TABU

Festschrift for Jack Hoeksema

A LONGITUDINAL EVENT-RELATED POTENTIALS STUDY OF IDIOM PROCESSING IN HEALTHY ELDERLY ADULTS

Amélie la Roi, Simone A. Sprenger and Petra Hendriks

Abstract

This study investigates the longitudinal effects of cognitive aging on elderly adults' sentence processing. To this aim, 16 Dutch elderly adults who participated in a previous study on idiom processing were tested again three years later, using the same event-related potential (ERP) paradigm. Compared to the previous study, we found the same reduced N400 for idiomatic sentences and for sentences preceded by a predictive context, and the same increased P600 for literal sentences preceded by a predictive context. However, the general N400 effect decreased over time, whereas the general P600 effect increased over time. Our findings suggest that elderly adults' ability to benefit from context information and the fixed character of idioms to facilitate sentence processing is preserved over time. However, the longitudinal changes in the N400 and P600 effect suggest that elderly adults adopt alternative processing strategies to compensate for age-related cognitive decline.

Keywords: cognitive aging, Dutch, ERP, idioms

1. Introduction[1,2]

As average life expectancies across the globe are increasing, we are facing an increasing number of people who show signs of cognitive decline. Even in the absence of neurodegenerative diseases, performance on cognitive tasks measuring working memory

Center for Language and Cognition Groningen, University of Groningen
Corresponding author: Petra Hendriks, p.hendriks@rug.nl

Published by University of Groningen Press,
How to cite this article: La Roi, A., Sprenger, S.A. & Hendriks, P. (2024). A longitudinal event-related potentials study of idiom processing in healthy elderly adults. *TABU Festschrift for Jack Hoeksema*. 282-311.
https://doi.org/10.21827/tabu.2023.41249

capacity, inhibition skills and processing speed has been shown to decline in elderly adults. In younger adults, language processing abilities are strongly related to performance on these cognitive tasks, suggesting that an important side effect of age-related cognitive decline could be that it makes language and communication increasingly difficult as well. Several studies have indeed linked the effects of cognitive aging to age-related decline in language abilities. For example, the larger the working memory capacity in elderly adults, the better their reading comprehension (e.g., Daneman & Carpenter, 1980; DeDe et al., 2004) and the more they benefit from context information in sentence processing (e.g., Federmeier & Kutas, 2005). In addition, elderly adults with low inhibition skills have shown a reduced ability to resolve ambiguity (e.g., Lee & Federmeier, 2012).

Thus, decline in elderly adults' cognitive abilities indeed seems to affect language skills. However, it remains unknown how fast language abilities decline as a consequence of cognitive aging. In addition, it is unclear to what extent individual differences play a role in the slope of the decline. It is conceivable that there is substantial variability between elderly individuals regarding the way in which language abilities change over time, for example due to intelligence, level of education, social economic status, or genetics. These factors may not only influence the magnitude of the decline in elderly adults' language abilities, but also the slope of the decline. Therefore, studies on the effects of cognitive aging on language abilities would benefit from a longitudinal approach, because this provides insight in the speed at which language abilities change as a consequence of cognitive aging. Furthermore, a longitudinal design takes into account the effects of interindividual variability by examining age-related changes in language abilities within the same group of individuals.

Yet, although studying language skills longitudinally is quite common in child research, longitudinal studies on elderly adults' language skills are scarce. The few longitudinal studies that examine linguistic abilities across the lifespan focus on oral or written language production (e.g., Kemper et al., 2001) or lexical processing (e.g., Rönnlund et al., 2005). Although a study by Vonk et al. (2020) investigated longitudinal changes in elderly adults' sentence processing, this study only includes behavioral measures. As it is expected that age-related cognitive decline reduces the speed at which the information in sentences is processed, potentially leading to comprehension difficulties in later stages, we need a method that is sensitive to the fast pace and complexity of the processes involved in sentence processing.

Therefore, this longitudinal study uses event-related potentials (ERPs) to examine how language processing abilities change over time as a consequence of cognitive aging. To investigate to what extent increasing age in elderly adults leads to difficulties in language processing due to cognitive decline, we focus on a type of language processing that has been shown to be particularly dependent on cognitive processing, namely the processing of ambiguous expressions. Ambiguity resolution has been shown to depend on inhibition skills (e.g., Gunter et al., 2003; Lee & Federmeier, 2012; Zempleni et al., 2007), and inhibition skills have been found to decline with age (e.g., Van Hooren et al., 2007). As a measure of ambiguity resolution, we focus on the processing of idiomatic expressions such as *to kick the bucket,* which are ambiguous between a figurative interpretation ('to die') and a literal, compositional interpretation ('to kick the pail'). In addition, we investigate the processing of these expressions in context, as the ability to benefit from context information has been linked to working memory capacity (e.g., Federmeier & Kutas, 2005), and working memory capacity declines with age as well (e.g., Bopp & Verhaeghen, 2005).

1.1 Using ERPs to investigate effects of cognitive aging on sentence processing

In a previous study, la Roi et al. (2020) used ERPs to compare younger and elderly adults' processing of literal and idiomatic sentences in context. They found that 1) both younger and elderly adults showed facilitated processing of idiomatic relative to literal sentences, consistent with findings from previous studies (e.g., Beck & Weber, 2020; Canal et al., 2015; Carrol & Littlemore, 2020; Laurent et al., 2006; Rommers et al., 2013; Titone & Connine, 1999), 2) elderly adults were as able as younger adults to use context to facilitate the retrieval of word meanings in both literal and idiomatic sentences, and 3) elderly adults, but not younger adults, relied on context information for the subsequent integration of these word meanings in the sentence. As the ERP paradigm used in la Roi et al. (2020) has proven its value in investigating sentence processing in elderly adults, we investigate the same elderly adults who participated in la Roi et al. (2020) again three years later, using the same paradigm, to examine how cognitive aging in these elderly adults affects sentence processing. To our knowledge, our longitudinal ERP study on healthy elderly adults' sentence processing constitutes a novel approach to studying the effects of age-related cognitive decline on language processing.

Longitudinal changes in healthy elderly adults' online processing of language have remained largely unexplored. Most longitudinal ERP studies on elderly adults focus on changes in cognitive processing in clinical populations, such as patients with Mild Cognitive Impairment or Alzheimer's dementia (Ball et al., 1989; Papaliagkas et al., 2011). We found only one longitudinal ERP study that investigated healthy elderly adults. In this study, the researchers examined the effect of repeated exposure to speech on elderly adults' auditory processing (Giroud et al., 2018). However, this study investigated the processing of syllables instead of sentences. Thus, none of the previous studies has examined the longitudinal effects of cognitive aging on sentence processing.

1.2 The present study

To investigate longitudinal changes in healthy elderly adults' processing of literal and idiomatic sentences in context, we focus on the N400 and P600 components. We selected the N400 because it indexes the ability to benefit from context information to facilitate the retrieval of word meanings in literal and idiomatic sentences (e.g., Federmeier & Kutas, 2005; Kutas & Federmeier, 2011). In addition, the N400 has been found to reflect the facilitated processing of idiomatic relative to literal sentences (e.g., Canal et al., 2015; Laurent et al., 2006; Rommers et al., 2013), possibly due to the fact that an idiom's figurative meaning can be retrieved from the mental lexicon as a whole (e.g., Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Sprenger et al., 2006; Titone & Connine, 1999). The P600 was selected because it has been argued to indicate the effort involved in the integration of information that was retrieved in earlier processing stages (e.g., Brouwer et al., 2017; Delogu et al., 2019; Ness & Meltzer-Asscher, 2018). Supportive context information can facilitate the integration of information by enabling the reader to predict and pre-activate upcoming words.

