Reanalyzing language expectations: Native language knowledge modulates the sensitivity to intervening cues during anticipatory processing

Francesco Giannelli[1, 2], Nicola Molinaro[3, 4]

[1] Department of Cognition, Development and Educational Psychology, University of Barcelona, Barcelona, Spain

[2] Cognition and Brain Plasticity Unit, IDIBELL, L'Hospitalet de Ll., Spain

[3]BCBL, Basque center on Cognition Brain and Language, Donostia-San Sebastian, Spain

[4]Ikerbasque, Basque Foundation for Science, Bilbao, Spain

Correspondence to:

Francesco Giannelli

Cognition and Brain Plasticity Unit

Campus Bellvitge, Pavelló de Govern

Feixa Llarga s/n (08907 - L'Hospitalet de Llobregat)

Tel. +34657142472

e-mail:

francescogiannelli1@gmail.com

telephone:

+34 65714247

Abstract

We investigated how native language experience shapes anticipatory language processing. Two groups of bilinguals (either Spanish or Basque natives) performed a word matching task (WordMT) and a picture matching task (PictureMT). They indicated whether the stimuli they visually perceived matched with the noun they heard. Spanish noun endings were either diagnostic of the gender (transparent), or ambiguous (opaque). ERPs were time-locked to an intervening gender-marked determiner preceding the predicted noun. The determiner always gender-agreed with the following noun but could also introduce a mismatching noun, so that it was not fully task diagnostic. Evoked brain activity time-locked to the determiner was considered as reflecting updating/reanalysis of the task- relevant pre-activated representation. We focused on the timing of this effect by estimating the comparison between a gender-congruent and an – incongruent determiner. In the WordMT both groups showed a late N400 effect. Crucially, only Basque natives displayed an earlier P200 effect for determiners preceding transparent nouns. In the PictureMT both groups showed an early P200 effect for determiners preceding opaque nouns. The determiners of transparent nouns triggered a negative effect at ~430 ms in Spanish natives, but at ~550 ms in Basque natives. This pattern of results supports a "retracing hypothesis" according to which the neuro-cognitive system navigates through the intermediate (sub-lexical and lexical) linguistic representations available from previous processing to evaluate the need of an update in the linguistic expectation concerning a target lexical item.

Keywords: language anticipation, bilingualism, hierarchical representations, word expectation, grammatical gender, ERP

Introduction

General world knowledge and contextual constraints constantly influence the processing system of speakers. Top-down operations not only facilitate the incorporation of new inputs into the preceding structure after they have been observed (*integration*), but also pre-activate upcoming inputs before they are even perceived (*prediction*; e.g., Altmann & Mirkovic, 2009; Jaeger & Snider, 2013; Kuperber & Jaeger, 2016; Levy, 2008; MacDonald, 2013).

Anticipatory behaviour in the native language has been attested by a great number of studies. Experiments using the visual word paradigm have shown that when given a constraining context, participants tend to move their eyes towards objects or images before they are explicitly revealed by the input (Altmann & Kamide, 1999; Kamide, Altmann & Haywood, 2003). Research using Event-Related Potentials (ERPs) manipulated context sentences in order to measure what happens in the brain right before participants perceive an unexpected lexical item (in comparison with an expected one), and found replicable electrophysiological correlates of lexical prediction (DeLong, Urbach & Kutas, 2005; Wicha, Moreno & Kutas, 2003).

When they listen or read, native speakers can pre-activate the phonological (DeLong et al., 2005), orthographic (Hawelka et al., 2015), syntactic (Levy, 2008; Dikker et al., 2010) and semantic (Altmann & Kamide, 1999; Kutas & Hillyard, 1984) properties of the input coming next. However, even if "predictive processing" (see Clark, 2013) during language comprehension has been widely studied, a comprehensive picture of this phenomenon is still far from being complete.

The general aim of the present study is to investigate which cognitive mechanisms support anticipatory language processing, with a detailed focus on their precise timing of activation.

In order to properly present our experiment and illustrate the rationale behind it, we divided the Introduction in four parts. First, we will briefly bring up the importance of exploring anticipatory mechanisms in second language (L2) processing. In the second paragraph, we will review three key studies that are functional to our experiment, with a special focus on the third one, which directly inspired the present study. In the third paragraph, we will describe how multiple levels of analysis hierarchically might participate to shape the anticipation processes. In the fourth part, we will introduce the present study and its logic.

