EFFECTS OF HEALTHY AGING AND GENDER ON THE ELECTROPHYSIOLOGICAL CORRELATES OF SEMANTIC SENTENCE COMPREHENSION: THE DEVELOPMENT OF DUTCH NORMATIVE DATA

Elissa-Marie Cocquyt[a]*, Emma Depuydt[b], Patrick Santens[c], Pieter van Mierlo[b], Wouter Duyck[d], Arnaud Szmalec[d,e] & Miet De Letter[a]

[a] Department of Rehabilitation Sciences, Ghent University, Belgium

[b] Department of Electronics and Information Systems, Medical Image and Signal Processing Group, Ghent University, Belgium

[c] Department of Neurology, Ghent University Hospital, Belgium

[d] Faculty of Psychology and Educational Sciences, Department of Experimental Psychology, Ghent University, Belgium

[e] Psychological Sciences Research Institute, Université catholique de Louvain, Louvain-la-Neuve, Belgium

* Corresponding author. Address: Department of Rehabilitation Sciences, Ghent University, Corneel Heymanslaan 10, B-9000 Ghent, Belgium. E-mail address: ElissaMarie.Cocquyt@UGent.be

Abstract

Purpose: The clinical use of event-related potentials in patients with language disorders is increasingly acknowledged. For this purpose, normative data should be available. Within this context, healthy aging and gender effects on the electrophysiological correlates of semantic sentence comprehension were investigated. **Method**: One hundred and ten healthy subjects (55 men and 55 women), divided among three age groups (young, middle-aged and elderly), performed a semantic sentence congruity task in the visual modality during electro-encephalographic recording. **Results**: The early visual complex was affected by increasing age as shown by smaller P2 amplitudes in the elderly compared to the young. Moreover, the N400 effect in the elderly was smaller than in the young, and was delayed compared to both middle-aged and young subjects. The topography of age-related amplitude changes of the N400 effect appeared to be gender specific. The late positive complex (LPC) effect was increased at frontal electrode sites from middle-age on, but this was not statistically significant. No gender effects were detected regarding the early P1, N1 and P2, or the LPC effect. **Conclusions**: Especially aging effects were found during semantic sentence comprehension, and this from the level of perceptual processing on. Normative data are now available for clinical use.

Conflict of Interest

The authors declare that there is no conflict of interest.

Keywords

Electrophysiology – semantic sentence comprehension – N400 – aging – gender – normative data

1. Introduction

1.1 Event-related potentials in a clinical context

Event-related potentials (ERPs) elicited during language paradigms are increasingly used in a variety of clinical populations, e.g. patients with aphasia (PWA), Alzheimer's disease (AD) and schizophrenia (SCZ). Because of their excellent temporal resolution, ERPs provide an objective measure of the sensory and cognitive sub-stages of phonological, semantic and grammatical processing. Each ERP component is characterized by its amplitude (magnitude of the electrical response), latency (timing of the electrical response) and topography (distribution of the electrical response among the scalp), which are all parameters sensitive to neural changes. As such, reduced amplitudes, increased latencies and/or topographical changes are often observed in PWA (Meechan et al., 2021), AD (Joyal et al., 2020) and SCZ (Wang et al., 2011). This electrophysiological information can be combined with results from standardized language batteries to obtain a complete picture of language capacities and, if applicable, to develop a personalized therapy protocol (McWeeny et al., 2020). Moreover, there is increasing evidence that ERPs 1) are sensitive to detect subtle differences in cognitive processing that are not detected by behavioral tests (Cocquyt et al, 2021; Stoodley et al., 2006), 2) can be used to monitor (therapy-induced) neuroplasticity (Cocquyt et al., 2020; Stalpaert et al., 2022) and 3) can be applied as biomarkers for early disease detection (AD: Horvath et al., 2018; SCZ: Du et al., 2015) or recovery (aphasia: Ehlers et al., 2015; Nolfe et al., 2006; Silkes & Anjum, 2021).

An implementation of ERPs as a clinical tool requires the availability of well-developed paradigms on the one hand and normative (i.e. reference) data on the other hand. Previous research showed that the parameters 'age' and 'gender' should be taken into account when developing normative data for linguistic ERP tasks. For the Flemish population, this research has been performed for phoneme perception (Aerts et al., 2013; 2015), semantic word priming (Cocquyt et al., 2022) and (grammatical) word order processing (Dorme et al., revised), revealing specific results for each aspect of language (cf. Table 1). However, this type of research remains lacking for semantic sentence comprehension. Since the latter is a crucial aspect in daily life communication, the current study aims to fill this specific research gap by investigating the effects of increasing age and gender on the electrophysiological correlates of semantic sentence processing.

1.2 Electrophysiological correlates of semantic sentence comprehension

In ERP studies, semantic sentence comprehension is often examined by means of a semantic sentence congruity task (SSCT), with or without behavioral verification. In the visual modality, multiple sentences are presented word by word in the center of a laptop or computer screen. Half of them are semantically correct (e.g. 'Bob buys bread at the bakery'), whereas the other half contain a semantic violation at the end (e.g. 'Jack has a bucket full of fever'). After the presentation of each sentence, participants can be asked to indicate whether or not it was semantically correct by means of a button press response. During an SSCT, the ERP component of utter interest is the **N400**. The N400 is a centro-parietal, negative-going waveform reaching its maximum amplitude around 400 ms after the onset of a critical written word (Kutas & Hillyard, 1982). Its amplitude is sensitive to the preceding context, i.e. semantic violations elicit larger responses in comparison to words that fit well within the prior sentence. Generally, the amplitude difference between the ERPs elicited during a semantically violated and correct condition is called the 'N400 effect'. Even though the N400 (effect) has been extensively described, its exact functional correlates remain debated. According to the 'memory retrieval view', the N400 amplitude reflects the effort with which the meaning of words is retrieved from semantic memory (Lau et al., 2008; Brouwer et al., 2012). The 'prediction account' states that healthy individuals make specific predictions regarding the continuation of a sentence. Whenever such a prediction is fulfilled, this will result in a reduced N400 amplitude (Delong et al., 2005). Finally, the 'integration account' refers to the ease with which a critical word is integrated within the prior context (Brown & Hagoort, 1993; Brown et al., 2000).

Before and beyond the N400, multiple other ERP components are elicited during an SSCT. Nonetheless, these components have been described to a much lesser extent. The visual **P1-N1-P2 complex** precedes the N400 and is a composite of temporally overlapping activations involving the (extra)striate cortex (Pratt, 2011). Previous ERP studies have shown that the N1 is the earliest component that is sensitive to the type of visual information, as its amplitude is larger for words (and word-like stimuli) than for symbols (Bentin et al., 1999; Maurer et al., 2005). In general, the N1 has been linked to orthographic processing, i.e. the perceptual expertise for recognizing letters and typical sequences of letters within written word forms (Maurer & McCandliss, 2008). The latter view is supported by significant correlations between N1 amplitudes and reading proficiency skills in children (Brem et al., 2013; Maurer et al., 2006). The P2 on the other hand, has been associated with perceptual matching. As summarized by Evans & Federmeier (2007), this entails the comparison of visual incoming stimuli to representations stored in memory (Luck & Hillyard, 1994; Nieuwland, 2019) or built from an accumulating sentence context (Federmeier et al., 2005a). Finally, a post N400 positivity, also termed **late positive complex/LPC or 'semantic P600'**, has been observed in multiple studies. Corresponding to the N400, the amplitude of the LPC is larger when elicited by semantic violations than by semantically correct words (i.e. LPC effect). Similarly to the syntactic P600 (Kuperberg, 2007), the LPC is assumed to index re-analysis (i.e. revising a problematic sentence), repair and/or updating (Brouwer et al., 2012; Kuperberg et al., 2020; for a review see Van Petten & Luka, 2012).

