Simulating event-related potentials in bilingual sentence comprehension: syntactic violations and syntactic transfer

Stephan Verwijmeren (stephan.verwijmeren@ru.nl)
Department of AI, Radboud University
Nijmegen, The Netherlands

Stefan L. Frank (stefan.frank@ru.nl)
Centre for Language Studies, Radboud University
Nijmegen, The Netherlands

Hartmut Fitz (hartmut.fitz@mpi.nl)
Max Planck Institute, Radboud University
Nijmegen, The Netherlands

Yung Han Khoe (yunghan.khoe@ru.nl)
Centre for Language Studies, Radboud University
Nijmegen, The Netherlands

Abstract

Event-related potentials (ERPs) are used to study how language is processed in the brain, including differences between native (L1) and second-language (L2) processing. A P600 ERP effect can be measured in proficient L2 learners in response to an L2 syntactic violation, indicating native-like processing. Cross-language similarity seems to be a factor that modulates P600 effect size. This manifests in a reduced P600 effect in response to a syntactic violation in the L2 when the syntactic feature involved is expressed differently in two languages. We investigate if this reduced P600 effect can be explained by assuming that ERPs reflect learning signals that arise from mismatches in predictive processing; and in particular that the P600 reflects the error that is back-propagated through the language system (Fitz & Chang, 2019). We use a recurrent neural network model of bilingual sentence processing to simulate the P600 (as back-propagated prediction error) and have it process three types of syntactic constructions differing in cross-language similarity. Simulated English-Spanish participants displayed a P600 when encountering constructions that are similar between the two languages, but a reduced P600 for constructions that differ between languages. This difference between the two P600 responses mirrors what has been observed in human ERP studies. Unlike human participants, simulated participants showed a small P600 response to constructions unique to the L2 (i.e., grammatical gender), presumably because of how this grammatical feature is encoded in the model. Our modelling results shed further light on the viability of error propagation as an account of ERPs, and on the effects of syntactic transfer from L1 to L2.

Keywords: Event-related potential; P600; bilingualism; cross-language similarity; syntactic transfer; recurrent neural network; sentence processing

Introduction

Event-related potentials in bilingualism

Electroencephalography is a technique for recording electrical voltage potentials produced by neural activity. Recorded potentials can be analyzed in relation to cognitive events, yielding interpretable patterns called event-related potentials (ERPs; Morgan-Short, 2014). ERP effects have for instance been observed in response to reading words in sentence-comprehension studies. More specifically, syntactic violations result in an increased positivity in the ERP waveform that starts at around 600 ms after observing an anomalous word, as compared to its correct counterpart (Osterhout & Mobley, 1995). This ERP effect is called a P600.

The P600 effect has been used to investigate if second-language (L2) learners show similar ERP effects as native (L1) speakers for morpho-syntactic processing. L2 proficiency is the most important factor determining P600 size (Antonicelli & Rastelli, 2022; Caffarra, Molinaro, Davidson, & Carreiras, 2015; McLaughlin et al., 2010; Morgan-Short, 2014) but similarities and differences between the L1 and L2 often modulate the effect of proficiency. Some ERP studies showed reduced P600 effects, or no P600 effect, for syntactic features that are instantiated differently between languages (Antonicelli & Rastelli, 2022; Liu, Dunlap, Tang, Lu, & Chen, 2017; Morgan-Short, 2014), while others found P600 effects for syntactic L2 features regardless of the (dis)similarity between L1 and L2 (Caffarra et al., 2015; McLaughlin et al., 2010; Morgan-Short, 2014). There appears to be a complex influence of L1-L2 similarity. Native-like L2 processing (i.e., showing a native-like P600 response) of syntactic features that are unique to the L2 is possible (Foucart & Frenck-Mestre, 2012; McLaughlin et al., 2010; Morgan-Short, 2014), as is native-like L2 processing of syntactic features that are expressed similarly in the L1 and L2 (Foucart & Frenck-Mestre, 2011; McLaughlin et al., 2010; Morgan-Short, 2014). But when a syntactic feature is present but expressed differently in the two languages, the P600 seems to be less sensitive to syntactic violation in the L2 (Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005).