Anticipatory processes in the second language

Research on second language (L2) predictive processing is fundamental for comprehending the general mechanisms underlying anticipation during language processing. When comprehending a message in an L2, most speakers are not fast enough to keep pace with the speed with which a sentence unfolds and generally their comprehension is impaired (e.g. Hahne, 2001; Hahne & Friederici, 2001). Non-native speakers differ from natives in terms of quantity and quality of exposure to the L2: this distinct linguistic experience not only leads to reduced proficiency, but also to delayed, impaired or qualitatively different predictive processing (Kaan, 2014). According to the "shallow structure hypothesis", the syntactic representations computed by L2 learners during comprehension are less detailed than those computed by native speakers: for example, late learners are not able to map detailed syntactic representations onto semantic representations (Clahsen & Felser, 2006a, 2006b). It therefore seems reasonable to extend this shallower L2 analysis to a shallower anticipatory linguistic ability in the L2.

However, some interactive views suggest that L2 predictive processing is not an all-or-none issue. In connectionist psycholinguistic models sentence meaning is interpreted by taking into account linguistic cues in the sentence context, establishing cue weights, computing probabilistic values for each interpretation, and choosing the most likely option (Bates & MacWhinney, 1989). Because different languages use different cues to interpret meanings, cue weights differ between languages, and speakers of a given language use the cue weights associated with that language for the interpretation of an utterance. When people acquire a L2, they must learn which cues are important in which languages in order to effectively interpret the sentences. Thus, L1 and L2 cues compete in the mind of the learner during acquisition and usage of the language (MacWhinney, 1987, 1992, 2015). This competition leads to transfers and interferences at many linguistic levels, including anticipation mechanisms. Different linguistic experiences might therefore determine different predictions as the weights of the competing representations are computed and adjusted in real-time according to the experience of the user with the linguistic input. As a result, as the exposure of the learner to a new language grows, she/he will obtain an increasingly better understanding of the linguistic mechanisms of the target language. As the efficiency of the anticipatory system would directly depend on the proficiency in each language, it seems reasonable to assume that as proficiency in L2 gets higher, language prediction processes refine accordingly (Kaan, 2014).

The present study examines anticipatory processes that are at play during L2 lexical pre-activation when L2 proficiency is very high (thus comparable with L1 prediction processes). We show that even if proficiency is comparable and predictive processing efficient in L2 (as compared to L1), the native language exposure still affects the format of the anticipated representation (as reported also in Molinaro et al., 2017). Based on this, we propose a model of how the cognitive system dynamically adjusts the pre-activated representation based on available cues. Thus, through the analysis of bilingual speakers this experiment not only aims to explore the differences between L1 and L2 in anticipatory processing, but also has the goal of explaining these differences (if any) within a more general and unified theory of the predictive processing that supports language comprehension.

Anticipation in bilingual speakers

Few studies have investigated anticipation mechanisms in bilingual speakers.

Martin and colleagues (2013) employed a clever paradigm introduced by De Long and colleagues (2005). Participants read highly constraining English sentences that could only be followed by either an expected (e.g., *a kite*) or an unexpected (but correct) noun phrase (e.g., *an airplane*). Crucially, the experiment utilized the phonological property of English according to which the indefinite article "*a*" changes to "*an*" if the noun that follows it starts with a vowel. Taking advantage of this rule to detect how the pre-activation is adjusted based on the "article cue" previous to the target (expected) word onset, the authors compared a group of late proficient Spanish-English bilinguals and a group of English monolinguals. While monolinguals showed a differential ERP response at reading "*an*" vs. "*a*" (a negative effect around 300 ms, as did De Long et al., 2005), bilinguals did not. The authors concluded that L2 users do not predict to the same extent native speakers do.

This result (see also Ito et al., 2017) did not however take into account either the typological differences between the languages, or the impact that cross-linguistic similarities may have on anticipation processes. Foucart et al.(2014) compared Spanish monolinguals, Spanish-Catalan early bilinguals and French-Spanish late bilinguals, using a similar paradigm in Spanish by taking advantage of the gender value of determiners that could introduce more (*la puerta*, the [fem] door [fem]) or less (*el armario*, the [masc] closet [masc]) expected target words that were gender marked (see also Wicha et al., 2004). Interestingly, the same ERP effect (as in De Long et al., 2005) on the unexpected vs. expected article was found in all three groups. In contrast with the previous experiment, in this study the conclusion was that bilinguals can predict upcoming words in L2 as native speakers do.

The main difference between the two experiments was the linguistic feature manipulated to detect prediction mechanisms. While in the first experiment the phonotactic agreement feature ("*a*", "*an*") exists in English but not in Spanish, the gender marking between article and noun is present in Spanish, French and Catalan. This suggests that L2 speakers are more sensitive during anticipatory processing to linguistic cues that are also available in their L1.