1.3 Healthy aging effects on semantic sentence comprehension

Healthy aging implies neuroanatomical (Gunning & Brickmann, 2010; Oschwald et al., 2020; Raz et al., 2005) and functional changes (Cabeza, 2002; Davis et al., 2008; Hoffman & Morcom, 2018; Reuter-Lorenz & Park, 2014; Spreng et al., 2010), and these have been shown to be related to cognitive performance. As such, older adults often show a decline in functions such as working memory, processing speed and task switching (Salthouse, 2010). However, language comprehension is assumed to remain relatively intact. At word level, lexical-semantic knowledge (including categorical and associative relations) is retained with increasing age as shown by behavioral (Giffard et al., 2003) and electrophysiological priming studies (Cocquyt et al., 2022). At sentence level, behavioral results are more heterogeneous since both lower (Xu et al., 2017; Zhu et al., 2018, 2019) and similar (or even higher) accuracy scores (Federmeier et al., 2010; Iragui et al., 1996) in elderly compared to young subjects have been reported. The latter heterogeneity is probably related to between-task differences in the required cognitive resources. Electrophysiological studies targeting semantic congruity effects (i.e. the comparison of ERPs elicited during a congruent and incongruent condition) in the visual modality reveal significantly reduced amplitudes (Federmeier et al., 2010; Gunter et al., 1992; Iragui et al., 1996; Kutas & Iragui, 1998; Xu et al., 2017; Zhu et al., 2018) and increased latencies of the N400 effect (Federmeier et al., 2010; Gunter et al., 1992; Iragui et al., 1996; Kutas & Iragui, 1998; Xu et al., 2017; Zhu et al., 2018, 2019) in older compared to young participants. Gunter et al. (1992) showed that these age-related changes were already present in middle-aged subjects (mean age: 56 years), which clearly emphasizes the need to investigate aging effects on language processing across a broad age range. Moreover, previous studies did not reveal healthy aging effects on the anterior-posterior or left-right distribution of the N400 congruity effect. Both young and older participants showed a dominant N400-effect at centro-parietal electrode sites which was characterized by a right hemispheric lateralization (Gunter et al., 1992; Iragui et al., 1996; Kutas & Iragui, 1998).

Although less investigated, the effects of healthy aging on the early components elicited during an SSCT seem to be rather limited. Previous studies reported significantly delayed latencies of the N1 (Kutas & Iragui, 1998) and P2 (Gunter et al., 1992), although the latter applied to the incongruent sentence condition only. Conversely, no significant effect of aging on the N1 and P2 amplitudes has been found (Gunter et al., 1992; Kutas et al., 1998). Finally, aging effects on the LPC effect elicited during an SSCT have been reported by Zhu et al. (2018, 2019) and Xu et al. (2017). The authors observed a significant LPC effect in elderly participants only. Even though this LPC effect is generally largest at parietal electrode sites when elicited during an SSCT (Van Petten & Luka, 2012), a topographical shift towards frontal electrode sites was present with increasing age (Xu et al., 2017, Zhu et al., 2019). The interpretation of this finding remains somewhat ambiguous: Zhu et al. (2018) found a significant positive correlation between the amplitude of the LPC effect and the behavioral accuracy of participants in the incongruent condition, whereas no significant correlations were found by Xu et al. (2017) and Zhu et al. (2019). Similarly, Federmeier et al. (2010) observed a frontal LPC effect in older adults only, but the authors remained inconclusive about the functional significance of this phenomenon.

1.4 Gender effects on semantic sentence comprehension

In addition to aging effects, differences between men and women in language processing have been a topic of major interest (Wallentin, 2009). Although anatomical differences between the linguistic cortex of men and women have been found, e.g. larger superior temporal volumes in women compared to men (Sowell et al., 2007), these findings should be interpreted with caution since men show greater variance in brain structures than women across the entire lifespan (Wierenga et al., 2022). At the functional level, however, men and women seem to use different cognitive strategies during specific verbal tasks. It has been suggested that women rely on more detailed semantic processing, and pay more attention to semantic features and relationships (Guillem & Mograss, 2005; Meyers-Levi & Sternthal, 1991). The latter view is strengthened by electrophysiological research with women showing larger N400 effects or shorter latencies compared to men during semantic word priming tasks in the auditory (Cocquyt et al., 2022; Daltrozzo et al., 2007) and visual modality (Wirth et al., 2007). At sentence level, similar findings (i.e. larger N400 effects in women than in men) have been observed during an auditory congruity (Daltrozzo et al., 2007) and a who-/what-question-answer paradigm. In the latter study (Wang et al., 2011), it was shown that women elicited significant N400 effects independent of external cues (i.e. context induced focus and accentuation of the critical word), whereas men only showed an N400 effect in case of a match between the context induced focus and accentuation. Hence, the authors concluded that women 'engaged in more elaborate semantic processing compared to men'. Contrastingly to findings on the N400 effect, the LPC effect was larger in men than in women during the auditory SSCT of Daltrozzo et al. (2007). To the best of our knowledge, gender effects on the ERPs elicited during an SSCT in the visual modality have not been investigated.

1.5 Research aims and hypotheses

The goal of the current study is to investigate healthy aging and gender effects on the electrophysiological correlates of semantic sentence comprehension. More precisely, the effects of age and gender on the amplitude and latency of different ERP components (P1, N1, P2, N400 and LPC) elicited during a visual SSCT will be investigated. Moreover, age- and gender-related differences in the amplitude, latency and topographical distribution of the N400 and LPC effects observed in the difference waveforms will be examined. This research is performed within the context of developing normative data for clinical purposes.

2. Methods

2.1 Participants

One hundred and ten healthy individuals (55 men and 55 women) between 21 and 84 years old participated in this study. All subjects were native speakers of Dutch, right-handed (according to the questionnaire of Van Strien, 1992) and without a neurological history. Based on self-report, all participants had intact or corrected vision. Cognitive abilities were screened by means of the Dutch version of the Montréal Cognitive Assessment (MoCA, Nasreddine et al., 2005). To be included, a minimum score of 26/30 was required. Two subtests of the Semantic Association Test (SAT, Visch-Brink et al., 2005) were used to evaluate linguistic (semantic) abilities, i.e. naming and verbal (written word) association. All participants reached a minimum score of 25/30, confirming intact semantic processing skills.

To investigate aging and gender effects, the participants were clustered into three age groups, i.e. the young (20-39 years, n=40), middle-aged (40-59 years, n=40) and elderly (≥ 60 years, n=30) (following Karayanidis et al., 1993 and Kutas et al., 1994). Each age group consisted of an equal number of male and female subjects. The demographic characteristics of the subjects, as well as the MoCA and SAT scores, can be found in Table 2. This study was approved by the Ethical committee of the University Hospital Ghent and all participants signed an informed consent. As a reward, they received a small gift at the end of the experiment.

