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

Von Grebmer Zu Wolfsthurn, S.

CHAPTER 4

Cross-language effects in comprehension and production: the case of Italian-Spanish speakers

4.1 Introduction

In Chapters 2 and 3 of this thesis, we explored cross-linguistic influence (CLI) between German and Spanish in non-native comprehension and production in German late language learners of Spanish with intermediate proficiency levels (Von Grebmer Zu Wolfsthurn et al., 2021a; Von Grebmer Zu Wolfsthurn, Pablos-Robles & Schiller, 2021b). In this chapter, we expanded on this previous work and applied the experimental design to a linguistically more similar language pair: Italian and Spanish. Both of these languages show a significant overlap in terms of morphosyntax, cognates and phonology (Serratrice, Sorace, Filiaci & Baldo, 2012; Schepens, Dijkstra, Grootjen & Van Heuven, 2013; Van der Slik, 2010). Therefore, they can be considered as linguistically more similar compared to German and Spanish. In line with the previous chapters, the primary aim of the present chapter was the following: we examined how congruency type (i.e., gender congruent or incongruent across Italian and Spanish) and cognate status (i.e., cognate or non-cognate across Italian and Spanish) influenced behavioural and electrophysiological measures of non-native comprehension and production in a syntactic violation task and in a picture-naming task. We focused on the gender congruency effect (Klassen, 2016; Morales et al., 2016) and the cognate facilitation effect (Costa et al., 2005). As discussed in the previous chapters, the gender congruency effect represents CLI at the level of gender and is typically reflected in more accurate and faster processing of gender congruent compared to incongruent nouns, e.g., [il$_M$ cane$_M$ - el$_M$ perro$_M$] *"the dog"* vs. [il$_M$ tavolo$_M$ - la$_F$ mesa$_F$] *"the table"*. On the other hand, the cognate facilitation effect represents CLI at the orthographic and the phonological level and manifests itself in more accurate and faster processing of cognates compared to non-cognates, e.g., [trattore - tractor] *"tractor"* vs. [viso - cara] *"face"* (Costa et al., 2005; Lemhöfer et al., 2008, 2004; Paolieri et al., 2020). In the current chapter, we subsequently explored the interplay between Italian and Spanish at the level of gender, and at the level of orthography and phonology.

From the perspective of non-native comprehension, we embedded our study within the broader theoretical framework of how multilingual speakers represent gender. On the one hand, the gender-integrated representation hypothesis predicts CLI of the gender systems (Bordag & Pechmann, 2007; Morales et al., 2016; Salamoura & Williams, 2007) due to shared gender systems across languages. On the other hand, the gender-autonomous representation hypothesis does not predict such an interplay between languages and proposes independent gender systems for each language (Costa et al., 2003). Similar to Chapter 2, in this chapter we studied how gender congruency influenced non-native syntactic processing in noun phrases such as [el$_M$ perro$_M$] *"the dog"*. We additionally explored cognate status as a potential modulator of syntactic processing in order to examine a possible interaction effect alongside gender congruency. Moreover, we aimed to characterise the neural correlates of syntactic processing in late language learners: some studies suggested that language learners may be less sensitive to syntactic violations compared to more proficient speakers, as reflected in smaller event-related potentials or ERPs (Foucart & Frenck-Mestre, 2012; Gillon-Dowens et al., 2010; Hahne, 2001; Tokowicz & MacWhinney, 2005; Weber-Fox & Neville, 1996; Zawiszewski & Laka, 2020), but see Von Grebmer Zu Wolfsthurn et al. (2021a) in Chapter 2 for contrasting results. Results from the study conducted in Chapter 2 suggested a separate influence of gender congruency and cognate status in German-Spanish and also provided evidence for the gender-integrated representation hypothesis (Bordag & Pechmann, 2007; Salamoura & Williams, 2007). Critically, those previous results for German-Spanish speakers showed the ERP effect typically linked to syntactic violation processing, namely the P600 component (Steinhauer et al., 2009; Von Grebmer Zu Wolfsthurn et al., 2021a). However, it is unclear whether these results are also applicable to a linguistically more similar language pair such as Italian and Spanish.

From the perspective of non-native production, the primary motivation behind both the current study and previous work in Chapter 3 (Von Grebmer Zu Wolfsthurn et al., 2021b) was to examine the effect of CLI on the time course of non-native production. We again focused on the gender congruency effect and the cognate facilitation effect. Both features were shown to significantly modulate non-native production (Lemhöfer et al., 2008; Midgley et al., 2011; Paolieri et al., 2020; Peeters et al., 2013). By extension, this implied that speakers experienced CLI during non-native production. Similar to Chapter 3, the current chapter was concerned with the modulation of the time course of non-native production in light of CLI and the locus of target language selection. We used the LRM model (Levelt et al., 1999) of single word production to test two contrasting accounts of target language selection: one theoretical account previously suggested that the target language was selected before or upon lexical retrieval (Hermans et al., 1998; Lee & Williams, 2001). In contrast, a second account postulated that the target language was selected after lexical retrieval (Christoffels et al., 2007; Colomé, 2001; Hoshino & Thierry, 2011). To discriminate between these two accounts, in this chapter we used the gender congruency effect to examine the lexical retrieval stage, and the cognate facilitation effect to explore the phonological encoding stage described in the LRM model (Levelt et al., 1999). Results from the

German-Spanish speakers from Chapter 3 suggested first, that CLI was traceable at the level of gender and cognates; and second, that CLI continued beyond lexical retrieval into phonological encoding. Moreover, we provided evidence for the P300 ERP component as an index for the mitigation of CLI. Yet, it remains unclear whether these findings were applicable to speakers of highly similar languages (e.g., Italian and Spanish).

Taking previous chapters from this thesis as its starting point, the current chapter extends on the findings reported in Chapters 2 and Chapter 3 and applies an almost identical theoretical framework and methodology to a linguistically highly similar language pair, namely Italian and Spanish. The tasks used in this study were modelled after previous work by Von Grebmer Zu Wolfsthurn et al. (2021a) for the syntactic violation task, and after Von Grebmer Zu Wolfsthurn et al. (2021b) for the picture-naming task. In the next sections, we separately discuss the syntactic violation task and the picture-naming task, followed by a more general discussion.

4.2 Syntactic violation task

4.2.1 Research questions

For the syntactic violation task, the research questions were identical to the ones outlined in Von Grebmer Zu Wolfsthurn et al. (2021a). The questions were as follows: first, whether there was an effect of gender congruency (congruent vs. incongruent) and cognate status (cognate vs. non-cognate) on processing syntactic violations; second, whether there was a P600 effect in late language learners; and finally, whether the P600 effect was modulated by gender congruency and cognate status. Therefore, our focus in terms of the neural correlates of CLI in non-native comprehension was the P600 effect, see Chapter 2 and Von Grebmer Zu Wolfsthurn et al. (2021a).

Hypotheses

We first predicted that participants would be more accurate and faster during non-violation trials compared to violation trials. Second, we also predicted participants to be more accurate and faster for congruent and cognate trials compared to incongruent and non-cognate trials. Critically, we expected *gender congruency* and *cognate status* to have a joint effect on syntactic violation processing: we expected participants to be most accurate and fastest for congruent cognates compared to incongruent non-cognates, in line with the hypotheses in Von Grebmer Zu Wolfsthurn et al. (2021a). We empirically tested this by aggregating *gender congruency* and *cognate status* into the variable *condition* and by including an interaction term for *violation type* and *condition* in the statistical model.

For the EEG data for the syntactic violation task, we expected larger voltage amplitudes in centro-parietal regions around 500 ms to 900 ms post-stimulus onset for violation trials compared to non-violation trials. This would be evidence for a classical P600 effect (Von Grebmer Zu Wolfsthurn et al., 2021a). Moreover, we expected an interactive effect of *gender congruency* and *cognate status* on P600 effect amplitudes. More specifically, we predicted a larger P600 effect for congruent cognates compared to incongruent non-cognates. This would not only be evidence for CLI at the level of gender and cognates, but also that these two linguistic features have a joint influence on the neural correlates underlying syntactic violation processing. Similar to the behavioural analysis, we tested for this by including an interaction effect for *violation type* and *condition* in our statistical analysis.

4.2.2 Methods

Prior to the experiment, participants were asked to complete the Language Proficiency and Experience Questionnaire, LEAP-Q (Kaushanskaya, Blumenfeld & Marian, 2020; Marian et al., 2007). This questionnaire was designed to establish a detailed picture about the linguistic profile of each participant with respect to the acquired languages. During the experiment, participants completed the LexTALE-Esp (Izura et al., 2014), a lexical decision task to measure vocabulary size in Spanish. Participants then alternated between completing the syntactic violation task, as described in Von Grebmer Zu Wolfsthurn et al. (2021a), and the picture-naming task, as described in Von Grebmer Zu Wolfsthurn et al. (2021b). The picture-naming task is described separately in section 4.3. We recorded participants' EEG during both of these latter tasks.

Participants

Participants were 33 native Italian late learners of Spanish living in Barcelona and tested at Pompeu Fabra University. Twenty-four of our participants were female, and participants' mean age was 27.12 years (SD = 4.08). Recruitment criteria were the following: right-handedness, no language, reading or psychological impairments, no second language learnt before five years of age, between 18 and 35 years old and acquisition of Spanish after fourteen years of age. Moreover, participants who had lived in a Spanish-speaking country for more than one year were not included in this study. Critically, participants had to have a B1/B2 level of Spanish (Council of Europe, 2001). This proficiency level was established first, by recruiting participants directly from Spanish language courses for this specific level; and second, by using the measures obtained from the LEAP-Q and the LexTALE-Esp as an additional proxy indicator for their proficiency in Spanish. Participants acquired Spanish at the age of M = 23.94 (SD = 5.07). They reported oral fluency in Spanish at the age of M = 24.89 (SD = 4.48). Reading onset age was M = 24.36 (SD = 4.91) and reading fluency was reached at the age of M = 24.24 (SD = 4.82). On average, participants had spent M = 0.46 years (SD = 0.34) in a Spanish-speaking country. Their self-rated mean speaking proficiency was M = 6.09 (SD = 1.76), their comprehension proficiency M = 7.26 (SD = 1.67) and their reading proficiency M = 7.36 (SD = 1.48) on a scale from one to ten, with ten being maximally proficient. Two participants acquired Spanish as their first foreign language, eighteen participants as their second, ten participants as their third and three participants as their fourth foreign language. Further, thirteen participants reported Spanish as their current second most dominant language after Italian, fourteen as their third, five as their fourth and one participant as their fifth most dominant language. See Appendix 4.A for an overview of the other languages participants reported in the LEAP-Q.

