexity of sentences. Our aim was to determine whether syntactic simplification in healthy ageing reflects impairments in more global or more local planning processes.

As mentioned above, in studies of connected speech, pauses greater than 250 ms have been defined as planning pauses and their duration is related to the complexity of the following stretch of fluent speech. Here we directly test for effects of age on the relationship between such planning pauses and syntactic complexity. Research has indicated a complex relationship

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

between filled and silent pauses (Gósy, 2023). There is also significant inter-speaker variation in the ratio of silent to filled pausing (e.g., Braun et al., 2023; van Donzel & Koopmans-van Beinum, 1996). We therefore do not distinguish between filled and silent pauses in our duration measures. We also analyse shorter medial pauses (100-249ms occurring between planning pauses). Pauses of less than 100ms are usually related to articulatory functions (e.g., plosive closures, Hieke et al., 1983; Stathopoulos & Weismer, 1985). However, pauses between 100-250ms have been shown to relate to linguistic rather than simply articulatory functions (e.g., Bennett, 1986; Campione & Véronis, 2002; Hieke et al., 1983), and filled pause durations have been reported from 100ms upwards (e.g., Clark & Fox Tree, 2002; Gósy, 2023). We hypothesised that planning pauses are associated with more global sentence planning processes whereas shorter medial pauses reflect more local planning. Our aim was to investigate the effects of healthy ageing on syntactic complexity and pausing patterns and to determine the relationship between them.

Methods

Participants

The data for this publication were collected as part of a larger preregistered study, the **[REDACTED]** Project (**[REDACTED URL]**). The data in the present sub-study are from a sub-group of the larger participant cohort, with unique outcome measures to other publications related to the larger project **[CITATIONS REDACTED FOR REVIEW]** and address *a priori* questions (as per preregistration hypothesis 3). The study was approved by the STEM Ethical Review Committee at the **[REDACTED FOR REVIEW]** Before enrolment, participants received an information sheet and provided informed consent. The sample is described in Table 1 below. All were British-English monolinguals with no speech or language disorders. All older participants scored > 23 on the MoCA (Carson et al., 2018; Nasreddine et al., 2005). Age- and education descriptive statistics are given in Table 1.

Table 1: Participant group descriptives.

	Younger (*n* = 37; 17 Female)			Older (*n* = 89; 41 Female)		
	M	SD	Range	M	SD	Range
Age	23.35	5.38	[18, 35]	65.59	4.84	[60, 81]
Education[1]	2.92	1.14	[2, 6]	2.36	1.16	[1, 5]

[1] 1 = Compulsory education, 2 = Further education (e.g., A levels, college), 3 = Undergraduate degree (e.g., BA, BSc), 4 = Postgraduate degree (e.g., MA, MSc), 5 = Higher degree (e.g., PhD, DSc).

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Task

Participants were shown one of four scenes (assignment was random) depicted in Figure 1 below. The scenes were inspired by the cookie theft picture first created by Goodglass and Kaplan (1983). The scenes were matched for number of animate parties depicted as well as number of inanimate objects and actions. In the first screen, a written text instructed participant to speak for a few minutes, describing the picture that they would be presented with, and to press the spacebar when they had finished. Then the picture was presented. No timer was shown, in order to encourage relaxed speech production. As such, no timeout was enforced. Participants were seated in a sound-attenuated booth. Speech was recorded from the onset of the image with a sampling rate of 48.0 kHz.

Figure 1: Stimuli scenes. Participants were shown one of the four possible pictured scenes.

Transcription and Segmentation

The recordings were manually transcribed by English native speakers and then verified by a fluent L2 English-speaking researcher. Transcribers were instructed to include dysfluencies such as false starts and word repetitions as well as filled pauses. We next sought to compare these manual

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

transcriptions with ones generated automatically using OpenAI's Whisper (version 20250625, Radford et al., 2022). To do so, we calculated Levenshtein distances for both transcripts at both the character- and word level and both are reported in Table 2 below.

Levenshtein distances (LD) are typically used to measure the minimal number of changes it takes to change from one word to another. Valid changes are substitutions, removals, and additions. For example, "cat" to "hat" would have an LD of 1. This can be normalised by dividing the LD by the maximal possible number of changes (i.e., the length of the longest word in a pair) which makes the LD range from 0 (complete overlap) to 1 (no overlap). In the case of "cat" to "hat" this normalised distance would be 0.33. To more thoroughly compare the two transcriptions, we expanded this methodology to also calculate LDs at the word level (i.e., the minimum number of whole words that need to be replaced to make the two transcriptions identical).

Manual inspection of the discrepancies between the two methods showed that they largely stem from inconsistencies in using contracted and non-contracted verb-forms (e.g., it's vs. it is) as well as Whisper not recording filled pauses and dysfluencies as reliably as the manual transcribers. Nevertheless, the two methods align well at both the character and word levels as indicated by the comparisons in Table 2.

Table 2: Normalised Levenshtein distance transcription comparisons between manual and automatic transcriptions.

	Character LD	Word LD
M	.042	.072
SD	.032	.046
Range	[.000, .239]	[.000, .273]

The manual transcriptions were then segmented into sentences using wtpsplit (Frohmann et al., 2024) via the 3-layer sat-sm model. These automatically segmented sentences were reviewed by a researcher who corrected instances of erroneous single-word syntactic units. To assess the performance of this automatic method on a speech sample which has not been cleaned of dysfluencies we also segmented the transcripts into syntactic units by hand. This was done by fluent L2 English-speaking MA students, and the segmentations were reviewed by a researcher. The manual segmentation was done in accordance with the following protocol:

1) A complete syntactic unit should contain at least an independent clause with possible exceptions for list-like structures.
2) Subordinate clauses were counted as being part of the larger syntactic unit. This also applied to non-finite subclauses and relative clauses.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

3) Parentheticals were treated as a separate syntactic unit if they contained at least one lexical verb.
4) Coordinated independent clauses were treated as separate syntactic units but coordinated subordinated clauses were counted as being part of the same syntactic unit pursuant to criterion 2.
5) Filled pauses between two syntactic units were tagged as belonging to the second syntactic unit.
6) All recordings were listened to during segmentation and speaker intonation was used to determine syntactic boundaries. In cases where intonation conflicted with steps 1-5, intonation was given priority.

Because automatic segmentation tools rely on text without intonation, we treated the manual transcriptions as the standard against which the automatically segmented sentences were compared. To do so, we investigated how often boundaries placed by the two methods appeared in the same location[2]. This was done using three metrics. First, precision was defined as the proportion of boundaries placed by the automatic method which match a manually placed boundary. As such[3]:

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$$

Second, recall was defined as the proportion of manually placed boundaries which were found by the automatic method. As such:

$$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

The two metrics produced similar values with Precision = .68 and Recall = .69. This gave an F1 score of .68[4]. Examining the overall boundary proportions, the two methods produced similar results with manual = .088 and automatic = .090. Lastly, to further quantify the level of agreement between the two methods, we calculated Cohen's k for the two methods. This gave an estimate of *k* = .65 which constitutes substantial agreement. We proceeded to analyse the automatically segmented parses here as they are more realistic for applications in analyses of large volumes of free speech. However, we also present results from the manual parses for sensitivity purposes in the appendices.

[2] A researcher manually reviewed the automatic segmentations in cases of single-word sentences. Without this correction, the metrics were worse. Precision = .55, Recall = .70, F1 = .61, and k = .57

[3] TP = True Positive. Boundaries which were placed in the same place by both methods.
FP = False Positive. Boundaries that were placed but the automatic method but not by the manual one.
FN = False Negative. Boundaries present in the manual method but absent in the automatic method.

[4] F1 scores represent the harmonic mean between precision and recall thus capturing both metrics.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Sentence Complexity

For analyses of sentence complexity we began with the segmented sentences described above. We used the BLLIP parser (Charniak et al., 2021) to parse the individual sentences. Relative clauses were clauses which functioned as postmodifiers in noun phrases and could be both finite or non-finite (e.g., the man **who is doing the dishes**; the man **doing the dishes**) while centre embeddings were clauses or phrases which were embedded within a larger unit of the same type (e.g., "the jar **that the cookies were in** has fallen over"). From those parses we calculated six measures of complex syntactic structures that are cognitively demanding and might, therefore, be less frequently used by healthy older adults than by younger adults (see Agmon et al., 2024):

- Number of clauses
- Number of relative clauses
- Number of noun phrases
- Size of noun phrases
- Number of centre embeddings
- Size of centre embeddings

Forced Alignment

We next forced aligned the transcripts with the source speech files to get reliable timing data. This was done using the Montréal Forced Aligner (henceforth MFA, McAuliffe et al., 2017). The input transcripts were, again, not cleaned for dysfluencies and the MFA produced text grids which showed the onset and offset of each word. Silent pauses were defined as any gap where articulation was not detected, and which was not defined as being within a word by the MFA alignment. The temporal resolution of these alignments was 5ms. These alignments were analysed to examine the effects of healthy ageing on language production across multiple levels of processing with an emphasis on syntactic planning and complexity, speech fluency and articulation rate, speech dysfluencies including both filled and silent pauses, and lexical diversity. However, to assess the reliability of these alignments, two trained phoneticians manually segmented a subset of the files (n = 12) using Praat (Boersma et al., 2026, version 6.4.23). To assess the agreement between human raters and the MFA we calculated the interclass correlation coefficient (ICC) both for lexical words and pauses combined and for pauses separately (ICC = .976, ICC = .980 respectively)[5]. This indicates a large degree of structural

[5] It could be that larger human-to-human agreement affects the overall comparison. However, we also conducted pairwise ICC comparisons between each human coder and the MFA. The range of ICC strengths did not noticeably differ (Range = [.967, .972])

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

reliability between raters. We next calculated absolute mean differences between the human raters and between the MFA as reported in Table 3 below.

Table 3: Summary of absolute mean differences between human coders and the MFA.

	Lexical words and Pauses *Absolute Mean Difference (ms)*	Only Pauses *Absolute Mean Difference (ms)*
Human Coders	34.23	31.35
Coder A x MFA	56.36	72.84
Coder B x MFA	53.38	73.63

Lastly, we sought to quantify whether this disagreement between the MFA and human raters reflected systematic bias or noise. To do so, we calculated the mean human-measured duration for each lexical word and pause and subtracted this from mfa durations. A strongly positive or negative mean and small standard deviation would indicate systematic bias. Instead, we found the opposite (M = 3.13, SD = 412.91). Taken together, the MFA boundary placements were susceptible to noise with no clear directional bias. However, the large ICC demonstrates that the internal ranking of relative durations was consistent between human raters and the MFA.

General Data Handling and Exclusions

In this and subsequent sections, all sentence-level groupings were done based on the automatically parsed sentences. Prior to analysis, one younger participant was removed due to lack of task engagement (i.e., listing nouns absent of syntactic structure). One younger participant was also removed for having produced only one sentence. This left 126 participants (37 younger, 89 older). The raw data comprised 27388 words, filled pauses, and dysfluencies. Although no time-out was enforced, we chose to trim recordings which lasted more than 180 seconds. In these cases, recordings were trimmed once the current syntactic unit was finished which led to the exclusion of 848 words and dysfluencies. This trimming applied to 10 participants (9 older, 1 younger) and the range of recording durations exceeding 180 seconds was [184, 260]. A further 1424 data points were removed for being part of the first sentence that each participant produced since these were assumed to be potentially confounding due to strongly reflecting processes such as initial scene perception and task adjustment. Sentences not containing at least one lexical word were removed next (n = 39 data points) which resulted in an overall data loss of 3359 rows or 12% of the data. Next, we addressed values unlikely to represent actual processing and instead being caused by misalignment or hallucinations. We first removed the durations of any words that exceeded 5000 ms in length (n = 9). Lastly, we identified silent pauses and added these to the data. Prior to analyses, we also removed articulatory pauses that were shorter than 100ms in length (n = 4047).

Design

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Our data comprised dependent variables related to the speech produced by our participants. These were (i) lexical and morphemic complexity measures at the transcript, sentence, and word levels, (ii) syntactic and lexical complexity measures of syntactic and fluent speech units (including the six syntactic complexity measures listed above), and (iii) frequency and number of pauses (both planning and medial pauses). For the pause measures, we divide analyses into planning pauses (i.e., above 250 ms in length) and medial pauses (i.e., shorter than 250 ms in length) while fluent speech units will be defined as any unit of speech uninterrupted by a planning pause. We return to these definitions in the relevant results sections below.

Our primary independent variable of interest was the age contrast (younger vs. older) and we also included a covariate of ordinal education level. For some models, pause durations were used as independent variables to examine whether pausing behaviour interacted with age in predicting lexical or syntactic complexity.