Tokowicz and MacWhinney (2005) presented native English speaking learners of L2 Spanish with Spanish sentences containing syntactic violations. There were three types of syntactic violations: verb-tense violation, determiner gender violation, and determiner number violation (see Table 1). A sentence with a *tense violation* contained a verb in the progressive tense without an auxiliary verb. The syntactic con-

Table 1: Constructions containing syntactic violations with Spanish example sentences and their English translation. Words indicated with an asterisk are experimentally manipulated (here shown in the violation condition). Critical words are underlined. Table adapted from Tokowicz and MacWhinney (2005).

Violated feature	Similarity	Example sentence Spanish	English translation
Tense	Similar	Su abuela *cocinando muy bien	His grandmother *cooking very well
Determiner gender	Unique	Ellos fueron a *un fiesta	They went to *a-MASC party
Determiner number	Different	*El niños están jugando	*The-SING boys are playing

struction for the progressive tense is **similar** between Spanish and English. In a sentence with a *determiner gender violation*, the gender of a noun phrase was switched to the incorrect gender, resulting in a violation at the following noun. This syntactic construction is **unique** to Spanish compared to English, since the English language does not express grammatical gender. In a sentence with a *determiner number violation*, the number of the determiner was switched to the incorrect number, resulting in a violation at the following noun. In both languages, plurality of a noun is expressed by an inflectional morpheme suffix on the noun. However, unlike English, Spanish also expresses plurality in the determiner preceding a noun, which makes the syntactic construction **different** from English. Tokowicz and MacWhinney (2005) found that the P600 effect was reduced (in fact, it was not statistically significant) for determiner number violations compared to the other two types, which suggests that aspects of L1 syntax affect L2 processing; a phenomenon known as syntactic transfer. Specifically, the fact that number is not expressed on the determiner in English would make native English speakers less sensitive to determiner number in L2 Spanish. The same does not apply to determiner gender because there is no English grammatical gender to transfer to L2 Spanish.

Computational models of P600 effects

Although ERPs are a useful in psycholinguistic research, their precise functional interpretation is still unclear (Beres, 2017; Kaan, 2007). Several computational cognitive models have been proposed to account for ERPs (Eddine, Brothers, & Kuperberg, 2022) although only few provide an interpretation of the P600 (Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Fitz & Chang, 2019; Li & Futrell, 2023).

Fitz and Chang (2019) propose that P600 size corresponds to the amount of backpropagated word-prediction error in a recurrent neural network model of word-by-word sentence processing. They used Chang's (2002) Dual-path model to compute backpropagated error on sentences based on stimuli from ERP studies. The simulated P600 effects corresponded to the effect in humans across a wide range of studies, providing support for the hypothesis that ERPs reflect learning signals in the language system. This account of ERPs is known as the Error Propagation account.

The Dual-path model is a connectionist model of sentence production and syntactic development. The model has two pathways. The first pathway is the sequencing system that learns how words are ordered in a sentence and is based on the Simple Recurrent Network (Elman, 1990). The second pathway is a meaning system that learns how to map message content onto words in a target language. The model has also been extended to the bilingual case (Janciauskas & Chang, 2018; Tsoukala, Broersma, Van den Bosch, & Frank, 2021). Verwijmeren, Frank, Fitz, and Khoe (2023) used the Bilingual Dual-path model to simulate ERP responses to syntactic violations in L2 learning. These simulated ERPs depended on L2 proficiency in a manner that resembled human subjects, adding further support to the Error Propagation account.

The present study

We use the Bilingual Dual-Path model to investigate whether the Error Propagation account can explain the P600 results from Tokowicz and MacWhinney (2005). The model simulates native speakers of English (L1) who start learning Spanish (L2) from a later age. At every point in L2 learning, we run an experiment similar to that of Tokowicz and MacWhinney, presenting simulated participants with sentences containing a verb-tense violation, a determiner gender violation, or a determiner number violation, or with a control sentence without any violation.