Tasks and stimuli

We used the original LexTALE-Esp by Izura et al. (2014), but excluded three stimuli words for both groups due to overlap with the stimuli from the syntactic violation task. The critical manipulation in this task was *condition* (word vs. pseudoword). We then selected stimuli from the MultiPic database (Duñabeitia et al., 2018) and the Spanish Frequency Dictionary (Davies & Davies, 2017). Both databases included common nouns and pictures of objects in Spanish. We chose highly frequent nouns and pictures where the highest percentage of the correct name of the object was provided in the norming phase. Next, each selected noun was assigned a congruency type (i.e., either *congruent* or *incongruent* across Italian and Spanish), and a cognate status (i.e., either a *cognate* or a *non-cognate* in Italian and Spanish). Cognate status was defined based on orthographic and phonological overlap, and only recognisable cognates were included as stimuli in the respective tasks. Importantly, we did not include identical cognates (e.g., il taxi - el taxi [the taxi]), plural forms of nouns (e.g., gafas [glasses]), professions, English loan words or words with multiple translation equivalents (e.g., el asno/el burro [the donkey]). We modelled the distribution of terminal phonemes of the nouns according to the natural terminal phoneme distribution in Spanish to increase the ecological validity of our stimuli (Clegg, 2011). Further, we included a balanced ratio of feminine-to-masculine nouns, and controlled for syllable length in our stimuli. See Table 4.2.1 for an example set of stimuli for the syntactic violation task. The design was a 2 x 2 x 2 fully factorial within-subjects design with *violation type* (non-violation vs. violation), *congruency type* (congruent vs. incongruent) and *cognate status* (cognate vs. non-cognate) as our three critical manipulations.

Table 4.2.1: *Example stimuli for the syntactic violation task, illustrating the three manipulations of violation type, gender congruency and cognate status.*

		non-violation	
		congruent	**incongruent**
cognate	**Italian** **Spanish**	il$_M$ trattore$_M$ el$_M$ tractor$_M$	il$_M$ grasso$_M$ la$_F$ grasa$_F$
		the tractor	*the fat*
non-cognate	**Italian** **Spanish**	il$_M$ cane$_M$ el$_M$ perro$_M$	il$_M$ tavolo$_M$ la$_F$ mesa$_F$
		the dog	*the table*
		violation	
		congruent	**incongruent**
cognate	**Italian** **Spanish**	il$_M$ pane$_M$ *la$_F$ pan$_M$	la$_F$ labbra$_F$ *la$_F$ labio$_M$
		the bread	*the lip*
non-cognate	**Italian** **Spanish**	il$_M$ fiume$_M$ *la$_F$ río$_M$	il$_M$ viso$_M$ *el$_M$ cara$_F$
		the river	*the face*

Procedure

During the experiment, participants were comfortably seated in front of a computer screen in an experimental booth. They were provided with an information sheet and gave informed consent before proceeding to the tasks, in line with the ethics guidelines at the Faculty of Humanities at Leiden University. Participants completed the LexTALE-Esp before the syntactic violation task. Both tasks were programmed in E-prime2 (Psychology Software Tools, Inc).

The procedure for the LexTALE-Esp and the syntactic violation task for each task was identical to Von Grebmer Zu Wolfsthurn et al. (2021a). For the LexTALE-Esp, participants made a lexical decision on whether or not the string on the screen was a Spanish word while accuracy was measured. Post-task, we calculated a LexTALE-Esp vocabulary scores by subtracting the percentage of yes-answers to pseudowords from the percentage of yes-answers to words. The resulting vocabulary size score (*LexTALE-Esp score*) was included as a covariate in the analysis of the syntactic violation task. For the syntactic violation task, participants were instructed to use keyboard buttons to indicate their familiarity with the noun, and to then provide a judgement as accurately and fast as possible of whether or not the presented noun phrase was correct. during this task, we measured both accuracy and response times (RTs), as well as voltage amplitudes. The two major differences to the procedure in Chapter 2 were first, that the stimuli differed significantly for the Italian-Spanish group due to constraints by gender congruency type and cognate status; and second, that the information sheet, the consent form, the oral and written task instructions and the debrief were provided in the participants' native language Italian.

EEG recordings

EEG data were collected from 32 active Ag/AgCl channels at 500 Hz configured in a 10/20 montage from BrainProducts. Channel FT9 was placed underneath the left eye to record the vertical electrooculogram (VEOG), and channel FP10 on the outer canthus of the left eye to record the horizontal electrooculogram (HEOG). We positioned the ground electrode on the participants' right cheek. The original reference channel was FCz, and the impedances for all channels were configured to be below 10kΩ for optimal EEG signal conductivity. EEG data was recorded using BrainVision Recorder (BrainProducts GmbH).

4.2.3 Results

Behavioural data analysis

The behavioural data analysis procedure was identical as in Von Grebmer Zu Wolfsthurn et al. (2021a). Moreover, we included the same participants in the behavioural analyses and in the EEG analyses, see the next section. Subsequently, 29 of the 33 participants were included for the analysis of this task. To model accuracy and

RTs, we employed a linear mixed effects model (LMM) approach (Baayen et al., 2008) in R via RStudio (R Core Team, 2020) using the *lme4* package (Bates et al., 2020). Our model selection procedure was as follows: first, we separately specified a theoretically plausible maximal model for accuracy and for RTs. More concretely, we specified a generalised linear mixed effects model (GLMM) with a binomial distribution to model accuracy for familiar trials, and a GLMM with a gamma distribution and the identity link function to model positively skewed RTs for familiar and correct trials (Lo & Andrews, 2015). Each maximal model included the interaction term for *violation type* and *condition*, the variable aggregating both *congruency type* and *cognate status*. Further, we included several covariates, such as *LexTALE-Esp score*, *terminal phoneme* of the target word, *target noun gender* and *order of acquisition of Spanish*. Moreover, we included random intercepts for each *participant* and individual *item*, as well as by-participant random slopes for the effect of *condition*. In a second step, in the case of non-convergence or singular fit of the model, we first simplified the random effects structure. We then tested for the statistical relevance of the covariates and interaction effects by systematically comparing models with an without a particular term by using the *anova()* function. For each model, we performed model diagnostics to assess the goodness of fit via the *DHARMa* package (Hartig, 2020). Absolute test-statistics larger than 1.96 were interpreted as being statistically significant at $\alpha = 0.05$ (Alday et al., 2017). Treatment coding was our default contrast.

EEG data exclusion

Our inclusion criteria for the EEG analysis for this task were the following: we only included trials where participants had indicated familiarity with the noun prior to the experimental trial. Second, trials which were incorrectly identified as violations or non-violations were excluded, as were trials containing artefacts. Taking these criteria together, we only included participants with more than 60% of (valid) trials left in their data. Subsequently, we excluded four participants from the EEG analysis, adding to a total of 29 included datasets. The same datasets were included in the behavioural data analysis.

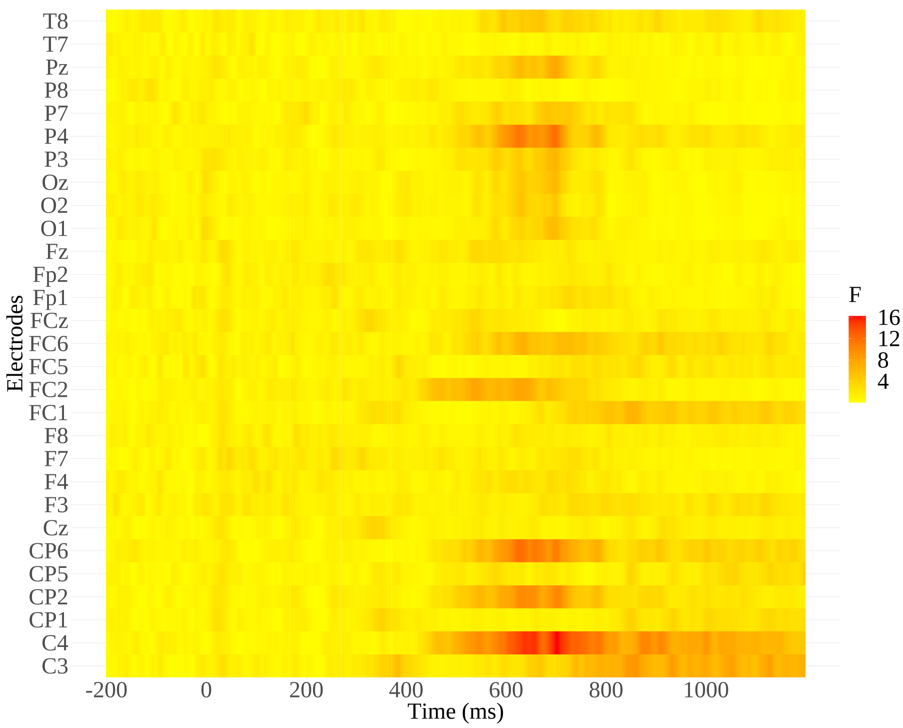
EEG data pre-processing

We thoroughly pre-processed the EEG data to increase the signal-to-noise ratio and to minimise noise related to artefacts, for example jaw muscle movement, eye blinks, or other external interferences (Ganushchak, Christoffels & Schiller, 2011; Porcaro et al., 2015). For this, we used BrainVision Analyzer (BrainProducts GmbH). As the first pre-processing step, we applied the average of the mastoid electrodes TP9 and TP10 as the new references. In this, FCz was reused as a regular data channel. Next, we filtered the data using a high-pass filter of 0.1 Hz, and a low-pass filter of 30 Hz. Channel interpolation was performed if deemed appropriate given the quality of the surrounding channels. We then performed residual drift detection in preparation for ocular independent component analysis (ICA) for blink correction. After this step, we performed an artefact search across all data channels. The criteria for artefact detection were the following: for gradient, the maximal voltage step was defined as 50 μV/ms, the maximal difference in 100 ms - intervals as 200 μV, the maximal amplitude as $\pm$ 200 μV, and the lowest allowable amplitudes in 100 ms - intervals as 0.5 μV. In a final step, we segmented our EEG signal on the basis of the stimulus onset, thereby generating segments between -200 ms prior and 1,200 ms post-stimulus. Segments were then baseline-corrected using the activity in the 200 ms prior to the stimulus onset. We exported all available valid segments for each channel and participant. As described above, we defined valid trials as those trials where participants had indicated familiarity with the noun, provided a correct response and were artefact-free for the statistical analysis (Christoffels et al., 2007).

EEG data analysis

We performed a cluster-based permutation analysis on the exported EEG data to determine our region of interest (ROI). For this, we used the *permutes* package (Voeten, 2019) in R. This analysis is particularly powerful because it reveals potentially significant differences in voltage amplitudes by condition for each channel. The outcome is measured in F-values, with larger F-values indicating an increased likelihood for a statistically relevant effect of our manipulations on voltage amplitudes. The permutation analysis suggested channels *C4, CP2, CP6, P3, P4, P7, P8* and *Pz* in centro-parietal regions as ROI in the time-window between 500 ms and 800 ms. We then performed the statistical analysis using these channels as our ROI. See Figure 4.2.1 for the permutation analysis outcome for this task.