General Analysis Workflow

The analyses reported in the results section were all conducted using mixed effects models. While the distributions of dependent variables varied, the general model selection workflow remained consistent. In the case of count-data (whole integers, typically with a right-hand skewed distribution), an initial Poisson GLMM was fit and diagnosed. In cases of overdispersion, a negative binomial model was fit instead. Continuous or semi-continuous data was modelled using LMMs (Baayen et al., 2008) with assumptions inspected visually. In cases of assumption violations (i.e., residual normality and homoscedasticity of variance), box-cox analyses were conducted to select an appropriate transformation (Box & Cox, 1964). In cases of extreme positive skew where transformations did not sufficiently rectify assumption violations, continuous and positively skewed DVs were instead analysed using log-linked Gamma GLMMs (Lo & Andrews, 2015) or lognormal GLMMs. The full analysis documentation is available on OSF. Binary dependent variables were always analysed using binomial GLMMs (e.g., Jaeger, 2008) and all analyses were conducted with the least amount of data aggregation possible to minimise power loss.

Models included random intercepts for subject and scene which were inspected visually for normality. In cases of models failing to converge, random effects were removed one at a time and re-fit with scene being removed over participant. Age-group was always entered as a factor of two levels with a positive contrast assigned to the older group while the ordinal and continuous IVs were scaled and mean-centered prior to fitting the model. This applied to the ordinal covariate of education level as well as pause duration measures. Education level was imputed by the mean of the group in cases of missing data (n = 8 older, n = 0 younger). [6]

[6] Example model code: DV ~ age + education + (1 | scene) + (1 | participant)

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

In the case of significant 2-way interactions involving a non-categorical variable we calculated Estimated Marginal Trends (EMTs) to identify the driving components of the interaction. EMTs provide the model-based estimates of how a continuous independent variable affects the dependent variable while controlling for the other variables in the model. As such, EMTs give the predicted slope of a continuous variable. This allows for further analyses such as pairwise comparisons of continuous effects (trends) between groups in the case of an interaction (see the emmeans package in R; Lenth & Piaskowski, 2026)

Results

We first report age-group effects for measures of lexical and morphemic complexity at the transcript, sentence, and word-levels. We then turn to analyses of the syntactic complexity of sentences produced by both age-groups (see the list on p. 11). Next, we report analyses of speech fluency as well as measures of lexical and syntactic complexity within participants' fluent speech units. Lastly, we analyse the relationship between speech pauses and complexity at both the lexical and syntactic levels.

Lexical and Morphemic Complexity Measures

Our first analyses relate to general differences in lexical and morphemic complexity between the older and younger age-groups. Descriptive statistics of these measures are given in Table 4 while model outputs are given in Table 5 below.

Exclusions followed the procedure outlined in the section above with no additional removals necessary for these analyses. The numbers divide into three levels: The entire transcript, within each sentence, and where appropriate within each individual word. At the transcript level, we report the recording length in seconds and the number of sentences produced by participants[7]. The remaining two transcript measures distinguish between the number of lexical words (i.e., all content word tokens produced) and number of unique lexemes (i.e., word-forms without inflectional morphemes). We calculated the length of each sentence in terms of number of lexical words and number of morphemes. For all lexical word types, we also calculated morphemic length (aggregating inflectional and derivational morphemes), syllabic length, word frequency on the Zipf scale (retrieved from the SubtLEX-UK corpus, van Heuven et al., 2014), as well as phonological and orthographic neighbourhood densities retrieved from the CELEX corpus (Baayen et al., 1995).

[7] Number sentenes in each transcript as identified by our segmentation with wtpsplit.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Table 4: Descriptive statistics of transcript, sentence, and word-level variables.

Measure	Level	Younger			Older		
		M	*SD*	Range	*M*	*SD*	Range
Recording length (seconds)	Transcript	66.75	33.04	[22.12, 151.24]	91.98	39.83	[23.19 177.20]
Number of Sentences	Transcript	14.89	7.73	[4, 38]	21.33	8.17	[7, 48]
Number of Unique Lexemes	Transcript	81.70	27.59	[35, 148]	99.39	29.15	[32, 168]
Number of Lexical Words	Transcript	82.86	40.76	[23, 188]	110.88	46.75	[20, 244]
Number of Lexical Words	Sentence	5.62	3.64	[1, 23]	5.23	3.44	[1, 25]
Morphemic Length	Sentence	13.29	8.75	[2, 53]	11.79	7.78	[1, 54]
Lexical Word Zipf Frequency	Word	5.24[8]	1.04	[1.17, 7.18]	5.24	1.05	[1.30, 7.18]
Lex. Word Orth. NBD	Word	9.56	8.44	[0, 37]	9.36	8.50	[0, 40]
Lexical Word Phonemic NBD	Word	19.20	15.30	[0, 60]	19.00	15.50	[0, 60]
Lex. Word Morph. Length	Word	1.25	0.45	[1, 3]	1.26	0.45	[1, 3]
Lexical Word Syllabic Length	Word	1.42	0.59	[1, 5]	1.45	0.64	[1, 5]

For the models in Table 5 below we did not analyse recording length in seconds because transcript length is captured by the remaining transcript-level measures. We do not report analyses of the number of sentences per transcript for the same reason. For neighbourhood density at the word level, we only report analyses of phonemic neighbourhood density as this was a spoken task, and orthographic and phonemic neighbourhood density showed a strong correlation ($\tau = .65$). We observed no significant effects of age for any of these lexical measures (all $ps > .05$).

As can be seen, older participants were generally more verbose producing longer transcripts measured in number of lexical words ($p < .001$). Older participants also produced significantly more unique lexemes showing that that the effect of increased length is not due to a repetition

[8] Zipf scale

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

effect (p < .001). Averaged across sentences, younger participants produced significantly more morphemes than older adults (p = .005) but there was no overall difference in number of lexical words (p = .20).

Table 5: Summary of regression models of speech content.

Measure	Level	Predictor	Est.	SE	*z/t*	*p*	sig.
Number of Unique Lexemes	Transcript	Age-group	0.12	0.03	3.60	< .001	*
		Education	0.06	0.03	1.98	.048	*
Number of Lexical Words	Transcript	Age-group	0.06	0.02	3.33	< .001	*
		Education	-0.01	0.01	-0.51	.61	
Number of Lexical Words	Sentence	Age-group	-0.03	0.02	-1.29	.20	
		Education	0.03	0.02	1.57	.12	
Morphemic Length[9]	Sentence	Age-group	-0.06	0.02	-2.82	.005	*
		Education	0.03	0.02	1.43	.15	
Lexical Word Zipf Frequency	Word	Age-group	-0.00	0.02	-0.10	.92	
		Education	0.02	0.02	1.01	.32	
Lexical Word Phonemic NBD	Word	Age-group	-0.01	0.02	-0.71	.48	
		Education	-0.00	0.01	-0.24	.81	
Lex. Word Morphemic Length	Word	Age-group	0.00	0.01	0.10	.92	
		Education	-0.01	0.01	-1.25	.21	
Lexical Word Syllabic Length	Word	Age-group	0.01	0.01	1.27	.21	
		Education	-0.00	0.01	-0.21	.84	

Syntactic Complexity of Sentences

Data Handling

For syntactic complexity analyses, we removed sentences that contained 0 noun phrases (n = 15, .01 of sentences). This was done to exclude sentences which do not capture information related to processing and that may be due to parsing or segmentation errors. We next identified one sentence in the older age-group which was tagged as containing 10 relative clauses (far outside the remaining 0-4 range) and we removed this sentence as abnormal. Lastly, any sentence whose complexity measured in the number of NPs fell in the 99th percentile of the data was considered an outlier and removed (n = 20, .01 of the data).

Analyses

[9] The morphemic length model was run using a truncated negative binomial model due to persistent overdispersion in the initial negative binomial model. The zero-inflation model specified zero inflation of intercepts (ziformula = ~1).

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Descriptive statistics of syntactic complexity are given in Table 6 below with all measures showing larger values for the younger group.

Table 6: Syntactic complexity descriptives by age-group.

Measure	Younger			Older		
	M	*SD*	Range	*M*	*SD*	Range
Number of Clauses	1.79	1.48	[0, 11]	1.54	1.28	[0, 12]
Number of Rel. Clauses	0.23	0.49	[0, 2]	0.19	0.47	[0, 4]
Number of NPs	4.39	2.86	[1, 14]	3.87	2.51	[1, 14]
NP Length	2.86	1.35	[1, 8.62)	2.76	1.24	[1, 12.40]
Centre Embed Presence	.23	.43	NA	.19	.39	NA
Number of Centre Embeds	0.32	0.67	[0, 4]	0.23	0.50	[0, 3]
Length of Centre Embeds	2.94	2.06	[1, 16]	2.64	1.61	[0, 14]

The model results are given in Table 7 below. Due to the overall scarcity of centre embeddings (*n* = 482, .20 of sentences) we analysed them in two stages. First, centre embeddings were re-coded as a binary variable indicating the presence or absence of a centre embedding in each sentence. Second, we analysed only sentences containing at least one centre embedding to examine whether the number of centre embeddings used differed by age group.

Table 7: Sentence syntactic complexity regression model summaries.

Measure	Predictor	Est.	*SE*	*z*	*p*	sig.
Number of Clauses	Age-group	-0.06	0.03	-2.49	.01	*
	Education	0.03	0.02	1.38	.17	
Number of Rel. Clauses	Age-group	-0.11	0.07	-1.51	.13	
	Education	0.04	0.07	0.60	.55	
Number of NPs	Age-group	-0.06	0.02	-3.22	.001	*
	Education	0.03	0.02	1.74	.08	
NP Length	Age-group	-0.01	0.01	-1.08	.28	
	Education	0.00	0.01	0.42	.67	
Centre Embed Presence[10]	Age-group	-0.13	0.08	-1.58	.11	
	Education	0.08	0.07	1.11	.27	
Number of Centre Embeds[11]	Age-group	-0.05	0.05	-1.13	.26	
	Education	0.05	0.04	1.15	.25	

[10] Due to the overall low number of sentences that contained centre embed, the analysis of centre embed number was run in two stages. First we re-coded the centre embed variabled into a binary variable and this analysis thus represents the proportion of sentences that contain one or more centre embeds.

[11] In this second step we only inspected sentences that contained one or more centre embeds to look at whether the age-groups differ in number of centre embeds used when they are present in a sentence.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Length of Centre Embeds	Age-group	-0.04	0.02	-1.89	.06
	Education	0.00	0.02	0.21	.83

We found significant age-differences in syntactic complexity with number of clauses (p = .01) and number of NPs (p = .001) per sentence. However, while all measures showed patterns in favour of the younger cohort, most comparisons did not reach significance (ps > .05) although there was a trend towards significance in the length of centre embeds (p = .06). For sensitivity purposes we performed the same analyses of sentence syntactic complexity on the manually parsed sentences (see Appendix B). This did not affect significance but the trend for length of centre embeds was no longer present (p = .75)

Pause Durations and Number

Data Handling

We examined the patterns of fluent articulation and pausing within each age group. For these and subsequent timing analyses, we excluded 11 participants whose forced alignments showed clear misalignment upon manual inspection. We first turn to pause durations. Any pauses of 100 ms or less were removed from analyses as their short duration meant they likely represented articulatory processes. Due to the heavy positive skew observed in the pause duration data, a method for outlier exclusion based on standard deviations is problematic. Instead, based on the distribution given above we implemented a cut-off point at 3000 ms with pauses above this threshold being considered extreme and therefore removed from analyses of fluency. Meanwhile, the density peaks were at 175 ms for the older group and 202 ms for the younger group meaning that our 250 ms cut-off point for planning pauses sits at the right side of the modal region of the pause distributions. Short medial pauses thus comprised the densest regions of the distributions while the majority of pauses overall were planning pauses. Remaining pauses are shown in the density plot in Figure 2 below.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

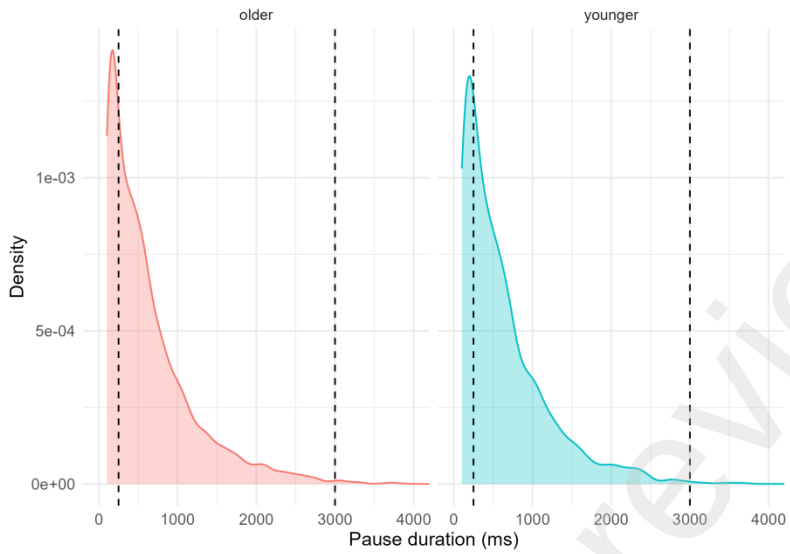


Figure 2: Density plots of pause distributions by age-groups. The dashed lines represent the 250 ms thresholds for medial/planning pauses and the upper 3000 ms cut-off point for planning pauses.

