Cognition 164 (2017) 61–73

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Hierarchical levels of representation in language prediction: The influence of first language acquisition in highly proficient bilinguals

Nicola Molinaro [a,b,*], Francesco Giannelli [c], Sendy Caffarra [a], Clara Martin [a,b]

[a] BCBL, Basque Center on Cognition, Brain and Language, Donostia-San Sebastian, Spain
[b] Ikerbasque, Basque Foundation for Science, Bilbao, Spain
[c] Dipartimento di Psicologia, Universitá di Milano-Bicocca, Milano, Italy

A R T I C L E I N F O

Article history:
Received 31 March 2016
Revised 22 November 2016
Accepted 24 March 2017
Available online 3 April 2017

Keywords:
Prediction
Multilingualism
N200
Beta-band activity
Reading

A B S T R A C T

Language comprehension is largely supported by predictive mechanisms that account for the ease and speed with which communication unfolds. Both native and proficient non-native speakers can efficiently handle contextual cues to generate reliable linguistic expectations. However, the link between the variability of the linguistic background of the speaker and the hierarchical format of the representations predicted is still not clear. We here investigate whether native language exposure to typologically highly diverse languages (Spanish and Basque) affects the way early balanced bilingual speakers carry out language predictions. During Spanish sentence comprehension, participants developed predictions of words the form of which (noun ending) could be either diagnostic of grammatical gender values (transparent) or totally ambiguous (opaque). We measured electrophysiological prediction effects time-locked both to the target word and to its determiner, with the former being expected or unexpected. Event-related (N200–N400) and oscillatory activity in the low beta-band (15–17 Hz) frequency channel showed that both Spanish and Basque natives optimally carry out lexical predictions independently of word transparency. Crucially, in contrast to Spanish natives, Basque natives displayed visual word form predictions for transparent words, in consistency with the relevance that noun endings (post-nominal suffixes) play in their native language. We conclude that early language exposure largely shapes prediction mechanisms, so that bilinguals reading in their second language rely on the distributional regularities that are highly relevant in their first language. More importantly, we show that individual linguistic experience hierarchically modulates the format of the predicted representation.

1. Introduction

The neural system strongly relies on feedback-recurrent neural projections to interact with evidence from the environment (Arnal, Doelling, & Poeppel, 2015; Bastos et al., 2015; Michalareas et al., 2016). This supports the idea that much of our brain activity focuses on the prediction of upcoming sensory information rather than its passive integration, and that the system mainly encodes only the unpredicted portion of the sensory signal (Bar, 2007; Clark, 2013; Friston, 2005). Predictive processes thus are a valuable resource contributing to the ease and speed with which language comprehension incrementally builds upon contextual information and internal knowledge (Federmeier, 2007; Levy, 2008; Pickering & Garrod, 2013). Some authors consider prediction so fundamental that it has been suggested that learning to speak in infants arises directly from learning to predict (Mani & Huettig, 2012). Even so, the role of internal linguistic knowledge and more specifically of early language exposure in modulating prediction across different speakers has not been adequately studied. In the present study we focus on this topic, evaluating how multilingual experience (and more specifically native language knowledge) affects language prediction (e.g., Chang, Dell, & Bock, 2006).

Research on language prediction in bilinguals has provided clear evidence that low proficiency second language (L2) speakers who cannot rely on life-long experience consequently have not developed robust prediction processes, as native (L1) speakers do (for a review, Kaan, 2014). Factors such as reduced proficiency and reduced experience make language prediction in L2 "weaker" than in L1. Nonetheless, if proficiency levels are balanced, prediction in L2 should be similar to prediction in L1 (e.g., Hopp, 2013). In the present study we test this hypothesis, focusing on the prediction abilities of balanced Basque-Spanish speakers who have

---

* Corresponding author at: BCBL, Basque center on Cognition, Brain and Language, Paseo Mikeletegi, 69, 2°, 20009 San Sebastian/Donostia, Spain

E-mail address: n.molinaro@bcbl.eu (N. Molinaro).

http://dx.doi.org/10.1016/j.cognition.2017.03.012

been primarily exposed (before age of 3) to one of the two languages, but are fluently proficient in both. Focusing on these populations, we study how the language background provided by the early language exposure affects language prediction. We show that when L2 proficiency is high, prediction in L2 is not necessarily equivalent to prediction in L1, since it is sensitive to the typological characteristics of the native language of each speaker. This can trigger anticipations of even more specific linguistic representations for non-native compared to native speakers.

1.1. Bilingual prediction

There are multiple factors which affect prediction in a second language (Kaan, 2014). The role of proficiency has been highlighted in a series of studies employing the visual word eye-tracking paradigm (see Altmann & Kamide, 1999, for evidence of prediction in natives). Mitsugi and MacWhinney (2015) reported that non-native Japanese learners did not show predictive saccades during speech listening, as native speakers do. On the other hand, Hopp (2013) reported that English learners of German show such predictive looks if they are proficient enough in German. Complementary evidence comes from the ERP (Event-Related Potentials) sentence comprehension literature. In a recent study, Martin et al. (2013) studied language prediction in L2 similarly to DeLong, Urbach, and Kutas (2005) who provided evidence of prediction in natives. In the former study, while reading sentences such as *The day was breezy so the boy went outside to fly*... L2 late English speakers did not show a lexical prediction effect (i.e., a modulation of the N400 ERP component, assumed to reflect lexical/semantic processing) for determiners (*a*) that match a following highly expected noun (*kite*) compared to determiners introducing a low expected noun (*an* introducing *airplane*).

While proficiency appears to be a critical factor modulating prediction, interaction of the predicted representations with the ones available in the native language has also been shown to be relevant. Foucart, Martin, Moreno, and Costa (2014) observed that both early Spanish-Catalan and late French-Spanish bilinguals showed prediction effects during sentence reading, as did Spanish monolinguals (see Wicha, Moreno, & Kutas, 2004, for evidence of prediction in natives). The authors recorded ERPs time-locked to a gender-marked determiner preceding a highly expected noun in sentences such as *El pirata tenía el mapa pero no encontró*... ("The pirate had a map but did not find..."). The following determiner could either gender-match with the predicted Spanish noun (such as the grammatically masculine determiner *el* preceding *tesoro* – "the treasure"), or it could gender-mismatch with the predicted noun (the feminine *la*, introducing a non-anomalous noun such as *gruta* – "cave"). This study thus highlights that L1–L2 similarity (both French and Catalan have grammatical gender, as does Spanish) boost language prediction, independently of language proficiency (early or late L2 speaker).

Individual differences (Dussias & Pinar, 2010; Kaan, 2014) and task-related processing strategies (Ferreira, Foucart, & Engelhardt, 2013; see also Clahsen & Felser, 2006) are additional elements shown to modulate language prediction. An important point raised by Kaan (2014) is that differences in prediction between native and second language speakers are mainly due to the same factors that account for individual differences in natives. We here provide a step further in the research on bilingual prediction. Until now, most studies have focused on the presence/absence of prediction effects in late/low proficiency bilinguals. We here accept Kaan's position, that prediction is possible in L2 providing proficiency levels are high enough. One unsolved issue in such a scenario, however, is what kind of prediction bilinguals develop: Is prediction fully tuned to the properties of the L2 or can we find traces of influence from the native language? In other words, is prediction mainly a question of proficiency, or is there any influence of the language background even in highly proficient early L2 readers?

Studies on highly proficient bilinguals so far have not answered this question. In Foucart et al. (2014) there was no reason for Catalan-Spanish bilinguals to show differences from the native Spanish speakers, since Catalan and Spanish are typologically highly similar and share an overlapping gender system. Hopp (2013) did report similar prediction effects for English-German bilinguals and German natives, but that experimental paradigm was not designed to highlight differences of prediction in the two groups. Here, we focus on balanced Basque-Spanish bilinguals who are Basque natives (compared to Spanish-Basque bilinguals), since the large typological difference between the two languages could affect the way prediction processes are at work in the two groups while they process Spanish sentences.

This study on prediction in bilinguals can inform research on language prediction, since it focuses on the link between the language background of a speaker and the way prediction abilities develop during comprehension. In addition, based on the observation that prediction has considerable advantages for learning (Rescorla & Wagner, 1972; Schultz, Dayan, & Montague, 1997; see also Kuperberg & Jaeger, 2016), it is relevant for language learning research to evaluate how multilingual experience modulates language prediction.

1.2. The present study

In the present study, Basque (L1)-Spanish (L2) and Spanish (L1)-Basque (L2) very early bilinguals read Spanish sentences word by word for comprehension. We tested participants who were highly proficient in Spanish but were primarily exposed either to Spanish or to Basque before the age of 3. We avoided the comparison between monolinguals and bilinguals since this latter group has a huge amount of competing linguistic information (another language) that can alter the prediction processing dynamics as compared to monolinguals. The present design, comparison between two groups of early balanced bilinguals, resolves this confound.

The interaction between Spanish and Basque was considered as highly informative to address our research question. Spanish and Basque are two rich-morphology languages that are, however, typologically very different on a large number of dimensions (mainly lexical and syntactic). A relevant difference is the way in which these two languages instantiate the relation between content and function words. As an example, determiners (articles, quantifiers and prepositions) precede their nouns in Spanish (*la mesa*, "the table"), while in Basque these function words are consistently implemented as post-nominal bound suffixes (*mahai-a*, "the table"; de Rijk & de Coene, 2008; for corpus evidence, Gervain et al., 2013). These functional elements are relevant, since they are prominent cues for speech segmentation and signal syntactic boundaries within a sentence. This difference makes Basque speakers focus more on the morphological structure of nouns, given their syntactic diagnosticity, and more specifically on noun endings (both infants, Molnar, Lallier, & Carreiras, 2014, and adults, Gervain et al., 2013).