In order to assess the relationship between age-related cognitive decline and longitudinal changes in elderly adults' ability to benefit from the facilitative effect of context and idioms on language processing, we include a measure of cognitive processing. Many studies of the effects of cognitive aging on language processing have used verbal fluency as a proxy for a large variety of cognitive processes, including processing speed (Grindrod & Raizen, 2019), predictive processing (e.g., DeLong et al., 2012), and effortful processing that depends on frontal lobe functioning (Dave et al., 2018). Although no consensus has been reached yet on the exact mechanisms underlying performance on verbal fluency tasks, multiple studies have indicated a relationship between verbal fluency and the ability to use context in sentence processing (e.g., Dave et al., 2018; Grindrod & Raizen, 2019), as well as the ability to resolve ambiguity (e.g., Lee & Federmeier, 2011). As the verbal fluency task is a verbal task, it does not provide a measure of cognitive processing that is strictly independent from language processing. However, it enables us to compare our results to the findings from previous studies. Moreover, verbal fluency has been shown to decline rapidly in old age (e.g., Clark et al., 2009; Hatta et al., 2020). Therefore, using verbal fluency as a measure of cognitive processing enables us to link a potential longitudinal decline in elderly adults' language processing abilities to their level of cognitive processing.

In la Roi et al.'s (2020) study, performance on the Verbal Fluency task significantly predicted participants' N400 and P600 amplitudes, indicating that the mechanisms underlying verbal fluency play a role in the use of context in literal and idiomatic sentence processing. La Roi et al. (2020) also carried out a Dutch Reading Span task (Van Den Noort et al., 2008) to measure working memory capacity, and a Dutch version of the Paired-Associates task (Shimamura et al., 1995) to measure inhibition skills. However, performance on these two tasks did not predict participants' brain responses and therefore are not included in the present study.

1.3 Research questions and predictions

This study investigates to what extent the longitudinal effects of cognitive aging affect elderly adults' language processing. It addresses the following two research questions: 1) To what extent do the longitudinal effects of cognitive aging affect elderly adults' ability to benefit from the fixed character of idioms in sentence processing?, and 2) To what extent do the longitudinal effects of cognitive aging affect elderly adults' ability to benefit from context information in sentence processing?

To investigate these questions, we apply a longitudinal ERP paradigm and test the same elderly adults who participated in la Roi et al. (2020) again three years later. That is, we compare the N400 and P600 measured in the elderly adults in la Roi et al. (2020), forming the *baseline* session, with the N400 and P600 elicited with the same paradigm and in the same elderly adults three years later (factor *Session*), constituting the *follow-up* session. The N400 and P600 will be measured in response to literal and idiomatic sentences (factor *Idiomaticity*) preceded by a neutral or a predictive context sentence (factor *Context*). Furthermore, given the fact that ERPs are usually studied in response to a violation, half of the literal and idiomatic sentences will contain the correct target word, whereas in the other half of the sentences an incorrect target word will elicit a semantic violation (factor *Correctness*).

La Roi et al. (2020) found that the elderly adults they tested were as able as younger adults to benefit from context and the fixed character of idioms to facilitate sentence processing, despite the dependency of context processing and ambiguity resolution on cognitive processing. However, the question is what happens when the elderly adults' cognitive abilities decline further with increasing age. If language processing depends on cognitive processing, as is often assumed (e.g., Gunter et al., 2003; Huettig & Janse, 2016; Nieuwland & Van Berkum, 2006), and the processing of idioms and the use of context in sentence processing depend on cognitive processing, we hypothesize that elderly adults' sentence processing will decline over a period of three years. In addition, we hypothesize that the decline in elderly adults' sentence processing is predicted by their level of cognitive processing.

With respect to our first question about the ability to benefit from the fixed character of idioms in sentence processing, we expect to find effects in the N400. Specifically, we expect the amplitude of the N400 to be predicted by an interaction between Session and Idiomaticity. La Roi et al. (2020) found a decrease in N400 amplitude in response to idiomatic compared with literal sentences, showing that the retrieval of word meanings requires less cognitive effort in idiomatic compared with literal sentences, because the figurative meaning of an idiom can be retrieved from the mental lexicon as a whole (e.g., Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Sprenger et al., 2006; Titone & Connine, 1999). This idiom-related decrease in the N400 is expected to decrease or even disappear in the elderly adults tested three years later.

With respect to our second question about the ability to benefit from context information in sentence processing, we expect that the amplitude of the N400 will be predicted by an interaction between Session and Context. La Roi et al. (2020) found a reduction in N400 amplitude for literal and idiomatic sentences preceded by a predictive compared with a neutral context, indicating that elderly adults are able to use context information to predict and pre-activate upcoming words, thereby facilitating their retrieval upon encountering them. This context-related reduction in the N400 is expected to decrease or disappear in the elderly adults tested three years later. Furthermore, we expect that the amplitude of the P600 will be predicted by an interaction between Session, Context, and Idiomaticity. La Roi et al. (2020) reported an increase in the P600 for literal but not idiomatic sentences preceded by predictive contexts, showing that elderly adults relied on context information to construct the meaning of literal sentences. This increase in the P600 is predicted to decrease or even disappear in the elderly adults tested three years later.

2. Methods

2.1 Participants

For this study, we tested the elderly adults again who participated in the study of la Roi et al. (2020), with an interval of three years between the first (baseline) and second (follow-up) experimental session (mean: 34 months, range: 33 - 36 months). Participants who participated in the study of la Roi et al. (2020) were invited to participate in the present study as well. All participants were right-handed, monolingual native speakers of Dutch without any history of language or neurological disorders. Out of the 25 elderly adults whose data was included for analysis in the first experimental session, 19 volunteered to participate in the second experimental session. The data of three participants was rejected because of too many artifacts in the EEG recordings, leaving 16 elderly adults in the analysis. The elderly adults received a monetary reward for their participation. The study was approved by the Research Ethics Review Committee (CETO) of the University of Groningen. Table 1 summarizes the methodological and participant characteristics of the two experimental sessions.

Table 1. Methodological and participant characteristics of the first (baseline, la Roi et al., 2020) and second (follow-up, three years later) experimental session with the same participants.

		Baseline	**Follow-up**
Participants	*N* (total)	28	19
	N (included in analysis)	25	16
	Mean age	68	71
	Age range	61 - 74	64 - 77
	Sex	10 women, 15 men	8 women, 8 men
	Mean years of education	16.26	16.25
Design	Design type	Between-subjects + within-subjects	Within-subjects
	Factors	Group (Younger adults/Elderly adults), Context (Neutral/Predictive),	Session (Baseline/Follow-up), Context (Neutral/Predictive),

		Idiomaticity (Literal/Idiomatic), Correctness (Correct/Incorrect)	Idiomaticity (Literal/Idiomatic), Correctness (Correct/Incorrect)
Procedure	Offline cognitive tasks	Reading Span, Paired-Associates, Verbal Fluency	Verbal Fluency
EEG recording	Number of electrodes	62	62
	Sampling rate	500 Hz	500 Hz
ERP analysis	Amplifier	8-72 average reference Refa amplifier	8-72 average reference Refa amplifier
	Preprocessing software	Brain Vision Analyzer 2.1.1.	Brain Vision Analyzer 2.2
	Re-referenced to	Average of mastoids	Average of mastoids
	Filter settings	0.1 – 40 Hz bandpass	0.1 – 40 Hz bandpass
	Ocular correction	Gratton & Coles	Gratton & Coles
	% Interpolated channels	1.97	1.80
	Average % of rejected trials due to artefacts	12	8
	Epoch length	-150 – 1200 ms	-150 – 1200 ms
	Time windows	200-300 ms, 300-400 ms, 400-500 ms, 500-800 ms, 800-1100 ms	300-400 ms, 400-500 ms, 500-800 ms
	Analysis software	R (3.3.2)	R (3.6.2)

2.2 Materials and design

The study had a 2 x 2 x 2 x 2 within-subjects design. The factor Session (Baseline/Follow-up) specified whether mean voltages in EEG were recorded in the first or second experimental session, Context (Neutral/Predictive) indicated the type of context sentence, Idiomaticity (Literal/Idiomatic) described the type of test sentence, and Correctness (Correct/Incorrect) specified whether the target word was the expected word or an incorrect substitute, resulting in a semantic violation.