By the above criteria, 70% of pauses were defined as planning pauses. These pauses included both silent and filled pauses. Where silent and filled pauses followed one another with no content in between their durations were merged. The remaining 30% of the pauses (between 100-250 ms) were defined as medial pauses (older = 0.30, younger = 0.30).

We next created fluent speech units (henceforth FSUs) which were defined as any unit of speech which was uninterrupted by a planning pause. We next implemented the following inclusion criteria to omit FSUs which do not contain lexical or syntactic information as well as FSUs whose durations are so short that they may indicate possible misalignment.

1) Must contain 1 or more words (*n* = 506)
2) must contain 1 or more syllables (*n* = 361)
3) Must have a speech rate measured in syllables per second less than 10 (*n* = 77)
4) Must have a duration above 300ms (*n* = 129)
5) Cannot be part of the first FSU due to parts possibly being removed in the previous first-sentence removal (*n* = 118).

Due to overlap between the criteria with some FSUs being excluded on multiple grounds, the sum of eliminated FSUs was 625 or .17 of the data.

Fluent Speech Unit Pause Analyses

We calculated four measures of pause durations within FSUs. For planning pauses we calculated the mean planning pause duration per FSU. For medial planning pauses, we calculated the mean number of medial pauses per FSU as well as two measures of duration. Mean pause duration

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

reflects the length of medial pauses while total medial pause length represents the mean sum duration of medial pause durations within FSUs – FSUs that contain no medial pauses are entered as 0 for this latter measure. These descriptives are given in Table 8 below. As can be seen the age differences were small and regression analyses yielded no significant effects (all *p*s > .05).

Table 8: Descriptive statistics of pausing measures.

Measure	Younger			Older		
	M	*SD*	Range	*M*	*SD*	Range
FSU Planning pause length	795	501	[252, 2997]	810	524	[252, 2817]
FSU mean total medial pause length	74	119	[0, 696]	73	123	[0, 972]
Mean medial pause duration	159	42	[102, 249]	156	41	[102, 249]
FSU n medial pauses	0.46	0.72	[0, 4]	0.47	0.78	[0, 6]

We then examined the effects of age-group on the complexity of FSUs produced. For each FSU we calculated length in terms or number of syllables, morphemes and lexical words. We also calculated the syntactic complexity of FSU using the measures reported above. The descriptives for FSUs by age group are shown in Table 9 below.

Table 9: Descriptive FSU statistics.

Measure	Younger			Older		
	M	*SD*	Range	*M*	*SD*	Range
Syllabic Length	8.95	5.91	[2, 41]	9.09	5.98	[2, 41]
Number of Morphemes	8.58	5.74	[2, 41]	8.56	5.61	[2, 36]
Number of Lexical Words	3.65	2.43	[0, 18]	3.81	2.52	[0, 16]
Number of Clauses	1.20	1.13	[0, 6]	1.17	1.07	[0, 7]
Number of Rel. Clauses	0.14	0.39	[0, 3]	0.14	0.41	[0, 5]
Number of NPs	2.91	2.26	[0, 16]	2.82	2.19	[0, 15]
NP Length	2.39	1.24	[0, 10.4]	2.39	1.23	[0, 8.2}
Centre Embed Presence	0.13	0.34	NA	0.12	0.33	NA
Number of Centre Embeds	0.16	0.43	[0, 3]	0.12	0.36	[0, 3]
Length of Centre Embeds	2.20	1.25	[1, 11]	2.26	1.27	[1, 13]

As with the pausing data, the differences between age groups were small and showed no significant age-effects (all *p*s > .05, see Appendix C).

Finally, we examined speech rates (i.e. syllables per second within FSUs without the inclusion of planning pauses but including medial pauses). The regression model showed a significant age-effect (Est. = -0.05, *SE* = 0.02, *t* = -2.84, *p* = .005) with younger participants having significantly

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

faster rates of articulation than older adults (Younger: *M* = 4.54, *SD* = 1.31, Range = [1.10, 9.17]; Older: *M* = 4.17, *SD* = 1.12, Range = [1.05, 9.21]).

Pausing and Complexity

We next sought to examine the relationships between pauses and the complexity of speech units produced. We report these analyses both for lexical-complexity (i.e., the number of lexical words) and the syntactic complexity measures reported above. For syntactic complexity, we chose to limit analyses to sentences as the shorter FSUs are unlikely to capture larger syntactic structures and we limit the syntactic complexity measures to number of clauses and number of NPs within sentences as these were the only two to show significant age differences.

Fluent Speech Unit Lexical Complexity

Beginning with FSUs, 31 planning pauses falling in the 99th percentile were removed as outliers and a further 44 FSUs were removed which contained no lexical words. We then conducted regression models to see whether planning pause durations as well as the number- and sum-duration of medial pauses predicted the number of lexical words within FSUs and, if so, whether this interacted with age. The results of these models are reported in Table 10 below.

Table 10: Regression summaries of FSU lexical complexity models.

Measure	Predictor	Est.	*SE*	*z*	*p*	Sig.
FSU n lexical words	Planning Pause Duration	0.05	0.01	3.77	< .001	*
	Age	0.02	0.02	1.01	.31	
	Planning Pause Duration:Age	0.02	0.01	1.25	.21	
	Education	-0.00	0.02	-0.10	.92	
FSU n lexical words	Medial Pause Sum Duration	0.21	0.01	19.28	< .001	*
	Age	0.02	0.02	1.01	.31	
	Medial Pause Sum Duration:Age	-0.01	0.01	-0.96	.34	
	Education	-0.01	0.02	-0.49	.62	
FSU n lexical words	Number of Medial Pauses	0.21	0.01	19.29	< .001	*
	Age	0.02	0.02	0.96	.34	
	Number of Medial Pauses:Age	-0.01	0.01	-1.23	.22	
	Education	-0.01	0.02	-0.49	.63	

All three measures of pausing showed significant main effects on the number of lexical words in FSUs (all *p*s < .05) however none of the interactions with age reached significance (*p*s > .05). Longer FSU-initial planning pauses thus predict more lexical words within the upcoming FSU. For medial pauses, both sum duration and number predict more lexical words within the FSU.

Sentence Lexical Complexity

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

For sentences, we examined the sum durations of all planning pauses or all medial pauses that occurred in a sentence regardless of position. If a planning pause occurred between two sentences, then it was counted as the part of the second sentence. The descriptive statistics in given in Table 11 below are thus given as the mean sum duration within sentences.

Table 11: Descriptive statistics of pauses in sentences.

Measure	Younger			Older		
	M	*SD*	Range	*M*	*SD*	Range
Sentence Sum Planning Pause Duration	1080	1071	[0, 8283]	974	958	[0, 6594]
Sentence Number of Planning Pauses	1.31	1.14	[0, 8]	1.14	0.98	[0, 7]
Sentence Sum Medial Pause Duration	102	148	[0, 828]	86	130	[0, 729]
Sentence Number of Medial Pauses	0.66	0.92	[0, 5]	0.56	0.83	[0, 5]

As can be seen younger adults' pausing within sentences is both longer and more frequent for both pause types. However, regression analyses did not show significant contrasts between age groups for any of the four pause measures (all *p*s > .05), although we did observe a trend in favour of the younger cohort for the number of planning pauses per sentence (p = .07). However, the critical question to be addressed is if pausing predicts complexity. We therefore conducted further regression analyses as reported in Table 12 below. In these models we analysed planning pauses and medial pauses separately and limited analyses to the two measures of syntactic complexity which showed significant age differences in our earlier analyses (i.e., number of clauses and number of noun phrases).

Table 12: Regression summaries of lexical complexity in sentences.

Measure	Predictor	Est.	*SE*	*z*	*p*	Sig.
Number of Lexical words	Planning Pause Duration	0.23	0.01	18.01	< .001	*
	Age	-0.01	0.02	-0.51	.61	
	Planning Pause Duration:Age	0.01	0.01	0.66	.51	
	Education	0.03	0.02	1.73	.08	
Number of Lexical words	Number of Planning Pauses	0.03	0.01	27.44	< .001	*
	Age	-0.01	0.02	-0.33	.74	
	Number of Planning Pauses:Age	0.01	0.01	1.28	.20	
	Education	0.01	0.02	0.79	.43	
Number of Lexical words	Medial Pause Duration	0.21	0.01	15.31	< .001	*
	Age	-0.01	0.02	-0.39	.69	
	Number of Medial Pauses:Age	-0.03	0.01	-1.85	.07	

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

	Education	0.01	0.02	0.62	.54	
Number of Lexical words	Number of Medial Pauses	0.21	0.01	15.42	< .001	*
	Age	-0.01	0.02	-0.27	.78	
	Number of Medial Pauses:Age	-0.03	0.01	-2.46	.01	*
	Education	0.01	0.02	0.73	.47	

For all four pause measures the main effects were significant in predicting the number of lexical words with longer and more frequent pauses predicting sentences that are more lexically rich, mirroring the FSU findings (all *p*s < .05). We also observed a significant 2-way interaction between number of medial pauses and age (p = .01). EMTs showed a significant trend in both age groups (*p*s < .05) and the contrast between the two age groups was also significant (z = -2.46, p = .01) with a significantly steeper slope for the younger age-group as illustrated in Figure 3 below.

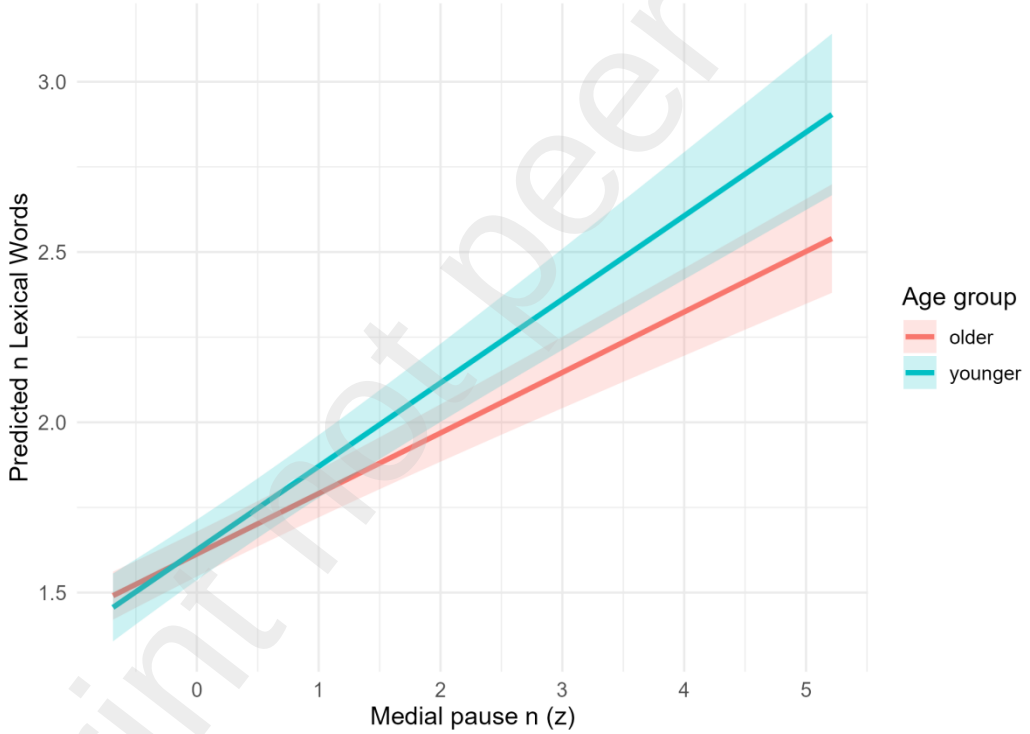


Figure 3: Significant interaction between age and number of medial pauses on the predicted number of lexical words in sentences.

In summary, we observe reliable main effects of pausing on both FSU and sentence-level lexical complexity with medial pauses showing a significant interaction with age for sentence-level lexical complexity. We next turn to sentence-level syntactic complexity.

Sentence-Level Syntactic Complexity

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

The analyses are separated into analyses of planning pauses and analyses of medial pauses. Beginning with planning pauses, the results are reported in Table 13 below.