Figure 4.2.1: *Permutation test outcome for the syntactic violation task (n = 29). Larger F-values are shown in darker colours and denote an increased likelihood for a statistically relevant effect of our manipulations on voltage amplitudes.*



Similar to the behavioural analyses, we followed a single-trial LMM approach for the EEG data analysis using the *lme4* package (Bates et al., 2020). We based this approach on work by Frömer et al. (2018). In this, we examined voltage amplitudes from all exported valid segments. The modelling procedure was as follows: first, we specified a theoretically feasible maximal LMM model using a Gaussian distribution. The maximal model consisted of the interaction between *violation type* and *condition* (which combined *congruency type* and *cognate status* into a single variable), the covariates *channel*, *LexTALE-Esp score*, *terminal phoneme* of the target word, *target noun gender* and *order of acquisition of Spanish*, random intercepts for each participant and individual item, and finally, correlated by-participant random slopes for the effect of *violation type* and *condition*. We did not specify an interaction random slope with *violation type* to avoid over-parametrisation (Matuschek et al., 2017). Next, we thoroughly evaluated the goodness of fit of this model using the *DHARMa* package (Hartig, 2020). In the event of non-convergence, we simplified the random effects structure, followed by the systematic evaluation of the contribution of the covariates in the fixed effects structure. We then performed model comparisons using the *anova()* function to establish our model of best fit. We again used treatment coding as our contrast, and absolute test-statistics larger than 1.96 were interpreted as showing a significant effect on voltage amplitudes at $\alpha = 0.05$ (Alday et al., 2017).

Data results

We first calculated descriptive statistics for mean accuracy and RTs. See Table 4.2.2 for mean accuracy per condition and Table 4.2.3 for mean RTs per condition.

Table 4.2.2: *Mean accuracy for each condition for the syntactic violation task (n = 29).*

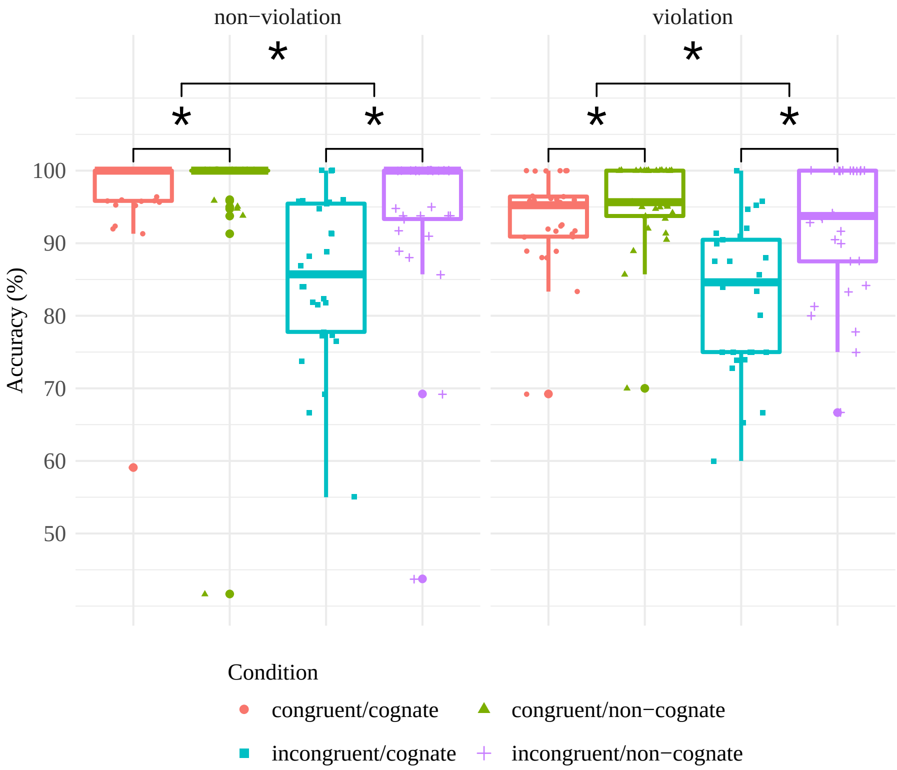
Condition	**Mean accuracy (%)**	**SD**
non-violation/congruent/cognate	96.65	18.01
non-violation/congruent/non-cognate	96.47	18.47
non-violation/incongruent/cognate	85.97	34.76
non-violation/incongruent/non-cognate	93.72	24.28
violation/congruent/cognate	93.17	25.25
violation/congruent/non-cognate	95.56	20.63
violation/incongruent/cognate	82.75	37.81
violation/incongruent/non-cognate	91.99	27.17

Table 4.2.3: *Mean RTs for each condition for the syntactic violation task (n = 29).*

Condition	**Mean RTs (ms)**	**SD**
non-violation/congruent/cognate	816.76	328.44
non-violation/congruent/non-cognate	817.31	347.17
non-violation/incongruent/cognate	931.11	452.49
non-violation/incongruent/non-cognate	854.38	377.15
violation/congruent/cognate	953.55	401.33
violation/congruent/non-cognate	929.73	356.10
violation/incongruent/cognate	1049.37	489.65
violation/incongruent/non-cognate	991.48	439.02

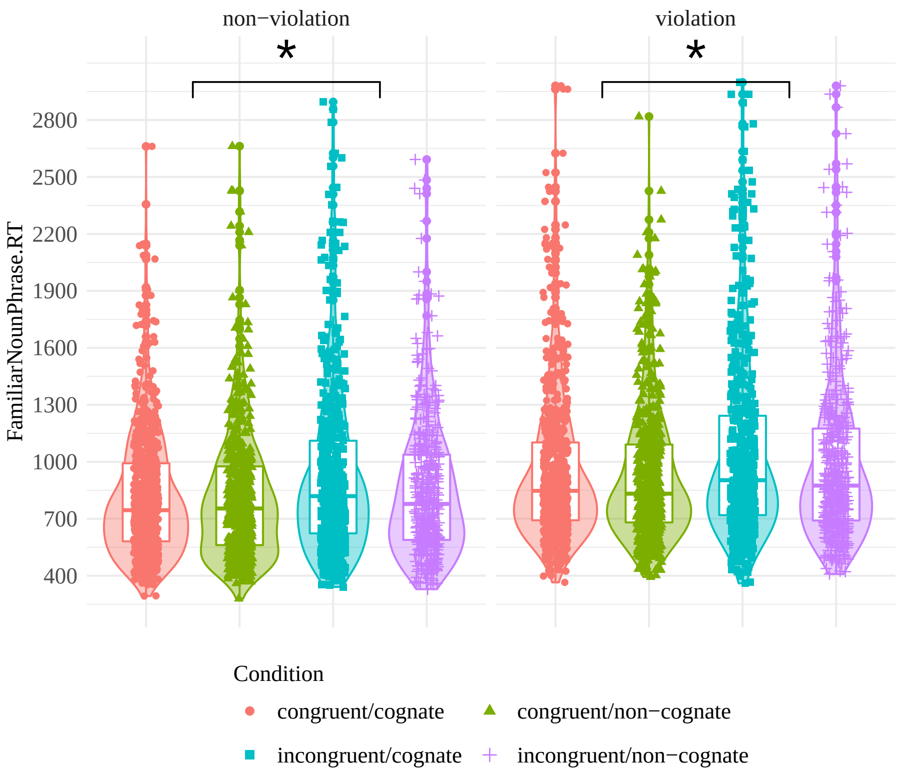
Accuracy. For accuracy, the maximal model did not converge and was subsequently simplified. The simplified model, which included the interaction between *violation type* and *condition*, did not yield a better model fit compared to a model without the interaction $\chi^2(1, \text{n} = 29) = 0.894$, $p = 0.827$. We then only included a main effect for *violation type* and an interaction effect between *gender congruency* and *cognate status* in the subsequent model and compared it with a model without this interaction term but only main effects. This comparison yielded a better model fit for the latter model, with $\chi^2(1, \text{n} = 29) = 1.28$, $p = 0.258$. Therefore, our model of best fit included a main effect of *violation type*, *gender congruency* and *cognate status*, as well as correlated random slopes for *gender congruency* and *cognate status* for the random effect of *participant*. Moreover, *item* was included as a random effect. None of the covariates significantly contributed to a better model fit and were therefore not included in the best-fitting model. Taken together, the model of best fit was as follows: accuracy $\sim$ violation type (violation vs. non-violation) + gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + (gender congruency + cognate status|participant) + (1|item). Participants were more accurate for *congruent* compared to *incongruent* trials with $\beta = 0.321$, *95% CI*[0.195, 0.531], $z = -4.44$, $p < 0.001$, and for *non-cognates* compared to *cognates* with $\beta = 2.11$, *95% CI*[1.31, 3.40], $z = 3.07$, $p = 0.002$. Contrary to our predictions, we did not find evidence that our participants were more accurate for *non-violation* trials compared to *violation* trials with $\beta = 0.679$, *95% CI*[0.449, 1.03], $z = -1.84$, $p = 0.066$. See Appendix 4.B for model parameters and Figure 4.2.2 for a visualisation of the accuracy results. Note that model parameters are reported as odds ratios.

Figure 4.2.2: *Visualisation of mean accuracy for each condition for the syntactic violation task (n = 29). The brackets indicate statistical significance between conditions.*



Response times. For RTs, the maximal model yielded non-convergence and was therefore simplified. Our best-fitting model for RTs included main effects for *violation type*, *gender congruency* and *cognate status*, and an interaction effect for *gender congruency* and *cognate status*. Further, the model also included correlated by-participant random slopes for *gender congruency* and *cognate status* in addition to a random effect for *item*. The best-fitting model was the following: RTs ∼ violation type (violation vs. non-violation) + gender congruency (congruent vs. incongruent) * cognate status (cognate vs. non-cognate) + (gender congruency + cognate status|participant) + (1|item). Participants were significantly faster for *non-violation* trials compared to *violation trials* with $\beta = 112.80$, *95% CI*[102.54, 123.06], $t = 21.55$, $p < 0.001$. Further, the main effect for *gender congruency* was significant with $\beta = 123.42$, *95% CI*[110.16, 136.69], $t = 18.24$, $p < 0.001$ for *congruent* compared to *incongruent* trials. The main effect of *cognate status* was not significant with $\beta =$ -6.83, *95% CI*[-15.95, 2.29], $t =$ -1.47, $p = 0.142$. The interaction effect for *gender congruency* and *cognate status* was significant, with $\beta =$ -68.33, *95% CI*[-79.05, -57.60], $t =$ -12.49, $p < 0.001$. See Appendix 4.C for model parameters and Figure 4.2.3 for a visualisation of the RT results.

