Event-Related Potentials Reveal Increased Dependency on Linguistic Context Due to Cognitive Aging

la Roi, Amélie; Sprenger, Simone A.; Hendriks, Petra

Published in:
Journal of experimental psychology. Learning, memory, and cognition

DOI:
10.1037/xlm0000777

ERPs Reveal Increased Dependency on Linguistic Context due to Cognitive Aging

Amélie la Roi, Simone Sprenger, & Petra Hendriks

University of Groningen

Author note

Amélie la Roi, Center for Language and Cognition Groningen, University of Groningen, the Netherlands and Neuroimaging Center, University of Groningen, the Netherlands

Simone Sprenger, Center for Language and Cognition Groningen, University of Groningen, the Netherlands and Neuroimaging Center, University of Groningen, the Netherlands

Petra Hendriks, Center for Language and Cognition Groningen, University of Groningen, the Netherlands

Correspondence concerning this article should be addressed to Amélie la Roi, Center for Language and Cognition Groningen, University of Groningen, Oude Kijk in 't Jatstraat 26, 9700 AS, Groningen, the Netherlands.

Contact: a.la.roi@rug.nl

This work was funded by a PhDs in the Humanities grant from the Netherlands Organisation for Scientific Research NWO (grant no. 322-75-008).

Abstract

Whereas executive functions are known to be closely tied to successful language processing in children and younger adults, less is known about how age-related decline in these functions affects language processing in elderly adults. Since the abilities to use linguistic context and resolve potential ambiguities such as between an idiom's figurative and literal meaning depend on executive functions, we investigated this issue by examining elderly adults' processing of idioms in context. We recorded event-related potentials of 25 younger (age 18-28) and 25 elderly adults (age 61-74) while they read literal sentences and sentences containing an idiom (e.g., the Dutch idiom *to walk against the lamp,* meaning 'to get caught'), each preceded by a neutral or predictive context sentence. Participants' use of context was hypothesized to relate to working memory capacity, while their ability to disambiguate idioms was hypothesized to depend on inhibition skills. Both groups showed facilitated processing for idioms compared to literal sentences and for sentences preceded by predictive compared to neutral contexts, indexed by a reduced N400. However, only elderly adults showed an increased P600 for literal but not idiomatic sentences preceded by a predictive context, suggesting that they rely on linguistic context when a sentence's meaning needs to be computed word by word, but not when a large part is retrieved from memory (as in idioms). Our findings suggest that in both younger and elderly adults processing literal sentences requires more cognitive effort than processing idiomatic sentences, and that cognitive aging affects language when processing is effortful.

Keywords: Aging, Executive functions, Idioms, Context processing, Event-related potentials

INTRODUCTION

By 2050, the percentage of people older than 60 will have almost doubled to 22% compared to 12% in 2015 (World Health Organization, 2015). The expected global increase in the elderly population urges researchers to investigate the physical, but also cognitive consequences of aging. Effects of cognitive aging are mostly attributed to a decline in the brain's executive functions: the higher-level cognitive skills used in planning and controlling thoughts and behavior, such as working memory (Baddeley, 2003) and inhibition (Miyake et al., 2000). At the same time, language processing is thought to be strongly dependent on executive functions such as working memory and inhibition (e.g., Daneman & Carpenter, 1980; Gunter, Wagner, & Friederici, 2003; Hoenig & Scheef, 2009; Van der Linden et al., 1999), suggesting that a decrease in executive functions should have a negative effect on the ability to process language. Therefore, the question is how decline in elderly adults' working memory capacity (e.g., Salthouse, Kausler, & Saults, 1988; Bopp & Verhaeghen, 2005) and inhibition skills (e.g., Hasher & Zacks, 1988; Hasher, Quig, & May, 1997) affects their ability to successfully process complex linguistic constructions.

In this study, we use ERPs to compare how younger and elderly adults use linguistic context, requiring working memory, to resolve ambiguity, involving inhibitory skills. We focus on the ambiguity in idioms, such as *to kick the bucket*, which have both a figurative meaning ('to die'), and a literal meaning that is constructed on the basis of the literal meanings of the idiom constituents ('to kick the pail'). Our study is the first to apply event-related potentials to investigate the effects of cognitive aging on the processing of idioms in context. Online idiom comprehension has been shown to be correlated with inhibition skills (Cacciari, Corrardini, & Ferlazo, 2018), possibly because the reader or listener needs to suppress the literal meanings of the idiom constituents in order to select the idiom's intended, figurative meaning. Supportive context information can facilitate the selection of an idiom's figurative meaning, but only if the reader or listener has sufficient working memory capacity to retain it for later use.

We found that both younger and elderly adults showed facilitated processing for idioms compared to literal sentences and that elderly adults only used context for the processing of literal sentences. This finding adds a novel insight to the literature by generalizing previous findings on the facilitated processing of idioms and proverbs (Conklin & Schmitt, 2008; Fernández, Shalom, Kliegl, & Siegman, 2014; Pawley & Syder, 1983; Rommers, Dijkstra, & Bastiaansen, 2013; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Swinney & Cutler, 1979; Underwood, Schmitt, & Galpin, 2004) to an elderly population.

Context-dependent language processing

To be able to interpret a single sentence within the context of a longer piece of text, the reader or listener needs to retain context information in working memory. If working memory decreases with age, language processing should therefore become more effortful. Previous research on elderly adults' use of context in literal language processing has yielded mixed results. On the one hand, offline studies investigating the use of context within sentence boundaries indicate that elderly adults benefit more from context (e.g., Pichora-Fuller, Schneider, & Daneman, 1995; Speranza, Daneman, & Schneider, 2000; Stine & Wingfield, 1994). On the other hand, studies applying online methods, such as event-related potentials, show that the use of context in language processing changes with age. However, those changes have been attributed to different causes: reduced working memory capacity (Federmeier & Kutas, 2005), context processing strategies (e.g., Federmeier, McLennan, Ochoa, and Kutas, 2002), or age-related slow-down of language processing (Dagerman, MacDonald, & Harm, 2006). These age-related changes even occur in simple tasks such as listening for comprehension (Federmeier, Van Petten, Schwartz, & Kutas, 2003) and especially reduce elderly adults' performance when context information has to be used across sentence boundaries, which increases processing demands (Federmeier et al., 2002; Light & Capps, 1986).

Idiom processing

To examine the processing of ambiguous language we study idioms, such as *to kick the bucket* (literally: 'to kick the pail', but figuratively: 'to die'). A special characteristic of idioms is that the literal meanings of an idiom's constituents, in this case 'to kick' and 'bucket', are irrelevant to deriving its figurative meaning (Weinreich, 1969). Early models of idiom processing claim that the figurative meaning of an idiom is accessed without activating the literal meanings of idiom constituents (Bobrow & Bell, 1973; Gibbs, 1980), whereas more recent models agree on the fact that these literal word meanings are also activated, at least in initial stages of processing (e.g., Cacciari & Tabossi, 1988; Giora, 1997; Sprenger, Levelt, & Kempen, 2006; Swinney & Cutler, 1979). Evidence for the activation of an idiom's components has been accumulated since (e.g., Snider & Arnon, 2012; Konopka & Bock, 2009; Peterson, Burgess, Dell, & Eberhard, 2001). The activation of an idiom's constituents during processing entails that to interpret the sentence *He kicked the bucket* as 'He died', the reader or listener has to switch from processing the constituents' literal meanings 'to kick' and 'pail' to processing the figurative meaning of the idiom as a whole. After this switch, word by word processing of the literal meanings of idiom constituents becomes irrelevant, and the idiom's figurative meaning is retrieved from the mental lexicon. The ambiguity between the idiom's figurative meaning and its compositional interpretation computed from the literal meanings of idiom constituents increases processing load (see Vulchanova, Milburn, Vulchanov, & Baggio, 2019 for a discussion on the role of compositional meaning in idiom processing).

Although switching between the literal meanings of idiom constituents and an idiom's figurative meaning adds a processing step and requires the inhibition of the constituents' literal meanings, idioms are suggested to reduce processing costs in younger adults (Conklin & Schmitt, 2008; Pawley & Syder, 1983; Siyanova-Chanturia, Conklin, & Smith, 2011; Swinney & Cutler, 1979; Underwood et al., 2004; but see Van Ginkel & Dijkstra, 2019 for a contrasting view). This could be due to the fact that an idiom's figurative meaning can be retrieved from the mental lexicon as a whole, once sufficient evidence has been accumulated.

The idea that idioms are stored as separate units in the mental lexicon is supported by frequency effects. In an eye-tracking study Carrol and Conklin (2019) showed that the reduced reading times for idioms compared to novel phrases was largely explained by the idioms' phrase frequency. In this respect, idiom processing resembles the processing of other types of multi-word expressions such as proverbs (Fernández et al., 2014) and formulaic expressions such as 'I don't know' or 'you don't have to worry' (Arnon & Cohen Priva, 2013; Arnon & Cohen Priva 2014; Arnon & Snider, 2010; Molinaro, Canal, Vespignani, Pesciarelli, & Cacciari, 2013; Pawley & Syder, 1983). Although the reader or listener activates the individual components of these expressions, processing times are reduced compared to regular compositional phrases due to the high frequency of co-occurrence.

A crucial difference between idioms and other highly frequent multi-word constructions is that the meanings of idiom constituents are often irrelevant to the meaning of the expression as a whole and therefore need to be suppressed. Here, the reader/listener needs to recruit executive functions (inhibition) to successfully interpret the idiom, making this process a function of age. With respect to idiom processing in context, we therefore predict that the elderly face two types of challenges: first they benefit less from context due to reduced working memory capacity, and second they find it more effortful to suppress irrelevant literal word meanings.

Processing ambiguity in idioms

The link between domain-general inhibition skills and the processing of language is supported by the finding that the ability to resolve ambiguity in language can be improved by training people's general ability to resolve conflicts in non-linguistic cognitive tasks (Novick et al., 2014). Furthermore, similar to the involvement of inhibitory control skills in the processing of lexical ambiguity (Gunter, Wagner, & Friederici., 2003; Mason & Just, 2007; Zempleni, Renken, Hoeks, Hoogduin, & Stowe, 2007; Hoenig & Scheef, 2009), the relation between inhibition and online idiom comprehension has been established in younger adults as well (Cacciari et al., 2018). In elderly adults, inhibition skills have been related to successful processing of ambiguity at the word level (Meyer & Federmeier, 2010), as well as ambiguity at the sentence level, such as irony (Gaudreau et al., 2013). Patients with Alzheimer's disease, whose executive functions decline rapidly, show a decreased ability to inhibit the literal meaning of idioms (Papagno, Luchelli, Muggia, and Rizzo, 2003).

In general, the processing of idioms relative to literal language has been shown to increase activation of the frontal brain regions (e.g., Zempleni, Haverkort, Renken, & Stowe, 2007), a brain area important for executive processes (West, 1996), such as the ability to suppress irrelevant information. However, since frontal brain regions are also the first and most affected by aging (e.g., Cabeza, Nyberg, & Park, 2005), the inhibition skills required to successfully process the ambiguity in idioms are expected to be affected in elderly adults. Thus, age-related decline in the executive functions needed to suppress the literal meanings of idiom constituents may require elderly adults to depend more heavily on linguistic context to select an idiom's figurative meaning.

Using ERPs to study context-dependent idiom processing

In this study, we use ERPs to investigate the time course of idiom processing. We chose ERPs instead of the eye-tracking methodology that related studies have applied previously (e.g., Carrol & Conklin, 2019; Fernández et al., 2014; Holsinger, 2013; Rayner, Reichle, Stroud, Williams, & Pollak, 2006; Von der Malsburg, Kliegl, & Vasisthth, 2015), because we followed the experimental set-up by Rommers et al. (2013). Moreover, ERPs allow the investigation of specific components that have been linked to different cognitive processes (e.g., semantic retrieval, semantic integration), whereas the eye-tracking methodology merely yields reading times that are not linked to specific stages of processing.

Expanding on the study by Rommers et al. (2013) on the processing of idiomatic sentences in younger adults, we investigate the extent to which these processes are affected by context and age. A well-known ERP component is the N400, which is a negativity peaking around 400 ms after the onset of a stimulus. As the N400's amplitude increases for unexpected relative to expected words, it is typically studied in response to a semantic violation. Since its initial discovery by Kutas and Hillyard (1980), researchers have argued that the N400 reflects semantic retrieval (Brouwer, Crocker, Venhuizen, & Hoeks, 2016), integration (Brown & Hagoort, 1993), or both (e.g., Baggio & Hagoort, 2011; Kutas & Federmeier, 2011). Although no consensus has been reached yet with respect to the actual process that it reflects, there is wide consensus that it is a direct function of the extent to which a target word is semantically expected.

Besides responding to semantic retrieval and integration, the N400 has been shown to be sensitive to working memory load (Gunter, Jackson, & Mulder, 1995) and context (Federmeier & Kutas, 1999; Kutas & Federmeier, 2000; Molinaro, et al., 2010). Based on a predictive context (also called a *high constraint* or *strongly constraining context*, e.g., Federmeier & Kutas, 2005), the reader or listener can pre-activate upcoming words, which facilitates their retrieval and integration in the sentence. This facilitation effect is indexed by a reduced N400. In this study we use the context sensitivity of the N400 to study how elderly adults and younger adults use context to facilitate the retrieval of an idiom's figurative meaning.

In addition to investigating context effects in the N400, we focus on the N400 to examine to what extent elderly adults are able to inhibit literal word meanings in idiom processing. Previous studies showed that the N400 can index facilitated processing of idioms compared to literal language (e.g., Canal, et al., 2015; Laurent et al., 2006; Paulmann et al., 2015; Rommers, et al., 2013; Strandburg et al., 1993; but see Proverbio et al., 2009, and Molinaro & Carreiras, 2010; Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2010; Siyanova-Chanturia, Conklin, Caffarra, Kaan, & Van

Heuven, 2017 for the possible involvement of the P300 in addition to the N400).

The literature is equivocal about whether idiom-related effects in the N400 reflect facilitated retrieval of idioms due to their formulaic character, or facilitated integration due to the irrelevance of literal word meanings for deriving an idiom's figurative meaning. Yet, a prerequisite for both processes is that the literal meanings of idiom constituents are inhibited. Thus, comparing younger and elderly adults' N400 following idioms and literal expressions enables us to study elderly adults' ability to suppress irrelevant meanings in language processing, and to relate this ability to their potentially reduced inhibition skills.

In addition to studying the N400, we study effects in the P600 to investigate how younger compared to elderly adults use context to integrate an idiom in a discourse. Compared to the retrieval of an idiom's figurative meaning, integrating an idiom's figurative meaning in a sentence involves and additional processing step: the contents of working memory must be updated and the meaning of the sentence as a whole must be constructed by combing the contexts of working memory with the meanings of upcoming words. The P600 was originally found in response to syntactic violations (Hagoort, Brown, & Groothusen, 1993), but has also been demonstrated in response to garden-path sentences (e.g., Osterhout, Holcomb, & Swinney, 1994), inferences (e.g., Burkhardt, 2006), irony (Regel, Gunter, & Friederici, 2011), and idioms (Rommers et al., 2013). Based on these findings, the P600 has been described as reflecting a general process of integration of the information that was retrieved in earlier stages (Brouwer and Hoeks, 2013; Brouwer et al., 2016). In this view, the P600 that Regel et al. (2011) found in response to ironic but not literal sentence could indicate that the integration of the ambiguous meaning of ironic sentences in a discourse requires additional effort. Similar to ironic sentences, sentences containing an idiom also have an ambiguous meaning. Therefore, a P600 following idioms but not literal expressions may index the additional processing effort involved in integrating an idiom's figurative meaning in a sentence. As a predictive context is likely to facilitate this integration process by pre-activating upcoming words, we expect the P600 to reflect the extent to which elderly adults are able to use context in idiom processing.

The present study

The present study investigates how age-related decline in executive functioning affects the ability to use context information in language processing. We focus on the processing of ambiguous language, in this case idioms, because context is essential to deriving the intended meaning of an ambiguous expression. We aim to answer our research question by posing two sub questions, namely: 1) Does elderly adults' decreased working memory capacity impair their ability to benefit from predictive context information to facilitate idiom processing? and 2) Does age-related decline in inhibition affect elderly adults' ability to suppress the literal meanings of idiom constituents? To investigate these questions, we expand the study by Rommers et al. (2013) and measure younger and elderly adults' brain potentials (factor *Age group*) in response to literal and idiomatic sentences (factor *Idiomaticity*) that are preceded by either a neutral or a predictive context sentence (factor *Context*). As we are interested in ERP components that are modulated in response to a violation, half of the experimental items contain the literal or idiomatic sentence in its correct form (*correct* condition), whereas the other half of the items contain a semantic violation (*incorrect* condition) (factor *Correctness*).

Given the results of previous studies showing a relation between executive functions and the processing and comprehension of language (e.g., Van der Linden et al., 1999; Gunter et al., 2003), our general hypothesis is that the successful processing of idioms in context requires executive functions. That is, we hypothesize that the reduced working memory capacity of elderly adults impairs their ability to retain context information that can be used to facilitate the retrieval and integration of an idiom's figurative meaning. In terms of ERP components, we predict that Age group interacts with Context: while in elderly adults predictive compared to neutral context sentences will not change N400 and P600 amplitude in response to literal, nor idiomatic sentences, in younger adults N400 and P600 amplitudes are expected to reduce following predictive context sentences. If supportive context information is especially helpful to resolve ambiguity in idiomatic sentences, Age group and Context will also interact with Idiomaticity. However, if supportive context information has the same facilitative effect on the processing of literal and idiomatic sentences Age group will only interact with Context.

Second, we hypothesize that elderly adults' reduced inhibition skills hamper the suppression of the literal meanings of idiom constituents. In terms of ERP components, we predict that Age group interacts with Idiomaticity: while in elderly adults the N400 will be similar in response to literal and idiomatic sentences, in younger the N400 is expected to reduce for idiomatic compared to literal sentences.

Third, we expect the N400 and P600 amplitudes in younger and elderly adults to correlate with offline measures of working memory, inhibition, and verbal fluency, and thus to reflect the involvement of executive functions in language processing.

