From oscillations to language: behavioural and electroencephalographic studies on cross-language interactions

Von Grebmer Zu Wolfsthurn, S.

Citation
Von Grebmer Zu Wolfsthurn, S. (2023, January 17). *From oscillations to language: behavioural and electroencephalographic studies on cross-language interactions*. *LOT dissertation series*. LOT, Amsterdam. Retrieved from https://hdl.handle.net/1887/3512212

CHAPTER 5

Processing non-native syntactic violations: different ERP correlates as a function of typological similarity

This article was submitted for review as: Von Grebmer Zu Wolfsthurn, S., Pablos-Robles, L., & Schiller, N. O. (2022). Processing syntactic violations in the non-native language: different ERP correlates as a function of typological similarity. Neuropsychologia.

Abstract: Despite often featured in theoretical accounts, the exact impact of typological similarity on non-native language comprehension and its corresponding neural correlates remain unclear. Here, we examined the modulatory role of typological similarity in syntactic violation processing, e.g., [el volcán] (the volcano) vs. [*la volcán] in the non-native language Spanish, as well as in cross-linguistic influence. Participants were either Italian late learners of Spanish (highly similar language pair) or German late learners of Spanish (less similar language pair). We measured P600 component amplitudes, accuracy and response times. In line with our predictions, we found a larger P600 effect and differential CLI effects for Italian-Spanish speakers compared to German-Spanish speakers. Interestingly, Italian-Spanish speakers responded overall more slowly compared to German-Spanish speakers. Taken together, the results reflect a typological similarity effect in non-native comprehension in the form of a processing advantage for typologically similar languages, but only at the neural level. These findings have critical implications for the interplay of different languages in the multilingual brain.

Keywords: *typological similarity, non-native comprehension, cross-linguistic influence, gender congruency effect, cognate facilitation effect, EEG, ERPs, P600 effect, generalised additive mixed models*

5.1 Introduction

A fundamental characteristic of multilingual language comprehension is *cross-linguistic influence* (CLI) between the native language (L1) and the non-native language (Kroll et al., 2015; Lago et al., 2021; Lemhöfer et al., 2008). In this study, we considered individuals who were able to communicate in two or more languages as multilinguals (Cenoz, 2013). In language comprehension, CLI is often conceptualised as the parallel activation of both the L1 and the non-native language (Hamers & Lambert, 1972; Lago et al., 2021), even when the circumstances only require the use of one language (Blumenfeld & Marian, 2013; Lago et al., 2021; Marian & Spivey, 2003b; Nozari & Pinet, 2020). CLI was demonstrated at the level of (morpho)syntax (Grüter, Lew-Williams & Fernald, 2012; Lemhöfer et al., 2008; Tolentino & Tokowicz, 2011; Zawiszewski et al., 2011), for grammatical gender (Lemhöfer et al., 2008; Paolieri et al., 2020) and for cognate processing (Midgley et al., 2011; Peeters et al., 2013). Moreover, CLI was reported for different ages of non-native acquisition (AoA), with some evidence suggesting that CLI may be more pronounced in early acquisition stages (Gillon-Dowens et al., 2010; Ringbom, 1987; Sunderman & Kroll, 2006). One important question is whether CLI is modulated by the *typological similarity*, that is, the syntactic and structural similarities between the L1 and the non-native language (Foote, 2009; Putnam, Carlson & Reitter, 2018; Tolentino & Tokowicz, 2011). In other words, does similarity at the level of, for example, grammatical gender or orthographic and phonological form overlap have an impact on non-native language processing? This is a critical issue because it is intimately linked to the functional organisation of multilinguals' languages and the question of how cross-language similarities can facilitate or hinder non-native processing (Tolentino & Tokowicz, 2011). As will be discussed below, it has long been proposed that typological similarity is a crucial factor in multilingual language processing (Casaponsa & Duñabeitia, 2016; MacWhinney, 2005; Odlin, 1989; Sabourin & Stowe, 2008; Tolentino & Tokowicz, 2011; Weinreich, 1953; Zawiszewski & Laka, 2020). Yet, there is a distinct lack of studies directly tackling the impact of typological similarity on some of the most fundamental cognitive aspects of multilingual language processing such as CLI.

This study focused on examining the role of typological similarity via two CLI effects. The first CLI effect we investigated was the *gender congruency effect*, which reflects CLI at the level of grammatical gender (hereafter gender). Gender refers to a noun classification system which is featured in several Indo-European languages (Corbett, 1991). Among those languages are Italian, German and Spanish, which are the languages of interest in this study. The gender systems of both Italian and Spanish feature a feminine and masculine gender value, marked by [la$_F$] and [il$_M$], and [la$_F$] and [el$_M$], respectively. In contrast, German has a three-way gender system characterised by a feminine, masculine and neuter gender value marked by [der$_M$], [die$_F$] and [das$_N$], respectively (Schiller & Caramazza, 2003; Schiller & Costa, 2006). The so-called gender congruency effect manifests itself in more accurate and faster processing of gender congruent items, e.g., [il$_M$ cane$_M$] and [el$_M$ perro$_M$] *"the dog"* compared to incongruent items, e.g., [il$_M$ latte$_M$] and [la$_F$ leche$_F$] *"the milk"* in Italian and Spanish (Lemhöfer et al., 2008; Paolieri et al., 2019; Sá-Leite et al., 2020). In other words, similarity at the level of gender results in a measurable processing advantage for gender congruent items vs. incongruent items across the L1 and the non-native language.

The second CLI effect we examined in this study was the *cognate facilitation effect*. It reflects CLI at the level of orthographic and phonological overlap, i.e., cognates. More specifically, this effect entails more accurate and faster processing of cognates, i.e., words with a significant overlap in terms of orthographic and phonological word form, e.g., [vulcano] and [volcán] *"volcano"*; compared to non-cognates, e.g., [viso] and [cara] *"face"* in Italian and Spanish (Comesaña et al., 2014; Costa et al., 2005; Marian, Blumenfeld & Boukrina, 2008; Midgley et al., 2011; Lemhöfer et al., 2008). With respect to typological similarity, Marian et al. (2008) showed that a larger phonological overlap for native Russian speakers with high proficiency in English was linked to higher performance and shorter response times (RTs) in an auditory lexical decision task. In turn, this particular effect highlights the processing advantage for orthographically and phonologically similar word forms, i.e., cognates compared to non-cognates. Taking both effects together, the gender congruency effect and the cognate facilitation effect tentatively indicate a processing advantage for typologically more similar structures compared to less similar structures, as reflected by higher accuracy and faster RTs for congruent items and cognates compared to incongruent items and non-cognates.

In this study, we used both effects to closely examine the impact of typological similarity on non-native comprehension, specifically in terms of gender similarity and orthographic and phonological word form overlap between the L1 and the non-native language. Directly relevant to this study is the *Language Distance Hypothesis*, LDH (Zawiszewski & Laka, 2020), which provides a theoretical account of the interaction between typological similarity and CLI effects. The core prediction of this account is the modulation of CLI on the basis of (morpho)syntactic similarity between the L1 and the non-native language. Concretely, the LDH predicts more native-like behavioural patterns and event-related components (ERPs) emerging in the non-native language for highly morphologically similar structures across the L1 and the non-native language compared to less similar structures. This would be reflected in higher accuracy, shorter RTs and larger (more native-like) ERP components for morphologically similar structures across languages.

Zawiszewski and Laka (2020) systematically tested this account in a recent experiment on morphological processing in grammatical and ungrammatical sentences in highly proficient Basque-Spanish speakers and Spanish-Basque speakers. The critical manipulation was the presence or absence of a particular morphological feature in the non-native language compared to the L1. Consistent with the LDH, their results indicated a link between shorter RTs and larger ERP effects (i.e., native-like ERP effects) in the non-native language for some morphologically similar structures compared to less similar structures. In turn, this suggested an overall processing advantage in the non-native language for morphologically similar structures. Critically, the authors also acknowledged that AoA and non-native proficiency could modulate typological similarity effects. This is in line with previous studies which have highlighted the impact of non-native proficiency on typological similarity effects (Gillon-Dowens et al., 2010; Ringbom, 1987; Tokowicz & MacWhinney, 2005; Weber-Fox & Neville, 1996). For example, Tokowicz and MacWhinney (2005) examined low proficient and highly proficient English-Spanish speakers and their sensitivity to the correctness of syntactic structures. In addition to non-native proficiency, the second critical manipulation was that some syntactic structures were similar across the languages (auxiliary marking), whereas the other structures were not (gender and number agreement). Results demonstrated that increased typological similarity was linked to shorter RTs, in particular for lower proficient speakers. In contrast, highly proficient speakers in this study appeared to remain largely unaffected by typological similarity. This finding suggests that typological similarity effects may be more pronounced in earlier acquisition stages (Sunderman & Kroll, 2006; Zawiszewski & Laka, 2020). Therefore, in this study we focused on late language learners to examine typological similarity effects more closely, i.e., individuals who acquired a non-native language later during development after the age of fourteen (S. Rossi et al., 2006).

Before the formulation of the LDH, earlier work by Sabourin and Stowe (2008) examined the impact of typological similarity on gender processing. In their study, they compared gender agreement processing in Dutch across native Dutch speakers vs. native German and Romance language speakers, who were all late learners of Dutch (AoA $>$ 14 years of age). In terms of typological similarity, German and Dutch have a greater linguistic overlap compared to Romance languages and Dutch (Schepens et al., 2013; Van der Slik, 2010). Therefore, in their study, the German-Dutch speakers represented the typologically similar language pair, and the Romance language-Dutch speakers the typologically less similar language pair. Importantly, the authors also explored the effects of typological similarity on neural correlates of gender processing, with a specific focus on P600 component amplitudes. The P600 component is an event-related brain potential (ERP) and is characterised as a positive-going waveform reaching its peak approximately 600 ms post-stimulus onset in centro-parietal regions (Friederici et al., 1999; Friederici, Hahne & Saddy, 2002; Swaab et al., 2011). The so-called P600 effect has been reported in the context of higher voltage amplitudes for syntactic violations such as [*la$_F$ volcán$_M$] vs. syntactically correct structures such as [el$_M$ volcán$_M$] *"the volcano"* (Hagoort et al., 1993; Friederici, Gunter, Hahne & Mauth, 2004; Hahne, 2001; Weber-Fox & Neville, 1996). Critically, Sabourin and Stowe (2008) found that P600 effects were modulated by syntactic similarity between the L1 and Dutch: only native German speakers showed a clear P600 effect for syntactic violations in Dutch, whereas the native Romance language speakers did not. The results suggested that typologically similar languages (e.g., German-Dutch) were linked to an enhanced sensitivity to gender violations in comparison to less typologically similar languages (e.g., Romance language-Dutch) and a larger P600 effect. Behaviourally, the German-Dutch speakers outperformed the Romance language-Dutch speakers in terms of accuracy in gender assignment, which indicates differential CLI effects as a function of typological similarity. These results are in line with the predictions by the LDH (Zawiszewski & Laka, 2020) and are also compatible with studies linking increased CLI to typologically similar languages compared to typologically less similar languages (Mosca, 2017; Tolentino & Tokowicz, 2011).

In sum, current research strongly suggests that typological similarity plays a significant role in modulating both behavioural and neural measures of non-native language comprehension. More specifically, typological similarity was shown to influence non-native gender processing as well as orthographic and phonological processing. In this, previous studies suggest the following: first, behavioural effects of typological similarity were found for cross-linguistic gender processing, suggesting a gender processing advantage for typologically similar languages compared to less similar languages (Paolieri et al., 2020). Secondly, typological similarity effects were also found for cross-linguistic cognate processing, whereby a higher typological similarity was linked to more efficient and faster processing of orthographically and phonologically similar structures (Costa et al., 2005; Comesaña et al., 2014; Lemhöfer et al., 2008). Critically, this suggests that CLI is influenced by typological similarity, with more pronounced CLI for typologically similar language combinations compared to less similar combinations (Sabourin & Stowe, 2008; Tolentino & Tokowicz, 2011). Third, studies have also reported a typological similarity effect on the neural correlates of cross-linguistic non-native gender processing (Sabourin & Stowe, 2008). Specifically, larger, more native-like P600 effects were linked to a higher typological similarity (Sabourin & Stowe, 2008).

5.1.1 The current study

The aim of the current study was to systematically investigate the effect of typological similarity on syntactic violation processing and on CLI in non-native comprehension in late language learners using behavioural measures (accuracy and RTs) and ERP measures (P600 component voltage amplitudes). For this, we tested two groups of late learners of Spanish speakers with a varying degree of typological similarity: representing the typologically similar group, we tested native Italian speakers; and representing the typologically less similar group, we tested native German speakers (Schepens, Dijkstra & Grootjen, 2012; Schepens et al., 2013). Further, we focused on two CLI effects: the gender congruency effect, which reflects CLI of the gender systems (Lemhöfer et al., 2008; Paolieri et al., 2019; Sá-Leite et al., 2020), and the cognate facilitation effect, reflecting CLI of the orthographic and phonological systems (Costa et al., 2005; Comesaña et al., 2014; Lemhöfer et al., 2008). To test these typological similarity effects, we employed a syntactic violation paradigm, whereby participants judged the grammatical correctness of noun phrases such as *el volcán* [the volcano] (non-violation trial) vs. **la volcán* (violation trial) while we recorded their ERPs.

