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

Von Grebmer Zu Wolfsthurn, S.

CHAPTER 6

Does your native language matter? Neural correlates of typological similarity in non-native production

This article is published as: Von Grebmer Zu Wolfsthurn, S., Pablos, L., & Schiller, N. O. (2022). Does your native language matter? Neural correlates of typological similarity in non-native production. Lingue e Linguaggio, 21(1), 143-169.

Abstract: Cross-linguistic influence (CLI) and typological similarity are key features in multilingual language processing. Here, we study whether CLI effects in language production are more pronounced in typologically similar vs. dissimilar languages in late language learners. In a picture-naming task, we manipulated gender congruency and cognate status as indices for CLI in a group of Italian learners of Spanish and a group of German learners of Spanish. Further, we explored modulations of P300 amplitudes indexing inhibitory control. Behaviourally, we observed effects of CLI, but not of typological similarity. At the neural level, P300 amplitudes were modulated by CLI effects. However, we did not find evidence for a typological similarity effect on P300 amplitudes. There-

fore, our results suggest a limited role of typological similarity. This study has crucial implications for non-native language production mechanisms in light of the similarity between the native and the non-native language.

Keywords: *typological similarity, non-native production, cross-linguistic influence, P300 ERP effect, late language learners*

6.1 Introduction

Anecdotally, multilingual language learners sometimes describe learning a particular language as "easy" because their native language (L1) is "similar" to the non-native language they are acquiring. Here, the term "multilinguals" describes individuals from diverse linguistic backgrounds who have acquired two or more languages at varying proficiency levels (Cenoz, 2013). In turn, proficiency refers to the extent to which language abilities match the age-based standard in comparison to native speaker (Bedore et al., 2012). Beyond anecdotal accounts, the notion of similarity between the L1 and the non-native language refers to *typological similarity*, i.e., the structural cross-linguistic similarities with respect to lexico-semantics and morphosyntax (Foote, 2009; Rothman & Cabrelli Amaro, 2010). For example, Italian and Spanish may be considered as more typologically similar than German and Spanish due to more overlap in terms of morphosyntax and lexicon (Schepens et al., 2012). The question of how much typological similarity affects language processing is crucial because it directly addresses the debate of the functional organisation of the L1 and a non-native language in the multilingual brain (Costa, Heij & Navarrete, 2006; Tolentino & Tokowicz, 2011; Zawiszewski & Laka, 2020).

At the root of so-called *typological similarity effects* are the shared cognitive representations and neurocognitive resources of the L1 and the non-native language (MacWhinney, 2005). Several neuroimaging and electroencephalographic (EEG) studies on highly proficient speakers suggest that more neurocognitive resources are shared across the L1 and the non-native language for typologically similar linguistic features compared to typologically less similar features (De Diego Balaguer, Sebastián-Gallés, Díaz & Rodríguez-Fornells, 2005; Jeong et al., 2007). In turn, these shared representations are linked to language co-activation and *cross-linguistic influence*, hereafter CLI (Lago et al., 2021; Nozari & Pinet, 2020). CLI refers to the bi-directional influence of the L1 and the non-native language on the underlying processing mechanisms. Moreover, CLI effects are found for different ages of acquisition (AoA) and may be larger and more pronounced at lower non-native proficiency levels (Heidlmayr et al., 2021; MacWhinney, 2005; Ringbom, 1987). Connected to typological similarity, research also suggests a link between high typological similarity and increased CLI (Cenoz, 2001; Costa, Heij & Navarrete, 2006; Yamasaki et al., 2018). Relevant for the purpose of our study, CLI was found to significantly impact non-native production (Lemhöfer et al., 2008; Paolieri et al., 2019). Multilinguals are said to recruit a language control network to manage language co-activation and CLI, in order to select the appropriate target language and obtain successful communication (D. W. Green, 1998; Stocco et al., 2014). A prominent theoretical framework for language control is the Inhibitory Control (IC) model (D. W. Green, 1998), which postulates the suppression of the non-target language prior to any linguistic output. Further, some evidence also suggests that language control forms part of domain-general cognitive control (Declerck et al., 2021) paramount to everyday functioning.

6.2 Background

A crucial question at this point is the following: do typologically similar languages bear a processing advantage over typologically less similar languages, or does typological similarity between languages result in a processing disadvantage instead? The relevant literature remains inconclusive with respect to this particular question, as will be discussed below. The *Conditional Routing Model* (CRM) by Stocco et al. (2014) proposes that increased CLI for typologically similar languages effectively trains and strengthens the control network throughout development (Yamasaki et al., 2018). This implies that speakers of typologically similar languages develop overall superior cognitive control skills to mitigate CLI effects and therefore have an advantage over speakers of typologically less similar languages (Declerck, Koch, Duñabeitia, Grainger & Stephan, 2019; Yamasaki et al., 2018; Zawiszewski & Laka, 2020). In this study, we used this theoretical framework to generate testable predictions about the role of typological similarity in non-native production. Participants were late learners and speakers of languages with differing degrees of typological similarity. We defined late learners as speakers who have acquired Spanish at a later age (AoA $>$ 14 years), and who have not yet reached high proficiency levels. We tested Italian-Spanish speakers for the typologically similar language pair, and German-Spanish speakers for the typologically less similar language pair. To directly probe the influence of typological similarity on non-native production, we examined its effects on two features shown to be prone to CLI from a behavioural and neural perspective: grammatical gender (hereafter gender) and cognates. Both of these CLI effects offer a unique window into the mechanisms of non-native production in light of differing degrees of typological similarity. According to the CRM, we should find that Italian-Spanish speakers show overall smaller CLI effects compared to the German-Spanish speakers as a function of increased training in mitigating CLI effects over time. This would imply that speakers of typologically similar languages indeed bear a processing advantage over speakers of typologically less similar languages.

The first CLI effect we explored in this study was the *gender congruency effect*. It was previously proposed to reflect CLI of the gender systems (Costa et al., 2003; Lemhöfer et al., 2008). Relevant to this study, Italian and Spanish both feature a feminine and masculine gender value, marked by *la* and *il* in Italian, and *la* and *el* in Spanish, respectively. In contrast, German has a three-way gender system characterised by masculine, feminine and neuter marked by *der*, *die* and *das*, respectively (Schiller & Caramazza, 2003). The core feature of the gender congruency effect is that *congruent* items, i.e., items belonging to the same nominal gender category across languages (e.g., *die*$_F$ *Kerze*$_F$– *la*$_F$ *vela*$_F$ [the candle] in German and Spanish), are processed more accurately and faster compared to *incongruent* items, i.e., items mismatching in gender categories across languages (e.g., *der*$_M$ *Schlüssel*$_M$ – *la*$_F$ *llave*$_F$ [the key] in German and Spanish)[1]. To this date, few studies have investigated the effect of typological similarity on gender processing in language production. Further, those who have studied it have in many cases looked at highly proficient speakers, and have not always obtained consistent results. For example, a study by Paolieri et al. (2019) explored the gender congruency effect in a translation task in highly proficient Italian-Spanish (typologically similar pair) and Russian-Spanish (typologically dissimilar pair) speakers. Results revealed a gender congruency effect in both groups, but the effect was more consistently found in the Italian-Spanish group than in the Russian-Spanish group. This finding suggested that typological similarity facilitated gender processing in Italian-Spanish speakers compared to Russian-Spanish speakers, in line with the CRM account (Stocco et al., 2014). By contrast, Costa et al. (2003) conducted a picture-naming study with highly proficient Spanish-Catalan, Catalan-Spanish, Italian-French and Croatian-Italian speakers to study the gender congruency effect. The authors found no evidence for a gender congruency effect, and therefore argued that typological similarity may not play a role in non-native production (see also Costa et al., 2006). This evidence therefore suggests a limited effect of typological similarity on non-native production, in contrast to the CRM account.

The second CLI effect examined in the context of typological similarity was the *cognate facilitation effect.* It reflects CLI of orthographic and phonological systems (Lemhöfer et al., 2008; Peeters et al., 2013). *Cognates*, i.e., items with a large semantic, phonological and orthographic form overlap across languages (e.g., *Tomate*

[1]Note that we are not assuming that gender values are conceptually identical across languages. However, as a result from explicit instructions in language courses, learners may perceive some gender values as sufficiently similar to each other.

– *tomate* [tomato] in German and Spanish), were found to be processed faster and more accurately compared to *non-cognates* (e.g., *Erdbeere* – *fresa* [strawberry]). Italian and Spanish share a larger amount of cognates compared to German-Spanish (Schepens et al., 2012). In turn, this suggests a strong structural overlap for the typologically similar Italian-Spanish pair. Therefore, the cognate facilitation effect per se provisionally suggests a processing advantage for orthographically similar and form-related structures compared to more less similar structures. To the authors' knowledge, no study has previously directly investigated the role of typological similarity in cognate production.