Given the disparities in these experiments, we designed a research project that tested two groups of very early bilinguals who spoke two typologically distant languages daily. We aimed to investigate whether bilinguals can predict the features of the L2 even though those features are not part of their L1, and to discern whether typological language differences have an impact on L2 prediction processes. In a recent EEG study (Molinaro et al., 2017), we tested Basque (L1)-Spanish (L2) and Spanish (L1)-Basque (L2) speakers; all of them were highly proficient in both languages but they were different in terms of age of acquisition, with the L2 acquired at the age of three years old. Participants read sentences (in Spanish) in which a constraining context was followed by an expected or an unexpected noun phrase (gender marked article-noun). Mirroring Foucart et al., (2014), expected and unexpected nouns differed in gender so that it was possible to observe anticipation effects on the article. However, in comparison with previous studies that examined anticipation mechanisms in bilingual speakers, here there was a relevant difference mainly originated by the particular linguistic characteristics of the bilingual speakers considered, namely: Spanish and Basque.

In fact, there are two main typological differences, among others, between the Spanish language and the Basque language. First, while in Spanish function words precede nouns (e.g. *la mesa* [the table]), in Basque morphological regularities are based on post-nominal suffixes and nouns precede function words (e.g. *mahai-a* [table the]) (Laka, 1996; Rijk, 2008). As a consequence, Basque speakers have to pay more attention to noun endings in order to extract syntactic cues, compared to Spanish speakers who can process them earlier.

Second, the Basque language does not have grammatical gender features. Instead, in the Spanish language, grammatical gender is assigned to any inanimate noun, but the noun ending is not always diagnostic of the gender as there is a substantial amount of irregularity. Two thirds of Spanish nouns are gender-transparent (ending in *–a* for feminine and in *–o* for masculine nouns), but the remaining nouns are gender-opaque (ending in a consonant, or in a different non-diagnostic vowel) (R.A.E., 2010).

It has been hypothesized (Gollan & Frost, 2001) that grammatical gender can be computed on the basis of two different sets of cues: one based on the lexical properties (abstract features) of the noun stored in the mental lexicon, and one deriving the gender features from the formal properties of the noun (in Spanish *–a/–o*). According to this theory, as Spanish opaque nouns do not have formal cues for gender retrieval, grammatical gender necessarily has to be derived from the abstract lexical representation. In contrast, transparent nouns present reliable formal cues (e.g. gender-related noun endings), therefore both the lexical cues and the form-based cues are utilizable for gender extraction (see Caffarra et al., 2014, for supporting data).

In Molinaro et al. (2017), not only did expected and unexpected nouns differ in gender, but transparency was also taken into account. Participants read sentences that could end with a transparent or an opaque noun, such as *Acabo de salir de la casa y no recuerdo si he cerrado la puerta / el armario"* ("I just got out of the house, and I don't remember whether I closed the [fem] door [fem, transp] / the [masc] wardrobe [masc, transp]") or *Prefiero que el te esté muy dulce, puedes pasarme el azúcar / la miel por favor?* ("I prefer my tea very sweet, would you pass me the [masc] sugar [masc, opaque] / the [fem] honey [fem, opaque]?").

ERPs recorded on the determiners preceding the unexpected nouns (in comparison with the ones preceding the expected) revealed an ERP negative effect starting ~300 ms in both groups (identified as a N400 ERP component; Kutas & Hillyard, 1984), independently of the transparency of the pre-activated target noun. Given its onset timing, this effect was interpreted as reflecting the reanalysis of the lexical pre-activation induced by the unexpected article. In addition, since gender information can be extracted only on a lexical basis from gender-opaque items, we considered this later effect as reflecting lexical mismatch processing. This result not only demonstrated that highly proficient bilinguals can pursue predictions the way monolinguals do (also replicating Foucart et al. (2014), who found very similar N400 effects comparing different bilingual speakers), but importantly did so by using a feature, namely gender, which is not available in the L2 of the participants. In addition, the ERP analysis presented an even more striking outcome. On the unexpected determiners that preceded transparent words (but not opaque), Basque natives showed an earlier negative effect starting ~170 ms (identified as a N200). This earlier negative component was interpreted as reflecting orthographic mismatch detection in the visual domain (Holcomb & Grainger, 2006; Brothers et al., 2015; Federmeier, Mai, & Kutas, 2005; Kim & Lai, 2012).