Figure 4.2.3: *Visualisation of mean RTs for each condition for the syntactic violation task (n = 29). The brackets indicate statistical significance between conditions.*



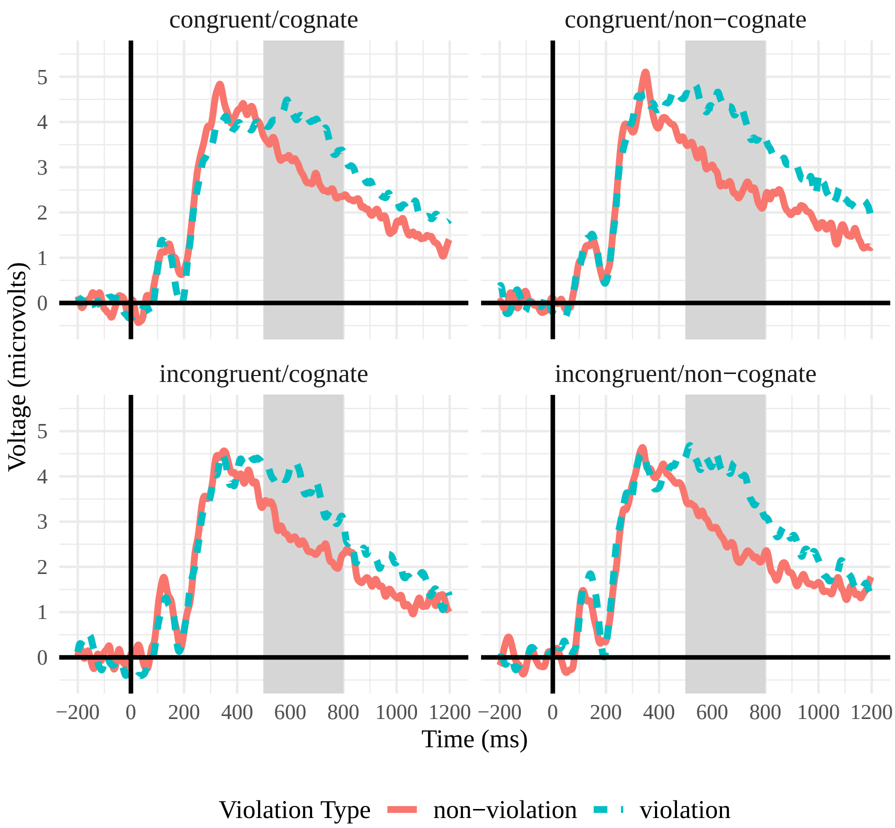
Voltage amplitudes. We calculated mean voltage amplitudes for each condition for the time window between 500 ms and 800 ms post-stimulus onset for the selected channels in centro-parietal regions (Table 4.2.4).

Table 4.2.4: *Voltage amplitudes by condition for the time window of interest (500 ms - 800 ms) for channels C4, CP2, CP6, P3, P4, P7, P8 and Pz for the syntactic violation task (n = 29).*

Condition	**Mean voltage (μV)**	**SD**
non-violation/congruent/cognate	1.99	8.51
non-violation/congruent/non-cognate	2.16	8.23
non-violation/incongruent/cognate	1.78	8.06
non-violation/incongruent/non-cognate	1.92	7.93
violation/congruent/cognate	2.61	8.27
violation/congruent/non-cognate	2.81	8.83
violation/incongruent/cognate	2.24	8.06
violation/incongruent/non-cognate	2.47	8.37

Following the analysis procedure outlined above, the model of best fit for voltage amplitudes included the main effects for *violation type*, *gender congruency* and *cognate status*, the covariates *channel*, *terminal phoneme* and *LexTALE-Esp score*, correlated random slopes for by-participant effects of *violation type* and *condition* and random intercepts for both *participant* and *item*. Therefore, the final model was: voltage amplitudes ∼ violation type (violation vs. non-violation) + gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + channel + terminal phoneme + LexTALE-Esp score + (violation type + gender congruency * cognate status|participant) + (1|item). See Appendix 4.D for the full model parameters. The interaction effect between *violation type* and *condition* on P600 voltage amplitudes was neither significant nor did it significantly improve the model fit. It was therefore dropped from the model-fitting procedure. Subsequently, these results did not provide evidence for a significant modulation of P600 voltage amplitudes as a function of the CLI effects. In addition, the main effects for *gender congruency* and *cognate status* in the best-fitting model were insignificant with β = -0.072, *95% CI*[-0.551, 0.406], t = -0.296, p = 0.767 for congruent compared to incongruent trials, and with β = 0.108, *95% CI*[-0.330, 0.545], t = 0.483, p = 0.629 for cognates compared to non-cognates. Nevertheless, voltage amplitudes were significantly higher for violation trials compared to non-violation trials with β = 1.48, *95% CI*[0.893, 2.08], t = 4.92, p < 0.001 (Appendix 4.D). Figure 4.2.4 visualises mean voltage amplitudes over time for each condition for our ROI.

Figure 4.2.4: *Visualisation of voltage amplitudes for each condition for channels C4, CP2, CP6, P3, P4, P7, P8 and Pz (n = 29). The time window of interest is highlighted in grey.*



4.2.4 Discussion

The aim of this study was to examine CLI in non-native comprehension in Italian late learners of Spanish. First, we explored whether and how gender congruency and cognate status affected non-native language comprehension in the context of a syntactic violation task; second, we examined whether there was evidence for a sensitivity to syntactic errors in the form of a P600 effect in Italian late language learners of Spanish; and finally, we studied whether P600 amplitudes were modulated by gender congruency and cognate status. We predicted that participants would be more accurate and faster for non-violation compared to violation trials. We also predicted participants to be most accurate and fastest at detecting congruent cognates compared to incongruent non-cognates for syntactic violations. Finally, we predicted larger P600 voltage amplitudes for violation compared to non-violation trials, as well as larger amplitudes for congruent cognates compared to incongruent non-cognates in an interaction effect with violation type.

With respect to our first research question, our behavioural results suggested that participants were significantly faster, but not more accurate for non-violation compared to violation trials. Critically, participants were more accurate and faster for congruent compared to incongruent items. This reflects the classical gender congruency effect (Klassen, 2016; Lemhöfer et al., 2008). Interestingly, participants were also more accurate for non-cognates compared to cognates. This reflects a reverse cognate facilitation effect, with higher accuracy for non-cognates instead of cognates. In contrast, we found no evidence for an effect of cognate status on RTs. Importantly, we also did not find evidence for an interaction effect between gender congruency and cognate status with violation type. This suggested that the performance in detecting syntactic violations was not significantly modulated by a joint effect of gender congruency and cognate status. Therefore, with respect to our first research question, we found some evidence for differential processing of violation vs. non-violation trials and congruent vs. incongruent trials, thereby reflecting the effects of violation and gender congruency on behavioural measures of non-native comprehension. However, as we had previously predicted, we did not find evidence that gender congruency and cognate status had a joint effect on detecting syntactic violations. In addition, we found a significant interaction effect of gender congruency and cognate status on RTs, suggesting that the effect of one factor was dependent on the other and vice versa.

Comparing the current findings with the results from Chapter 2 on German-Spanish speakers (Von Grebmer Zu Wolfsthurn et al., 2021a), the results from both groups are highly fascinating for several reasons: first, the findings across both groups are compatible in that both participant groups were faster for non-violation trials compared to violation trials. This reflects differential processing of NPs with a syntactic error vs. syntactically correct NPs. Secondly, both the Italian-Spanish speakers and the German-Spanish speakers displayed the classical gender congruency effect, with more accurate and faster processing for congruent compared to incongruent items (Bordag & Pechmann, 2007; Lemhöfer et al., 2008). In turn, this particular finding supports the gender-integrated representation hypothesis (Bordag & Pechmann, 2007; Morales et al., 2016; Salamoura & Williams, 2007). In other words, the results from the current study support a theoretical framework whereby the gender systems are shared between two languages, irrespective of the linguistic similarity of the languages. Thirdly, we found comparable effects of cognate status in both the Italian-Spanish speakers and the German-Spanish speakers: the former were more accurate for non-cognates compared to cognates, while the latter were faster for non-cognates compared to cognates. This is an interesting finding as it suggests that cognate status may have a similar reverse effect on non-native comprehension across two different language pairs in late learners. Finally, in line with findings from the German-Spanish speakers, we did not find evidence for an interaction effect of gender congruency and cognate status on detecting syntactic violations in the Italian-Spanish speakers. Overall, the behavioural results from both studies show that participants were more sensitive to an overlap in gender rather than to cognate status. They also showed that gender congruency and cognate status together had a limited effect on detecting syntactic violations in this task. However, one significant difference to the German-Spanish speakers was that gender congruency and cognate status did emerge as having an overall interaction effect on RTs in the Italian-Spanish speakers, while there was no interaction of gender congruency and cognate status with violation type in either group. Subsequently, this indicates a small joint influence of these two linguistic features on RTs in the syntactic violation task.

With respect to our second and third research question regarding the P600 effect, results from the EEG data suggested that violation trials elicited larger voltage amplitudes compared to non-violation trials. This reflected the classical P600 effect (Hahne & Friederici, 2001; Steinhauer et al., 2009) and was in line with our original hypotheses. Therefore, our Italian-Spanish speakers were indeed sensitive to syntactic violations even at moderate proficiency levels. This contrasts with previous research reporting no P600 effect for late language learners (Hahne & Friederici, 2001; Weber-Fox & Neville, 1996), but is consistent with more recent research presenting evidence for a P600 effect in late learners (S. Rossi et al., 2006; Tokowicz & MacWhinney, 2005; Von Grebmer Zu Wolfsthurn et al., 2021a). Critically, matching the behavioural results, we did not find differential P600 effects as a function of the interaction effect between gender congruency and cognate status. In other words, we found no evidence that CLI effects modulated P600 effect sizes. This is in contrast to our predictions and suggests that the P600 effect was comparable in size across our experimental conditions.

The EEG findings from the Italian-Spanish speakers in this chapter are highly similar to the findings reported for the German-Spanish speakers. First, we again found evidence for a P600 effect for syntactic violations in our Italian-Spanish speakers. Secondly, we similarly did not find evidence that CLI effects modulated P600 effects and instead found similar P600 amplitudes across our conditions. In this, our results were compatible with our previous work because they suggested, on the one hand, a sensitivity to syntactic violations even at moderate proficiency levels. On the other hand, they did not show that CLI for gender or cognates had a significant impact on these processes per se. Therefore, our results support the gender-integrated representation hypothesis (Salamoura & Williams, 2007). Importantly, our results add to the findings from Chapter 2: independently of whether the languages were linguistically highly similar (Italian-Spanish) or less similar (German-Spanish), speakers displayed a P600 effect despite their early acquisition stages and intermediate proficiency levels. The P600 effect, in turn, was largely unaffected by the linguistic features of congruency type and cognate status we put to test in Chapter 2 and in the current chapter. Questions remain as to whether or not there was a statistical difference in terms of the P600 effect sizes across the Italian-Spanish and the German-Spanish group, and whether CLI effects differed across these two groups. These issues were the scope of Chapter 5 of this thesis, which includes a direct comparison of the data from the Italian-Spanish and German-Spanish speakers in light of language similarity.