Based on findings from human ERP studies (Foucart & Frenck-Mestre, 2011, 2012; McLaughlin et al., 2010; Morgan-Short, 2014), we expect a clear P600 effect of violations expressed similarly in L1 and L2 (i.e, verb-tense violations) and a clear P600 effect to grammaticality violations expressed uniquely in L2 (i.e., determiner gender violations). We expect a reduced P600 effect (in line with Sabourin & Stowe, 2008) or even an absent P600 effect (in line with Tokowicz & MacWhinney, 2005) to the determiner number violations compared to the other two violation types. The results from our simulations were largely in line with these expectations, although they did not clearly confirm our expectations for the determiner gender violations. We therefore conduct a second simulated experiment with simulated monolinguals to further explore this discrepancy. Differences between the monolingual and bilingual model predictions suggest the bilingual model does display syntactic transfer from L1 to L2.

Methods

In Experiment 1, we simulate native speakers of English who are learning L2 Spanish. We train instances of the Bilingual

Table 2: Example of an experimental sentence in all for conditions. The bold morphemes indicate the sentence position where the violation occurs.

Example sentence	Violation condition
el padre hacer -a-e una bañera	none (control)
el padre hacer **-ger** una bañera	Tense
los padre **hacer** -a-e una bañera	Number
la **padre** hacer -a-e una bañera	Gender

Dual-path model[1], using a similar model configuration as in Verwijmeren et al. (2023), to learn English from "infancy" and Spanish as L2 at a later stage. The model configuration in this paper differs from the configuration in Verwijmeren et al. (2023) in how to model's next-word prediction is fed back into the model, forming its input signal at the next time step. Following Fitz and Chang (2019) closely, the input of the current model is set to the single highest activation value of the sum of the output vector (i.e., the distribution over possible next words) and the target vector (representing the single target word) . This method emphasizes correct word prediction over actual word prediction.

The model's training input consisted of sentences in artificial versions of Spanish and English that were paired with messages that expressed their meaning. The model learned to express messages as sentences in the target language (Spanish or English) by repeatedly predicting the next word. When presented with a message, corresponding nodes in the model are activated. One of the two target-language nodes is activated, and tense and aspect nodes are activated in the Event-semantics layer. Nodes in the Concept layer are activated for content-words, and a plural node is activated for plurality of a content-word. Corresponding thematic role nodes in the Role layer are activated and fixed connections are formed with the nodes in the Concept layer depending on their thematic role.

After each training epoch, the model is evaluated to measure proficiency, and tested in experimental trials to measure simulated ERPs. For Experiment 2, we trained a monolingual Spanish model. The simulated monolingual participants are trained, evaluated, and tested in the same way as the simulated L2 learners, except that they received only Spanish.

Artificial languages and model training

The artificial languages had the same constructions as the lagnagues created by Verwijmeren et al. (2023). The two artificial languages together consisted of 259 lexical items: 121 nouns, 11 adjectives, 6 pronouns, 6 determiners, 12 prepositions, 87 verbs, 8 auxiliary verbs, 6 verb inflectional morphemes, 1 plural noun marker, and the period. Using the inflectional morphemes, verbs were generated in present or past tense, with simple, progressive or perfect aspect. Plural nouns were generated using the plural noun marker. Plural determiners in Spanish were individual words, namely "los" and "las". For example, the semantic message: AGENT: LADY; ACTION-LINKING: CARVE; PATIENT: CAKE; AGENT-MODIFIER: OLD; TARGET-LANGUAGE: EN would be expressed in English by the sentence: "the old lady carves a cake". The semantic message AGENT: ORANGE, PL; ACTION-LINKING: DISAPPEAR; TARGET-LANGUAGE: ES would be expressed in Spanish by the sentence: "las naranja -s desaparecer -an-en".

We generated 10,000 unique message-sentence pairs for training and a different set of 200 message-sentence pairs for testing, for English and Spanish combined, for each of the 60 simulated L2 learners. The message-sentence pairs are approximately equally divided over the two languages, with the percentage of English sentences being sampled from a uniform distribution between 48% and 52% and the rest in Spanish. Sentence constructions were distributed uniformly in the training input. Following Fitz and Chang (2019), we excluded the message from 70% of the training items. Each model instance iterated five times over its monolingual English training set first, before iterating for 45 epochs over its bilingual training set. The training set's order was randomized at the start of each epoch. The model learned by steepest descent backpropagation, with momentum set to 0.9. The learning rate was first set to 0.1, it then decreased linearly to 0.02 over the 5 epochs of monolingual training, and it stayed at that value during bilingual training. The simulated monolinguals were trained in the same way as the simulated L2 learners, except that that all the message-sentence pairs were in Spanish.