METHODS

Participants

Twenty-eight healthy elderly adults were recruited from the *Senioren Academie Groningen*. Twenty-eight students of the University of Groningen and the Hanze University of Applied Sciences Groningen served as a control group. All were right-handed, monolingual native speakers of Dutch with no history of language or neurological disorders. The data of three elderly adults and three students was rejected because of too many artifacts. The data of the remaining twenty-five elderly adults (mean age = 68 years, range = 61-74 years, 15 men) and twenty-five students (mean age = 22 years, range = 18-28 years, 6 men) was included in the analysis. Elderly adults and students had an average of 16.26 and 17.73 years of education ($t(44.64) = -2.05$, $p = .047$), respectively. For their participation in the study elderly adults received a small gift and students received a gift voucher. The study was approved by the Research Ethics Committee Faculty of Arts (CETO), University of Groningen.

Materials and Design

The experiment had a 2 x 2 x 2 x 2 mixed design involving the between-subjects factor Age group (Younger adults/Elderly adults) and three within-subject factors. The factor Context (Neutral/Predictive) described the nature of the context sentence, Idiomaticity (Literal/Idiomatic) specified whether the test sentence was literal or idiomatic, and Correctness (Correct/Incorrect) indicated whether the idiom's last noun, the *critical word*, was the correct word in that position or not. If the critical word was incorrect, the test sentence contained a semantic violation. Table 1 summarizes the conditions.

The experimental materials were created by adding context sentences to the test sentences from Rommers et al. (2013). Rommers et al.'s stimuli consisted of single literal and idiomatic sentences. In their study, each idiomatic sentence was matched to a literal equivalent in which the same critical word was presented in a literal context. The critical word was never the last word of the sentence to avoid potential sentence wrap-up effects influencing the results.

Table 1. Overview of experimental stimuli. Words printed in bold illustrate the correct and incorrect critical word (marked by *) of the factor *Correctness* in the test sentence.

Context	Idiomaticity	Context sentence	Test sentence
Predictive	Idiomatic	Op de bovenverdieping betrapte de schoonmaakster een nieuwsgierige vrouw. *On the upper floor the cleaning lady caught an inquisitive woman.*	Bij het stiekeme afluisteren ging haar mobieltje af en viel ze door de **mand/*beer** door het geluid. *While secretly eavesdropping her cell phone rang and she fell through the* ***basket/*bear*** *because of the sound.*
Predictive	Literal	Op haar fiets had Linda niet alleen een boodschappenkrat. *On her bicycle Linda did not only have a shopping crate.*	Op weg naar het ziekenhuis nam ze veel fruit mee in een rieten **mand/*beer** met hengsels. *On her way to the hospital she brought a lot of fruit in a wicker* ***basket/*bear*** *with handles.*
Neutral	Idiomatic	Achter de deur zag de loodgieter een jonge vrouw. *Behind the door the plumber saw a young woman.*	Bij het stiekeme afluisteren ging haar mobieltje af en viel ze door de **mand/*beer** door het geluid. *While secretly eavesdropping her cell phone rang and she fell through the* ***basket/*bear*** *because of the sound.*
Neutral	Literal	Na haar werk bezocht Linda haar ernstig zieke vader. *After work Linda visited her seriously ill father.*	Op weg naar het ziekenhuis nam ze veel fruit mee in een rieten **mand/*beer** met hengsels. *On her way to the hospital she brought a lot of fruit in a wicker* ***basket/*bear*** *with handles.*

In our study, the context sentences preceding the test sentences were either neutral or predictive. The predictive context sentences provided the reader with information that could be used to predict the content of the following idiomatic or literal test sentence. The neutral context sentences did not contain any information predicting the content of the subsequent test sentence. To ensure that predictive context sentences had a higher predictive value than the neutral context sentences, a norming study was conducted. Participants read sentence pairs consisting of a test sentence preceded by either a predictive or a neutral context sentence and were asked to judge how well the second sentence matched the first sentence on a scale from 1 (the second sentence does not match at all with the first sentence) to 5 (the second sentence matches very well with the first sentence). Statistical analysis confirmed that mean ratings for test sentences preceded by a predictive context sentence (M=4.07, SD=1.02) were significantly higher than mean ratings for test sentences preceded by a neutral context sentence (M=2.92, SD=1.33) (t(2248)=-23.73, $p < .001$).

All context sentences had a length of nine words and contained a subject NP, a transitive verb with its corresponding direct or prepositional object, and a PP. Out of each item set, two sentence pairs appeared in one list, but in different conditions. The grey rows in Table 2 give an example of two sentence pairs that occurred in one list. The underlined word indicates the critical word that was used in the sentence. Although the two sentence pairs belong to the same item set, the two sentence pairs are completely different from each other. With this set-up we ensured that the same critical word, test sentence and context sentence occurred only once within a list, so that all trials presented to one participant were unique and effects in the ERP signal were not confounded by effects of item repetition.

In total, each list contained 192 experimental items and 60 fillers. Out of the 192 experimental item sets, six were created by the researchers to guarantee an equal division of items over the four lists. Filler items resembled test items but did not contain semantic violations.

Procedure

Participants were tested in a quiet cabin where they were seated comfortably on a chair at an approximate distance of 60 cm from a computer screen. Prior to the start of the experiment, the participant was told to carefully read the sentence pairs, some of which would be followed by a question, while minimizing blinks and other movements during presentation of the sentences. At the start of the experiment, the instructions were repeated on the screen.

The experiment began with a practice block of ten trials. All stimuli were presented in black letters (Tahoma font) on a white screen. Each trial started with a 1000 ms fixation cross. Subsequently, the context sentence was presented in full for 4500 ms, followed by a 1000 ms fixation cross. Next, the sentence containing the critical word was presented word by word in the center of the screen, each word being presented for 300 ms, separated by a 300 ms blank screen. A 3000 ms screen displaying three asterisks (***) concluded each sentence pair. Here, participants were encouraged to blink. Filler items were followed by a simple control question about the preceding sentence pair, to which participants had to respond 'yes' or 'no' by pressing a designated key on the keyboard. We only inserted questions after filler items to prevent the ERPs in response to experimental items from being influenced by key presses.

Each experimental list was divided into seven blocks of 36 sentence pairs, with each block being followed by a short break with a duration determined by the participants. After reading sentences under EEG recording, participants completed three offline cognitive tasks presented to them in counterbalanced order. Including the application of electrodes, instruction, and cognitive tasks, the test session took approximately three hours.

The *working memory* task was a Dutch computerized version of the reading span task (Van den Noort, Bosch, Haverkort, & Hugdahl, 2008) based on the original reading span task of Daneman and Carpenter (1980). In this task participants have to read aloud sentences and remember the sentence-final word. After a number of sentences, ranging from two to six, participants are asked to recall the final words of the preceding set of sentences.

Inhibition was measured with a Dutch version of the paired-associates test of Shimamura, Jurica, Mangels, Gershberg, and Knight (1995). This task asks participants to learn a list of word pairs. Next, participants are presented with the first words of a pair and have to produce the matching words. After the first recall session, the first word of each word pair is matched to a new word. Participants have to learn this new set of word pairs and subsequently recall the second word of the new pair upon presentation of the word pair's first word. To produce the correct word during this second recall phase, participants have to inhibit the matching word that was learned in the first list of word pairs. Therefore, we used the number of correctly recalled items as a measure of participants' inhibition skills.

We conducted a shortened version of the *verbal fluency* task modeled on Spaan (2012) in which participants had five times one minute to list as many animals, kitchen utensils and words starting with the letter 'D', 'A', and 'T' as possible.

EEG recording

EEG was recorded through 62 electrodes in an Electro-Cap, positioned on the head according to the International 10-20 system (Jasper, 1958). Two separate electrodes were attached to the left and right mastoid. The ground electrode was attached to the sternum. Horizontal eye movements were monitored through two electrodes placed at the outer canthus of both eyes. One electrode above the right eye and one below the eye's infraorbital ridge monitored blinks. Electric impedance was kept below 20 kΩ. An 8-72 average reference Refa amplifier (TMSi, Oldenzaal, The Netherlands) amplified the EEG and EOG recordings with a 140 Hz cut-off filter, digitizing the data online with a sampling frequency of 500 Hz.

ERP analysis

Brain Vision Analyzer 2.1.1 (Brainproducts Gilching, Germany) software was used for the preprocessing of the EEG data. The EEG recordings were re-referenced offline to the average of the two mastoids and filtered with a bandpass filter of 0.1-40 Hz. Channels containing a high amount of noise were replaced by the average voltage of the channels surrounding it (1.97% of all channels in total). Whenever two adjoining channels both showed excessive noise, one of the two was removed from the dataset and the other was interpolated based on the voltage of the surrounding channels. This was done for only two channels in the complete data set. If necessary, ocular artefacts were corrected using the Gratton & Coles method (Gratton, Coles, & Donchin, 1983), which calculates a propagation factor between the voltage of an artefact recorded at the eyes and at the different electrode sites. Subsequently, a proportion of the voltage recorded at the eyes that corresponds to the calculated propagation factor is subtracted from the voltage measured at the various electrode sites, thereby reducing or removing ocular artefacts from the EEG signal (Luck, 2014, pp. 212-213). Trials contaminated by other artifacts, such as muscle movements, were detected and removed by means of semi-automatic artifact rejection. Subjects of whom more than 25% of the data were rejected were excluded from analysis. Based on this criterion, six participants (three younger adults) were excluded and the data of the remaining 50 participants (25 younger adults) was used for the analysis. The average percentage of rejected trials of the final set was 12% for both younger and elderly participants.

All statistical analyses were carried out in R (R Core Team, 2016; version 3.3.2). Given that one of our aims was to replicate the findings by Rommers et al. (2013), we selected the same time windows for analysis. Therefore, epochs of 1350 ms, from 150 ms prior until 1200 ms after critical word onset, were used for segmentation of the continuous EEG signal, containing three time windows to be used for analysis: 300-400 ms, 400-500 ms, and 500-800 ms after critical word onset. Furthermore, we added two extra time windows ranging from 200-300 ms and 800-1100 ms following critical word onset. These time windows were selected to capture processes related to the early recognition of idioms and the potential delay of effects in elderly adults, respectively. Following Rommers et al. (2013), within each time window mean voltages were averaged over four regions of interest (ROIs), each containing nine electrodes (see Figure 1). Voltages measured in electrodes on the midlines (front to back and left to right) were not included in the analysis to ensure sufficient distance between neighboring ROIs.

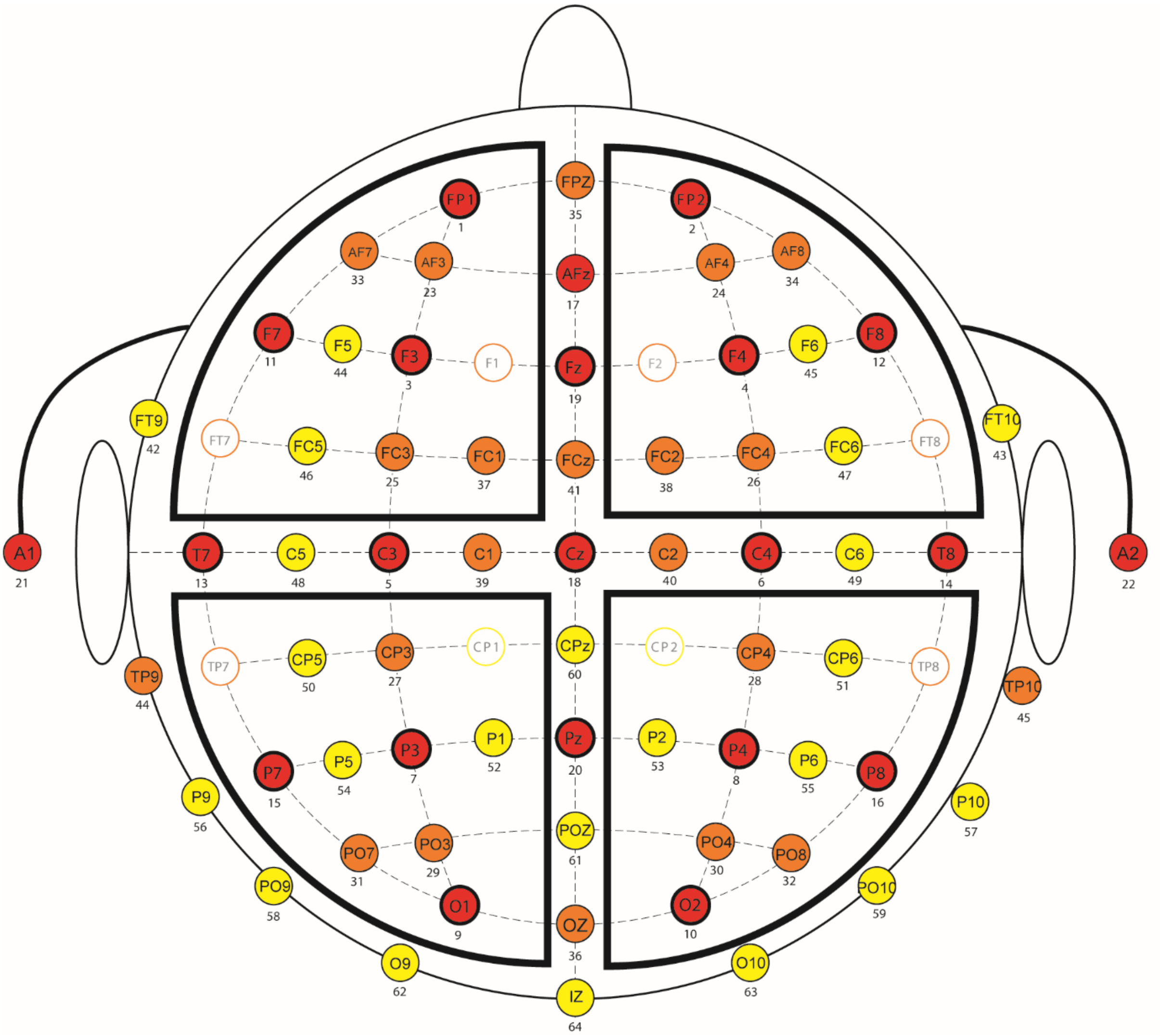

Figure 1. Schematic representation of the electrode lay out. For statistical analysis, mean voltages were averaged over electrodes within four regions of interest (anterior left, anterior right, posterior left, posterior right), indicated by thick solid lines. Each ROI included nine electrodes.

The EEG data were analyzed by linear mixed effects regression modeling. For each of the five time windows a model was fitted using the R package *lme4* (Bates, Maechler, Bolker, & Walker, 2015). Using pairwise contrast coding in the package *lsmeans* (Lenth, 2016), we calculated pairwise contrasts between the levels of the fixed effects specified in each final model. Starting out with a basic model only including random intercepts for subjects and items, we tested which of the fixed effect factors Age group (Younger adults/Elderly adults), Context (Neutral/Predictive), Idiomaticity (Literal/Idiomatic), Condition (Correct/Incorrect), Anteriority (Anterior/Posterior), and Hemisphere (Left/Right) significantly increased the amount of explained variance in the ERPs. After this procedure we added random slopes to account for differences due to subjects and items to reduce the amount of unexplained variance in the model (cf. Wieling, Nerbonne, & Baayen, 2011). For example, one participant may be more familiar with a specific idiom than another or some context sentences may more strongly predict the critical word in the test sentence than other context sentences.

The contribution of each factor was evaluated by comparing the model with and without the factor in question on the basis of Akaike's information criterion (AIC). Whenever the addition of a factor decreased a model's AIC by more than 2, the factor in question was considered to significantly contribute to the fit of the model to the data and was therefore kept in the model (cf. Akaike, 1974; Burnham & Anderson, 2002:170; Wieling, Montemagni, Nerbonne, & Baayen, 2014). The model with the best fit was checked for a normal distribution of residuals. If model residuals were not normally distributed due to outliers in the tails of the distribution, model criticism was applied by removing residuals that deviated by more than two and a half standard deviations from the mean (see Wieling et al., 2011). The removal of outliers improved the distribution of residuals without changing the pattern of effects in the final models.

RESULTS

Behavioral results

The control questions were answered correctly by 96% (sd = .02) and 91% (sd = .05) of the younger adults and elderly adults, respectively, indicating that both groups read the sentences attentively.

Table 2 shows the mean group scores on each of the cognitive tasks. Statistical tests confirmed that elderly adults scored significantly lower than younger adults on the reading span task ($t(52.99) = -2.34$, $p = .023$) and the paired-associates task ($U = 74$, $p < .001$), but not on the verbal fluency task ($t(46.23) = -0.34$, $p = .738$).

Table 2. Younger and elderly adults' mean scores on the reading span task, the paired-associates task, and the verbal fluency task.

	Younger adults			Elderly adults		
Task	***m***	***sd***	***range***	***m***	***sd***	***range***
Reading span task	61	10.4	43 - 83	55	9.1	37 - 74
Paired-associates task	35	6.7	23 - 47	25	5.6	12 - 33
Verbal fluency task	84	12.4	67 - 113	83	15.1	58 - 116

EEG results

We investigated whether younger adults and elderly adults differ in their ability to suppress the literal meanings of idiom constituents and in how far linguistic context is used to facilitate this process. Following Rommers et al. (2013), we measured mean voltages in a 300-400 ms and a 400-500 ms time window to study effects in the N400. Furthermore, we added a time window (200-300 ms) to capture early effects in the N400 that are due to the fast recognition of idioms. Besides examining the N400, we measured mean voltages in a 500-800 ms time window to investigate effects in the P600. In addition, we added a later time window (800-1100 ms) to detect delayed effects in elderly adults' processing of idioms in context. In the following, we will discuss our findings per time window, first for the N400, followed by those for the P600.

N400

Table 3 provides an overview of the final models fitted to the 200-300 ms, 300-400 ms, and 400-500 ms time windows. Figures 2A and 3A present younger adults' and elderly adults' grand average ERPs in response to correct versus incorrect critical words in literal and idiomatic sentences. As can be seen from Figures 2B and 3B, we found a negativity between 200-500 ms following the onset of the critical word that was maximal in posterior regions, which corresponds to the typical distribution of the N400 as demonstrated in previous studies (e.g., Kutas & Hillyard, 1980; Kutas & Federmeier, 2011; Rommers et al., 2013).

An overview of the N400 effects in single electrodes can be found in Figures A1-A4 in Appendix I. The complete set of the final models' coefficients can be found in Table A2 in Appendix II. Here, we focus on the significant differences that are directly related to our predictions.