Table 13: Syntactic complexity model summaries of regressions examining planning pauses and ageing

Measure	Predictor	Est.	*SE*	*z*	*p*	Sig.
Number of Clauses	Planning Pause Duration	0.16	0.02	9.11	< .001	*
	Age	-0.04	0.03	-1.57	.12	
	Planning Pause Duration:Age	-0.03	0.02	-1.79	.07	
	Education	0.04	0.02	1.66	.10	
Number of NPs	Planning Pause Duration	0.23	0.02	15.14	< .001	*
	Age	-0.06	0.02	-2.55	.01	*
	Planning Pause Duration:Age	0.01	0.02	0.85	.40	
	Education	0.04	0.02	1.89	.06	
Number of Clauses	Number of Planning Pauses	0.22	0.02	12.55	< .001	*
	Age	-0.03	0.03	-1.23	.22	
	Number of Planning Pauses:Age	-0.02	0.02	-1.21	.23	
	Education	0.02	0.02	1.08	.28	
Number of NPs	Number of Planning Pauses	0.29	0.01	21.70	< .001	*
	Age	-0.05	0.02	-2.43	.02	*
	Number of Planning Pauses:Age	0.02	0.01	1.68	.09	
	Education	0.02	0.02	0.95	.34	

Planning pause measures significantly predicted syntactic complexity in all four models (*p*s < .05) consistent with increased planning being associated with sentences that are more syntactically complex. None of the 2-way interactions between planning pauses and age reached significance (all *p*s > .05) but we did observe a trend for number of NPs predicted by number of planning pauses (*p* = .09). We next examined the effects of medial pauses on the same two measures of syntactic complexity as summarised in Table 14 below.

Table 14 Syntactic complexity model summaries of regressions examining medial pauses and ageing.

Measure	Predictor	Est.	*SE*	*z*	*p*	Sig.
Number of Clauses	Medial Pause Duration	0.18	0.02	10.10	< .001	*
	Age	-0.04	0.03	-1.61	.11	
	Medial Pause Duration:Age	-0.04	0.02	-2.00	.045	*
	Education	0.02	0.02	1.01	.31	
Number of NPs	Medial Pause Duration	0.20	0.02	13.07	< .001	*
	Age	-0.06	0.02	-2.61	.01	*
	Medial Pause Duration:Age	-0.01	0.02	-0.60	.55	

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

	Education	0.02	0.02	0.75	.45	
Number of Clauses	Number of Medial Pauses	0.19	0.02	9.94	< .001	*
	Age	-0.04	0.03	-1.49	.14	
	Number of Medial Pauses:Age	-0.05	0.02	-2.39	.02	*
	Education	0.02	0.02	1.09	.28	
Number of NPs	Number of Medial Pauses	0.21	0.02	13.35	< .001	*
	Age	-0.06	0.02	-2.52	.01	*
	Number of Medial Pauses:Age	-0.02	0.02	-1.19	.24	
	Education	0.02	0.02	0.82	.42	

For medial pauses, medial pause duration and number interacted with age in predicting the number of clauses of a sentence (p = .045, p = .02). For medial pause durations, longer durations were associated with more clauses within a sentence. However, this slope was steeper for younger participants than for older ones, a contrast which EMTs showed was significant (z = -2.00, p = .045). The model predictions for this interaction on simulated data is given in Figure 4 below. This pattern of results is consistent with younger participant gaining more of a planning benefit from medial pauses compared to their older counterparts.

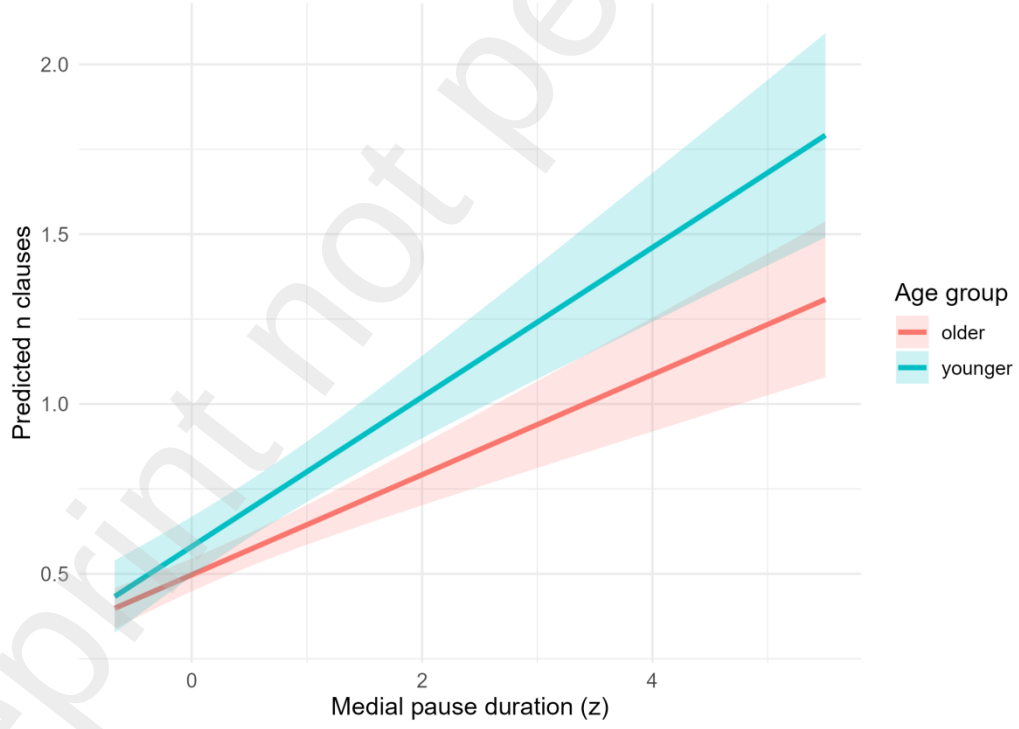


Figure 4: Significant interaction between sum medial pause duration and age-group in predicting the number of clauses in sentences.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Number of medial pauses displayed a similar pattern with more medial pauses being associated with sentences containing more clauses. EMTs again showed a significant difference with a steeper slope for the younger group (z = -2.39 p = .02). This interaction is shown in Figure 5 below.

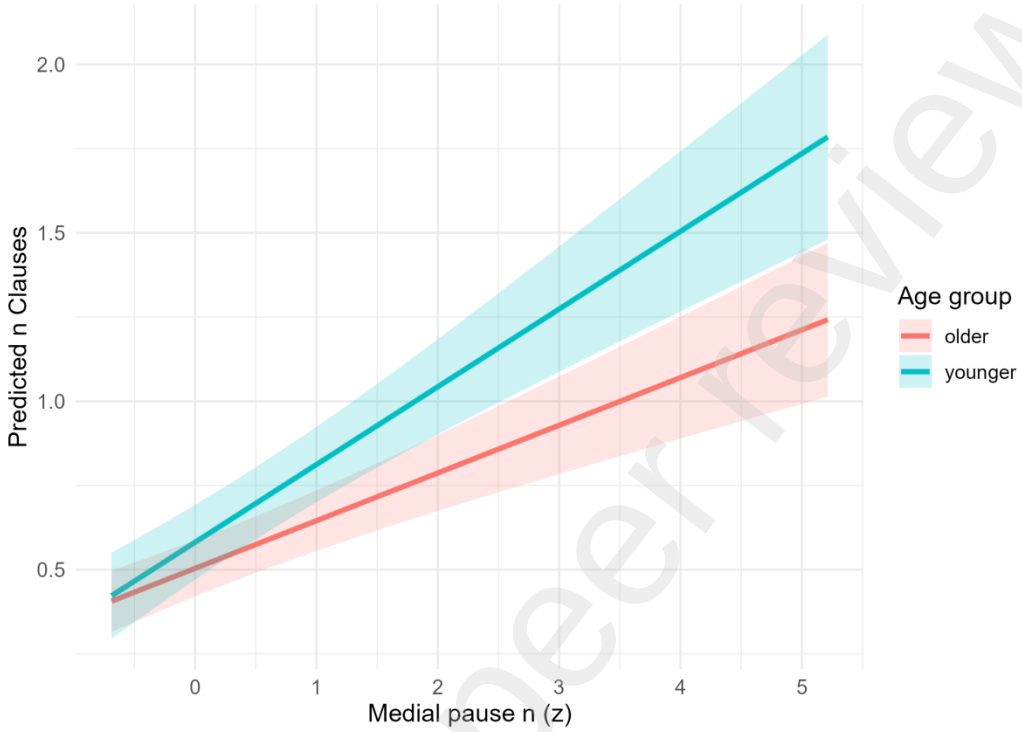


Figure 5: Significant interaction between number of medial pauses and age-group in predicting the number of clauses in sentences.

Discussion

Healthy ageing reduces the syntactic complexity of spoken language, but the source of that reduction has been unclear: does it reflect a narrowing of global syntactic planning, or a difficulty in the more local processes that assemble words and structure as a sentence unfolds? Combining transcript-based measures of complexity with fine-grained pause timing in semi-constrained connected speech, we found a consistent pattern. Longer planning pauses predicted syntactic complexity equally in younger and older adults, indicating that global, structure-level planning is largely preserved in healthy ageing. In contrast, shorter medial pauses predicted both clausal complexity and lexical content more strongly in younger than older adults. We argue this points to a local, rather than global, planning limitation as the locus of age-related simplification. Below we summarize our findings in greater detail and consider what this implies for language production theories and the ageing literature more broadly.

Our key findings are summarised in Table 15 below. Firstly, our findings replicate previous results showing that older adults have increased lexical knowledge but decreased syntactic complexity. In their overall speech output, our older adults produced more unique lexical items than the

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

younger adults, but within sentences, the number of unique lexical items was the same. The words younger adults produced were significantly more morphologically complex than those of older adults. Consistent with this, the sentences produced by younger adults were also more syntactically complex: they comprised significantly more clauses and noun phrases than those of the older adults. Secondly, we examined the language planning process driving the reduction in syntactic complexity observed for older adults. To do this we analysed the pausing patterns of our speakers and used them as predictors of our syntactic complexity measures. The medial pauses, but not planning pauses, predicted syntactic complexity differentially for young vs. older adults. Medial pausing was a significantly stronger predictor of syntactic complexity for younger compared to older adults. Medial pauses also predicted the number of lexical items in a sentence, again more strongly so for young than older adults. We conclude that the planning that occurred during medial pauses was more successful for younger than older adults. These findings are the first to demonstrate that age-related reductions in syntactic complexity may be predominantly driven by problems with local planning issue such as lexical integration rather than higher level global syntactic planning processes.

Table 15: Summary of our key findings and their implications for language and ageing.

Measure	Young Adults	Older adults	Key interpretation
Unique lexemes (overall description)	Lower	Higher	Older adults have broader lexical knowledge
Number of lexical items (overall description)	Lower	Higher	Older adults were more loquacious in their descriptions
Number of lexical items (Sentence)	Same	Same	While more loquacious overall, older adults did not produce sentences that were more lexically rich.
Number of morphemes (per sentence)	Higher	Lower	Younger adults make more frequent use of morphologically complex words.
Number of clauses per sentence and Number of NPs per sentence	Higher	Lower	Younger adults produce sentences that are more syntactically complex.
Syntactic complexity measures in FSUs	Same	Same	Syntactic complexity effects are located at the sentence-level with fluent units being too short to capture such complexity differences.
FSU and Sentence pause durations and frequencies	Same	Same	Differences in relationships between pause measures and speech content are not attributable to pause length differences between age-groups.
Pauses → FSU lexical items	Significant predictor	Significant predictor	The lexical length of fluent speech units is not affected by healthy ageing although pauses are a significant overall predictor.
Planning pauses → Sentence syntax	Significant predictor	Significant predictor	Higher-level more global planning processes do not differ by age.
Planning Pauses → Sentence lexical items	Significant predictor	Significant predictor	Lexical preparation processes do not proceed differently between age groups during longer, more global planning windows.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Medial pauses→ Sentence number of clauses	Stronger predictor	Weaker predictor	Differences in syntactic complexity between age-groups are driven by local, not global planning processes.
Medial pauses → Sentence lexical items	Stronger predictor	Weaker predictor	Lexical preparation is a central part of the local planning represented by medial pauses.

Our findings are consistent with previous demonstrations of a larger vocabulary being accrued with age (e.g., Kavé, 2022). This previous knowledge motivated our choice of speech elicitation task. We elicited natural speech production data using a semi-constrained scene description task. This task was designed to restrict older adults' ability to benefit from their increased vocabulary and accrued language skills (e.g., Hilviu et al., 2025). Benefits that could be called upon to mask impairments in wholly unconstrained narrative or conversational speech. Though we confirmed that in their general speech output older adults used more unique words, within sentences, no significant age effects at the lexical level were observed. This suggests that the task successfully limited lexical variation and that older adults' increased vocabulary did not translate to a lexically driven benefit in sentence length or complexity, allowing for a more direct examination of syntactic complexity between age groups.

Our results also confirmed previous findings showing a reduction in syntactic complexity with age. Using Agmon et al's (2024) algorithm, we replicated their effect of age on syntactic complexity. Similarly to Agmon et al. (2024), all complexity measures showed patterns in favour of the younger cohort. In contrast however, the majority of our comparisons did not reach significance. A possible explanation for this discrepancy may be a difference in data preprocessing. While Agmon et al. used stringently cleaned data (e.g., removing filled pauses and dysfluencies) the present study retained such features to better reflect more naturalistic speech production. The inclusion of dysfluencies and filled pauses may introduce additional variability into measures of syntactic complexity via e.g., incomplete sentences. It is also not clear how the inclusion of dysfluencies and filled pauses affect the automatic sentence parsing although the sensitivity analysis of manually parsed data (Appendix B) showed only minor differences with no changes to significance. From this perspective, our pattern of effects may reflect a trade-off between ecological validity and data cleaning processes. It is also worth noting that, of the syntactic complexity measures, three pertained to centre embeddings which were scarce in our current dataset. Nevertheless, we found significant age-differences in both the number of clauses and number of NPs per sentence suggesting that age differences in these forms of syntactic complexity are highly robust.