Based on this typological distinction, the present study capitalized on bilingual sensitivity to the "unsystematic" distributional properties of grammatical gender, a feature that is present in Spanish but not in Basque. In Spanish inanimate nouns, grammatical gender is an arbitrary feature (either masculine or feminine) that is uniquely assigned to individual lexical items. This feature is informative of structural relations such as the one between a determiner and its head noun (see Foucart et al., 2014, study described above) and it has been used to study lexical prediction during sentence processing (Wicha et al., 2004). Intriguingly, noun ending information in Spanish is diagnostic of grammatical gender in only two thirds of cases (–*a* for feminine and –*o* for masculine nouns: cue availability, Harris, 1991). However, since there are plenty of irregularities (∼1/3 of the nouns are gender opaque – i.e., *flor*, "flower", is feminine - or gender irregular - i.e., *mano*, "hand", is feminine), it has been suggested that proficient Spanish speakers do not rely on formal cues (i.e., the –*a*/–*o* noun ending alternation) to compute agreement dependencies involving grammatical gender, but rely on lexical cues (i.e., the gender value that is lexically associated to each individual noun; Caffarra & Barber, 2015; Caffarra, Janssen, & Barber, 2014; Molinaro, Barber, Pérez, Parkkonen, & Carreiras, 2013; see Gollan & Frost, 2001, for a dual route proposal).

The present study tested the prediction of gender-transparent (e.g., *mes-a*, "table", feminine noun) and gender-opaque nouns (e.g., *flor*, "flower", feminine) by focusing on the processing of the preceding article in Spanish. The preceding article did not differ in the two cases (e.g., *la*, "the", feminine) so that possible differences between the two conditions would mainly be due to the prediction of the following noun.

Based on previous studies, balanced Basque (L1)-Spanish (L2) and Spanish (L1)-Basque (L2) bilinguals should show prediction effects, independently of their initial language exposure, since both groups tested are highly proficient in Spanish and, more specifically, in grammatical gender processing in Spanish (see Section 2). This hypothesis is supported by available studies that report similar prediction effects for highly proficient bilinguals and native speakers (Foucart et al., 2014; Hopp, 2013) and proposals stating that prediction in L2 for highly proficient speakers should be similar to L1 (Kaan, 2014).

Crucially, the Basque language does not have grammatical gender and its morphological regularities (specifically, post-nominal suffixes) are highly diagnostic of the underlying linguistic structure (de Rijk & de Coene, 2008; Laka, 1996). Spanish grammatical gender (with its large amount of irregularities, ∼1/3) provides an interesting test case to evaluate whether the native knowledge of Basque differentially affects the specificity of the predicted representation for transparent and opaque words. It is possible that Basque natives show more sensitivity to transparent gender cues (noun endings) compared to Spanish natives, since the distribution of nominal terminations is statistically relevant in their native language.

In the present experiment we did not expect prediction differences between transparent and opaque words for Spanish natives. Since they extract grammatical gender information from lexical representations independently of the transparency of the noun, they were expected to develop similar lexical predictions in the two cases. For Basque natives it could be hypothesized that prediction effects would also be similar for transparent and opaque words. Overall, the effects could be weaker as compared to Spanish natives, since their first language does not have grammatical gender (differently from the study by Foucart et al. (2014), in which prediction effects emerged independently of L2 proficiency, but where L1 and L2 were typologically very similar). Since the level of proficiency in Spanish was controlled across groups, this evidence would support the hypothesis that Basque natives predict in the same way as Spanish natives: they mainly rely on the lexical gender of the target nouns (for morphosyntactic integration effects see Caffarra & Barber, 2015; Caffarra et al., 2014). This scenario would support the claim that prediction is independently tuned to the distributional properties of each known language. On the contrary, however, Basque natives could show differences in the prediction of transparent and opaque words. It is possible that Basque natives rely more on word form properties (noun endings) in predicting gender-transparent words. They would thus predict to a larger extent in the case of transparent words compared to opaque words (but still show prediction effects for both gender categories). This would support the hypothesis that the native language properties interact with prediction in L2 even for highly proficient bilingual speakers.

We measured prediction taking advantage of electrophysiology (EEG), since this provides the necessary high-temporal resolution to detect pre-target noun effects time-locked to the preceding determiner. Similar to previous studies (discussed above) we recorded the ERPs time-locked to the target expected determiner as compared to an unexpected determiner (with opposite gender) that introduced a non-anomalous unexpected noun (*En el mapa que tenían los piratas la cruz indicaba donde estaba el tesoro secreto/la perla mágica*, "In the map that the pirates had, the cross indicated where the secret treasure/the magic pearl was."). In line with previous studies, we expected a larger N400 effect for determiners whose gender does not agree with the value of the expected target noun (but see Molinaro, Vespignani, Canal, Fonda, & Cacciari, 2008, for an alternative functional interpretation of this effect). Also, we explored ERPs on the critical noun (expected vs. unexpected) to explore the consequences of prediction (or absence of prediction, or deceived prediction) on word integration (Molinaro, Conrad, Barber, & Carreiras, 2010; in L2 literature see Martin et al., 2013).

In addition to previous studies, we also estimated the oscillatory activity time-locked to the target determiner. Differently from ERPs, time–frequency estimation provides complementary evidence about neural activity that is not phase-locked to a target event but presents a relative amount of jittering (variability in its time-course across trials). Even more important, from a theoretical point of view, there is mounting evidence that oscillatory activity in the beta band (13–30 Hz) plays a relevant role in predictive processing. Wang (2010) initially proposed that feedforward visual processing is mediated by feedback-recurrent connection sending top-down information through a beta-band channel. Bastos et al. (2015) (Michalareas et al., 2016, for MEG evidence from humans) analyzed the oscillatory dynamics of the primate visual system employing electrocorticography recordings from grids implanted throughout the whole visual system of monkeys. They reported that the beta band activity was associated with feedback influences from higher processing regions to primary visual regions during pre-stimulus visual processing. Similar proposals have been advanced even in the sentence comprehension domain (Lewis & Bastiaansen, 2015; Molinaro, Monsalve, & Lizarazu, 2016). Consequently, we considered it relevant to estimate the beta band components time-locked to the target (unexpected vs. expected) determiner to quantify the strength of the prediction in our experimental design across groups. Evidence of beta-band modulations in our reading design would indicate that more detailed predictions are at work for a specific group/condition (possibly at the visual word form level, based on the timing of the effects).

To sum up, in the present sentence-reading study we estimated prediction of gender marked nouns whose ending was gender informative (transparent nouns) or not (opaque nouns). Prediction effects were recorded time-locked to the previous gender-marker determiner that could be either gender-consistent or not with the expected noun. Balanced Basque-Spanish bilinguals who were either Basque or Spanish natives took part in the study, our aim being to evaluate how the native language background affects prediction. Similar prediction effects across differently transparent items and groups of balanced bilinguals would provide evidence for the hypothesis that prediction in a second language is just a matter of proficiency; differential prediction effects, depending on transparency, for the two groups of bilinguals would support the hypothesis that prediction is tuned to the distributional regularities of the native language even in fluently proficient second language speakers.

Table 1
General proficiency assessment of the participants in the two groups.

Measure		Spanish natives (N = 24)	Basque natives (N = 24)	
Self-evaluation	Span	9.79 (0.41)	9.50 (0.97)	*n.s.*
(0–10)	Basq	8.67 (0.96)	9.67 (0.56)	$p < 0.01$
LexTALE	Span. (0–60)	55.08 (3.21)	54.47 (3.95)	*n.s.*
	Basq. (0–50)	38.03 (4.54)	45.00 (3.40)	$p < 0.05$
Picture naming	Span.	64.67 (0.70)	64.17 (1.09)	*n.s.*
(0–65)	Basq.	54.92 (3.27)	64.54 (0.78)	$p < 0.05$
Interview	Span.	4.73 (0.38)	4.70 (0.46)	*n.s.*
(0–5)	Basq.	4.64 (0.60)	4.85 (0.28)	*n.s.*

2. Materials and methods

2.1. Participants

Forty-eight early bilingual speakers took part in the experiment. They were divided in two groups. Twenty-four native speakers of Basque (14 females; age range 18–35, mean: 25, SD: 5.10; Age of acquisition of Spanish: 3.75 y, SD: 1.36) formed the first group. They were first exposed to Spanish after the age of 3 and interacted in Basque with both parents. Twenty-four native speakers of Spanish (19 females) formed the second group (age range: 19–41, mean: 24, SD: 4.54; Age of acquisition of Basque: 4.04 y, SD: 1.57); they started to learn Basque after the age of 3 and interacted in Spanish with both parents. Participants received a payment of 10€ per hour for their collaboration. All subjects were right handed and their vision was normal or corrected to normal. All participants signed an informed consent form before taking part in the study that was approved by the BCBL ethics committee.