The experiment contained 192 experimental items and 60 filler items, all consisting of a context sentence followed by a test sentence. Test sentences were either literal or idiomatic and were taken from Rommers et al. (2013). Context sentences were either predictive or neutral. Table 2 illustrates the experimental items. Filler items were similar to experimental items, except that they did not contain semantic violations.

Table 2. Example stimuli for each experimental condition. Words in bold illustrate the correct and incorrect (marked by *) target words that are part of the factor Correctness.

Context	**Idiomaticity**	**Context sentence**	**Test sentence**
Predictive	Idiomatic	In de concertzaal zong iedereen luidkeels alle hits mee. *In the concert hall everybody sang along with all the hits at the top of their voice.*	Het uitzinnige publiek ging volledig uit zijn ***dak***/****blad*** bij de show gisteravond. *The hysterical crowd went completely out of their* ***roof/*leaf*** *at the show last night.*
Predictive	Literal	Bovenop het gebouw rondde een arbeider zijn werk af. *On top of the building a worker finished his work.*	De bouwvakker monteerde de schoorsteen op een ***dak***/****blad*** van een huis gisteravond. *The construction worker installed the chimney on the* ***roof/*leaf*** *of a house last night.*
Neutral	Idiomatic	Bij de ingang controleerden de portiers nauwkeurig alle toegangskaarten. *At the entrance the porters checked all entrance passes carefully.*	Het uitzinnige publiek ging volledig uit zijn ***dak***/****blad*** bij de show gisteravond. *The hysterical crowd went completely out of their* ***roof/*leaf*** *at the show last night.*
Neutral	Literal	Sinds vanochtend vroeg kampt de man met hevige rugpijn.	De bouwvakker monteerde de schoorsteen op een ***dak***/****blad*** van een huis gisteravond.

	Since early this morning the man struggled with severe back pain.	_The construction worker installed the chimney on the **roof/*leaf** of a house last night._

2.3 Procedure

The second experimental session took place in the same quiet cabin as the first experimental session. Each participant was instructed orally to carefully read the sentence pairs appearing on the screen and to minimize blinks and other movements while the sentences were presented. If a sentence pair was followed by a comprehension question, which only happened after filler items, participants could respond to this question with ‘yes’ or ‘no’ by pressing one of the two designated keys on the keyboard.

At the start of the experiment, participants completed ten practice trials to familiarize themselves with the experiment. A 1000 ms fixation cross indicated the start of a new trial. Next, the full context sentence was presented on the screen for 4500 ms, followed by another 1000 ms fixation cross. Subsequently, the test sentence with the target word was presented word by word, with each word being presented in black letters in the center of a white screen for 300 ms. Words were separated by a 300 ms blank screen. After each sentence pair, three asterisks (***) that were displayed for 3000 ms encouraged participants to blink.

The experiment consisted of seven blocks of 36 sentence pairs. In between blocks, participants could take a break for as long as they wanted. After completing the EEG experiment, participants carried out the same verbal fluency task as was used in la Roi et al. (2020), namely a shortened version of the verbal fluency task used by Spaan (2012). The verbal fluency task was divided in two parts. In the first part, participants were given five times one minute to list as many animals and kitchen utensils as possible, providing an index of *category fluency* (Benton, 1968) In the second part, participants had three times one minute to list as many words as possible starting with the letter ‘D’, ‘A’, and ‘T’, providing an index of *letter fluency* (Newcombe, 1969). Category and letter fluency were collected as separate measures, because they appear to play different roles in context-dependent processing of ambiguous expressions: whereas category fluency plays a role in predictive language processing, which affects the ability to benefit from context (Federmeier et al., 2010), letter fluency has been shown to be involved in ambiguity resolution (Lee & Federmeier, 2011). The entire test session at follow-up, including electrode application, instruction, and the verbal fluency tasks, took approximately two-and-a-half hours.

2.4 EEG recording

The EEG signal was recorded from 62 tin electrodes mounted in an Electro-Cap, placed on the head following the expanded International 10-20 system (Jasper, 1958). In addition, two separate electrodes were placed on the left and right mastoids. Horizontal eye movements (EOGh) and blinks (EOGvl) were monitored by two electrodes placed at the outer canthus of both eyes and by one electrode above and below the right eye, respectively. The ground electrode was attached to the sternum. Electrode impedances were kept below 20 kΩ. The continuous EEG and EOG recordings were amplified by an 8-72 Refa amplifier (TMSi, Oldenzaal, The Netherlands) with a 140 Hz cut-off filter, recording the data online with a sampling rate of 500 Hz.

2.5 ERP analysis

The details of the ERP analysis are summarized in Table 1 above. Below, we discuss the analysis in general terms and indicate how the analysis in this longitudinal study differs from the analysis of la Roi et al. (2020).

Following Rommers et al. (2013), the continuous EEG signal was segmented into epochs of 1350 ms, ranging from 150 ms prior to target word onset to 1200 ms after target word onset. We selected three time windows that had shown relevant effects in la Roi et al. (2020): 1) between 300-400 ms to test for early effects in the N400, 2) between 400-500 ms to test for late effects in the N400, and 3) between 500-800 ms to test for effects in the P600. Mean voltages in the three time windows were averaged over four regions of interest (ROIs), each containing nine electrodes that divided the scalp into four quadrants (see la Roi et al., 2020). We excluded voltages recorded in midline electrodes (front to back and left to right) to avoid spill-over effects between voltages measured in neighboring ROIs.

We used the R package *lmerTest* (version 3.1.1, Kuznetsova et al., 2017) to fit a linear mixed-effects regression model on the mean voltages in each time window. As the N400 and P600 effects are typically found in centroparietal and posterior brain regions (e.g., Kutas & Federmeier, 2011; Kutas & Hillyard, 1980), we performed the analysis on the data in the two posterior ROIs. As fixed effects, we included Session (Baseline/Follow-up), Context (Neutral/Predictive), Idiomaticity (Literal/Idiomatic) and Correctness (Correct/Incorrect). Furthermore, we used pairwise contrast coding in the package *emmeans* (version 1.4.5, Lenth, 2020) to carry out planned comparisons between the levels of the factors that interacted significantly in each final model.

Before fitting models to test our hypotheses, we fitted three simple models (one for each time window) to test for general longitudinal changes in the *N400 effect* and *P600 effect*. The N400 effect and P600 effect are computed by calculating the difference in mean voltages in response to correct compared with incorrect target words in the 300-400 ms and 400-500 ms time windows (N400 effect), and the 500-800 ms time window (P600 effect). After testing for general longitudinal changes in the N400 and P600 effect, we fitted a linear mixed-effect model on the mean voltages in each time window. Starting out with the full model including random intercepts for subjects and items, we selected the final model by applying a backward-fitting model selection procedure using chi-square tests as well as evaluation of Akaike's Information Criterion (Akaike, 1974) to compare models. Random slopes were added to the final models to account for unexplained variance introduced by subjects and items. Subsequently, final models were checked for a normal distribution of residuals. If model residuals were not normally distributed, we applied model criticism by removing residuals exceeding two and a half standard deviations from the mean (cf. Wieling et al., 2011).

3. Results

3.1 Behavioral results

During the first experimental session, which formed the baseline for our study, elderly adults correctly answered 91% (SD = 0.05) of the control questions. During the follow-up session, which took place three years later, elderly adults also correctly answered 91% (SD = 0.29) of the control questions. The high percentage of correct responses to the control questions shows that at both baseline and follow-up, elderly adults read the sentences carefully.