Model evaluation

Linguistic proficiency of the model was tested using the 200-message-sentence-pairs test set after each epoch. Sentences produced by the model were compared to the target sentence. The model's L1 and L2 proficiency was evaluated with two accuracy measures. Following Tsoukala et al. (2021), syntactic accuracy was measured as the percentage of sentences for which all words had the correct part of speech. Meaning accuracy was measured as the percentage of sentences that are syntactically accurate and also correctly conveyed the target message without additions. As pre-registered[2], we only included the 40 simulated participants with the highest meaning accuracy in our analysis.

Differences between simulated participants

Weights are initialized randomly, and differed between simulated participants. The percentage of English versus Spanish (training and testing) sentences varied between simulated participants, ranging from 48/52 to 52/48. The distribution of constructions is the same for all simulated participants. Training, testing and experimental trial sentences in the same language with the same constructions can differ between simulated participants in two ways: singular nouns that are direct

[1] The model code and script for the GAMMs be accessed here: https://osf.io/nbxu6/

[2] The pre-registration can be found here: https://aspredicted.org/HSR_NKN

objects can differ in definiteness of the article, and sentences can differ in content-words resulting in different meaning of sentences. Consequently, a different content-word can result in a different grammatical gender of a noun phrase.

Experimental trials

We generated 30 Spanish control sentences to obtain simulated ERPs on. For each of the control sentences we constructed a version for every violation type (see Table 2). The control sentence was a syntactically correct, active transitive sentence. There were three violation types: (1) Tense violations, where the inflectional marker for singular verbs (-a-e) was changed to progressive verbs (-ger). (2) Determiner number violation, where the singular determiner was changed to a plural determiner. (3) Determiner gender violation, where the determiner's grammatical gender was changed. These three violations involve features that are similar to English, different from English, or unique to Spanish, respectively.

Measuring simulated ERPs

The simulated participants were tested on the experimental sentences after every training epoch. Following Fitz and Chang (2019), learning was turned on in the model while processing the experimental and control sentences, but connection weights were reset to the weights of the respective training epoch after each of those sentences to prevent learning effects during the experiment. Therefore, the simulated participants encountered each trial in the same state for all of the sentences.

We measured prediction error at the hidden layer (see Fitz & Chang, 2019, for details). The prediction error of output unit j is the difference between its activation y_j and the target activation t_j, or: $\delta_j = y_j - t_j$, with $y_j \in [0,1]$ and $t_j \in \{0,1\}$. In the same way as during training, error backpropagated through the network to generate error at deeper layers. Error for units connected to the output layer was calculated as shown in Eq. 1, where k indexes the units connected to the output layer with weight w_{kj}, and j references the units that are backpropagating error.

$$\delta_k = y_k(1-y_k)\sum_{j=1}^{n}\delta_j w_{kj} \quad y_k \in [0,1] \tag{1}$$

Error was also calculated this way for other layers backpropagating error through the network. The simulated P600 sizes are the sums over $|\delta|$ of the recurrent-layer units. The error resulting from a violation was collected at the first position where the sentence becomes ungrammatical (see Table 2). These errors were compared to errors at the same position of control sentences.

Results

Experiment 1: simulated L2 learners

Figure 1 displays the proficiency of the model at the start and the end of bilingual training. The model learns both languages to a high degree, although (unsurprisingly) it remains more proficient in L1 English than L2 Spanish.

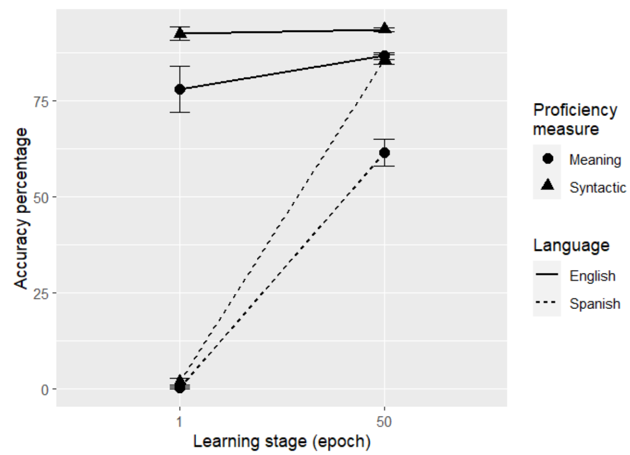

Figure 1: Mean proficiency of the bilingual model. The syntactic and meaning accuracy are displayed for the first and last epoch of bilingual training. The error bars show the 95% confidence interval.