Hierarchical levels of analysis in bilingual speakers

In order to explain our results, we formulated a multifaceted hypothesis. Spanish natives do not strongly rely on formal cues (i.e. *–a*/*–o* noun ending) to compute agreement dependencies involving grammatical gender (since ~1/3 of the nouns are gender opaque) (Caffarra & Barber, 2015), but also rely on the lexical information stored in the mental lexicon to estimate the gender of the noun that is coming next. This is true for both transparent and (critically) opaque target nouns as evident from the later (~300 ms) N400 effect. This would explain the similar lexical reanalysis (triggered by an unexpected determiner) of the pre-activated target noun representation.

On the other hand, Basque natives rely more strongly on sub-lexical, word-form related information for the reanalysis of the pre-activated transparent nouns, as reflected in the early (~170 ms) N200 effect. Since this information is not available for opaque nouns, they rely on the lexical cues in order to reanalyse the pre-activated (gender-opaque) target noun pre-activation. The reasons for this processing difference would reside in the fact that in their native language, Basque speakers are driven by default post-nominal suffix analysis to bootstrap syntactic cues in their mother tongue (Molnar et al., 2014), and they apply the same strategy in the L2 (Spanish) when cues are available.

This hypothesis implies two fundamental issues. First, it provides a step further in the research on multilingual predictive processing that goes beyond recent paradigms that have focused on the presence or absence of effects in L2 depending on the speakers' proficiency. These results show that the native language experience hierarchically influences the way expected linguistic information is pre-activated; in other words, the environmental language regularities available during early childhood influence reliance on different levels of linguistic representations for prediction. This leads to the second point: the experiment provides electrophysiological evidence that language prediction is not a "representation encapsulated" phenomenon (dealing only with high-level linguistic representations) but it flexibly takes advantage of the linguistic representations made available by different cognitive processes across the form-to-meaning (perception-to-abstraction) hierarchy.

The present study

The aim of the present study is to test the above hypothesis that the language processing system can "navigate" across different hierarchical level of representation to adjust the on-going expectations about a target noun. We did so by using two tasks: a word-matching task (referred to as WordMT in the following) and a picture-matching task (PictureMT), and combining them with the EEG to have the necessary high-temporal resolution to detect anticipation effects. The basic idea is that word and picture will stimulate a different chain of hierchically ordered processes, with the sub-lexical analysis preceding the lexical analysis for the WordMT and a lexical analysis preceding the sub-lexical in the PictureMT.

We analysed two similar groups of highly proficient bilinguals: participants were Spanish (L1) - Basque (L2) speakers and Basque (L1) - Spanish (L2) speakers, with the L2 acquired at the age of three years old.

We avoided comparing monolinguals and bilinguals. This is because bilinguals have a considerable amount of competing linguistic information (due to their other language) that can modify the prediction processing dynamics compared to monolinguals. Our present comparison between two groups of early balanced bilinguals, prevents this confound.

In two different blocks of the same experimental session, participants read a noun in Spanish on a screen, or they saw the picture of a noun on the screen (e.g. *cuchillo* (knife) or the image of a knife). These stimuli were followed by a voice saying a noun phrase (NP, a determiner followed by a noun). The determiner could be gender congruent, incongruent or neutral (e.g. *el* “the [masc]”, *la* “the [fem]”, *mi* “my” [neut]) in relation to the initial priming stimulus (printed word or picture), but it was always in agreement with its own noun. The noun could be congruent or incongruent with the prime stimuli visually presented. When the noun was incongruent, also its gender disagreed with the prime. Furthermore, nouns could be transparent or opaque. Participants had to indicate whether what they read or what they saw (the prime) corresponded to what they heard (the target).

We recorded the ERPs time-locked to both the determiner (and the noun) that the participant heard in the match compared to the mismatch condition, as an index of the predictive processing.

The reasoning behind the use of different prime formats comes from models of visual word perception and word production. On presentation of a written word stimulus, an orthographic (and a phonological) analysis of its visual features is rapidly performed simultaneously and in parallel. The orthographic analysis combines letters to reach a unique whole-word orthographic representation. This process provides the entry to a lexical level and from there onto semantics. In short: on reading a word sensory information first feeds forwards activation to the sub-lexical features of that item (reliably after 200 ms post-stimulus onset), and then to the lexical/semantic information related to that word (after 350 ms, Grainger & Holcomb, 2009; for theoretical proposals see Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981).