Table 3. Specification of the linear mixed effects regression models fitted to the data of the 200-300ms, 300-400 ms, and 400-500 ms time windows.

Time window	Best fit model
200-300 ms	µV ~ Age group * Idiomaticity * Correctness + Context + Anteriority + Hemisphere + Verbal fluency + (1 + Idiomaticity|Subject) + (1 + Context|Critical Word)
300-400 ms	µV ~ Age group * Idiomaticity * Correctness + Context + Anteriority + Hemisphere + Verbal fluency + (1 + Idiomaticity|Subject) + (1 + Context|Critical Word)
400-500 ms	µV ~ Age group * Idiomaticity * Correctness + Context + Anteriority + Hemisphere + (1 + Idiomaticity|Subject) + (1 + Context|Critical Word)

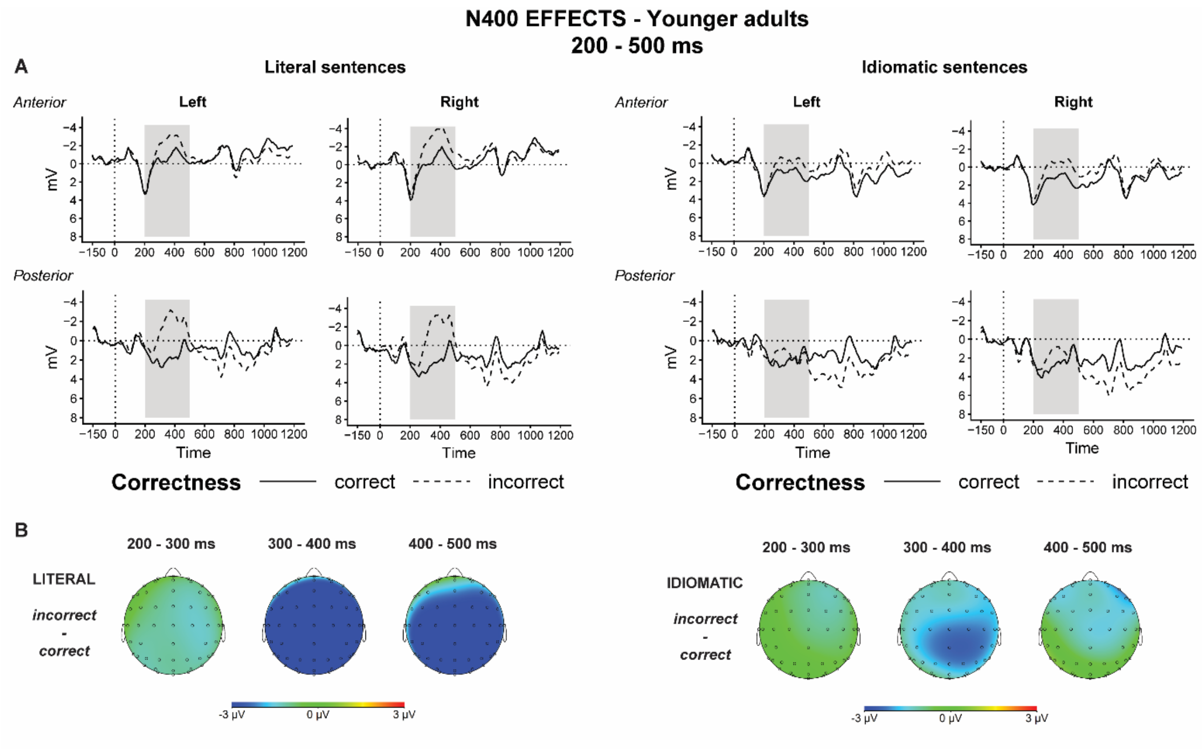

Figure 2. A) Younger adults' grand average ERPs evoked by critical words in literal and idiomatic sentences, in the four ROIs (anterior left, anterior right, posterior left, posterior right) that were used in the statistical analysis for the 200-300 ms, 300-400 ms, and 400-500 ms time windows. Solid and dashed lines represent mean voltages in response to correct and incorrect critical words, respectively. Grey areas indicate the selected time window. Mean voltages are averaged over the two context conditions. B) Younger adults' scalp distributions of the difference waves for literal and idiomatic sentences, plotted per time window. Difference waves were computed by subtracting mean voltages in the correct condition from mean voltages in the incorrect condition.

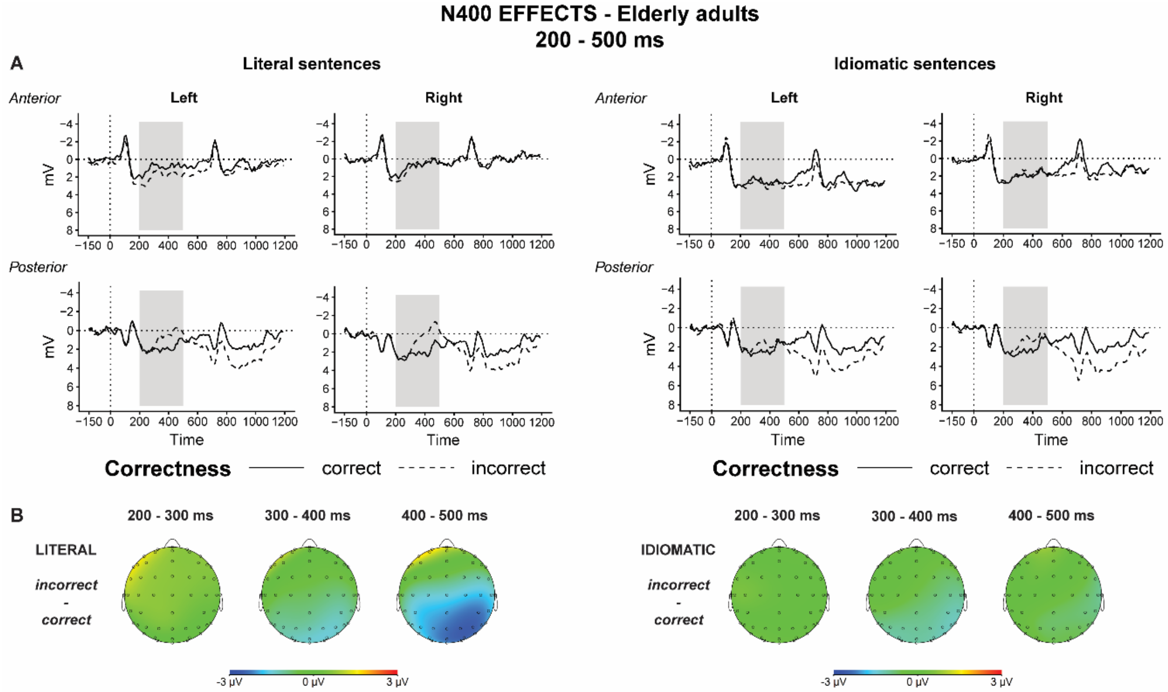

Figure 3. A) Elderly adults' grand average ERPs evoked by critical words in literal and idiomatic sentences, in the four ROIs (anterior left, anterior right, posterior left, posterior right) that were used in the statistical analysis for the 200-300 ms, 300-400 ms, and 400-500 ms time windows. Solid and dashed lines represent mean voltages in response to correct and incorrect critical words, respectively. Grey areas indicate the selected time window. Mean voltages are averaged over the two context conditions. B) Elderly adults' scalp distributions of the difference waves for literal and idiomatic sentences, plotted per time window. Difference waves were computed by subtracting mean voltages in the correct condition from mean voltages in the incorrect condition.

In line with our predictions, Context significantly predicted mean voltages in the 300-400 ms (β = 0.40, *SE* = 0.13, *t* = 2.94) and 400-500 ms time window (β = 0.37, *SE* = 0.15, *t* = 2.52), with voltages being significantly less negative for test sentences preceded by a predictive compared to a neutral context sentence. However, Context did not interact significantly with Age group.

Also consistent with our predictions, Age group and Idiomaticity significantly interacted in all three time windows (200-300 ms: β = -0.71, *SE* = 0.29, *t* = -2.44; 300-400 ms: β = -2.32, *SE* = 0.32, *t* = -7.33; 400-500 ms: β = -1.49, *SE* = 0.33, *t* = -4.54), as part of a significant three-way interaction with Correctness. Since adding the factor Context to this three-way interaction did not significantly improve model fit, Context was only kept as a main effect. Consequently, the planned comparisons and plots presented below will exclude the factor Context, and mean voltages are averaged over the neutral and predictive context condition.

The fact that Age group, Idiomaticity and Correctness significantly interacted with each other indicates that the response to violations depends on the specific conditions. We therefore examined the average voltages per time window per condition. Figure 4 presents the fitted mean voltages in the different conditions for each time window.

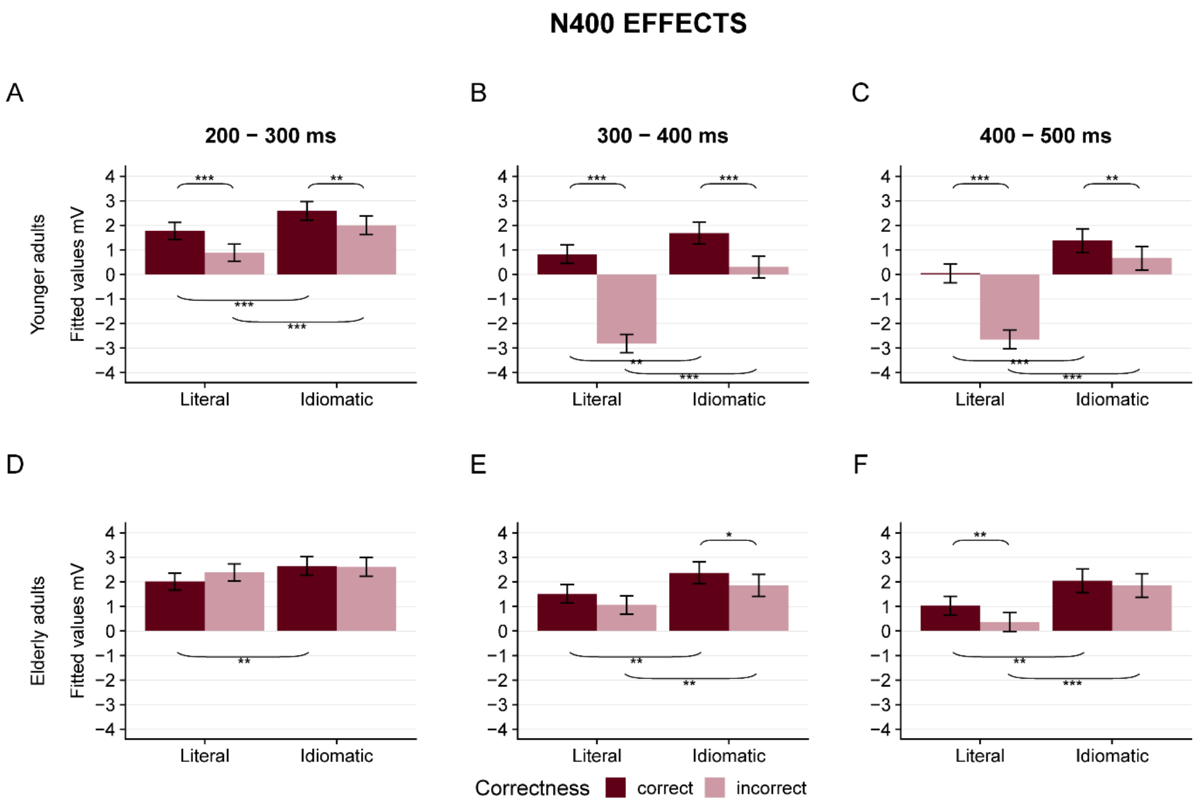

Figure 4. Mean voltages per condition illustrating the interaction Age group * Idiomaticity * Correctness as fitted to the data in the 200-300 ms, 300-400 ms, and 400-500 ms time windows. Each plot shows mean voltages for literal versus idiomatic sentences in response to correct versus incorrect critical words. The top row presents mean voltages per time window for younger adults (A, B, C). The bottom row presents mean voltages per time window for elderly adults (D, E, F).

Table 4 provides an overview of the planned comparisons testing differences in mean voltages as a function of Idiomaticity and Correctness in younger and elderly adults. In line with our predictions, younger adults' mean voltages were significantly more negative in literal compared to idiomatic sentences in all three time windows (Figure 4, panel A-C). Furthermore, incorrect critical words elicited larger N400s than correct critical words in both literal and idiomatic sentences.

Contrary to our predictions, elderly adults also showed significantly more negative mean voltages in response to literal compared to idiomatic sentences in both correct and incorrect conditions, with the exception of the 200-300 ms time window (Figure 4, panel D-F). Furthermore, incorrect critical words in literal sentences only elicited significantly more negative mean voltages than correct critical words in the 400-500 ms time window. In idiomatic sentences this difference started earlier, in the 300-400 ms time window, but disappeared in the subsequent time window.

Summary N400 results

In contrast with our predictions, Context had a similar influence on younger and elderly adults' processing of the test sentences, with mean voltages being significantly less negative when test sentences were preceded by a predictive compared to a neutral context. However, Context did not have a significantly different effect on literal versus idiomatic sentences. Both younger and elderly adults showed facilitated processing for idiomatic compared to literal sentences, although the effect started later in elderly adults.

Table 4. Planned comparisons between the levels of the significant interaction between Age group, Idiomaticity, and Correctness. Significant p-values are printed in bold.

Age group	Contrast		200-300 ms				300-400 ms				400-500 ms			
			β	SE	*z*-value	*p*	β	SE	*z*-value	*p*	β	SE	*z*-value	*p*
Younger adults	Literal	Incorrect - Correct	-0.89	0.21	-4.33	**< .001**	-3.64	0.24	-15.13	**<. 001**	-2.70	0.25	-10.66	**< .001**
	Idiomatic	Incorrect - Correct	-0.59	0.21	-2.86	**.004**	-1.38	0.24	-5.70	**< .001**	-0.72	0.25	-2.86	**.004**
	Correct	Literal - Idiomatic	-0.82	0.23	-3.52	**< .001**	-0.85	0.28	-3.05	**.002**	-1.34	0.33	-4.01	**< .001**
	Incorrect	Literal - Idiomatic	-1.11	0.23	-4.80	**< .001**	-3.11	0.28	-11.10	**< .001**	-3.30	0.33	-9.88	**< .001**
Elderly adults	Literal	Incorrect - Correct	0.37	0.21	1.80	.072	-0.45	0.24	-1.86	.062	-0.66	0.25	-2.62	**.009**
	Idiomatic	Incorrect - Correct	-0.04	0.21	-0.19	.847	-0.51	0.24	-2.13	**.033**	-0.19	0.25	-0.74	.459
	Correct	Literal - Idiomatic	-0.64	0.23	-2.76	**.006**	-0.86	0.28	-3.08	**.002**	-1.02	0.33	-3.05	**.002**
	Incorrect	Literal - Idiomatic	-0.23	0.23	-0.98	.326	-0.80	0.28	-2.85	**.004**	-1.49	0.33	-4.48	**< .001**

P600

Table 5 lists the final models that were fitted to the 500-800 ms and 800-1100 ms time windows. Figures 5A and 6A present younger adults' and elderly adults' grand average ERPs and scalp distributions in response to literal and idiomatic sentences preceded by a neutral or predictive context sentence in the 500-800 ms time window. Figures 5B and 6B show that the positivity between 500-800 ms following critical word onset was maximal in posterior regions, consistent with the distribution of the P600 components in previous studies (e.g., Hagoort et al., 1993; Osterhout et al., 1994; Regel et al., 2011; Rommers et al., 2013). An overview of the P600 effects in single electrodes can be found in Figures A5-A10 in Appendix III. Table A2 in Appendix IV contains the full set of coefficients for the final models fitted on the data of the 500-800 ms and 800-1100 ms time windows.

In line with our predictions, there was a significant interaction between Age group, Context, and Idiomaticity in the 500-800 time window ($\beta = -0.81$, $SE = 0.31$, $t = -2.60$). Although in general sentences including an incorrect critical word elicited more positive mean voltages than sentences including a correct critical word ($\beta = 0.73$, $SE = 0.20$, $t = 3.72$), adding Correctness to the three-way interaction did not significantly improve model fit. Therefore, Correctness was only kept in the model as a main effect. As a result, in subsequent analyses on the interaction effect in the 500-800 ms time window, mean voltages were averaged over correct and incorrect conditions.

Table 5. Specification of linear mixed effects regression models for the 500-800 ms time window and the 800-1100 time window.