The mean backpropagated error over L2 learning stages at the hidden layer are displayed in Figure 2. As pre-registered, we analyzed the data from our experiment with two generalized additive mixed-effects models (GAMMs; Hastie, 2017), using the bam function from the package mgcv (Wood & Wood, 2015) in R (R Core Team, 2018). Both GAMMs fit the simulated P600 effect, that is, the difference between violation and control sentences in the backpropagated error in the Bilingual Dual-path model. We fit a GAMM to determine if P600 effects differs between violation conditions Similar and Different (i.e., tense and number violations), and we fit a second GAMM to determine if P600 effects differ between conditions Unique and Different (i.e., gender and number violations).

The first GAMM[1] included the predictors of interest: DIFFERENT, LEARNING_STAGE, and their interaction. DIFFERENT indicated violation type and was dummy-coded with levels Similar and Different, coded as 0 and 1 respectively. LEARNING_STAGE is the number of L2 training epochs (standardized). We included by-participant random slopes for NOT_SIMILAR and by-participant random smooths for LEARNING_STAGE. See Table 3 (left-hand side) for a summary of the fitted GAMM. We clearly see predicted P600 effects in the Similar and Different conditions, but it is reduced in the Different compared to the Similar condition, in line with our expectations. The simulated P600 effect grows significantly over LEARNING_STAGE ($F = 33.60$, edf = 8.61, $p < .001$) and this growth differs between the violation types ($F = 2202.45$, edf = 8.39, $p < .001$).

The second GAMM[1] is the same as the first model, except for one predictor of interest, namely DIFFERENT which in this case had the levels Unique and Different, coded as 0 and 1,

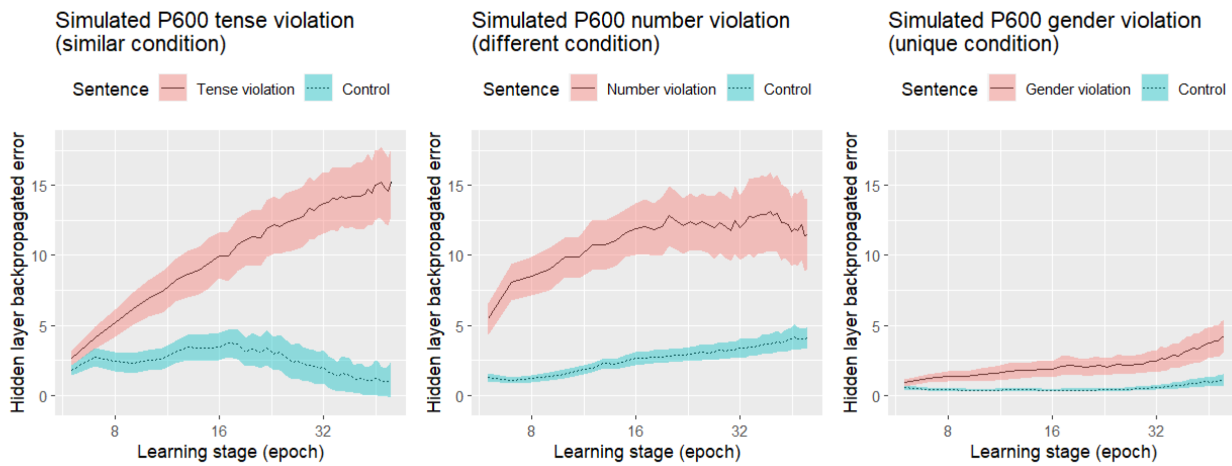

Figure 2: Mean backpropagated error (averaged over all bilingual trained model subjects) as a function of learning stage in the hidden layer, split between the three violation types. Learning stage is log-scaled. Shaded areas represent the 95% CI.

Table 3: Summary of the components in the generalized additive mixed-effects models fit on data from bilingual participants, comparing violation conditions Similar and Different (left; predictor DIFFERENT: Similar = 0, Different = 1) and the conditions Unique and Different (right; predictor DIFFERENT: Unique = 0 and Different = 1).