2.1.1. General proficiency assessment

In order to participate in the experiment, all participants went through a proficiency evaluation in Spanish and Basque (results in Table 1). On a self-rating of their comprehension levels (10-points scale: 0, unintelligible; 10, native-like) they rated themselves very high in both. Importantly, an analyses of variance considering the factors Language (Basque vs. Spanish language) and Group (Basque vs. Spanish natives) revealed a robust interaction [$F(1, 46) = 14.10$, $p < 0.001$, $\eta^2_G = 0.23$]: for Spanish comprehension there was no difference between groups, while there was a difference for Basque comprehension. We then tested the vocabulary size of our participants in a lexical decision task (no time constraint) in Spanish and Basque (for details of the Spanish version: LexTALE, Izura, Cuetos, & Brysbaert, 2014; Lemhofer & Broersma, 2012). Both groups showed native-range scores for Spanish, and high proficiency scores for Basque. The interaction between Language and Group [$F(1, 46) = 14.34$, $p < 0.001$, $\eta^2_G = 0.29$] was due to the fact that we observed no difference in Spanish level but difference in Basque level (see Table 1). Finally, participants had to name a set of pictures of increasing difficulty in Spanish and Basque. Also for this task an interaction between Language and Group emerged [$F(1, 46) = 20.92$, $p < 0.001$, $\eta^2_G = 0.31$]: participants had native-like scores in Spanish; in Basque they also had very high scores that differed between groups. After the proficiency test, all participants were rated (based on an interview on a 0-to-5 point scale) as fluently proficient in both languages by the experimenters (who were balanced bilinguals). No participant had a score below 4 in either Basque or Spanish, and there was no group difference. Overall, there was no group difference in the general proficiency assessment of Spanish. It should be noted that all the participants also knew English as a third language. This additional language is not relevant in the present design, since participants were largely more proficient in the other two languages. Their proficiency in English did not differ between groups; proficiency was overall rated as good (LexTALE, score 0–40: 23.53, SD: 5.05; picture naming: 46.06, SD: 8.03; final interview: 3.29, SD: 1.31), but still lower than Spanish and Basque.

2.1.2. Grammatical gender proficiency assessment

After the EEG experimental session, we further tested individual proficiency in processing the grammatical gender of the target Spanish nouns employed in the sentence comprehension task. In two complementary tasks we recorded accuracy and response times (reported in Fig. 1). In a gender decision task the participants had to identify the gender of the target items as soon and as correctly as possible. Participants were visually presented with isolated words (120 transparent and 120 opaque; same words used in the EEG experiment). Response hand was counterbalanced across participants. They had to press left (or right) for feminine words and right (or left) for masculine words. We analyzed the data with a two-way ANOVA considering the within factor Transparency (transparent, opaque) and the between factor Group (Spanish and Basque natives). Both accuracy [$F(1, 46) = 12.83$, $p < 0.001$, $\eta^2_G = 0.08$] and response times [$F(1, 46) = 13.46$, $p < 0.001$, $\eta^2_G = 0.08$] showed a main effect of transparency. In both groups opaque words were more difficult than transparent words, but there was no interaction involving Group (Fig. 1).

Since we studied language prediction focusing on a determiner-noun gender agreement relation, we also evaluated gender agreement in a grammaticality judgement task. Participants were presented with 240 determiner-noun phrases (60 opaque-congruent; 60 opaque-incongruent; 60 transparent-congruent; 60 transparent-incongruent; same words used in the EEG experiment). They had to decide if the determiner-noun phrase was grammatically correct or not by pressing the left or right key of the keyboard (counterbalanced across participants). A two-way ANOVA (Transparency and Group) showed no significant effects in the accuracy. Response times were slower for the opaque items as showed by the main Transparency effect [$F(1, 46) = 5.9$, $p < 0.05$, $\eta^2_G = 0.01$]. In addition, a Group effect emerged [$F(1, 46) = 4.41$, $p < 0.05$, $\eta^2_G = 0.09$], indicating that Basque natives were slower in their judgements (Fig. 1). The two factors did not interact.

Overall, these two tasks indicate that both groups handle grammatical gender similarly, being sensitive to the transparency factor. A main effect of group was observed in the grammaticality task (but not in the accuracy) indicating that Spanish grammar was more complex for Basque natives. However, the lack of interaction with Transparency (that showed reliable effects in all tasks) does not suggest differential processing of the two types of gender at the syntactic level between the two groups.

2.2. Materials

Two lists of NPs were created. In the first one, 120 transparent nouns were selected, where 60 were masculine and 60 were

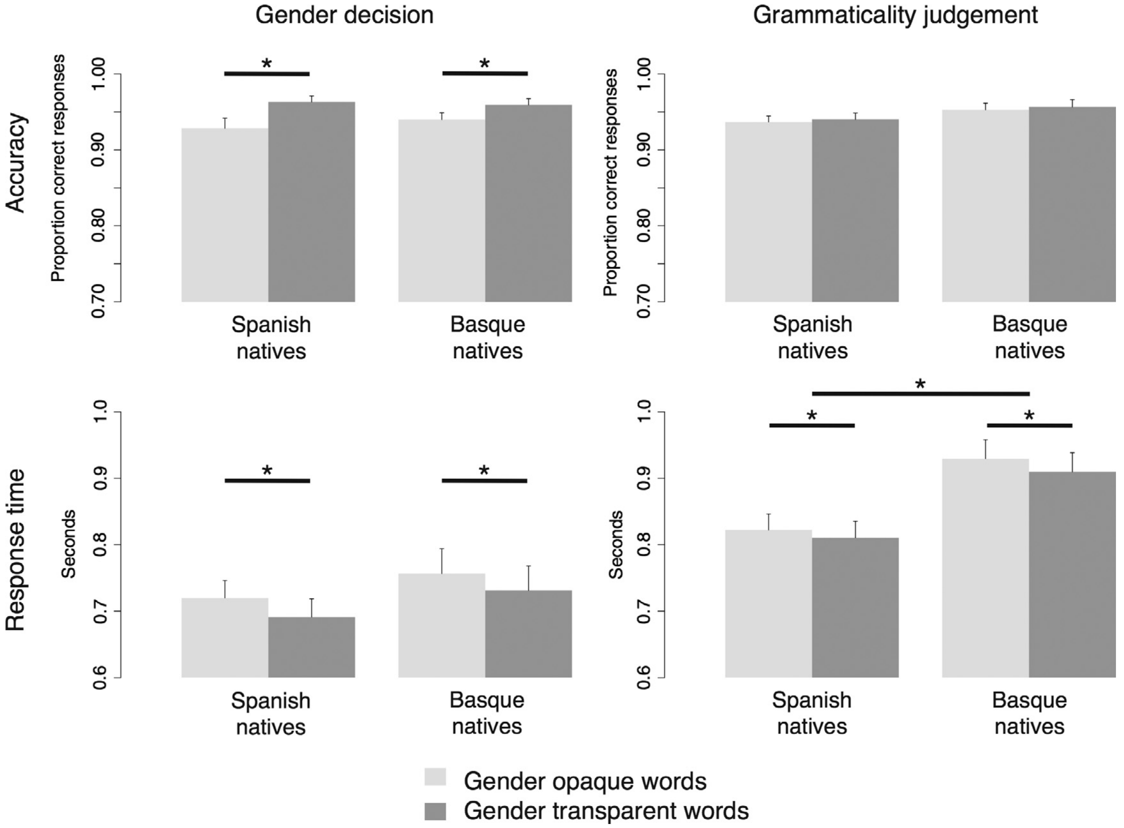

Fig. 1. Behavioural results of the grammatical gender proficiency assessment. We here report both the accuracy (proportion of correct responses) and the response times (in ms) for the Gender Decision and the Grammaticality Judgement task in the two experimental groups. Asterisks and horizontal lines indicate the statistically significant differences.

feminine, with the masculine nouns ending in "–*o*", and the feminine nouns ending in "–*a*". The second list had 120 opaque nouns that could have different endings. Here also, half of the nouns were masculine and half were feminine, but the word ending was not informative of the gender value. All nouns referred to inanimate entities. It can be noticed that the equal proportion of gender transparent/opaque items in the present design does not mirror the real distribution of these items in Spanish (majority of transparent items). However, Gollan and Frost (2001; Experiment 1A, 1B) showed that transparency effects emerge independently of the proportion of gender transparent/opaque items in the experimental materials.

An article ("*el*", "*la*", the; "*un*", "*una*", a) preceded the nouns in both lists. These NPs were employed to construct the sentence stimuli. A hundred and twenty sentence contexts were highly constraining towards a NP (expected condition). The unexpected condition was created substituting the target noun phrase with a different one of opposite gender (but same transparency), resulting in a total of 240 experimental sentences. All the sentences were semantically correct and the target nouns were never in sentence final position. Across sentences the target word (determiner) was on average in position 13.22.

The mean cloze probability of expected and unexpected words was assessed by Basque-Spanish bilinguals (N = 20) who did not take part in the experiment. They had to read sentence contexts and continued them with the very first continuation that came up to their mind. The sentences stopped before the article that should have preceded the word, so that participants were free to use or not the article before the noun. The mean cloze probability for expected nouns and for expected whole NPs was respectively 0.87 (SD: 0.08), and 0.84 (SD: 0.10); the mean cloze probability for unexpected words and unexpected NPs was 0.02 (SD: 0.03), and 0. No cloze-probability differences were observed between sentences preceding opaque and transparent items (all $p > 0.4$).