Time window	Best fit model
500-800	μV ~ Age group * Context * Idiomaticity + Correctness + Anteriority + (1+Idiomaticity|Subject) + (1+Context|Critical Word)
800-1100	μV ~ Age group * Context + Idiomaticity + Correctness + Anteriority + Verbal fluency + (1+Idiomaticity|Subject) + (1+Context|Critical Word)

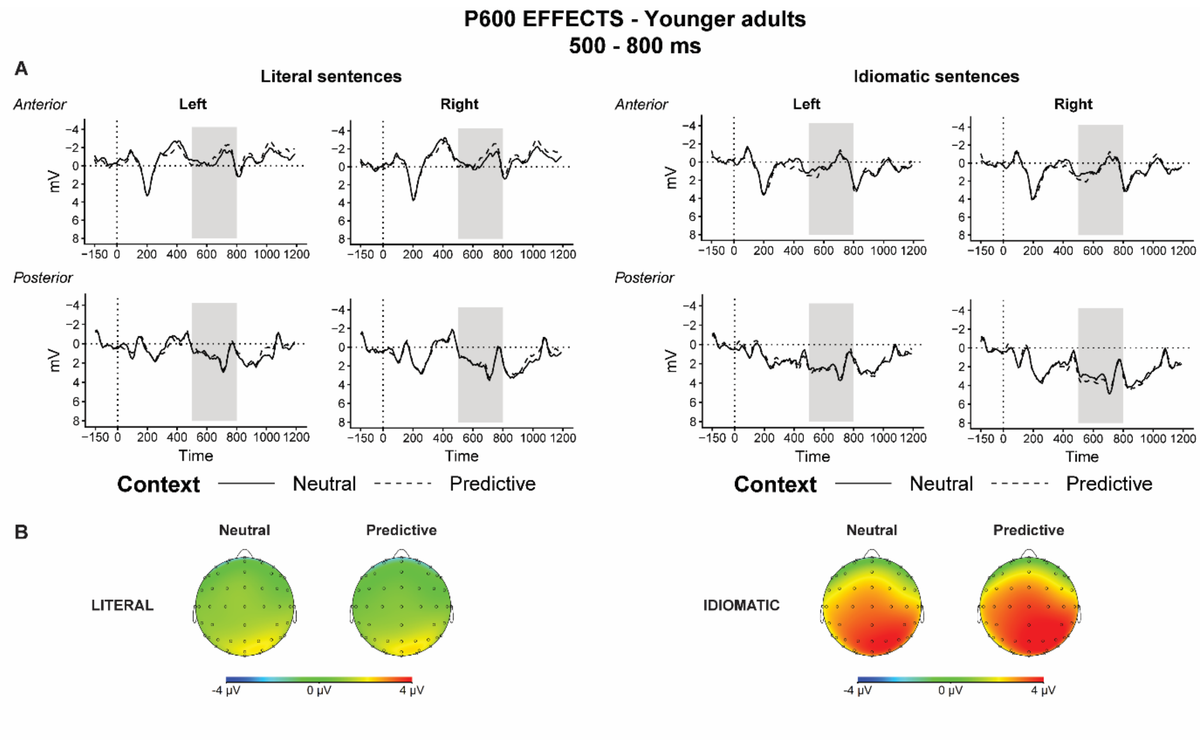

Figure 5. A) Younger adults' grand average ERPs evoked by critical words in literal and idiomatic sentences, in the four ROIs that were used in the statistical analysis for the 500-800 ms time window. Mean voltages are averaged over correct and incorrect conditions. Solid lines represent mean voltages evoked by test sentences preceded by a neutral context, dashed lines represent mean voltages evoked by test sentences preceded by a predictive context. Grey areas indicate the selected time window. B) Younger adults' scalp distributions of the context effects in literal and idiomatic sentences in the 500-800 ms time window. Mean voltages are averaged over correct and incorrect conditions.

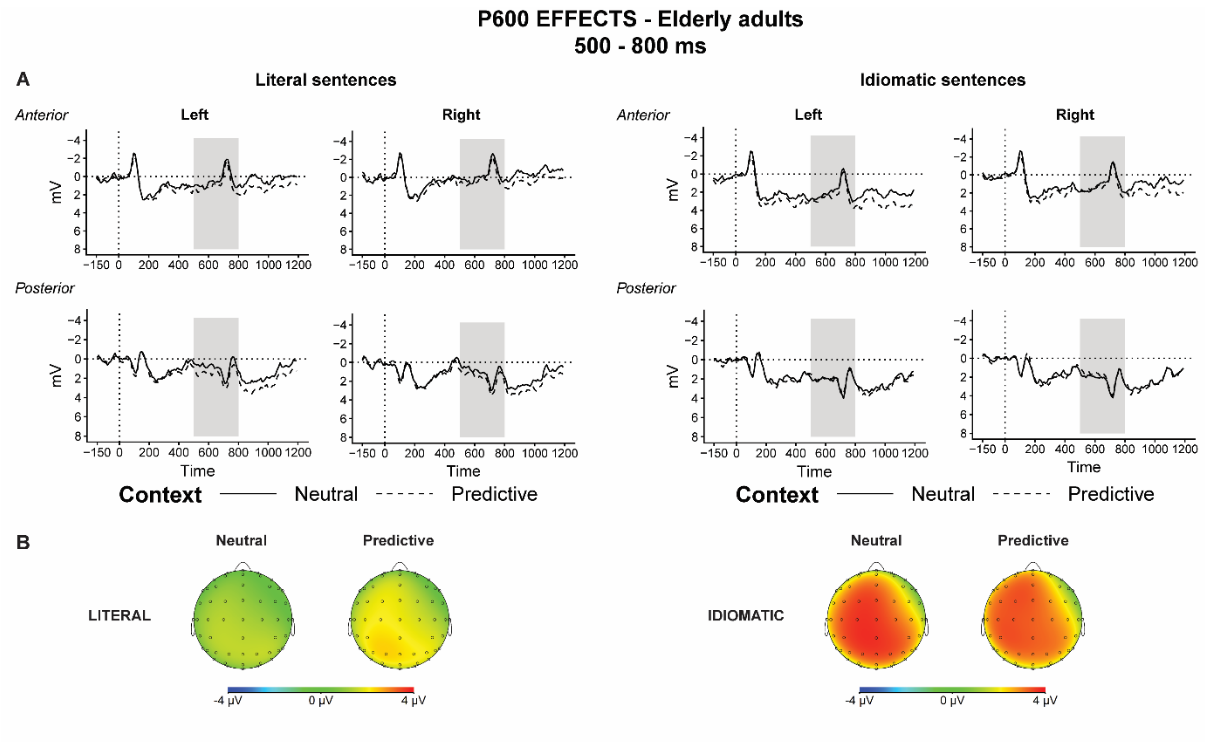

Figure 6. A) Elderly adults' grand average ERPs evoked by critical words in literal and idiomatic sentences, in the four ROIs that were used in the statistical analysis for the 500-800 ms time window. Mean voltages are averaged over correct and incorrect conditions. Solid lines represent mean voltages evoked by test sentences preceded by a neutral context, dashed lines represent mean voltages evoked by test sentences preceded by a predictive context. Grey areas indicate the selected time window. B) Elderly adults' scalp distributions of the context effects in literal and idiomatic sentences in the 500-800 ms time window. Mean voltages are averaged over correct and incorrect conditions.

Consistent with our predictions, Age group also significantly interacted with Context ($\beta = 0.98$, $SE = 0.17$, $t = 5.91$) in the 800-1100 ms time window. However, adding Idiomaticity and Correctness to the interaction between Age group and Context did not significantly improve model fit in this late time window. Therefore, in the 800-1100 ms time window the difference between literal and idiomatic sentences was only analyzed as a main effect, indicating that idiomatic sentences elicited significantly more positive mean voltages than literal sentences ($\beta = 1.38$, $SE = 0.20$, $t = 6.90$). Furthermore, mean voltages were significantly more positive in response to incorrect compared to correct critical words ($\beta = 0.84$, $SE = 0.20$, $t = 4.26$).

As Idiomaticity and Correctness did not significantly add to the interaction between Age group and Context, further analysis on this interaction was performed on the combined voltages of literal and idiomatic sentences and of correct and incorrect sentences. Figures 7 and 8 present the grand average ERPs and scalp distribution for the interaction between Age group and Context in the 800-1100 ms time window. Although mean voltages in literal and idiomatic sentences did not significantly differ as a function of Age group and Context, separate plots are made for literal and idiomatic sentences to facilitate comparison to the results in the 500-800ms time window.

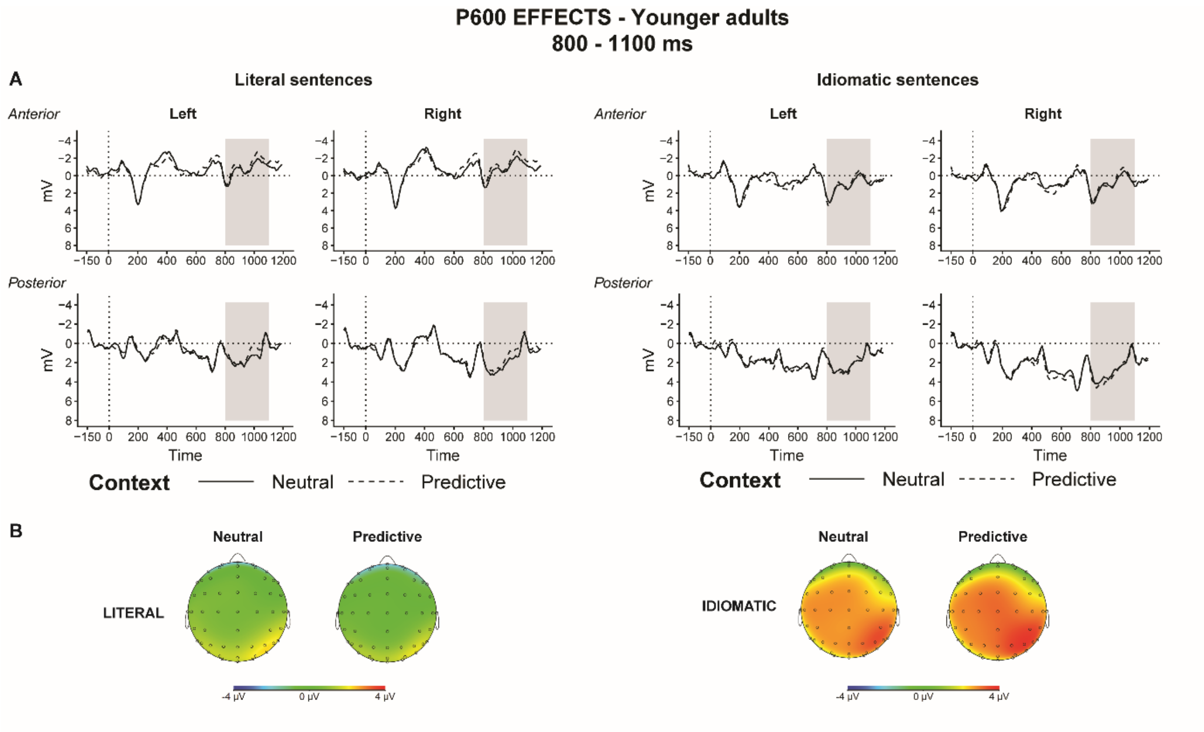

Figure 7. A) Younger adults' grand average ERPs evoked by critical words in literal and idiomatic sentences, in the four ROIs that were used in the statistical analysis for the 800-1100 ms time window. Mean voltages are averaged over correct and incorrect conditions. Solid lines represent mean voltages evoked by test sentences preceded by a neutral context, dashed lines represent mean voltages evoked by test sentences preceded by a predictive context. Grey areas indicate the selected time window. B) Younger adults' scalp distributions of the context effects in literal and idiomatic sentences in the 800-1100 ms time window. Mean voltages are averaged over correct and incorrect conditions.

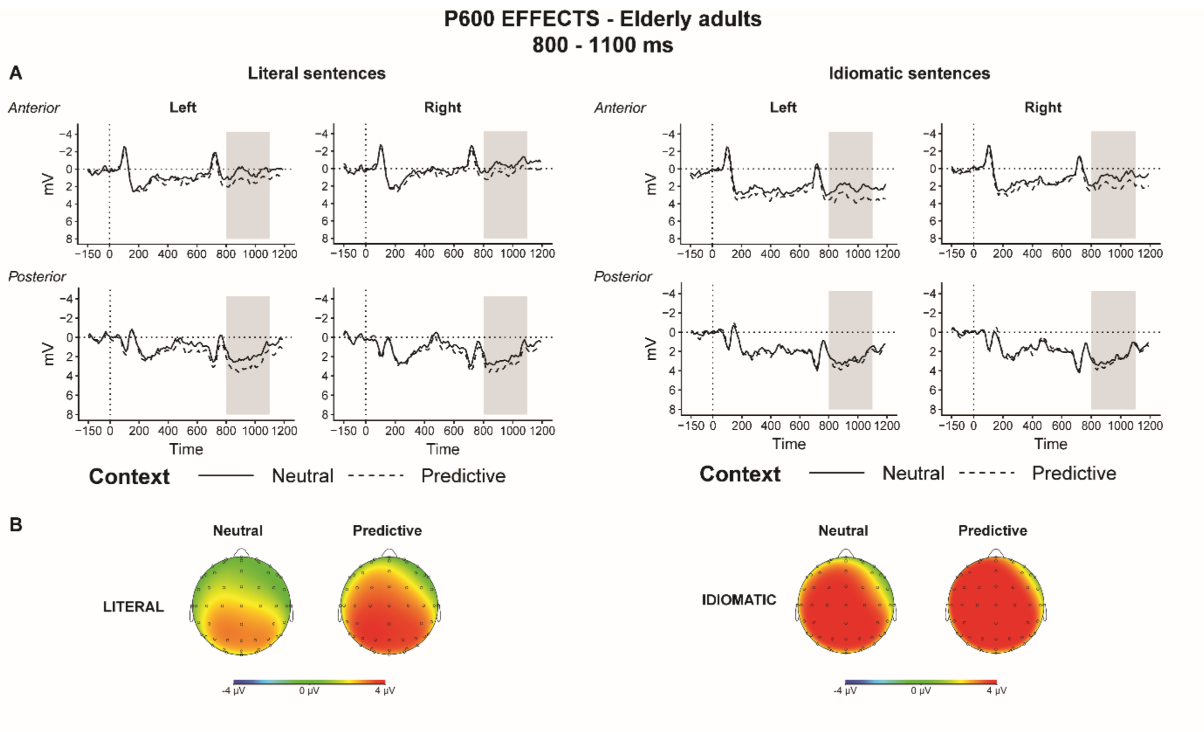

Figure 8. A) Elderly adults' grand average ERPs evoked by critical words in literal and idiomatic sentences, in the four ROIs that were used in the statistical analysis for the 800-1100 ms time window. Mean voltages are averaged over correct and incorrect conditions. Solid lines represent mean voltages evoked by test sentences preceded by a neutral context, dashed lines represent mean voltages evoked by test sentences preceded by a predictive context. Grey areas indicate the selected time window. B) Elderly adults' scalp distributions of the context effects in literal and idiomatic sentences in the 800-1100 ms time window. Mean voltages are averaged over correct and incorrect conditions.

Figure 9 shows the fitted mean voltages in the different conditions included in the interaction effects in the 500-800 ms and 800-1100 ms time window. Table 6 provides an overview of the planned comparisons contrasting younger and elderly adults' mean voltages in literal versus idiomatic sentences as a function of the type of context sentence, averaged over correct and incorrect conditions.

In both groups, idiomatic sentences showed an increased positivity between 500-800 ms after the onset of the critical word (panel A and B). Contrary to our predictions, younger adults' mean voltages were unaffected by the type of preceding context sentences in the 500-800 as well as the 800-1100 time window (panel A and C). In contrast, in the 500-800 ms time window, elderly adults' responses were significantly more positive when preceded by a predictive compared to a neutral context sentence in literal, but not idiomatic sentences (panel B). Although we could not perform statistical tests on the difference between elderly adults' processing of correct and incorrect literal sentences preceded by a predictive context because Correctness did not interact with Context in the 500-800 ms time window, the fact that mean voltages in response to literal sentences preceded by a predictive context were more positive in incorrect (1.55 mV) than correct sentences (0.58 mV) suggests that the context effect in elderly adults' literal sentence processing was mostly carried by the incorrect condition.

In the 800-1100 ms time window, we contrasted mean voltages in younger and elderly adults in response to sentences preceded by a neutral or predictive context. Here, mean voltages were averaged over literal and idiomatic sentences and over correct and incorrect conditions. Elderly adults' significantly increased positivity following predictive sentences remained significant, but context did not differentially affect literal and idiomatic sentences (panel C).

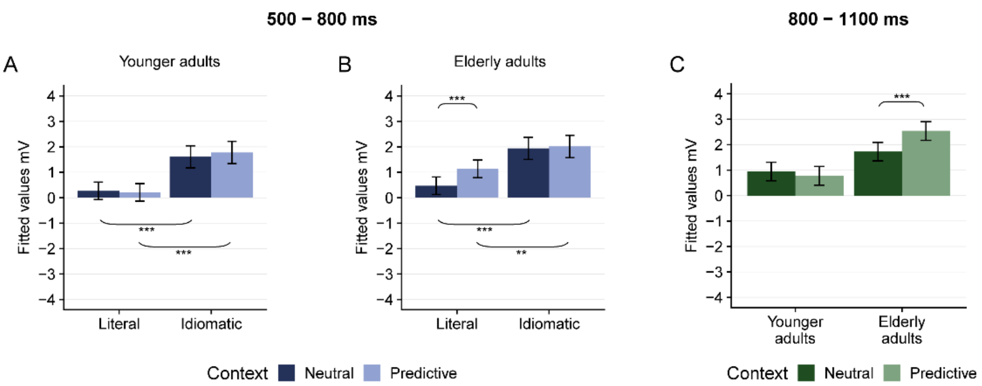

Figure 9. Mean voltages per condition illustrating the interaction Age group * Context * Idiomaticity as fitted to the data in the 500-800 ms time window for younger (A) and elderly adults (B). Plot A and B show mean voltages for literal versus idiomatic sentences preceded by a neutral versus a predictive context sentence. Mean voltages in A and B are averaged over correct and incorrect sentences, because Correctness did not significantly interact with Age group, Context, and Idiomaticity in the 500-800 ms time window. Panel C presents mean voltages in the 800-1100 ms time window in response to sentences preceded by a neutral versus a predictive context in younger and elderly adults. Mean voltages in panel C are averaged over literal and idiomatic sentences and across correct and incorrect sentences, because Idiomaticity and Correctness did not significantly interact with Group and Context in the 800-1100 ms time window.

Table 6. Planned comparisons between the levels of the significant interactions between Age group, Context, and Idiomaticity in the 500-800 ms time window and between Group and Context in the 800-1100 ms time window.

Age group		Contrast	500-800 ms β	SE	*z*-value	*p*
Younger adults	Neutral	Literal - Idiomatic	-1.34	0.30	-4.43	**< .001**
	Predictive	Literal - Idiomatic	-1.56	0.30	-5.19	**< .001**
	Literal	Predictive - Neutral	-0.05	0.19	-0.29	.771
	Idiomatic	Predictive - Neutral	0.17	0.20	0.87	.385
Elderly adults	Neutral	Literal - Idiomatic	-1.47	0.30	-4.87	**< .001**
	Predictive	Literal - Idiomatic	-0.88	0.30	-2.92	**.004**
	Literal	Predictive - Neutral	0.66	0.19	3.42	**< .001**
	Idiomatic	Predictive - Neutral	0.08	0.19	0.40	.692

Age group	Contrast	800-1100 ms β	SE	*z*-value	*p*
Younger adults	Predictive - Neutral	-0.17	0.16	-1.03	.304
Elderly adults	Predictive - Neutral	-0.17	0.16	-1.03	.304

Summary P600 results

In contrast with our predictions, Context only influenced elderly adults' processing of the test sentences in both the 500-800 ms and the 800-1100 ms time window. In the 500-800 ms time window, predictive context sentences modulated elderly adults' processing of literal, but not idiomatic sentences. However, in the 800-1100 ms time window, the effect of Context on elderly adults' processing of test sentences generalized to idiomatic sentences.