Analysis of participants' performance on the verbal fluency tasks showed no significant difference in elderly adults' performance on the category fluency task at the follow-up session (mean = 35.79) compared with the baseline session (mean = 35.52, $t(38.69) = -0.11$, $p = .910$).

Similarly, participants' performance on the letter fluency task was not significantly different at follow-up (mean = 45.05) compared with baseline (mean = 44.36, $t(40.04) = -0.20$, $p = .840$).

3.2 EEG results

Table 3 presents the final models fitted to mean voltages in the 300-400 ms, 400-500 ms, and 500-800 ms time windows. We first discuss the results for the N400 (time windows 300-400 ms and 400-500 ms) and P600 (time window 500-800 ms). Next, we discuss the effects of verbal fluency on the N400 and P600 amplitudes. The scalp distributions of the differences in mean voltages in the 300-400 ms and 400-500 ms time windows were consistent with the typical distribution of the N400, and in the 500-800 ms time window were consistent with the typical distribution of the P600 (see *Supplementary Materials*).

Table 3. Specification of the linear mixed-effects regression models fitted to the data of the 300-400 ms, 400-500 ms, and 500-800 ms time windows.

Time window	**Best fit model**
300-400 ms	µV ~ Session + Context * Idiomaticity * Correctness + Hemisphere + (1 + Idiomaticity|Subject) + (1 + Context|Target word)
400-500 ms	µV ~ Session + Context * Idiomaticity * Correctness + Hemisphere + (1 + Idiomaticity|Subject) + (1 + Context|Target word)
500-800 ms	µV ~ Session * Context * Idiomaticity * Correctness + Hemisphere + Category fluency + (1 + Idiomaticity|Subject) + (1 + Context|Target Word)

3.2.1 Longitudinal changes in the general N400 effect

To test for age effects in the N400 effect, we tested the interaction between Session and Correctness in two linear mixed-effects models that were fitted on the mean voltages measured in the two posterior ROIs in the 300-400 ms and 400-500 ms time windows. In addition to the interaction between Session and Correctness, the models included a fixed effect for Hemisphere, accounting for differences in mean voltages between the left and right hemisphere, and random intercepts for subjects and items. Mean voltages were averaged over neutral and predictive context sentences and over literal and idiomatic test sentences.

We found that in the 400-500 ms time window, but not in the 300-400 ms time window, the interaction between Session and Correctness significantly improved the model fit, compared with a model including Session and Correctness as main effects (300-400 ms: $\chi2(1)=1.09$, p = .296; Δ AIC = 0.91; 400-500 ms: $\chi2(1)=4.24$, p = .039; Δ AIC = 2.24). Furthermore, Session significantly interacted with Correctness in the 400-500 ms time window (β = 0.41, *SE* = 0.20, *t* = 2.06, *p* = .040), but not in the 300-400 ms (β = -0.19, *SE* = 0.19, *t* = -1.05, *p* = .296).

As the interaction between Session and Correctness did not reach significance in the 300-400 ms time window, nor did it improve model fit, we only performed planned comparisons on the interaction in the 400-500 ms time window. This showed that incorrect target words elicited significantly more negative mean voltages than correct target words in both the baseline (β = -1.20, *SE* = 0.20, z = -6.17, $p < .001$) and follow-up session (β = -0.79, *SE* = 0.21, *z* = -3.70, $p < .001$), but that the N400 effect was larger in the baseline than the follow-up session (Δ 1.20 µV at baseline compared with Δ 0.79 µV at follow-up). Direct comparison of mean voltages in the baseline compared with the follow-up session revealed that incorrect target words elicited significantly more negative mean voltages in the baseline compared with the follow-up session (β = -0.60, *SE* = 0.15, *z* = -4.06, $p < .001$), but correct target words did not (β = -0.19, *SE* = 0.15, *z* = -1.30, *p* = .193). This means that the larger N400 effect in the baseline compared with the follow-up session in the 400-500 ms time window was driven by elderly adults' brain responses to incorrect target words. Figure 1 shows the grand averages of the N400 effects in the 300-400 ms and 400-500 ms time windows (green) and of the P600 effect in the 500-800 ms time window (red).

Figure 2 shows the fitted mean voltages for the interaction between Session and Correctness in the 300-400 ms and 400-500 ms time windows.

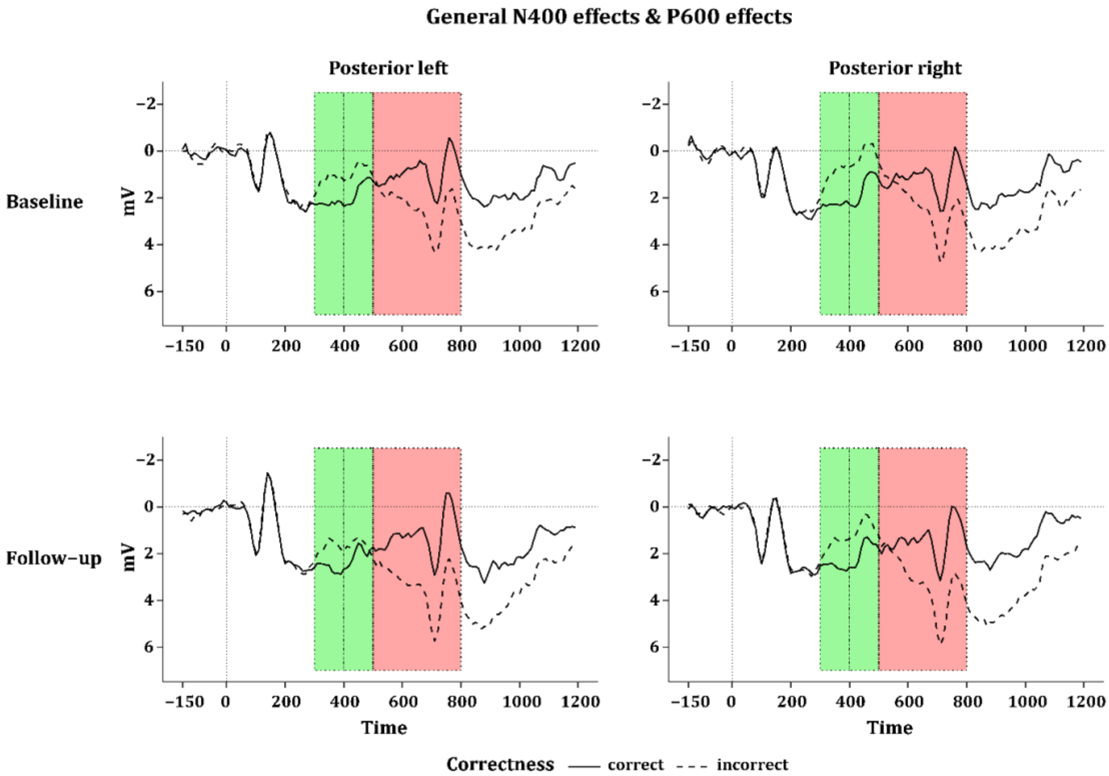

Figure 1. Grand averages of the N400 effects (green) and the P600 effect (red). The dotted vertical line in the green rectangle indicates the end of the 300-400 ms time window and the beginning of the 400-500 ms time window. Negative values are plotted upwards.