The 240 sentences were divided in two lists. Each list had 120 sentence contexts followed by 30 transparent expected NPs, 30 transparent unexpected NPs, 30 opaque expected NPs, and 30 opaque unexpected NPs. Each sentence context, as well as each NP, could appear only once in each list in order to avoid repeated presentations, but each sentence context appeared in both lists. We balanced the target words across conditions and lists employing independent ANOVAs (two-way: Expectedness by Transparency). Within each list the target words were balanced (all $p > 0.2$; based on EsPal, Duchon, Perea, Sebastián-Gallés, Martí, & Carreiras, 2013) for grammatical gender, word frequency (log-values: List1: 1.52, SD: 0.56; List 2: 1.52, SD: 0.59), number of letters (List1: 6.09, SD: 1.91; List 2: 5.93, SD: 1.89) and number of neighbors (List1: 6.71, SD: 7.27; List 2: 7.17, SD: 7.55), average position of the determiner (13.22 in both Lists). No differences emerged between lists. Examples of sentences used as experimental material, with the expected vs. the unexpected NP, are reported below:

Transparent item example:

Expec. Acabo de salir de casa y no recuerdo si | he | cerrado | ***la***$_{\mathbf{[FEM]}}$ | ***puert-a***$_{\mathbf{[FEM]}}$ | cuando | me | he | ido.
Unexp. Acabo de salir de casa y no recuerdo si | he | cerrado | ***el***$_{\mathbf{[MAS]}}$ | ***armari-o***$_{\mathbf{[MAS]}}$ | cuando | me | he | ido.
[I just left home and I don't remember if I closed the$_{\text{[FEM]}}$ *door*$_{\text{[FEM]}}$ *(Expec.) / the*$_{\text{[MAS]}}$ *closet*$_{\text{[MAS]}}$ *(Unexp.) when I left]*

Opaque item example:

Expec. Prefiero que el te esté muy dulce, | puedes | pasarme | ***el*$_{\textbf{[MAS]}}$** | ***azúcar*$_{\textbf{[MAS]}}$** | por | favor?
Unexp. Prefiero que el te esté muy dulce, | puedes | pasarme | ***la*$_{\textbf{[FEM]}}$** | ***miel*$_{\textbf{[FEM]}}$** | por | favor?
[I prefer the tea very sweet, could you please pass me the$_{\text{[MAS]}}$ sugar$_{\text{[MAS]}}$ (Expec.) / the$_{\text{[FEM]}}$ honey$_{\text{[FEM]}}$ (Unexp.)?]

2.3. Procedure

The EEG experiment was run in a soundproof electrically shielded chamber. Participants were seated in a chair about sixty centimeters in front of a computer screen. Stimuli were delivered with the Presentation software (https://www.neurobs.com/). Participants read sentences displayed in white letters on a grey background. We divided the sentences in two parts, so that the second part was equally long across conditions (6–8 words). The target determiner was in variable sentence positions in the second part of the sentence. After a fixation cross (500 ms), the first part of the sentence was presented as a whole on the screen (average length: 8.2 words, no difference between conditions) for participants to read. After button press, the second part of the sentence was presented word by word (200 ms + 500 ms inter-stimulus blank interval) until the end. In order to make sure that participants were paying attention to the sentence content, a yes–no comprehension question followed one fourth of the trials: they could answer using the corresponding yes–no buttons on the computer keyboard; response hand was counterbalanced across participants and lists. A brief practice session included three sentences and the relative yes–no questions. Participants were asked to stay still and to try to reduce blinking and eyes movement to a minimum, especially during the word-by-word presentations. Stimuli were presented in three blocks of 40 sentences, with a small break between the blocks. Overall, the experiment lasted one hour and 40 min on average.

2.4. EEG recording

Electrophysiological activity was recorded from 27 tin electrodes (Fp1/2, F7/8, F3/4, FC5/6, FC1/2, T7/8, C3/4, CP1/2, CP5/6, P3/4, P7/8, O1/2, F/C/Pz) arranged in an elastic cap (Easycap) according to the extended 10–20 international system. Additional electrodes were placed over the left (on-line reference) and right mastoids. A forehead electrode served as the ground. In addition, four electrodes were placed around the eyes (VEOL, VEOR, HEOL, HEOR) in order to detect blinks and eye movements. Data were amplified (Brain Amp DC) with a bandwidth of 0.01–100 Hz, at a sampling rate of 250 Hz. The impedance of the scalp electrodes was kept below 5 kΩ, while the eye electrodes impedance was below 10 kΩ.

Further data analyses were pursued using Matlab toolboxes (Fieldtrip, Oostenveld, Fries, Maris, & Schoffelen, 2011; http://www.fieldtriptoolbox.org/) and R (R Core Team, 2015; https://www.r-project.org/). Collected recordings were off-line re-referenced to the average activity of the two mastoids. Raw data were visually inspected and artifacts such as muscular activity and ocular artifacts marked for subsequent rejection. Epochs of interest were computed from −0.3 s to 2 s with respect to the determiner onset. Two participants were excluded (and replaced) because more than 20% of the epochs were rejected. On average 5.50% of epochs were considered artifacts. No difference between conditions and groups emerged in terms of artifact rejection.

2.5. ERP data analysis

After baseline correction (−0.3 to 0 s) epochs were averaged independently for each condition and subject. We initially focused on a reduced time interval (−0.3 to 1 s) to select time windows of interest for the analysis of the prediction effects time-locked to the determiner. To this aim we ran for each electrode a point-by-point ANOVA in R considering three factors: Prediction (expected, unexpected), Transparency (transparent, opaque) and Group (Spanish natives, Basque natives). Type-1 error was controlled applying the Guthrie and Buchwald (1991) correction: p-values were plotted if they extended consecutively over a period of at least 48 ms (see Janssen, Hernández-Cabrera, van der Meij, & Barber, 2015). We specifically focused on the main effect and interactions involving the Prediction factor.

Significant interactions were resolved focusing on the average ERP activity across nine groups of electrodes (Left Anterior: F3, F7, FC1; Medial Anterior: Fp1, Fp2, Fz; Right Anterior: F4, F8, FC2; Left Central: T7, FC5, CP5; Medial Central: C3, Cz, C4; Right Central: T8, FC6, CP6; Left Posterior: CP1, P7, O1; Medial Posterior: P3, Pz, P4; Right Posterior: CP2, P8, O2) in the time interval of interest. We ran an ANOVA (Greenhouse-Geisser corrected) with the experimental factors of interest and two additional factors reflecting the electrodes' topographical distribution: Longitude (Anterior, Central, Posterior) and Laterality (Left, Medial, Right). The factors Prediction, Transparency and the topographic factors were nested under Group. Post-hoc analysis mainly focused on the Prediction effects employing FDR corrected t-tests.

Further analyses were pursued on the longer time-window (−0.3 to 2 s) to evaluate possible integration effects time-locked to the target noun presentation. Less relevance is given to this latter analysis, since the ERP effects could be affected by earlier modulations time-locked to the determiner. In this latter analysis, we focused on the N400 and the late positivity time-intervals (five-way ANOVA: Prediction, Transparency, Group, Longitude, Laterality) as indexes of successful integration of the target noun in the sentence context (see Foucart et al., 2014; Martin et al., 2013). Proficient readers should definitively show such effects.

2.6. Time-frequency data analysis

The data related to the prediction effects elicited by the determiner were further analyzed focusing on the beta band oscillatory activity (13–30 Hz). Artifact-free EEG data in the time-interval between the determiner and the noun onset (0–0.7 s) were selected. The time-varying power spectrum of single trials was obtained using a Hanning window approach (400 ms window, 0.5 Hz frequency steps, 5 ms time steps) for the overall frequency range between 2 and 40 Hz (as implemented in Fieldtrip). Power values were expressed as relative change from a baseline interval calculated from −0.3 to −0.05 ms. After power estimation single trials were averaged independently for each condition for further statistical analyses and grand-averaged for display purposes.

Statistical significance of the effects was evaluated by means of the cluster-based permutation approach as implemented in Fieldtrip (Maris & Oostenveld, 2007). This approach takes care of the multiple comparison problem by selecting clusters of electrodes, time points and frequency bands that are statistically different between conditions. The initial t-test was set at a probability threshold of 0.05 and the sum of the individual t-statistic in each cluster was employed to determine the cluster statistic. After the randomization procedure (1000 times), clusters exceeding the highest or lowest 2.5th percentile were considered significant. Pairwise comparisons were focused on the Prediction effect for each transparency level in each experimental group.

3. Results

3.1. Comprehension questions

Participants' responses to the comprehension questions during the EEG session were not significantly different in the two groups ($p > 0.1$). Spanish natives had an average accuracy of 86.98% (SD: 5.87), while Basque natives' accuracy was 88.28% (SD: 5.23).

3.2. ERP data

In Fig. 2 we report the point-by-point analyses considering the three experimental factors of interest. To better highlight possible prediction effects time-locked to the determiner presentation we report the data in the time window until 1 s. A strong effect of Prediction emerges in the time interval from around 250 ms until 400 ms. The effect is evident in most electrodes with a bilateral central-posterior distribution. Interestingly, the effect of Prediction re-emerges similarly across the whole scalp starting around 900 ms. The earlier (250–400 ms) Prediction effect reflects an early N400 effect that is evident across the two Transparency conditions in both levels of the factor Group (Fig. 3). The later Prediction effect (>900 ms) reflects the N400 effect time-locked to the noun following the determiner (Fig. 4). In fact, the onset of the following N400 effect is 200 ms after the presentation of the noun. Visual inspection of Fig. 4 also reveals a late positive component effect evident after the N400, supporting the claim that the effects observed reflect semantic integration (as in Molinaro, Carreiras, & Duñabeitia, 2012).