Research questions

The research questions we sought to answer in this study were the following: first, is there a P600 effect (i.e., a difference between non-violation and violation trials) for both the Italian-Spanish and the German-Spanish group? Second, is the P600 effect larger for one group compared to the other? Third, do CLI effects of gender congruency and cognate status vary across the two groups? This would reflect a typological similarity effect at the neural level, as well as a typological similarity effect on CLI between the native and the non-native language. Taking the LDH (Zawiszewski & Laka, 2020) as our theoretical basis, we predicted that speakers of typologically similar languages would bear a processing advantage in the non-native language compared to speakers of typologically less similar languages.

Hypotheses

Behavioural hypotheses. With respect to our first research question, we expected participants to be significantly more accurate and faster for non-violation trials compared to violation trials. Critically, for our second research question, we predicted an interaction effect of *L1* (Italian vs. German) with *violation type* (non-violation vs. violation) to indicate a typological similarity effect on processing syntactic (non-)violations. In other words, consistent with the LDH, we predicted that the Italian-Spanish group would be more accurate and faster at processing non-violation trials vs. violation trials compared to the German-Spanish group. For our third research question, we first predicted CLI effects, as manifested in more accurate and faster processing of congruent and cognate items compared to incongruent and non-cognate items. Importantly, here we aggregated our two main manipulations *gender congruency* (congruent vs. incongruent) and *cognate status* (cognate vs. non-cognate) into the variable *condition* with four levels: congruent/cognate, congruent/non-cognate, incongruent/cognate and incongruent/non-cognate items. In this, we investigated an interaction effect of *L1* with *condition*. We hypothesised that the Italian-Spanish group would be statistically more accurate and faster at processing congruent and cognate items vs. incongruent and non-cognate items compared to the German-Spanish group.

ERP hypotheses. In terms of our first research question, we expected a P600 effect in both groups, as indicated by smaller voltage amplitudes for non-violation trials compared to violation trials. For our second research question, we predicted an interaction effect between *L1* and *violation type* to indicate a typological similarity effect on the P600 effect size. More specifically, in line with the LDH, we hypothesised a larger P600 effect for the Italian-Spanish group compared to the German-Spanish group. For our third research question, we predicted an interaction effect between *L1* and *condition*, indicating an effect of typological similarity on CLI. Specifically, we expected to observe larger voltage amplitudes connected to larger CLI for the Italian-Spanish group compared to the German-Spanish group. Taken together, these findings would indicate a general processing advantage for the Italian-Spanish group compared to the German-Spanish group, with overall higher accuracy, shorter RTs and larger P600 amplitudes for the typologically similar language combination.

5.2 Methods

Before the experiment, participants filled out the Language Experience and Proficiency Questionnaire, LEAP-Q (Kaushanskaya et al., 2020; Marian et al., 2007). The LEAP-Q was used to establish proficiency and experience measures for the participants' known languages. During the experimental session, participants completed the LexTALE-Esp task (Izura et al., 2014), a lexical decision task that provides a vocabulary size score (*LexTALE-Esp score*) in Spanish. LexTALE-Esp scores were previously found to be highly correlated with overall proficiency levels, see Lemhöfer and Broersma (2012). Subsequently, participants completed the syntactic violation paradigm. Note that the German-Spanish participants included in this study as well as the procedures used are identical to the ones reported in Von Grebmer Zu Wolfsthurn et al. (2021a).

5.2.1 Participants

We recruited and tested 33 native speakers of Italian (24 females) with $M = 27.12$ years of age ($SD = 4.08$). We also tested 33 native speakers of German (27 females) with $M = 23.06$ years of age ($SD = 2.47$), previously described in Von Grebmer Zu Wolfsthurn et al. (2021a). All participants had an intermediate B1/B2 proficiency level in Spanish according to the Common European Framework of Reference for Languages, *CEFR* (Council of Europe, 2001). We established this proficiency level using various linguistic variables of the LEAP-Q, the LexTALE-Esp score and by recruiting directly from foreign language courses aimed at the B1/B2 level. Participants had to meet the recruitment criteria to be eligible for the study: dominant right-handed, between 18 and 35 years of age, absence of psychological, reading or language impairments, no second language learnt before the age of five and an age of acquisition of Spanish of more than fourteen years. We imposed additional recruitment criteria for the Italian-Spanish group because we tested them in the non-native environment: participants had to have lived in a Spanish-speaking country for less than one year and started learning Spanish shortly before or upon their arrival to Spain. We combined these criteria with the information of the LEAP-Q to establish our speakers within the category of late language learners with intermediate B1/B2 proficiency levels (Kaushanskaya et al., 2020). Note that not all participants were included in the data analyses, see section 5.3.4 for details about data exclusion.

Linguistic profile of participants

Below, we summarised several key linguistic variables related to Spanish from the LEAP-Q and the LexTALE-Esp (Table 5.2.1). We limited these descriptions to the participants included in the statistical analyses (section 5.3.4). In the Italian-Spanish group, twelve participants stated they perceived Spanish as their current first foreign language in terms of dominance, thirteen participants stated Spanish as their second, three participants as their third and finally, one participant as their fourth foreign language. For the German-Spanish group, four participants self-reported Spanish as their perceived first foreign language, twenty-one participants as their second, and three as their third foreign language. See Appendix 5.A and Appendix 5.B for a more detailed linguistic profile of the two groups.

Table 5.2.1: *Linguistic profile of Spanish for the Italian-Spanish group (n = 29) and the German-Spanish group (n = 28), including the LexTALE-Esp score. Self-reported proficiency measures (speaking, comprehension, reading) were rated on a scale from zero to ten (ten being equal to maximal proficiency) and are highlighted in bold.*

Measure	**Italian-Spanish**	**German-Spanish**
LexTALE-Esp score mean	27.29 (*SD* = 14.01)	18.91 (*SD* = 20.45)
LexTALE-Esp score range	-7.37 - 49.30	-23.16 - 60.18
AoA Spanish (years)	23.31 (*SD* = 4.86)	16.46 (*SD* = 2.33)
Fluency age Spanish (years)	24.52 (*SD* = 4.45)	18.59 (*SD* = 2.13)
Reading onset age Spanish (years)	23.79 (*SD* = 4.74)	17.36 (*SD* = 2.88)
Fluent reading age Spanish (years)	23.92 (*SD* = 4.84)	18.50 (*SD* = 2.52)
Immersion in Spanish-speaking country (years)	0.48 (*SD* = 0.35)	1.04 (*SD* = 0.69)
Daily exposure (%)	41.38 (*SD* = 18.27)	9.86 (*SD* = 9.73)
Speaking proficiency	6.31 (*SD* = 1.73)	6.85 (*SD* = 0.93)
Comprehension proficiency	7.32 (*SD* = 1.76)	7.50 (*SD* = 0.88)
Reading proficiency	7.48 (*SD* = 1.48)	7.18 (*SD* = 1.12)

5.2.2 Materials and design

We used the Italian and the German version of the LEAP-Q for our two groups, respectively. Further, we generated E-prime (Version 2) scripts (Schneider et al., 2002) for the LexTALE-Esp and the syntactic violation paradigm.

Stimuli

LexTALE-Esp. In line with the original lexical decision task by Izura et al. (2014), stimuli consisted of 60 Spanish words varying in terms of frequency, as well as 30 pseudowords with different degrees of similarity to real Spanish words, for example [*alardio*]. Therefore, the critical manipulation was *condition* (word vs. pseudoword), and we measured accuracy during this task.

Syntactic violation paradigm. The stimuli selection procedure for the Italian-Spanish and the stimuli for the German-Spanish group were identical as outlined in Von Grebmer Zu Wolfsthurn et al. (2021a). However, the selected stimuli differed between the two groups due to the constraints by our main manipulations: stimuli were selected separately for each group based on their gender congruency and cognate status across Italian and Spanish, and across German and Spanish. As a result, the stimuli were different for the Italian-Spanish compared to the German-Spanish group. We selected a total of 224 stimuli for each group. We followed a 2 x 2 x 2 fully factorial design, with *violation type* (non-violation vs. violation), *gender congruency* (congruent vs. incongruent) and *cognate status* (cognate vs. non-cognate) as our critical manipulations. Half of all trials were violation trials, and the other half non-violation trials. Half of our stimuli were gender congruent, and half gender incongruent. In turn, half of the stimuli nouns were cognates, and the rest non-cognates. Therefore, each experimental condition contained 28 stimuli, adding to a total of 224 stimuli for each group. The task was a grammaticality judgment task embedded within a syntactic violation paradigm, whereby participants determined whether a noun phrase such as *el volcán* was grammatically correct. We recorded participants' EEG during this task, as well as accuracy and RTs.

EEG recordings

Italian-Spanish group. We used 32 active electrodes in a standard 10/20 montage to collect EEG data at a sampling rate of 500 Hz via the BrainVision Recorder software (Version 1.10) by BrainProducts. We placed one electrode (FT9) under the participant's left eye to record the vertical electrooculogram (VEOG), and one electrode (FT10) at the outer canthus of the left eye for the horizontal electrooculogram (HEOG). All electrodes were referenced to FCz. A ground electrode was positioned on the participant's right cheek. We used BrainVision Recorder to keep our impedances for each electrode below 10 kΩ for an enhanced signal.

German-Spanish group. We sampled the EEG data from 32 passive electrodes configured in a 10/20 montage at a rate of 500 Hz and again using the BrainVision Recorder software (Version 1.23.0001). We placed one VEOG electrode underneath the left eye, two HEOG electrodes at the outer canthus of each eye, and the ground electrode on the right cheek of the participant. The original reference electrode was Cz. We used the actiCAP ControlSoftware (Version 1.2.5.3) to ensure that impedances were below 5 kΩ for the reference and ground electrode, and below 10 kΩ for the remaining electrodes.

5.2.3 Procedure

The experimental session was carried out on a computer screen in an experimental booth and took place in the CBC Laboratories at the Pompeu Fabra University for the Italian-Spanish group, and in the Neurolinguistic Laboratories at the University of Konstanz for the German-Spanish group. Prior to the start of the experiment, we provided participants with an information sheet and a consent form in their L1, complying with the ethics code for neurolinguistic research in the Faculty of Humanities at Leiden University. During the experiment, participants completed both the LexTALE-Esp and the syntactic violation paradigm. Written instructions for each task were provided on the screen in black font on a white background. The procedure for each task was identical for both groups, with the exception that the oral and written instructions were given in Italian to the Italian-Spanish group, and in German to the German-Spanish group. After the experiment, participants received a written and oral debrief in their L1, as well as a monetary compensation for their participation.

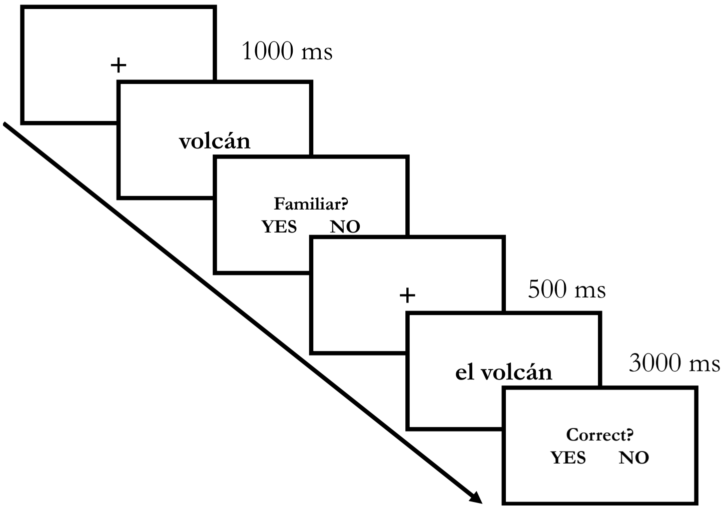
LexTALE-Esp

Participants were shown a fixation cross for 1,000 ms. Next, a letter string of either a Spanish word or pseudoword appeared on the screen. Participants decided whether or not the letter string was a Spanish word via a button press. The next trial was initiated following the participant's response. Prior to the experiment, we eliminated three word stimuli due to overlap with the stimuli from the syntactic violation paradigm. Therefore, we presented participants with 57 word stimuli, and 30 pseudoword stimuli, adding to a total of 87 trials. Each stimulus was only presented once, and trial order was fully randomised for each participant. In a final step, we calculated the LexTALE-Esp score in offline calculations by subtracting the percentage of incorrectly identified pseudowords from the correctly identified words for each participant (Izura et al., 2014).