Summary and conclusions

In this study, we explored non-native comprehension in the context of a linguistically similar language pair, Italian and Spanish. Therefore, this study represented an extension to Chapter 2, where we studied the linguistically less similar language pair German and Spanish. Behavioural results from the syntactic violation task in the current chapter showed that participants were sensitive to gender congruency across Italian and Spanish. This was reflected in a processing advantage for congruent items compared to incongruent items. In contrast, the feature of cognate status was less salient in the context of non-native comprehension. Further, we found ERP evidence for a P600 effect in our late language learners. This P600 effect, however, did not appear to be modulated by neither gender congruency nor cognate status. Generally speaking, we provided support for the gender-integrated representation hypothesis (Salamoura & Williams, 2007), and demonstrated that language learners with limited non-native proficiency displayed a clear sensitivity to syntactic violations. Our results are also highly similar to the findings reported in Chapter 2 and in Von Grebmer Zu Wolfsthurn et al. (2021a).

4.3 Picture-naming task

In this task, we tested the same participants outlined in section 4.2 on non-native production. As described above, participants completed the LexTALE-Esp (Izura et al., 2014) before the syntactic violation task and the picture-naming task. Importantly, participants alternated between first completing the syntactic violation task vs. the picture-naming task.

4.3.1 Research questions

We asked the question whether gender congruency (congruent vs. incongruent) and cognate status (cognate vs. non-cognate) modulated the behavioural and neural correlates of non-native production. Here, we focused specifically on P300 amplitudes, see Von Grebmer Zu Wolfsthurn et al. (2021b). Second, we used the gender congruency and the cognate facilitation effect to examine *when* during non-native production speakers experienced CLI between the native and the non-native language. By extension, our third research question was concerned with the locus of target language selection: when during non-native production is the target language selected? These questions are identical to the research questions in Chapter 3 and in Von Grebmer Zu Wolfsthurn et al. (2021b).

Hypotheses

Behaviourally, we were interested in whether gender congruency and cognate status would impact naming accuracy and naming latencies. More specifically, we predicted more accurate and faster naming of congruent cognate nouns compared to incongruent non-cognate nouns, in line with previous findings from Chapter 3 and Von Grebmer Zu Wolfsthurn et al. (2021b). In contrast, we did not expect significant behavioural differences in naming accuracy and latencies between congruent non-cognates and incongruent cognates.

For the EEG data, we first examined whether a P300 effect would be also elicited in the Italian-Spanish group. Moreover, we predicted P300 amplitudes to be modulated by gender congruency and cognate status: in line with the results reported in Chapter 3, we hypothesised larger P300 amplitudes for congruent cognate nouns compared to incongruent non-cognate nouns. In contrast, we expected similar amplitudes for congruent non-cognates and incongruent cognates. Therefore, we predicted the smallest P300 amplitudes for incongruent non-cognates, and largest amplitudes for congruent cognates.

4.3.2 Methods

Participants

The participants were the same as the ones described in section 4.2.2.

Tasks and stimuli

The stimuli selection procedure for the picture-naming task was identical to the procedure described for the syntactic violation task in section 4.2.2. Critically however, the stimuli differed from the stimuli used in the previous task and in Chapter 3. The design for the picture-naming task was a 2 x 2 fully factorial within-subjects design with *gender congruency* (congruent vs. incongruent) and *cognate status* (cognate vs. non-cognate) as our critical manipulations. See Table 4.3.1 for an example stimuli set.

Table 4.3.1: *Example picture stimuli for the picture-naming task, illustrating the two manipulations of gender congruency and cognate status.*

Condition	**Noun phrase**	**Italian translation**	**English translation**
congruent/ cognate	la_F llave_F	la_F chiave_F	*the key*
congruent/ non-cognate	la_F gonna_F	la_F falda_F	*the skirt*
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*

Procedure

The task procedure was identical to Chapter 3 and Von Grebmer Zu Wolfsthurn et al. (2021b). Participants were placed in an experimental booth in front of a computer screen and presented with the stimuli pictures. During the task, we recorded participants' voice using a built-in microphone while they produced the name of the object together with the corresponding determiner as accurately and fast as possible. In this task, we measured naming accuracy, naming latencies and voltage amplitudes.

EEG recordings

The EEG recording set-up was identical to the syntactic violation task, see section 4.2.2.

4.3.3 Results

We closely followed the behavioural data analysis procedure described in section 4.2.3 and in Chapter 3, see also Von Grebmer Zu Wolfsthurn et al. (2021b). Here we also included the same participants in both the behavioural analysis and in the EEG analysis, see the next section. Therefore, we included 28 participants in the behavioural analysis. We followed a LMM approach (Baayen et al.,

2008) in R via RStudio (R Core Team, 2020) using the *lme4* package (Bates et al., 2020) to model naming accuracy and naming latencies. We used a GLMM with a binomial distribution to model naming accuracy, and a GLMM with a gamma distribution and the identity link function (Lo & Andrews, 2015) to model correctly named stimuli pictures. The model selection procedure was identical to section 4.2.3, with the following exceptions: first, the fixed effects structure of the maximal models for naming accuracy and naming latencies included an interaction effect between *gender congruency* and *cognate status* and the covariates *LexTALE-Esp score, familiarisation phase performance, order of acquisition of Spanish, target noun gender, word length* and *terminal phoneme*. Second, the random effects consisted of random slopes for the interaction effect between *gender congruency* and *cognate status* for each participant, as well as random intercepts for each *participant* and *item*.

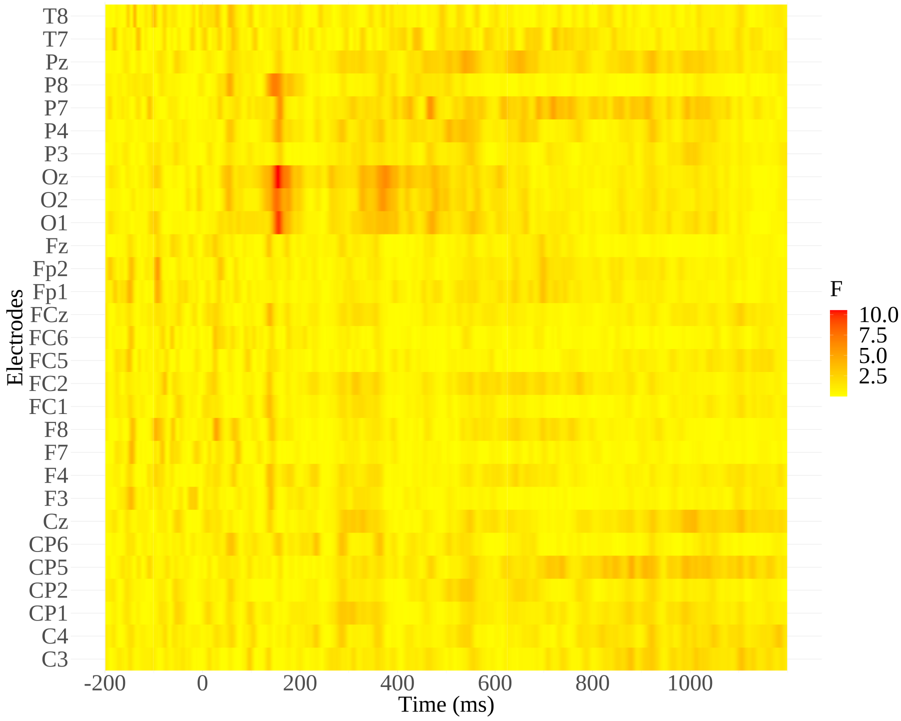
EEG data exclusion

For the EEG analysis, the inclusion criteria consisted of correctly named trials, as well as artefact-free trials and a sufficiently high data quality in terms of artefacts. In this, we excluded participants with less than 60% of (valid) trials left after the application of these criteria. Subsequently, five participants were excluded for the picture-naming task, thereby including 28 datasets in the behavioural and in the EEG analysis of this task.

EEG data analysis

The EEG data analysis procedure was modelled after section 4.2.3. Data pre-processing was especially critical as production data is often characterised by large articulatory artefacts (Grözinger et al., 1975). Therefore, we set an upper threshold of 600 ms post-stimulus onset to avoid those artefacts in our signal as much as possible. We again used a cluster-based permutation test to determine our ROI. This permutation analysis outcome suggested channels *P3, P4, P7, P8, Pz, O1, O2* and *Oz* in centro-parietal regions as ROI in the time window between 350 to 600 ms for the picture-naming task, see Figure 4.3.1.

Figure 4.3.1: *Permutation test outcome for the picture-naming task (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.*



The EEG data analysis procedure was identical as in section 4.2.3. We used a LMM approach to model voltage amplitudes. The maximal model included an interaction term for *gender congruency* and *cognate status*, as well as the covariates *hemisphere*, *LexTALE-Esp score*, *terminal phoneme* of the target word, *target noun gender*, *order of acquisition of Spanish* and *familiarisation phase performance*. Finally, we also included random slopes for the interaction effect of *gender congruency* and *cognate status* for each participant, and random intercepts for each *participant* and *item*.

Data results

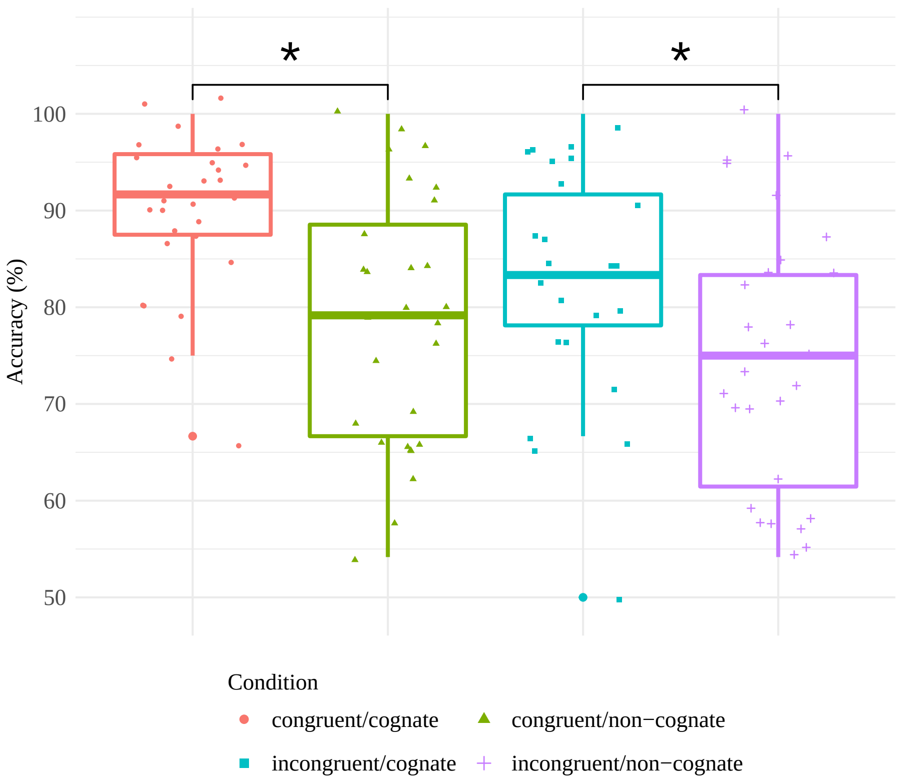
We first computed descriptive statistics for naming accuracy and naming latencies. Mean naming accuracy and mean naming latencies for each condition are shown in Table 4.3.2.