	Similar vs. Different				Unique vs. Different			
Predictor (coefficient)	Est.	SE	t-value	Pr($>\|t\|$)	Est.	SE	t-value	Pr($>\|t\|$)
(Intercept)	9.12	0.27	33.30	<0.001	5.26	0.267	20.83	<0.001
DIFFERENT	0.70	0.39	1.81	0.07	4.76	0.31	15.27	<0.001
Predictor (smooth)	edf	Ref.df	F-value	Pr($>\|t\|$)	edf	Ref.df	F-value	Pr($>\|t\|$)
s(LEARNING_STAGE)	8.61	8.72	33.60	<0.001	7.44	7.78	8.94	<0.001
s(LEARNING_STAGE:DIFFERENT)	8.39	8.89	2202.45	<0.001	8.79	8.98	334.19	<0.001
s(LEARNING_STAGE, participant)	295.03	359.00	48.34	<0.001	307.02	359.00	2748.53	0.05
s(DIFFERENT, participant)	77.83	78.00	447.96	<0.001	68.57	78.00	283.45	<0.001

to determine if models respond differently between violation conditions Unique (i.e., gender violation) and Different (i.e., number violation). See Table 3 (right-hand side) for a summary of the fitted GAMM. We see a weak simulated P600 effect in the Unique condition, which is smaller than the P600 effect in the Different condition. This is not in line with our expectations. The simulated P600 grows significantly over LEARNING_STAGE ($F = 8.94$, edf = 7.44, $p < .001$) and this growth differs between the violation types ($F = 334.19$, edf = 8.79, $p < .001$).

Experiment 2: simulated monolinguals

Mean Spanish meaning accuracy and mean Spanish syntactic accuracy were 99.98% and 99.99%, respectively, at the end of training.

The mean backpropagated error over learning stages at the hidden layer are displayed in Figure 3.

Similar to our pre-registered analysis, we analyzed the data from our experiment with two GAMMs, to determine if participants respond differently between conditions Similar and Different, and between Unique and Different. Both GAMMs fit the simulated P600 effect from the Bilingual Dual-path model, here trained only on Spanish input. For the GAMM comparing Similar and Different violations, there is a larger simulated P600 effect for the Different condition compared to the Similar condition. This P600 effect significantly grows over LEARNING_STAGE ($F = 1141.37, \text{edf} = 8.61, p < .001$) and this growth differs between the violation types ($F = 488.73, \text{edf} = 8.39, p < .001$). For the GAMM comparing Unique and Different violations, there is a larger simulated P600 effect in the Different condition compared to the Unique condition. In fact, the simulated P600 effect in the Unique condition is very small. The simulated P600 effect over LEARNING_STAGE ($F = 301.10, \text{edf} = 7.44, p < .001$) and this growth differs between the violation types ($F = 1864.80, \text{edf} = 8.79, p < .001$).

Discussion

In the present work, we investigated whether syntactic (dis)similarities between L1 and L2 affect simulated L2 learners in the same way as human L2 learners. We simulated English-Spanish bilinguals and, throughout L2 learning, ex-