Syntactic violation paradigm

The task procedure was identical for both groups, as is outlined in detail in Von Grebmer Zu Wolfsthurn et al. (2021a). It was as follows: participants were first presented with a fixation cross for 1,000 ms. Then, they were instructed that they would see a bare noun (e.g., *volcán* [volcano]) on the screen. Here they had to determine their familiarity with the noun by responding to a yes/no question during its presentation. This was followed by the display of a fixation cross for 500 ms. We then visually presented participants with determiner + noun constructions, e.g., *el volcán* [the volcano] for a maximum time of 3,000 ms and asked participants to determine the grammatical correctness of each noun phrase as accurately and fast as possible via a button press. The next trial was initiated upon participant's response. Each stimulus was only shown once within a noun phrase, adding to a total of 224 trials. Trial order was fully randomised, and we incorporated two self-paced breaks for our participants. At the beginning of the task, we included eight practise trials to familiarise participants with the trial procedure. Within-experiment instructions and prompts were displayed in Spanish. See Figure 5.2.1 for example trials of this task.

Figure 5.2.1: *Example trial for the syntactic violation paradigm. Within-trial prompts in the figure were translated to English for convenience. The final prompt was added to the figure for visualisation purposes only.*



5.3 Results

5.3.1 Behavioural data exclusion

We included the same participants in the behavioural analysis as in the EEG analysis (see section 5.3.4). This meant that we analysed data from 29 Italian-Spanish speakers after excluding four participants, and data from 28 German-Spanish speakers after excluding five participants, thereby analysing data from a total of 57 participants.

5.3.2 Behavioural data analysis

The behavioural data analysis procedure matched the analysis described in Von Grebmer Zu Wolfsthurn et al. (2021a), with the exception that our maximal model in this study reflected our research questions. Here, we used a generalised linear mixed effects modelling (GLMM) approach to model *accuracy* and *RTs* for the grammaticality judgement. All analyses were implemented in R, Version 4.1.2, and in RStudio, Version 2021.09.0 (R Core Team, 2021) using the *lme4* package (Bates et al., 2020). We specified a binomial distribution to model *accuracy*, and a gamma distribution with an identity link function to model positively skewed *RTs* from correct trials (Lo & Andrews, 2015). We initially built a theoretically plausible maximal model with an elaborate fixed effects structure. This included the interaction effect for *L1* (Italian vs. German) and *violation type* (violation vs. non-violation), as well as the interaction effect for *L1* and *condition* (congruent/cognate vs. congruent non-cognate vs. incongruent/cognate vs. incongruent non-cognate), representing the CLI effects. Next, our model further included the covariates *LexTALE-Esp score*, *order of acquisition of Spanish*, *terminal phoneme*, *target noun gender* and *word length*. Finally, we included random intercepts for *participant* and *item*, as well as random slopes for *violation type* and *condition* for a maximal random effects structure (Barr, 2013). Upon model non-convergence or singular fit, we simplified our random effects structure. We then tested for the relevance of each covariate and the significance of the other fixed effects terms by systematically examining their statistical significance in a model comparison approach using the *anova()* function. A significant χ^2-test indicated that a particular term significantly contributed to an improved goodness of fit and was subsequently kept in the model. For accuracy, the models were fitted with the Laplace approximation. For RTs, we used the default maximum likelihood estimation (Bates et al., 2020) for unbiased estimates for the model comparisons, but re-fitted the final model with the restricted maximum likelihood method (Mardia, Southworth & Taylor, 1999). We determined treatment coding as our default contrast, and vigorously checked the model diagnostics using the *DHARMa* package (Hartig, 2020). P-values were derived using the *lmerTest* package (Kuznetsova, Brockhoff, Christensen & Pødenphant-Jensen, 2020), and test statistics above ±1.96 were interpreted as significant at $\alpha = 0.05$ (Alday et al., 2017). Note that we report model parameters for accuracy as odds ratios.

5.3.3 Behavioural data results

We calculated mean accuracy and RTs for each condition and each group in Table 5.3.1.

Accuracy. The maximal model described above in section 5.3.2 did not converge and was subsequently simplified. The simplified model contained both interaction effects, but yielded an insignificant interaction effect for *L1* and *violation type* with $\beta = 0.946$, $z = -0.145$, $p = 0.885$. We therefore compared this model to a model which included only the interaction effect between *L1* and *condition*, but not *L1* and *violation type*. There was no significant difference in model fit between these two models with $\chi^2(1, n = 57) = 0.021$, $p = 0.885$, and we subsequently selected the simpler model as our best-fitting model (Appendix 5.D). This best-fitting model included the interaction effect between *L1* and *condition*, and main effects for *L1* and *violation type*. Further, the model included *LexTALE-Esp score* and *target noun gender* as covariates, by-*participant* random slopes for *violation type*, and random intercepts for both *participant* and *item* (Appendix 5.D). Therefore, the final model was: accuracy ∼ L1 (Italian vs. German) + violation type (violation vs. non-violation) + L1 * condition (congruent/cognate vs. congruent/non-cognate vs. incongruent/cognate vs. incongruent/non-cognate) + LexTALE-Esp score + target noun gender (feminine vs. masculine) + (violation type|participant) + (1|item).

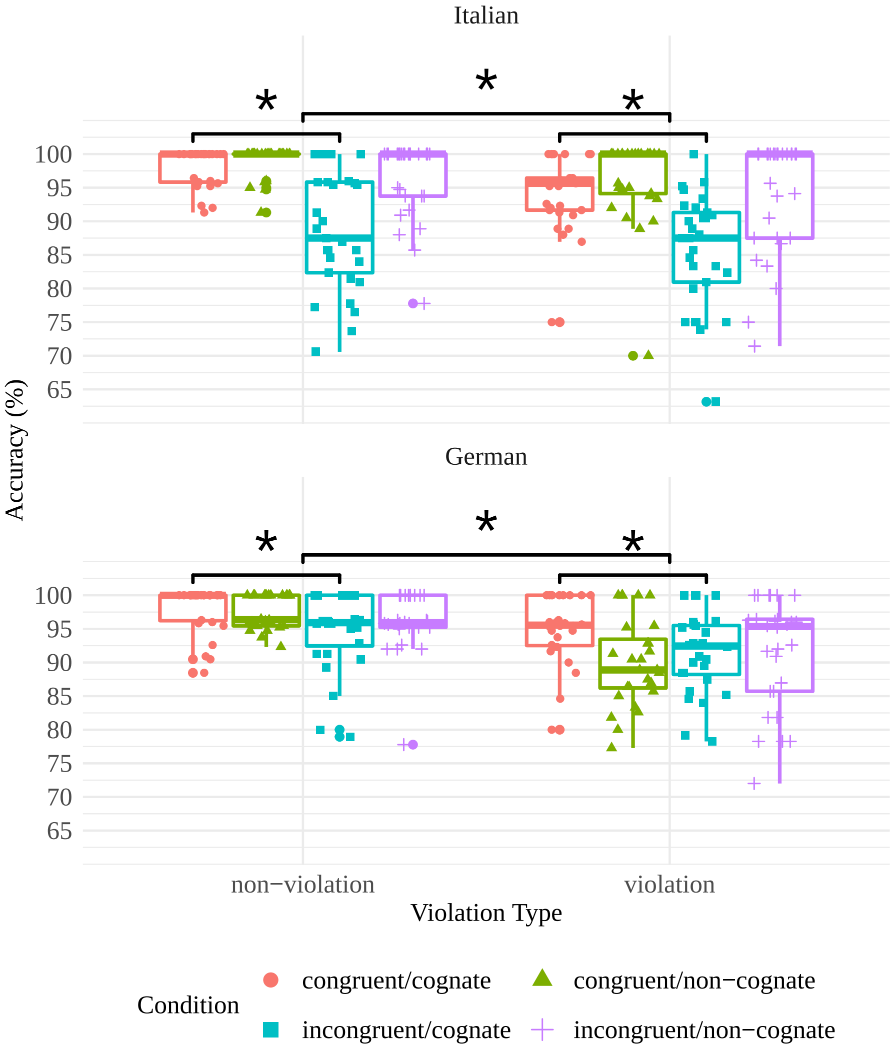
Participants were more accurate for non-violation trials compared to violation trials with $\beta = 0.412$, *95% CI*[0.279, 0.609], $z = -4.45$, $p < 0.001$. Further, there was a main effect of *condition* with participants being more accurate for congruent/cognate items compared to incongruent/cognate items with $\beta = 0.258$, *95% CI*[0.132, 0.504], $z = -3.96$, $p < 0.001$ (Figure 5.3.1). Despite being included in the final model, the main effect for *L1* was not significant with $\beta = 1.50$, *95% CI*[0.673, 3.35], $z = 0.993$, $p = 0.321$ for the Italian-Spanish group compared to German-Spanish group. Critically, the interaction effect between *L1* and *condition* was insigni-

Table 5.3.1: *Mean accuracy and mean RTs for each condition for each group (n = 57).*

L1	**Violation type**	**Condition**	**Mean (%)**	**SD**	**Mean (ms)**	**SD**
Italian	non-violation	congruent/cognate	98.08	13.74	816.76	328.44
		congruent/non-cognate	98.97	10.13	817.31	347.17
		incongruent/cognate	88.76	31.62	931.11	452.49
		incongruent/non-cognate	96.54	18.31	854.38	377.15
	violation	congruent/cognate	94.08	23.61	953.55	401.33
		congruent/non-cognate	96.38	18.70	929.73	356.10
		incongruent/cognate	85.60	35.14	1049.37	489.65
		incongruent/non-cognate	93.92	23.92	991.48	439.02
German	non-violation	congruent/cognate	98.08	13.72	721.69	320.48
		congruent/non-cognate	97.52	15.57	724.44	318.78
		incongruent/cognate	94.78	22.25	791.09	370.86
		incongruent/non-cognate	96.02	19.56	737.29	351.83
	violation	congruent/cognate	95.06	21.70	852.12	387.34
		congruent/non-cognate	90.23	29.71	833.22	368.91
		incongruent/cognate	91.47	27.95	875.94	386.87
		incongruent/non-cognate	91.92	27.27	859.32	406.40

ficant for all levels contrasted with the Italian-Spanish group and congruent/cognate items with $\beta = 0.349$, *95% CI*[0.121, 1.01], $z =$ -1.94, $p = 0.052$ for the German-Spanish group and congruent/non-cognate items, $\beta = 1.68$, *95% CI*[0.631, 4.46], $z = 1.04$, $p = 0.300$ for the German-Spanish group and incongruent/cognate items, and finally, $\beta = 0.661$, *95% CI*[0.238, 1.84], $z =$ -0.792, $p = 0.428$ for the German-Spanish group and incongruent/non-cognate items. Taken together, we found a main effect of *violation type* and a small main effect for *condition* on accuracy. However, we found neither a significant interaction effect of *L1* and *violation type*, nor of *L1* and *condition*. We also did not find a main effect of *L1* on accuracy, either. This indicated that accuracy levels were comparable for the two groups. See Appendix 5.D for the full model parameters for accuracy.

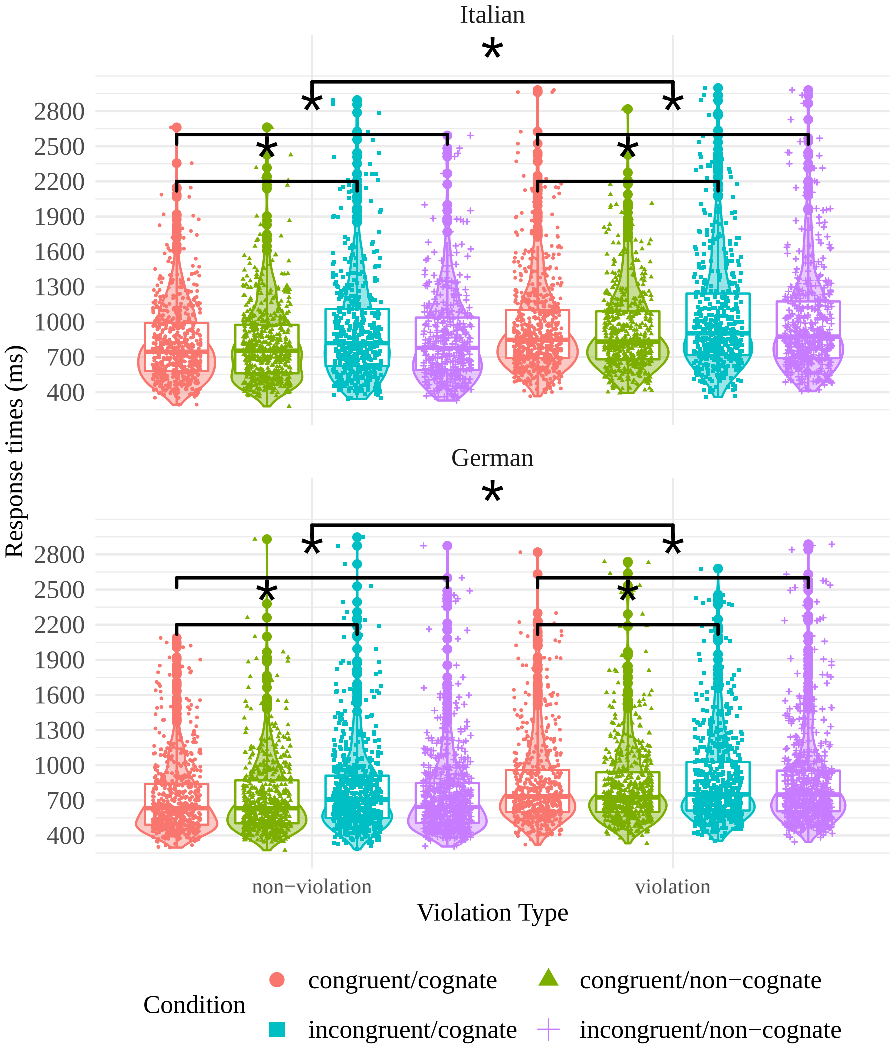
Figure 5.3.1: *Mean accuracy (%) for each group for each condition (n = 57).*



Response times. The maximal model described in section 5.3.2 that included both interaction terms yielded non-convergence. We subsequently simplified the random effects structure and also excluded *LexTALE-Esp score* as a covariate. This simplified model yielded an insignificant interaction effect for *L1* and *violation type* with β = -1.51, t = -0.407, p = 0.684 for Italian and non-violation items compared to German and violation items. We then compared this model to a model which included only the interaction effect between *L1* and *condition*, but not *L1* and *violation type*. This comparison showed no difference in model fit with $\chi^2(1, \text{n} = 57) = 0.001$, $p = 0.971$. We therefore declared the model containing the interaction effect between *L1* and *condition* and main effects of *L1* and *violation type* as our best-fitting model (Appendix 5.E). Similar to the best-fitting model for accuracy, this model also included *target noun gender* as covariate, by-*subject* random slopes for *violation type* and random intercepts for *participant* and *item* (Appendix 5.E). Subsequently, the best-fitting model was: RTs ~ L1 (Italian vs. German) + violation type (violation vs. non-violation) + L1 * condition (congruent/cognate vs. congruent/non-cognate vs. incongruent/cognate vs. incongruent/non-cognate) + target noun gender (feminine vs. masculine) + (violation type|participant) + (1|item).