In sum, current evidence suggests first, a processing advantage for typologically similar structures (e.g., gender congruent items and cognates) compared to typologically dissimilar structures (e.g., gender incongruent items and non-cognates). Second, studies on the gender congruency effect suggest a tentative trend towards a production advantage for typologically similar languages compared to typologically less similar languages in highly proficient speakers, supporting the CRM (Stocco et al., 2014). However, in light of the conflicting results with respect to the gender congruency effect and typological similarity and the lack of direct evidence on the role of typological similarity on cognate production, we systematically explored typological similarity effects on both gender and cognate processing in this EEG study.

EEG and event-related potentials (ERPs) are critical tools in examining the temporal unfolding of the cognitive mechanisms underlying multilingual language processing (D. W. Green & Kroll, 2019). In this study, we probed the neural correlates of the typological similarity effect by focusing on the P300 component. This ERP component was previously linked to more general cognitive mechanisms such as inhibitory control, working memory and allocation of attentional resources in control network paradigms, such as language switching or the Flanker task (Declerck et al., 2021; González Alonso et al., 2020; Polich, 2007). In light of its functional involvement in control network processes, and because non-native production heavily relies both on the successful mitigation of CLI effects and on language control to inhibit the non-target language prior to articulation, the P300 component is the most relevant ERP component for our study. We predicted the P300 component to be an index of CLI effect mitigation where any differences in P300 amplitudes across groups could reflect a typological similarity effect. To the authors' knowledge, only the study by Von Grebmer Zu Wolfsthurn et al. (2021b) reported a P300 effect in an overt non-native picture-naming task. The present study is the first ERP study exploring whether typological similarity effects and CLI effects are detectable at the neural level in the form of distinct P300 correlates in non-native production.

The aim of this study was twofold: first, we systematically studied the gender congruency effect and the cognate facilitation effect in non-native language production in late learners with intermediate proficiency levels. Secondly, and more importantly, we investigated the effect of typological similarity on CLI in two language pairs with differing degrees of typological similarity: Italian-Spanish and German-Spanish. The research question investigated in this study was the following: does typological similarity have an impact on the potential effects of gender congruency and cognate status in non-native language production? In this, we employed an overt picture-naming task where participants named pictures in the non-native language Spanish using determiner + noun constructions, e.g., *la llave* [the key]. This was done to ensure the processing of grammatical gender since the correct determiner had to be produced alongside the noun (Schiller & Caramazza, 2003). During this task, we measured naming accuracy, naming latencies and P300 amplitudes.

6.2.1 Hypotheses

Hypotheses for behavioural data

We predicted higher accuracy and faster naming latencies for congruent and cognate items compared to incongruent and non-cognate items as an index of CLI. Next, and in line with the CRM

(Stocco et al., 2014), we predicted a non-native production advantage for the Italian-Spanish group compared to the German-Spanish group, reflected in overall higher accuracy and naming latencies. Finally, we predicted CLI effects to vary as a function of typological similarity, i.e., smaller CLI effects for the Italian-Spanish group compared to the German-Spanish group.

Hypotheses for ERP data

We first hypothesised P300 amplitudes to be modulated by gender congruency and cognate status. Specifically, we predicted less positive P300 amplitudes for congruent and cognate items compared to incongruent and non-cognate items. Second, as a reflection of a typological similarity effect and in line with the CRM (Stocco et al., 2014), we expected a production advantage for the typologically similar over the typologically less similar languages. Accordingly, we predicted different neural signatures of CLI between groups in the form of overall smaller P300 amplitudes for the Italian-Spanish group compared to the German-Spanish group.

6.3 Methods

6.3.1 Participants

The Italian-Spanish group included 33 participants (24 females) with $M = 27.12$ years of age ($SD = 4.08$), recruited from the Universitat Pompeu Fabra (Barcelona, Spain). The German-Spanish group consisted of 33 participants (27 females) with $M = 23.06$ years of age ($SD = 2.47$) recruited from the University of Konstanz (Germany). To avoid confounding effects by proficiency, we restricted our participant selection to late learners with intermediate proficiency levels in the B1/B2 range according to the Common European Framework of Reference for Languages, *CEFR* (Council of Europe, 2001). Note that the majority of participants were recruited from Spanish language courses specifically aimed at B1/B2 proficiency levels. Further eligibility criteria for this study were the following: right-handedness, no additional language learnt before the age of five, AoA of Spanish from fourteen years onwards, no language, reading, vision or hearing impairments and no psychological or neurological issues at the time of testing. Finally, the age limit was between 18 and 35 years. Given that the Italian-Spanish speakers were tested in Spain and the German-Spanish speakers were tested in their native language environment, we balanced any potential differences in immersion by imposing additional inclusion criteria for the Italian-Spanish group. These extra criteria were to only accept those individuals who had started learning Spanish shortly before or upon their arrival to Barcelona and those who had been living in a Spanish-speaking country for less than one year that conformed to B1/B2 proficiency levels.

Linguistic profile

Participants' linguistic profile including self-reported proficiency and experience with Spanish from the LEAP-Q is summarised in Table 6.3.1. The Italian-Spanish group spent on average 0.46 years (SD = 0.343) in a Spanish-speaking country. This compares to an average of 0.96 years (SD = 0.690) for the German-Spanish group. In the Italian-Spanish group, thirteen participants reported Spanish as their strongest language after Italian, fourteen participants as their second, five participants as their third, and one participant as their fourth strongest. Of the German-Spanish group, four participants stated that Spanish was the strongest language after German, twenty-six participants as their second, and three as their third strongest. The proficiency measures were rated on a ten-point scale, ten corresponding to being maximally proficient. The linguistic profiles for both groups were therefore highly comparable. See Appendix 6.A for an overview of the languages acquired by the Italian-Spanish speakers, and Appendix 6.B for the German-Spanish speakers.

Table 6.3.1: *Linguistic profile of participants' Spanish proficiency and experience from the LEAP-Q for the Italian-Spanish group (N = 33) and the German-Spanish group (N = 33). The self-reported proficiency measures are highlighted in bold.*

Native language	**Italian**	**German**
Mean AoA (years)	23.93 (SD = 5.07)	16.29 (SD = 2.39)
Mean fluency age (years)	24.88 (SD = 4.48)	18.53 (SD = 2.29)
Mean reading onset age (years)	24.36 (SD = 4.91)	17.27 (SD = 3.03)
Mean reading fluency age (years)	24.24 (SD = 4.82)	18.42 (SD = 2.62)
Exposure (%)	40 (SD = 18.37)	10 (SD = 9.48)
Speaking proficiency	6.09 (SD = 1.76)	6.76 (SD = 1.00)
Comprehension proficiency	7.26 (SD = 1.67)	7.34 (SD = 0.92)
Reading proficiency	7.36 (SD = 1.48)	7.18 (SD = 1.07)

6.3.2 Materials and design

Prior to the experimental session, participants completed the LEAP-Q background questionnaire which provides information regarding language proficiency and language experience from multilinguals (Marian et al., 2007). During the experimental session, participants were first presented with the LexTALE-Esp[2] (Izura et al., 2014), a lexical decision task to measure vocabulary size in Spanish, followed by the picture-naming task. Both these tasks were programmed in E-prime2 (Schneider et al., 2002).

Stimuli

Picture-naming task. We followed an identical stimulus selection procedure for both groups. We selected the picture stimuli from the MultiPic database (Duñabeitia et al., 2018). We chose

[2]For our study, we transformed the LexTALE-Esp version by Izura et al. (2014) into an E-prime equivalent using the same instructions and stimuli words.

those stimuli with the largest proportion of valid and correct responses during the validation phase from the database. We selected 96 stimuli pictures for each group. There were 24 stimuli pictures for each one of the four conditions: congruent cognates, congruent non-cognates, incongruent cognates, and incongruent non-cognates, see Table 6.3.2 and Table 6.3.3 for examples for each group. Identical cognates (e.g., *das Taxi - el taxi* for German and *il taxi - el taxi* for Italian [the taxi]) were not included in our sample; neither were items with biological gender (e.g., *il judice - el juez* [the judge], *der Sänger - el cantante* [the singer]), English loanwords (e.g., *el boomerang* [the boomerang]), or gender-ambiguous nouns (*la mar/el mar* [the sea]). The classification of cognates as such was based on semantic, orthographic and phonological overlap. To increase the validity of our study, we modelled the distribution of terminal phonemes of the Spanish stimuli after work by Clegg (2011), for example, approximately 30% of the stimuli nouns ended in [*a*] and 10% of nouns in [*e*]. Nevertheless, terminal phoneme was included as a covariate in the statistical analyses (see section 6.4.2).