To further validate such analyses and make sure that the Prediction effect is statistically robust for the two groups, we ran the statistics (four-way ANOVA: Prediction, Transparency, Longitude, Laterality) independently for the two experimental groups. Both in the 250–400 ms post-determiner-onset time window [Spanish natives: $F(1, 23) = 11.81$, $p < 0.001$, $\eta^2_G = 0.05$; Basque natives: $F(1, 23) = 15.31$, $p < 0.001$, $\eta^2_G = 0.03$] and in the 200–500 ms post-noun-onset time window [Spanish natives: $F(1, 23) = 32.60$, $p < 0.001$, $\eta^2_G = 0.13$; Basque natives: $F(1, 23) = 9.24$, $p < 0.001$, $\eta^2_G = 0.04$] reliable effects of Prediction emerged. In the late positivity time window (600–900 ms post-noun-onset) we also observed a strong Prediction effect [Spanish natives: $F(1, 23) = 24.17$, $p < 0.001$, $\eta^2_G = 0.16$; Basque natives: $F(1, 23) = 40.33$, $p < 0.001$, $\eta^2_G = 0.19$] that was slightly more pronounced in frontal regions given the significant interaction between Prediction and Longitude [Spanish natives: $F(1, 23) = 4.78$, $p < 0.05$, $\eta^2_G = 0.01$; Basque natives: $F(1, 23) = 13.88$, $p < 0.01$, $\eta^2_G = 0.02$]. When the Group factor was also included in these analyses, no interaction of the experimental factor with Group emerged.

In Fig. 1 a triple interaction is also visible in the time interval between 170 and 250 ms in central and parietal electrodes. We further explored this effect statistically (five-way ANOVA: Group, Prediction, Transparency, Longitude, Laterality) in this time interval reporting the triple interaction between Group, Prediction and Transparency [$F(1, 46) = 6.42$, $p < 0.01$, $\eta^2_G = 0.01$]. To evaluate this interaction, we ran separate analyses in the two groups. Spanish natives did not show any main effect or interaction with the Prediction factor in this time interval (all $ps > 0.1$). Basque natives, on the other hand, showed an interaction between Prediction and Transparency [$F(1, 23) = 12.89$, $p < 0.001$, $\eta^2_G = 0.02$]. When considering the Prediction factor independently from the two levels of the Transparency factor, the effects were seen to be highly robust for transparent items [$F(1, 23) = 8.14$, $p < 0.01$, $\eta^2_G = 0.02$] but not for opaque items [$F(1, 23) = 2.38$, $p > 0.1$]. This effect is evident in Fig. 3 as a negative effect around 200 ms (N200) for unexpected vs. expected transparent items read by Basque natives.

3.3. Time-frequency data

The oscillatory power analyses pursued in the determiner time-window (0–700 ms) were aimed at evaluating possible beta-band (13–30 Hz) effects as an index of word-form level prediction (Bastos et al., 2015; Michalareas et al., 2016). Across the four unexpected vs. expected comparisons reported in Fig. 5 only the one involving the Basque group for transparent items revealed significant results. More specifically, two clusters emerged in this comparison. An earlier one (Cluster 1) mainly involved central electrodes and was significant in the 196–256 ms time interval and between 15 and 17 Hz (lower beta band). The later one (Cluster 2) emerged in the same lower beta frequency band and involved slightly more right-lateralized electrodes between 438 and 496 ms. Both clusters show more power in the low beta band for the unexpected condition as compared to the expected. We further explored other frequency bands (from 2 to 40 Hz and from 20 to 100 Hz by means of a multi-taper approach) but no reliable effects were observed.

It is interesting to note that Cluster 1 highly overlapped in time (ms) and space (scalp topography) with the early negative effect (N200) for this same comparison in the ERP analyses. We further explored the nature of the early oscillatory effect (Cluster 1) by means of Pearson correlation (R software) between the beta band effect and the tasks involving grammatical gender for the Basque natives group. We thus computed for each Basque native participant the transparency advantage, i.e., the difference in response time for opaque minus transparent items in the gender decision task. Then we extracted the difference between the beta power for the unexpected minus the expected condition for transparent items in the most representative electrode (C4 reported in Fig. 5). The two measures positively correlated ($r(23) = 0.53$, $p < 0.001$) showing that the participants who had a stronger sensitivity to transparency in the gender decision task were the ones showing a larger effect in the lower beta range (no correlation for Spanish natives). No relevant correlation was observed involving response times for the grammaticality judgement task.

4. Discussion

In the present experiment, participants read sentences that were highly constraining towards a specific lexical item. Grammatical gender features are encoded in the lexical representation of nouns (Harris, 1991), since each inanimate noun has its own grammatical gender (*mes-a*, table, is only feminine; see also Levelt, Roelofs, & Meyer, 1999). This makes such a feature a relevant constraint for lexical prediction.

In the present design, when bottom-up information provided by the determiner interacts with top-down information provided by the predicted noun the two sources of information can either match or not. The timing of the ERP Prediction effect (unexpected vs. expected determiner) reveals the processing stage at which the two representations interact. The unexpected determiner provides a grammatical gender value that contrasts with the information encoded in the predicted lexical item. This "representational contrast" triggers a conflict effect that takes place at the lexical level of processing for Spanish natives, independently of the transparency of the predicted lexical element. The 250–400 ms effect (Fig. 3) likely represents a lexical-related negativity (earlier N400) already reported for function words (King & Kutas, 1998; Molinaro et al., 2008; Osterhout, Bersick, & McKinnon, 1997). This effect replicates what Foucart et al. (2014) reported for Spanish monolinguals and Spanish-Catalan early bilinguals, employing a similar design.

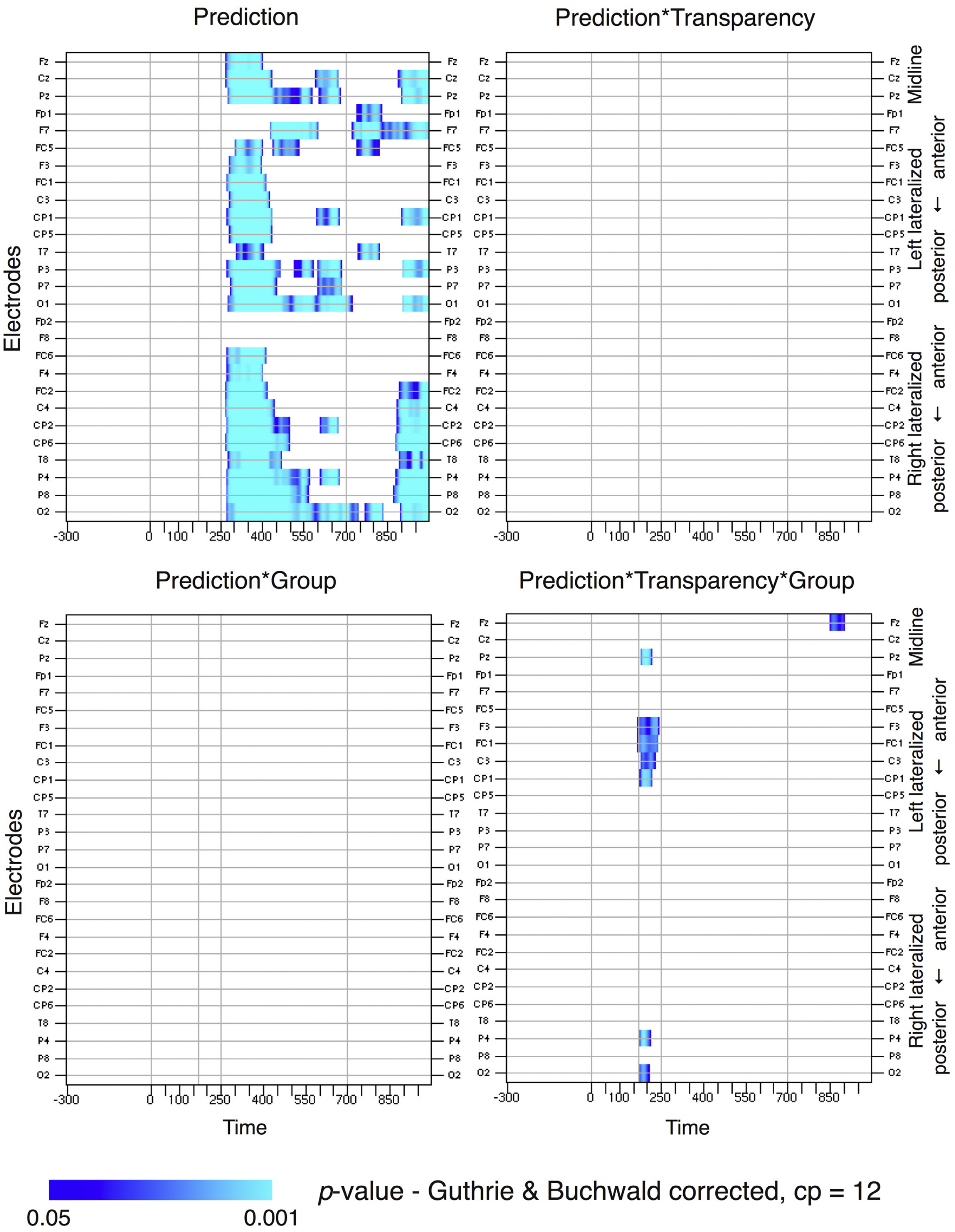

Fig. 2. Point-by-point split-plot analysis of variance (Guthrie & Buchwald, 1991, corrected) for each electrode considering the three experimental factors (within: Prediction and Transparency; between: Group). We report the main effect and the interactions involving the Prediction factor. Vertical grey lines at 0 and 700 ms indicate respectively the onset of the determiner and the onset of the predicted noun. Vertical grey lines at 170 and 250 ms indicate the time interval in which the triple interaction emerged.