Participants were faster for non-violation trials compared to violation trials with β = 128.18, *95% CI*[93.90, 162.45], t = 7.33, $p < 0.001$. Participants were also significantly faster for congruent/cognate items compared to incongruent/cognate items with β = 105.64, *95% CI*[89.30, 121.98], t = 12.67, $p < 0.001$, and for incongruent/non-cognates with β = 36.02, *95% CI*[25.35, 46.68], t = 6.62, $p < 0.001$. Importantly, participants in the German-Spanish group were statistically faster compared to the Italian-Spanish group with β = -82.55, *95% CI*[-100.54, -64.56], t = -8.99, $p < 0.001$. Moreover, the interaction effect between *L1* and *condition* was significant for Italian and congruent/cognate items compared to German and incongruent/cognate items with β = -63.19, *95% CI*[-102.49, -23.89], t = -3.15, p = 0.002, with Italian participants being significantly slower (Figure 5.3.2). In sum, we found first, that participants were faster for non-violation compared to violation items; second, that participants were faster for congruent/cognate items than for incongruent/cognate and incongruent/non-cognate items; third, that the German-Spanish group was overall faster compared to the Italian-Spanish group; and fourth, that the German-Spanish group was faster for incongruent/cognate items compared to congruent/cognate items than the Italian-Spanish group. This indicated an effect of *L1* on CLI across the two groups for RTs. See Appendix 5.E for the full model parameters for RTs.

Figure 5.3.2: *Mean response times (ms) for each group for each condition (n = 57).*



5.3.4 EEG data exclusion

EEG trials were excluded based on one of the following reasons: first, the participant had indicated that they were unfamiliar with the noun; second, because the participant made an incorrect grammatical judgement; and third, the trial segment contained an artefact. Therefore, we only included familiar, correct and uncontaminated trials in our analysis, provided that the trial rejection threshold did not exceed 60% of trials per participant. Subsequently, we excluded four participants from the Italian-Spanish group, and four participants from the German-Spanish group. Moreover, one participant from the German-Spanish group was lost due to a recording failure. In total, we included 57 datasets, 29 from the Italian-Spanish group, and 28 from the German-Spanish group. We included the same participants in the behavioural analyses (see previous section 5.3.1).

5.3.5 EEG data pre-processing

We pre-processed our EEG data before the statistical analysis using BrainVision Analyzer (Brain Products, GmbH, Munich). For both groups, we re-referenced to the mastoid electrodes TP9 and TP10 and re-used the original reference channel as a data channel. For the German-Spanish group, we additionally implemented linear derivation to obtain an average HEOG signal. Next, we applied a high-pass filter of 0.1 Hz and a low-pass filter of 30 Hz. We then corrected for residual drift using a maximum amplitude of ± 200 μV for the HEOG channel, and ± 800 μV for the VEOG channel. We used ocular independent component analysis to correct for blink activity using both the VEOG and the HEOG channel as a baseline. We performed artefact correction according to the following criteria: we allowed a maximal voltage step of 50 μV/ms for the gradient, a maximal difference in 100 ms - intervals of 200 μV; maximal amplitudes of $\pm$ 200 μV, and the lowest allowable amplitude in 100 ms - intervals of 0.5 μV. Next, we segmented our data from -200 ms prior to the onset of the stimulus to 1,200 ms after the onset of the stimulus for familiar and correct trials. We applied a baseline correction to each segment using the signal in the 200 ms before stimulus onset. In a final step, we exported all available voltage amplitude samples for each time point, segment, data channel (excluding HEOG, VEOG and the reference channels) and participant to perform our statistical analysis. In this, we exported 29 data channels for the Italian-Spanish group (Fp1, Fp2, Fz, F3, F4, F7, F8, FCz, FC1, FC2, FC5, FC6, Cz, C3, C4, T7, T8, CP1, CP2, CP5, CP6, Pz, P3, P7, P4, P8, Oz, O1 and O2) and 31 channels for the German-Spanish group (Fp1, Fp2, AFz, Fz, F3, F4, F7, F8, FCz, FC3, FC4, FT7, FT8, Cz, CPz, CP3, CP4, C3, C4, T7, T8, TP7, TP8, Pz, P3, P4, P7, P8, Oz, O1 and O2). Each channel was assigned to one of the following topographic regions: left anterior, mid anterior, right anterior; left central, mid central, right central; and finally, left posterior, mid posterior and right posterior regions.

5.3.6 EEG data analysis

For the statistical analysis, we employed a data-driven approach to model *voltage amplitudes* over time. For this, we first conducted a permutation analysis to determine our region of interest (ROI) in terms of channels. Second, we used generalised additive mixed models (GAMMs) to establish our time window of interest for a potential P600 effect (Meulman et al., 2015) and to model group differences in terms of the P600 effect and CLI effects.

To determine our ROI, we performed a cluster-based permutation analysis using the *permutes* package (Voeten, 2019) in R to highlight potentially significant effects of *violation type* and *condition* on *voltage amplitudes*. We visualised the outcomes of the permutation analysis in Figure 5.3.3 for the Italian-Spanish speakers, and in Figure 5.3.4 for the German-Spanish speakers. Potentially significant effects of *violation type* and *condition* are highlighted in red colours. Note that the figure for the German-Spanish speakers is identical to Von Grebmer Zu Wolfsthurn et al. (2021a). For the Italian-Spanish group, the outcome tentatively suggested channels *C4, CP2, CP6, Pz, P3, P4, P7* and *P8* as ROI, these channels were located in centro-parietal regions with a slight left lateralisation. In contrast, for the German-Spanish group the outcome yielded *CPz, CP3, CP4, TP7, TP8, Pz, P3, P4, P7, P8, Oz, O1* and *O2* as a potential ROI. These electrodes were located in left posterior, central posterior and right posterior regions, consistent with the classical topography of the P600 component (Steinhauer et al., 2009).

Figure 5.3.3: *Permutation analysis outcome for the Italian-Spanish group (n = 29). Note that higher F-values are visualised in red colours, and lower F-values in yellow.*

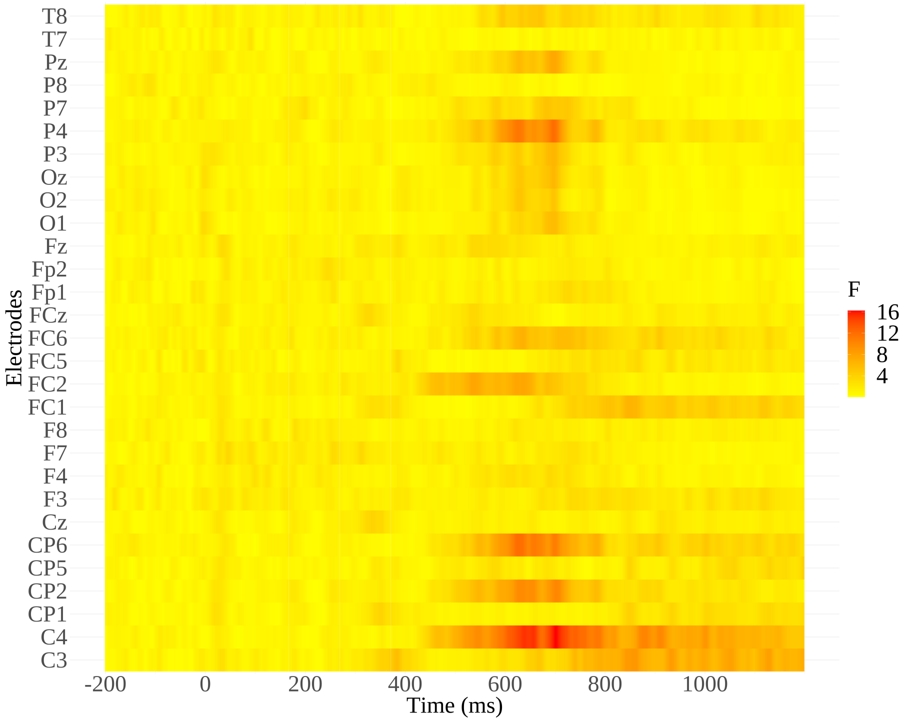
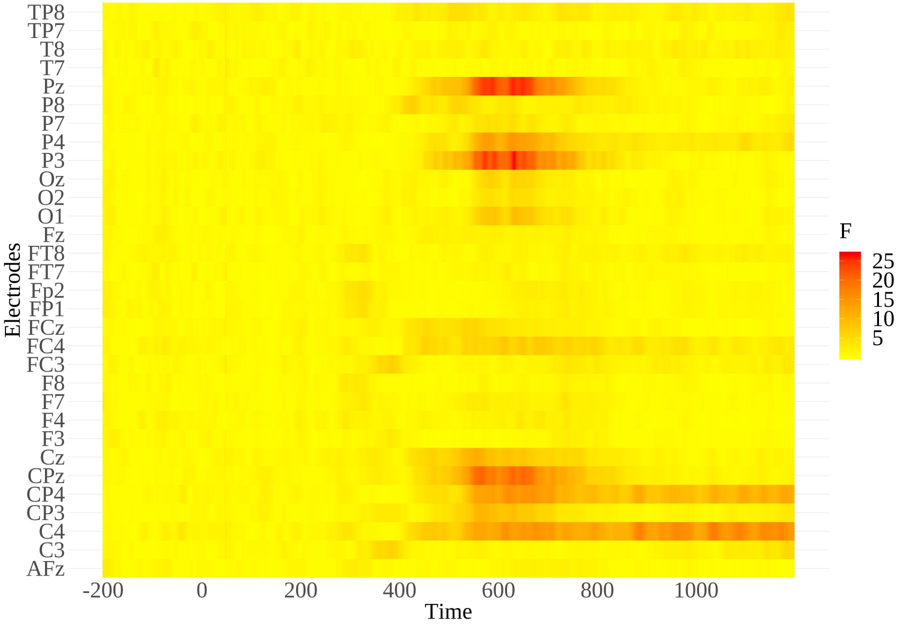


Figure 5.3.4: *Permutation analysis outcome for the German-Spanish group (n = 28). Note that higher F-values are visualised in red colours, and lower F-values in yellow.*



Pooling the ROI channels for both groups, we selected only channels which were present in the montage of each group, namely *Pz, P3, P4, P7* and *P8* as our ROI (Appendix 5.C). In a second step, we modelled *voltage amplitudes* over time in our ROI using a generalised additive mixed model (GAMM) to determine our time window of interest. A detailed discussion of this method and its application in EEG research can be found in Meulman et al. (2015) and in Tremblay and Newman (2015). Briefly, GAMMs not only allow for the inclusion of by-participant and by-item random effects (as do GLMMs), but are also robust against missing data following the missing-at-random mechanism and unbalanced observations per participant. Most importantly, GAMMs allow for the inclusion of non-linear terms to flexibly model the non-linear effects of voltage amplitudes over time, which cannot be captured with linear functions. Here, the non-linear term *time* is modelled flexibly using (penalised) splines, resulting in a smooth fit for the oscillatory trend of voltage amplitudes over time (Meulman et al., 2015). To avoid over-fitting our data, we constructed a simpler, theoretically plausible model which included the interaction effect of *L1* and *violation type*, the interaction effect of *L1* and *condition*, as well as *channel* as a covariate. Next, we added a non-linear term for *time*, and interaction effects between: *time* and *L1*, *time* and *violation type*, *time* and *condition*, and *time* and *channel*. We further created additional variables to test for our critical interaction effects over *time*, namely *L1* and *violation type*, and *L1* and *condition*. Finally, we added random intercepts for *participant* and *item*, random slopes for each participant for the effects of *time*, *violation type*, *condition* and *channel*; and random slopes for each item for the effects of *time* and *channel*. This model was fitted using the *mgcv* package (Wood, 2021) with the fast restricted likelihood estimation (fREML) using a scaled t-distribution to account for heavy tails in the residuals (Meulman et al., 2015). For storage efficiency reasons, we further applied discretisation. We carefully checked the model diagnostics for problematic residual patterns, the appropriate number of basis functions (k-parameter), the goodness of fit and for strong autocorrelation (De Cat et al., 2015). Further, we assumed missing data to be following the missing-at-random (MAR) mechanism (Ibrahim, Chen & Lipsitz, 2001).