Individual differences

To test whether potential differences between younger and elderly adults' processing of idioms in context can be linked to between-group differences in cognitive performance, we tested the extent to which participants' scores on the reading span task, the paired-associates task and the verbal fluency task were significant predictors of N400 and P600 amplitude. A significant effect was only found for the verbal fluency task: high performance on the verbal fluency task was associated with a less negative amplitude of the N400 in the 200-300 ms time window (β = 0.05, SE = 0.02, t = 2.70) and 300-400 ms time window (β = 0.05, SE = 0.02, t = 2.33), but not in the 400-500 ms time window. In the 800-1100 ms time window, performance on the verbal fluency task showed a significant relation with mean voltage (β = 0.04, *SE* = 0.02, *t* = 2.19), with high scores being associated with more positive mean voltages in the P600.

DISCUSSION

This study examined how age-related changes in executive functions influence the way in which elderly adults use linguistic context when processing ambiguous expressions, in this case idioms. We investigated electrophysiological signatures of language processing that have previously been shown to be sensitive to contextual factors (e.g., Federmeier & Kutas, 2005; Kutas & Federmeier, 2011) as well as to semantic retrieval (Brouwer et al., 2016), and integration (e.g., Brown & Hagoort, 1993). Semantic retrieval and integration processes in particular were expected to play a key role in the resolution of ambiguity that arises during idiom comprehension. Predictive context information is expected to facilitate the processes of semantic retrieval and integration by pre-activating upcoming words. Since the use of context in language processing has been related to working memory (e.g., Federmeier & Kutas, 2005; Light & Capps, 1986) we hypothesized that, compared to younger adults, elderly adults' use of context across sentence boundaries would be affected because of their age-related decline in working memory capacity. Moreover, as idiom processing and comprehension are correlated with inhibition skills in younger adults (Cacciari et al., 2018), age-related decline in inhibition was hypothesized to hamper elderly adults' ability to suppress the literal meanings of idiom constituents. Our findings indeed show a difference between younger adults' and elderly adults' processing of language in context, but only with respect to literal sentences.

In the following, we first discuss the relation between working memory capacity and the effect of context on the retrieval of meaning in literal and idiomatic sentences, which we studied in the N400. Second, we elaborate on the effect of context on the integration of sentence meaning, which we studied in the P600. Third, we discuss the link between inhibition and the suppression of literal word meanings in younger and elderly adults' idiom processing in the N400. Finally, we relate younger and elderly adults' processing of literal and idiomatic sentences in context to individual differences in cognitive abilities.

Using linguistic context to facilitate the retrieval of idiom meaning (N400 effects)

The extent to which younger adults and elderly adults use linguistic context information to facilitate the retrieval of an idiom's figurative meaning was studied via the N400. For younger adults, we expected the N400 to decrease for sentences preceded by a predictive compared to a neutral context, consistent with previous studies on the processing of literal language in context (e.g., Federmeier & Kutas, 1999; Federmeier et al., 2002). Moreover, since supportive context information is particularly important in idiomatic sentences to resolve the ambiguity of the idiom, we tested whether the facilitating effect of context was stronger in idiomatic compared to literal sentences. As the use of context in language processing has been related to working memory capacity (Federmeier & Kutas, 2005), which typically declines with age (e.g., Salthouse et al., 1988), we hypothesized that elderly adults would not benefit from a predictive context to disambiguate an idiom as much as younger adults.

As expected, the elderly adults in our study had a significantly smaller working memory capacity than the younger adults. Nevertheless, contrary to our predictions, we found that a preceding predictive context sentence decreased the amplitude of the N400 in both groups. Moreover, context sentences had a similar influence on the processing of literal and idiomatic sentences. These findings indicate that for younger as well as elderly adults a predictive context facilitates the retrieval of upcoming words from the lexicon, regardless of whether these words are part of an idiom or not.

While our findings do not meet our expectations, they are consistent with previous eye-tracking studies showing that elderly adults are able to maintain their level of sentence processing in the face of memory decline, by adopting a top-down processing strategy instead of a bottom-up strategy in which each incoming word is processed (Rayner et al., 2006; Von der Malsburg et al., 2015). By relying more heavily on predictive processing, elderly adults can reduce the burden on working memory capacity, because they do not have to process each single word. In our study, elderly adults may have used the context information to predict and thereby pre-activate upcoming words in the subsequent sentence. When actually encountering the predicted word, its retrieval is facilitated, as indexed by a reduction in N400 amplitude (e.g. Baggio & Hagoort, 2011; Kutas & Federmeier, 2011; Brouwer et al., 2016).

The extent to which compensatory processing strategies can be adopted has also been argued to depend on elderly adults' executive functions, such as attention (Risse & Kliegl, 2011) or working memory capacity (Kemper, Crow, & Kemtes, 2004). Nevertheless, whereas working memory capacity was significantly reduced in the elderly adults, we found that their ability to use context information for the processing of subsequent sentences was unaffected. Although Federmeier and Kutas (2005) found that elderly adults with lower reading spans showed larger age-related changes in the use of context within a sentence, it could be that the type of working memory measured by the Reading Span does not match the type of working memory capacity that is required for using context across sentence boundaries. The challenge for future studies is to find tasks of working memory that are sensitive to the type of working memory that is involved in the use of context information across sentence boundaries.

Our finding that context facilitated ambiguity processing in both younger and older adults seemingly contradicts previous work suggesting that the ability to use context declines with age (Dagerman et al., 2006; Federmeier et al., 2002; Federmeier et al., 2003; Federmeier & Kutas, 2005). It also contradicts studies arguing the opposite (e.g., Pichora-Fuller et al., 1995; Speranza et al, 2000; Stine & Wingfield, 1994). However, close inspection of the methodologies used in these studies suggests that the contradictions arise from methodological differences.

First, context has been manipulated both within and across sentence boundaries. For example, similar to Light and Caps (1986) and Federmeier et al. (2002), in our study we investigated the use of predictive context information across sentence boundaries. Pichora-Fuller et al. (1995), Speranza et al. (2000), and Stine and Wingfield (1994) however, investigated the use of context within a sentence and found stronger effects of context for elderly adults than for younger adults. Second, online and offline measures differ with respect to their sensitivity to context effects. While studies using offline measures do not find an age-related decline in context-dependent language processing (e.g., Pichora-Fuller et al., 1995; Speranza et al, 2000; Stine & Wingfield, 1994; but see Dagerman et al., 2006), online measures, such as EEG and eye-tracking, have shown that the use of context in language processing changes with age (e.g., Federmeier et al., 2002; Federmeier et al., 2003; Federmeier & Kutas, 2005).

However, Federemeier et al. (2002), who used ERPs to study the use of context across sentence boundaries, only reported a difference between younger and elderly adults' use of context for target words that were unexpected though categorically related to the expected word. Younger and elderly adults used context in the same way in processing correct critical words and incorrect, semantically unrelated critical words, the same two types of critical words that were used in our study. Our results corroborate thus the findings by Federmeier et al. (2002). Crucially, our study highlights the importance of using online measures to get insight in the effect of cognitive aging on language processing.

Using linguistic context to facilitate the integration of sentence meaning (P600 effects)

In the previous section, we focused on the N400 to study the effect of context on the retrieval of idiom meaning. Here, we discuss the facilitative effect of context on the integration of an idiom's figurative meaning in a sentence, which we studied in the P600. Relative to retrieving an idiom's figurative meaning, integrating it in a sentence requires an additional processing step: the content of working memory (i.e., at least the preceding sentence context) must be updated and the meaning of the sentence as a whole must be computed from the combined information.

In our study, a predictive context was hypothesized to support these integration processes. We tested whether the effect of context on integration was stronger in idiomatic compared to literal sentences, because the successful integration of idioms requires context-dependent disambiguation. In terms of the P600, the facilitative effect of a preceding predictive context on the integration of an idiomatic sentence's meaning was expected to decrease the amplitude of the P600. However, as elderly adults' reduced working memory capacities would make it difficult to retain context information, the predicted decrease in P600 amplitude was expected to be weaker or absent in elderly adults. Our findings, however, show the opposite effect, with context only affecting semantic integration processes in our group of elderly participants. Also, contrary to our expectations, the effect is more pronounced in the processing of literal sentences and appears in an earlier time window.

How can these findings be reconciled with what we know about working-memory decline in old age? To compute the meaning of literal sentences, readers must retrieve and integrate the meanings of the individual words one by one. Younger adults' speed of sentence processing is sufficiently high to be able to update the contents of working memory with each individual word from the input. However, potential slowing down in elderly adults' sentence processing could make this process relatively more effortful for them. They might therefore develop a processing strategy that relies more strongly on the use of context information to predict upcoming words in literal sentences (e.g., Rayner et al., 2006; Von der Malsburg et al., 2015).

For idioms, different processing strategies are required. Contrary to literal sentences, the meanings of idioms can be retrieved for the phrase as a whole from the mental lexicon (e.g., Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Sprenger et al., 2006; Titone & Connine, 1999). That is, idioms have a processing advantage above literal phrases (e.g., Carrol & Conklin, 2019; Conklin & Schmitt, 2008; Rommers et al., 2013; Siyanova-Chanturia et al., 2011; Swinney & Cutler, 1979; Underwood et al., 2004). This may compensate for the age-related slowing down in sentence processing (see Dagerman et al., 2006). In addition, we recently showed that idiom familiarity is particularly well established in the elderly (Sprenger, la Roi, & van Rij, 2019). Since familiarity has also been found to play an important role in the processing advantage of idioms (Carrol & Conklin, 2019), elderly adults may rely much less on additional context information to construct the meaning of idiomatic sentences than younger adults.

Interestingly, our results showed that a predictive context resulted in an increased instead of the expected decreased P600 following literal sentences in elderly adults. This is likely due to the fact that the context effect was driven by incorrect sentences. Using a predictive context to formulate strong expectations for an upcoming word actually hampers processing when the strong expectation is violated, as in the incorrect sentences. Consequently, P600 amplitude increased for incorrect sentences preceded by a predictive compared to a neutral context. Although the predictive context seemed to generally hamper elderly adults' processing, this finding nevertheless indicates that elderly adults make use of context to predict upcoming words. However, since the type of critical word did not interact with Age group in the 500-800 ms and 800-1100 ms time windows, we could not statistically test the difference between elderly adults' mean voltages in response to correct versus incorrect sentences.

In conclusion, whereas the effects in the N400 indicated that a predictive context facilitates the retrieval of idioms and literal phrases in both younger and elderly adults, we found that the process of integrating word meanings to construct a representation of sentence meaning, indexed by the P600, was influenced by context only in elderly adults. This indicates that processing literal sentences is more effortful for elderly adults than processing idiomatic sentences. Idioms reduce the load on elderly adults' working memory, thereby leaving more resources available to construct a representation of sentence meaning.

Additionally, our study shows that although literal language processing is effortful for elderly adults, they can facilitate this process by using linguistic context information. The discrepancy in the effect of context on the N400 and P600 suggests that context information has a different influence on the processes of semantic retrieval and semantic integration and that these processes may not be equally affected by cognitive aging.

Disambiguating idiom meaning (N400 effects)

By investigating the use of context in younger and elderly adults' literal and idiomatic sentence processing, we aimed to get insight in the effect of age-related changes in working memory capacity on language processing abilities. On top of that, we were interested in the effect of age-related decline in inhibition skills on language processing. To study inhibition processes, we examined the way in which younger and elderly adults retrieve and process an idiom's figurative meaning. As the N400 has been used as an index of retrieval processing (Brouwer et al., 2016), we focused on younger and elderly adults' amplitude of the N400 in response to idioms versus literal sentences. We predicted that elderly adults would have reduced inhibition skills, making it more difficult for them to suppress the literal word meanings in idiom processing. Consequently, elderly adults were predicted *not* to show an N400 reduction for idioms relative to literal sentences, as previously found for younger adults (Canal et al., 2015; Laurent et al., 2006; Rommers et al., 2013; Strandburg et al., 1993).

However, the decreased N400 that we found for idiomatic compared to literal sentences in younger as well as elderly adults indicates that both groups were able to quickly inhibit the literal meanings of the idiom constituents and process the meaning of the idiom as a whole. Although unexpected, this result adds a novel insight to the existing literature by generalizing previous findings on the facilitative effect of idioms on language processing in younger adults (e.g., Canal, et al., 2015; Laurent et al., 2006; Conklin & Schmitt, 2008; Paulmann et al., 2015; Rommers, et al., 2013; Siyanova-Chanturia et al., 2011; Siyanova-Chanturia et al., 2017; Strandburg et al., 1993; Swinney & Cutler, 1979; Underwood et al., 2004; Vespignani et al., 2010) to a population of elderly adults. Again, these findings are in line with our recent study showing that idiom familiarity increases with age (Sprenger et al., 2019). Although being familiar with an idiom is not the same as being able to process its meaning in context, it is conceivable that elderly adults' larger experience with idiomatic expressions throughout their life compensates for the potential difficulties that may be involved in selecting the idiom's intended, figurative meaning.

Even though both younger and elderly adults distinguished between literal and idiomatic sentences starting from 200 ms after the onset of the critical word, one interesting finding is that elderly adults did not show a significantly smaller N400 for correct compared to incorrect critical words. While this absence of an N400 difference was not predicted, it is compatible with the results of Gunter et al. (1995). They found that in middle-aged adults, N400 differences between sentences with a correct compared to an incorrect ending disappeared when participants had to retain an incomplete main clause while processing a subordinate clause inserted before the final word. In our study, it was possibly cognitively effortful for elderly adults to keep the context sentence in working memory while processing the literal or idiomatic test sentence, causing N400 differences between correct and incorrect critical words to disappear. In this way, the absence of a difference in elderly adults' processing of correct and incorrect sentence can be seen as an illustration of how their limited working memory capacity may affect their ability to process a sentence while retaining context information from a previous sentence.

Another possibility is that the lack of a difference between elderly adults' N400 in response to incorrect versus correct sentences is driven by general age-related changes in electrophysiological activity, such as a posterior to anterior shift of brain activity (e.g., Davis, Dennis, Daselaar, Fleck, & Cabeza, 2008) or a general age-related reduction in N400 amplitude (Kutas & Iragui, 1998; Wlotko, Lee, & Federmeier, 2010), or effects of fatigue. However, within the groups of younger and elderly adults we found similar patterns of amplitude changes as well as a similar distribution of effects in the N400 and P600 following the processing of idioms versus literal sentences. Thus, the finding that younger and elderly adults' brain responses were similarly modified by idiomatic versus literal sentences refutes the possibility that all our effects are merely caused by these general changes in electrophysiological activity related to the effect of age or fatigue.

Individual differences

To relate electrophysiological differences between younger and elderly adults to between-group differences in executive functioning, we examined the extent to which participants' scores on offline cognitive tasks predicted the mean amplitudes of their N400s and P600s. Although elderly adults scored significantly lower on the working memory task and the inhibition task than younger adults, only verbal fluency scores significantly explained the amplitude of the N400 and P600.

Since verbal fluency scores did not significantly differ between younger and elderly adults, differences between younger and elderly adults' use of context in online literal and idiomatic sentence processing cannot be attributed to differences in verbal fluency. Nevertheless, the mechanisms involved in verbal fluency tasks may match the working memory and inhibition processes at play in context-dependent language processing. In the verbal fluency task, to avoid repetition of previously retrieved items, participants must retain and suppress words that have become irrelevant because they were already listed, which resembles the suppression of the literal meanings of idiom constituents. At the same time participants must continue to retrieve new words and add them to the contents of working memory, a process that shows similarities with word by word sentence integration.

Previous findings confirm the idea that the mechanisms underlying verbal fluency support the processing of literal and ambiguous language in context. For example, Federmeier et al. (2002) showed that elderly adults who score high on verbal fluency tasks benefit in the same way as younger adults from supportive context information in sentence processing. Furthermore, Lee and Federmeier (2011) found that compared to elderly adults with low verbal fluency scores, those with high verbal fluency scores more easily inhibit the irrelevant meaning of ambiguous words. Thus, our findings suggest that verbal fluency scores may be an important measure to take into account when studying the effect of cognitive aging on language processing. Furthermore, since we did not find a correlation between the offline working memory and inhibition tasks and the N400 and P600, future studies should aim to find psychometric tests that are able to measure the type of working memory capacity and inhibition skills that are involve in the use of context in the processing of ambiguous language. The Listening Span used by Cacciari et al. (2018) may be a useful alternative to test working memory capacity and inhibitory control skills that are involved in the processing of idioms.

Finally, it should be noted that the elderly adults participating in this study were all highly educated, matching the education level of the younger adults. As education has been identified as a protective factor against age-related cognitive decline (e.g., Gollan et al., 2011; Karp et al., 2004), it is possible that participants' high level of education had a beneficial effect on their executive functions The preserved level of executive functioning may have preserved elderly adults' language processing abilities in general, or allowed them to adopt compensatory processing strategies, as has been found in previous studies (e.g., Rayner et al., 2006; Risse & Kliegl, 2011; Von der Malsburg et al., 2015). However, as we did not specifically investigate the relation between language abilities and education, we cannot draw conclusions about the role that education played in our study. To gain more insight in the way in which education modulates the relation between executive functioning and language processing in elderly adults, future research needs to include elderly participants from a variety of educational backgrounds. At the same time, the fact that we found age effects on language processing even in a group of elderly adults with a high level of education illustrates the pervasiveness of cognitive aging on language performance.

CONCLUSION

We found that both younger adults and elderly adults use context information to facilitate the retrieval of words in upcoming sentences. Furthermore, both groups showed facilitated processing for idioms compared to literal sentences. However, with regards to the integration of words in a sentence, we found that in neither in younger nor elderly adults context information specifically contributes to the integration of an idiom's figurative meaning when constructing a representation of sentence meaning. Regarding the processing of literal sentences, we found that only elderly adults relied on context information to compute the meaning of literal sentences.