Table 6.3.2: *Example noun phrase stimuli for each condition for the Italian-Spanish group.*

Condition	**Noun phrase**	**Italian translation**	**English translation**
congruent/ cognate	la_F llave_F	la_F chiave_F	*the key*
congruent/ non-cognate	la_F fresa_F	la_F fragola_F	*the strawberry*
incongruent/ cognate	el_M bolso_M	la_F borsa_F	*the handbag*
incongruent/ non-cognate	el_M caracol_M	la_F lumaca_F	*the slug*

Table 6.3.3: *Example noun phrase stimuli for each condition for the German-Spanish group.*

Condition	**Noun phrase**	**German translation**	**English translation**
congruent/ cognate	la$_F$ jirafa$_F$	die$_F$ Giraffe$_F$	*the giraffe*
congruent/ non-cognate	la$_F$ pera$_F$	die$_F$ Birne$_F$	*the pear*
incongruent/ cognate	el$_M$ melón$_M$	die$_F$ Melone$_F$	*the melon*
incongruent/ non-cognate	el$_M$ tenedor$_M$	die$_F$ Gabel$_F$	*the fork*

EEG recordings

Italian-Spanish group. EEG data were collected via 32 active electrodes using the BrainVision Recorder software (Version 1.10) by BrainProducts using a standard 10/20 montage with a 500 Hz sampling rate. We recorded the vertical electrooculogram (VEOG) from an additional facial electrode placed underneath the participant's left eye (FT9), and the horizontal electrooculogram (HEOG) from one electrode at the outer canthus of the left eye (FT10). The original reference electrode was FCz. The ground electrode was placed on the right cheek of the participant. Electrodes were configured via the BrainVision Recorder software to ensure optimal conductivity. Impedances were kept below 10 kΩ.

German-Spanish group. EEG data were collected from 32 passive electrode locations via the BrainVision Recorder software (Version 1.23.0001) at a sampling rate of 500 Hz. We used the standard 10/20 montage with an EasyCap electrode cap. The HEOG was measured from two electrodes at the outer canthus of the left and right eye. The VEOG was recorded using an electrode underneath the left eye. The ground electrode was placed on the right cheek of the participant. Electrodes were initially referenced to the Cz electrode. Impedances of the electrodes were checked and configured using actiCAP ControlSoftware (Version 1.2.5.3). We kept impedances below 5 kΩ. for the reference and ground electrode. For the remaining channels, impedances were below 10 kΩ.

6.3.3 Procedure

Complying with the ethics code for linguistic research in the Faculty of Humanities at Leiden University, participants signed an informed consent form prior and after participation. Further, they were given an information sheet prior to the experiment and a written and oral debrief upon termination of the experiment in their respective L1. Participants received a monetary reimbursement for their participation. The procedure for both the Italian-Spanish and the German-Spanish group was identical, with the difference that for the former group oral instructions were provided in Italian, and for the latter in German by a native speaker.

LexTALE-Esp

For the LexTALE-Esp, instructions were provided in black font on a white screen. Next, a fixation cross was displayed for 1,000 ms. Then, a letter string corresponding to either a Spanish word or a pseudoword was displayed in the centre of the screen. Participants were asked to make a lexical decision via a button press about whether or not the string was a Spanish word. The letter string remained on the screen until the participant's response. The original stimuli from Izura et al. (2014) consisted of 60 words and 30 pseudowords. For both the Italian-Spanish and the German-Spanish group, three stimuli were eliminated from the stimuli list before the experiment due to overlap with the picture stimuli. Therefore, the total number of trials was 87 for both groups. Each letter string was only shown once, and trial order was randomised for each participant. Offline, we computed LexTALE-Esp vocabulary size scores by subtracting the percentage of incorrectly identified pseudowords from the correctly identified words (Izura et al., 2014). LexTALE-Esp scores were subsequently included as a covariate in our statistical analyses (see section 6.4.2).

Picture-naming task

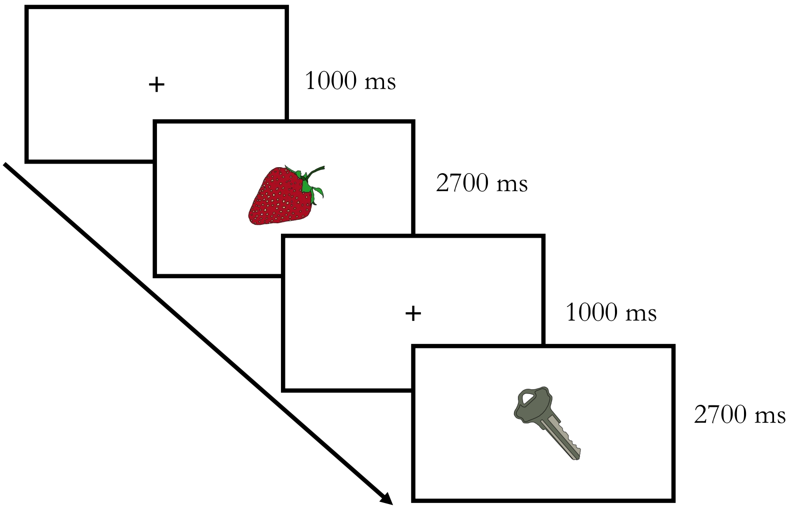
For the picture-naming task, we manipulated gender congruency and cognate status in a 2 x 2 fully factorial within-subjects design. Half of the stimuli were congruent items, whereas the other half were incongruent items. Half of the congruent and incongruent items were cognates, respectively, whereas the other half were noncognates. We divided the task into a familiarisation phase and an experimental phase, during which we recorded the EEG signal. The familiarisation phase consisted of three rounds. In each round, participants were shown each picture and had to overtly produce the corresponding noun together with the correct definite determiner (e.g., *la llave* [the key]). If either the noun or the determiner, or both, were incorrect, the experimenter provided oral feedback to the participant to ensure the participants' familiarity with each picture. The total number of trials in the familiarisation phase was 288. In the experimental phase, a trial was initiated by a black fixation cross on a white screen, which was displayed for 1,000 ms. Next, a picture appeared on the screen for 2,700 ms (Figure 6.3.1). Participants were instructed to overtly name each picture on the screen as fast and accurately as possible. They were explicitly encouraged to minimise all movements during the experiment. Each picture was shown once in a unique trial order during the experimental phase, resulting in a total of 96 trials. There were two self-paced breaks to restore participants' engagement with the task.

6.4 Results

6.4.1 Behavioural data exclusion

Data for one participant of the German-Spanish group were lost due to a recording failure. For naming latencies, we included only correct trials in the analysis. Moreover, the behavioural analyses consisted of the identical data sets as the EEG analysis (see section 4.4 for details on EEG data exclusion). Therefore, we included a total of 28 participants from the Italian-Spanish group and 30

Figure 6.3.1: *Trial sequence for the picture-naming task.*



participants from the German-Spanish group in the behavioural analyses (n = 58).

6.4.2 Behavioural data analysis

We calculated naming accuracy and naming latencies in Praat (Broersma & Weenink, 2019). We pooled the behavioural data for both groups to probe modulatory effects of CLI and typological similarity on *naming accuracy* and *naming latencies* across the Italian-Spanish speakers and the German-Spanish speakers. We employed a single-trial linear mixed effects modelling (LMM) approach using the *lme4* package (Bates et al., 2020) in RStudio (Team, 2020). To model *naming accuracy*, we used a generalised linear mixed effect model (GLMM) with a binomial distribution using the *glmer()* function. For our positively skewed *naming latencies*, we used a GLMM with a gamma distribution and the identify link function with the *glmer()* function (Lo & Andrews, 2015). See Appendix 6.C for details about the included interaction effects and the model fitting procedure, which is also described in Von Grebmer Zu Wolfsthurn et al. (2021b). Our fixed effects structure consisted of the following: *typological similarity* (typologically similar vs. typologically dissimilar), *gender congruency* (congruent vs. incongruent) and *cognate status* (cognate vs. non-cognate). Moreover, we controlled for the covariates *LexTALE-Esp score, familiarisation phase performance, target noun gender, word length, order of acquisition of Spanish* and *terminal phoneme* of the target word in both analyses[3]. Finally, our random effects structure included *subject* and *item* (i.e., the individual picture) as well as random slopes for our main manipulations.

6.4.3 Behavioural data results

We first computed descriptive statistics for the remaining participants of each group. See Table 6.4.1 for descriptives for *naming accuracy*, reported as percentages, and for *naming latencies* reported in ms. Mean LexTALE-Esp scores were $M = 25.92$ ($SD = 13.69$, range between -7.37 and 49.30) for the Italian-Spanish group, and $M = 18.45$ ($SD = 20.52$, range between -23.16 and 60.18) for the German-Spanish group. Scores above 60 were previously linked to C1/C2 levels, thereby confirming the B1/B2 range of our participants (Lemhöfer & Broersma, 2012). A two-sample t-test yielded no statistical difference in LexTALE-Esp scores between the two groups with $t(50.83) = 1.64$, *95% CI*[1.68, 16.60], $p = 0.101$. Nevertheless, we included LexTALE-Esp scores as a covariate in our analysis. Further, we plotted naming accuracy for both groups in Figure 6.4.1 and naming latencies in Figure 6.4.2.