The unique contribution of our study lies in analysing syntactic complexity effects in the context of the articulation and pausing characteristics of our participants' speech. Overall, our age groups did not differ in their fluency and pausing patterns when sentence level information was not

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

considered. We observed similar pausing patterns across age groups, and the complexity and length of fluent stretches of speech did not significantly differ. The only significant difference was in articulation rate (syllables per second in fluent stretches of speech) which was greater for younger than older adults.

However, the critical findings were those linking pausing behaviour to sentence complexity. We defined pauses of greater than 250 as planning pauses. Our analyses showed that both the number and duration of planning pauses significantly predicted the syntactic complexity measures that were affected by age (i.e., number of clauses and number of NPs). However, this effect did not vary between our older and younger speakers. Planning pauses therefore cannot explain the observed reduction of syntactic complexity with age. The shorter medial pauses also predicted syntactic complexity. Importantly however, these effects did significantly interact with age group. Both the number and duration of the medial paused differentially predicted the number of clauses in sentences for older and younger speakers. For younger speakers both measures predicted complexity significantly more strongly than for older adults. These findings suggest that the planning that occurred during medial pauses was more successful for younger than older adults. We suggest that the shorter medial pauses index local planning, and specifically the cost of integrating retrieved lexical material into the syntactic frame being built. This hypothesis is supported by the finding that the number of medial pauses significantly predicted the number of lexical items in sentences. Importantly this effect also interacted with age – again showing a stronger effect for younger than older adults suggesting that the lexical planning occurring during medial pauses was more successful for our younger than older speakers.

Despite the critical nature of planning scope for fluent language production, experimental studies of the effects of healthy aging on speech planning are rare. Two experimental studies conducted on a local locus. Hardy et al. (2020) elicited spoken sentences from younger and older adults using moving picture displays. Planning scope was tested by eliciting sentences with either simple or coordinated initial phrases (e.g., *The dog moves above the hat and the tree/The dog and the hat move above the tree*). Both age groups showed a similar effect of initial phrase size, with longer onset latencies to produce initial coordinate phrases. Therefore, no effect of age on planning scope was observed. However, on half of the trials the second picture (hat) was previewed, and the age groups differed in how they benefited from the additional information. When the previewed picture occurred in the initial phrase (initial coordinate phrases), responses were speeded in both age groups. When the previewed picture occurred beyond the first phrase (initial simple phrases), only the younger adults benefited while the older adults' error rates increased. Hardy et al. (2020) argued that older adults show increased difficulty with managing the activation and integration of lexical content during sentence planning. Fernandes et al. (2025) also used picture descriptions to examine planning scope in older and younger adults. They

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

demonstrated that older speakers were slower than younger to initiate small-scope prepositional phrases (e.g., the cone above the grape), suggesting they engage in more extensive rather than more limited planning prior to speech onset, when speeded and fluent output is required. This again is consistent with integration of lexical information into syntactic units being costly, which older speakers may compensate by front-loading planning. Both studies are consistent with our current findings which suggest that syntactic simplification in healthy ageing does not reflect a decrease in global planning scope but instead a local planning limitation with syntactic consequences.

To further examine what the medial pauses may reflect, as a post-hoc measure, we manually tagged each medial pause in accordance with what type of structure followed. We then grouped these measures together resulting in the plot displayed in Figure 6 below which shows each category and the proportion of pauses within that category grouped by age. A more detailed description of the contents of these categories is available in Appendix D.

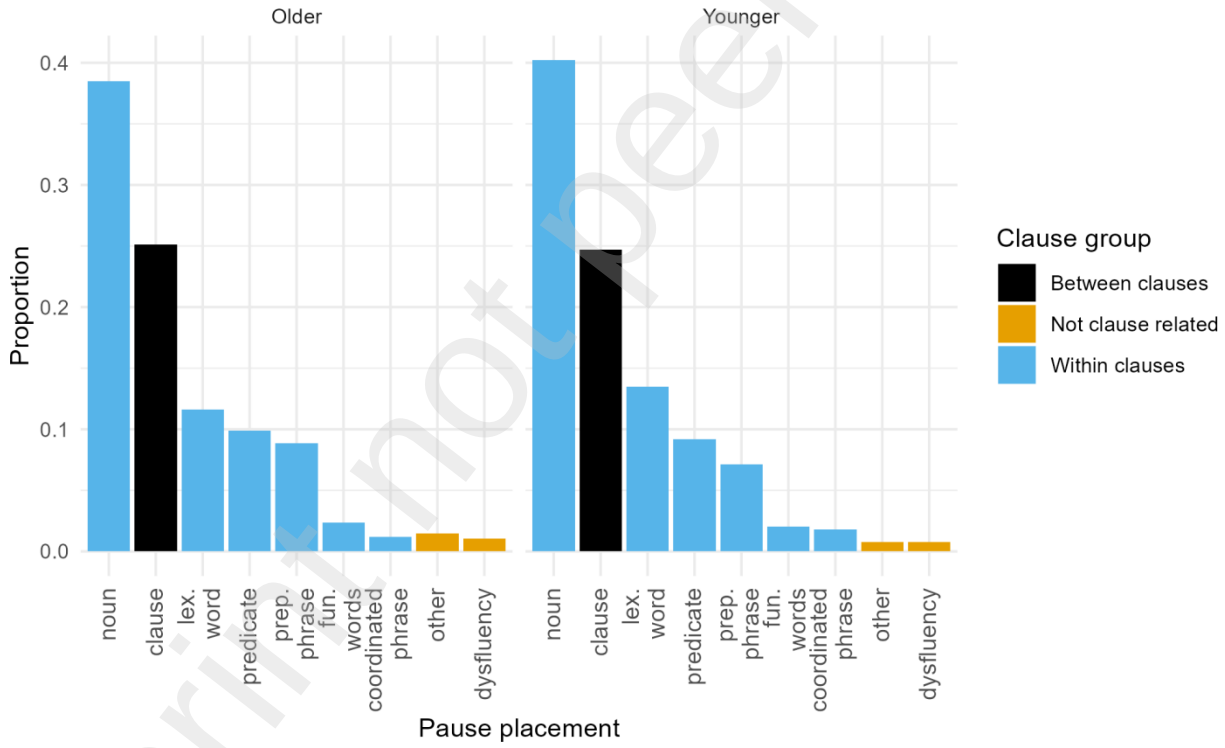


Figure 6: Distribution and categories of medial pauses by age-group.

The overall distributions appear similar between the two age-groups. In both cases, a large portion of medial pauses occurred in relation to nouns and noun phrases while the second-largest grouping is clauses (either subclauses or main clauses). Meanwhile, most medial pauses occurred within clauses suggesting a more local scope. It is also worth noting that when pauses occurred clause-initially it may also have been in relation to more local planning as opposed to larger

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

clausal structures. Word-finding may play a role in this effect given the prevalence of nouns. However, this includes all nouns produced regardless of their perceived difficulty of production. Lexical words beyond nouns, meanwhile, accounts for a much smaller proportion of pauses (i.e., adverbs, verbs, and adjectives where colours may have been particularly challenging word-finding wise). Instead, the current results may reflect local preparation at a phrasal level with nouns being a disproportionally large part of the data due to the large number of noun phrases which are produced in natural speech. That medial pauses cluster at phrasal and clausal boundaries, and are not confined to nouns or to plausibly hard-to-retrieve words, argues against pure word-finding. Instead, it points to a cost incurred when retrieved lexical information must be integrated into the unfolding structure.

Our study therefore examined language and ageing using a data set with a unique combination of variables. However, several limitations to our approach require some discussion. First, the unbalanced nature of our sample may have resulted in less precise estimates for the smaller (younger) group than for the larger (older) group. However, LMMs and GLMMs are well suited to handling unequal group sizes and unbalanced data and are generally robust to such imbalances (Baayen et al., 2008). A second limitation lies in the necessarily indirect inferences about planning processes. Our aim was to investigate natural connected speech rather than tightly controlled experimentally elicited sentences. Our study therefore infers local vs. global planning mechanisms primarily from pause duration and placement, rather than controlling for planning stages directly. Pausing during speaking is of course multifunctional, reflecting a mixture of conceptual, syntactic and prosodic and lexical planning processes, as well as monitoring and articulation-related processes (e.g., Bennett, 1986; Butterworth, 1980; Campione & Véronis, 2002; Clark & Fox Tree, 2002; Goldman-Eisler, 1968; Gósy, 2023; Grosjean et al., 1979; Henderson et al., 1966). This of course limits the precision with which specific planning stages can be isolated. Our operational distinction between planning and medial pause categories, relies on temporal thresholds rather than categorical cognitive markers. Shorter pauses may also reflect global planning processes and longer pauses may include lower-level planning difficulties and significant problems with word retrieval. Moreover, as discussed in the introduction, all such thresholds will be affected by multiple individual differences (e.g., Braun et al., 2023; van Donzel & Koopmans-van Beinum, 1996). As such, the mapping between pause type and cognitive process is probabilistic rather than absolute. Nevertheless, we demonstrate that shorter rather than longer pauses show significant age effects in the degree to which they predict syntactic complexity. Moreover, this age effect was replicated in the relationship between shorter pauses and number of lexical items. These effects are consistent with our hypothesis of local rather than global age-related planning impairments.

Finally, a third limitation involves the choice of task. Our semi-constrained picture description task, which balances experimental control and naturalness, does not capture the full range of

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

cognitive demands involved in everyday communicative discourse (e.g., topic selection, pragmatic demands, discourse structure). Our aim was to limit the use of compensatory strategies which might mask effects of ageing. However, it remains possible that our data might also underestimate age effects on global aspects of planning in fully spontaneous discourse.

Speech fluency timing measures are increasingly mined for early markers of neurodegenerative disease, where connected-speech pausing is treated as a candidate signal for Parkinson's and dementia (e.g., D'Ascanio et al., 2025; Nasreen et al., 2021). Our results add a note of specificity: in healthy ageing, the age signal was driven by *local* planning (short medial pauses), while *global* planning pauses were age-invariant. If healthy ageing and early-stage pathology load on different components of the pause distribution, then aggregate measures of "pausing" or overall speech rate may obscure the markers that are most diagnostic of disease. Distinguishing local from global pausing — as we do here — may therefore sharpen connected-speech measures as early markers, by separating the timing signatures of normal ageing from those of pathology.

In conclusion, our study highlights the importance of examining language complexity in the light of timing dimensions of language production relating to planning processes. Together, our findings suggest that age-related decline in syntactic complexity is largely a downstream consequence of more local planning issues, rather than a fundamental impairment of syntactic knowledge or global syntactic planning mechanisms.

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

References

Agmon, G., Pradhan, S., Ash, S., Nevler, N., Liberman, M., Grossman, M., & Cho, S. (2024). Automated Measures of Syntactic Complexity in Natural Speech Production: Older and Younger Adults as a Case Study. *Journal of Speech, Language, and Hearing Research*, *67*(2), 545–561. https://doi.org/10.1044/2023_JSLHR-23-00009

Allum, P. H., & Wheeldon, L. R. (2007). Planning Scope in Spoken Sentence Production: The Role of Grammatical Units. *Journal of Experimental Psychology: Learning Memory and Cognition*, *33*(4), 791–810. https://doi.org/10.1037/0278-7393.33.4.791

Allum, P. H., & Wheeldon, L. R. (2009). Scope of Lexical Access in Spoken Sentence Production: Implications for the Conceptual-Syntactic Interface. *Journal of Experimental Psychology: Learning Memory and Cognition*, *35*(5), 1240–1255. https://doi.org/10.1037/a0016367

Andreetta, S., & Marini, A. (2015). The effect of lexical deficits on narrative disturbances in fluent aphasia. *Aphasiology*, *29*(6), 705–723. https://doi.org/10.1080/02687038.2014.979394

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baayen, R., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database. In *Linguistic Data Consortium*.

Beier, E. J., Chantavarin, S., & Ferreira, F. (2023). Do Disfluencies Increase With Age? Evidence From a Sequential Corpus Study of Disfluencies. *Psychology and Aging*, *38*(3), 203–218. https://doi.org/10.1037/pag0000741

Bennett, P. R. (1986). The role of pause in discourse and its place in linguistics: Some evidence from Eastern Bantu. *Language Sciences*, *8*(1), 63–79. https://doi.org/10.1016/S0388-0001(86)80006-7

Bock, K. J., & Levelt, W. J. M. (1994). Language production: grammatical encoding. In M. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 945–984). Academic Press.