[Insert Table 2 here]

*2.2 Experimental procedure*The paradigm was an SSCT consisting of 120 sentences. Half of the sentences were semantically correct (e.g. ‘The carpenter got a compliment from his boss’), whereas the other half contained a semantic violation at the end (e.g. ‘The plastic bucket is full of fever’). The stimuli were adapted from Swaab, Brown and Hagoort (1997) and presented in the visual modality. The syntactic structure of the sentences was matched between both conditions. Moreover, semantically correct and incorrect sentences were similarly constraining as shown by the results from a cloze probability test (correct condition: 60.3%, range: 29.9% – 100%, incorrect condition: 54.6%, range: 29.2% - 100%; see Swaab, Brown and Hagoort, 1997 for a detailed description). The average sentence length was 8.0 words (correct condition: 8.2 words, incorrect condition: 7.9 words) and most of the sentences were active (active: n=114, passive: n=6). To be sure that observed differences between conditions were related to the semantic fit of the final noun, these nouns were matched on a number of psycholinguistic variables across conditions (Table 3): word frequency (SUBTLEX-NL, Keuleers et al., 2010), orthographic length, number of orthographic neighbors (CLEARPOND, Marian et al., 2012), concreteness (Brysbaert et al., 2014), imageability (Van Loon-Vervoorn, 1985), age of acquisition, valence, arousal, and dominance (Moors et al., 2013).

[Insert Table 3 here]

Participants were comfortably seated in front of a Dell laptop screen. They were instructed to read the sentences internally and to judge their semantic plausibility. The 120 sentences were presented at random by means of E-Prime 3 (Psychology Software Tools, Pittsburgh, PA) and divided among seven blocks with pauses in between. The inter-trial interval was 1500 ms. Before the start of the actual experiment, participants were familiarized with the test procedure during a practice block consisting of eight sentences (half of them were correct). These sentences were not used in the analyses.

Each trial started with a fixation cross in the middle of the screen for 1500 ms. The sentences appeared word by word (white text color on a black background) with each word being displayed for 500 ms, followed by a black screen for 500 ms. A semantic plausibility judgment had to be made after reading the final word of each sentence, when the Dutch word 'Druk' ('press') appeared. Participants had to press a green (correct condition) or red button (incorrect condition) with their left or right index finger. The button press response was delayed (i.e. minimum 1 second after the presentation of the final word), such that action-related potentials could not contaminate the ERPs of experimental interest (Van Vliet et al., 2014). The allocation of the green and red button was switched for half of the men and women within each age group. The button press accuracy scores were logged automatically in E-Prime 3.

2.3 EEG recording

Continuous EEG was recorded from 32 Ag/AgCl scalp electrodes mounted in Easycap electrode caps (Brain Products, Munich, Germany). The electrode sites were the following: Fp1/2, Fpz, F3/4, F7/8, Fz, FC1/2, FC5/6, C3/4, T7/8, Cz, CP1/2, CP5/6, P3/4, Pz, TP9/10, POz, O1/2 and Oz. During the recording, all electrodes were referenced to FCz, whereas AFz was used as the ground electrode. We aimed to keep all electrode impedances below 10 kΩ by using an abrasive electrolyte gel (Abralyt 2000, Easycap). The continuous EEG data were amplified using a BrainAmp DC/MR plus (Brain Products, Munich, Germany) and digitized with a sample rate of 500 Hz.

2.4 ERP data analysis

MNE-Python (v1.0.0, Gramfort et al., 2013) was used to analyze the recorded EEG data offline. The first step was to remove the practice block. Second, the continuous EEG data were band-pass filtered using an infinite impulse response filter (i.e. zero phase shift Butterworth filter) with half-amplitude cut-off frequencies of 0.3 and 30Hz and a slope of 12 dB/octave. Also, a notch filter of 50Hz was applied to remove power line noise. Components related to eye blinks and horizontal eye movements were automatically isolated and removed by means of Independent Component Analysis (ICA). The data were re-referenced to the average of the left (TP9) and right mastoids (TP10) to avoid the presence of a hemispheric asymmetry in the reference (Duncan et al., 2009). Subsequently, the responses elicited by the critical nouns (i.e. the last word in each sentence) were segmented into epochs of 300 ms pre-stimulus and 1200 ms post-stimulus, for the semantically correct and incorrect trials separately. Only trials that were evaluated accurately (button press response), were included for further analyses. All epochs were baseline corrected, using the 300 ms pre-stimulus period. Next, epochs with artefacts were automatically rejected using the following criteria: 1) a gradient criterion of 75 µV (the absolute difference between two adjacent samples cannot exceed 75 µV), 2) a maximal allowed difference of 150 µV between the minimum and maximum voltage in intervals of 200 ms, 3) a minimal and maximal allowed amplitude of -75 and 75 µV respectively and 4) a low activity criterion of 0.5 µV during 100 ms (i.e. the difference between the minimum and maximum voltage cannot be less than 0.5 µV during a time window of 100 ms). Finally, average waveforms were computed for the semantically correct and incorrect condition separately. In all participants, these averages were based on more than 75% of the presented trials (Table 4). In addition, difference waves were created by subtracting the average ERP to the semantically correct condition from the average ERP elicited during the incorrect condition.

[Insert Table 4 here]

Amplitude and latency values were extracted in component specific time windows at nine electrode positions (F3, Fz, F4, C3, Cz, C4, P3, Pz and P4, consistent with Cocquyt et al., 2022). The time windows were determined through the collapsed localizers approach, i.e. using the topography of the average waveform across both conditions and all participants (Luck & Gaspelin, 2017). For the correct and incorrect condition, mean amplitudes and 50% fractional area latencies (Luck, 2014) were obtained in the following time windows: P1 (120-170 ms), N1 (170-230 ms), P2 (230-300 ms), N400 (300-500 ms) and LPC (500-900 ms). In addition, difference waves (i.e. a subtraction of the average ERPs to the semantically correct condition from the average ERPs elicited during the incorrect condition) were created to investigate the N400 and LPC effects. The amplitude and latency of the N400 and LPC effects were defined as the mean amplitude and 50% signed (N400: negative, LPC: positive) area latency within their corresponding time windows (N400: 300-500 ms, LPC: 500-900 ms).

2.5 Statistical analyses

IBM SPSS statistics (version 28) was used to perform all statistical analyses. The 0.05 alpha level was chosen to determine statistical significance. First, effects of aging and gender on the behavioral accuracy of participants (i.e. number of correct button press responses) during the SSCT was examined by a univariate analysis of variance (ANOVA) with age group and gender as independent variables.

Second, mean amplitudes of the early ERP components (P1, N1 and P2), the N400 and LPC across the nine electrode positions (F3, Fz, F4, C3, Cz, C4, P3, Pz and P4) were subjected to a repeated measures ANOVA with condition (semantically correct versus incorrect) as a within-subject variable and age group and gender as between-subject variables. For the N400 effect, the within-subject factor condition did not apply, but the within-subject factors region (frontal, central and parietal) and lateralization (left, middle and right) were added to examine possible (aging or gender related) differences in the topographical distribution. For the LPC effect, a similar analysis was performed, but to replicate results from previous studies (Xu et al., 2017; Zhu et al., 2018, 2019) only the within-subject factor region (frontal, central and parietal) was targeted.