In contrast, after picture presentation, the processing hierarchy could be rather different. In a language task, upon seeing a picture, the cognitive system directly activates a conceptual representation that can be mapped onto a lexical representation (after 200 ms post-stimulus, affecting the P200 component: Strijkers & Costa, 2011) - similar to what happens in picture naming studies (see Dell & Chang, 2014; Pickering & Garrod, 2013; Molinaro et al., 2016). If we follow the parallel with the word production literature, the lexical representation (often defined as *lemma*) could be potentially mapped onto a sub-lexical one, i.e., a sequence of speech sounds (after around 300 ms; Indefrey & Levelt, 2004; Strijkers & Costa, 2016). Importantly, while this last step is mandatory for production, it could be relevant also in the present task in which the target word (where the matching task is performed) is presented acoustically.

Therefore, the rationale of the study is to track the time course of the mismatching effect (if we observe any) time-locked to the determiner, depending on the format of the initial prime (either a written word or a picture). In order to perform the matching task, participants pre-activate the word they expect to hear, according to the prime they perceived. The gender values of the prime will be consequently activated in time depending on the hierarchical level in which they are represented. In the case of WordMT, we expect that the prime (the written word) will sequentially activate the following representational levels, i.e., visual > sub-lexical > lexical. Importantly, access to lexical representation is mediated by sub-lexical (word form related) processing. If the gender value is represented at the sub-lexical level, the gender information provided by a mismatching (vs. matching) determiner can potentially trigger the analysis of the pre-activated representation earlier in time compared to a lexical-level gender representation.

In contrast, in the PictureMT, the feature pre-activation process triggered by the prime (the picture) would be visual > lexical > sub-lexical; in other words, full lexical access occurs earlier and it is not mediated by sub-lexical units. If the gender value is represented at the lexical level of processing it will trigger an earlier reanalysis of the pre-activated representation compared to a scenario in which gender is strictly sub-lexically encoded.

Crucially, our experiment did not only manipulate gender but also transparency. Let us consider the opaque words first. For these items, the reanalysis of the pre-activated representation of the target (triggered by a gender unexpected determiner) will be pursued at the lexical level, as there are no formal cues to give them clues about the gender. Therefore, in both groups of bilinguals the unexpected determiner should elicit the same pre-activation timing, but the ERP effect is expected to be earlier in the PictureMT (~200 ms, possibly a P200 modulation) than in the WordMT (~300 ms, possibly a N400 modulation) since the lexical information is earlier available in the former (Strijkers & Costa, 2011). As opaque-gender could be only evaluated on a lexical-basis, the timing effect elicited by the opaque words mismatch is not expected to be different across groups, therefore it represents our reference point for the lexical reanalysis of the target word pre-activation. On the other hand, transparent-gender items can elicit a different mismatch effect in the two groups, so their timing will be compared with the timing of the opaque-gender items.

Let us examine the transparent words now. In line with the previous experiment (Molinaro et al, 2017), we assume that Spanish (L1) speakers rely more on lexical information for the pre-activation of gender, so they should show a lexical ERP effect (with timing more similar to the timing of the effect observed for opaque items) after they hear an unexpected article, both in the WordMT and in the PictureMT.

In contrast, Basque (L1) natives performing the WordMT should present an earlier ERP effect for the unexpected determiner (compared to the mismatch effect that Spanish natives show for transparent items), as they rely more on sub-lexical information for the pre-activation of gender. On the other hand, when they see an image in the PictureMT, the lexical information will be accessed before the sub-lexical one. Thus, we expect Basque natives (who rely more on the sub-lexical analysis for gender extraction) to show a delayed ERP effect in comparison to the mismatch effect that Spanish natives show for transparent items.

If our expectations are correct, we will have further evidence that bilinguals can navigate the language hierarchy quickly adjusting to the input; and that the language acquired first has an impact on language anticipation processes, even in early bilinguals with very high proficiency

Material and methods

Participants

Forty-two early bilinguals participated in the experiment. They were divided in two groups. The first group was composed by twenty-one native speakers of Spanish who were first exposed to Basque after the age of three (11 females; age range 19-29, mean: 23.38, SD: 3.24: Age of acquisition of Basque: 3.61 y.o., SD: 1.46).

Twenty-one native speakers of Basque (13 females; age range 18-33, mean: 25.66, SD: 5.45: Age of acquisition of Spanish: 4.23 y.o., SD: 1.33) who started to learn Spanish after the age of three formed the second group. All participants were right handed, their vision was normal or corrected to normal, and they had no history of neurological disorder. Before taking part in the experiment, participants signed an informed consent. They received a payment of 10 € per hour for their participation. The study was approved by the Basque Center of Cognition, Brain and Language (BCBL) ethics committee.