Boersma, P., Weenink, D., & Schupak, A. (2026). *Praat: doing phonetics by computer*.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, *44*(2), 123–147. https://doi.org/10.1177/00238309010440020101

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, *26*(2), 211–252.

Braun, A., Elsässer, N., & Willems, L. (2023). Disfluencies Revisited—Are They Speaker-Specific? *Languages*, *8*(3), 155. https://doi.org/10.3390/languages8030155

Burke, D. M., Mackay, D. G., & Worthley, J. (1991). On the Tip of the Tongue: What Causes Word Finding Failures in Young and Older Adults? In *JOURNAL OF MEMORY AND LANGUAGE* (Vol. 30).

Butterworth, B. (1980). Evidence from Pauses in Speech. In B. Butterworth (Ed.), *Language Production: Speech and Talk* (pp. 155–176). Academic Press.

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. *Speech Prosody 2002*, 199–202. https://doi.org/10.21437/SpeechProsody.2002-35

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–94. https://doi.org/10.1017/S0140525X99001788

Carrol, J. B. (1964). *Language and Thought*. Prentice-Hall.

Carson, N., Leach, L., & Murphy, K. J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *International Journal of Geriatric Psychiatry*, *33*(2), 379–388. https://doi.org/10.1002/gps.4756

Chang, F., Dell, G. S., & Bock, K. J. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. https://doi.org/10.1037/0033-295X.113.2.234

Charniak, E., Johnson, M., & McClosky, D. (2021). *BLLIP Parser* (2021.11.7). https://pypi.org/project/bllipparser/

Clark, H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73–111. https://doi.org/10.1016/S0010-0277(02)00017-3

Dell, G. S., & Chang, F. (2014). The P-chain: relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1634). https://doi.org/10.1098/rstb.2012.0394

Dell, G. S., Oppenheim, G. M., & Kittredge, A. K. (2008). Saying the right word at the right time: Syntagmatic and paradigmatic interference in sentence production. *Language and Cognitive Processes*, *23*(4), 583–608. https://doi.org/10.1080/01690960801920735

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*(2), 210–233. https://doi.org/10.1016/0749-596X(91)90004-4

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*(1), 57–84. https://doi.org/10.1006/jmla.2001.2797

Feyereisen, P. (1997). A Meta-Analytic Procedure Shows an Age-Related Decline in Picture Naming. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1328–1333. https://doi.org/10.1044/jslhr.4006.1328

Frohmann, M., Sterner, I., Vulić, I., Minxhofer, B., & Schedl, M. (2024). Segment Any Text: A Universal Approach for Robust, Efficient and Adaptable Sentence Segmentation. *Computer Science*. https://doi.org/https://doi.org/10.48550/arXiv.2406.16678

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: telephone speech corpus for research and development. *[Proceedings] ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 517–520 vol.1. https://doi.org/10.1109/ICASSP.1992.225858

Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in Spontaneous Speech*. Academic Press.

Goodglass, H., & Kaplan, E. (1983). *The Assessment of Aphasia and Related Disorders* (2nd ed.). Lea and Febiger.

Gósy, M. (2023). Occurrences and Durations of Filled Pauses in Relation to Words and Silent Pauses in Spontaneous Speech. *Languages*, *8*(1), 79. https://doi.org/10.3390/languages8010079

Griffin, Z. M. (2001). *Gaze durations during speech reflect word selection and phonological encoding*. *82*(1), 1–16.

Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, *11*(1), 58–81. https://doi.org/10.1016/0010-0285(79)90004-5

Hardy, S. M., Segaert, K., & Wheeldon, L. (2020). Healthy Aging and Sentence Production: Disrupted Lexical Access in the Context of Intact Syntactic Planning. *Frontiers in Psychology*, *11*. https://doi.org/10.3389/fpsyg.2020.00257

Henderson, A., Goldman-Eisler, F., & Skarbek, A. (1966). Sequential Temporal Patterns in Spontaneous Speech. *Language and Speech*, *9*(4), 207–216. https://doi.org/10.1177/002383096600900402

Hieke, A. E., Kowal, S., & O'Connell, D. C. (1983). The Trouble with "Articulatory" Pauses. *Language and Speech*, *26*(3), 203–214. https://doi.org/10.1177/002383098302600302

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Hilviu, D., Parola, A., Bosco, F. M., Marini, A., & Gabbatore, I. (2025). Grandpa, tell me a story! Narrative ability in healthy aging and its relationship with cognitive functions and Theory of Mind. *Language, Cognition and Neuroscience*, *40*(1), 103–121. https://doi.org/10.1080/23273798.2024.2401027

Horton, W. S., Spieler, D. H., & Shriberg, E. (2010). A corpus analysis of patterns of age-related change in conversational speech. *Psychology and Aging*, *25*(3), 708–713. https://doi.org/10.1037/a0019424

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Jongman, S. R., Meyer, A. S., & Roelofs, A. (2015). The Role of Sustained Attention in the Production of Conjoined Noun Phrases: An Individual Differences Study. *PLOS ONE*, *10*(9), e0137557. https://doi.org/10.1371/journal.pone.0137557

Jongman, S. R., Roelofs, A., & Meyer, A. S. (2015). Sustained attention in language production: An individual differences investigation. *Quarterly Journal of Experimental Psychology*, *68*(4), 710–730. https://doi.org/10.1080/17470218.2014.964736

Kavé, G. (2022). Vocabulary changes in adulthood: Main findings and methodological considerations. *International Journal of Language & Communication Disorders*, *59*(1), 58–67. https://doi.org/10.1111/1460-6984.12820

Kavé, G., & Goral, M. (2017). Do age-related word retrieval difficulties appear (or disappear) in connected speech? *Aging, Neuropsychology, and Cognition*, *24*(5), 508–527. https://doi.org/10.1080/13825585.2016.1226249

Kavé, G., Knafo, A., & Gilboa, A. (2010). The rise and fall of word retrieval across the lifespan. *Psychology and Aging*, *25*(3), 719–724. https://doi.org/10.1037/a0018927

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science*, *11*(2), 201–258. https://doi.org/10.1016/S0364-0213(87)80006-X

Kemper, S., Herman, R., & Lian, C. (2003). Age Differences in Sentence Production. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *58*(5), 260–268. https://doi.org/10.1093/geronb/58.5.P260

Konopka, A. E. (2012). Planning ahead: How recent experience with structures and words changes the scope of linguistic planning. *Journal of Memory and Language*, *66*(1), 143–162. https://doi.org/10.1016/j.jml.2011.08.003

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Konopka, A. E., & Kuchinsky, S. E. (2015). How message similarity shapes the timecourse of sentence formulation. *Journal of Memory and Language*, *84*, 1–23. https://doi.org/10.1016/j.jml.2015.04.003

Konopka, A. E., & Meyer, A. S. (2014). Priming sentence planning. *Cognitive Psychology*, *73*, 1–40. https://doi.org/10.1016/j.cogpsych.2014.04.001

Konopka, A. E., Meyer, A. S., & Forest, T. A. (2018). Planning to speak in L1 and L2. *Cognitive Psychology*, *102*, 72–104. https://doi.org/10.1016/j.cogpsych.2017.12.003

Krivokapić, J. (2007). Prosodic planning: Effects of phrasal length and complexity on pause duration. *Journal of Phonetics*, *35*(2), 162–179. https://doi.org/10.1016/j.wocn.2006.04.001

Kuchinsky, S. E., Bock, K. J., & Irwin, D. E. (2011). Reversing the Hands of Time: Changing the Mapping From Seeing to Saying. *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(3), 748–756. https://doi.org/10.1037/a0022637

Lenth, R. V., & Piaskowski, J. (2026). emmeans: Estimated Marginal Means, aka Least-Squares Means. *R Package Version 2.0.3*. https://rvlenth.github.io/emmeans/.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75. https://doi.org/10.1017/S0140525X99001776

Liu, H., Xu, C., & Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, *21*, 171–193. https://doi.org/10.1016/j.plrev.2017.03.002

Lo, S., & Andrews, S. (2015). To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01171

Marini, A., Boewe, A., Caltagirone, C., & Carlomagno, S. (2005). Age-related Differences in the Production of Textual Descriptions. *Journal of Psycholinguistic Research*, *34*(5), 439–463. https://doi.org/10.1007/s10936-005-6203-z

Marini, A., Petriglia, F., Gasparotto, G., D'Ortenzio, S., Andreetta, S., & Gobbo, M. (2026). Standardization of the MultiLevel discourse Analysis (MLA) and identification of data-driven age bands for research on language in healthy aging. *Journal of Neurolinguistics*, *79*, 101343. https://doi.org/10.1016/j.jneuroling.2026.101343

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Marlin, R. C., & Slevc, L. R. (2014). Language production and working memory. In M. A. Goldrick, V. S. Ferreira, & M. Miozzo (Eds.), *The Oxford handbook of language production* (pp. 437–450).

Martin, R. C., Yan, H., & Schnur, T. T. (2014). Working memory and planning during sentence production. *Acta Psychologica*, *152*, 120–132. https://doi.org/10.1016/j.actpsy.2014.08.006

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *Interspeech 2017*, 498–502. https://doi.org/10.21437/Interspeech.2017-1386

Momma, S., & Ferreira, V. S. (2021). Beyond linear order: The role of argument structure in speaking. *Cognitive Psychology*, *128*. https://doi.org/10.1016/j.cogpsych.2021.101397

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Payne, B. R., & Stine-Morrow, E. A. L. (2014). Adult age differences in wrap-up during sentence comprehension: Evidence from ex-Gaussian distributional analyses of reading time. *Psychology and Aging*, *29*(2), 213–228. https://doi.org/10.1037/a0036282

Piai, V., & Roelofs, A. (2013). Working memory capacity and dual-task interference in picture naming. *Acta Psychologica*, *142*(3), 332–342. https://doi.org/10.1016/j.actpsy.2013.01.006

Pickering, M. J., & Branigan, H. P. (1998). The Representation of Verbs: Evidence from Syntactic Priming in Language Production. *Journal of Memory and Language*, *39*(4), 633–651. https://doi.org/10.1006/jmla.1998.2592

Radford, A., Kim, J. W., Xu, T., McLeavey, C., & Sutskever, I. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. *Electrical Engineering and Systems Science* . https://doi.org/https://doi.org/10.48550/arXiv.2212.04356

Samani, M., Abnavi, F., & Ghasisin, L. (2017). Do older adults experience changes in their speech fluency? Some evidence from Iranian elderly people. *Journal of Clinical Gerontology and Geriatrics*, *8*(4), 127–133. https://doi.org/10.24816/jcgg.2017.v8i4.05

Schmitter-Edgecombe, M., Vesneski, M., & Jones, D. W. R. (2000). Aging and Word-Finding A Comparison of Spontaneous and Constrained Naming Tests. *Archives of Clinical Neuropsychology*, *15*(6), 479–493. https://doi.org/10.1016/S0887-6177(99)00039-6

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language*, *29*(1), 86–102. https://doi.org/10.1016/0749-596X(90)90011-N

Shao, Z., Roelofs, A., & Meyer, A. S. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *Quarterly Journal of Experimental Psychology*, *65*(10), 1927–1944. https://doi.org/10.1080/17470218.2012.670252

Shriberg, E. (1996). *DISFLUENCIES IN SWITCHBOARD*.

Smith, M., & Wheeldon, L. R. (1999). High level processing scope in spoken sentence production. *Cognition*, *73*(3), 205–246. https://doi.org/10.1016/S0010-0277(99)00053-0

Stathopoulos, E. T., & Weismer, G. (1985). Oral Airflow and Air Pressure during Speech Production: A Comparative Study of Children, Youths and Adults. *Folia Phoniatrica et Logopaedica*. https://doi.org/10.1159/000265794

Sung, J. E., Jo, E., Choi, S., & Lee, J. (2024). Coordinating Words and Sentences: Detecting Age-Related Changes in Language Production. *Journal of Speech, Language, and Hearing Research*, *67*(1), 211–220. https://doi.org/10.1044/2023_JSLHR-23-00222

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64–81. https://doi.org/10.1037/0096-3445.136.1.64

Swets, B., Fuchs, S., Krivokapić, J., & Petrone, C. (2021). A Cross-Linguistic Study of Individual Differences in Speech Planning. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.655516

Swets, B., Jacovina, M. E., & Gerrig, R. J. (2014). Individual differences in the scope of speech planning: evidence from eye-movements. *Language and Cognition*, *6*(1), 12–44. https://doi.org/10.1017/langcog.2013.5

van de Velde, M., Meyer, A. S., & Konopka, A. E. (2014). Message formulation and structural assembly: Describing “easy” and “hard” events with preferred and dispreferred syntactic structures. *Journal of Memory and Language*, *71*(1), 124–144. https://doi.org/10.1016/j.jml.2013.11.001

van Donzel, M. E., & Koopmans-van Beinum, F. J. (1996). Pausing strategies in discourse in Dutch. *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP ’96*, 1029–1032. https://doi.org/10.1109/ICSLP.1996.607779

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190.