Table 4.3.2: *Mean naming accuracy and latencies by condition for the picture-naming task (n = 28).*

Condition	**Mean naming accuracy (%)**	**SD**	**Mean naming latencies (ms)**	**SD**
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

Naming accuracy. The model of best fit included the following terms: main effects for *gender congruency* and *cognate status*, as well as *LexTALE-Esp score* and *familiarisation phase performance* as covariates. The random slopes for our main manipulations for each participant led to singular fit. Similarly, the model with included the interaction effect between *gender congruency* and *cognate status* did not yield a better model fit with χ^2(1, n = 28) = 0.494, p = 0.482. We subsequently simplified our fixed effects and random effects structures and included random intercepts for *participant* and *item*. Our best-fitting model was therefore as follows: naming accuracy ~ gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + LexTALE-Esp score + familiarisation phase performance (none correct vs. one correct vs. two correct vs. three correct) + (1|participant) + (1|item). Participants were significantly more accurate for cognates over non-cognates with β = 0.647, *95% CI*[0.443, 0.946], z = -2.24, p = 0.025. In contrast, participants were not more accurate for congruent items compared to incongruent items with β = 0.807, *95% CI*[0.550, 1.18], z = -1.10, p = 0.270. See Appendix 4.E for details about the model parameters, and Table 4.3.2 for mean naming accuracy across the conditions.

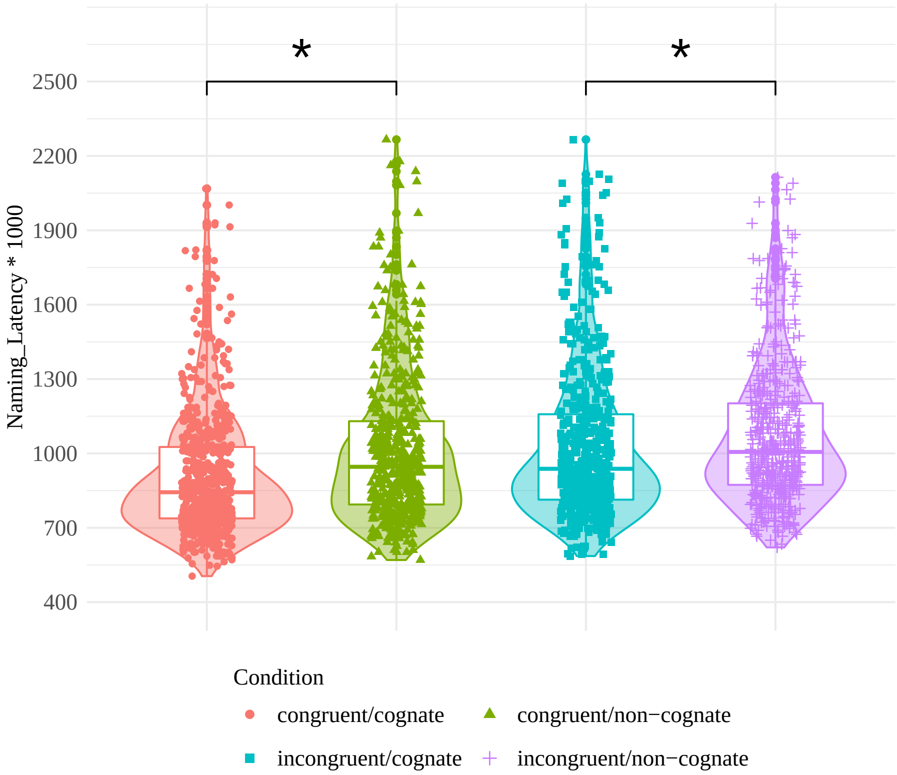
Figure 4.3.2: *Visualisation of naming accuracy for each condition for the picture-naming task (n = 28). The brackets indicate statistical differences across conditions.*



Naming latencies. Our best-fitting model included a main effect for both *gender congruency* and *cognate status*, *familiarisation phase performance* as covariate, correlated random slopes for *gender congruency* and *cognate status* for each participant, and random intercepts for *participant* and *item*. The model containing the interaction effect between *gender congruency* and *cognate status* was not statistically better compared to the model without with $\chi^2(1, \text{n} = 28) = 0.256$, $p = 0.613$. Therefore, the best-fitting model was: naming latencies $\sim$ gender congruency (congruent vs.

incongruent) + cognate status (cognate vs. non-cognate) + familiarisation phase performance (none correct vs. one correct vs. two correct vs. three correct) + (gender congruency + cognate status|participant) + (1|item). Critically, participants were faster at naming cognates compared to non-cognates, with $\beta = 0.071$, *95% CI*[0.007, 0.135], $t = 2.18$, $p = 0.030$. Participants were marginally not faster at naming congruent compared to non-congruent items, with $\beta = 0.067$, *95% CI*[-0.001, 0.135], $t = 1.94$, $p = 0.052$. See Appendix 4.F for the parameters of the best-fitting model, and Figure 4.3.3 for a visualisation of mean naming latencies across conditions. Note that model estimates are reported in seconds. Taken together, *cognate status*, but not *gender congruency*, significantly influenced both naming accuracy and naming latencies in this study.

Figure 4.3.3: *Visualisation of naming latencies for each condition for the picture-naming task (n = 28). The brackets indicate statistical differences across conditions.*



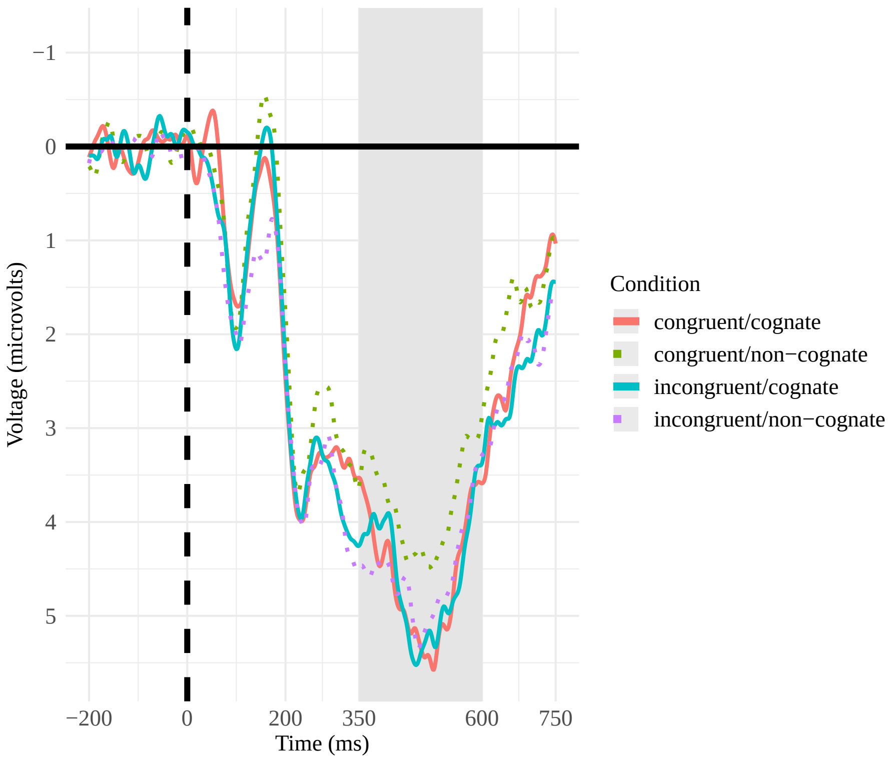
Voltage amplitudes. Visual inspection of the EEG data across the entire segment showed the classical N1/P2/N2 ERP complex for early visual processing (Eulitz et al., 2000) in our ROI (Figure 4.3.4). Voltage amplitudes reached a peak around 550 ms, which was followed by a downward trend back to baseline. Descriptive statistics for the time window between 350 ms and 550 ms for channels P3, P4, P7, P8, Pz, O1, O2 and Oz can be found in Table 4.3.3. Descriptively, we found the highest voltage amplitudes in this particular time window for congruent cognates, followed by incongruent non-cognates, incongruent cognates and finally, congruent non-cognates.

Table 4.3.3: *Voltage amplitudes by condition for the time window of interest (350 ms - 550 ms) for channels P3, P4, P7, P8, Pz, O1, O2 and Oz for the picture-naming task (n = 28).*

Condition	**Mean voltage(μV)**	**SD**
congruent/cognate	4.77	8.71
congruent/non-cognate	3.96	8.91
incongruent/cognate	4.72	8.16
incongruent/non-cognate	4.74	8.35

Our maximal model as outlined in section 4.3.3 failed to converge. We subsequently simplified our fixed and random effects structures. The best-fitting model for voltage amplitudes included *condition* as fixed effect, as well as the covariate *hemisphere* and *familiarisation phase performance*. The best-fitting model was as follows: voltage amplitudes ∼ condition (congruent cognate vs. congruent non-cognate vs. incongruent cognate vs. incongruent non-cognate) + hemisphere (left vs. midline vs. right) + familiarisation phase performance (none correct vs. one correct vs. two correct vs. three correct) + (condition|participant) + (1|item). Despite being included in the final model, there was no statistical difference between the condition levels; see Appendix 4.G for the exact model parameters. Therefore, statistically speaking, there was no significant modulation of voltage amplitudes as a function of *condition*, and voltage amplitudes were statistically comparable across *congruency type* and *cognate status*. In addition, the model included a random slope for the effect of *condition* for each random intercept of *participant*, as well as a random intercept for *item*.

Figure 4.3.4: *Visualisation of voltage amplitudes for each condition for channels P3, P4, P7, P8, Pz, O1, O2 and Oz (n = 28).*



4.3.4 Discussion

In this study, we conducted an almost identical experiment to Von Grebmer Zu Wolfsthurn et al. (2021b), except that instead of German-Spanish speakers, we tested Italian-Spanish speakers and we used stimuli which would fit with the constraints of gender congruency and cognate status across Italian and Spanish. The aim was to examine whether CLI, in particular with respect to congruency type (congruent vs. incongruent) and cognate status (cognate vs. non-cognate), had an effect on non-native production. We were particularly interested in naming accuracy, naming latencies and P300 voltage amplitudes. Taking the LRM model (Levelt et al., 1999) as the theoretical basis, we investigated during which stage of non-native language production speakers would experience measurable CLI. By extension, the current study tapped directly into the question at which stage during non-native production the target language was selected over the non-target language. On the basis of this theoretical framework and the findings from Chapter 3 of this thesis, we predicted the following: higher naming accuracy, shorter naming latencies and larger P300 voltage amplitudes for congruent cognate items, followed by congruent non-cognates and incongruent cognates, and the lowest naming accuracy, longest RTs and smallest P300 voltage amplitudes for incongruent non-cognates.