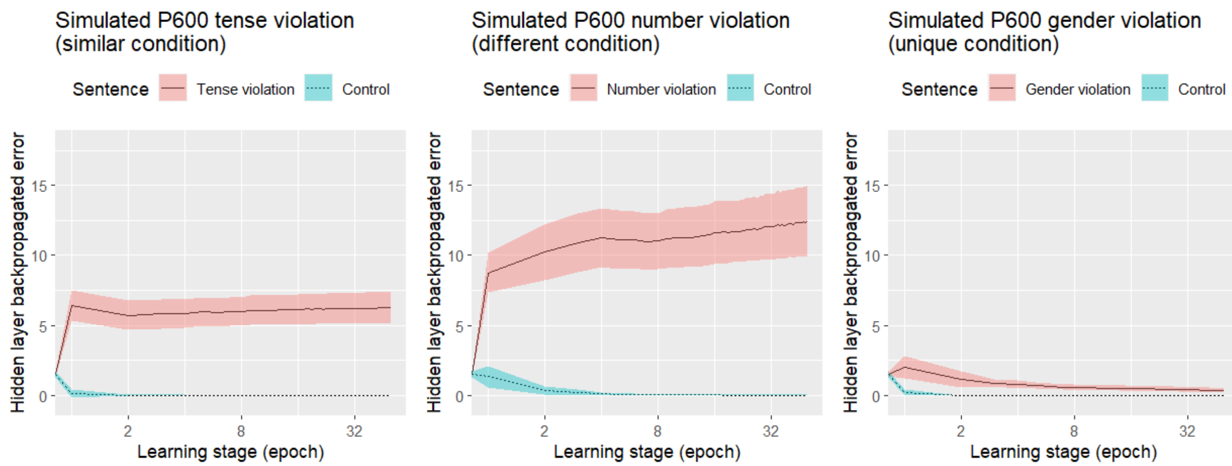

Figure 3: Mean backpropagated error (averaged over all monolingual trained model subjects) as a function of learning stage in the hidden layer, split between the three violation types. Learning stage is log-scaled. Shaded areas represent the 95% CI computed over items.

posed them to three types of syntactic L2 violations that differ in their relation to the L1. We recorded simulated P600s in response to these syntactically anomalous sentences by calculating propagated prediction error at the hidden layer, following the Error Propagation account in Fitz and Chang (2019). On this account, ERPs are summary signals of brain activity that index the propagation of prediction error during comprehension whose functional role is to support learning.

The results of our bilingual simulations are only partially in alignment with our expectations. As expected, our results reveal stronger P600 effects when syntactically anomalous sentences in the L2 contain a tense violation (similar between English and Spanish) compared to a number violation (different between English and Spanish). However, the simulated P600 effect when the L2 sentences contain a gender violation (unique to Spanish) was very weak, especially compared other two types of syntactic violations, in contrast with our expectations.

We did run our model on a simulated L1 control group and found that it predicts a *larger* P600 effect in the number violation condition compared to the tense violation condition. This is the opposite from what was found for the bilingual model's L2 and therefore support the idea that properties from the L1 affect processing in the L2 (i.e., syntactic transfer) in our model, as also appears to happen in humans (De Garavito & White, 2002; Ionin, Zubizarreta, & Philippov, 2009; Montrul, 2010; White, Valenzuela, Kozlowska-Macgregor, & Leung, 2004).

Moreover, the monolingual model showed an even smaller P600 effect in the gender violation compared to the bilingual model; an effect that reduced over L1 training whereas it increased over L2 training. Thus, it appears there is also syntactic transfer from L1 to L2 going on in the processing of gender violations.

It is not entirely clear why backpropagated error is low in response to a gender violation but not in response to a number violation. A possible explanation is the implementation of syntactic features in the model. The messages that accompany sentences during training encode tense as well as plurality of nouns, but not gender. Grammatical gender is present and expressed in our artificial language of Spanish, but there is no representation of gender in the concept layer of the model. Specifically, there is no gender node in the concept layer preceding the hidden layer, to backpropagate error to. Furthermore, verb conjugation indicating tense, as well as plurality of nouns, are expressed by morphemes that follow verbs or nouns, respectively. The model treats these morphemes as words. We have no such morphemes for gender, only separate gendered determiners for Spanish.

Conclusion

The error propagation account explained key findings from a considerable number of monolingual ERP studies (Fitz & Chang, 2019). Previous work on simulating bilingual ERPs and how they change over development (Verwijmeren et al., 2023) added further support to his account. In our present work, the reduced P600 for the number compared to tense violation supports a theory of syntactic transfer affecting ERP effects in L2 learners. The model in its present state, however, was unable to produce a strong P600 in response to a grammatical gender violation, in contrast with human participants (Antonicelli & Rastelli, 2022; Caffarra et al., 2015; McLaughlin et al., 2010; Foucart & Frenck-Mestre, 2011; Frenck-Mestre, Foucart, Carrasco-Ortiz, & Herschensohn, 2009; Morgan-Short, 2014; Tokowicz & MacWhinney, 2005). Further work is needed to determine if the Error Propagation account, as implemented in the Bilingual Dual-path model, simulates a strong P600 effect in response to a grammatical gender violation when gender is implemented in the message in the same way as plurality and tense.