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging*, *18*(2), 332–339. https://doi.org/10.1037/0882-7974.18.2.332

Wang, M., & Wang, M. (2024). Age-related differences in the interplay of fluency and complexity in Chinese-speaking seniors' oral narratives. *International Journal of Language and Communication Disorders*, *59*(5), 1672–1690. https://doi.org/10.1111/1460-6984.13023

Wei, H. T., Kulzhabayeva, D., Erceg, L., Robin, J., Hu, Y. Z., Chignell, M., & Meltzer, J. A. (2024). Cognitive components of aging-related increase in word-finding difficulty. *Aging, Neuropsychology, and Cognition*, *31*(6), 987–1019. https://doi.org/10.1080/13825585.2024.2315774

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Appendix A: Descriptive Statistics of Syntactic Complexity for Automatic and Manually Parsed Sentences

Table 16: Descriptive statistics for syntactic complexity in sentences comparing the automatic (left) and manual (right) segmentation methods for the younger adults.

Measure	Younger - Automatic			Younger - Manual		
	M	*SD*	Range	*M*	*SD*	Range
Number of Clauses	1.79	1.48	[0, 11]	1.63	0.93	[0, 6]
Number of Rel. Clauses	0.23	0.49	[0, 2]	0.22	0.46	[0, 3]
Number of NPs	4.39	2.86	[1, 14]	4.32	2.64	[1, 12]
NP Length	2.86	1.35	[1, 8.62)	2.86	1.43	[1, 10.40]
Centre Embed Presence	.23	.43	NA	0.16	0.37	NA
Number of Centre Embeds	0.32	0.67	[0, 4]	0.17	0.39	[0, 2]
Length of Centre Embeds	2.94	2.06	[1, 16]	2.60	1.33	[1, 17]

Table 17: Descriptive statistics for syntactic complexity in sentences comparing the automatic (left) and manual (right) segmentation methods for the older adults.

Measure	Older - Automatic			Older - Manual		
	M	*SD*	Range	*M*	*SD*	Range
Number of Clauses	1.54	1.28	[0, 12]	1.48	0.96	[0, 7]
Number of Rel. Clauses	0.19	0.47	[0, 4]	0.19	0.46	[0, 3]
Number of NPs	3.87	2.51	[1, 14]	4.00	2.45	[1, 13]
NP Length	2.76	1.24	[1, 12.40]	2.82	1.27	[1, 9.09]
Centre Embed Presence	.19	.39	NA	0.15	0.35	NA
Number of Centre Embeds	0.23	0.50	[0, 3]	0.16	0.40	[0, 3]
Length of Centre Embeds	2.64	1.61	[0, 14]	2.50	1.25	[0, 13]

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Appendix B: Regression Analyses of Manually Parsed Sentences

Following the same exclusion criteria as in the main text, 9 sentences were removed for containing zero NPs (.003 of the data) with an additional 20 being removed for being in the 99th percentile (.01 of the data).

Table 18: Sensitivity analysis of syntactic complexity measures in manually segmented sentences.

Measure	Predictor	Est.	*SE*	*z/t*	*p*	sig.
Number of Clauses	Age-group	-0.05	0.02	-2.02	.04	*
	Education	0.03	0.02	1.27	.21	
Number of Rel. Clauses	Age-group	-0.10	0.07	-1.34	.18	
	Education	-0.00	0.07	-0.01	.99	
Number of NPs	Age-group	-0.05	0.02	-2.39	.02	*
	Education	0.02	0.02	0.87	.38	
NP Length	Age-group	-0.00	0.01	-0.20	.84	
	Education	0.00	0.01	0.22	.83	
Centre Embed Presence	Age-group	-0.11	0.09	-1.16	.25	
	Education	0.08	0.08	0.95	.34	
Number of Centre Embeds	Age-group	-0.03	0.06	0.44	.66	
	Education	0.01	0.05	0.18	.86	
Length of Centre Embeds	Age-group	0.01	0.03	0.32	.75	
	Education	0.00	0.02	0.07	.94	

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Appendix C: FSU Syntactic Complexity Output

Table 19: Syntactic complexity model summaries for FSUs.

Measure	Predictor	Est.	*SE*	*z/t*	*p*	sig.
Number of Clauses	Age-group	-0.01	0.03	-0.26	.80	
	Education	0.02	0.03	0.85	.40	
Number of Rel. Clauses	Age-group	-0.04	0.08	-0.57	.57	
	Education	-0.01	0.07	-0.07	.94	
Number of NPs	Age-group	-0.02	0.02	-0.91	.36	
	Education	-0.00	0.02	-0.17	.86	
NP Length	Age-group	-0.00	0.01	-0.62	.53	
	Education	-0.00	0.01	-0.33	.75	
Centre Embed Presence	Age-group	-0.10	0.08	-1.22	.22	
	Education	.0.05	0.07	0.74	.46	
Number of Centre Embeds	Age-group	-0.02	0.06	-0.31	.76	
	Education	0.02	0.05	0.49	.62	
Length of Centre Embeds	Age-group	0.03	0.03	1.17	.24	
	Education	0.00	0.02	0.09	.93	

Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

Appendix D: Overview of categories following medial pauses

Table 20: Medial pause category overview.

Category	Constituents	Comment
Clause	Sentence boundary, main clauses, subclause, coordinated clause	Clauses defined as any subject + predicate combination.
Coordinated phrase	Coordinated phrases	Any phrase in a coordinated structure.
Dysfluency	E.g., corrections, repetitions, stutters	
Function words	Stranded conjunctions, demonstratives, stranded determiners, particles, stranded prepositions, pronouns, and numerals[12]	Conjunctions, determiners, and prepositions that were not stranded were grouped with other categories[13]
Lexical Word	Adjectives, adverbs, verbs	Nouns grouped separately due to their prevalence.
Noun	Nouns, noun phrases, and post-modifiers of nouns.	
Other	Comments, lists, and time-telling	In one picture it was to some participants unclear whether the time showed 10 or 11 o'clock which may have incurred pausing.
Predicate	VPs or obligatory complements to VPs.	
Prepositional Phrase	Prepositional phrases	Also likely related heavily to noun due to typical P + NP structure of PPs.

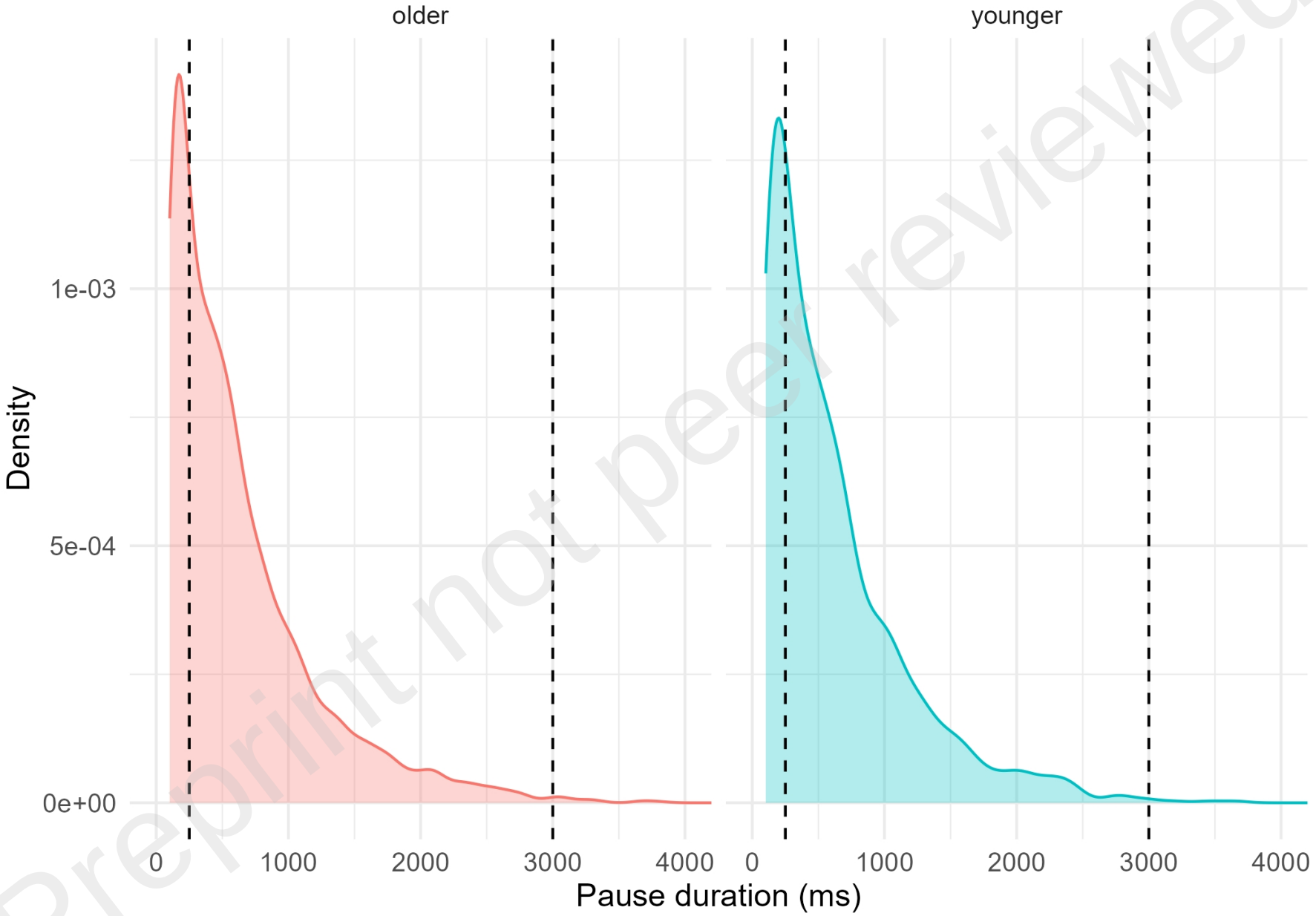
[12] Stranding examples:

Conjunction (*n* = 4): "a mum <medial pause> and a there's a (...)
Preposition (*n* = 18): "not sure what that is <medial pause> on <end of sentence>.
Determiner (*n* = 4): "looks as if there is <medial pause> a <end of sentence> it looks just like (...)
[13] Conjunctions: Coordinated phrases or causes
Determiners: Noun phrases
Prepositions: Prepositional phrases

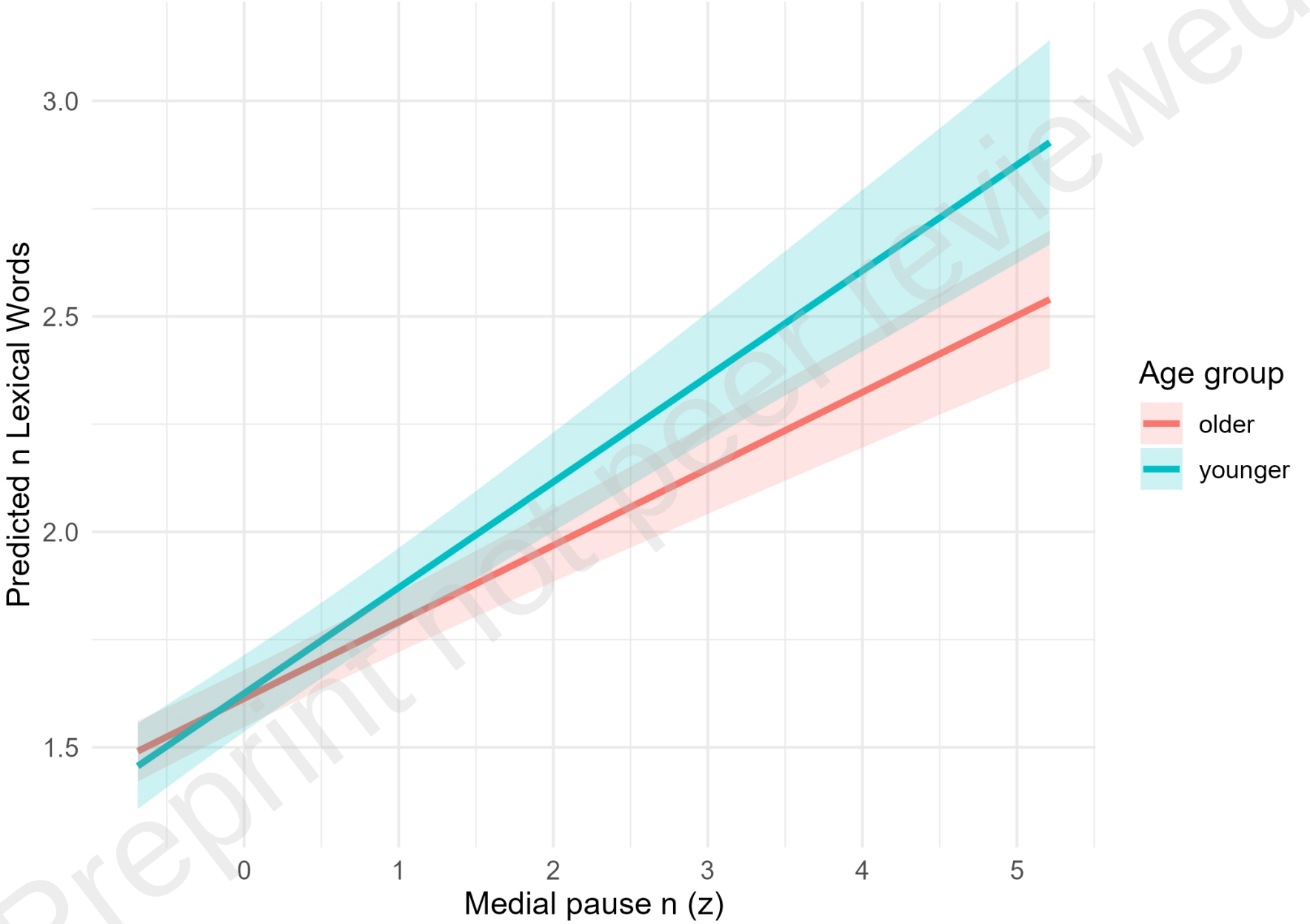
Preprint not peer reviewed

This preprint research paper has not been peer reviewed. Electronic copy available at: https://ssrn.com/abstract=7206250