Third, the latency of the P1, N1 and P2 across the nine electrode positions (F3, Fz, F4, C3, Cz, C4, P3, Pz and P4), of the N400 across the centroparietal electrode sites (C3, Cz, C4, P3, Pz and P4) and of the LPC across the parietal electrodes (P3, Pz and P4)[1] were subjected to a repeated measures ANOVA with condition as a within-subject variable and age group and gender as between subject variables. For all repeated measures, inhomogeneous covariances (as determined by Mauchly's test) were compensated for by applying the 'Greenhouse-Geisser' (GG) or 'Huynh-Feldt' (HF) correction (epsilon < 0.75 or > 0.75 respectively; Verma, 2015) if there were more than two levels of a within-subject factor. In this case, the adjusted *p*-values, unadjusted degrees of freedom and GG or HF epsilon values were reported. Finally, latencies of the centroparietal N400 effect and parietal LPC effect were investigated by a univariate ANOVA with age group and gender as independent variables. Significant main effects were investigated by post hoc multiple comparisons with a Bonferroni correction, whereas the nature of significant interactions was explored by Bonferroni corrected simple main effects (Field, 2009).

[1] Latency values were averaged across C3, Cz, C4, P3, Pz and P4 for the N400 (effect) and across P3, Pz and P4 for the LPC (effect) as these components are typically largest at centroparietal and parietal electrode sites respectively during semantic sentence comprehension tasks (Kutas & Federmeier, 2011; Van Petten & Luka, 2012).

3. Results

3.1 Behavioral results

Table 5 provides an overview of the mean accuracy scores per condition for men and women of each age group. No significant main effects of age group ($F(2, 104) = 0.07$, $p > 0.05$) and gender ($F(1, 104) = 3.59$, $p > 0.05$), nor a significant interaction between both variables ($F(2, 104) = 0.84$, $p > 0.05$), could be detected regarding the button press accuracy scores (total). The same findings apply to the accuracy scores in the correct (main effect of age group: $F(2, 104) = 0.67$, $p > 0.05$; main effect of gender: $F(1, 104) = 3.33$, $p > 0.05$; age group x gender: $F(2, 104) = 0.56$, $p > 0.05$) and incorrect condition (main effect of age group: $F(2, 104) = 0.36$, $p > 0.05$; main effect of gender: $F(1, 104) = 0.90$, $p > 0.05$; age group x gender: $F(2, 104) = 0.84$, $p > 0.05$) separately.

[Insert Table 5 here]

3.2 Electrophysiological results

The grand average waveforms elicited during by the critical words during the correct and incorrect condition within each age group are presented in Figure 1. The target words elicited an early visual complex consisting of a C1, P1, N1 and P2. In this paper, the C1 component is not of interest and will not be further discussed. Beyond the P2, an N400 and post-N400 positivity (LPC) could be observed.

[Insert Figure 1 here]

3.2.1 P1-N1-P2

Table 6 provides an overview of the statistical results regarding the early components. The corresponding normative data can be found in Appendix A. P1 amplitudes were significantly higher in the incorrect than in the correct condition (main effect of condition: $F(1, 104) = 4.67$, $p < 0.05$), whereas for the N1 there was no significant difference between conditions. For both components, neither age group nor gender had a significant effect on the amplitude. Regarding the P2, mean amplitudes were higher in the correct than incorrect condition (main effect of condition: $F(1, 104) = 28.05$, $p < 0.001$). This difference was larger in the young and middle-aged group compared to the elderly group (condition by age group interaction: $F(2, 104) = 4.54$, $p < 0.05$). Moreover, the main effect of age group was significant ($F(2, 104) = 3.80$, $p < 0.05$). Post hoc pairwise comparisons revealed significantly lower P2 amplitudes in the elderly compared to the young group (mean difference: 1.4 µV, 95% CI: 0.1 – 2.7, see Figure 1).

P1, N1 and P2 latencies did not statistically differ between the correct and incorrect condition. Also, no significant age group or gender effects were detected.

[Insert Table 6 here]

3.2.2 N400 (effect)

The statistical results and normative data can be found in Table 7 and Appendices B (amplitude and latency) and C (topography) respectively. Across conditions, there were no significant effects of age group and gender on the N400 amplitude and latency. As expected, N400 amplitudes were significantly larger for the incorrect than for the correct condition (main effect of condition: $F(1, 104) = 220.21$, $p < 0.001$). The interaction between condition and age group was significant ($F(2, 104) = 8.10$, $p < 0.001$) and even though simple main effects showed that a significant effect of condition (i.e. the N400 effect) was present in each of the three age groups ($p < 0.001$), the between condition difference became smaller with increasing age. Moreover, the latency of the N400 was significantly shorter for the correct than for the incorrect condition (main effect of condition: $F(1, 104) = 8.10$, $p < 0.01$), independent of the age group or gender.

Although there was an N400 effect in each age group (observed in the difference waveforms), its amplitude was modulated by increasing age (Figure 2, Appendix D), as reflected by the significant main effect of age group ($F(2, 104) = 8.10$, $p < 0.001$). Post hoc pairwise comparisons showed that the elderly had a significant smaller N400 effect in comparison to the young subjects (mean difference: 1.2 µV, 95% CI: 0.5 - 2.0). There was no significant main effect of gender, nor a significant interaction between age group and gender (Figure 4, Appendix E). Similar to findings on the amplitude, the latency of the N400 effect was subjective to aging as shown by the significant main effect of age group ($F(2, 104) = 10.15$, $p < 0.001$). Post hoc pairwise comparisons revealed that the N400 effect was significantly delayed in elderly compared to both the young (mean difference: 21 ms, 95% CI: 10-32 ms) and middle-aged subjects (mean difference: 13 ms, 95%CI: 2-24 ms). Regarding the topographical distribution of the N400 effect (Figure 3), a significant main effect of region ($F(2, 208) = 18.88$, $p < 0.001$, GG epsilon = 0.61) and lateralization ($F(2, 208) = 26.37$, $p < 0.001$, HF epsilon = 0.84) was found as well a significant region by lateralization interaction ($F(4, 416) = 11.04$, $p < 0.001$, HF epsilon = 0.87). Simple main effects revealed that the N400 effect was significantly larger at central ($p < 0.001$) and parietal electrodes ($p < 0.05$) than at frontal sites. Moreover, the N400 effect was significantly larger at the midline ($p < 0.001$) and right electrode positions ($p < 0.001$) in comparison to left electrode positions. The amplitude of the midline and right N400-effect did not statistically differ. The latter applied to centroparietal sites only since the frontal N400 effect was significantly larger at the right than at the midline ($p < 0.001$). Finally, a significant three way interaction between region, age group and gender ($F(4, 208) = 3.53$, $p < 0.05$, GG epsilon = 0.61) showed that young women had significantly smaller N400 effects compared to elderly women at frontal, but not at central and parietal electrode sites. Whereas in men, the opposite result was observed, i.e. significantly smaller N400 effects in young men compared to elderly men at central and parietal, but not at frontal electrode sites.

[Insert Figures 2, 3 and 4 here]

3.2.3 Late positive complex (effect)

Across conditions, the LPC amplitude was not affected by age group, but it was larger in women than in men (mean difference: 0.9 µV, 95% CI: 0.1 – 1.7) as shown by the main effect of gender ($F(1, 104) = 4.93$, $p < 0.05$). For the LPC latency across conditions, there were no significant between-gender differences, but the main effect of age group ($F(2, 104) = 9.23$, $p < 0.001$) revealed a significantly higher latency in the elderly compared to the young group (mean difference = 34 ms, 95% CI: 15-53 ms).

Regarding between-condition differences, the main effect of condition ($F(1, 104) = 30.51$, $p < 0.001$) revealed significantly larger LPC amplitudes for the incorrect than for the correct condition (i.e. LPC effect), independent of the age group or gender. Parallel with the N400 latency, the latency of the LPC was significantly shorter for the correct than for the incorrect condition (main effect of condition: $F(1, 104) = 8.99$, $p < 0.01$), in both young, middle-aged and elderly men and women.