Regarding the behavioural data, we found that gender congruency and cognate status did not yield an interaction effect on neither naming accuracy nor naming latencies, contrary to our hypotheses. However, in line with our predictions, we found a significant effect of cognate status on naming accuracy and naming latencies: participants were both more accurate and faster at naming cognates compared to non-cognates. This reflects the classical cognate facilitation effect (Christoffels et al., 2007; Peeters et al., 2013). From the perspective of target language selection, this indicates that the speaker experienced CLI at the orthographic and phonological level. In other words, lexical entries from both Italian and Spanish competed at the *phonological encoding* stage of production. In turn, this suggested that the target language was not selected when lexical retrieval was completed (Christoffels et al., 2007; Colomé, 2001). Interestingly, we found no effect of gender congruency on naming accuracy or naming latencies. Despite a statistical trend, the main effect of gender congruency marginally failed to reach statistical significance. Therefore, our results did not provide evidence for differential production of congruent vs. incongruent items, and therefore also not for a gender congruency effect.

The behavioural results from this study are highly relevant for multiple reasons: first, results from the previous study on German-Spanish speakers suggested that participants were more accurate and faster for congruent vs. incongruent nouns but did not yield an effect of cognate status (Von Grebmer Zu Wolfsthurn et al., 2021b). In contrast, the results of the present study reflect the opposite pattern, namely that participants were more accurate and faster for cognates compared to non-cognates, but not for congruent compared to incongruent nouns. This suggests that during non-native production, German-Spanish speakers were more sensitive to similarities and dissimilarities at the gender level, whereas Italian-Spanish speakers were more sensitive to similarities at the orthographic and phonological level. In turn, CLI effects were traceable primarily during the stage of *lexical retrieval* for the German-Spanish speakers, and during the *phonological encoding* stage for the Italian-Spanish speakers. In other words, our findings from both studies suggest that German-Spanish speakers battled CLI mostly when processing gender, whereas Italian-Spanish speakers faced CLI primarily when processing cognates at the behavioural level. Subsequently, this suggested qualitative and quantitative differences in the underlying non-native production mechanisms for the linguistically similar languages, i.e., Italian and Spanish, compared to linguistically less similar languages, i.e., German and Spanish. Importantly, the statistical comparison of the behavioural differences in non-native production between the German-Spanish speakers and the Italian-Spanish speakers can be found in Chapter 6.

As for the EEG results, the oscillatory pattern in centro-parietal regions between 350 ms and 600 ms post-stimulus onset was consistent with a P300 component, in line with our original hypothesis. Therefore, similar to the results reported in Chapter 3 and in Von Grebmer Zu Wolfsthurn et al. (2021b), the P300 emerged as a critical component during non-native language production. However, we did not find an effect of gender congruency or cognate status on P300 voltage amplitudes: the descriptive trend of smaller voltage amplitudes for congruent non-cognates compared to the other three conditions was not statistically significant. Therefore, we did not find neural evidence for a modulation of P300 amplitudes by gender congruency or cognate status. In turn, these results do not support the notion of CLI during non-native production, as was indicated in the behavioural modulation by cognate status in the behavioural results. A possible interpretation of this finding is that the modulation of P300 amplitudes was too subtle and potentially masked by the remaining noise in our data, despite meticulous pre-processing of the ERP data. Another interpretation is that effects of gender congruency and cognate status on P300 amplitudes had a counterbalancing effect; thereby effectively cancelling any neural effects but preserving a behavioural effect of cognate status. Taken together, we did not find traceable CLI effects at the level of P300 component amplitudes. In turn, this did not allow for the exploration of the locus of target language selection. Subsequently, we were unable to examine during which production stage our speakers experienced CLI from either gender congruency or cognate status. This particular notion therefore remains an open issue for future research. Nevertheless, our results again highlight the relevance of the P300 component during non-native production.

Comparing these EEG results with those from Chapter 3 and Von Grebmer Zu Wolfsthurn et al. (2021b), where incongruent non-cognates descriptively elicited the smallest voltage amplitudes for the German-Spanish speakers, the present study linked congruent non-cognates to the smallest elicited voltage amplitudes. However, unlike the German-Spanish speakers, the Italian-Spanish speakers did not display a significant difference in voltage amplitudes as a function of condition. In other words, ERP results from the German-Spanish speakers indicated that both the native and non-native language were active beyond the stage of lexical retrieval until at least the stage of phonological encoding. Conversely, we did not find any evidence for this in the current study. These are relevant findings because they indicate a potentially different neural signature of non-native production for German-Spanish speakers compared to Italian-Spanish speakers. The contrast in voltage amplitudes across the two groups is examined in more detail in Chapter 6 of this thesis, which describes a direct comparison of the EEG data for the German-Spanish speakers and the Italian-Spanish speakers. Finally, results from both studies provide evidence for the critical role of the P300 component in non-native production, both in linguistically similar and less similar language combinations. Future research should therefore examine the exact characteristics and involvement of the P300 component in non-native production in a more nuanced manner.

Summary and conclusions

In this chapter, we examined the effect of gender congruency (congruent vs. incongruent) and cognate status (cognate vs. non-cognate) on non-native production. Within the framework of the LRM model (Levelt et al., 1999), we probed the effect of CLI on non-native production and the locus of target language selection. We used a picture-naming task in native Italian late learners of Spanish. We were particularly focused on the impact of CLI on naming accuracy, naming latencies and P300 voltage amplitudes. Our behavioural results showed that participants were more accurate and faster at naming cognates compared to non-cognates. In contrast, they did not show an effect of gender congruency. From a neural perspective, we found no evidence that gender congruency or cognate status modulated P300 voltage amplitudes. Behaviourally, our results therefore suggested that participants faced CLI until the *phonological encoding* stage, which is consistent with Chapter 3 and Von Grebmer Zu Wolfsthurn et al. (2021b). However, we did not find complementary evidence in the EEG data. We were therefore unable to examine the locus of target language selection in non-native production in our Italian late learners of Spanish. Importantly, this notion therefore warrants a closer investigation in future studies.

4.4 General discussion

The aim of this chapter was to expand on the cross-linguistic evidence from Chapters 2 and 3 and to investigate the gender congruency effect and the cognate facilitation effect in a linguistically highly similar language pair, namely Italian-Spanish. More specifically, we used a syntactic violation paradigm to quantify CLI effects in non-native comprehension, and a picture-naming task to investigate CLI effects in non-native production. We explored several different issues: in terms of non-native comprehension, we asked the question whether and how gender congruency (congruent vs. incongruent) and cognate status (cognate vs. non-cognate) influenced syntactic violation processing. Moreover, we characterised the corresponding neural correlates and investigated whether our late language learners with moderate proficiency levels would show the P600 effect, which is typically reported for syntactic violations (Steinhauer et al., 2009). Finally, we were interested in whether P600 effect sizes would be modulated by gender congruency and cognate status as representatives for cross-linguistic influence (CLI) effects in non-native comprehension. In contrast, for non-native production, our goals were to explore whether and how gender congruency and cognate status modulated the non-native production process. Further, we also studied until when speakers experienced CLI effects related to gender congruency and cognate status and when the target language was selected over the non-target language during non-native production. Here, our main ERP component of interest was the P300 component.

The general picture emerging from our findings in this chapter is the following: first, in non-native comprehension, gender congruency emerged as the primary salient linguistic feature to impact performance on a syntactic violation task. This suggests that there is interaction between the Italian and Spanish gender systems, resulting in measurable CLI effects at the behavioural level in non-native comprehension. Next, we provided evidence for a P600 effect in our Italian late learners of Spanish with moderate proficiency levels. This is a critical finding because it suggests that there are distinct neural signatures for processing syntactically correct vs. incorrect structures. In turn, this has implications for the characterisation of the comprehension processes in late language learners. Finally, the P600 effect was statistically comparable independently of any potential influences from gender congruency and cognate status. This suggests that at earlier acquisition stages, we are not yet able to describe distinct neural signatures as a function of CLI effects. Second, in non-native production, behavioural results showed that cognate status was the more salient cue during the production process. This suggested that both languages were active until the later stages of non-native production, and that CLI persisted at least until the phonological encoding stage. This is consistent with previous research suggesting that the target language is selected after lexical retrieval (Christoffels et al., 2007; Colomé, 2001; Hoshino & Thierry, 2011). By extension, this finding implied that Italian late learners of Spanish with moderate proficiency levels may have faced CLI for a large part of the production process. However, in the absence of complementary evidence at the neural level, further research is needed to corroborate these findings.

Comparing these findings from both experiments on non-native comprehension and production in the Italian-Spanish speakers, the striking difference was that gender congruency was a modulator for non-native comprehension, but that cognate status played a more significant role in non-native production. This is a critical finding because it speaks directly to the respective relevance of linguistic features such as gender congruency and cognate status across the linguistic domains of comprehension vs. production. In turn, this suggests that speakers use different linguistic cues to successfully manage non-native comprehension and non-native production. The systematic comparison of the CLI effects across the two domains is beyond the scope of this chapter (but see Chapter 5 and Chapter 6). However, we argue that future research should investigate this particular notion more closely to provide a more detailed and nuanced picture of CLI effects across comprehension and production.

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 the CDC lab group at University of Pompeu Fabra, in particular Nuría Sebastián-Gallés and Ege Ekin Özer. We are also grateful to Julian Karch for his support in the statistical analyses. The analyses were performed using the computing resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University. A special thanks goes to all of our participants for their time and participation.

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/tjea9/?view_only=cc2a2daf17bd4cf38a42ec8fc06b20ce`

Citation diversity statement

This citation diversity statement serves the purpose of highlighting the systematic bias in academia against citing from women and members of minorities (Dworkin et al., 2020; Zurn et al., 2020). We therefore calculated the distribution of gender in terms of the first and last author in our reference list. We included papers written by 17.6% woman/woman authors, 31.4% man/man, 23.5% woman/man and finally, 15.7% man/woman authors. For lack of a direct comparison, in the field neuroscience the reference lists typically contain 6.7% woman/woman authors, 58.4% man/man, 25.5% woman/man and finally, 9.4% man/woman authors (Dworkin et al., 2020).