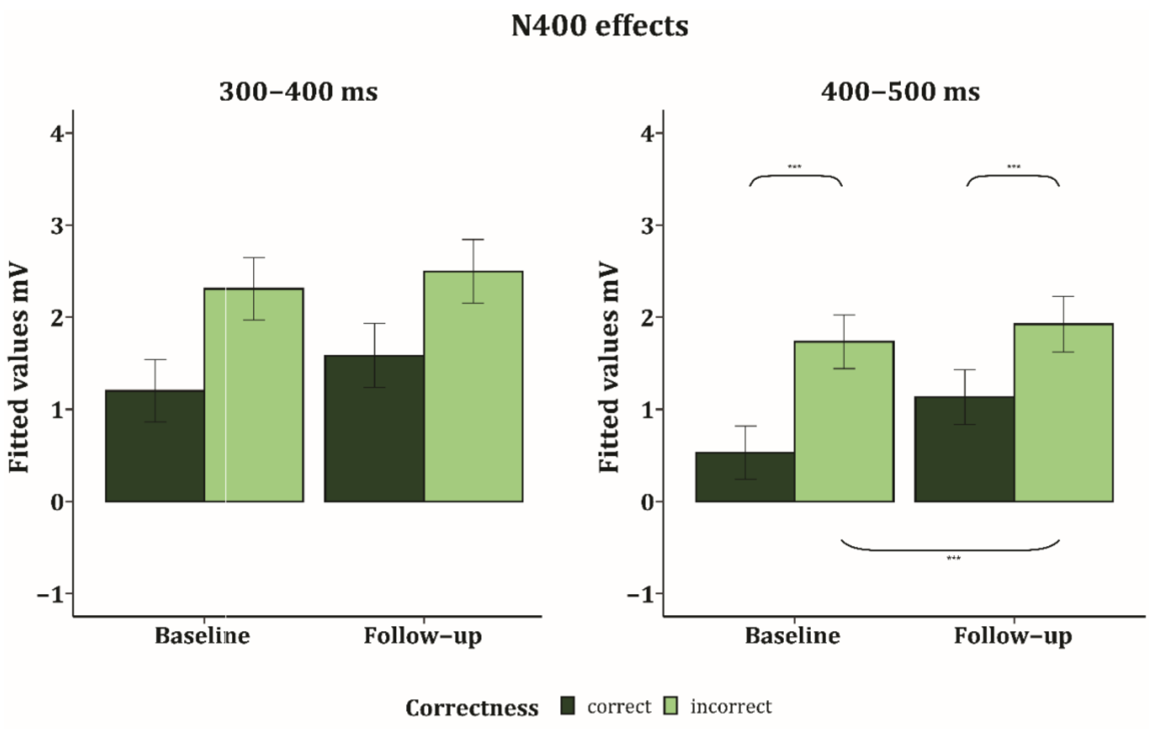

Figure 2. Mean voltages per condition illustrating the interaction between Session and Correctness as fitted to the data in the 300-400 ms and 400-500 ms time windows. Mean voltages are averaged over literal and idiomatic sentences, as well as over neutral and predictive contexts. Note: *** p < .001

3.2.2 Longitudinal effects of Context and Idiomaticity on the N400

The three-way interaction between Context, Idiomaticity, and Correctness included in the final models fitted on the mean voltages in the 300-400 ms and 400-500 ms time windows significantly improved model fit compared with models only including a two-way interaction between Context and Idiomaticity (300-400 ms: $\chi 2(3)=11.06$, $p = .011$; Δ AIC = 5.06; 400-500 ms: $\chi 2(3)=47.29$, $p < .001$; Δ AIC = 41.29), between Idiomaticity and Correctness (300-400 ms: $\chi 2(3)=10.22$, $p = .017$; Δ AIC = 4.21; 400-500 ms: $\chi 2(3)=8.29$, $p = .040$; Δ AIC = 2.29), or between Context and Correctness (300-400 ms: $\chi 2(3)=11.07$, $p = .011$; Δ AIC = 5.07; 400-500 ms: $\chi 2(3)=47.68$, $p < .001$; Δ AIC = 41.68). Although in the 300-400 ms time window the interaction between Context, Idiomaticity, and Correctness was significant ($\beta = 1.05$, $SE = 0.34$, $t = 3.12$, $p = .002$), it just missed significance in the 400-500 ms time window ($\beta = 0.70$, $SE = 0.36$, $t = 1.95$, $p = .051$). In the 300-400 ms time window mean voltages did not significantly differ between the left and right hemisphere ($\beta = -0.04$, $SE = 0.08$, $t = -0.47$, $p = .641$), but in the 400-500 ms time window mean voltages were significantly more negative in the right compared with the left hemisphere ($\beta = -0.39$, $SE = 0.09$, $t = -4.38$, $p < .001$). A complete overview of the final model's coefficients can be found in the *Supplementary Materials*.

The fact that there was no significant interaction between Session and Context or Idiomaticity in the 300-400 ms and 400-500 ms time windows means that we did not find a change over time in the N400 in response to literal and idiomatic sentences in context. As the present study focuses on longitudinal changes in elderly adults' brain responses, we do not provide a discussion of the N400 effects that remained the same over time. A detailed description of the effects of Context and Idiomaticity on the N400 in the follow-up session is provided in the *Supplementary Materials*.

3.2.3 Longitudinal changes in the general P600 effect

To test for age effects in the P600 effect, we tested the interaction between Session and Correctness in a linear mixed-effects model that was fitted on the mean voltages in the 500-800 ms time window. In addition to the interaction between Session and Correctness, the models included a fixed effect for Hemisphere, accounting for differences in mean voltages between the left and right hemisphere, and random intercepts for subjects and items. Mean voltages were averaged over neutral and predictive context sentences and over literal and idiomatic test sentences. We found that the interaction between Session and Correctness significantly improved model fit compared with a model only including Session and Correctness as main effects ($\chi2(1)=10.28$, p = .001; Δ AIC = 8.28). Furthermore, Session significantly interacted with Correctness in the 500-800 ms time window (β = 0.63, *SE* = 0.20, t = 3.21, p = .001).

Planned comparisons showed that incorrect target words elicited significantly more positive mean voltages than correct target words in both the baseline (β = 1.25, *SE* = 0.19, z = 6.68, $p < .001$) and the follow-up session (β = 1.88, *SE* = 0.21, z = 9.09, $p < .001$), but the P600 effect was larger in the follow-up session (Δ 1.88 µV at follow-up compared with Δ 1.25 µV at baseline). Furthermore, direct comparison of mean voltages elicited by correct and incorrect target words in the baseline compared with the follow-up session revealed that incorrect target words elicited significantly more positive mean voltages in the follow-up compared with the baseline session (β = -0.85, *SE* = 0.15, z = -5.80, $p < .001$), but correct target words did not (β = -0.22, *SE* = 0.15, z = -1.50, p = .133). Thus, the larger P600 in the follow-up session was driven by elderly adults' responses to incorrect target words. Grand averages of the P600 effect in the baseline and follow-up session are plotted in Figure 1 (red area). Figure 3 shows the fitted mean voltages for the interaction between Session and Correctness in the 500-800 ms time window.

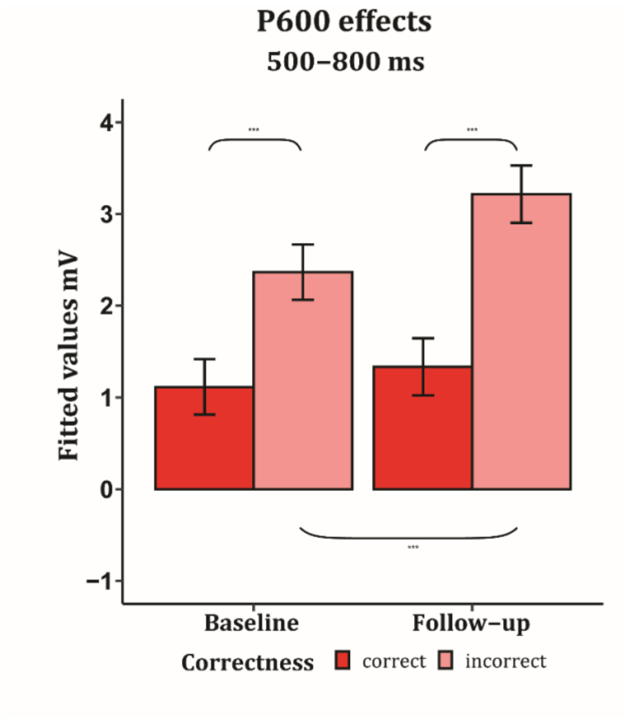

Figure 3. Mean voltages per condition illustrating the interaction between Session and Correctness as fitted to the data in the 500-800 ms time window. Mean voltages are averaged over literal and idiomatic sentences, as well as over neutral and predictive contexts. Note: *** p < .001