The topographical distribution of the LPC effect (i.e. difference wave) was characterized by a parietal maximum as shown by the main effect of region ($F(2, 208) = 18.09$, $p < 0.001$, GG epsilon = 0.62). Visual inspection suggested a larger LPC effect in the elderly and middle-aged compared to the young, and this especially at frontal electrode sites (Figure 3). However, this was not borne out by a significant main effect of age group ($p = 0.09$) or region by age group interaction ($p = 0.2$). In addition, no significant effect of gender could be detected (Figure 4). Finally, the latency of the LPC effect did not statistically differ between age groups, or between male and female subjects. The statistical findings and normative data are reported in Table 7 and Appendices B (amplitude and latency) and C (topography) respectively.

[Insert Table 7 here]

4. Discussion

The current study aimed to investigate healthy aging and gender effects on the electrophysiological correlates of semantic sentence comprehension in the visual modality. The results showed that aging has an important effect on multiple ERP components elicited during an SSCT, whereas between-gender differences are minimal. This research occurred within the context of developing normative data for subsequent clinical use in patients with brain damage.

4.1 Aging effects on semantic sentence comprehension

In contrast to the absence of significant aging effects on the early **P1** and **N1** components, the amplitude of the **P2** was reduced in elderly compared to young adults. Although significant P2 reductions have not been observed in previous studies using an SSCT (Gunter et al., 1992; Kutas & Iragui, 1998), the authors reported on the difficulty to examine the P2 statistically due to overlap with subsequent ERPs (Kutas & Iragui, 1998). Interestingly, our findings do correspond to Federmeier et al. (2005b) and Wlotko & Federmeier (2012). Their tasks, consisting of semantically congruent sentences presented in either two constraint conditions (strong versus weak, Federmeier et al., 2005b) or across a broad range of cloze probabilities (Wlotko et al., 2012), revealed significantly smaller, frontal P2 responses in the older subjects than in the young. Our study expands these findings by showing that age-related reductions of the P2 occur for both semantically congruent and incongruent conditions, and are not limited to frontal electrode positions. Together, these results suggest that aging effects during semantic sentence comprehension start at the level of perceptual matching, i.e. the comparison of incoming information to representations built from the available context. However, we should consider the possibility of a partial overlap between the P2 and N400 (see Nieuwland, 2019) since the latter can start as early as 250 ms after the onset of a critical word (Hagoort et al., 2003). This could also explain why P2 amplitudes were more positive (or less negative in terms of the N400) in the semantically correct condition. Moreover, all the participants reported normal or corrected vision, but this was not properly evaluated, which can be seen as a limitation of this study. Even though no behavioral accuracy differences between age groups were present and no aging effects were detected for the P1 and N1 components, age related changes in visual word processing (e.g. Finnigan et al., 2011) might have contributed to the observed P2 amplitude reduction (as well as to the N400 effect amplitude reduction and latency increase) in the elderly.

A significant **N400 effect** was present in each of the three age groups. However, with increasing age, there was a clear amplitude reduction, leading to a significantly smaller N400 effect in older compared to young adults. In women, this was observed at frontal electrode sites and, in men, at centroparietal sites. Based on the behavioral findings (button press accuracy scores), no aging related decline in semantic sentence comprehension could be detected. However, ERPs revealed (gender specific) age related neuroplasticity and, possibly the use of compensation mechanisms (see 4.1, LPC effect). A higher sensitivity of ERPs to detect (subtle) deficits has been suggested in previous research (Cocquyt et al, 2021; Stoodley et al., 2006), hence, this cannot be ruled out in the elderly group. Decreased amplitudes of the N400 effect in older adults correspond to the (fMRI-based) findings of Hoffman & Morcom (2018). In their meta-analysis on age-related neural changes during semantic processing, a reduced activation of (among others) the left inferior frontal gyrus and the posterior temporal cortex (including the medial temporal gyrus) was found, regions that have been postulated as potential generators of the N400 (Halgren et al., 1994; Johnson & Hamm, 2000; Lau et al., 2008; Van Petten & Luka, 2006). A significantly decreased N400 effect in older adults has been observed in previous experiments with both low (Federmeier et al., 2010; Iragui et al., 1996; Kutas & Iragui, 1998) and higher working memory demands (Xu et al., 2017; Zhu et al., 2018, 2019), and were generally interpreted as a decline in semantic processing (Xu et al., 2017), or more specifically as a decreased/less efficient reliance on prediction mechanisms during language comprehension (Federmeier et al., 2010; Zhu et al., 2018) or poor integration of information within a semantic context (Iragui et al., 1996; Kutas &

Iragui, 1998; Zhu et al., 2018, 2019), corresponding to the prediction (Delong et al., 2005) and integration account (Brown & Hagoort, 1993; Brown et al., 2000) to explain the underlying mechanisms of the N400 (effect). Since the current paradigm was not designed to specifically target one of these accounts, both declined prediction and integration abilities could have contributed to the reduced N400 effect in the elderly. Remarkably, Gunter et al. (1992) observed amplitude reductions of the N400 effects from middle age on, but this was not confirmed in our study. A possible explanation for this discrepancy could be the age range under investigation. Gunter and colleagues included participants between 50 and 67 years, whereas 'middle-aged' corresponded to subjects between 40 and 59 years in our study. Moreover, Gunter et al. matched their participants regarding verbal IQ, but no cognitive screening was performed. In our study, we used a rather strict cutoff score on the MoCA, namely 26/30 as recommended by Nasredinne et al. (2005). More recently, it was shown that a cutoff score of 23/30 is more optimal in terms of specificity and diagnostic accuracy (Carson et al., 2018). Hence, our participants can be considered as cognitively high performers, which could also have contributed to the observed difference with the results from Gunter et al. (1992).

Regarding the latency of the N400, shorter latencies for the correct than for the incorrect condition were present in the three age groups. The same result was observed for the LPC occurring later in time. These findings show that words that fit within the prior context are processed more efficiently, and this independent of age. Contrarily, the latency of the N400 effect was significantly delayed in the elderly compared to both the young and middle-aged subjects. The latter finding is in line with previous research (Federmeier et al., 2010; Iragui et al., 1996; Kutas & Iragui, 1998; Xu et al., 2017; Zhu et al., 2018), although peak latency measures were consistently used in these studies to quantify the timing of the N400 effect. Currently, it is well known that ERP peaks are not very reliable as they are highly influenced by the amount of included trials and the presence of noise. Therefore, fractional area latencies provide a better alternative (Luck, 2014). In addition to peak measures, Iragui et al. (1996) also examined aging effects on the 50% fractional signed (negative) area latency in the time-window 200-800 ms, and found significantly higher latencies in older compared to young subjects as well. Our findings in the time-window 300-500 ms correspond to theirs. Moreover, the observed delay cannot be explained by a delay of earlier processes, as we could not find a significant delay in the P1, N1 and P2 latencies. The latter is in contrast with the observations by Kutas et al. (1998) and Gunter et al. (1992), who found a significant delay of the N1 and P2 (incongruent condition) respectively. These authors extracted peak latencies at different electrode sites (N1: occipital, P2: frontal) and in slightly different time-windows (N1: 100-200ms, P2: 200-300 ms), which could explain our divergent results.