To answer our first research question about the presence of a P600 effect in both groups, we used the *itsadug* package (Van Rij, Wieling & Baayen, 2020) in R to plot the predicted differences in voltage amplitudes for non-violation vs. violation trials separately for both groups. This also provided us with a precise time window of interest for the P600 component (Appendix 5.G). For our second research question, we generated conditional plots for the interaction effect of *L1* and *violation type* over time. Similarly, we created conditional plots for the interaction effect of *L1* and *condition* over time to tackle our third research question.

5.3.7 EEG data results

We visualised raw voltage amplitudes for our ROI for each violation type for both groups in Figure 5.3.5, which illustrates the oscillatory trend of voltage amplitudes over time. The first 250 ms post-stimulus onset show the early visual processing response typical for visual stimuli (Eulitz et al., 2000). Critically, the signal yielded a deviation in voltage amplitudes around 450 ms post-stimulus onset across both groups. Descriptively speaking, voltage amplitudes appeared lower for non-violation trials compared to violation trials between 450 ms and 900 ms post-stimulus onset in both groups in Figure 5.3.5, which tentatively suggested a P600 effect for both groups. In contrast, Figure 5.3.6 shows mean voltage amplitudes for each condition for the Italian-Spanish and the German-Spanish group. Importantly, Appendix 5.F visualises the large variance and individual differences in the EEG signal across both groups, which is a critical aspect to keep in mind when dealing with large EEG datasets.

Figure 5.3.5: *Mean voltage amplitudes over time for each violation type for channels Pz, P3, P4, P7 and P8 for both groups.*

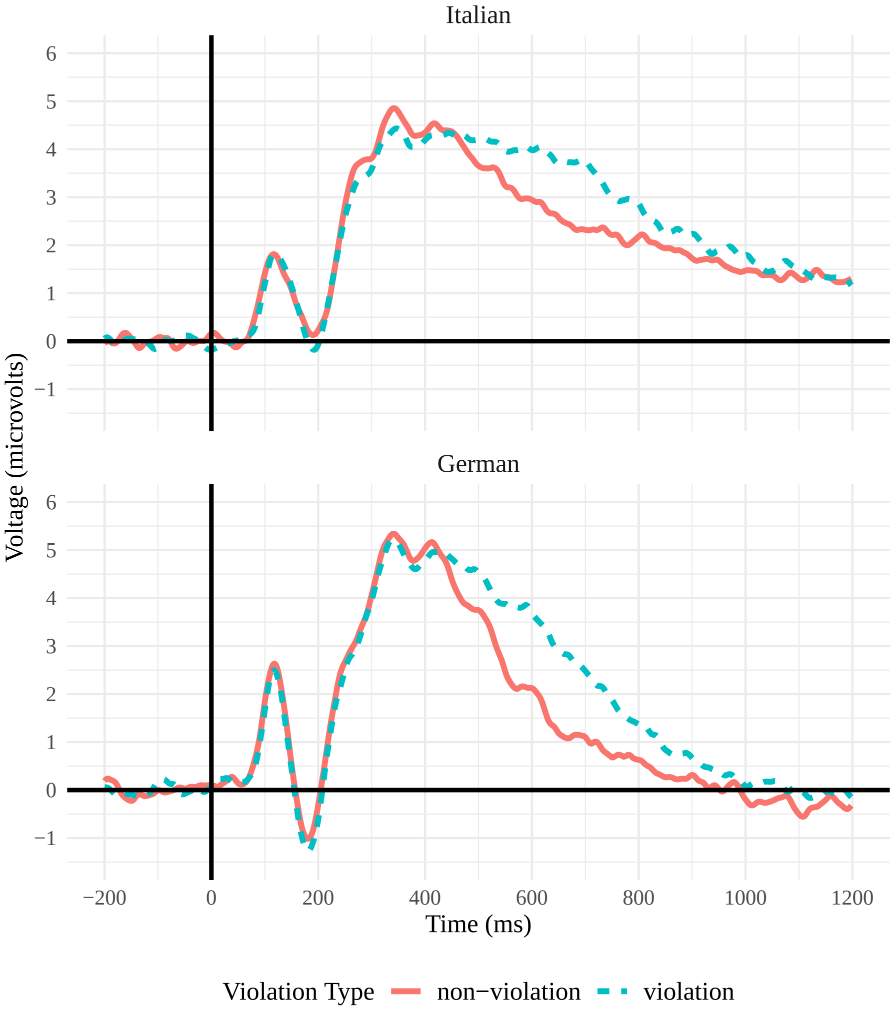
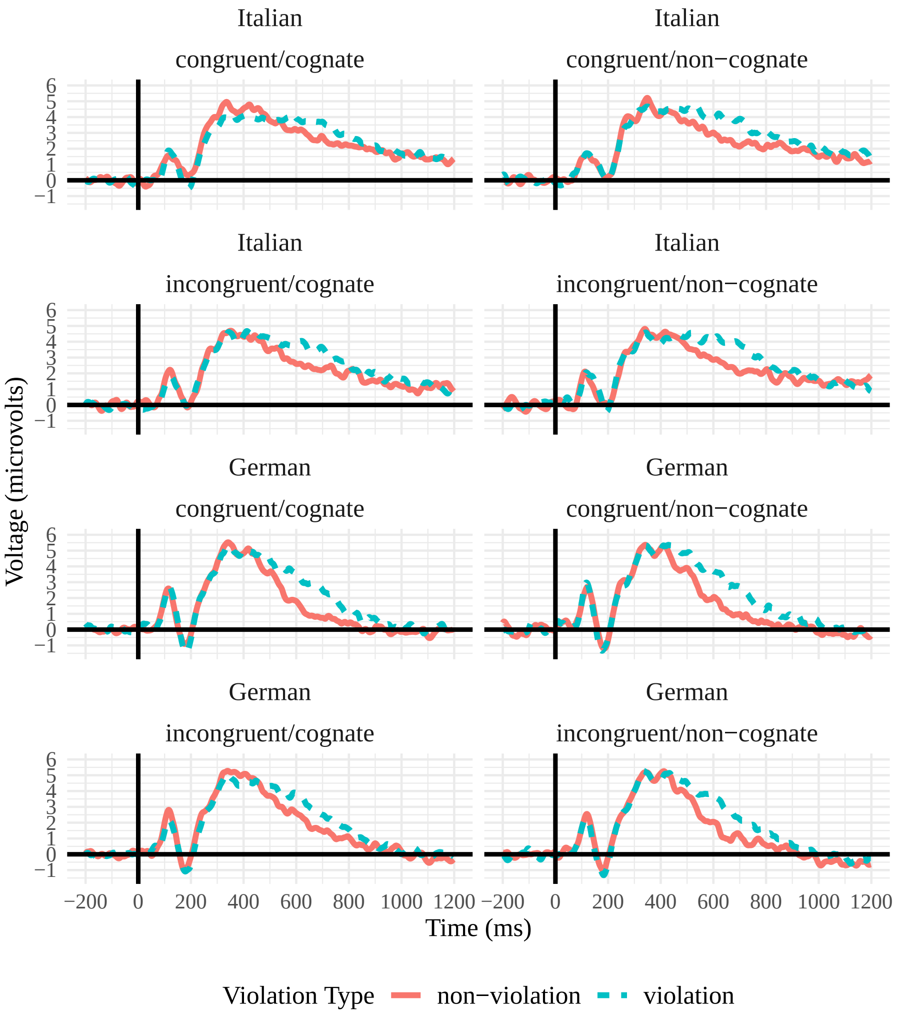


Figure 5.3.6: *Mean voltage amplitudes over time for each condition for channels Pz, P3, P4, P7 and P8 for both groups.*



As described above, our fitted GAMM model was as follows: voltage amplitudes $\sim$ L1 * violation type + L1 * condition + channel + s(time, k = 20) + s(time, by = L1, k = 20) + s(time, by = violation type, k = 20) + s(time, by = condition, k = 20) + s(time, by = L1 *violation type, k = 20) + s(time, by = L1 * condition, k = 20) + s(time, by = channel, k = 20) + s(participant, time, bs = "re") + s(participant, violation type, bs = "re") + s(participant, condition, bs = "re") + s(participant, channel, bs = "re") + s(participant, bs = "re") + s(item, time, bs = "re") + s(item, bs = "re")[1]. See Appendix 5.G for the exact model parameters. The model captured 9.61% of the variance in the data.

With respect to our first research question, we found a significant difference between non-violation and violation trials over time with $F = 636.46$, $p < 0.001$, which is indicative of a P600 effect. We examined this effect individually for each group and found a significant difference between non-violation and violation trials between 477.82 ms and 1056.79 ms post-stimulus onset for the Italian-Spanish group (Figure 5.3.7) and between 491.94 ms and 1056.79 ms for the German-Spanish group (Figure 5.3.8). In addition, the German-Spanish group showed a small difference at 350 ms post-stimulus onset, which is likely linked to the early visual response. Taken together, we found a P600 effect for both the Italian-Spanish group and the German-Spanish group.

---

[1] Our model diagnostics revealed autocorrelation and we subsequently generated a model where we corrected for this autocorrelation (De Cat et al., 2015). However, this model did not reach convergence and is therefore not reported here. Importantly, while the correction for autocorrelation may have a small impact on the model parameters, it likely does not affect the overall results.

Figure 5.3.7: *Marginal plot of predicted differences in voltage amplitudes over time for violation vs. non-violations for channels Pz, P3, P4, P7 and P8 for the Italian-Spanish group (n = 29).*