Although idioms are ambiguous between a literal and a figurative meaning, the fact that an idiom's figurative meaning is retrieved from the mental lexicon as a whole could explain why idiomatic sentences are processed more easily than literal sentences by both younger and elderly adults. If the reader does not have to retrieve and integrate the meaning of each single word especially elderly adults are left with sufficient cognitive resources to compute the meaning of an idiomatic sentence without the need for context information. In contrast, when sentence meaning needs to be computed word by word, as in literal sentences, elderly adults must rely more heavily on linguistic context information to predict the meaning of upcoming words. To conclude, our findings provide a novel perspective on idiom processing by showing that compared to literal sentences, idioms reduce the processing load not only in younger, but also in elderly adults, even though the suppression of literal word meanings in idioms requires cognitive effort. Our study furthermore suggests that when the processing load is high, age-related cognitive decline affects our ability to process language.

REFERENCES

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

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349-371.

Arnon, I., & Priva, U. C. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, *9*(3), 377-400.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of memory and language*, *62*(1), 67-82.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*(3), 189-208.

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes*, *26*(9), 1338-1367.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. doi:10.18637/jss.v067.i01.

Bobrow, S. A., & Bell, S. M. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, *1*(3), 343-346.

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The 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. (2016). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive Science*, *41*(S6), 1318-1352.

Brouwer, H., & Hoeks, J. C. (2013). A time and place for language comprehension: mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience*, *7*, 758.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, *5*(1), 34-44.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, *98*(2), 159-168.

Burnham, K. P., & Anderson, D. R. (2002). A practical information-theoretic approach. Model selection and multimodel inference, 2nd ed. Springer, New York.

Cabeza, R., Nyberg, L., & Park, D. (2005). *Cognitive Neuroscience of Aging: Linking Cognitive and Cerebral Aging*. Oxford University Press.

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

Cacciari, C., Corradini, P., & Ferlazzo, F. (2018). Cognitive and personality components underlying spoken idiom comprehension in context. An exploratory study. *Frontiers in Psychology*, *9*(659), 1-17.

Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., & Cacciari, C. (2015). Electrophysiological correlates idioms comprehension: semantic composition does not follow lexical retrieval. In *NetWordS* (pp. 98-101).

Carrol, G., & Conklin, K. (2019). Is All Formulaic Language Created Equal? Unpacking the Processing Advantage for Different Types of Formulaic Sequences. *Language and Speech*. https://doi.org/10.1177/0023830918823230

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers?. *Applied Linguistics*, *29*(1), 72-89.

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

Dagerman, K. S., MacDonald, M. C., & Harm, M. W. (2006). Aging and the use of context in ambiguity resolution: Complex changes from simple slowing. *Cognitive Science*, *30*(2), 311-

345.

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

Davis, S. W., Dennis, N. A., Daselaar, S. M., Fleck, M. S., & Cabeza, R. (2008). Que PASA? The posterior–anterior shift in aging. *Cerebral Cortex*, *18*(5), 1201-1209.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469-495.

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., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133-146.

Federmeier, K. D., Van Petten, C., Schwartz, T. J., & Kutas, M. (2003). Sounds, words, sentences: age-related changes across levels of language processing. *Psychology and Aging*, *18*(4), 858.

Fernández, G., Shalom, D. E., Kliegl, R., & Sigman, M. (2014). Eye movements during reading proverbs and regular sentences: The incoming word predictability effect. *Language, Cognition and Neuroscience*, *29*(3), 260-273.

Gaudreau, G., Monetta, L., Macoir, J., Laforce Jr, R., Poulin, S., & Hudon, C. (2013). Verbal irony comprehension in older adults with amnestic mild cognitive impairment. *Neuropsychology*, *27*(6), 702-712.

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, *8*(2), 149-156.

van Ginkel, W., & Dijkstra, T. (2019). The tug of war between an idiom's figurative and literal meanings: Evidence from native and bilingual speakers. *Bilingualism: Language and Cognition*, 1-17.

Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183-206.

Gollan, T. H., Salmon, D. P., Montoya, R. I., & Galasko, D. R. (2011). Degree of bilingualism predicts age of diagnosis of Alzheimer's disease in low-education but not in highly educated Hispanics. *Neuropsychologia*, *49*(14), 3826-3830.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology, 55*(4), 468-484.

Gunter, T. C., Jackson, J. L., & Mulder, G. (1995). Language, memory, and aging: An electrophysiological exploration of the N400 during reading of memory-demanding sentences. *Psychophysiology*, *32*(3), 215-229.

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.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439-483.

Hasher, L., Quig, M. B., & May, C. P. (1997). Inhibitory control over no-longer-relevant information: Adult age differences. *Memory & Cognition*, *25*(3), 286-295.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *Psychology of Learning and Motivation*, *22*, 193-225.

Hoenig, K., & Scheef, L. (2009). Neural correlates of semantic ambiguity processing during context verification. *Neuroimage*, *45*(3), 1009-1019.

Holsinger, E. (2013). Representing idioms: Syntactic and contextual effects on idiom processing. *Language and Speech*, *56*(3), 373-394.

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

Karp, A., Kåreholt, I., Qiu, C., Bellander, T., Winblad, B., & Fratiglioni, L. (2004). Relation of education and occupation-based socioeconomic status to incident Alzheimer's disease. *American Journal of Epidemiology*, *159*(2), 175-183.

Konopka, A. E., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology*, *58*(1), 68-101.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*(12), 463-470.

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*, 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 Section, 108*(5), 456-471.

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.

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

Lenth, R. V. (2016). Least-squares means: the R package lsmeans. *Journal of Statistical Software*, *69*(1), 1-33. doi:10.18637/jss.v069.i01.

Light, L. L., & Capps, J. L. (1986). Comprehension of pronouns in young and older adults. *Developmental Psychology*, *22*(4), 580-585.

Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.

von der Malsburg, T., Kliegl, R., & Vasishth, S. (2015). Determinants of scanpath regularity in reading. Cognitive science, 39(7), 1675-1703.

Mason, R. A., & Just, M. A. (2007). Lexical ambiguity in sentence comprehension. *Brain Research*, *1146*, 115-127.

Meyer, A. M., & Federmeier, K. D. (2010). Event-related potentials reveal the effects of aging on meaning selection and revision. *Psychophysiology*, *47*(4), 673-686.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49-100.

Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of collocational complex prepositions. *Language and Cognitive Processes*, *28*(6), 762-788.

Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, *83*(3), 176-190.

Molinaro, N., Conrad, M., Barber, H. A., & Carreiras, M. (2010). On the functional nature of the N400: Contrasting effects related to visual word recognition and contextual semantic integration. *Cognitive Neuroscience*, *1*(1), 1-7.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, *29*(2), 186-217.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 786-803.

Papagno, C., Lucchelli, F., Muggia, S., & Rizzo, S. (2003). Idiom comprehension in Alzheimer's disease: The role of the central executive. *Brain*, *126*(11), 2419-2430.

Paulmann, S., Ghareeb-Ali, Z., & Felser, C. (2015). Neurophysiological markers of phrasal verb processing: Evidence from L1 and L2 speakers. In Roberto R. Heredia, & Anna B. Cieslicka (Eds.). *Bilingual figurative language processing* (pp. 245-267). Cambridge: Cambridge University Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (191-225). London: Longman.

Peterson, R. R., Burgess, C., Dell, G. S., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1223.

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America, 97*(1), 593-608.

Proverbio, A. M., Crotti, N., Zani, A., & Adorni, R. (2009). The role of left and right hemispheres in the comprehension of idiomatic language: an electrical neuroimaging study. *BMC Neuroscience*, *10*(1), 116.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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.

Risse, S., & Kliegl, R. (2011). Adult age differences in the perceptual span during reading. *Psychology and Aging, 26*(2), 451-460.

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.

Salthouse, T. A., Kausler, D. H., & Saults, J. S. (1988). Investigation of student status, background variables, and feasibility of standard tasks in cognitive aging research. *Psychology and Aging, 3*(1), 29-37.

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & Van Heuven, W. J. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language, 175*, 111-122.

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27*(2), 251-272.

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.

Snider, N., & Arnon, I. (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. *Frequency Effects in Language*, 127-163.

Spaan, P. E. J. (2012). Cognitieve achteruitgang bij normale veroudering en de ziekte van Alzheimer: Een continue of discontinue overgang?. *Tijdschrift voor Neuropsychologie, 7*(1), 3-15.

Speranza, F., Daneman, M., & Schneider, B. A. (2000). How aging affects the reading of words in noisy backgrounds. *Psychology and Aging*, *15*(2), 253-258.

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

Sprenger, S. A., la Roi, A., & Van Rij, J. (2019). The Development of Idiom Knowledge across the

Lifespan. *Frontiers in Communication*, *4*, 29.

Stine, E. A., & Wingfield, A. (1994). Older adults can inhibit high-probability competitors in speech recognition. *Aging and Cogntion*, *1*(2), 152-157.

Strandburg, R. J., Marsh, J. T., Brown, W. S., Asarnow, R. F., Guthrie, D., & Higa, J. (1993). Event-related potentials in high-functioning adult autistics: Linguistic and nonlinguistic visual information processing tasks. *Neuropsychologia*, *31*(5), 413-434.

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Vehavior, 18*(5), 523-534.

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

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), Formulaic sequences (pp. 153–172). Amsterdam: John Benjamins.

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 der Linden, M., Hupet, M., Feyereisen, P., Schelstraete, M. A., Bestgen, Y., Bruyer, R., Lories, G., El Ahmadi, A., & Seron, X. (1999). Cognitive mediators of age-related differences in language comprehension and verbal memory performance. *Aging, Neuropsychology, and Cognition*, *6*(1), 32-55.

Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience, 22*(8), 1682-1700.

Vulchanova, M., Milburn, E., Vulchanov, V., & Baggio, G. (2019). Boon or Burden? The role of compositional meaning in figurative language processing and acquisition. *Journal of Logic, Language and Information*, *28*(2), 359-387.

Weinreich, U. (1969). Problems in the analysis of idioms. *Substance and Structure of Language*, 23-81.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, *120*(2), 272-292.

Wieling, M., Montemagni, S., Nerbonne, J., & Baayen, R. H. (2014). Lexical differences between Tuscan dialects and standard Italian: Accounting for geographic and sociodemographic variation using generalized additive mixed modeling. *Language, 90*(3), 669-692.

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., Lee, C. L., & Federmeier, K. D. (2010). Language of the Aging Brain: Event-Related Potential Studies of Comprehension in Older Adults. *Language and Linguistics Compass, 4*(8), 623-638.

World Health Organization (2015). *Fact sheet 404 - Ageing and health*, viewed March 13, 2017. http://www.who.int/mediacentre/factsheets/fs404/en/

Zempleni, M. Z., Haverkort, M., Renken, R., & Stowe, L. A. (2007). Evidence for bilateral involvement in idiom comprehension: An fMRI study. *Neuroimage*, *34*(3), 1280-1291.

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.

APPENDIX I

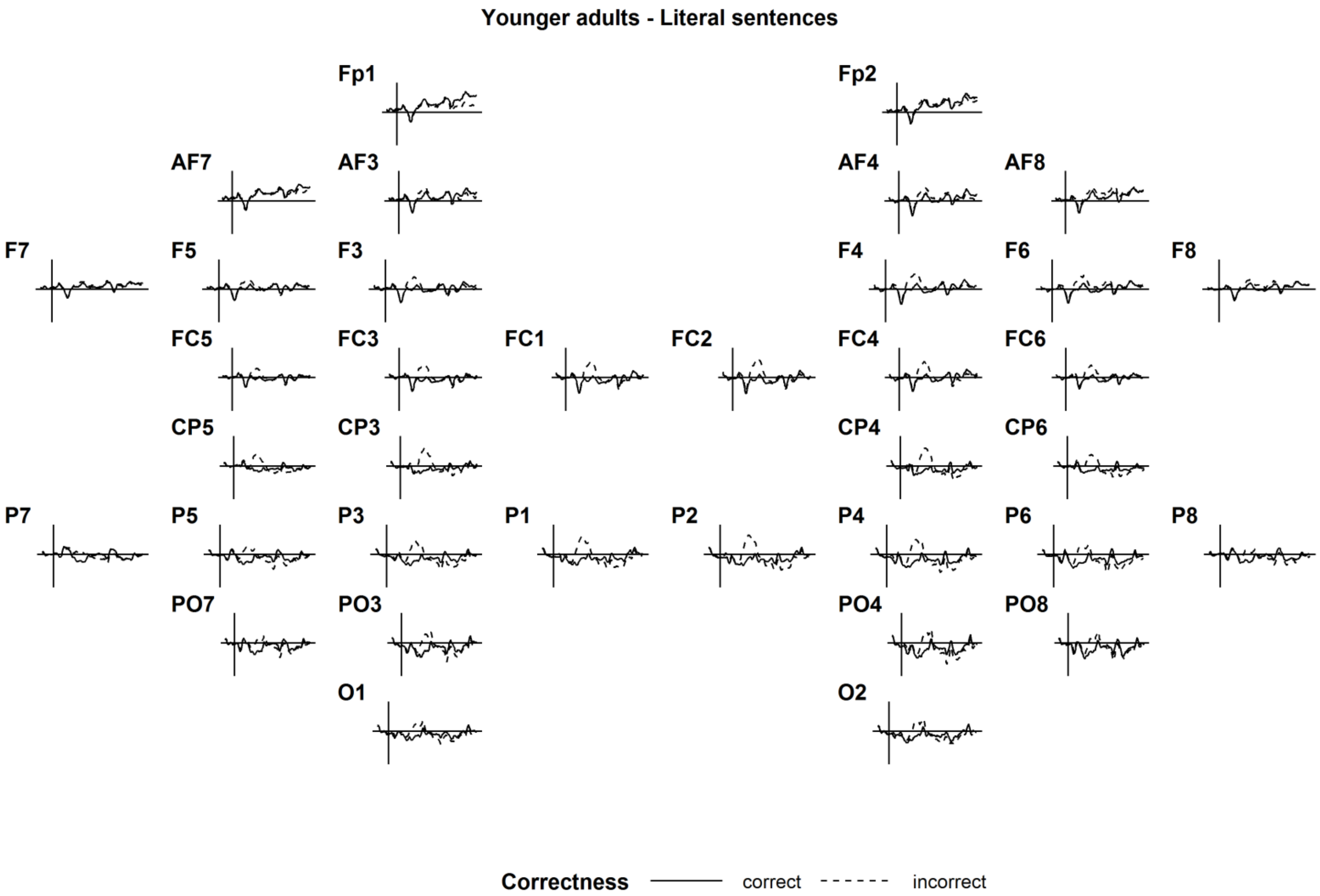

Figure A1. Younger adults' grand average ERPs in the 200-500 ms time window in response to correct (solid line) versus incorrect words (dotted line) in literal sentences. Effects are displayed for each electrode included in the statistical analysis.

APPENDIX I

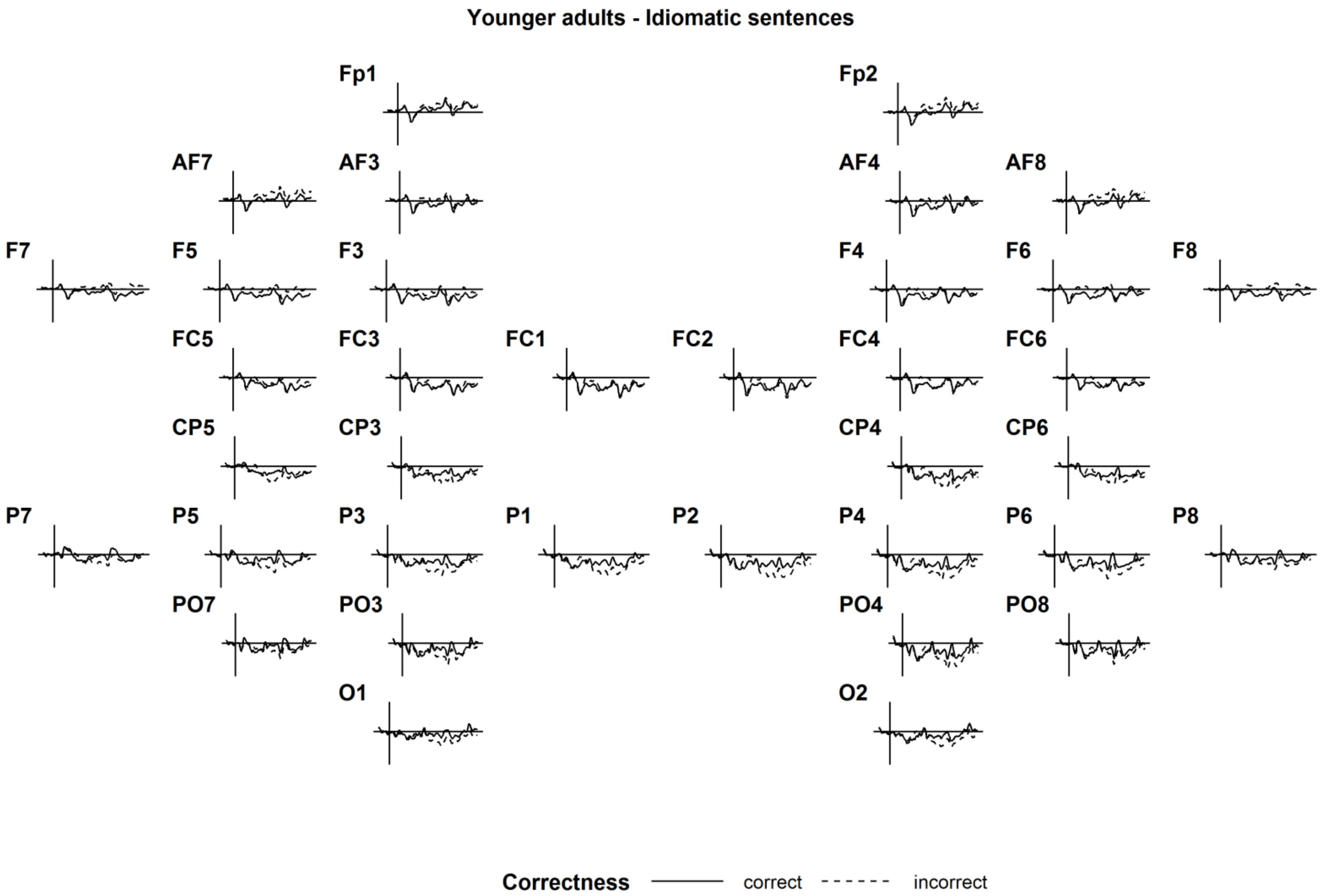

Figure A2. Younger adults' grand average ERPs in the 200-500 ms time window in response to correct (solid line) versus incorrect words (dotted line) in idiomatic sentences. Effects are displayed for each electrode included in the statistical analysis.