Interestingly, Federmeier & Kutas (2005b) stated that individual differences in the timing of the N400 effect in older adults could be linked to between-subject differences in executive abilities. In more detail, the authors found that the N400 effect peak latency was negatively correlated with the reading span, such that older adults with higher reader spans had a smaller delay of the N400 effect. We were not able to (dis)confirm these findings since working memory capacities were not explicitly assessed in this study. To gain insights into inter-individual differences in the timing (and size) of the N400 effect, future research should include a proper neuropsychological test battery targeting multiple executive abilities (e.g. inhibition, attention and working memory). Executive test scores should then be correlated with more reliable ERP parameters (e.g. mean amplitudes and fractional area latencies).

Regarding the topographical distribution, the N400 effect was characterized by a centroparietal maximum and was largest over the right hemispheric electrode sites, which corresponds to previous studies using a semantic congruity paradigm in the visual modality (Gunter et al., 1992; Hagoort et al., 2003; Kutas & Hillyard, 1982). Remarkably, no significant main effect of aging on the topography of the N400 effect was found, which is also in line with previous ERP research (Gunter et al., 1992; Iragui et al., 1996; Kutas & Iragui, 1998). In the context of numerous descriptions of age-related changes in semantic neural networks (Hoffman & Morcom, 2018), this result may seem rather odd. However, the latter observations mainly stem from spatial imaging techniques, and such insights are not obtained by basic ERP research.

Finally, although not statistically significant, an interesting observation in this study was the increase of the **LPC effect** in middle-aged and elderly participants compared to the young, and this especially at frontal electrode sites. The latter pattern has been observed multiple times in previous studies (Federmeier et al., 2010; Xu et al., 2017; Zhu et al., 2018, 2019), but a clear interpretation remains lacking. The LPC is generally linked to re-analysis, repair and/or updating (Kuperberg et al., 2020; Van Petten & Luka, 2012). Although it remains debated whether a posterior to anterior shift of neural activity is a compensatory or maladaptive mechanism (see Grady, 2012 for a review), the observed increase of the (frontal) LPC effect in our study could be compensatory in nature, as along significantly reduced amplitudes of the P2 and N400 effect, no significant differences between the behavioral accuracy scores (button press responses) of young, middle-aged and elderly subjects were found. To confirm this hypothesis, future research should include correlations between accuracy scores and LPC effect amplitudes, as the currently available findings are contradictory (Xu et al., 2017; Zhu et al., 2018, 2019). The latter analysis was not performed in this study due to the (very) high accuracy in all subjects (Table 5). Moreover, a comparison of LPC effects elicited by (a similar amount of) correctly and incorrectly judged trials would be beneficial to gain insights on this matter.

4.2 Gender effects on semantic sentence comprehension

Compared to aging effects, research towards the effects of gender on semantic processing is more limited. At the behavioral level, women and men often obtain similar accuracy scores (e.g. Cocquyt et al., 2022), parallel with the observations in this study. However, previous studies reported on significantly larger or earlier **N400 effects** in women than in men, elicited during auditory (Cocquyt et al., 2022; Daltrozzo et al., 2007) and visual word priming (Wirth et al., 2007), as well as during an auditory sentence congruity (Daltrozzo et al., 2007) and a who-/what-question-answer paradigm (Wang et al., 2011). Differences in cognitive strategy, i.e. more attention to semantic features in women than in men, have been premised as a possible explanation. Contrastingly, our sentence congruity task in the visual modality revealed no significant main effect of gender regarding the size or timing of the N400 effect. Although it is difficult to draw conclusions from our null result, several variables might have influenced our result. First of all, a different input modality and differences in the ERP analysis (e.g. chosen time-windows) could be possible explanations for our deviating result. Second, it is remarkable that three of the four previous experiments (Daltrozzo et al., 2007; Wang et al., 2011; Wirth et al., 2007) did not require a behavioral reaction, like the congruity evaluation (button press response) in our study. More precisely, all our participants were explicitly instructed to judge the semantic congruity of each sentence, which might have led to an increase of attention to semantic properties in men, possibly reaching a similar level as women. Apparently, to reach this goal, the task should require a relatively high amount of cognitive resources (e.g. working memory, attention) as well since Cocquyt et al. (2022) observed significantly larger N400 effects during associative word priming in women than in men, even though this task included a semantic relatedness evaluation. This should, however, be affirmed in future studies. Hence, whereas women spontaneously seem to engage in elaborated/detailed semantic processing, the used experimental stimuli and presence of explicit task instructions may have an important influence in men.

Lastly, we observed larger LPC amplitudes in women than in men across both conditions, but no significant influence of gender was found on the **LPC effect**. Conversely, in one previous study using an auditory sentence congruity task (Daltrozzo et al., 2007), men displayed larger LPC effects compared to women, due to men's higher suppression of the LPC during congruent sentences. Although the findings of Daltrozzo and colleagues should be interpreted with some caution because their sample size was smaller (i.e. 10 men and 10 women), the absence of task demands accompanied smaller N400 effects in men than in women, which could have led to a higher need for repair or updating (as reflected by the LPC effect) in men. Again, a replication of this finding in future studies is needed.

4.3 Clinical added value of normative electrophysiological data for semantic sentence comprehension

In this study, healthy aging and gender related differences in the electrophysiological correlates of semantic sentence comprehension were examined in the context of developing normative data for clinical purposes. In future research, the current SSCT will be used in patients with stroke-related aphasia. Significant correlations between the N400 congruity effect and the severity of language comprehension deficits can be expected (Chang et al., 2016; Kawohl et al., 2010). Importantly, our paradigm has already been used successfully in patients with primary progressive aphasia (PPA). More precisely, the elicited N400 effect revealed additional insights on top of the behavioral language tests, and helped to differentiate patients with the nonfluent/agrammatic PPA variant from patients with the logopenic or semantic variant (Stalpaert et al., 2021). Moreover, the N400 effect seems to be sensitive to therapy-induced neuroplastic changes (Stalpaert et al., 2022). Both studies of Stalpaert and colleagues support the use of ERPs (including the N400) during the diagnostic evaluation and follow-up of semantic processing abilities. The availability of normative data will facilitate the interpretation of ERP parameters in patients.

5. Conclusion

This study shows that especially aging has an important effect on the electrophysiological correlates of semantic sentence processing. Whereas no behavioral deterioration could be detected, amplitudes of the P2 and N400 effect were smaller in elderly than in young subjects. Moreover, the N400 effect was significantly delayed in older compared to middle-aged and young adults. Future research should target whether an increased LPC effect is present in middle-aged and older adults and whether it is compensatory in nature. Finally, the availability of normative data will be an added value to map and monitor language functions in patients with brain damage.

Acknowledgements

This study has been supported by the Special Research Fund (BOF) and the Industrial Research Fund (IOF) of Ghent University. We would like to express our gratitude to Professor Peter Hagoort (Radboud University, Nijmegen, The Netherlands) who gave permission to use and adapt his ERP paradigm.

Data availability statement

The datasets generated and analyzed during the current study are not publicly available due to an ongoing patent application.

References

Aerts, A., van Mierlo, P., Hartsuiker, R. J., Hallez, H., Santens, P., & De Letter, M. (2013). Neurophysiological investigation of phonological input: Aging effects and development of normative data. *Brain and Language*, 125(3), 253-263.

Aerts, A., van Mierlo, P., Hartsuiker, R. J., Santens, P., & De Letter, M. (2015). Sex differences in neurophysiological activation patterns during phonological input processing: an influencing factor for normative data. *Archives of sexual behavior*, 44(8), 2207-2218.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M. H., Echallier, J. F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. *Journal of Cognitive Neuroscience*, 11(3), 235-260.