Naming accuracy. The model of best fit included both *gender congruency* and *cognate status* as main effects. Participants were more accurate for congruent items compared to incongruent items ($\beta = 0.728$, *95% CI*[0.542, 0.978], $z = -2.11$, $p = 0.035$). Further, participants were more accurate for cognates compared to non-cognates ($\beta = 0.696$, *95% CI*[0.519, 0.934], $z = -2.41$, $p =$

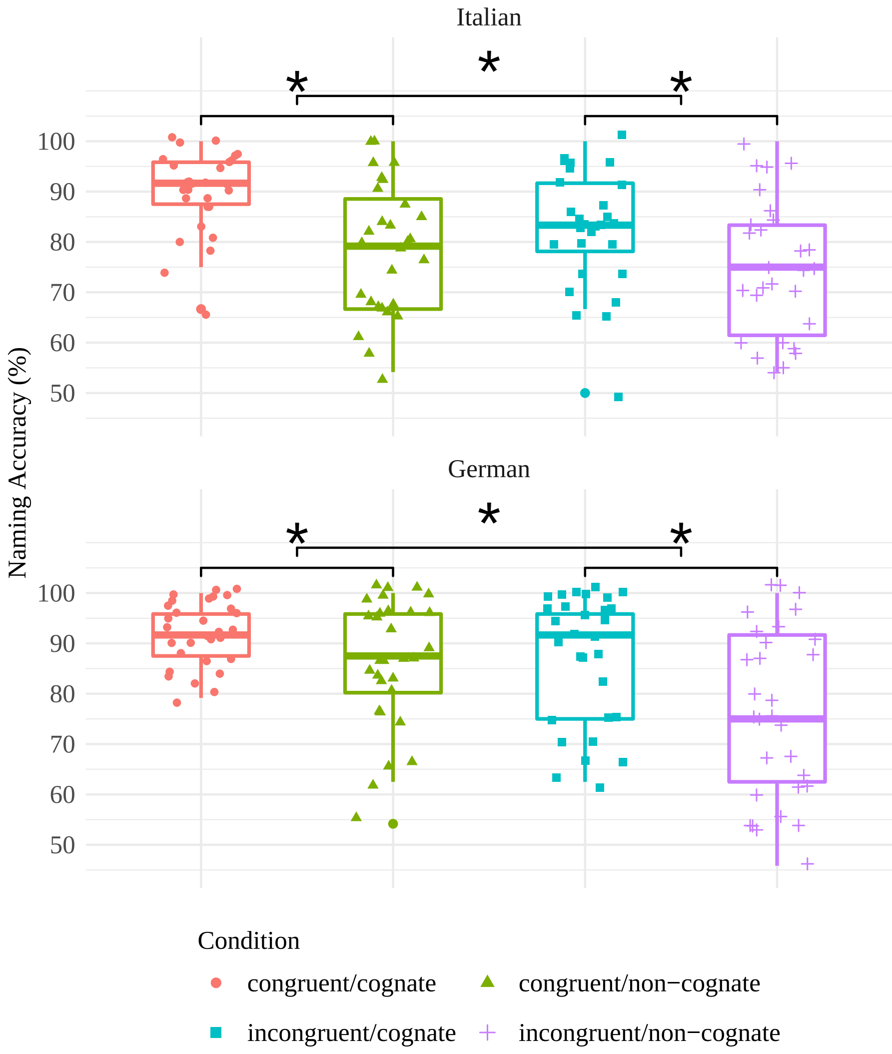
[3] Note that in Von Grebmer Zu Wolfsthurn et al. (2021b) acquisition of French was also considered as a potential covariate; however, no significant effect on our outcome variables was found. We therefore did not include acquisition of French in the current analyses.

Table 6.4.1: *Mean naming accuracy (%) and mean naming latencies (ms) for each condition for the Italian-Spanish group (n = 28) and the German-Spanish group (n = 30).*

Native language	**Condition**	**Mean (%)**	**SD**	**Mean (ms)**	**SD**
Italian	congruent/ cognate	89.88	30.18	911.12	253.91
	congruent/ non-cognate	78.57	41.06	1011.90	297.67
	incongruent/ cognate	82.59	37.95	1027.61	305.52
	incongruent/ non-cognate	75.00	43.33	1068.59	281.19
German	congruent/ cognate	92.08	27.02	891.46	236.85
	congruent/ non-cognate	86.39	34.31	932.52	293.93
	incongruent/ cognate	87.22	33.41	971.27	312.85
	incongruent/ non-cognate	75.83	42.84	977.70	303.22

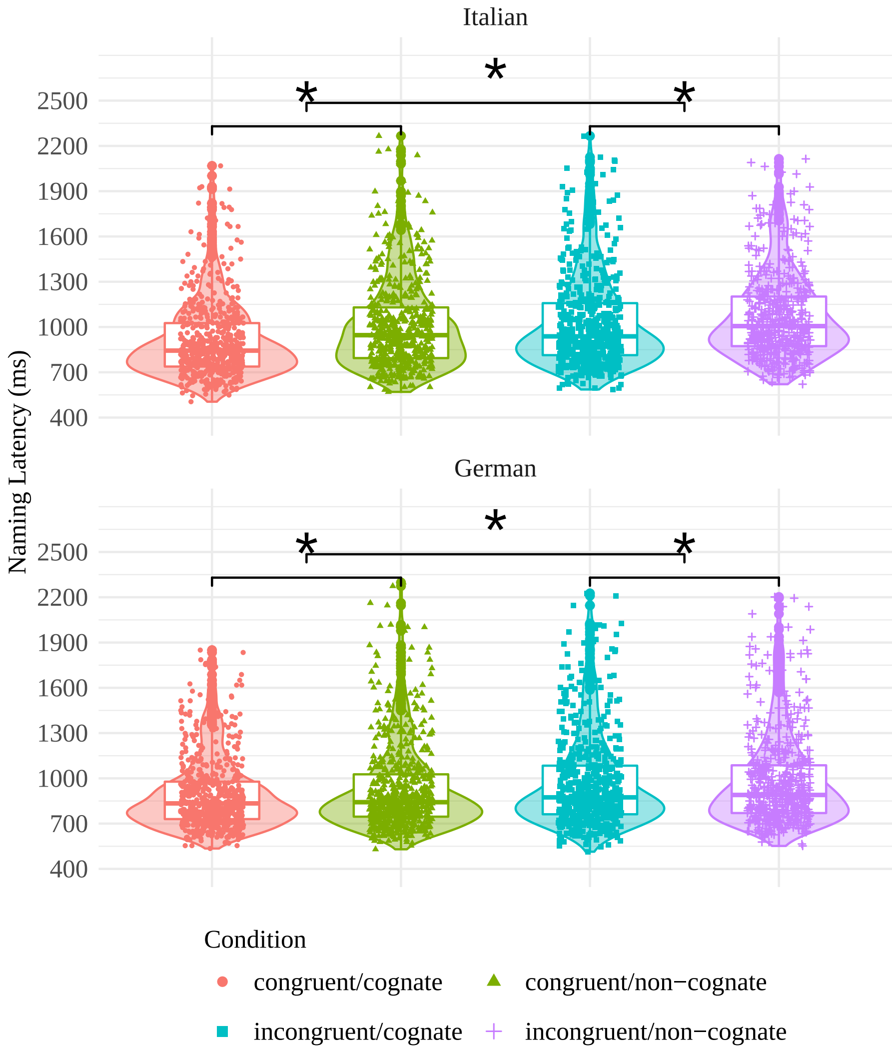
0.016). The model also included *familiarisation phase performance* as a covariate, and *subject* and *item* as random effects. *Typological similarity* did not significantly improve the model fit with χ^2(1, n = 58) = 0.235, p = 0.125 and was therefore not included in the final model. The best-fitting model was the following: naming accuracy ~ gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + familiarisation phase performance (zero correct vs. one correct vs. two correct vs. three correct) + (1|subject) + (1|item). See Appendix 6.D and Figure 6.4.1.