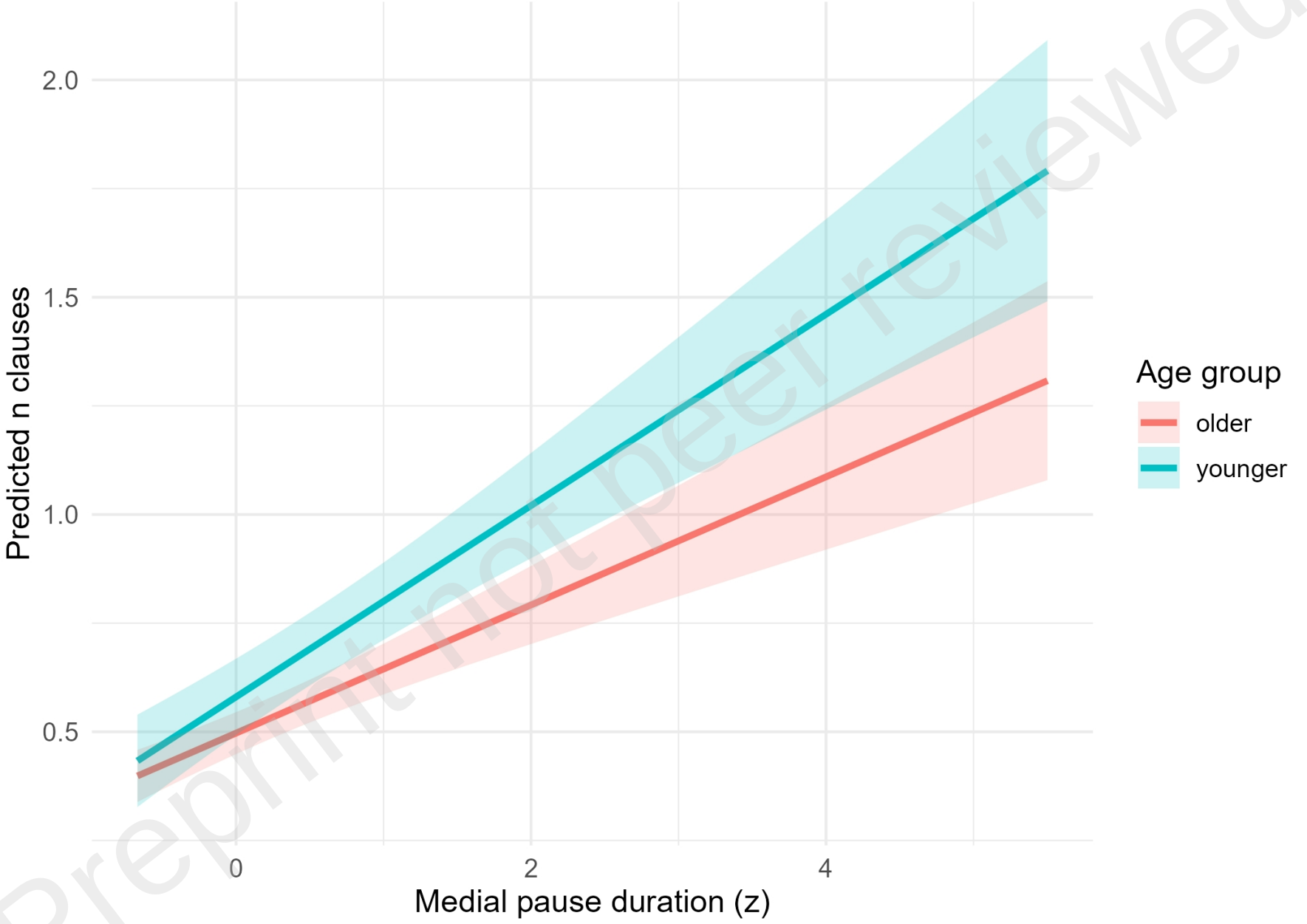
Preprint not peer reviewed



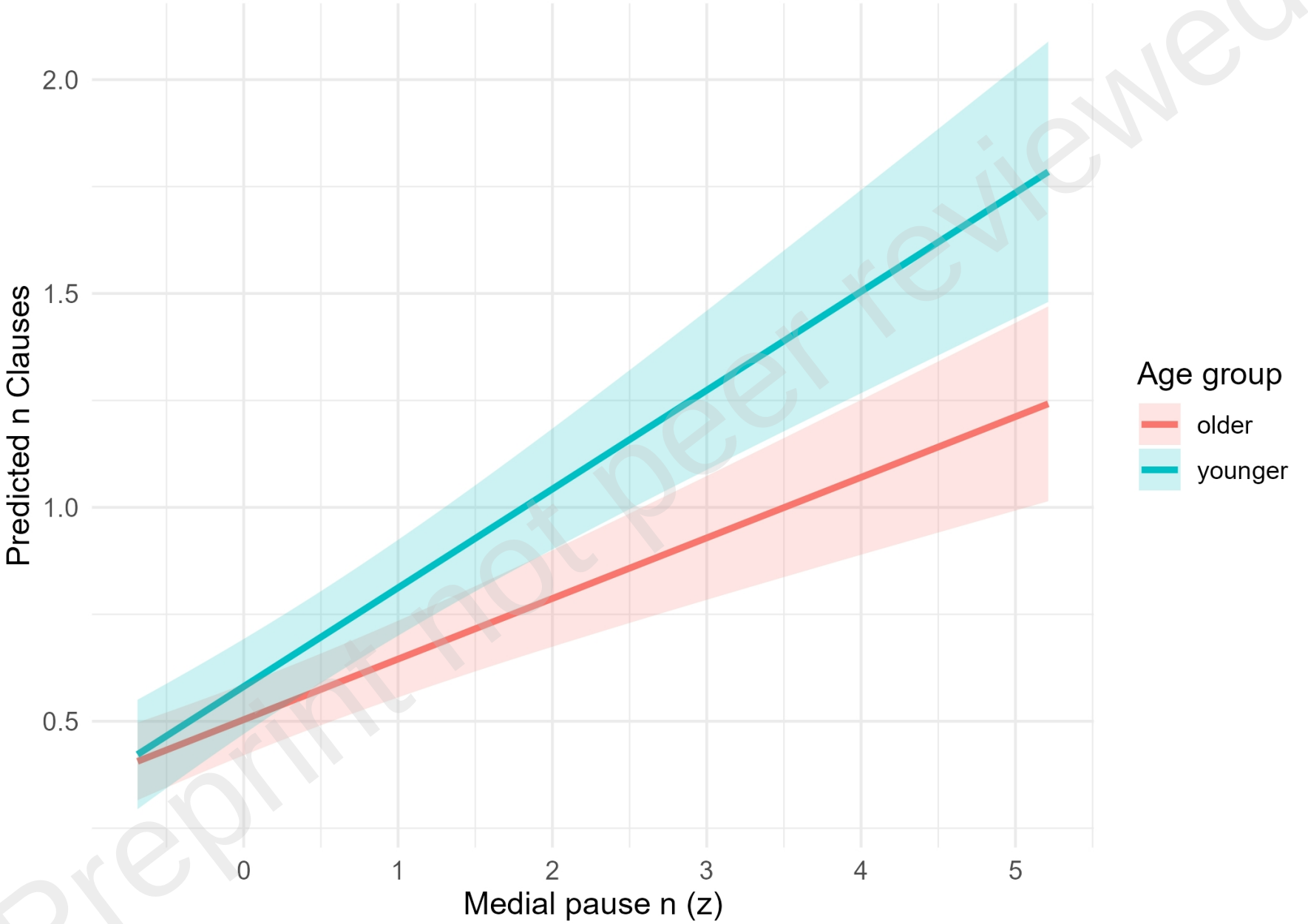
Preprint not peer reviewed



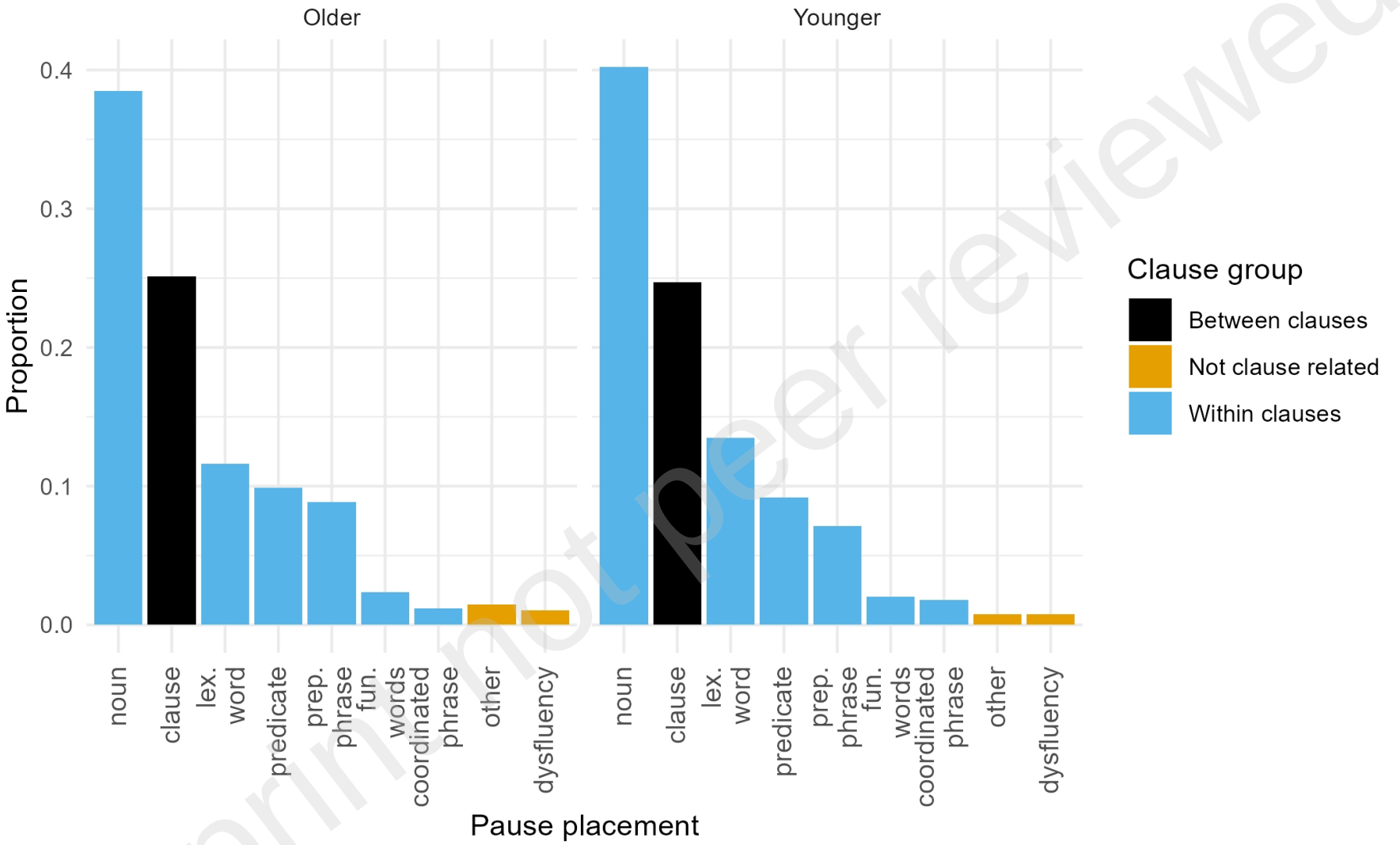
Preprint not peer reviewed



Preprint not peer reviewed



Preprint not peer reviewed



Preprint not peer reviewed