Brem, S., Bach, S., Kujala, J. V., Maurer, U., Lyytinen, H., Richardson, U., & Brandeis, D. (2013). An electrophysiological study of print processing in kindergarten: the contribution of the visual N1 as a predictor of reading outcome. *Developmental neuropsychology*, 38(8), 567-594.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: rethinking the functional role of the P600 in language comprehension. *Brain research*, 1446, 127-143.

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

Brown, C. M., Hagoort, P., & Kutas, M. (2000). Postlexical integration processes during language comprehension: Evidence from brain-imaging research. In The new cognitive neurosciences (pp. 881-895). MIT Press.

Brysbaert, M., Stevens, M., De Deyne, S., Voorspoels, W., & Storms, G. (2014). Norms of age of acquisition and concreteness for 30,000 Dutch words. *Acta psychologica*, 150, 80-84.

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: the HAROLD model. *Psychology and aging*, 17(1), 85.

Carson, N., Leach, L., & Murphy, K. J. (2018). A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *International journal of geriatric psychiatry,* 33(2), 379-388.

Chang, C. T., Lee, C. Y., Chou, C. J., Fuh, J. L., & Wu, H. C. (2016). Predictability effect on N400 reflects the severity of reading comprehension deficits in aphasia. *Neuropsychologia*, 81, 117-128.

Cocquyt, E. M., Vandewiele, M., Bonnarens, C., Santens, P., & De Letter, M. (2020). The sensitivity of event-related potentials/fields to logopedic interventions in patients with stroke-related aphasia. *Acta Neurologica Belgica*, 120(4), 805-817.

Cocquyt, E. M., Knockaert, N., van Mierlo, P., Szmalec, A., Duyck, W., Santens, P., & De Letter, M. (2021). The phonological Mismatch Negativity and P300 as diagnostic tools in stroke-related aphasia recovery: a longitudinal multiple case study. *Aphasiology*, 35(10), 1263-1280.

Cocquyt, E. M., Santens, P., van Mierlo, P., Duyck, W., Szmalec, A., & De Letter, M. (2022). Age-and gender-related differences in verbal semantic processing: the development of normative electrophysiological data in the Flemish population*. Language, Cognition and Neuroscience*, 1-27.

Daltrozzo, J., Wioland, N., & Kotchoubey, B. (2007). Sex differences in two event-related potentials components related to semantic priming. *Archives of Sexual Behavior*, 36(4), 555-568.

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.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117-1121

Dorme, A., Van Oudenhove, B., Criel, Y., Depuydt, E., De Groote, E., Huysman, E., …, & De Letter, M. Effect of Healthy Aging and Gender on Syntactic Input Processing: a P600 Event-Related Potential Study [Manuscript submitted for publication].

Du, X. D., Zhang, G. Y., Yang, Y., Li, Z., Pan, W., Yin, G. Z., ... & Chen, X. S. (2015). Follow-up of N400 in the Rehabilitation of First-episode Schizophrenia. *Chinese Medical Journal*, *128*(16), 2215-2219.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., ... & Van Petten, C. (2009). Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clinical Neurophysiology*, 120(11), 1883-1908.

Ehlers, M. R., Herrero, C. L., Kastrup, A., & Hildebrandt, H. (2015). The P300 in middle cerebral artery strokes or hemorrhages: Outcome predictions and source localization. *Clinical Neurophysiology*, 126(8), 1532-1538.

Evans, K. M., & Federmeier, K. D. (2007). The memory that's right and the memory that's left: Event-related potentials reveal hemispheric asymmetries in the encoding and retention of verbal information. *Neuropsychologia*, 45(8), 1777-1790.

Federmeier, K. D., Mai, H., & Kutas, M. (2005a). Both sides get the point: Hemispheric sensitivities to sentential constraint. *Memory & cognition*, 33(5), 871-886.

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

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, 115(3), 149-161.

Field, A. (2009) Discovering Statistics Using SPSS. 3rd Edition, Sage Publications Ltd., London.

Finnigan, S., O'Connell, R. G., Cummins, T. D., Broughton, M., & Robertson, I. H. (2011). ERP measures indicate both attention and working memory encoding decrements in aging. *Psychophysiology*, *48*(5), 601-611.

Giffard, B., Desgranges, B., Kerrouche, N., Piolino, P., & Eustache, F. (2003). The hyperpriming phenomenon in normal aging: A consequence of cognitive slowing?. *Neuropsychology*, 17(4), 594.

Grady, C. (2012). The cognitive neuroscience of ageing. *Nature Reviews Neuroscience*, 13(7), 491-505.

Guillem, F., & Mograss, M. (2005). Gender differences in memory processing: evidence from event-related potentials to faces. *Brain and cognition*, 57(1), 84-92.

Gunning, F. M., & Brickmann, A. M. (2010). Structural Brain Changes Associated With Normal Aging. Neuroimaging Research in Geriatric Mental Health, 101.

Gunter, T. C., Jackson, J. L., & Mulder, G. (1992). An electrophysiological study of semantic processing in young and middle-aged academics. *Psychophysiology*, 29(1), 38-54.

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., ... & Hämäläinen, M. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, 267.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of cognitive neuroscience*, 15(6), 883-899.

Halgren, E., Baudena, P., Heit, G., Clarke, M., Marinkovic, K., & Chauvel, P. (1994). Spatio-temporal stages in face and word processing. 2. Depth-recorded potentials in the human frontal and Rolandic cortices. *Journal of Physiology-Paris*, 88(1), 51-80.

Hoffman, P., & Morcom, A. M. (2018). Age-related changes in the neural networks supporting semantic cognition: A meta-analysis of 47 functional neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, 84, 134-150.

Horvath, A., Szucs, A., Csukly, G., Sakovics, A., Stefanics, G., & Kamondi, A. (2018). EEG and ERP biomarkers of Alzheimer's disease: a critical review. *Frontiers in bioscience (Landmark edition)*, *23*, 183-220.

Iragui, V., Kutas, M., & Salmon, D. P. (1996). Event-related brain potentials during semantic categorization in normal aging and senile dementia of the Alzheimer's type. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 100(5), 392-406.

Johnson, B. W., & Hamm, J. P. (2000). High-density mapping in an N400 paradigm: evidence for bilateral temporal lobe generators. *Clinical neurophysiology* : official journal of the International Federation of Clinical Neurophysiology, 111(3), 532-545.

Joyal, M., Groleau, C., Bouchard, C., Wilson, M. A., & Fecteau, S. (2020). Semantic processing in healthy aging and Alzheimer's disease: a systematic review of the N400 differences. *Brain sciences*, *10*(11), 770.

Karayanidis, F., Andrews, S., Ward, P. B., & McConaghy, N. (1993). Event-related potentials and repetition priming in young, middle-aged and elderly normal subjects. *Cognitive Brain Research*, 1(2), 123-134.

Kawohl, W., Bunse, S., Willmes, K., Hoffrogge, A., Buchner, H., & Huber, W. (2010). Semantic event-related potential components reflect severity of comprehension deficits in aphasia. *Neurorehabilitation and neural repair*, 24(3), 282-289.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior research methods*, 42(3), 643-650.

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

Kuperberg, G. R., Brothers, T., & Wlotko, E. W. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, 32(1), 12-35.

Kutas, M., & Hillyard, S. A. (1982). The lateral distribution of event-related potentials during sentence processing. *Neuropsychologia*, 20(5), 579-590.