Figure 6.4.1: *By-condition naming accuracy (%) for each group. The significance brackets reflect the statistical difference in accuracy between congruent and incongruent items and between cognate and non-cognate items, with higher accuracy rates for congruent and cognate items, as well as for the German-Spanish group (n = 30) compared to the Italian-Spanish group (n = 28).*



Naming latencies. For naming latencies, the model of best fit included main effects for *gender congruency* and *cognate status*. Participants were faster at naming congruent compared to incongruent items, and cognate compared to non-cognate items with $\beta = 0.064$, *95% CI*[0.019, 0.108], $t = 2.82$, $p = 0.005$ and $\beta = 0.046$, *95% CI*[0.005, 0.087], $t = 2.18$, $p = 0.029$, respectively. We included *familiarisation phase performance* as a covariate, however, we found no evidence that *typological similarity* contributed to a better model fit. Similarly, the other covariates did not improve the model fit or led to non-convergence and were therefore excluded from the model. Furthermore, we included a by-*subject* random slope for the correlated effects of *gender congruency* and *cognate status* as well as *item* as random effect. The final model was therefore the following: naming latency $\sim$ gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + familiarisation phase performance (zero correct vs. one correct vs. two correct vs. three correct) + (gender congruency + cognate status|subject) + (1|item), see Appendix 6.E and Figure 6.4.2 for details. Taken together, we found no behavioural evidence that *typological similarity* modulated *naming accuracy* or *naming latencies* across the two groups.

Figure 6.4.2: *By-condition naming latencies (ms) for each the Italian-Spanish group (n = 28) and the German-Spanish group (n = 30). The significance brackets reflect the statistical difference in accuracy between congruent and incongruent items and between cognate and non-cognate items, with faster naming latencies for congruent and cognate items.*



6.4.4 EEG data exclusion

Upon completion of pre-processing of our EEG data, we determined a set of identical inclusion criteria for both groups: first, we only modelled the EEG signal for correct trials, i.e., trials where participants provided the correct noun phrase in the experimental phase (Christoffels et al., 2007). Secondly, we only included trials which were not contaminated by artefacts. Finally, we did not include participants with heavily contaminated datasets, i.e., $> 40\%$ of trials lost due to artefacts. The resulting threshold for inclusion was a remainder of at least 60% of trials at the end of pre-processing and the application of the inclusion criteria. Following these criteria, we excluded five EEG datasets from the Italian-Spanish group and two EEG datasets from the German-Spanish group. The corresponding behavioural data were also excluded from the behavioural analysis. This amounted to 28 Italian-Spanish datasets and 30 German-Spanish datasets for further statistical analyses ($n = 58$).

6.4.5 EEG data pre-processing

Prior to any statistical analyses, we performed a vigorous pre-processing procedure to separate the task-related EEG signal from articulatory artefacts frequently found in production paradigms (Ganushchak, Christoffels & Schiller, 2011). We followed a largely identical procedure for both groups. Note that as stated above, the recording parameters differed between the two groups. The EEG data were analysed in BrainVision Analyser V2.2. First, we re-referenced all of our data channels to the average of the mastoid channels, TP9 and TP10. For the Italian-Spanish group, the original reference channel FCz was reused as a data channel, adding to a new total of 29 data channels. FT9 and FT10 were used as VEOG and HEOG, respectively. For the German-Spanish group, we reused the reference Cz channel as a data channel, amounting to a total of 31 data channels. For this group, we performed linear derivation on the two HEOG channels to form a single HEOG channel. For both groups, we subsequently applied a high-pass filter of 0.1 Hz, and a low-pass filter of 30 Hz. We interpolated channels where appropriate. We then performed residual drift detection on our HEOG and VEOG channels in order to improve the precision of the subsequent blink correction using ocular ICA. Next, we performed artefact rejection on all data channels. After increasing the signal-to-noise ratio by means of pre-processing, we segmented our data at 200 ms prior and 1,200 ms after picture onset. Segments which included artefacts were excluded in this process. After segmentation, we applied a baseline correction using the 200 ms interval prior to picture onset. Finally, we exported all voltage amplitudes for all segments, channels and participants in order to perform single-trial LMM in RStudio (R Core Team, 2020). This method was recently introduced as powerful alternative to the more traditional grand-averaging of EEG data because it captures both by-subject and by-item individual variance and correlations between individual data points while preserving statistical power (Frömer et al., 2018).

6.4.6 EEG data analysis

Next, we tentatively explored our EEG data via a permutation analysis to visualise potential effects of condition on voltage amplitudes. First, we divided our 29 data electrodes 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 the 31 electrodes for our 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) into nine topographic areas: anterior, central and posterior, in turn divided into left, midline and right sections. Next, we used the *permu.test()* function from the *permutes* package (Voeten, 2019) for a permutation analysis on each group. See Figure 6.4.3 for the outcome of the permutation test for the Italian-Spanish group, and Figure 6.4.4 for the German-Spanish group. In both figures, darker colours correspond to higher F-values indicating potential differences in voltage between conditions, whereas lighter colours correspond to lower F-values following an F-distribution under the null hypothesis that there are no differences by condition (Maris & Oostenveld, 2007; Voeten, 2019). For the Italian-Spanish group, the output from the permutation analysis showed that a potential region of interest was clustered in centro-parietal regions around the following eight electrodes: *Pz, P3, P4, P7, P8, Oz, O1* and *O2*. Further, the permutation analysis suggested a time window of interest between 350 ms and 600 ms post-stimulus onset. On the other hand, the permutation test for the German-Spanish group showed potentially significant effects in centro-parietal regions at the following thirteen electrodes: *CPz, CP3, CP4, TP7, TP8, Pz, P3, P4, P7, P8, Oz, O1* and *O2* in a similar time window, around 350 ms to 600 ms post-stimulus onset. Due to increasing articulatory artefacts in proximity to the articulatory onset, we deliberately set the upper time window threshold to 600 ms post-stimulus onset to avoid signal contamination by motor articulation.

Figure 6.4.3: *Permutation test outcome for the Italian-Spanish group (n = 28). Larger F-values are shown in darker colours and denote an increased likelihood for a statistically relevant effect of our manipulations on voltage amplitudes.*

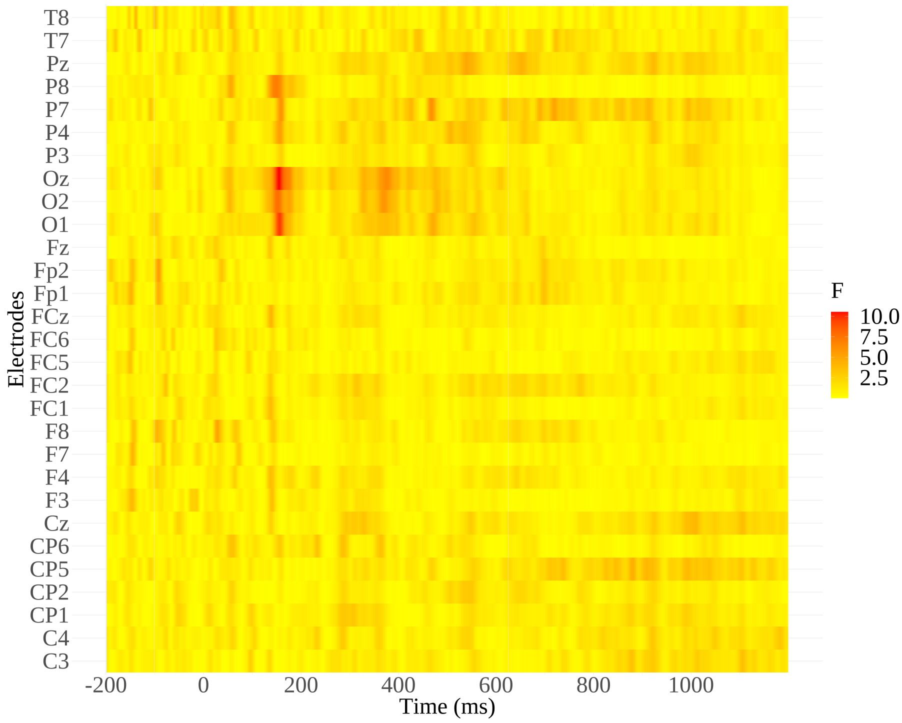
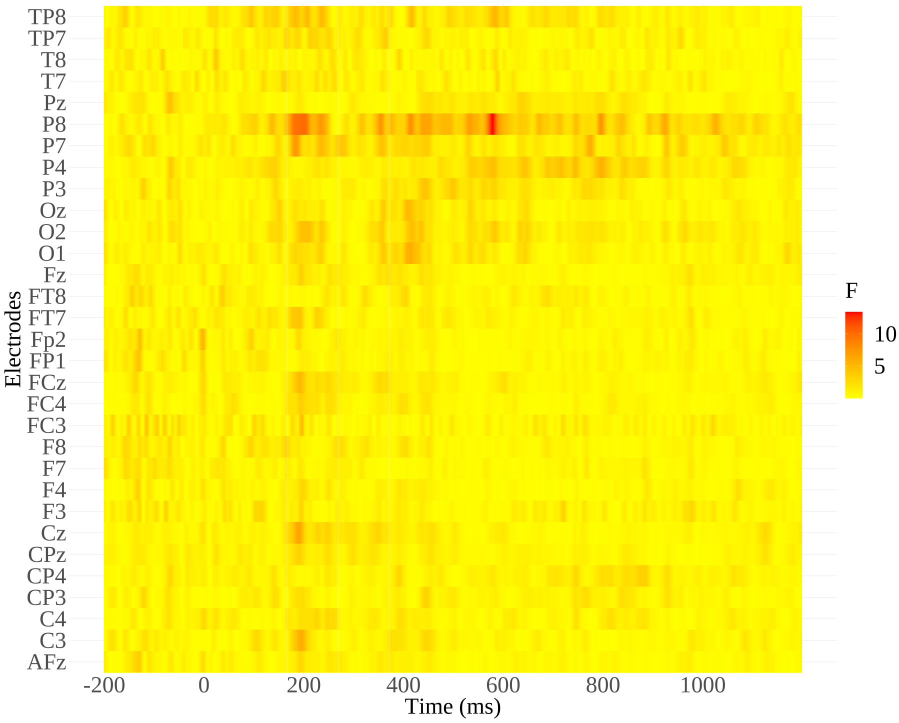