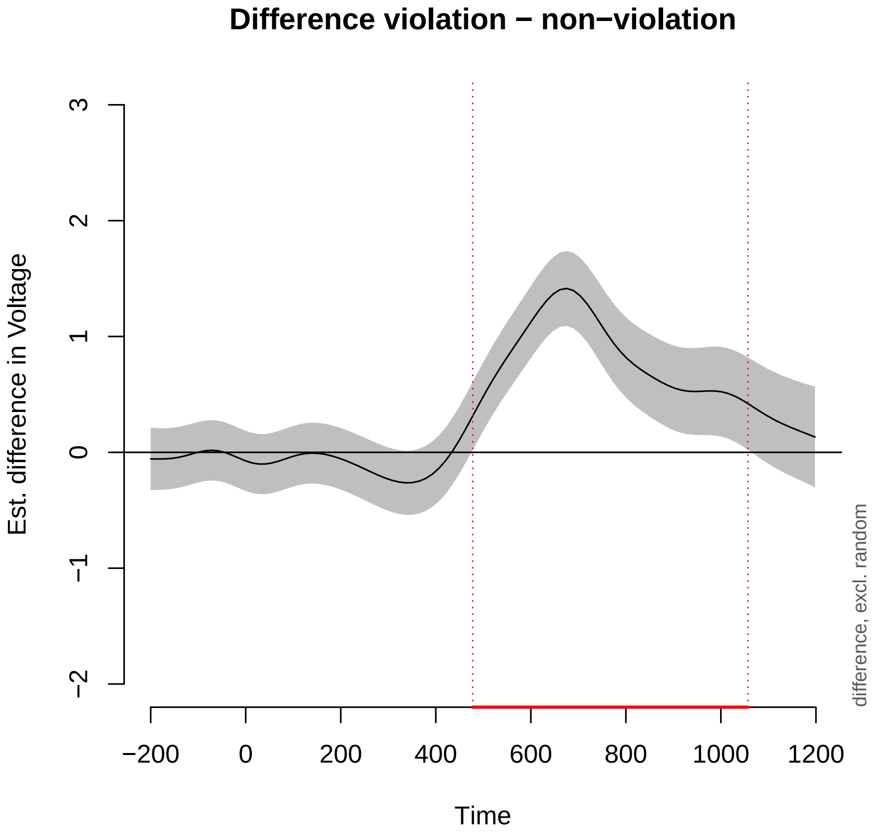
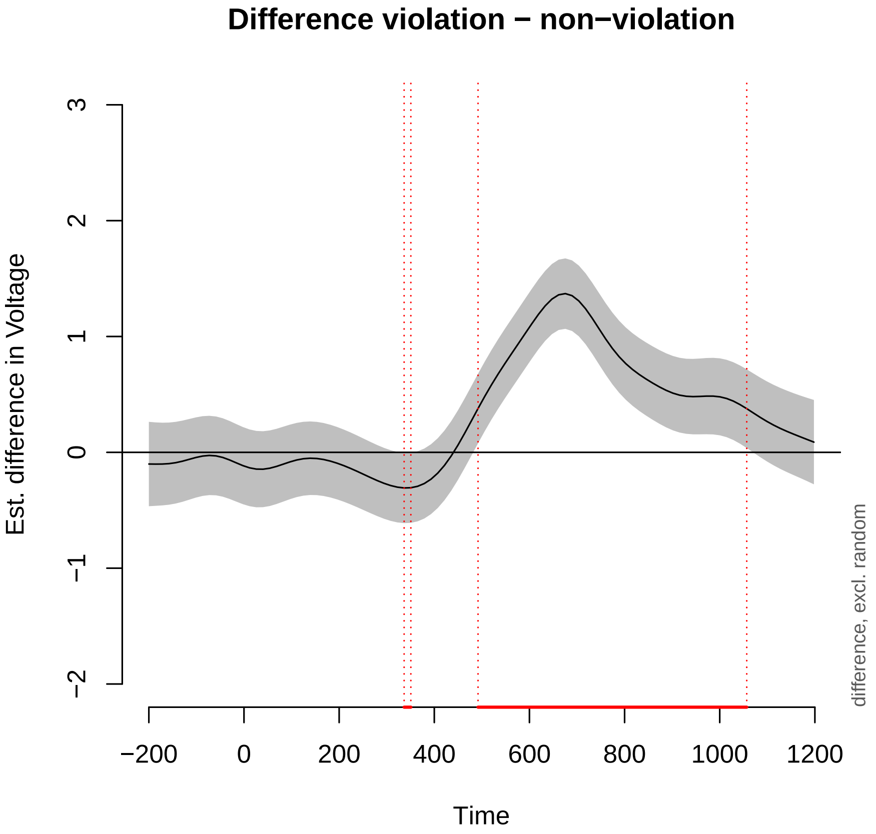
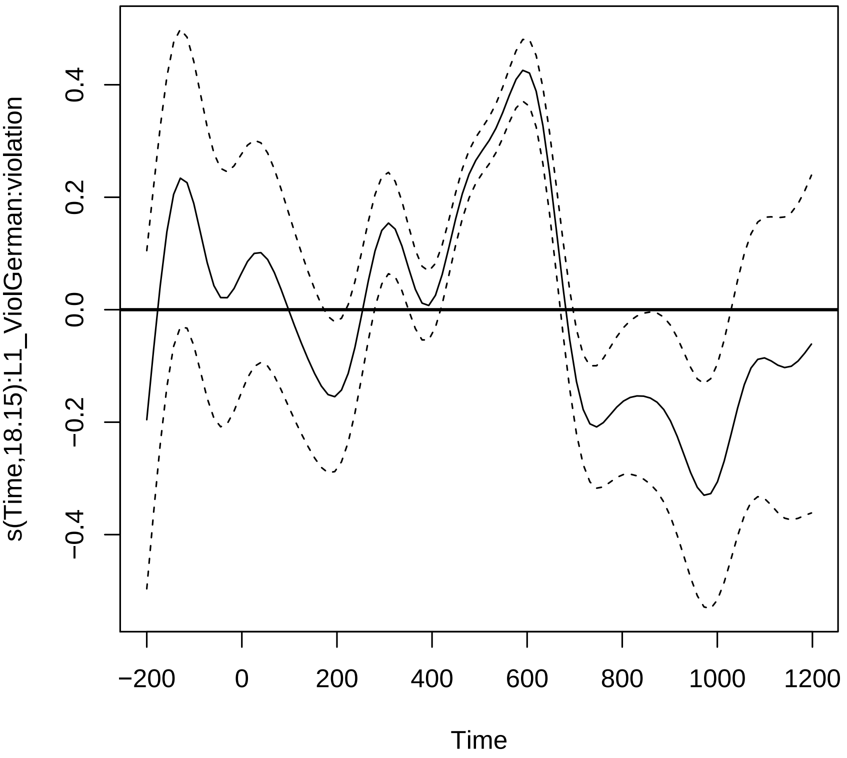


Figure 5.3.8: *Marginal plot of predicted differences in voltage amplitudes over time for violation vs. non-violations for channels Pz, P3, P4, P7 and P8 for the German-Spanish group (n = 28).*



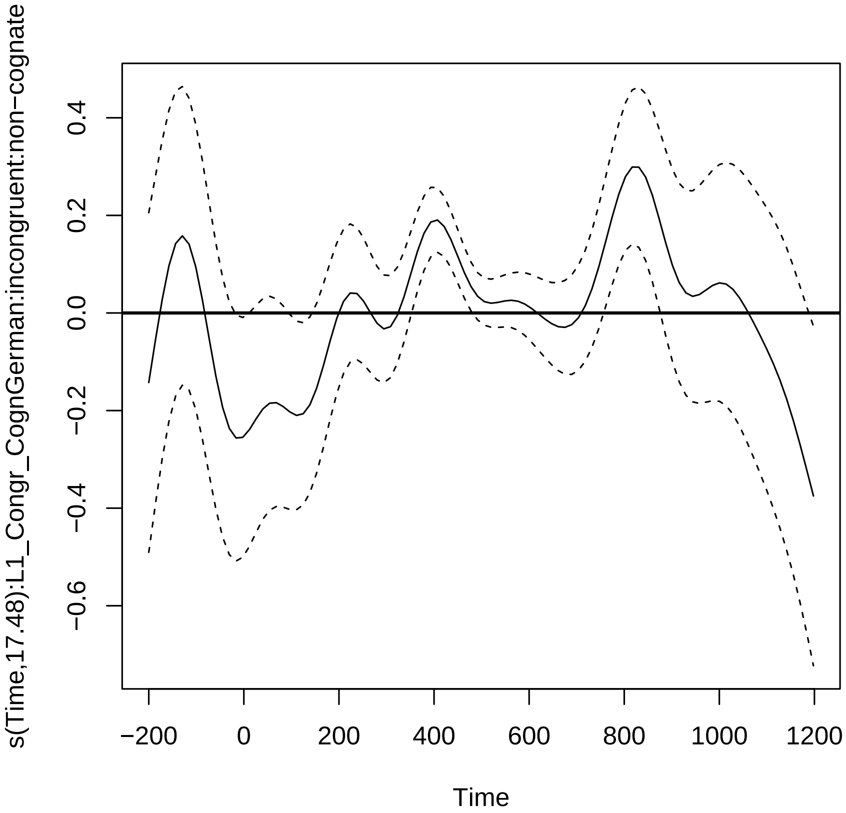
With respect to our second research question, the interaction effect of *L1* and *violation type* was significant over time with $F = 61.46$, $p < 0.001$. The conditional plot suggested a small, but robust difference in the P600 effect between the two groups (Figure 5.3.9). Figure 5.3.9 visualises this difference in voltage amplitudes between non-violation trials vs. violations trials over time for the Italian-Spanish group compared to the German-Spanish group. This figure shows a significant non-zero difference in P600 effects around 600 ms, with a larger P600 effect linked to the Italian-Spanish group compared to the German-Spanish group (Figure 5.3.9). The effect difference was close to zero for the the remaining time points and therefore not significant. Note that Figure 5.3.9 visually suggests a large difference in P600 effect size across the two groups, but was in fact much smaller as predicted by the model (Appendix 5.G). We captured this notion in Appendix 5.H, which shows this small difference in P600 effects in relation to our original scale.

Figure 5.3.9: *Conditional plot of predicted difference in voltage amplitudes over time for violations vs. non-violations for channels Pz, P3, P4, P7 and P8 across both groups (n = 57). The dashed lines represent the standard error.*



With respect to our third and final research question, the interaction effect of *L1* and *condition* was significant over time with $F = 29.30$, $p < 0.001$. This suggested that CLI effects differed over time between the groups. The conditional plot for this particular effect showed a small difference at two separate time points post-stimulus onset (Figure 5.3.10). More specifically, CLI effects were significantly larger around 400 ms and around 800 ms for the Italian-Spanish group compared to the German-Spanish group. For the remaining time points, the difference in CLI effects was close to zero and therefore not significant. Importantly, as Appendix 5.I shows, these differences in CLI effects across the two groups are small, but statistically significant according to the model. See Appendix 5.G for the exact model parameters.

Figure 5.3.10: *Conditional plot of predicted difference in voltage amplitudes over time for the CLI effects for channels Pz, P3, P4, P7 and P8 across both groups (n = 57). The dashed lines represent the standard error.*



In summary, our ERP findings were the following: first, we found evidence for a P600 effect for both groups. This was indicated by higher voltage amplitudes for violation trials compared to non-violation trials. Second, results suggested a statistically larger P600 effect around 600 ms for the Italian-Spanish compared to the German-Spanish group over time. Finally, voltage amplitudes connected to CLI effects were larger around 400 ms and around 800 ms for the Italian-Spanish compared to the German-Spanish group.

5.4 Discussion

The primary aim of the current study was to investigate typological similarity effects on cross-linguistic influence (CLI) and on the neural correlates of syntactic violation processing at the behavioural and neural level. More specifically, we examined typological similarity effects using a syntactic violation paradigm in speakers of typologically similar languages (Italian-Spanish) and of typologically less similar languages (German-Spanish), all of whom were late learners of Spanish. During the syntactic violation paradigm, we measured accuracy, RTs and voltage amplitudes over time, with a particular focus on the P600 component. We probed first, whether there was a P600 effect across both groups; second, whether this potential P600 effect was larger for one group compared to the other; and third, whether there were different CLI effects across the two groups. On the basis of the LDH (Zawiszewski & Laka, 2020) outlined in the introduction, we predicted an overall processing advantage for the Italian-Spanish group compared to the German-Spanish group.

From a behavioural perspective, we first predicted that speakers would be more accurate and faster for non-violation compared to violation trials, and for congruent/cognate items compared to incongruent/non-cognate items. Next, we hypothesised that the Italian-Spanish group would be more accurate and faster for non-violation than for violation trials compared to the German-Spanish group. Finally, we predicted that the Italian-Spanish group would be more accurate and faster at processing congruent/cognate items than for incongruent/non-cognate items compared to the German-Spanish group. This would reflect first, an advantage for speakers of typologically similar languages (Italian-Spanish) in detecting syntactic violations compared to speakers of typologically less similar languages (German-Spanish); and second, more pronounced CLI effects for the Italian-Spanish group.

Behavioural results suggested the following: for accuracy, we found that participants were indeed more accurate for non-violation trials compared to violation trials, in line with our hypothesis. Next, we also found a small effect of condition, indicating a difference in accuracy as a function of CLI. Here, participants were more accurate for congruent/cognate items compared to incongruent/cognate items, thereby suggesting a small effect of gender congruency. However, with respect to our second and third research question, results from accuracy indicated neither an influence of typological similarity on syntactic violation processing, nor on overall CLI effects as both critical interaction effects yielded non-significance.

In contrast, results from RTs provided us with a more extensive picture. Participants were faster for non-violation trials compared to violation items, and for congruent/cognate items compared to incongruent/cognate and incongruent/non-cognate items. This yields a processing advantage for congruent/cognates compared to incongruent/cognates both at the level of accuracy and RTs. One possible interpretation of this particular result could be that incongruent/cognates are potentially particularly difficult to process because of the simultaneous occurrence of similarity at the word form level and the unexpected mismatch at the gender level. Subsequently, the processing effort for incongruent/cognates may be comparatively high in contrast to cases where the similarity manifests itself at the word form level as well as at the gender level. Another critical finding was that Italian-Spanish speakers were overall slower compared to the German-Spanish speakers. This suggests a more general processing advantage for the typologically less similar German-Spanish pair compared to the Italian-Spanish pair in terms of RTs. Critically, with respect to our second research question about the differential processing of violation vs. non-violation trials across groups, we did not find evidence for this notion, contrary to our behavioural predictions. As for our third research question about differential CLI effects across groups, we found a difference in CLI across the two groups, but in the opposite direction to what we had predicted: Italian-Spanish speakers were significantly slower for incongruent/cognates compared to the German-Spanish speakers.

Taken together, the main finding from our behavioural results was the small effect of typological similarity both on overall RTs but also on CLI: the German-Spanish speakers, but not the Italian-Spanish speakers, displayed an overall behavioural processing advantage in this task. This was both in terms of faster RTs when detecting syntactic violations and in terms of overall smaller CLI effects. This notion is contrary to the predictions made by the LDH (Zawiszewski & Laka, 2020).