APPENDIX I

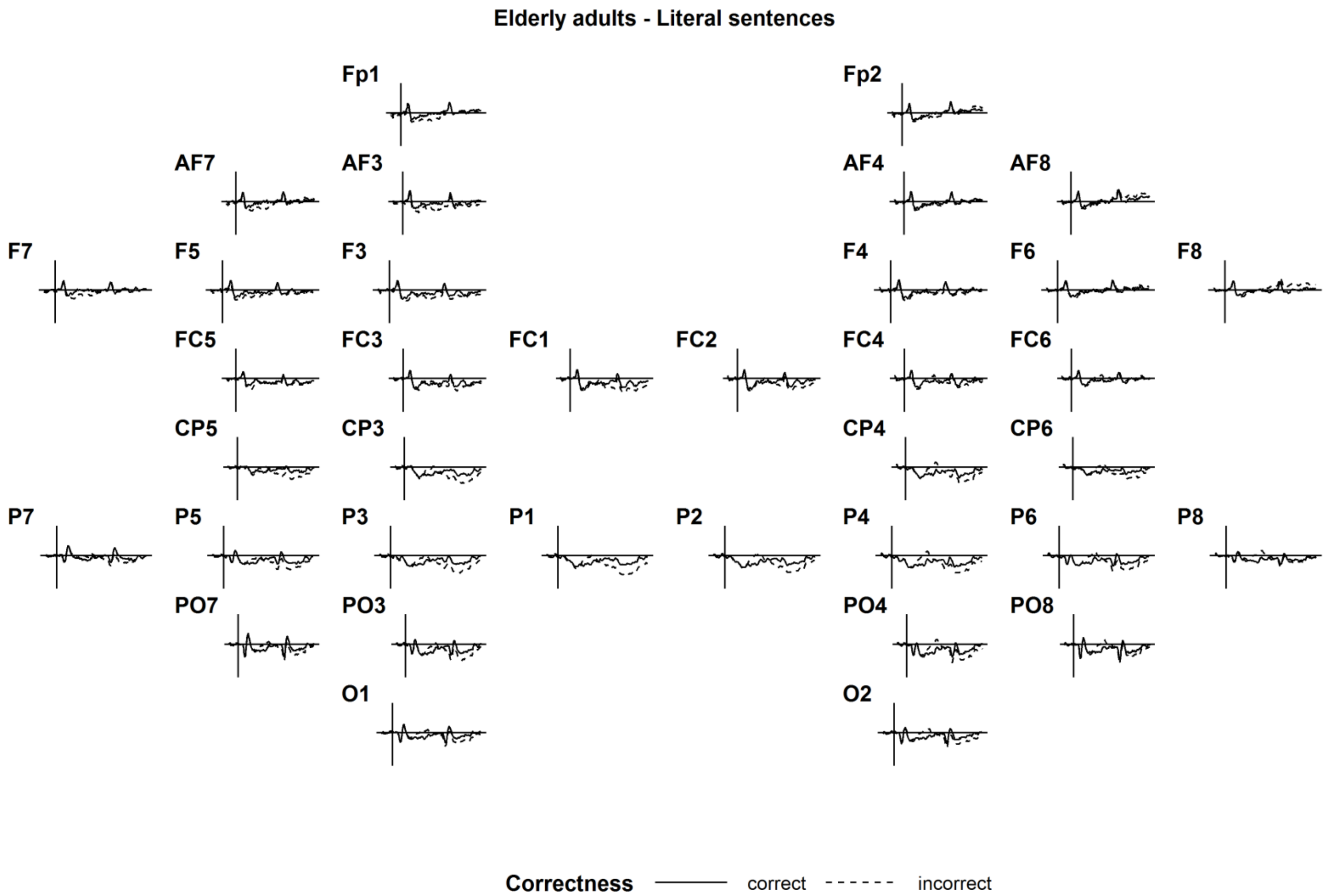

Figure A3. Elderly adults' grand average ERPs in the 200-500 ms time window response to correct (solid line) versus incorrect words (dotted line) in literal sentences. Effects are displayed for each electrode included in the statistical analysis.

APPENDIX I

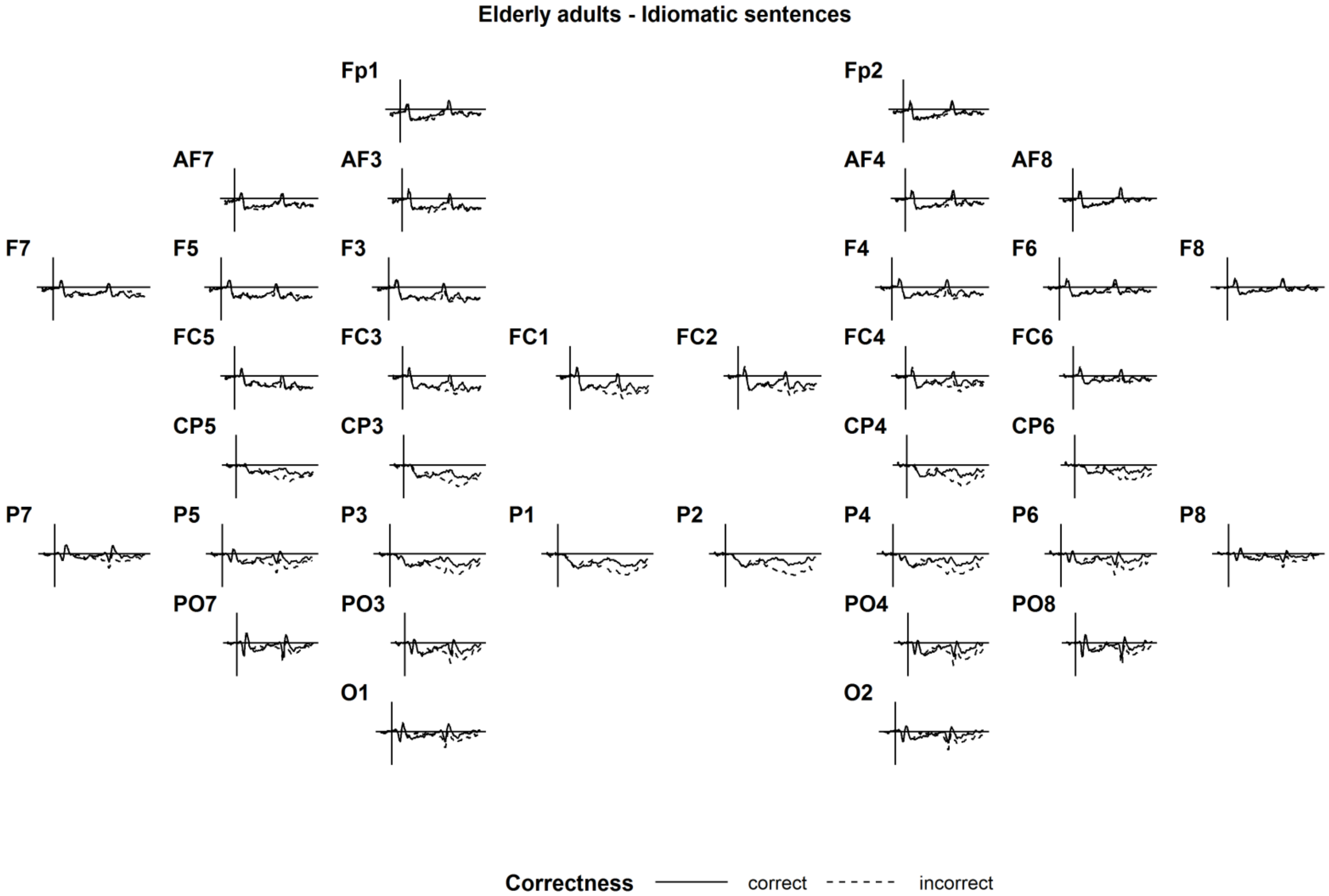

Figure A4. Elderly adults' grand average ERPs in the 200-500 ms time window in response to correct (solid line) versus incorrect words (dotted line) in idiomatic sentences. Effects are displayed for each electrode included in the statistical analysis.

APPENDIX II

Table A1. Summary of coefficients of the linear mixed effects regressions models fitted to the data of the 200-300 ms time window, the 300-400 ms time window, and 400-500 ms time window

Fixed effects			*200-300 ms*			*300-400 ms*			*400-500 ms*		
Factor			**Estimate**	**SE**	***t*-value**	**Estimate**	**SE**	***t*-value**	**Estimate**	**SE**	***t*-value**
1	Intercept		1.29	0.36	3.54	0.25	0.40	0.63	-0.09	0.40	-0.22
2	Group (elderly adults)		0.24	0.48	0.50	0.68	0.51	1.33	0.98	0.52	1.90
3	Context (predictive)		0.17	0.12	1.41	0.40	0.13	2.94	0.37	0.15	2.52
4	Idiomaticity (idiom)		0.81	0.23	3.64	0.85	0.28	3.00	1.34	0.34	3.93
5	Condition (incorrect)		-0.89	0.21	-4.32	-3.64	0.24	-15.10	-2.69	0.25	-10.64
6	Anteriority (posterior)		0.35	0.07	4.88	0.92	0.08	11.65	0.34	0.08	4.11
7	Hemisphere (right)		0.43	0.07	5.90	-0.19	0.08	-2.38	-0.45	0.08	-5.44
8	Group (senior) * Idiomaticity (idiom) * Condition (incorrect)		-0.71	0.29	-2.44	-2.32	0.32	-7.33	-1.49	0.33	-4.54
9	Verbal fluency		0.05	0.02	2.70	0.05	0.02	2.33	-	-	-
Random effects											
Groups		**Factor**	**Variance**	**SD**	**Corr.**	**Variance**	**SD**	**Corr.**	**Variance**	**SD**	**Corr.**
10	Subject	Intercept	2.62	1.62		2.99	1.73		3.02	1.74	
		Idiomaticity (idiom)	0.85	0.92	-0.08	1.39	1.18	0.02	2.20	1.48	-0.01
11	Critical word	Intercept	1.60	1.27		2.26	1.51		2.43	1.56	
		Context (predictive)	1.83	1.35	-0.60	2.26	1.50	-0.55	2.85	1.69	-0.52
12	Residual		43.28	6.58		51.40	7.17		55.43	7.45	

APPENDIX III

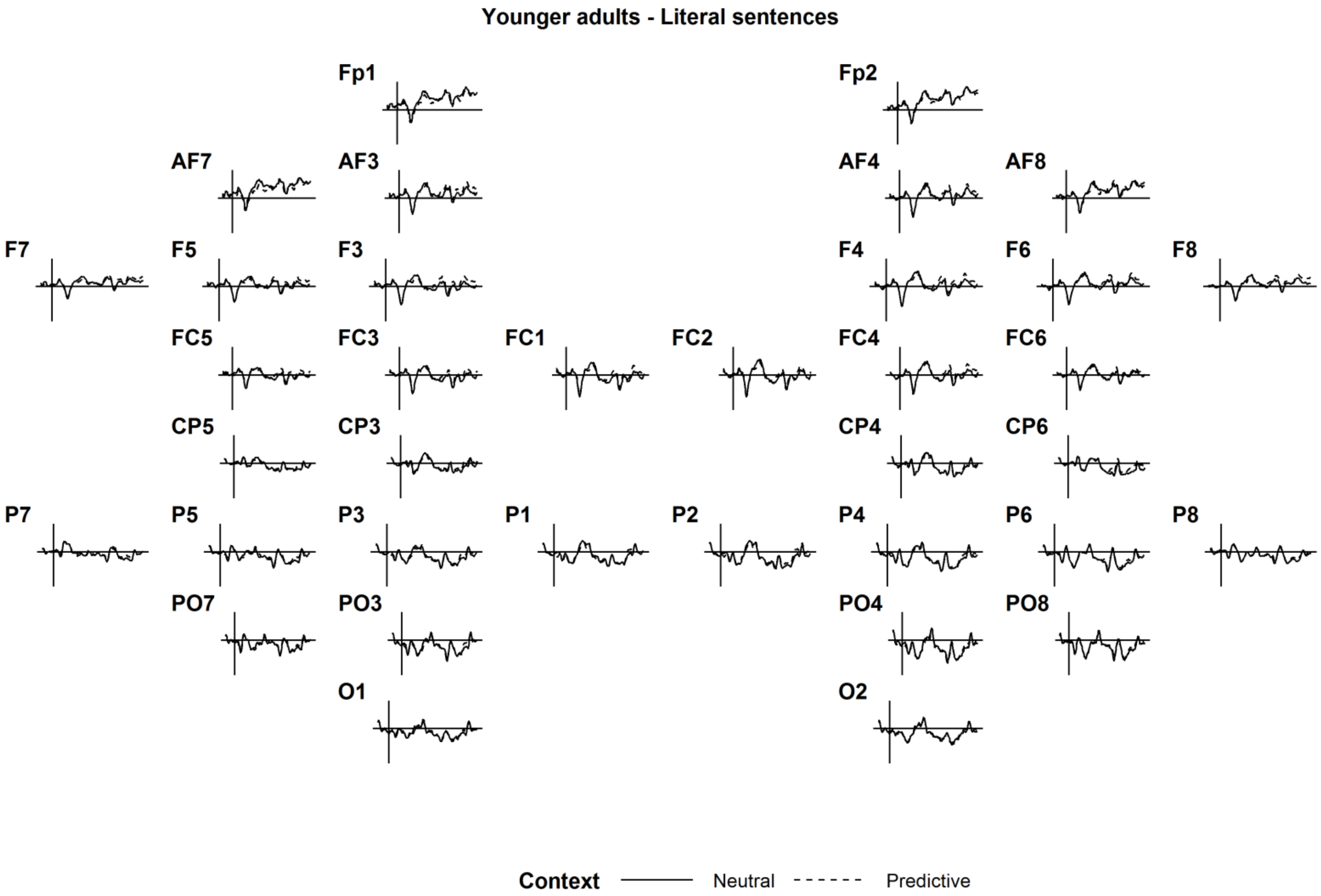

Figure A5. Younger adults' grand average ERPs in the 500-800 ms time window in response to literal sentences preceded by a neutral (solid line) versus a predictive context sentence (dotted line). Voltages are averaged across correct and incorrect sentences. Effects are displayed for each electrode included in the statistical analysis.

APPENDIX III

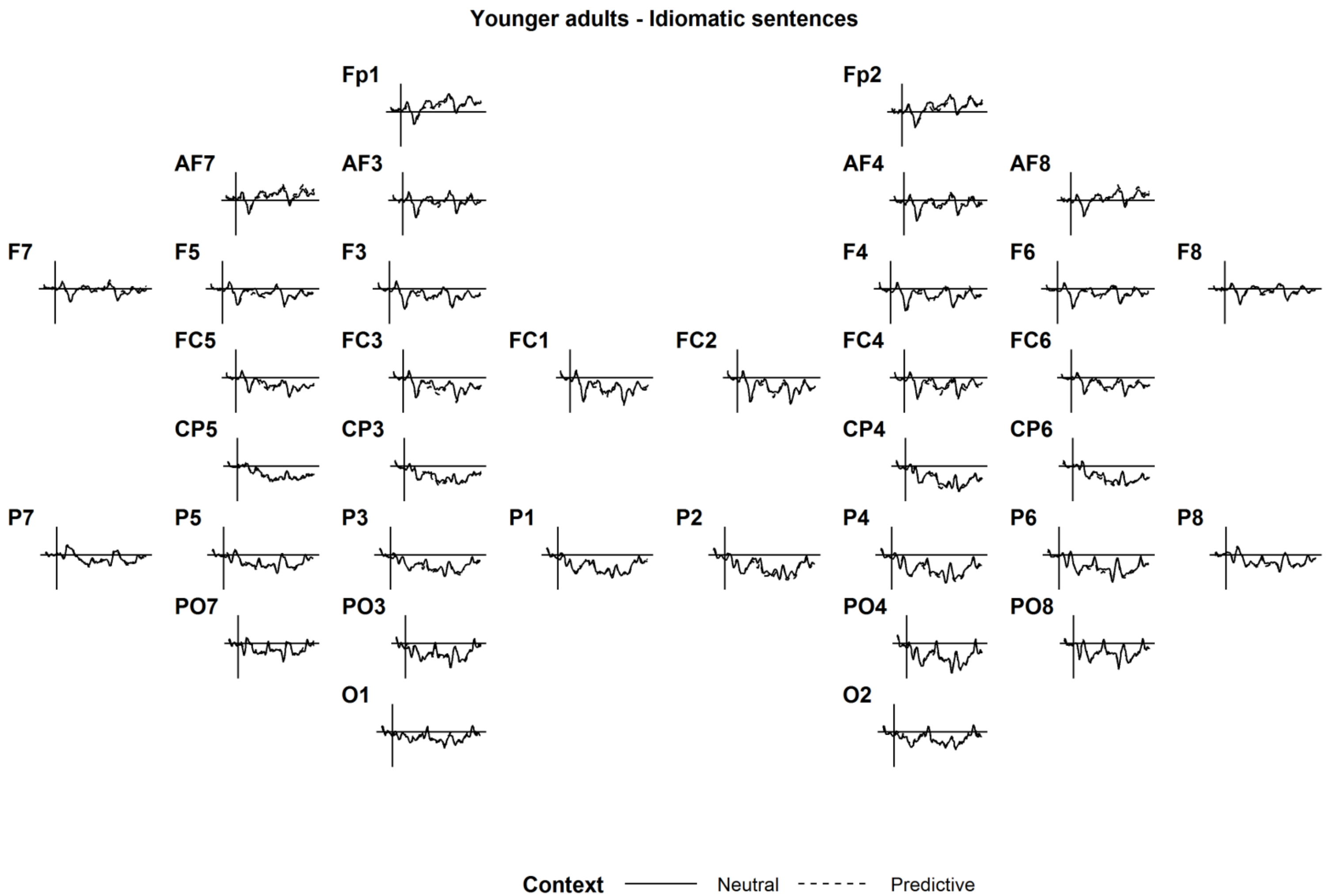

Figure A6. Younger adults' grand average ERPs in the 500-800 ms time window in response to idiomatic sentences preceded by a neutral (solid line) versus a predictive context sentence (dotted line). Voltages are averaged across correct and incorrect sentences. Effects are displayed for each electrode included in the statistical analysis.