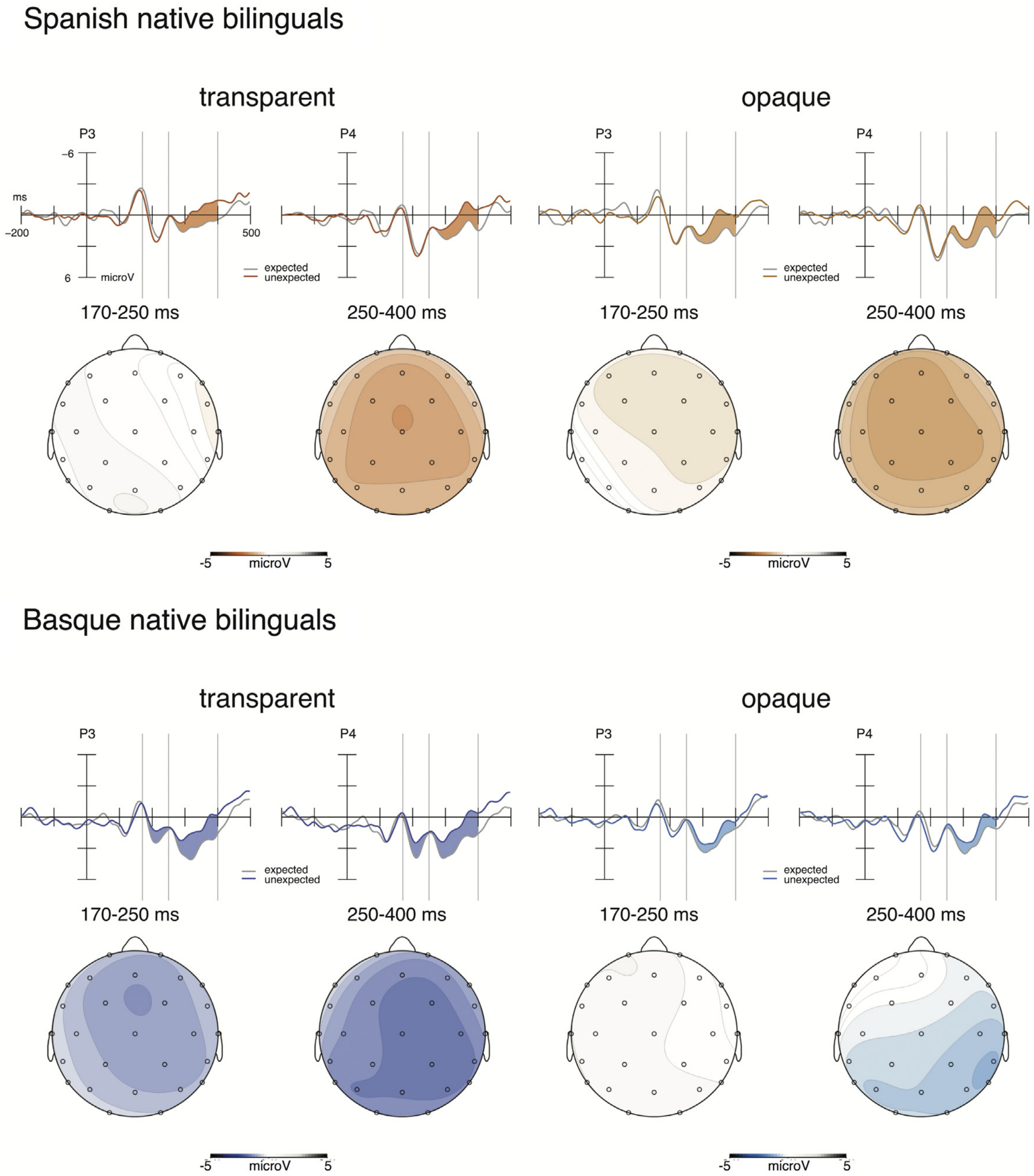

Fig. 3. ERPs for the expected and unexpected condition (Prediction effect) time-locked to the presentation of the determiner preceding the predicted noun (plotted in −200 to 500 ms for better displaying the ERP effects). We plotted separately the conditions based on the Transparency of the predicted noun and the Group of native speakers. For each plot we report in the upper panels the waveforms in two representative parietal electrodes (P3, P4; negative values plotted up) where we highlight the time intervals of interest (N200: 170–250 ms; N400 250–400 ms). Shadowed differences are the statistically significant ones. In the lower panels we present the topographical distribution of the difference effect (unexpected minus expected) in the two time intervals of interest.

4.1. Early prediction effect for Basque natives

The same lexical effect was also observed for Basque natives for both transparent and opaque items but it was preceded by an even earlier effect for transparent words. The timing (170–250 ms) of this earlier effect is indicative that the gender value of the determiner mismatches with the gender value of the predicted transparent items at a pre-lexical level. The topographical distribution of this effect is consistent with an increased N200, considered as reflecting orthographic processing in the visual domain (Holcomb & Grainger, 2006) and classically interpreted as a mismatch detector (see discussion about lexical prediction effects in Brother et al., 2015; see also Federmeier, Mai, & Kutas, 2005; Kim & Lai, 2012). The oscillatory evidence (involving the low beta-band channel, as in Bastos et al., 2015; Michalareas et al., 2016; see also Molinaro et al., 2016) for this transparent condition further supports our observation of a a prediction effect involving visual word form representations. In our view, the increased beta power for the unexpected determiner likely reflects an on-line update of the predicted representation: since the determiner is not gender-consistent with the predicted noun, the system inhibits such initial prediction and activates other possible (less predicted) lexical candidates.

The present data thus reveal that Basque natives activate word form level representations when predicting items whose ending is gender informative. The more sensitive these speakers were to the

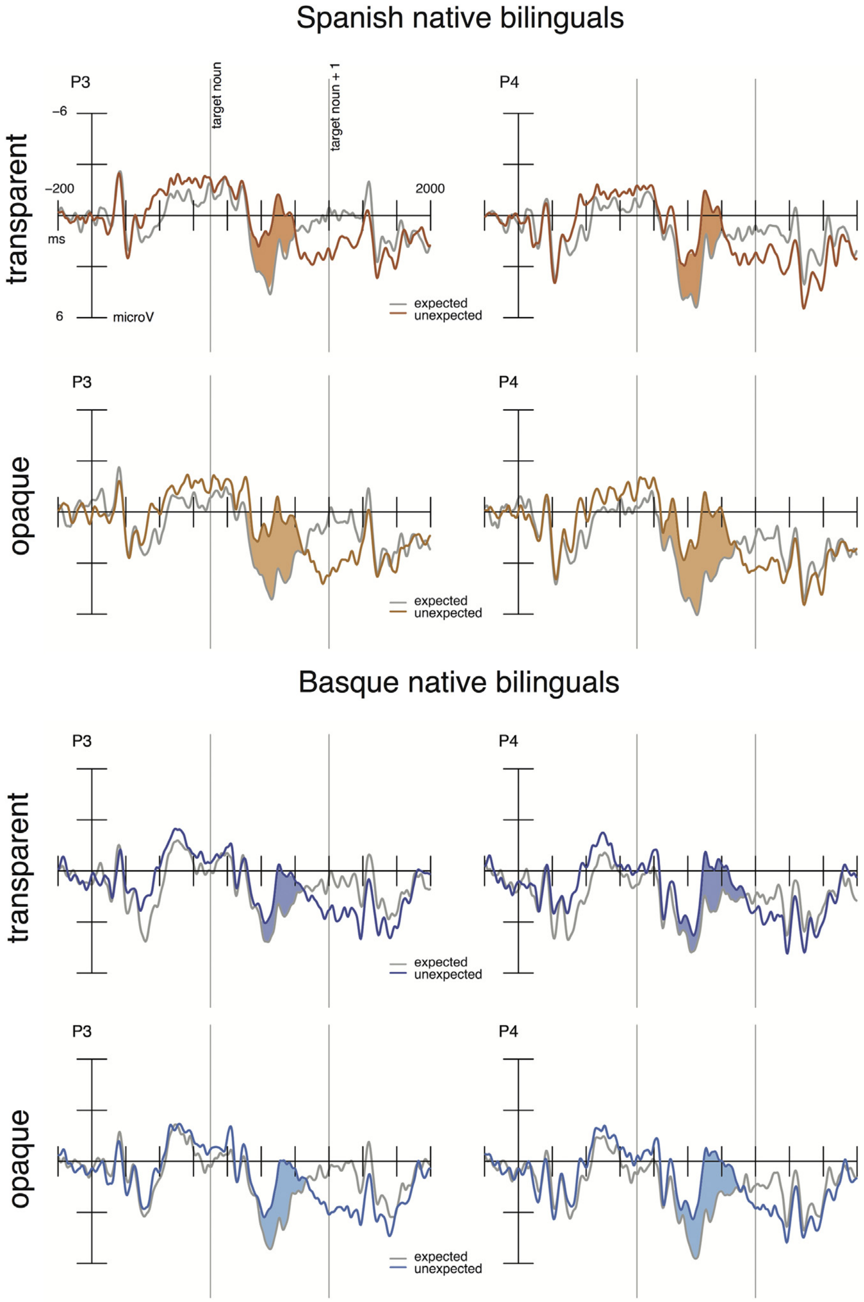

Fig. 4. ERPs for the expected and unexpected conditions considering the longer time interval (−300, 2000 ms) including both the determiner and the predicted noun in two representative parietal electrodes (P3, P4; negative values plotted up). Shadowed differences reflect the N400 effect post-noun onset, showing successful integration of the prediction. This effect is significant in all conditions.

gender transparency of the target items (in the gender decision task, Fig. 1) the stronger the word form prediction (as evidenced by the beta effect). In our opinion, this word-form prediction is not due to a reduced proficiency for Basque natives in Spanish, since in the general proficiency assessment the two groups of bilinguals were perfectly balanced for Spanish. In the grammatical gender proficiency assessment, we also did not find reliable differences between groups. The slower response times for Basque natives in the grammaticality judgment task would have predicted a later (and not earlier) electrophysiological reaction for this group compared to Spanish natives; importantly, the lack of interaction between transparency and group in the grammaticality judgment task (Fig. 1) does not parallel the strong interaction effect we observed in the EEG experiment involving transparency and group (Fig. 2). Furthermore, we cannot interpret the similar lexical effect for transparent and opaque items (the one emerging at 250–400 ms) as an index of lower proficiency for Spanish native bilinguals, since this overlaps (in time and scalp topography) with what has been reported for Spanish monolinguals with transparent items (Foucart et al., 2014). Previous studies have already suggested that Spanish speakers do not handle transparent and opaque items differently during sentence processing, given the large amount of irregularities (~1/3) in noun-endings within the Spanish lexicon (Caffarra et al., 2014, Caffarra & Barber, 2015; Molinaro, Barber, Pérez, Parkkonen, & Carreiras, 2013).