There are several possible interpretations of these findings: first, that there was less CLI for the German-Spanish speakers to begin with and therefore they were less subject to CLI effects compared to the Italian-Spanish speakers. Therefore, the processing advantage for the German-Spanish group could be a natural consequence of being less subject to CLI. A second interpretation is that CLI was equally pronounced in both groups, but the German-Spanish speakers employed a more efficient strategy to mitigate CLI effects compared to the Italian-Spanish speakers. Finally, the predictions of the LDH (Zawiszewski & Laka, 2020) may be limited to morphosyntactic similarity and may not apply to similarity at the level of gender and word form overlap as tested in this current study, at least in terms of behaviour. Our current design does not allow for the discrimination of these interpretations, but they should be subject to future research. Nevertheless, these results provide evidence for an effect of typological similarity on CLI favouring speakers of typologically less similar languages. To get a clearer interpretation of our findings, in the next section we corroborated these behavioural findings with the ERP findings.

In terms of ERPs, we first expected a P600 effect for both groups. In line with our predictions, we found significantly higher voltage amplitudes for violation trials compared to non-violation trials for both the Italian-Spanish and the German-Spanish group, which reflects the classical P600 effect (Friederici et al., 1999, 2002; Sabourin & Stowe, 2008; Swaab et al., 2011). In turn, this indicated that both groups were highly sensitive to syntactic violations at the level of gender. Notably, both groups displayed a highly similar onset of the P600 effect around 490 ms post-stimulus onset, as well as a comparable P600 effect latency until around 1,000 ms post-stimulus onset. Therefore, answering to our first research question, our data suggest a P600 effect for both the Italian and the German late learners of Spanish (S. Rossi et al., 2006; Tokowicz & MacWhinney, 2005).

For our second research question, we predicted a larger, more native-like P600 effect for the typologically more similar Italian-Spanish group compared to the typologically less similar German-Spanish group, in line with the LDH (Zawiszewski & Laka, 2020). Supporting this prediction, our data provided evidence for a small, but robust statistical difference in P600 effect sizes (around 600 ms post-stimulus onset), with a larger P600 effect for the Italian-Spanish group than for the German-Spanish group. This indicates a processing advantage for typologically more similar languages compared to less similar languages. Further, these findings corroborate the results by Sabourin and Stowe (2008), who reported a larger P600 effect for the typologically more similar language combination of German and Dutch compared to the combination of Romance languages and Dutch when processing syntactic violations in the non-native language Dutch. By extension, the results from our study support the notion of enhanced sensitivity to syntactic violations in speakers of typologically more similar languages compared to less similar languages, i.e., Italian-Spanish vs. German-Spanish, see also Sabourin and Stowe (2008). Therefore, as for our second research question, we provide evidence that typological similarity directly impacts P600 effect sizes. This notion expands on work by Zawiszewski and Laka (2020), who demonstrated a modulation of ERP effects by morphological similarity in highly proficient speakers. Therefore, our study contributes novel findings about the facilitatory role of gender similarity and word form similarity to existing accounts on the role of morphosyntactic similarity on non-native comprehension.

For our third research question, we predicted larger CLI for the Italian-Spanish group compared to German-Spanish group, as reflected in larger, more native-like voltage amplitudes for CLI effects for the typologically more similar group. In line with our predictions, we found that CLI effects were larger for the Italian-Spanish group compared to the German-Spanish group. Subsequently, this represents evidence for a modulation of CLI by typological similarity, as well as a processing advantage for typologically similar languages compared to less similar languages. These results extend the LDH by Zawiszewski and Laka (2020) in that we provide evidence that also similarity at the level of gender and orthographic and phonological form overlap (cognate status) elicited more native-like, larger ERP components. Moreover, these results are also in line with studies suggesting overall larger CLI for typologically similar languages compared to less similar languages (Mosca, 2017; Sabourin & Stowe, 2008; Tolentino & Tokowicz, 2011).

Taking both the behavioural and the ERP data together, results suggested a general typological similarity effect on non-native comprehension. Interestingly, however, they indicated a typological effect in opposite directions: on the one hand, the behavioural data suggested a behavioural processing disadvantage for the Italian-Spanish group in the form of overall slower RTs and slower RTs for processing CLI compared to the German-Spanish group. This contrasts with our predictions on the basis of the LDH (Zawiszewski & Laka, 2020). In turn, it could imply that the model's behavioural predictions were only applicable to morphosyntactic similarity but not to overlap at the level of gender, orthography and phonology. On the other hand, the ERP data suggested a processing advantage for the Italian-Spanish group at the neural level, with larger and more native-like P600 effects and larger CLI effects compared to the German-Spanish group. These results support the predictions of the LDH (Zawiszewski & Laka, 2020). Our interpretation is that the notion of larger, more native-like ERPs for similar languages holds not only for morphological similarity, but also for gender system similarity, and orthographic and phonological word form similarity.

Differential findings across behavioural data and ERP data are not uncommon in the non-native language processing literature

(Acheson et al., 2012; Bosma & Pablos, 2020; Jiao et al., 2020). In this current study, behavioural findings support a processing advantage for typologically less similar languages, whereas neural findings support a processing advantage for typologically similar languages. Critically, we argue that this contrasts highlights the complex association between behavioural and neural cognitive mechanism, which goes far beyond the more traditional interpretation that neural measures index ongoing processes and behavioural measures index the outcomes of those processes (White, Genesee & Steinhauer, 2012). Moreover, our contrasting results could also indicate that typological similarity effects differ not only across behavioural and neural measures, but potentially also in terms of the different linguistic domains, such as phonological similarity, orthographic similarity or lexico-semantic similarity. Our study design and research questions did not allow for a more nuanced investigation of whether the typological similarity effect is in fact an interplay between several similarity effects across different domains. Therefore, more refined research is needed first, to tease apart a potentially differential impact of typological similarity on behaviour and neural correlates; and second, to characterise typological similarity effects not as a unified effect, but as a combination of individual similarity effects.

Another direction for future research is concerned with examining the exact role of proficiency in modulating typological similarity effects more closely. As discussed in the introduction, some studies suggested more pronounced typological similarity effects at lower proficiency levels (Tokowicz & MacWhinney, 2005). However, more direct comparisons are needed between typological similarity effects at different levels of non-native proficiency and AoA. This was beyond the scope of the current study, but will be essential for characterising the role of typological similarity on non-native processing more broadly and to model the potentially dynamic effects of typological similarity over time with evolving proficiency levels.

Returning to our broader question of whether typological similarity impacts non-native processing, the results of this study suggest an affirmative answer. In turn, this notion indicates that the L1 and the non-native language are intrinsically linked with each other in our late language learners at this specific proficiency level. However, since studies on this particular topic are scarce, we argue for the need of more comprehensive studies to tackle this question in a more nuanced manner.

5.4.1 Conclusions

In this study, we investigated typological similarity effects in non-native comprehension in Italian-Spanish speakers (typologically similar group) and German-Spanish speaker (typologically less similar pair). On the basis of the Language Distance Hypothesis, LDH (Zawiszewski & Laka, 2020), we predicted a processing advantage for speakers of the typologically more similar language pair, as reflected in higher accuracy, shorter RTs and larger, more native like P600 amplitudes during a syntactic violation paradigm. We found different typological similarity effects: on the one hand, the Italian-Spanish speakers were overall slower during the task compared to the German-Spanish speaker. On the other hand, ERP evidence showed a larger P600 effect for the Italian-Spanish speakers as well as larger voltage amplitudes for CLI compared to the German-Spanish speakers. This latter finding was in line with the LDH (Zawiszewski & Laka, 2020). Therefore, our results indicate a general typological similarity effect at the level of both behavioural and neural measures. Moreover, our results suggest an intimate functional link between the L1 and the non-native language in the multilingual brain. Questions remain as to whether typological similarity effects are uniform across behavioural and neural measures and whether they are equally pronounced across different linguistic domains and proficiency levels.

CRediT author contribution statement

Sarah Von Grebmer Zu Wolfsthurn: Conceptualisation, Methodology, Validation, Investigation, Formal Analysis, Data Curation, Writing-Original Draft, Writing-Review and Editing, Visualisation. **Leticia Pablos-Robles**: Conceptualisation, Methodology, Writing-Review and Editing, Supervision. **Niels O. Schiller**: Conceptualisation, Writing-Review and Editing, Supervision, Funding Acquisition.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported here.

Acknowledgements

We thank all members of the Neurolinguistics Laboratory at the University of Konstanz for their support, in particular Carsten Eulitz, Oleksiy Bobrow, Charlotte Englert, Khrystyna Oliynyk, Anna-Maria Waibel, Zeynep Dogan and Fangming Zhang. We also thank Núria Sebastián Gallés and Ege Ekin Özer from the Speech Acquisition and Perception group at the Center for Brain and Cognition at Pompeu Fabra University. We further thank Marjolein Fokkema, Julian Karch and Franz Wurm for their support in the statistical analyses. A special thanks goes to Frank Doornkamp and Willem Hoogerworst, who consulted on the analysis for the EEG data and implemented the initial GAMM analysis. The statistical analyses were performed using the computing resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. Finally, we also want to thank all of our participants for taking part in this research.

Funding statement

This project has received funding from the European Union's Horizon2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N^o 765556 – The Multilingual Mind.

Data availability statement

The data that support the findings of this study are openly available in the Open Science Framework at: `https://osf.io/e6acy/?view_only=798087b3751b46d88654f76eaf26ec67`

Citation diversity statement

We included this statement to make readers aware of research suggesting that authors identifying as female or as members of a minority group are underrepresented in the reference list of scientific studies (Dworkin et al., 2020; Zurn et al., 2020). This is a particularly important topic to consider during a global pandemic (Viglione, 2020). Our references included 23% woman/ woman authors, 36% man/ man, 24% woman/ man and finally, 8% man/ woman authors. This compares to 6.7% for woman/ woman, 58.4% for man/ man, 25.5% woman/ man, and lastly, 9.4% for man/ woman authored references in the reference list of publications in the field of neuroscience (Dworkin et al., 2020).

Appendix

5.A Linguistic profile: Italian-Spanish group

Table 5.A.1: *Overview of the native and non-native languages acquired by the Italian-Spanish group (n = 29).*

	L1	**L2**	**L3**	**L4**	**L5**	**Total**
Italian	n = 29					**29**
Spanish		n = 2	n = 17	n = 8	n = 2	**29**
English		n = 24	n = 4			**28**
French		n = 3	n = 6	n = 3		**12**
German			n = 1	n = 1		**2**
Catalan				n = 1	n = 1	**2**
Portuguese					n = 3	**3**
Total	**29**	**29**	**28**	**13**	**6**	

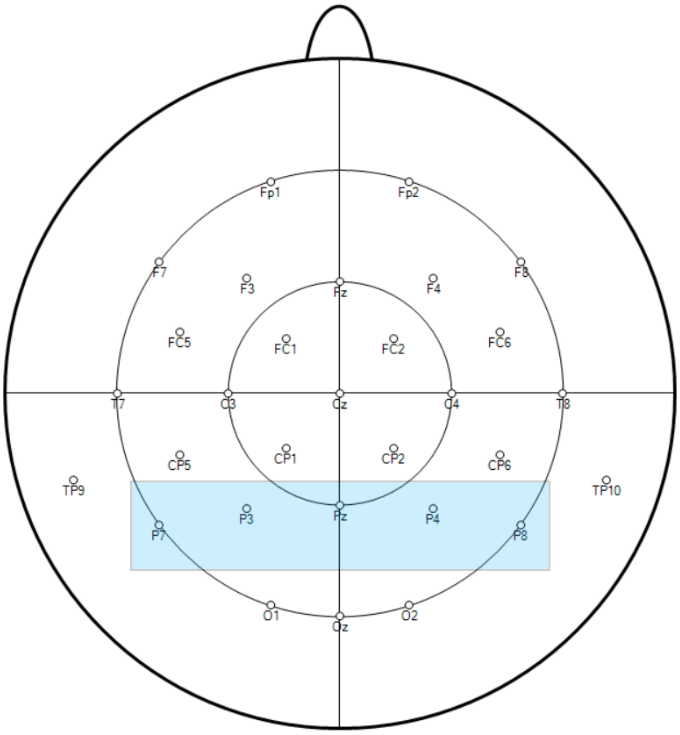
5.B Linguistic profile: German-Spanish group

Table 5.B.1: *Overview of the native and non-native languages acquired by the German-Spanish group (n = 28).*

	L1	**L2**	**L3**	**L4**	**L5**	**Total**
German	n = 28					**28**
Spanish			n = 15	n = 11	n = 2	**28**
English		n = 26	n = 2			**28**
French		n = 2	n = 8	n = 5		**15**
Latin			n = 2	n = 1	n = 1	**4**
Russian			n = 1		n = 1	**2**
Swedish				n = 1		**1**
Portuguese					n = 1	**1**
Arabic					n = 1	**1**
Catalan					n = 1	**1**
Italian					n = 1	**1**
Mandarin					n = 1	**1**
Total	**28**	**28**	**28**	**18**	**9**	

5.C EEG data: region of interest

Figure 5.C.1: *Region of interest and the corresponding channels Pz, P3, P4, P7 and P8 for the EEG analysis, shown in the shaded area in the montage of the Italian-Spanish group.*



5.D Model parameters: accuracy

Table 5.D.1: *Model parameters for the best-fitting model for accuracy (n = 57).*