3.2.4 Longitudinal effects of Context and Idiomaticity on the P600

The final model fitted on the mean voltages in the 500-800 ms time window included a four-way interaction between Session, Context, Idiomaticity, and Correctness that significantly improved model fit compared with a model only containing an interaction between Session, Idiomaticity, and Correctness ($\chi2(7)=22.65$, $p = .002$; Δ AIC = 8.65), between Session, Context, and Correctness ($\chi2(7)=35.31$, $p < .001$; Δ AIC = 21.31), between Context, Idiomaticity, and Correctness ($\chi2(7)=20.92$, $p = .004$; Δ AIC = 6.92), or between Session, Context, and Idiomaticity ($\chi2(7)=36.41$, $p < .001$; Δ AIC = 22.41). Although the four-way interaction improved model fit, the interaction itself did not reach significance ($\beta = 0.68$, $SE = 0.72$, $t = 0.95$, $p = .343$). However, we found three underlying two-way interactions that were significant: between Session and Correctness ($\beta = 1.04$, $SE = 0.36$, $t = 2.90$, $p = .004$), between Idiomaticity and Correctness ($\beta = 0.72$, $SE = 0.32$, $t = 2.25$, $p = .025$), and between Context and Idiomaticity ($\beta = -1.06$, $SE = 0.32$, $t = -3.32$, $p < .001$). Mean voltages did not significantly differ between the left and right hemisphere ($\beta = 1.04$, $SE = 0.36$, $t = 2.90$, $p = .004$). A complete overview of the final model's coefficients can be found in the *Supplementary Materials*.

The fact that there was no significant interaction between Session and Context or Idiomaticity in the 500-800 ms time window means that we did not find a change over time in the P600 in response to literal and idiomatic sentences in context. As the present paper focuses on longitudinal changes in elderly adults' brain responses, we do not provide a discussion of the P600 effects that remained the same over time. However, a detailed description of the effects of Context and Idiomaticity on the P600 in the follow-up session is provided in the *Supplementary Materials*.

3.2.5 Summary of the results of the N400 and P600

We found that the N400 effect in elderly adults decreases over time, whereas the P600 effect increases over time. Regarding elderly adults' N400 and P600 in response to literal and idiomatic sentences preceded by a neutral or predictive context, we did not find significant differences between the baseline and follow-up session.

3.2.6 Verbal fluency

To investigate to what extent longitudinal changes in elderly adults' processing of literal and idiomatic sentences in context can be attributed to longitudinal changes in cognitive abilities, we tested the degree to which participants' performance on the letter fluency and category fluency tasks predicted mean amplitudes in the N400 and P600. Letter fluency did not significantly predict mean voltages in any of the time windows. However, adding category fluency to the model significantly improved model fit in the 500-800 ms time window ($\chi2(1)=10.42$, p = .001; Δ AIC = 8.42), but not in the 300-400 ms and 400-500 ms time windows. Although we expected higher performance on the fluency task to be related to a reduction in P600 amplitude, we found instead that higher performance on the category fluency task correlated with significantly more positive mean voltages ($\beta = 0.05$, $SE = 0.02$, $t = 2.98$, $p = .003$).

4. Discussion

This study investigated to what extent language processing abilities decline over time due to cognitive aging. The same elderly adults who participated in the study of la Roi et al. (2020) were tested three years later with the same ERP experiment, thereby creating a longitudinal design. We measured the amplitude of the N400 and P600 in response to literal and idiomatic sentences in predictive and neutral contexts and compared the results from our follow-up study to those from la Roi et al. (2020), which formed the baseline of our study. Our study is the first to investigate the N400 and P600 in response to elderly adults' language processing in a longitudinal design. We hypothesized that elderly adults' ability to benefit from predictive context information and the fixed character of idiomatic expressions decreases over time, because the use of context information in language processing and the resolution of ambiguity have been shown to depend on general cognitive abilities such as working memory capacity (Federmeier & Kutas, 2005) and inhibition skills (e.g., Lee & Federmeier, 2012), which decline with age (e.g., Bopp & Verhaeghen, 2005; Salthouse, 2009; Van Hooren et al., 2007).

Contrary to our predictions, we found that elderly adults' ability to benefit from linguistic context information and the fixed character of idiomatic expressions to facilitate language processing seems to be preserved over time. In addition, we found no longitudinal decline in elderly adults' cognitive processing as indexed by their performance on a verbal fluency task. However, we found that the general N400 effect decreases with age, whereas the P600 increases with age. Below, we first discuss our findings in the N400 and P600 that did not appear to change over time. Next we discuss the longitudinal changes that we found in the N400 effect and the P600 effect, respectively.

4.1 Ability to benefit from context and idioms in sentence processing

In the study that formed the baseline for our study, la Roi et al. (2020) found that elderly adults showed facilitated processing of idioms compared with literal sentences (N400). Furthermore, elderly adults showed facilitated retrieval of word meanings following predictive context sentences (N400). With respect to the computation of sentence meaning, la Roi et al. (2020) showed that elderly adults depend on context information for the word-by-word computation of literal sentence meaning, but not for the integration of word meanings in idiomatic sentences (P600). Based on these findings, la Roi et al. (2020) argued that the fixed character of idioms facilitates elderly adults' sentence processing to such a degree that they have sufficient cognitive resources available to integrate the figurative meaning of the idiom in the sentence without having to rely on additional context information.

Given the effects of age-related cognitive decline on elderly adults' language processing abilities, as discussed above, we predicted that elderly adults' ability to benefit from predictive context information and the fixed character of idioms to facilitate sentence processing would decrease or even disappear in the same elderly adults three years later. Furthermore, the age-related reduction in elderly adults' language processing abilities was expected to correlate with a corresponding decline in cognitive processing. In contrast to our predictions, we did not find a longitudinal decline in elderly adults' ability to benefit from context and idioms. Specifically, with respect to the retrieval of word meanings, measured in the N400, we found no decline over time in elderly adults' use of context to predict upcoming words and facilitate the retrieval of word meanings, or in their ability to benefit from the holistic retrieval of an idiom's figurative meaning. Regarding the subsequent integration of individual word meanings to compute a sentence meaning, measured in the P600, we found that the elderly adults benefitted from context information to the same extent as three years earlier. Thus, elderly adults' language processing abilities seem to be preserved with age, at least over the course of a three-year interval.

We also did not find a longitudinal decline in elderly adults' level of cognitive processing, as measured in the verbal fluency tasks. Our findings are consistent with other studies that included verbal fluency as a measure of cognitive processing and that found that elderly adults with high verbal fluency showed the same contextual facilitation effects in sentence processing as younger adults (e.g., Dave et al., 2018; DeLong et al., 2012; Federmeier et al., 2010, but see Wlotko & Federmeier, 2012).

However, we should be careful interpreting our null findings. We only looked at a time interval of three years, which means that potential effect sizes are small, in changes in ERP responses as well as in verbal fluency. In addition, only 16 participants were included in our follow-up study, due to a non-response of 6 of the 25 elderly participants who were included in the analyses of the earlier study of la Roi et al. (2020) and the exclusion of 3 further participants in the analyses of the present study because of too many artifacts in the EEG recordings. As a consequence, the power of the study to detect potential effects is limited.