Figure 6.4.4: *Permutation test outcome for the German-Spanish group (n = 30). Larger F-values are shown in darker colours and denote an increased likelihood for a statistically relevant effect of our manipulations on voltage amplitudes.*



As a next step, we pooled our EEG data and proceeded to the statistical analysis. On the basis of the previous literature and the outcome of the permutation analysis, we selected centro-parietal regions as our region of interest, and 350 ms to 600 ms as our time window of interest. We only selected the eight electrodes which were present in the montage of both groups: *Pz, P3, P4, P7, P8, Oz, O1* and *O2*. See Appendix 6.F for a visualisation of the region of interest for the group analysis. Mirroring the behavioural data analysis, we modelled *voltage amplitudes* as a function of *gender congruency*, *cognate status* and *typological similarity* in our fixed effects structure using the *lme4* package (Bates et al., 2020). Moreover, we carefully controlled for potential confounding factors such as *hemisphere, LexTALE-Esp score, familiarisation phase per-*

formance, target noun gender, word length, order of acquisition and *terminal phoneme*. *Subject* and *item* were defined as random effects for the random effects structure and random slopes for our main manipulations were included. Lastly, the final model was re-fitted using the restricted maximum likelihood criteria (REML) for unbiased estimates (Mardia et al., 1999).

6.4.7 EEG data results

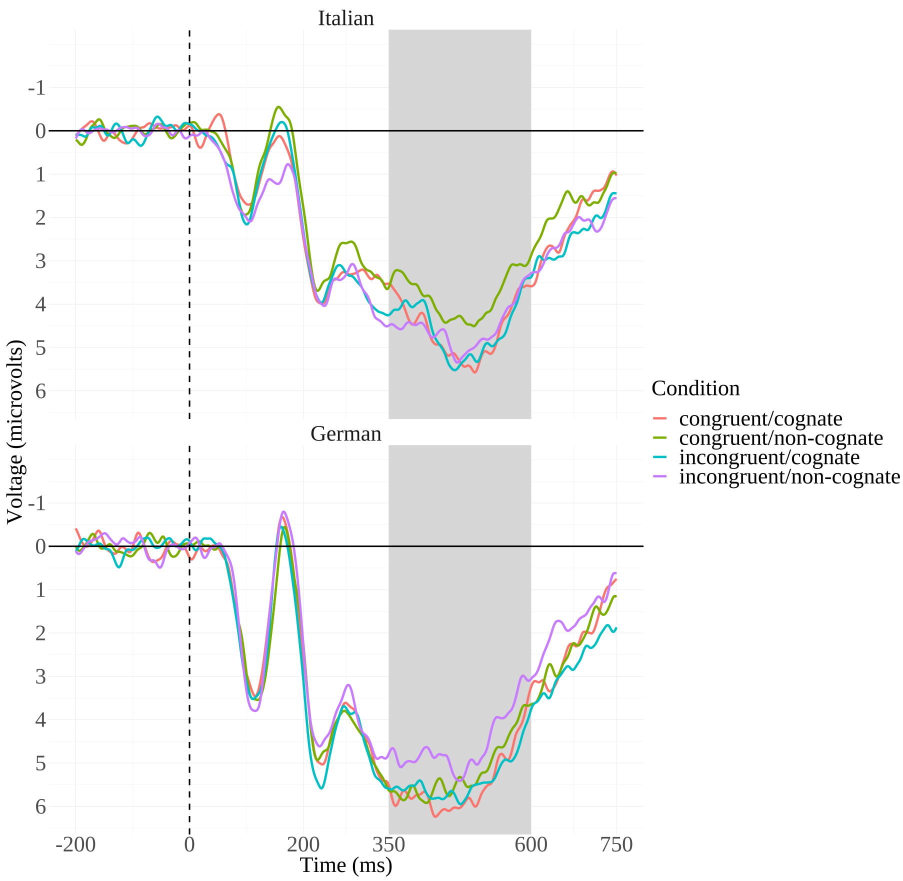
Mean voltage amplitudes by condition for each group are summarised in Table 6.4.2. Descriptively speaking, the Italian-Spanish group yielded overall lower voltage amplitudes compared to the German-Spanish group in our time window and region of interest. Visual inspection of voltage amplitudes in Figure 6.4.5 showed a positive oscillation followed immediately by a negative oscillation shortly after stimulus onset, which reflected the classical P1/N2 complex linked to early visual processing (P. Chen et al., 2017). In line with the outcomes of the permutation test, both groups then showed a positive-going waveform across centro-parietal regions between 350 ms and 600 ms after stimulus onset (indicated by grey shading in Figure 6.4.5). This particular waveform in our time window of interest, together with the topographic distribution of the channels, indicated a P300 component in both groups. Voltage amplitudes peaked around 500 ms post-stimulus onset. Finally, they visibly dropped back to baseline and became increasingly noisier closer to the articulatory onset around 700 ms.

Table 6.4.2: *Mean voltage amplitudes (μV) by condition for centro-parietal regions for the time window of interest (350 ms to 600 ms).*

Native language	**Condition**	**Mean (μV)**	**SD**
Italian	congruent/cognate	4.58	8.85
	congruent/non-cognate	3.80	9.07
	incongruent/cognate	4.57	8.23
	incongruent/non-cognate	4.54	8.54
German	congruent/cognate	5.48	9.59
	congruent/non-cognate	5.19	9.13
	incongruent/cognate	5.39	9.53
	incongruent/non-cognate	4.55	9.36

For voltage amplitudes, the model of best fit included *gender congruency* and *cognate status* as main effects, as well as *hemisphere* and *familiarisation phase performance* as covariates. *Subject* and *item* were included as random effects, as well as correlated by-*subject* random slopes for *gender congruency* and *cognate status.* Therefore, the final model was the following: voltage amplitudes $\sim$ gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + hemisphere (left vs. midline vs. right) + familiarisation phase performance (zero correct vs. one correct vs. two correct vs. three correct) + (gender congruency + cognate status|subject) + (1|item). *Voltage amplitudes* were significantly higher for cognates compared to non-cognates with β = -0.477, *95% CI*[-0.934, -0.021], t = -2.05, p = 0.040. In contrast, there was no significant effect of *gender congruency* on voltage amplitudes with β = 0.025, *95% CI*[-0.479, 0.529], t = 0.097, p = 0.922 for congruent compared to incongruent items. Despite a descriptive trend, *typological similarity* did not significantly improve the model fit, with χ^2(1, n = 58) = 1.31, p = 0.252 when comparing the model with and without typological similarity in the fixed effects structure, see Appendix 6.G and Figure 6.4.5 for further details. Therefore, mirroring the behavioural results, *typological similarity* did not appear to influence electrophysiological measures of non-native production.

Figure 6.4.5: *By-condition voltage amplitudes (μV) for centro-parietal regions for each group. The time window of interest (350 ms to 600 ms) is highlighted in light grey. Note that positive voltage amplitudes are plotted downwards, and negative voltage amplitudes are plotted upwards.*



6.5 Discussion

In this study, we investigated the effect of typological similarity on CLI in non-native production in Italian learners of Spanish (typologically similar group) and German learners of Spanish (typologically dissimilar group) with intermediate proficiency levels (B1/B2). More specifically, we examined a processing advantage for the typologically similar group within the framework of the CRM (Stocco et al., 2014) and explored how *gender congruency*, *cognate status* and *typological similarity* modulated overt picture-naming in intermediate learners of Spanish, e.g., when producing *la llave* [the key]. We modelled *naming accuracy* and *naming latencies*, as well as P300 *voltage amplitudes*. Outcomes of this study have crucial implications for non-native language processing, as well as the functional organisation of languages with respect to typological similarity.