Appendix

4.A Linguistic profile: Italian-Spanish group

Table 4.A.1: *Overview of the native and non-native languages acquired by the Italian-Spanish speakers (N = 33).*

	L1	L2	L3	L4	L5	**Total**
Italian	n = 33					**33**
Spanish		n = 2	n = 18	n = 10	n = 3	**33**
English		n = 27	n = 5			**32**
French		n = 4	n = 8	n = 3		**15**
German			n = 1	n = 2		**3**
Portuguese					n = 3	**3**
Catalan				n = 1	n = 1	**2**
Total	**33**	**33**	**32**	**16**	**7**	

4.B Model parameters: accuracy

Table 4.B.1: *Specification of model of best fit for accuracy for the syntactic violation task (n = 29). Note that estimates are reported as odds ratios.*

Formula: accuracy ∼ violation type (violation vs. non-violation) + gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + (gender congruency + cognate status|participant) + (1|item)

Term	**Odds Ratio [95% CI]**	**z-value**	**p-value**
(Intercept)	42.78 [24.70, 74.01]	13.40	< 0.001
Violation type [violation]	0.679 [0.449, 1.03]	-1.84	0.066
Gender congruency [incongruent]	0.321 [0.195, 0.531]	-4.44	$<$ **0.001**
Cognate status [non-cognate]	2.11 [1.31, 3.40]	3.07	**0.002**
Random effects			
σ^2	3.29		
$\tau_{00 Item}$	1.22		
$\tau_{00 Participant}$	0.63		
$\tau_{11 Participant[incongruent]}$	0.30		
$\tau_{11 Participant[non-cognate]}$	0.11		
$\rho_{01 Participant[incongruent]}$	-0.49		
$\rho_{01 Participant[non-cognate]}$	0.97		
ICC	0.38		
$N_{Participant}$	29		
N_{Item}	224		
Observations	4,754		
Marginal R^2/ Conditional R^2	0.087 / 0.435		

4.C Model parameters: response times

Table 4.C.1: *Specification of model of best fit for response times (RTs) for the syntactic violation task (n = 29). Note that estimates are reported in milliseconds.*

Formula: RTs ∼ violation type (violation vs. non-violation) + gender congruency (congruent vs. incongruent) * cognate status (cognate vs. non-cognate) + (gender congruency + cognate status|participant) + (1|item)

Term	**Estimate [95% CI]**	**t-value**	**p-value**
(Intercept)	864.35 [844.89, 883.81]	87.08	< 0.001
Violation type [violation]	112.80 [102.54, 123.06]	21.55	$<$ **0.001**
Gender congruency [incongruent]	123.42 [110.12, 136.69]	18.24	$<$ **0.001**
Cognate status [non-cognate]	-6.83 [-15.95, 2.29]	-1.47	0.142
Gender congruency [incongruent] * cognate status [non-cognate]	-68.33 [-79.05, -57.60]	-12.49	$<$ **0.001**
Random effects			
σ^2	0.14		
$\tau_{00 Item}$	5098.76		
$\tau_{00 Subject}$	7983.03		
$\tau_{11 Subject[incongruent]}$	3701.49		
$\tau_{11 Subject[non-cognate]}$	1095.85		
$\rho_{01 Subject[incongruent]}$	-0.04		
$\rho_{01 Subject[non-cognate]}$	-0.22		
ICC	1.00		
$N_{Subject}$	29		
N_{Item}	224		
Observations	4,374		
Marginal R^2/ Conditional R^2	0.293 / 1.000		

4.D Model parameters: P600 component

Table 4.D.1: *Specification of model of best fit for voltage amplitudes for the syntactic violation task (n = 29). Note that estimates are reported in microvolts.*

Formula: voltage amplitudes ∼ violation type (violation vs. non-violation) + gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + channel + terminal phoneme + LexTALE-Esp score + (violation type + gender congruency * cognate status|participant) + (1|item)

Term	**Estimate [95% CI]**	**t-value**	**p-value**
(Intercept)	3.03 [1.86, 4.20]	5.09	< 0.001
Violation type [violation]	1.48 [0.893, 2.08]	4.92	< **0.001**
Gender congruency [incongruent]	-0.072 [-0.551, 0.406]	-0.296	0.767
Cognate status [non-cognate]	0.108 [-0.330, 0.545]	0.483	0.629
Channel [CP2]	0.693 [0.667, 0.720]	51.11	< 0.001
Channel [CP6]	0.173 [0.146, 0.199]	12.75	< 0.001
Channel [P3]	0.549 [0.523, 0.576]	40.50	< 0.001
Channel [P4]	0.906 [0.879, 0.932]	66.78	< 0.001
Channel [P7]	-2.08 [-2.10, -2.05]	-153.14	< 0.001
Channel [P8]	-1.44 [-1.47, -1.41]	-106.21	< 0.001
Channel [Pz]	1.154 [1.13, 1.18]	85.10	< 0.001
Terminal phoneme [d]	2.94 [0.256, 5.63]	2.15	0.032
Terminal phoneme [e]	-0.896 [-1.54, -0.251]	-2.72	0.006
Terminal phoneme [ión]	-0.547 [-2.12, 1.03]	-0.682	0.495
Terminal phoneme [l]	0.318 [-0.604, 1.24]	0.676	0.499
Terminal phoneme [n]	0.020 [-0.627, 0.667]	0.060	0.952
Terminal phoneme [o]	-0.308 [-0.735, 0.118]	-1.42	0.157
Terminal phoneme [r]	-0.459 [-1.50, 0.580]	-0.866	0.387

Terminal phoneme [s]	0.968 [-0.914, 2.85]	1.01	0.314
Terminal phoneme [z]	-0.967 [-3.62, 1.69]	-0.713	0.476
LexTALE-Esp score	-0.012 [-0.046, 0.023]	-0.657	0.511
Random effects			
σ^2	59.10		
$\tau_{00 Item}$	1.78		
$\tau_{00 Participant}$	2.89		
$\tau_{11 Participant[violation]}$	1.71		
$\tau_{11 Participant[congr/non-cogn]}$	1.01		
$\tau_{11 Participant[incongr/cogn]}$	1.53		
$\tau_{11 Participant[incongr/non-cogn]}$	1.61		
$\rho_{01 Participant[violation]}$	-0.03		
$\rho_{01 Participant[congr/non-cogn]}$	-0.53		
$\rho_{01 Participant[incongr/cogn]}$	-0.61		
$\rho_{01 Participant[incongr/non-cogn]}$	-0.45		
ICC	0.08		
$N_{Participant}$	29		
N_{Item}	224		
Observations	5,141,248		
Marginal R^2/ Conditional R^2	0.027 / 0.110		

4.E Model parameters: naming accuracy

Table 4.E.1: *Specification of model of best fit for naming accuracy for the picture-naming task (n = 28). Note that estimates are reported as odds ratios.*

Formula: naming accuracy ∼ gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + LexTALE-Esp score + familiarisation phase performance (none correct vs. one correct vs. two correct vs. three correct) + (1|participant) + (1|item)

Term	**Odds Ratio [95% CI]**	**z-value**	**p-value**
(Intercept)	0.326 [0.159, 0.672]	-3.04	0.002
Gender congruency [incongruent]	0.807 [0.550, 1.18]	-1.10	0.270
Cognate Status [non-cognate]	0.647 [0.443, 0.946]	-2.24	**0.025**
LexTALE-Esp score	1.02 [1.00, 1.05]	2.35	0.019
Familiarisation phase performance [one correct]	5.54 [3.67, 8.35]	8.17	< 0.001
Familiarisation phase performance [two correct]	22.63 [14.91, 34.33]	14.66	< 0.001
Familiarisation phase performance [three correct]	43.57 [28.32, 67.03]	17.18	< 0.001
Random effects			
σ^2	3.29		
$\tau_{00 Item}$	0.47		
$\tau_{00 Subject}$	0.38		
ICC	0.20		
$N_{Subject}$	28		
N_{Item}	96		

Observations	2,688
Marginal R^2/ Conditional R^2	0.307/0.448

4.F Model parameters: naming latencies

Table 4.F.1: *Specification of model of best fit for naming latency for the picture-naming task (n = 28). Note that estimates are reported in seconds.*

Formula: naming latencies ∼ gender congruency (congruent vs. incongruent) + cognate status (cognate vs. non-cognate) + familiarisation phase performance (none correct vs. one correct vs. two correct vs. three correct) + (gender congruency + cognate status|participant) + (1|item)

Term	**Estimate [95% CI]**	**t-value**	**p-value**
(Intercept)	1.36 [1.25, 1.47]	24.98	< 0.001
Gender congruency [incongruent]	0.067 [-0.001, 0.135]	1.94	0.052
Cognate Status [non-cognate]	0.071 [0.007, 0.135]	2.18	**0.030**
Familiarisation phase performance [one correct]	-0.200 [-0.276, -0.124]	-5.16	< 0.001
Familiarisation phase performance [two correct]	-0.338 [-0.408, -0.268]	-9.47	< 0.001
Familiarisation phase performance [three correct]	-0.412 [-0.482, -0.341]	-11.46	< 0.001
Random effects			
σ^2	0.05		
τ_{00Item}	0.00		
$\tau_{00Subject}$	0.00		
$\tau_{11Subject[incongruent]}$	0.00		
$\tau_{11Subject[non-cognate]}$	0.00		
$\rho_{01Subject[incongruent]}$	-0.20		
$\rho_{01Subject[non-cognate]}$	-0.39		
ICC	0.17		

$N_{Subject}$	28
N_{Item}	96
Observations	2,191
Marginal R^2/ Conditional R^2	0.168/0.306

4.G Model parameters: P300 component

Table 4.G.1: *Specification of model of best fit for voltage amplitudes for the picture-naming task (n = 28). Note that estimates are reported in microvolts.*

Formula: voltage amplitudes ∼ Condition (congruent cognate vs. congruent non-cognate vs. incongruent cognate vs. incongruent non-cognate) + hemisphere (left vs. midline vs. right) + familiarisation phase performance (none correct vs. one correct vs. two correct vs. three correct) + (condition|participant) + (1|item)

Term	Estimate [95% CI]	t-value	p-value
(Intercept)	3.48 [2.14, 4.81]	5.10	< 0.001
Condition [congruent/cognate]	0.657 [-0.347, 1.66]	1.28	0.200
Condition [incongruent/cognate]	0.781 [-0.324, 1.89]	1.39	0.166
Condition [incongruent/non-cognate]	0.652 [-0.446, 1.75]	1.16	0.244
Hemisphere [midline]	0.243 [0.211, 0.274]	15.16	< 0.001
Hemisphere [right]	0.083 [0.055, 0.111]	5.84	< 0.001
Familiarisation phase performance [one correct]	0.231 [0.144, 0.318]	5.21	< 0.001
Familiarisation phase performance [two correct]	0.164 [0.084, 0.245]	4.004	< 0.001
Familiarisation phase performance [three correct]	0.684 [0.603, 0.766]	16.38	< 0.001
Random effects			
σ^2	64.64		
$\tau_{00 Item}$	1.70		

$\tau_{00 Participant}$	10.96
$\tau_{11 Participant[congr/cogn]}$	3.39
$\tau_{11 Participant[incongr/cogn]}$	4.92
$\tau_{11 Participant[incongr/non-cogn]}$	4.82
$\rho_{01 Participant[congr/cogn]}$	-0.60
$\rho_{01 Participant[incongr/cogn]}$	-0.79
$\rho_{01 Participant[incongr/non-cogn]}$	-0.62
ICC	0.12
$N_{Participant}$	28
N_{Item}	96
Observations	1,696,396
Marginal R^2/ Conditional R^2	0.002/0.123