In order to participate in the study, all the participants had to go through some language proficiency tests in both Spanish and Basque (results in Table 1). First, participants had to self-rate their language comprehension (on a scale from 1 to 10, where 10 was a native-like level; the result was averaged for speech comprehension, speech production, reading and writing). Basque speakers rated themselves very high in both Basque and Spanish, while Spanish speakers claimed they were slightly better in Spanish than Basque. Participants also went through a lexical decision task called LexTALE (Izura et al., 2014; Lemhofer & Broersma, 2012) to test their vocabulary knowledge. In Spanish, both groups showed the same very high score, but in Basque, Spanish speakers had lower scores than Basque natives.

Participants then had to name a sequence of pictures in both languages, which increased in difficulty over time. Again, both Spanish and Basque participants had the same native-range score in the Spanish language, but in the Basque language, Basque speakers were better than Spanish speakers. Finally, all the participants were interviewed by balanced bilingual linguists who rated them on a scale from 0 to 5: in both languages, no participants had a score below 4. All the participants live, study or work in a purely bilingual environment, and they speak both languages every day.

They all studied English at school, and claimed it was their third language. We applied the same measures to participants' English proficiency: there was no difference between groups, scores were good, but still much lower than Spanish and Basque (self-evaluation: 5.2, SD: 3.22; LexTALE, score 0–40: 22.43, SD: 5.15; picture naming: 44.07, SD: 7.22; interview: 3.88, SD: 1.44). Since participants were far more proficient in their two other languages than in English, we assumed that this third language could not influence the present design.

-- Please insert Table 1 around here --

Experimental design and materials

A list of 120 Spanish nouns was selected, where 60 nouns were transparent and 60 nouns were opaque (30 masculine and 30 feminine nouns per group).

The transparent masculine nouns ended with "-*o*", which is the typical Spanish ending for masculine, (e.g. *cuchillo*, "knife [masc]"), while the feminine nouns had the feminine ending "-*a*" (e.g. *silla*, "chair [fem]"). Gender irregular nouns were excluded. Opaque nouns showed endings that were not informative of the grammatical gender (i.e., "-*e*", "-*n*", "-*l*", "-*s*", "-*j*", "-*r*", "-*d*", "-*z*").

The mean number of letters for transparent and opaque nouns was identical (mean: 5.76 letters, SD 1.51; range: 4–9 letters). In addition, transparent and opaque nouns did not differ for measures of concreteness (transparent nouns mean: 5.86, SD 0.58; opaque nouns mean: 5.67, SD 0.63); imageability (transparent nouns mean: 6.11, SD 0.73; opaque nouns mean: 6.09, SD 0.75) and familiarity (transparent nouns mean: 6.16, SD 0.42; opaque nouns mean: 6.12.67, SD 0.53) (EsPal, Duchon et al., 2013).

Half of the nouns referred to artifacts, and half to natural objects.

The words selected were also used to create the stimuli for the picture-matching task. We found images that visually represented the nouns above. All the pictures were highly recognizable color photographs (.png extension, white background, 2000x2000 pixels) obtained from online image collections. In order to be sure that a picture could only be related to one possible noun, we ran a naming test. Spanish-Basque bilinguals (N=20) who did not take part in the experiment saw 240 images, and named them with the first noun that came up to their minds. We only chose the images whose name was univocally expressed by all the 20 participants, and we came up with a final 120 pictures that could only represent our original list of nouns.

Both the words and the images were followed by an auditory noun phrase formed by a determiner followed by a noun. The determiner could be a definite article (*el* "the [masc]", *la* "the [fem]"), or a possessive adjective (*mi* "my", *su* "his, her") that was gender unmarked, hence neuter. The noun could either match or mismatch with the previously presented stimulus, but always matched with its own determiner.

The experiment had 5 conditions. There were 2 main conditions (that represented 53% of the trials). In the first one, both the determiner and the noun matched with the prime stimuli (written or visual); in the second, both the determiner and the noun mismatched with the prime stimuli. The gender value of the determiner was balanced in the two conditions. In addition, we added a condition (26% of the trials) in which the determiner was neuter (not gender-marked), and the noun matched with the stimuli. This last condition was introduced to reduce strategic task-related effects in our experimental design time-locked to the determiner, but to focus participant's attention on the relevance of the noun. However, since the sound envelope of the neuter determiner was different compared to the experimental ones, we reasoned that the relative evoked response could not be directly compared with the experimental ones. Furthermore, there were 2 catch trial conditions (20% of the trials): in the first, a neuter determiner was followed by noun that mismatched with the previous stimulus; in the second, the determiner matched with the stimulus, but the noun did not.