If the apparently preserved language processing abilities of elderly adults can be confirmed by studies looking at larger time intervals or testing a larger sample of healthy elderly adults, a potential explanation of this finding could be that language experience compensates for the effects of age-related cognitive decline on language processing. In general, lexical knowledge, including idiom knowledge, increases with age (e.g., Brysbaert et al., 2016; Sprenger et al., 2019). The facilitated processing of idioms compared with literal sentences in particular may benefit from language experience, as the processing advantage of idioms has been shown to depend on the reader's familiarity with the idiom in question (Carrol & Littlemore, 2020) and idiom familiarity has been shown to increase with age (Sprenger et al., 2019; Carrol, 2023). Furthermore, especially adults with a high level of education are shown to have large vocabularies (e.g., Brysbaert et al., 2016). The fact that the elderly adults in our study all had a high level of education supports the assumption that they could have benefitted from their superior lexical knowledge in their sentence processing. Their superior lexical knowledge may compensate for potential age-related difficulties with the use of context resulting from deficient working memory capacity.

An additional explanation for the apparently preserved language processing abilities of elderly adults could be that the elderly adults participating in our study may have been too young to show a substantial decline in cognitive processing that affects language processing. Although longitudinal studies show that cognitive abilities start to decline between 60-65 years of age (e.g., Hatta et al., 2020), our findings are supported by a previous study by Vonk et al. (2020). They found that elderly adults' verbal fluency did not decline until age 73. In our follow-up session the elderly adults were between the ages of 64 and 77, with a mean age of only 71. Thus, to investigate the longitudinal effects of cognitive aging on language processing abilities, researchers may have to focus on elderly adults above 70 years old.

An interesting finding is that the effort involved in elderly adults' computation of sentence meaning was related to their performance on the category fluency task. Previous studies have linked category fluency in elderly adults to predictive processing in language (Federmeier et al., 2010) and to the ability to update the contents of working memory (Shao et al., 2014). Both prediction and updating benefit the computation of sentence meaning. When readers can predict upcoming words, it will be easier for them to retrieve the meanings of these words upon encountering them and add them to the contents of working memory in order to construct the sentence meaning. However, whereas we predicted that high verbal fluency scores would be correlated with a reduction of the P600, we found the opposite pattern. Nevertheless, findings are consistent with Ness and Meltzer-Asscher (2018), who found that the pre-updating of working memory contents with predicted words increased the amplitude of the P600 in younger readers. Furthermore, it is possible that the increased P600 was driven by elderly adults' response to incorrect target words, which increased in magnitude over time. Regardless of the direction of the effect, the fact that we found a link between elderly adults' computation of sentence meaning and category fluency suggests that elderly adults' successful sentence processing depends, at least partly, on general prediction and updating abilities.

4.2 Longitudinal decrease in the N400 effect

Despite the apparent absence of the predicted longitudinal changes in elderly adults' N400 and P600 in response to the processing of literal and idiomatic sentences in context, our study provides novel insights into cognitive aging by revealing general longitudinal changes in elderly adults' N400 and P600 effects. Specifically, we found that the N400 effect showed a general decrease over a period of three years, whereas the P600 effect actually increased over this period of three years.

The longitudinal decrease of the general N400 effect adds to the findings from cross-sectional studies showing a decrease in N400 amplitude in elderly compared with younger adults (e.g., Kutas & Iragui, 1998; Wlotko et al., 2010) by demonstrating how the general N400 effect declines within the same elderly individuals. Several explanations have been proposed for the age-related reduction in the N400, such as age-related changes in brain morphology that decrease the number of neural generators that fire simultaneously, or increased interindividual variability in the timing of the N400 that flattens the maximal peak of the component (Kutas & Iragui, 1998). Age-related reductions in the N400 effect could also be explained by the weakening of connections between nodes in the mental lexicon, which reduces the speed at which neural signals can travel through the language network (Burke et al., 1991), possibly resulting in a decrease in the amplitude of the N400 in response to linguistic input. Thus, the longitudinal reduction in the N400 may be the result of changes in the architecture of the aging brain that change the timing or strength of neural signals.

An interesting finding is that the longitudinal decrease in the N400 effect was driven by elderly adults' response to sentences containing a violation. This finding may be explained by a change in elderly adults' reading strategy that allows them to compensate for age-related deficiencies in processing speed or working memory capacity. Elderly adults have been shown to adopt a 'risky reading' strategy in which they skip words to compensate for age-related slowing in reading (Rayner et al., 2006). If elderly adults skip the words that introduce a semantic violation in the sentence, their brain response to this violation becomes less apparent. This may explain why elderly adults' brain response to semantic violation decreases longitudinally.

4.3 Longitudinal increase in the P600 effect

Besides shedding light on age-related changes in the N400 effect, our study adds to the literature on cognitive aging by finding an increase in the P600 effect over time in elderly adults. This is a novel and relevant finding, as cross-sectional evidence on age effects in the P600 is scarce. Kemmer et al. (2004) studied the P600 in response to syntactic violations in younger and elderly adults and found that the P600 remained unaffected by age. However, the incorrect target words in our study introduced semantic violations instead of syntactic violations. Although both the 'syntactic P600' and the 'semantic P600' have been argued to reflect a process of re-analysis (e.g., Kuperberg, 2007; Van Petten & Luka, 2012), studies also point to differences between the syntactic and semantic P600 (e.g., Regel et al., 2014). Regarding the effect of age on the syntactic and semantic P600, Leckey and Federmeier (2020) showed that whereas age changes the distribution of the syntactic P600, the distribution of the semantic P600 remains unaffected by age. However, Leckey and Federmeier (2020) do not explicitly discuss age effects on the amplitude of the syntactic and semantic P600. With respect to the amplitude of the P600, we found that elderly adults' P600 to semantically incorrect target words increased over time. Thus, our study not only shows how the P600 changes longitudinally as a consequence of aging, but also provides insight in the effects of age on the amplitude of the P600.

5. Conclusion

Our study is the first to use ERPs to investigate the longitudinal effects of cognitive aging on language processing. In our study, language processing was investigated through the processing of literal and idiomatic sentences in context. We did not find longitudinal changes in elderly adults' ability to benefit from context information and the fixed character of idioms to facilitate sentence processing. However, the general N400 effect was found to decrease longitudinally in elderly adults, whereas the P600 effect was found to increase longitudinally. We suggest that, if elderly adults' context-dependent language processing abilities are indeed preserved over time, this may be due to the protective effect of language experience or to a late onset of age-related decline in language processing, starting only after age 70.

Endnotes

[1] This article is part of a Festschrift in honor of Jack Hoeksema on the occasion of his retirement as a professor of Dutch linguistics. For this new phase of his life, we wish Jack the best of health and happiness and of course lots of interesting new words and idioms - they never get old.

[2] All supplementary materials of this study, including experimental materials, data and scripts, are available online at https://doi.org/10.17605/OSF.IO/7R26D

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723.

Ball, S. S., Marsh, J. T., Schubarth, G., Brown, W. S., & Strandburg, R. (1989). Longitudinal P300 latency changes in Alzheimer's disease. *Journals of Gerontology*, *44*(6), 195–200.

Beck, S. D., & Weber, A. (2020). Context and literality in idiom processing: Evidence from self-paced reading. *Journal of Psycholinguistic Research*, *49*(5), 837–863.

Benton, A. L. (1968). Differential behavioral effects in frontal lobe disease. *Neuropsychologia*, *6*(1), 53–60.

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *60*(5), 223–233.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2017). A Neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*, 1318–1352.

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*, 1116.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and older adults? *Journal of Memory and Language*, *30*(5), 542–579.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*(6), 668–683.

Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., & Cacciari, C. (2015). Electrophysiological correlates of idiom comprehension: Semantic composition does not follow lexical retrieval. In V. Pirrelli, C. Marazi, & M. Ferro (Eds.), *Word structure and word usage. Proceedings of the NetWordS final conference.* (Vol. 1347, pp. 98–101). http://ceur-ws.org/

Carrol, G. (2023). Old dogs and new tricks: Assessing idiom knowledge amongst native speakers of different ages. *Journal of Psycholinguistic Research*, 1-16.