Behaviourally, we expected more accurate and faster processing of congruent and cognate items compared to incongruent and non-cognate items. In addition, we predicted the typologically similar Italian-Spanish group to show an overall production advantage compared to the typologically dissimilar German-Spanish group in terms of naming accuracy and naming latencies. Finally, we predicted CLI effects to vary between our two groups. Our results partially supported our hypotheses. Participants were faster and more accurate at naming congruent compared to incongruent items and at naming cognates compared to non-cognates, displaying both the classical gender congruency effect (Paolieri et al., 2019) and the cognate facilitation effect (Lemhöfer et al., 2008). However, we found no main effect of typological similarity nor an interaction effect, indicating that the mitigation of CLI effects was equally successful in both groups. This is in line with previous studies (Costa, Heij & Navarrete, 2006; Costa et al., 2003).

From a theoretical point of view, our behavioural results showing more accurate and faster processing of congruent and cognate items indicated first that we found evidence for a sensitivity of late learners to both gender congruency and cognate status. Secondly, these results suggest that both groups experience comparable levels of CLI, as reflected by similar behavioural performances. One possible interpretation is that at intermediate proficiency levels, the facilitatory effects (when processing congruent items and cognates) and the hampering effects (when processing incongruent items and non-cognates) are balanced across the two groups. At this stage of learning, the overall similarity between the respective native languages and Spanish showed little influence on multilingual language production. Therefore, our behavioural results do not support the predictions of the CRM by Stocco et al. (2014). Based on these results, we postulate that gender congruency and cognate status are the main modulating factors of non-native production in this study and that typological similarity plays a limited role at intermediate proficiency levels. However, to obtain a more comprehensive interpretation of these data, we integrated the behavioural findings with the EEG findings.

At the neural level, we originally predicted a modulation of the P300 component as a function of *gender congruency* and *cognate status*. Moreover, we expected smaller P300 amplitudes for the Italian-Spanish group compared to the German-Spanish group to reflect an effect of *typological similarity*. The first critical finding was that the EEG signal in our time window and region of interest was consistent with a P300 component in both the Italian-Spanish and the German-Spanish group, although this P300 showed a delayed voltage peak compared to previous research (Polich, 2007). These results mirror Von Grebmer Zu Wolfsthurn et al. (2021b). As previously discussed, the P300 has been typically linked to conflict monitoring, cognitive control and interference, and more recently to attentional resources and working memory (González Alonso et al., 2020; Polich, 2007). On the basis of our findings, we argue that the P300 may be a critical index for cognitive control in the non-native production process because to succeed at this task, speakers need to mitigate language co-activation and CLI effects. This implies a strong element of cognitive control in non-native production for speakers at intermediate proficiency levels.

The second critical finding in line with our hypotheses was the modulation of the P300 voltage amplitudes by CLI of cognates: they elicited larger P300 voltage amplitudes compared to non-cognates. The notion of larger voltage amplitudes for cognates was found in previous work (Christoffels et al., 2007; Strijkers et al., 2010), but studies also reported the opposite pattern (Peeters et al., 2013).

Here, we argue that the difference in voltage amplitudes for cognates compared to non-cognates is reflective of the mitigatory processes to manage CLI, and that the P300 modulation may be a critical marker of the processes underlying CLI. The third important finding was that we found no evidence that gender congruency significantly modulated P300 voltage amplitudes. Therefore, our data suggest that cognate status is the more salient modulating feature during non-native production at the neural level.

Finally, we found no evidence for distinct neural signatures across the Italian-Spanish and the German-Spanish group: we did not find an interaction effect between typological similarity and the two CLI effects we tested that could indicate different CLI effects across groups. Statistically speaking, voltage amplitudes were comparable for both groups. Therefore, our ERP results suggest a limited effect of typological similarity on CLI effects and non-native production as a whole (Costa, Heij & Navarrete, 2006; Costa et al., 2003). This finding speaks directly to the theoretical framework regarding the directionality of the typological similarity effect in that it does not support an advantage of typologically similar languages (Stocco et al., 2014). These ERP findings are contrary to what we hypothesised, but are in line with the behavioural results we obtained. The question here is whether there could be another modulating factor at play that is potentially more powerful than typological similarity effects.

As discussed in the introduction, Costa et al. (2003) did not find evidence for an effect of typological similarity in groups of highly proficient Spanish-Catalan, Catalan-Spanish, Italian-French and Croatian-Italian speakers. This is in line with our findings. Similar results were found in a later study by Costa, Heij and Navarrete (2006) with highly proficient speakers, which also did not show evidence for typological similarity effect in a bilingual picture-naming task in Spanish-Catalan (typologically similar) and Spanish-Basque (typologically dissimilar) speakers. One possible interpretation of our findings is embedded within the context of proficiency, which is a key feature in mitigating CLI and language control (D. W. Green, 1998). More specifically, the IC model (D. W. Green, 1998) proposes similar language control mechanisms for the L1 and the non-native language when proficiency levels are highly similar. The speakers in our study were intermediate speakers of Spanish, implying that typological similarity effects on non-native production were limited when the difference in proficiency between the L1 and the non-native language was modest. Nevertheless, this does not exclude the possibility that typological similarity effects could increase with decreased non-native proficiency.

Taken together, our results suggest that CLI is a driving factor of the mechanisms underlying non-native production at intermediate proficiency levels in our study, and not typological similarity. However, the exact contribution of typological similarity may change dynamically as a function of other key factors in multilingual language processing, such as non-native proficiency. Given the scarcity of research, more empirical studies are needed to delve deeper into this issue, see section 6.5.1.

6.5.1 Conclusions

In this study, we asked the question whether or not typologically similar languages bear a processing advantage in non-native production. Our data showed CLI effects at the behavioural level in the form of a processing advantage of congruent and cognate items compared to incongruent and non-cognate items. Unexpectedly, we found no evidence for an effect of typological similarity on behavioural measures: the behavioural performance and CLI effects were comparable across the Italian-Spanish and German-Spanish group. For the EEG data, we found a P300 component across both groups. Further, P300 amplitudes were modulated by cognate status. This highlights the crucial involvement of the P300 component in mitigatory CLI processes. In contrast, there was no traceable effect of typological similarity on voltage amplitudes, reflecting highly similar neural signatures and CLI across the two groups. Taken together,these results suggest a limited role of typological similarity on non-native production at intermediate proficiency levels. We argue that this may be linked to the relatively small difference in proficiency between the L1 and the non-native language.

6.5.2 Future directions

Our study does not allow for a direct assessment of a typological similarity effect at lower non-native proficiency levels in driving CLI. Future research could incorporate varying groups of late learners at different acquisition stages to support our claims. Further, while we carefully recruited our participants to fit the B1/B2 proficiency range and controlled for a number of linguistic variables, future studies should incorporate a more direct measure of overall proficiency to include in the statistical models. In turn, this could be used for more nuanced analysis for participants in the lower vs. higher B1/B2 range. Finally, there is an urgent need for an objective measure of typological similarity to quantify any effects on multilingual language processing and cognition. While the classification of our groups was unambiguous in the current study, this remains a frequent debate in the literature (Van der Slik, 2010).

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 are grateful to the members of the neurolinguistics lab at the University of Konstanz, Carsten Eulitz, Oleksiy Bobrow, and the research assistants Charlotte Englert, Khrystyna Oliynyk, Zeynep Dogan, Fangming Zhang and Anna-Maria Waibel. We also thank Núria Sebastián-Gallés and the CBC team, in particular Ege Ekin Özer, at the Universitat Pompeu Fabra for providing us with the testing facilities. A special thanks goes to Theo Marinis and Maria del Carmen Parafita Couto for their feedback, to Julian Karch for his support in the statistical analyses and to the POLAR lab for their feedback. The statistical analyses were performed using the computing resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. Finally, we want to thank all of our participants for their time.

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/xraz2/?view_only=a708afaf8a684441a95d928829ed6733`

Citation diversity statement

We made a particular effort to include this Citation Diversity Statement in our manuscript. We aim to raise awareness about the systematic underrepresentation of female authors and authors from minority populations in academic publishing (Dworkin et al., 2020). Here, we report the proportion of gender representation in our reference list, wherever this information was available. Our reference list contained 19% woman/woman authors, 41% man/man, 15% woman/man and finally, 17% man/woman authors. In comparison, the equivalents for neuroscience are 6.7% for woman/woman, 58.4% for man/man, 25.5% woman/man, and lastly, 9.4% for man/woman authored references, see Dworkin et al. (2020). In the future, we are convinced that this binary gender classification system will be further improved and that data will become available for additional research areas.