For each word/image of the original list, 5 possible noun phrases were recorded corresponding to each of the 5 conditions (e.g. *BOTELLA* “bottle [fem]”: *la botella* “the [fem] bottle”; *mi botella* “my bottle”; *el mando* “the [masc] remote control” ; *su mando* “his/her remote control”; *la corona* “the [fem] crown”).

Sound strings were recorded by a Spanish female speaker at a natural but slow pace. The NPs were spliced and between the determiner and the noun there was a silence gap of about one second (range 1- 1.3s, the exact timing was measured for each item). This gap was considered necessary in order to have a clear time window to analyze expectation effects. All the items were checked for amplitude (recording, cuts, measures and standardization was done using a software for the analysis of speech in phonetics called "Praat"; Boersma, 2001; Boersma & Weenink, 2017).

With the purpose of assessing how many milliseconds a listener needs to distinguish between the determiner *el* “the [masc]”, and the determiner *la* “the [fem]”, we ran a discrimination test with 10 participants who did not take part in the experiment. We took five NPs whose determiner was *el*, and five whose determiner was *la*. The audio files corresponding to the determiners were cut in order to create smaller time windows. Participants listened to audio fragments containing the first 60 ms; 70 ms; 80 ms or 90 ms of the determiner of each NP for a total of 40 trials, and they had to indicate (by spelling it aloud to an experimenter) whether they thought it was an *el* or a *la*. For the 60 ms bits the inaccuracy percentage was 54% for the determiner *el*, and 64% for the determiner *la*. Participants listening to audio files lasting 70 ms had an inaccuracy of 16% for *el*, and 28% for *la*. When audio fragments lasted 80 ms participants recognized all trials with the determiner *el*, but they got wrong 3% of the *la* determiners. There was 100% of accuracy when the audio files lasted 90 ms for both determiners. Therefore, 90 ms is defined as the discrimination point (similar to the uniqueness point described by Marslen-Wilson, 1987) of the determiner gender value. Importantly, to avoid confusion in the description of the results, we decided to time-lock the emerging ERP components to the onset of the sound itself and not to the discrimination point. Readers might want to consider subtracting the time needed for discrimination from the timing of the components for a finer analysis of the ERP activation timings.

For both the WordMT, and the PictureMT, 3 lists were created. Each list had 240 visual stimuli, each of the 3 main conditions had 64 items, and the catch trial conditions had 24 items each. In each list, visual stimuli were repeated twice, but never in the same condition. Participants were never given the same list for both WordMT and PictureMT, furthermore, the task sequence was alternated among participants so that half of them first went through the WordMT and then the PictureMT, and the other half did the opposite. Within each list the words were balanced (all $p > 0.2$; based on EsPal, Duchon et al., 2013) for grammatical gender, word frequency (log-values: List1: 1.21, SD: 0.55; List 2: 1.26, SD: 0.52; List 3: 1.25, SD: 0.56), and number of letters (List1: 5.79, SD: 1.59; List 2: 5.79, SD: 1.46; List 3: 5.68, SD: 1.51). Finally, all the images were balanced among the lists, so that they were all equally distributed in all the conditions.

No differences emerged between lists.

Procedure

The EEG experiment was run in a soundproof electrically shielded chamber with a dim light. Participants were seated in a chair, about sixty centimeters in front of a computer screen. Stimuli were delivered with the PsychoPy software (Peirce, 2007).

In the WordMT, participants read words displayed in black letters on a white background. In the PictureMT, participants saw images in the center of the screen. After a fixation cross (lasting 500 ms), the words or the images appeared on the screen for 350 ms. After the visual stimuli disappeared, they were immediately followed by auditory stimuli played by two speakers (the article-noun NPs). After participants heard the NP, they had to indicate whether the word they read or the image they saw matched with the noun they heard. The question appeared in the center of the screen as soon as the sound finished, and the subject could answer using the relative buttons on the keyboard: the response hand was counterbalanced across participants and list. Numbers of correct responses were recorded, and RTs were calculated in milliseconds from the appearance of the question to the participant's key press. All trials were presented in a different random order for each participant (see Figure 1 for details).

WordMT and PictureMT were presented in two blocks of the same experimental session with a small break between the two.

A brief practice session included five words in one session, and five images in the other, followed by the relative auditory stimuli and the yes-no questions. Participants were asked to stay still and to try to reduce blinking and eye movements to a minimum, especially during the auditory presentation. Overall, the experiment lasted one hour on average.