Carrol, G., & Littlemore, J. (2020). Resolving figurative expressions during reading: The role of familiarity, transparency, and context. *Discourse Processes*, *57*(7), 609–626.

Clark, L. J., Gatz, M., Zheng, L., McCleary, C., MacK, W. J., & Chen, Y. L. (2009). Longitudinal verbal fluency in normal aging, preclinical, and prevalent alzheimer's disease. *American Journal of Alzheimer's Disease and Other Dementias*, *24*(6), 461–468.

Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition*, *25*(1), 57–71.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Dave, S., Brothers, T. A., Traxler, M. J., Ferreira, F., Henderson, J. M., & Swaab, T. Y. (2018). Electrophysiological evidence for preserved primacy of lexical prediction in aging. *Neuropsychologia*, *117*, 135–147.

DeDe, G., Kemtes, K., Caplan, D., & Waters, G. (2004). The relationship between age, verbal working memory, and language comprehension. *Psychology and Aging*, *19*(4), 601–616.

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition*, *135*, 103569.

DeLong, K. A., Groppe, D. M., Urbach, T. P., & Kutas, M. (2012). Thinking ahead or not? Natural aging and anticipation during reading. *Brain and Language*, *121*(3), 226–239.

Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, *42*(2), 133–141.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 379–390.

Giroud, N., Lemke, U., Reich, P., Bauer, J., Widmer, S., & Meyer, M. (2018). Are you surprised to hear this? Longitudinal spectral speech exposure in older compared to middle-aged normal hearing adults. *European Journal of Neuroscience*, *47*(1), 58–68.

Grindrod, C. M., & Raizen, A. L. (2019). Age-related changes in processing speed modulate context use during idiomatic ambiguity resolution. *Aging, Neuropsychology, and Cognition*, *26*(6), 842–864.

Gunter, T. C., Wagner, S., & Friederici, A. D. (2003). Working memory and lexical ambiguity resolution as revealed by ERPs: A difficult case for activation theories. *Journal of Cognitive Neuroscience*, *15*(5), 643–657.

Hatta, T., Hotta, C., Kato, K., Hatta, T., Hatta, J., Fujiwara, K., & Iwahara, A. (2020). Dissociation in age-related developmental trajectories between phonetic fluency and semantic fluency tests: Analysis of longitudinal data from the Yakumo study. *The American Journal of Psychology*, *133*(2), 197.

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93.

Jasper, H. H. (1958). The ten twenty system of the international federation. *Electroencephalography and Clinical Neurophysiology*, *10*, 371–375.

Kemmer, L., Coulson, S., De Ochoa, E., & Kutas, M. (2004). Syntactic processing with aging: An event-related potential study. *Psychophysiology*, *41*(3), 372–384.

Kemper, S., Thompson, M., & Marquis, J. (2001). Longitudinal change in language production : Effects of aging and dementia on grammatical complexity and prepositional content. *Psychology and Aging*, *16*(4), 600–614.

Kuperberg, G. R. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*(1), 23–49.

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117–129.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*(4427), 203–205.

Kutas, M., & Iragui, V. (1998). The N400 in a semantic categorization task across 6 decades. *Electroencephalography and Clinical Neurophysiology - Evoked Potentials*, *108*(5), 456–471.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1–26.

la Roi, A., Sprenger, S. A., & Hendriks, P. (2020). Event-related potentials reveal increased dependency on linguistic context due to cognitive aging. *Journal of Experimental Psychology: Learning Memory and Cognition*, *46*(7), 1226–1257.

Laurent, J. P., Denhières, G., Passerieux, C., Iakimova, G., & Hardy-Baylé, M. C. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research*, *1068*(1), 151–160.

Leckey, M., & Federmeier, K. D. (2020). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology*, *57*(7), e13351.

Lee, C.-L., & Federmeier, K. D. (2011). Differential age effects on lexical ambiguity resolution mechanisms. *Psychophysiology*, *48*(7), 960–972.

Lee, C.-L., & Federmeier, K. D. (2012). Ambiguity's aftermath: How age differences in resolving lexical ambiguity affect subsequent comprehension. *Neuropsychologia*, *50*(5), 869–879.

Lenth, R. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (1.4.5). https://cran.r-project.org/package=emmeans

Ness, T., & Meltzer-Asscher, A. (2018). Predictive preupdating and working memory capacity: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *30*(12), 1916–1938.

Newcombe, F. (1969). *Missile wounds of the brain: A study of psychological deficits*. Oxford University Press.

Nieuwland, M. S., & van Berkum, J. J. A. (2006). Individual differences and contextual bias in pronoun resolution: Evidence from ERPs. *Brain Research*, *1118*(1), 155–167.

Papaliagkas, V. T., Kimiskidis, V. K., Tsolaki, M. N., & Anogianakis, G. (2011). Cognitive event-related potentials: Longitudinal changes in mild cognitive impairment. *Clinical Neurophysiology*, *122*(7), 1322–1326.

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448–465.

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn’t it ironic - An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, *23*(2), 277–293.

Regel, S., Meyer, L., & Gunter, T. C. (2014). Distinguishing neurocognitive processes reflected by P600 effects: Evidence from ERPs and neural oscillations. *PLoS ONE*, *9*(5), e96840.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, *25*(5), 762–776.

Rönnlund, M., Nyberg, L., Bäckman, L., & Nilsson, L. G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging*, *20*(1), 3–18.

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514.

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772.

Shimamura, A. P., Jurica, P. J., Mangels, J. A., Gershberg, F. B., & Knight, R. T. (1995). Susceptibility to memory interference effects following frontal lobe damage: Findings from tests of paired-associate learning. *Journal of Cognitive Neuroscience*, *7*(2), 144–152.

Spaan, P. E. J. (2012). Veroudering en de ziekte van Alzheimer. *Tijdschrift Voor Neuropsychologie*, *7*(1), 3–15.

Spotorno, N., Cheylus, A., Van Der Henst, J. B., & Noveck, I. A. (2013). What’s behind a P600? Integration operations during irony processing. *PLoS ONE*, *8*(6), e66839.

Sprenger, S. A., la Roi, A., & van Rij, J. (2019). The development of idiom knowledge across the lifespan. *Frontiers in Communication*, *4*, 29.

Sprenger, S. A., Levelt, W. J. M., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, *54*(2), 161–184.

Titone, D. A., & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics*, *31*(12), 1655–1674.

van den Noort, M., Bosch, P., Haverkort, M., & Hugdahl, K. (2008). A standard computerized version of the reading span test in different languages. *European Journal of Psychological Assessment*, *24*(1), 35–42.

van Hooren, S. A. H., Valentijn, A. M., Bosma, H., Ponds, R. W. H. M., van Boxtel, M. P. J., & Jolles, J. (2007). Cognitive functioning in healthy older adults aged 64-81: A cohort study into the effects of age, sex, and education. *Aging, Neuropsychology, and Cognition*, *14*(1), 40–54.

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176–190.

Vonk, J. M., Higby, E., Nikolaev, A., Cahana-Amitay, D., Spiro, A., Albert, M. L., & Obler, L. K. (2020). Demographic effects on longitudinal semantic processing, working memory, and cognitive speed. *The Journal of Gerontology: Series B, 75*(9), 1850–1862.

Wieling, M., Nerbonne, J., & Baayen, R. H. (2011). Quantitative social dialectology: Explaining linguistic variation geographically and socially. *PLoS ONE*, *6*(9), e23613.

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology*, *49*(6), 770–785.

Wlotko, E. W., Lee, C.-L., & Federmeier, K. D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Linguistics and Language Compass*, *4*(8), 623–638.

Zempleni, M. Z., Renken, R., Hoeks, J. C. J., Hoogduin, J. M., & Stowe, L. A. (2007). Semantic ambiguity processing in sentence context: Evidence from event-related fMRI. *NeuroImage*, *34*(3), 1270–1279.