Appendix

6.A Linguistic profile: Italian-Spanish group

Table 6.A.1: *Overview of the native and non-native languages acquired by the Italian-Spanish speakers included in the analysis (n = 28).*

	L1	**L2**	**L3**	**L4**	**L5**	**Total**
Italian	n = 28					**28**
Spanish		n = 2	n = 15	n = 8	n = 3	**28**
English		n = 23	n = 4			**27**
French		n = 3	n = 7	n = 3		**13**
German			n = 1	n = 2		**3**
Portuguese					n = 2	**2**
Catalan					n = 1	**1**
Total	**28**	**28**	**27**	**13**	**6**	

6.B Linguistic profile: German-Spanish group

Table 6.B.1: *Overview of the native and non-native languages acquired by the Italian-Spanish speakers included in the analysis (n = 30).*

	L1	L2	L3	L4	L5	Total
German	n = 30					**30**
Spanish			n = 16	n = 12	n = 2	**30**
English		n = 28	n = 2			**30**
French		n = 2	n = 8	n = 5		**15**
Latin			n = 3	n = 1	n = 1	**5**
Russian			n = 1		n = 1	**2**
Swedish				n = 1		**1**
Portuguese					n = 1	**1**
Catalan					n = 1	**1**
Italian					n = 1	**1**
Mandarin					n = 1	**1**
Total	**30**	**30**	**30**	**19**	**8**	

6.C Model fitting procedure

The model-fitting procedure was as follows: we first constructed a theoretically plausible maximal model. This model included an elaborate fixed effects structure which consisted of our fixed effects and any pre-hypothesised interactions, as well as the covariates. For both our behavioural data and the EEG data, the maximal model included an interaction effect of *typological similarity* with *gender congruency* and *cognate status* in order to test the hypotheses of finding potential differential effects across groups. Our random effects structure was specified as maximally as possible with random slopes as well as random intercepts if supported by the data (Barr, 2013). In the case of non-convergence or singular fit, we first simplified our random effects structure. Next, we proceeded to simplify the fixed effects structure and systematically tested for statistical significance of the fixed effects and the covariates in a top-down fashion. By default, GLMMs were fitted using the maximum likelihood (ML) method with the Laplace approximation (Bates et al., 2020). Absolute test-statistic values larger than ± 1.96 at $\alpha = 0.05$ were defined as statistically significant (Alday et al., 2017). Model comparisons were performed to assess the contribution of each fixed effect using the *anova()* function. This function is based on the Information Criteria AIC and BIC and the loglikelihood ratio. Fixed effects which did not significantly improve the model fit were excluded from the model selection procedure. Model fit was assessed after each model by examining the model residuals using the *DHARMa* package (Hartig, 2020). Treatment coding was used as our default contrast.

6.D Model parameters: naming accuracy

Table 6.D.1: *Model outcome parameters for naming accuracy (n = 58).*

Formula: naming accuracy ~ gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + familiarisation phase performance (zero correct vs. one correct vs. two correct vs. three correct) + (1|subject) + (1|item)

Term	**Odds Ratio [95% CI]**	**z-value**	**p-value**
(Intercept)	0.641 [0.416, 0.987]	-2.02	0.043
Gender congruency [incongruent]	0.728 [0.542, 0.978]	-2.11	**0.035**
Cognate status [non-cognate]	0.696 [0.519, 0.934]	-2.41	**0.016**
Familiarisation phase performance [one correct]	6.342 [4.67, 8.61]	11.82	< 0.001
Familiarisation phase performance [two correct]	26.75 [19.50, 36.69]	20.39	< 0.001
Familiarisation phase performance [three correct]	58.02 [41.63, 80.87]	23.97	< 0.001
Random effects			
σ^2	3.29		
$\tau_{00 Item}$	0.55		
$\tau_{00 Subject}$	0.59		
ICC	0.26		
$N_{Subject}$	58		
N_{Item}	192		
Observations	5,568		
Marginal R^2/ Conditional R^2	0.286/0.469		

6.E Model parameters: naming latencies

Table 6.E.1: *Model outcome parameters for naming latencies (n = 58).*

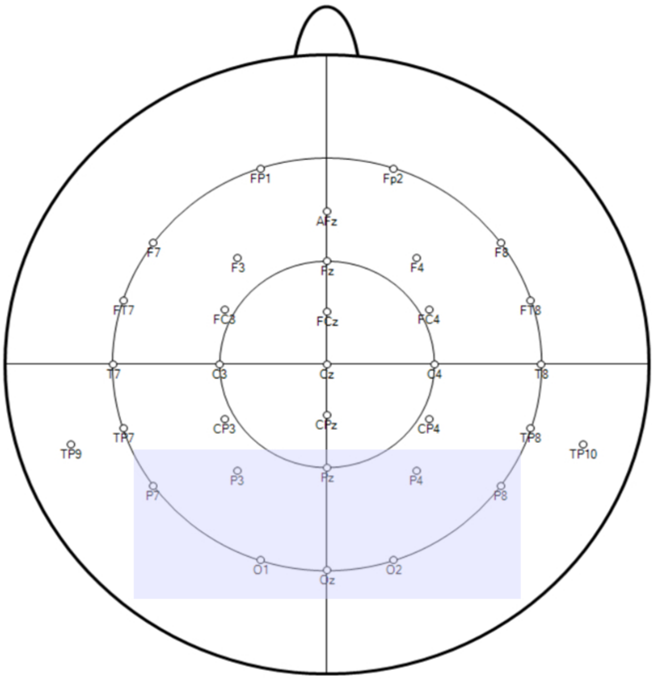
Formula: naming latency ∼ gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + familiarisation phase performance (zero correct vs. one correct vs. two correct vs. three correct) + (gender congruency + cognate status|subject) + (1|item)

Term	Estimate [95% CI]	t-value	p-value
(Intercept)	1.39 [1.31, 1.46]	34.80	< 0.001
Gender congruency [incongruent]	0.064 [0.019, 0.108]	2.82	**0.005**
Cognate status [non-cognate]	0.046 [0.005, 0.087]	2.18	**0.029**
Familiarisation phase performance [one correct]	-0.182 [-0.239, -0.125]	-6.25	< 0.001
Familiarisation phase performance [two correct]	-0.371 [-0.424, -0.318]	-13.71	< 0.001
Familiarisation phase performance [three correct]	-0.445 [-0.499, -0.392]	-16.39	< 0.001
Random effects			
σ^2	0.05		
$\tau_{00 Item}$	0.00		
$\tau_{00 Subject}$	0.00		
$\tau_{11 Subject[incongruent]}$	0.00		
$\tau_{11 Subject[non-cognate]}$	0.00		
$\rho_{01 Subject[incongruent]}$	-0.21		
$\rho_{01 Subject[non-cognate]}$	-0.25		
ICC	0.16		
$N_{Subject}$	58		
N_{Item}	192		

Observations	4,650
Marginal R^2/ Conditional R^2	0.174/0.310

6.F EEG data: region of interest

Figure 6.F.1: *Region of interest and the corresponding data channels for the group comparison, illustrated in the montage of the German-Spanish group.*



6.G Model parameters: P300 component

Table 6.G.1: *Model outcome parameters for voltage amplitudes (n = 58).*

Formula: voltage amplitudes ∼ gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + hemisphere (left vs. midline vs. right) + familiarisation phase performance (zero correct vs. one correct vs. two correct vs. three correct) + (gender congruency + cognate status|subject) + (1|item)

Term	Estimate [95% CI]	t-value	p-value
(Intercept)	4.96 [4.17, 5.74]	12.40	< 0.001
Gender congruency [incongruent]	0.025 [-0.479, 0.529	0.097	0.922
Cognate status [non-cognate]	-0.477 [-0.934, -0.021]	-2.05	**0.040**
Hemisphere [midline]	0.619 [0.598, 0.639]	59.14	< 0.001
Hemisphere [right]	-0.426 [-0.444, -0.408]	-45.63	< 0.001
Familiarisation phase performance [one correct]	0.009 [-0.052, 0.070]	0.292	0.770
Familiarisation phase performance [two correct]	0.113 [0.056, 0.170]	3.87	< 0.001
Familiarisation phase performance [three correct]	0.172 [0.114, 0.230]	5.81	< 0.001
Random effects			
σ^2	73.66		
$\tau_{00 Item}$	1.78		
$\tau_{00 Subject}$	7.60		
$\tau_{11 Subject[incongruent]}$	1.68		
$\tau_{11 Subject[non-cognate]}$	0.99		
$\rho_{01 Subject[incongruent]}$	-0.42		

$\rho_{01Subject[non-cognate]}$	-0.001
ICC	0.11
$N_{Subject}$	58
N_{Item}	192
Observations	4,500,216
Marginal R^2/ Conditional R^2	0.003/0.112