We therefore attribute the early effect for Basque natives to their early language exposure. Basque natives in their L1 strongly rely on post-nominal suffixes as highly frequent cues that are

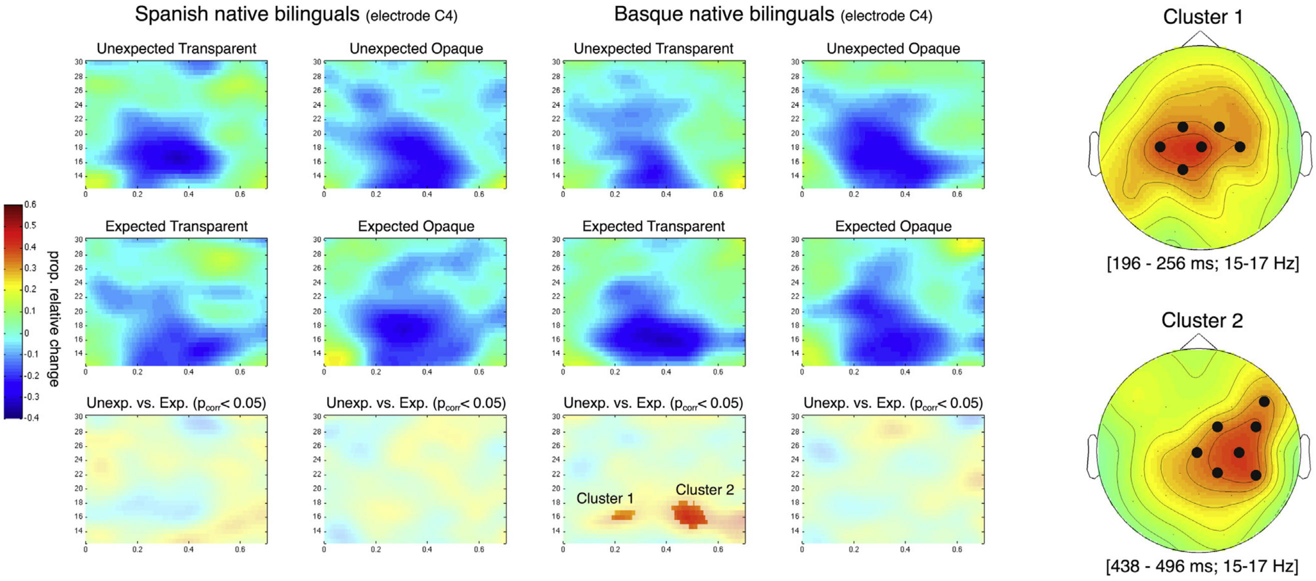

Fig. 5. Oscillatory power beta-band activity (13–30 Hz, relative change with respect to the baseline −300, −50 ms) time locked to the determiner preceding the predicted noun (at 700 ms). We report both the beta power decrease for all the conditions and the difference between unexpected minus expected determiner in a representative electrode (C4). Statistically significant time-frequency clusters are not shadowed, showing two significant effects in the 15–17 Hz frequency band for the transparent contrast in Basque natives. The topographical distribution of these two significant clusters is reported on the right, in which the electrodes contributing to the cluster are marked.

employed to bootstrap syntactic representations of Basque during language acquisition (Molnar et al., 2014) and for speech segmentation (as suggested by Gervain et al., 2013, see Gervain & Meheler, 2010). We have pointed out in the introduction that this morphological information stimulates more attention to the word-form properties (and more specifically to the noun endings) in the native language. It is likely that in our experiment Basque natives activate word-form representation also in Spanish: when the gender value is available at the word form level (for transparent items), an early word form electrophysiological effect emerges; when the gender value could be extracted only at the lexical level of processing (for opaque items), a lexical effect emerges.

4.2. The role of native exposure

The present findings thus speak for the idea that language prediction is mainly tuned to the native language characteristics (*native exposure*). Such properties largely vary across languages and speakers adapt their predictions mainly to the regularities of their native language. Recent proposals (Chang et al., 2006; Mani & Huettig, 2012) suggest that prediction is a relevant mechanism through which infants bootstrap the statistical regularities of the language to which they are initially exposed. Based on efficient prediction mechanisms, children can start to develop appropriate language production skills ("Prediction is Production" in the P-chain by Dell & Chang, 2014; see also Molinaro, Barraza, & Carreiras, 2013; Molinaro et al., 2016; Pickering & Garrod, 2013). Along the same lines, associative learning theories state that prediction stimulates learning (Rescorla & Wagner, 1972; Schultz et al., 1997) and this is possibly true also for language (Kuperberg & Jaeger, 2016).

This prediction mechanism applies even to an L2 (and for a feature not present in L1) when speakers can properly master it. Here we show that even when our Basque natives are daily exposed to the statistical regularities (and irregularities as in the case of grammatical gender) of Spanish, they still show signs of the influence of their native language during prediction. It is plausible that the different prediction effects we observed in our two groups (Spanish and Basque natives) are mainly due to the fact that those speakers were exclusively exposed to their native language until the age of three. Consequently, they learned to efficiently predict based mainly on the regularities of their native language, adjusting their prediction mechanisms either to Spanish or to Basque. Predictors initially tuned to Spanish handle transparent and opaque items similarly during prediction, given the reliability of lexical-level grammatical gender processing (Caffarra & Barber, 2015; Caffarra et al., 2014), while predictors initially tuned to Basque show more word-form effects even in an L2 when word-form (noun endings) cues are available.

As an alternative explanation for the findings of the present study, it could be argued that while our two groups were proficiency-balanced in Spanish, they were not so in Basque. Even if they both show high proficiency in this language, the Spanish natives were statistically less proficient in Basque than the Basque natives. This could have determined the stronger influence of Basque (L1) on Spanish (L2) for Basque natives, compared to the reduced influence of a weaker Basque (L2) on Spanish (L1) for Spanish natives (*L2 attrition*). We cannot completely exclude this hypothesis in the present study. Nonetheless, to examine this more closely, we selected the five Spanish native speakers that were most proficient in Basque (LexTALE mean score: 44.60; picture naming: 60.20; similar to Basque natives, see Table 1). We averaged their ERPs for the transparent conditions and we did not find quantitative differences in the 170–250 ms interval (amplitude across all electrodes: expected condition: 0.86 microV; unexpected: 0.88 microV; midline electrodes: expected: 1.77 microV; unexpected: 1.73 microV; electrodes showing the triple interaction in Fig. 2: expected: 1.51 microV; unexpected: 1.83 microV). This last exploratory analysis suggests that even for the Spanish natives who have higher proficiency scores in Basque there is no N200 expectation effect for transparent words. This does not support the idea that in the present experiment we observed an L2 attrition effect.

4.3. Developing prediction processes

In terms of proposals about L2 prediction, this study does not support hypotheses indicating that prediction in L2 is only a matter of proficiency (i.e., the more you know a language the more native-like your predictions will be; see review by Kaan, 2014). On the contrary, the present study shows that the native prediction mechanisms adapt to the properties of a second language, identifying regularities similar to those available in the native language. This view thus supports the idea that there are no separate domains of prediction, but that it is a unified mechanism looking for useful cues independently of the language processed. This can possibly provide clues and new directions for language learning research. Similar distributional regularities between a native and a new-learned language could serve to boost prediction in a second language and, consequently, facilitate its learning (Kuperberg & Jaeger, 2016; Molinaro et al., 2016).

We would like to emphasize that we do not think that prediction can only develop until the age of 3 and that it cannot further change through experience. Importantly, predictions can flexibly and rapidly adapt to the conditions of a new context (Bar, 2007; Sohoglu & Davis, 2016) by picking up all the available cues to construct an internal representation of the new environment (and develop predictions). However, early language exposure biases the way in which different cues are weighted to pursue optimal prediction mechanisms in the new experience settings.

4.4. Conclusion

The present study provides new insights into the mechanisms of prediction in sentence comprehension. Taking advantage of the typological distance between Spanish and Basque in early balanced bilinguals, we add an important piece of evidence to the puzzle on how multilingual experience shapes language prediction. Both evoked (N200-N400 prediction effects) and oscillatory electrophysiological evidence (15–17 Hz beta band activity) indicate that prediction can top-down reach the word-form hierarchical level of representation even in a second language. Based on this, we advance the hypothesis that prediction mechanisms are strongly influenced by the properties of early language exposure.