References

Antonicelli, G., & Rastelli, S. (2022). Event-related potentials in the study of L2 sentence processing: A scoping review of the decade 2010-2020. *Language Acquisition*, 1–38.

Beres, A. M. (2017). Time is of the essence: A review of electroencephalography (EEG) and event-related brain potentials (ERPs) in language research. *Applied Psychophysiology and Biofeedback*, *42*, 247–255.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. (2017). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, *41*, 1318–1352.

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience & Biobehavioral Reviews*, *51*, 31–47.

Chang, F. (2002). Symbolically speaking: A connectionist model of sentence production. *Cognitive Science*, *26*(5), 609–651.

De Garavito, J. B., & White, L. (2002). The second language acquisition of Spanish DPs: The status of grammatical features. In *The acquisition of Spanish morphosyntax: The L1/L2 connection* (pp. 153–178). Springer.

Eddine, S. N., Brothers, T., & Kuperberg, G. R. (2022). The N400 in silico: A review of computational models. *The Psychology of Learning and Motivation*, 123.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*(2), 179–211.

Fitz, H., & Chang, F. (2019). Language ERPs reflect learning through prediction error propagation. *Cognitive Psychology*, *111*, 15–52.

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*(3), 379–399.

Foucart, A., & Frenck-Mestre, C. (2012). Can late L2 learners acquire new grammatical features? Evidence from ERPs and eye-tracking. *Journal of Memory and Language*, *66*(1), 226–248.

Frenck-Mestre, C., Foucart, A., Carrasco-Ortiz, H., & Herschensohn, J. (2009). Processing of grammatical gender in French as a first and second language: Evidence from ERPs. *EuroSLA Yearbook*, *9*(1), 76–106.

Hastie, T. J. (2017). Generalized additive models. In *Statistical models in S* (pp. 249–307). Routledge.

Ionin, T., Zubizarreta, M. L., & Philippov, V. (2009). Acquisition of article semantics by child and adult L2-English learners. *Bilingualism: Language and Cognition*, *12*(3), 337–361.

Janciauskas, M., & Chang, F. (2018). Input and age-dependent variation in second language learning: A connectionist account. *Cognitive Science*, *42*, 519–554.

Kaan, E. (2007). Event-related potentials and language processing: A brief overview. *Language and Linguistics Compass*, *1*(6), 571–591.

Li, J., & Futrell, R. (2023). A decomposition of surprisal tracks the N400 and P600 brain potentials. In *Proceedings of the 45th Annual Meeting of the Cognitive Science Society.*

Liu, H., Dunlap, S., Tang, Y., Lu, Y., & Chen, B. (2017). The modulatory role of L1 and L2 morphosyntactic similarity during production of L2 inflected words: An ERP study. *Journal of Neurolinguistics*, *42*, 109–123.

McLaughlin, J., Tanner, D., Pitkänen, I., Frenck-Mestre, C., Inoue, K., Valentine, G., & Osterhout, L. (2010). Brain potentials reveal discrete stages of L2 grammatical learning. *Language Learning*, *60*, 123–150.

Montrul, S. (2010). Dominant language transfer in adult second language learners and heritage speakers. *Second Language Research*, *26*(3), 293–327.

Morgan-Short, K. (2014). Electrophysiological approaches to understanding second language acquisition: A field reaching its potential. *Annual Review of Applied Linguistics*, *34*, 15–36.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*(6), 739–773.

R Core Team. (2018). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–430.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, *27*(2), 173–204.

Tsoukala, C., Broersma, M., Van den Bosch, A., & Frank, S. L. (2021). Simulating code-switching using a neural network model of bilingual sentence production. *Computational Brain & Behavior*, *4*(1), 87–100.

Verwijmeren, S., Frank, S. L., Fitz, H., & Khoe, Y. H. (2023). A neural network simulation of event-related potentials in response to syntactic violations in second-language learning. In *Proceedings of the 21st International Conference on Cognitive Modelling.*

White, L., Valenzuela, E., Kozlowska-Macgregor, M., & Leung, Y.-K. I. (2004). Gender and number agreement in nonnative spanish. *Applied Psycholinguistics*, *25*(1), 105–133.

Wood, S., & Wood, M. S. (2015). Package 'mgcv'. *R package version*, *1*(29), 729.