APPENDIX III

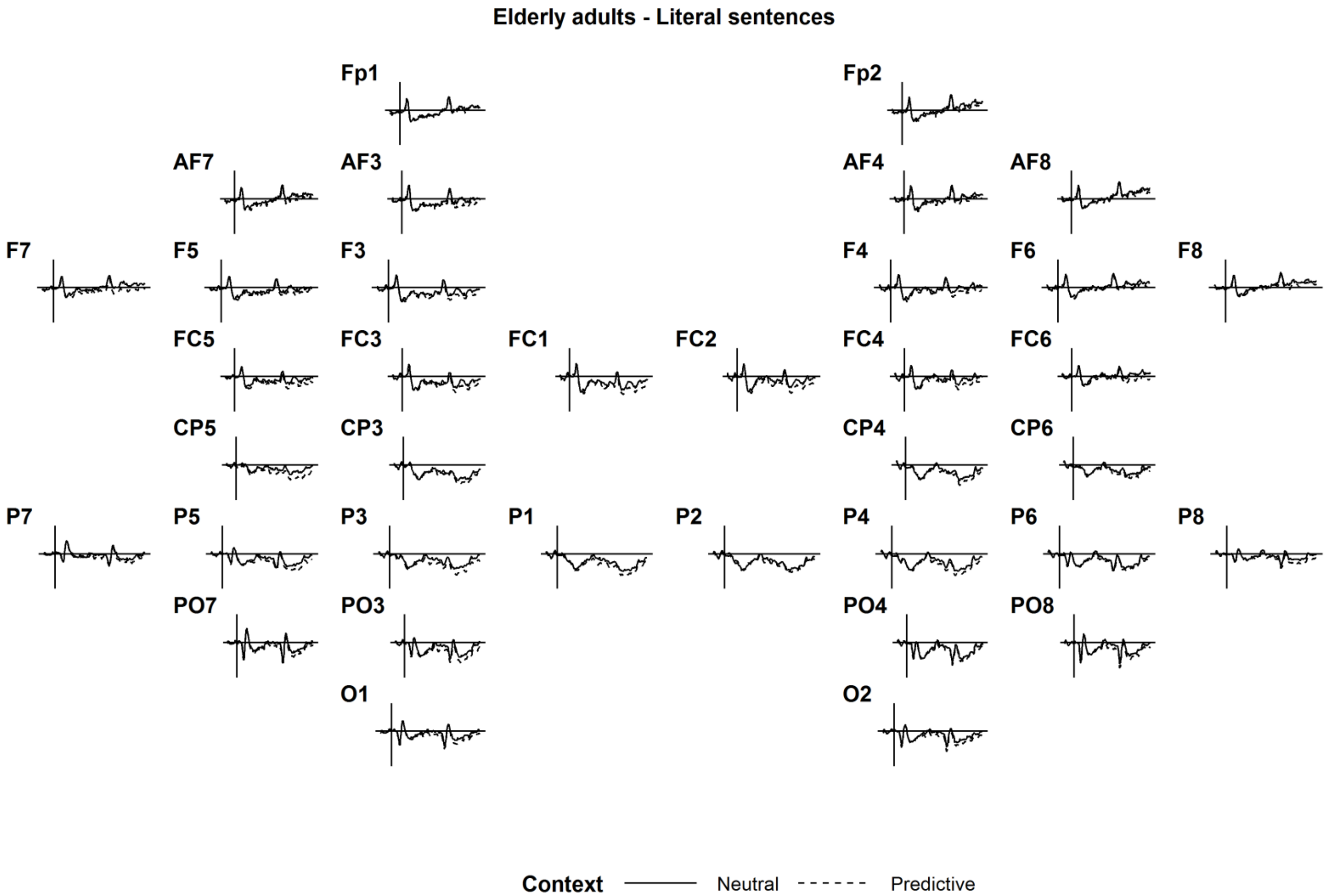

Figure A7. Elderly adults' grand average ERPs in the 500-800 ms time window in response to literal sentences preceded by a neutral (solid line) versus a predictive context sentence (dotted line). Voltages are averaged across correct and incorrect sentences. Effects are displayed for each electrode included in the statistical analysis.

APPENDIX III

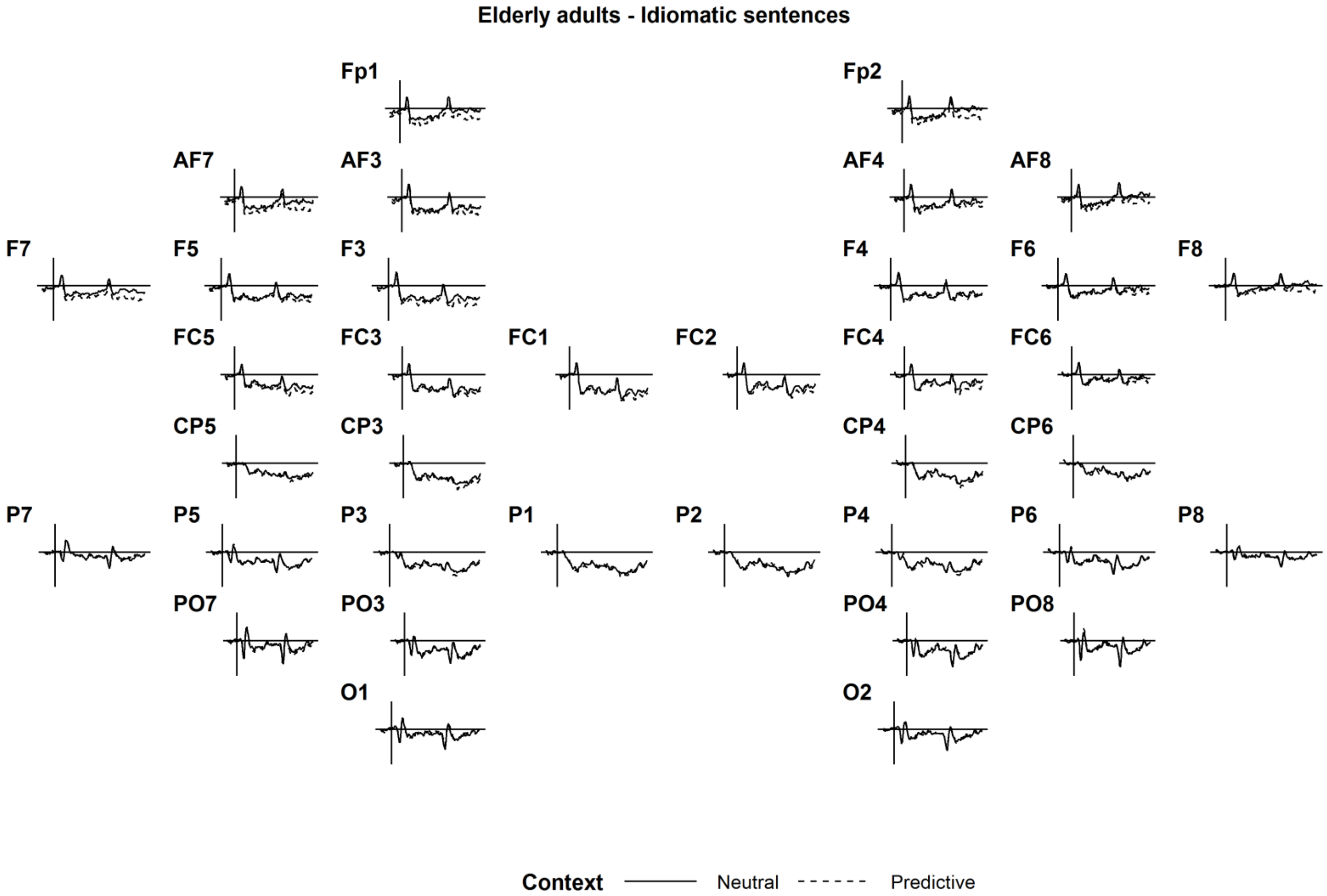

Figure A8. Elderly adults' grand average ERPs in the 500-800 ms time window in response to idiomatic sentences preceded by a neutral (solid line) versus a predictive context sentence (dotted line). Voltages are averaged across correct and incorrect sentences. Effects are displayed for each electrode included in the statistical analysis.

APPENDIX III

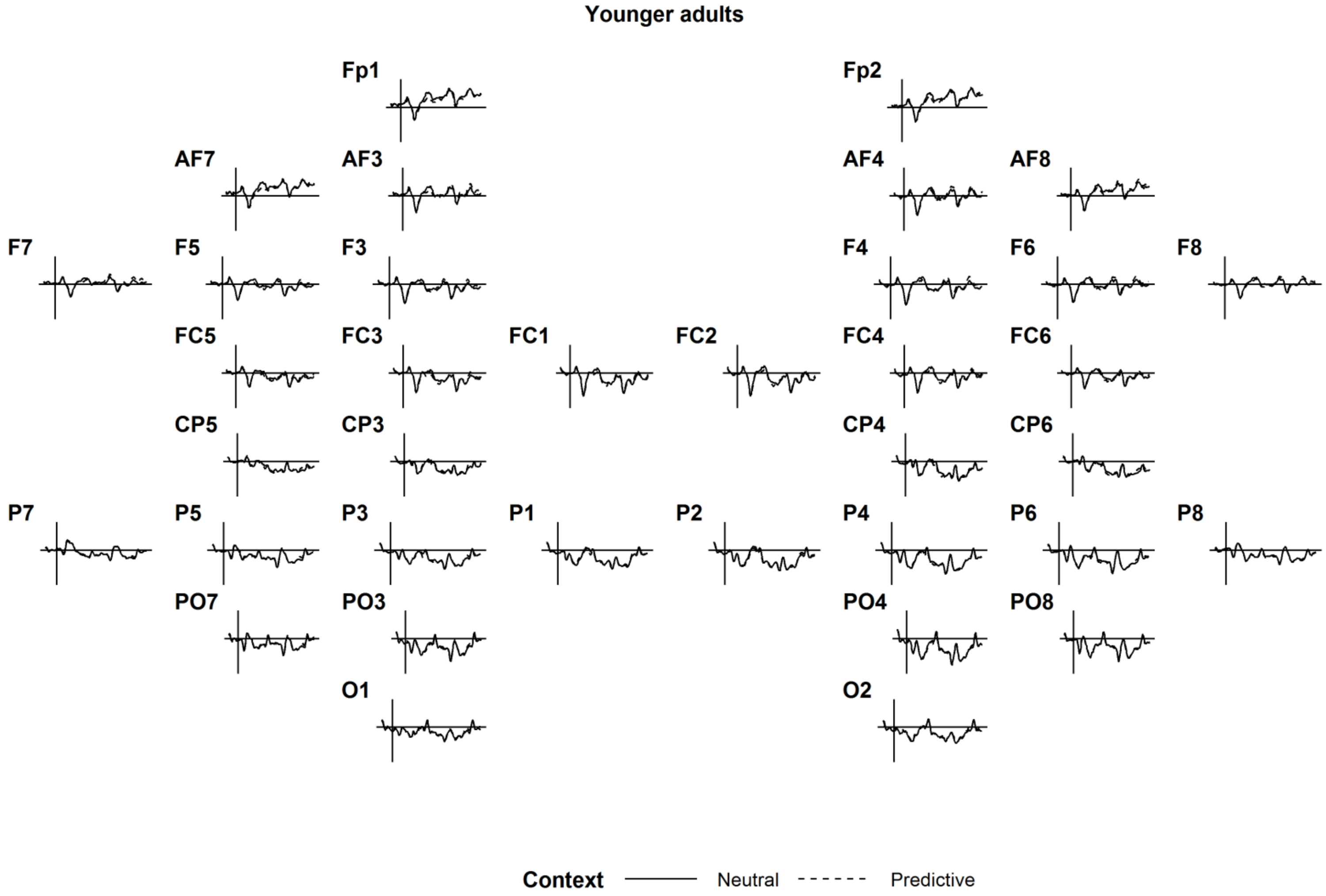

Figure A9. Younger adults' grand average ERPs in the 800-1100 ms time window in response sentences preceded by a neutral (solid line) versus a predictive context sentence (dotted line). Voltages are averaged across literal and idiomatic sentences and across correct and incorrect sentences. Effects are displayed for each electrode included in the statistical analysis.

APPENDIX III

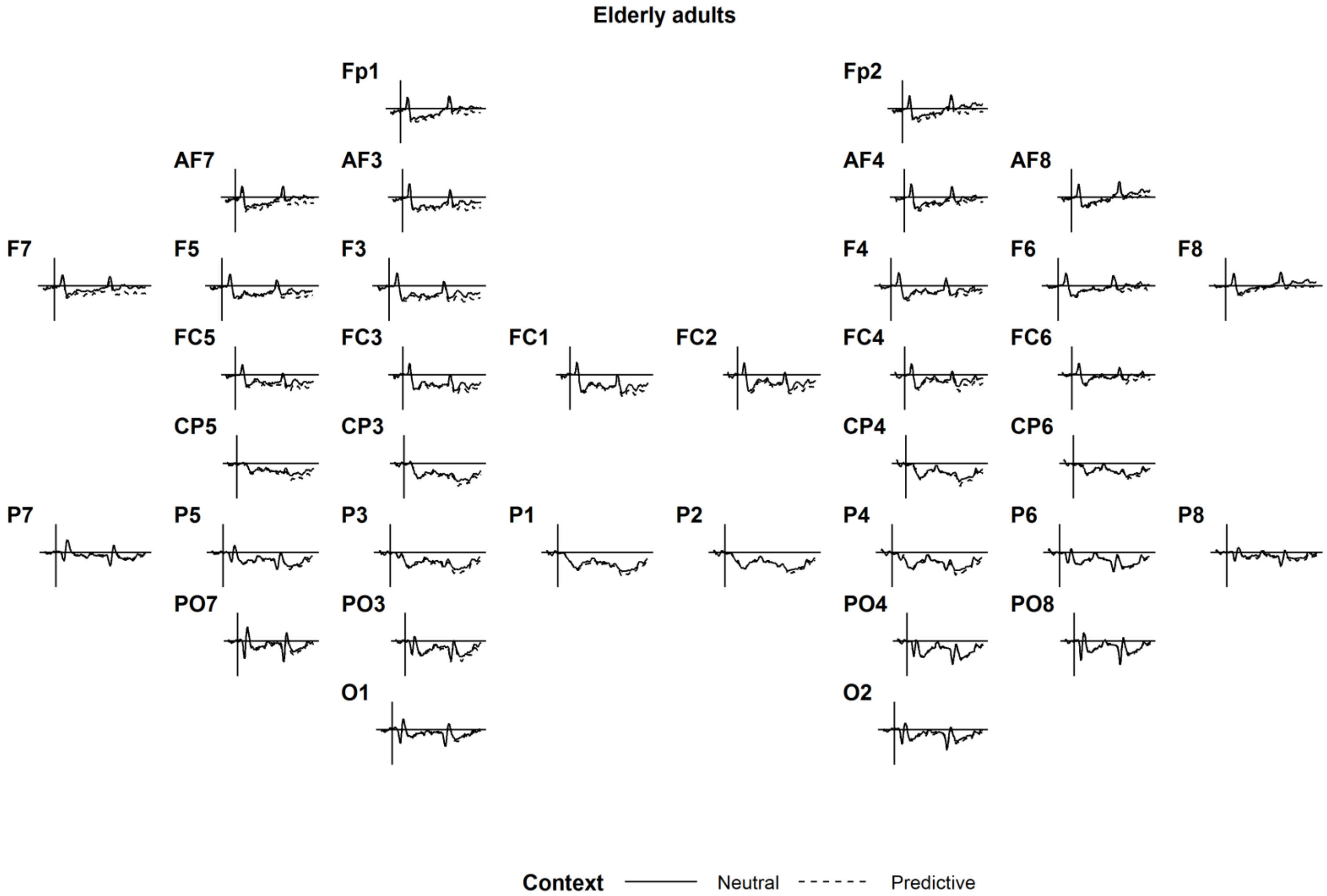

Figure A10. Elderly adults' grand average ERPs in the 800-1100 ms time window in response sentences preceded by a neutral (solid line) versus a predictive context sentence (dotted line). Voltages are averaged across literal and idiomatic sentences and across correct and incorrect sentences. Effects are displayed for each electrode included in the statistical analysis.

APPENDIX III

Table A2. Summary of coefficients of the linear mixed effects regressions models fitted to the data of the 500-800 ms time window and 800-1100 ms time window

Fixed effects		*500-800 ms*			**Fixed effects**		*800-1100 ms*		
Factor		**Estimate**	**SE**	***t*-value**	**Factor**		**Estimate**	**SE**	***t*-value**
1	Intercept	-0.84	0.36	-2.31	1	Intercept	-1.03	0.38	-2.67
2	Group (senior)	0.20	0.47	0.42	2	Group (senior)	0.78	0.50	1.57
3	Context (predictive)	-0.06	0.19	-0.29	3	Context (predictive)	-0.17	0.16	-1.03
4	Idiomaticity (idiom)	1.34	0.31	4.34	4	Idiomaticity (idiom)	1.38	0.20	6.90
5	Condition (incorrect)	0.73	0.20	3.72	5	Condition (incorrect)	0.84	0.20	4.26
6	Anteriority (posterior)	1.47	0.08	18.87	6	Anteriority (posterior)	1.70	0.08	20.49
7	Group (senior) * Context (predictive) * Idiomaticity (idiom)	-0.81	0.31	-2.60	7	Group (senior) * Context (predictive)	0.98	0.17	5.91
8	Verbal fluency	-	-	-	8	Verbal fluency	0.04	0.02	2.19

Random effects						**Random effects**					
Groups		**Factor**	**Variance**	**SD**	**Corr.**	**Groups**		**Factor**	**Variance**	**SD**	**Corr.**
9	Subject	Intercept	2.46	1.57		9	Subject	Intercept	2.91	1.71	
		Idiomaticity (idiom)	1.75	1.32	0.03			Idiomaticity (idiom)	1.65	1.29	-0.08
10	Critical Word	Intercept	1.88	1.37		10	Target	Intercept	1.99	1.41	
		Context (predictive)	2.59	1.61	-0.44			Context (predictive)	2.51	1.58	-0.48
11	Residual		49.92	7.07		11	Residual		56.63	7.53	