Acknowledgements

This work was partially supported by the Spanish Ministry of Economy and Competitiveness (grants PSI2012-32350 and PSI2015-65694-P to N.M. and grant PSI2014-54500-P to C.M. and S.C.), by the Basque Government (grant PI_2015_1_25 to C.M. and S.C.) and by the Ikerbasque Research Fellowships to N.M and C.M.

Further support derived from the AThEME project funded by the European Commission 7th Framework Programme, the ERC-2011-ADG-295362 from the European Research Council and the award "Centro de Excelencia Severo Ochoa SEV-2015-0490".

We would like to thank Margaret Gillon-Dowens and the Proactive group for comments on previous versions of this manuscript.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. http://dx.doi.org/10.1016/S0010-0277(99)00059-1.

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2015). Delta–beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex, 25*(9), 3077–3085. http://dx.doi.org/10.1093/cercor/bhu103.

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11*(7), 280–289. http://dx.doi.org/10.1016/j.tics.2007.05.005.

Bastos, A. M., Vezoli, J., Bosman, C. A., Schoffelen, J. M., Oostenveld, R., Dowdall, J. R., ... Fries, P. (2015). Visual areas exert feedforward and feedback influences through distinct frequency channels. *Neuron, 85*(2), 390–401. http://dx.doi.org/10.1016/j. neuron.2014.12.018.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition, 136*, 135–149. http://dx.doi.org/10.1016/j.cognition.2014.10.017.

Caffarra, S., & Barber, H. (2015). Does the ending matter? The role of gender-to-ending consistency in sentence reading. *Brain Research, 1605*, 83–92. http://dx.doi.org/10.1016/j.brainres.2015.02.018.

Caffarra, S., Janssen, N., & Barber, H. (2014). Two sides of gender: ERP evidence for the presence of two routes during gender agreement processing. *Neuropsychologia, 63*, 124–134. http://dx.doi.org/10.1016/j.neuropsychologia.2014.08.016.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234–272. http://dx.doi.org/10.1037/0033- 295X.113.2.234.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics, 27*, 3–42. http://dx.doi.org/10.1017/S0142716406060024.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204. http://dx.doi.org/10.1017/S0140525X12000477.

de Rijk, R. P. G., & de Coene, A. (2008). *Standard basque: A progressive grammar*. Cambridge, MA: MIT Press.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1634), 1–9. http://dx.doi.org/10.1098/rstb.2012. 0394.

DeLong, K., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117–1121. http://dx.doi.org/10.1038/nn1504.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: one-stop shopping for Spanish word properties. *Behavioral Research Methods, 45*(4), 1246–1258. http://dx.doi.org/10.3758/s13428-013-0326-1.

Dussias, P. E., & Pinar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh- gaps by Chinese-English second language speakers. *Second Language Research, 26*(4), 443–472. http://dx.doi.org/10.1177/0267658310373326.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*, 491–505. http://dx.doi.org/10.1111/j.1469-8986.2007.00531.x.

Federmeier, K. D., Mai, H., & Kutas, M. (2005). Both sides get the point: Hemispheric sensitivities to sentential constraint. *Memory and Cognition, 33*(5), 871–886. http://dx.doi.org/10.3758/BF03193082.

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language, 69*(3), 165–182. http://dx.doi.org/10.1016/j.jml.2013.06.001.

Foucart, A., Martin, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1461–1469. http://dx.doi.org/10.1037/a0036756.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences, 360*(1456), 815–836. http://dx.doi.org/10.1098/rstb.2005.1622.

Gervain, J., & Meheler, J. (2010). Speech perception and language acquisition in the first year of life. *Annual Reviews of Psychology, 61*, 191–218. http://dx.doi.org/10.1146/annurev.psych.093008.100408.

Gervain, J., Sebastian-Galles, N., Diaz, B., Laka, I., Mazuka, R., Yamane, N., ... Mehler, J. (2013). Word frequency cues word order in adults: cross- linguistic evidence. *Frontiers in Psychology, 4*, 689. http://dx.doi.org/10.3389/fpsyg.2013.00689.

Gollan, T., & Frost, R. (2001). Two routes to grammatical gender: Evidence from Hebrew. *Journal of Psycholinguistic Research, 30*, 627–651. http://dx.doi.org/10.1023/A:1014235223566.

Guthrie, D., & Buchwald, J. S. (1991). Significance testing of difference potentials. *Psychophysiology, 28*(2), 240–244. http://dx.doi.org/10.1111/j.1469-8986.1991.tb00417.x.

Harris, J. W. (1991). The exponence of gender in Spanish. *Linguistic Inquiry, 22*, 27–62.

Holcomb, P. J., & Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience, 18*(10), 1631–1643. http://dx.doi.org/10.1162/jocn.2006.18.10.1631.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research, 29*(1), 33–56. http://dx.doi.org/10.1177/0267658312461803.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica, 35*, 49–66.

Janssen, N., Hernández-Cabrera, J. A., van der Meij, M., & Barber, H. A. (2015). Tracking the time course of competition during word production: Evidence for a post-retrieval mechanism of conflict resolution. *Cerebral Cortex, 25*(9), 2960–2969. http://dx.doi.org/10.1093/cercor/bhu092.

Kaan, E. (2014). Predictive sentence processing in L1 and L2. *Linguistic Approaches to Bilingualism, 4*(2), 257–282. http://dx.doi.org/10.1075/lab.4.2.05kaa.

Kim, A., & Lai, V. (2012). Rapid interactions between lexical semantic and word form analysis during word recognition in context: evidence from ERPs. *Journal of Cognitive Neuroscience, 24*(5), 1104–1112. http://dx.doi.org/10.1162/jocn_a_00148.

King, W. K., & Kutas, M. (1998). Neural plasticity in the dynamics of human visual word recognition. *Neuroscience Letters, 244*, 61–64. http://dx.doi.org/10.1016/S0304-3940(98)00140-2.

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. http://dx.doi.org/10.1080/23273798.2015.1102299.

Laka, I. (1996). *A brief grammar of Euskara, the Basque language*. Leioa-Donostia (Spain): Euskal Herriko Unibertsitatea.

Lemhofer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavioral Research Methods, 44*(2), 325–343. http://dx.doi.org/10.3758/s13428-011-0146-0.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177. http://dx.doi.org/10.1016/j.cognition.2007. 05.006.

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex, 68*, 155–168. http://dx.doi.org/10.1016/j.cortex.2015.02.014.

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake – but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843–847. http://dx.doi.org/10.1037/a0029284.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(1), 177–190.

Martin, C., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language, 69*(4), 574–588. http://dx.doi.org/10.1016/j.jml.2013.08.001.

Michalareas, G., Vezoli, J., van Pelt, S., Schoffelen, J. M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron, 89*(2), 384–397. http://dx.doi.org/10.1016/j.neuron.2015.12.018.

Mitsugi, S., & MacWhinney, B. (2015). The use of case marking for predictive processing in second language Japanese. *Bilingualism: Language and Cognition, 19*(1), 19–35. http://dx.doi.org/10.1017/S1366728914000881.

Molinaro, N., Barber, H. A., Pérez, A., Parkkonen, L., & Carreiras, M. (2013). Left fronto-temporal dynamics during agreement processing: Evidence for feature-specific computations. *Neuroimage, 78*, 339–352.

Molinaro, N., Barraza, P., & Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: EEG correlates of processing expected words in sentences. *Neuroimage, 72*, 120–132. http://dx.doi.org/10.1016/j.neuroimage.2013. 01.031.

Molinaro, N., Carreiras, C., & Duñabeitia, J. A. (2012). Semantic combinatorial processing of non-anomalous expressions. *Neuroimage, 59*(4), 3488–3501. http://dx.doi.org/10.1016/j.neuroimage.2011.11.009.

Molinaro, N., Conrad, M., Barber, H. A., & Carreiras, M. (2010). On the functional nature of the N400: Contrasting effects related to visual word recognition and contextual semantic integration. *Cognitive Neuroscience, 1*(1), 1–7. http://dx.doi.org/10.1080/17588920903373952.

Molinaro, N., Monsalve, I., & Lizarazu, M. (2016). Is there a common oscillatory brain mechanism for producing and predicting language? *Language, Cognition and Neuroscience, 31*(1), 145–158. http://dx.doi.org/10.1080/23273798.2015.1077978.

Molinaro, N., Vespignani, F., Canal, P., Fonda, S., & Cacciari, C. (2008). Cloze-probability does not only affect N400 amplitude: The case of complex prepositions. *Psychophysiology, 45*(6), 1008–1012. http://dx.doi.org/10.1111/j.1469-8986.2008.00694.x.

Molnar, M., Lallier, M., & Carreiras, M. (2014). The amount of language exposure determines nonlinguistic tone grouping biases in infants from a bilingual environment. *Language Learning, 64*(s2), 45–64. http://dx.doi.org/10.1111/lang.12069.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 156869. http://dx.doi.org/10.1155/2011/156869.

Osterhout, L., Bersick, M., & McKinnon, R. (1997). Brain potentials elicited by words: Word length and frequency predict the latency of an early negativity. *Biological Psychology, 46*, 143–168. http://dx.doi.org/10.1016/S0301-0511(97)05250-2.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–347. http://dx.doi.org/10.1017/S0140525X12001495.

R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275*(5306), 1593–1599. http://dx.doi.org/10.1126/science.275.5306.1593.

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *PNAS*. http://dx.doi.org/10.1073/pnas.1523266113.

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiological Reviews, 90*(3), 1195–1268. http://dx.doi.org/10.1152/physrev.00035.2008.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16*(7), 1272–1288. http://dx.doi.org/10.1162/0898929041920487.