Formula: accuracy ~ L1 (Italian vs. German) + violation type (violation vs. non-violation) + L1 * condition (congruent/cognate vs. congruent/non-cognate vs. incongruent/cognate vs. incongruent/non-cognate) + LexTALE-Esp score + target noun gender (feminine vs. masculine) + (violation type|participant) + (1|item)

Term	**Odds Ratio [95% CI]**	**z-value**	**p-value**
(Intercept)	65.65 [32.57, 132.35]	11.70	< 0.001
L1 [German]	1.50 [0.673, 3.35]	0.993	0.321
Violation type [violation]	0.412 [0.279, 0.609]	-4.45	< **0.001**
Condition [congruent/non-cognate]	1.63 [0.752, 3.54]	1.24	0.215
Condition [incongruent/ cognate]	0.258 [0.132, 0.504]	-3.96	< **0.001**
Condition [incongruent/ non-cognate]	0.714 [0.342, 1.49]	-0.897	0.370
LexTALE-Esp score	1.02 [1.01, 1.03]	4.15	< 0.001
Target noun gender [m]	0.686 [0.478, 0.986]	-2.04	0.042
L1 [German] * Condition [congruent/ non-cognate]	0.349 [0.121, 1.01]	-1.94	0.052
L1 [German] * Condition [incongruent/ cognate]	1.68 [0.631, 4.46]	1.04	0.300
L1 [German] * Condition [incongruent/ non-cognate]	0.661 [0.238, 1.84]	-0.792	0.428

Random effects	
σ^2	3.29
$\tau_{00 Item}$	1.77
$\tau_{00 Participant}$	0.36
$\tau_{11 Participant[non-violation]}$	0.16
$\rho_{01 Participant}$	-0.35
ICC	0.39
$N_{Participant}$	57
N_{Item}	448
Observations	9,972
Marginal R^2/ Conditional R^2	0.111/0.461

5.E Model parameters: response times

Table 5.E.1: *Model parameters for the best-fitting model for RTs (n = 57).*

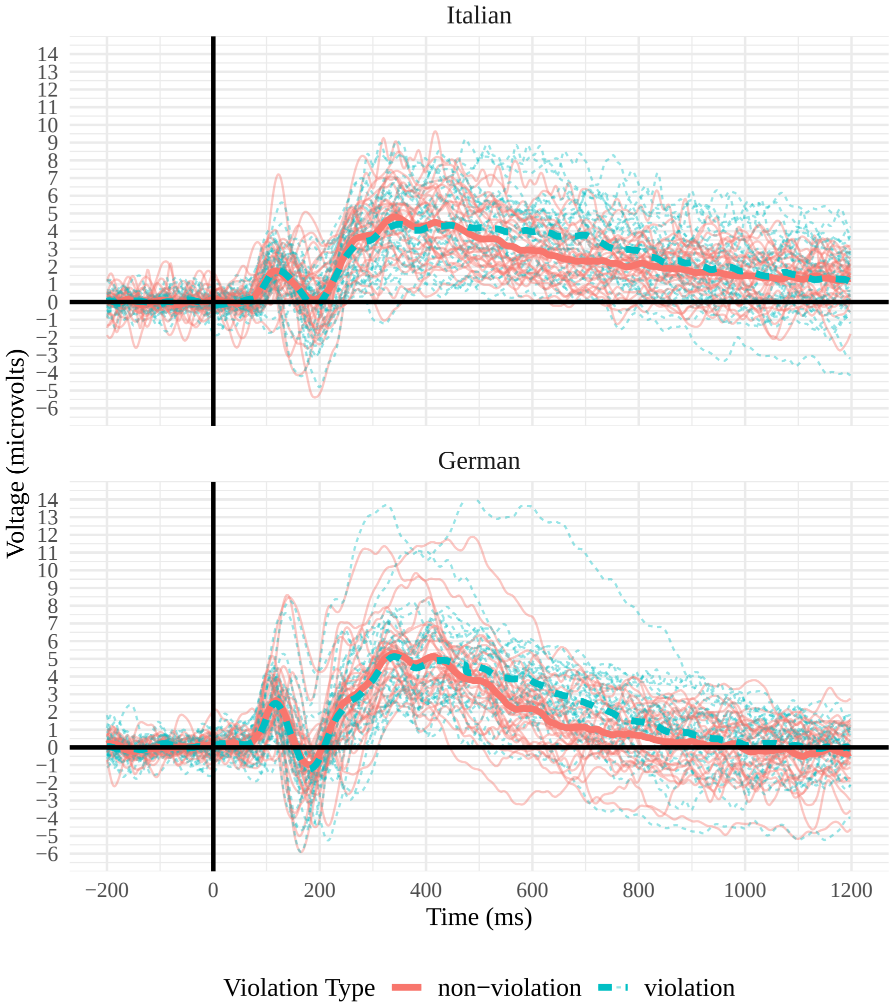
Formula: RTs ∼ L1 (Italian vs. German) + violation type (violation vs. non-violation) + L1 * condition (congruent/cognate vs. congruent/non-cognate vs. incongruent/cognate vs. incongruent/non-cognate) + target noun gender (feminine vs. masculine) + (violation type|participant) + (1|item)

Term	**Estimate [95% CI]**	**t-value**	**p-value**
(Intercept)	854.22 [791.17, 917.26]	26.56	< 0.001
L1 [German]	-82.55 [-100.54, -64.56]	-8.99	$<$ **0.001**
Violation type [violation]	128.18 [93.91, 162.45]	7.33	$<$ **0.001**
Condition [congruent/ non-cognate]	-2.05 [-47.75, 43.65]	-0.088	0.930
Condition [incongruent/ cognate]	105.64 [89.31, 121.98]	12.67	$<$ **0.001**
Condition [incongruent/ non-cognate]	36.02 [25.36, 46.67]	6.62	$<$ **0.001**
Target noun gender [m]	14.47 [-9.69, 38.63]	1.17	0.241
L1 [German] * Condition [congruent/ non-cognate]	0.297 [-49.29, 49.88]	0.012	0.991
L1 [German] * Condition [incongruent/ cognate]	-63.19 [-102.49, -23.89]	-3.15	**0.002**
L1 [German] * Condition [incongruent/ non-cognate]	-27.28 [-71.27, 16.71]	-1.22	0.224
Random effects			
σ^2	0.14		
$\tau_{00 Item}$	3966.19		

$\tau_{00 Participant}$	8558.85
$\tau_{11 Participant[non-violation]}$	5839.45
$\rho_{01 Participant}$	-0.18
ICC	1.00
$N_{Participant}$	57
N_{Item}	448
Observations	9,393
Marginal R^2/ Conditional R^2	0.359/1.00

5.F EEG data: by-violation type mean voltage amplitudes

Figure 5.F.1: *Mean voltage amplitudes over time for each violation type for each participant for channels Pz, P3, P4, P7 and P8 (n = 57). Mean amplitudes for violation type are shown as thicker lines.*



5.G Model parameters: P600 component

Table 5.G.1: *Model parameters of the GAMM model for the effect of L1 and time on voltage amplitudes for channels Pz, P3, P4, P7 and P8 (n = 57). Estimated degrees of freedom (edf) provide a measure for the complexity of the smooth terms. The edf parameters for our smooth terms suggested that voltage amplitudes follow a highly non-linear tendency.*

Formula: voltage amplitudes ∼ L1 * violation type + L1 * condition + channel + s(time, k = 20) + s(time, by = L1, k = 20) + s(time, by = violation type, k = 20) + s(time, by = condition, k = 20) + s(time, by = L1 * violation type, k = 20) + s(time, by = L1 * condition, k = 20) + s(time, by = channel, k = 20) + s(participant, time, bs = "re") + s(participant, violation type, bs = "re") + s(participant, condition, bs = "re") + s(participant, channel, bs = "re") + s(participant, bs = "re") + s(item, time, bs = "re") + s(item, bs = "re")

Linear terms	**Estimate**	**SE**	**t-value**	**p-value**
(Intercept)	2.32	0.243	9.58	< 0.001
L1 [German]	-0.541	0.301	-1.80	0.072
Violation type [violation]	0.328	0.151	2.17	0.029
Condition [congruent/ non-cognate]	0.044	0.171	0.255	0.798
Condition [incongruent/ cognate]	0.003	0.171	0.018	0.985
Condition [incongruent/ non-cognate]	0.083	0.181	0.460	0.645
Channel [P4]	0.490	0.176	2.78	0.005
Channel [P7]	-2.18	0.176	-12.39	< 0.001
Channel [P8]	-1.20	0.176	-6.81	< 0.001
Channel [Pz]	0.496	0.176	2.82	0.005
L1 [German] * Violation type [violation]	-0.044	0.214	-0.207	0.863
L1 [German] * Condition [congruent/ non-cognate]	0.039	0.220	0.176	0.860

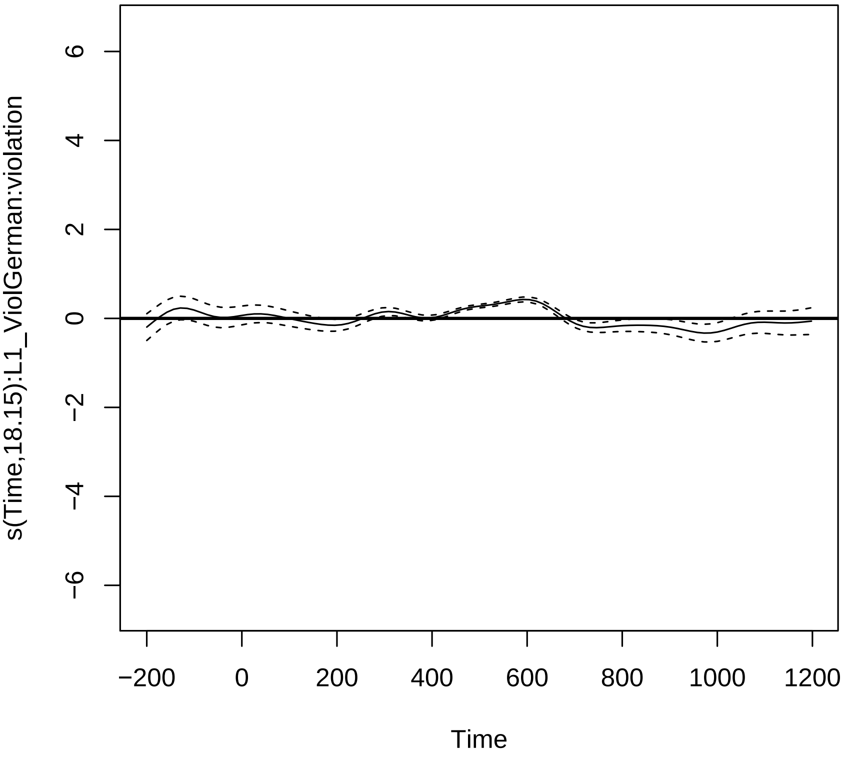
L1 [German] * Condition [incongruent/ cognate]	-0.016	0.220	-0.071	0.943
L1 [German] * Condition [incongruent/ non-cognate]	-0.152	0.250	-0.608	0.543

Non-linear terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Time)	17.97	18.00	3971.89	< 0.001
s(Time) * L1 [German]	18.89	18.99	653.49	< 0.001
s(Time) * Violation type [violation]	17.86	18.77	636.46	< **0.001**
s(Time) * Condition [congruent/ non-cognate]	17.51	18.66	25.98	< 0.001
s(Time) * Condition [incongruent/ cognate]	17.73	18.74	27.93	< 0.001
s(Time) * Condition [incongruent/ cognate]	17.98	18.76	25.37	< 0.001
s(Time) * [German/ violation]	18.15	18.85	61.46	< **0.001**
s(Time) * [German/ incongruent/ non-cognate]	17.48	18.62	29.30	< **0.001**
s(Time) * Channel [P3]	18.88	19.00	249.47	< 0.001
s(Time) * Channel [P4]	1.00	1.00	43.60	< 0.001
s(Time) * Channel [P7]	18.98	19.00	2619.93	< 0.001
s(Time) * Channel [P8]	18.98	19.00	1385.98	< 0.001
s(Time) * Channel [Pz]	18.98	19.00	474.41	< 0.001

s(Time, Participant	54.98	55.00	16133637.33	< 0.001
s(Violation type, Participant)	71.20	114.00	1476223.16	1.00
s(Condition, Participant)	174.60	226.00	275408.80	1.00
s(Channel, Participant)	252.74	284.00	1563283.21	1.00
s(Participant)	0.003	57.00	0.166	1.00
s(Time, Item)	437.67	441.00	77707.90	0.017
s(Item)	432.69	444.00	75336.34	0.217

5.H P600 effect sizes: unscaled predicted differences

Figure 5.H.1: *Conditional plot of predicted difference in voltage amplitudes over time for violation vs. non-violations for channels Pz, P3, P4, P7 and P8 across both groups (n = 57) on the original scale. The dashed lines represent the standard error.*



5.I CLI effect sizes: unscaled predicted differences

Figure 5.I.1: *Conditional plot of predicted difference in voltage amplitudes over time for the CLI effects for channels Pz, P3, P4, P7 and P8 across both groups (n = 57) on the original scale. The dashed lines represent the standard error.*