Kutas, M., Iragui, V., & Hillyard, S. A. (1994). Effects of aging on event-related brain potentials (ERPs) in a visual detection task. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 92(2), 126-139.

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.

Kutas, M., & Federmeier, K. D. (2011). Tfhirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of psychology*, *62*, 621.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920-933.

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

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, 54(1), 146-157.

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, 31(3), 291-308.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities.

Maurer, U., Brem, S., Bucher, K., & Brandeis, D. (2005). Emerging neurophysiological specialization for letter strings. *Journal of Cognitive Neuroscience*, 17(10), 1532-1552.

Maurer, U., Brem, S., Kranz, F., Bucher, K., Benz, R., Halder, P., ... & Brandeis, D. (2006). Coarse neural tuning for print peaks when children learn to read. *Neuroimage*, 33(2), 749-758.

Maurer, U., & McCandliss, B. (2008). The development of visual expertise for words: the contribution of electrophysiology. Single-word Reading: Behavioral and Biological Perspectives. Grigorenko, E. L., Naples, A. J., & Naples, A. J. (Eds.), Taylor & Francis, 43-64.

McWeeny, S., & Norton, E. S. (2020). Understanding event-related potentials (ERPs) in clinical and basic language and communication disorders research: a tutorial. *International Journal of Language & Communication Disorders*, 55(4), 445-457.

Meechan, R., McCann, C., & Purdy, S. (2021). The electrophysiology of aphasia: A scoping review. *Clinical Neurophysiology.*

Meyers-Levy, J., & Sternthal, B. (1991). Gender differences in the use of message cues and judgments. *Journal of marketing research*, 28(1), 84-96.

Moors, A., De Houwer, J., Hermans, D., Wanmaker, S., Van Schie, K., Van Harmelen, A. L., ... & Brysbaert, M. (2013). Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. *Behavior research methods*, 45(1), 169-177.

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695-699.

Nieuwland, M. S. (2019). Do 'early'brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience & Biobehavioral Reviews*, 96, 367-400.

Nolfe, G., Cobianchi, A., Mossuto-Agatiello, L., & Giaquinto, S. (2006). The role of P300 in the recovery of post-stroke global aphasia. *European journal of neurology*, 13(4), 377-384.

Oschwald, J., Guye, S., Liem, F., Rast, P., Willis, S., Röcke, C., ... & Mérillat, S. (2020). Brain structure and cognitive ability in healthy aging: a review on longitudinal correlated change. *Reviews in the Neurosciences*, 31(1), 1-57.

Pratt, H. (2011). Sensory ERP components. The Oxford handbook of event-related potential components, 89-114.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., ... & Acker, J. D. (2005). Regional brain changes in aging healthy adults: general trends, individual differences and modifiers. *Cerebral cortex*, 15(11), 1676-1689.

Reuter-Lorenz, P. A., & Park, D. C. (2014). How does it STAC up? Revisiting the scaffolding theory of aging and cognition. *Neuropsychology review*, 24(3), 355-370.

Salthouse, T. (2010). Major issues in cognitive aging (No. 49). Oxford University Press.

Silkes, J. P., & Anjum, J. (2021). The role and use of event-related potentials in aphasia: A scoping review. *Brain and Language*, 219, 104966.

Sowell, E. R., Peterson, B. S., Kan, E., Woods, R. P., Yoshii, J., Bansal, R., ... & Toga, A. W. (2007). Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. *Cerebral cortex*, 17(7), 1550-1560.

Spreng, R. N., Wojtowicz, M., & Grady, C. L. (2010). Reliable differences in brain activity between young and old adults: a quantitative meta-analysis across multiple cognitive domains. *Neuroscience & Biobehavioral Reviews*, 34(8), 1178-1194.

Stalpaert, J., Cocquyt, E. M., Miatton, M., Sieben, A., Van Langenhove, T., van Mierlo, P., & De Letter, M. (2021). A case series of verbal semantic processing in primary progressive aphasia: evidence from the N400 effect. *International Journal of Language & Communication Disorders*, 56(6), 1165-1189.

Stalpaert, J., Standaert, S., D'Helft, L., Miatton, M., Sieben, A., Van Langenhove, T., ... & De Letter, M. (2022). Therapy-Induced Electrophysiological Changes in Primary Progressive Aphasia: A Preliminary Study. *Frontiers in human neuroscience*, 16.

Stoodley, C. J., Hill, P. R., Stein, J. F., & Bishop, D. V. (2006). Auditory event-related potentials differ in dyslexics even when auditory psychophysical performance is normal. *Brain Research*, 1121(1), 190-199.

Swaab, T., Brown, C., & Hagoort, P. (1997). Spoken sentence comprehension in aphasia: Event-related potential evidence for a lexical integration deficit. *Journal of Cognitive Neuroscience*, 9(1), 39-66.

van Loon-Vervoorn, W. A. (1985). Voorstelbaarheidswaarden van Nederlandse woorden: 4600 substantieven, 1000 verba en 500 adjectieven. Swets & Zeitlinger.

Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, 97(3), 279-293.

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

Van Strien, J. W. (1992). Classificatie van links-en rechtshandige proefpersonen. Nederlands Tijdschrift voor de Psychologie en haar Grensgebieden, 47, 88-92.

Van Vliet, M., Manyakov, N. V., Storms, G., Fias, W., Wiersema, J. R., & Van Hulle, M. M. (2014). Response-related potentials during semantic priming: the effect of a speeded button response task on ERPs. *PloS one,* 9(2), e87650.

Visch-Brink, E., Stronks, D., & Denes, G. (2005). SAT. Semantische AssociatieTest. Pearson.

Verma, J. P. (2015). *Repeated measures design for empirical researchers*. John Wiley & Sons.

Wallentin, M. (2009). Putative sex differences in verbal abilities and language cortex: A critical review. *Brain and Language*, 108(3), 175-183.

Wang, L., Bastiaansen, M., Yang, Y., & Hagoort, P. (2011). The influence of information structure on the depth of semantic processing: How focus and pitch accent determine the size of the N400 effect. *Neuropsychologia*, 49(5), 813-820.

Wang, K., Cheung, E. F., Gong, Q. Y., & Chan, R. C. (2011). Semantic processing disturbance in patients with schizophrenia: a meta-analysis of the N400 component. *PLoS one*, *6*(10), e25435.

Wierenga, L. M., Doucet, G. E., Dima, D., Agartz, I., Aghajani, M., Akudjedu, T. N., ... & Wittfeld, K. (2022). Greater male than female variability in regional brain structure across the lifespan. *Human Brain Mapping*, 43(1), 470-499.

Wirth, M., Horn, H., König, T., Stein, M., Federspiel, A., Meier, B., ... & Strik, W. (2007). Sex differences in semantic processing: event-related brain potentials distinguish between lower and higher order semantic analysis during word reading. *Cerebral Cortex*, 17(9), 1987-1997.

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

Xu, N., Hou, X., Zhao, B., Zhu, Z., & Yang, Y. (2017). Age-related temporal-spatial dynamic ERP changes during sentence comprehension. *Neuroscience Letters*, 645, 74-79.

Zhu, Z., Hou, X., & Yang, Y. (2018). Reduced syntactic processing efficiency in older adults during sentence comprehension. *Frontiers in Psychology*, 9, 243.

Zhu, Z., Wang, S., Xu, N., Li, M., & Yang, Y. (2019). Semantic integration declines independently of working memory in aging. *Applied Psycholinguistics*, 40(6), 1481-1494.