-- Please insert Figure 1 around here --

Electrophysiological recording and data analysis

EEG was recorded from 27 electrodes placed in an elastic cap (Easycap, www.easycap.de): Fp1, Fp2, F7, F8, F3, F4, FC5, FC6, FC1, FC2, T7, T8, C3, C4, CP5, CP6, CP1, CP2, P3, P4, P7, P8, O1, O2, Fz, Cz, Pz. All sites were online referenced to the left mastoid (A1). Additional external electrodes were placed on the right mastoid (A2) and around the eyes (VEOL, VEOR, HEOL, HEOR) in order to detect blinks and eye movements. A forehead electrode served as the ground. 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. Collected recordings were off-line re-referenced to the average activity of the two mastoids. Artifacts exceeding 100 μV in amplitude were rejected. Raw data were visually inspected and artifacts such as muscular activity and ocular artifacts were marked for subsequent rejection. On average, 7.3% of epochs were excluded as considered artifacts. There was no difference between conditions and groups in terms of artifact rejection.

For the analysis of the determiner, epochs of 1600 ms (from –600 ms to 1000 ms) were obtained, considering a [–600 - 0] ms pre-stimulus baseline for baseline correction. The long baseline interval was chosen since in the target noun-locked ERP analysis we used a longer time interval of interest (details below) and we kept the same baseline in the two analyses. This is typical, for example, in CNV studies in which long-lasting effects can emerge between conditions (e.g., Walter et al., 1964; Weerts & Lang, 1973). For each condition, the average ERP waveforms were computed time-locked to the onset of the determiner. Epochs were averaged independently for each condition and subject.

Given the theoretical relevance of the effects time-locked to the determiner (reflecting reanalysis of a pre-activated representation), we mainly focus on the evoked EEG activity time-locked to this stimulus. The initial analysis of the data was performed through a six-way ANOVA evaluating the four-way interaction between Group (Spanish vs. Basque natives), Task (WordMT vs. PictureMT), Transparency (Transparent vs. Opaque target noun) and Match (Matching vs. Mismatching Noun). In this analysis, the dependent variable was the average ERP activity in time windows of interest. In addition to these experimental factors we included also the following topographical factors: Longitude (Anterior, Medial, Posterior) and Laterality (Left, Central, Right) following the clustering of electrodes in nine homogeneously distributed groups. We run this initial analysis in two specific time windows of interest (250-350 ms and 400-500 ms) based both on visual inspection and based on previous evidence on a similar design (Molinaro et al., 2017). In order to obtain a detailed estimation of the exact time course of the mismatch effects on the determiner and to define their evolution in time, we performed pairwise comparisons of the ERP waveforms (match vs. mismatch) with point-by-point (one point every 4 ms) t-test for each electrode. Single-subject ERPs were entered in this analysis. We ran separate comparisons for each task (WordMT and PictureMT), each gender type (transparent and opaque) and each language group (Spanish and Basque natives). To protect this analysis from false positives, we employed the Guthrie & Buchwald (1991) correction that filters out effects that last less than 50 ms (12 consecutive time points) in less than three sensors. Importantly, this approach does not constrain the selection of the time interval of interest. The exact timing of the ERP effects is crucial for disentangling the hierarchical format of representation processed.

Possible interactions between the independent variables of interest (Match, Transparency) within each experiment (Word/Picture matching task) and within each group (Basque/Spanish natives) were evaluated with further statistics. We selected 100 ms-long time intervals of interest (in theoretically relevant intervals emerging from the point-by-point analyses) after the determiner and entered the average ERP activity – across the electrodes in which a significant effect emerged – in a three-way post-hoc ANOVA with Match (two levels: match vs. mismatch), Transparency (two levels: transparent vs. opaque) and Electrode (variable number of levels depending on the electrodes of interest). P-values were Greenhouse-Geisser corrected. Given the perturbation of the ERP waveforms time-locked to previous determiner, we did not perform fine-grained timing analysis on the target noun. However, to estimate the mismatch effect time-locked to the target noun, we computed 3100 ms epochs (from –1600 ms to 1500 ms), applying the same pre-stimulus baseline used for the determiner, so that we could have all the NP electrophysiological time course, but the waveforms were time-locked at the exact onset of the noun for each condition. Here, no point-by-point analysis was performed since no relevant modulation was expected other than a lexical integration mismatch effect (namely the N400 for which the 300-700 ms post-target noun was selected). We evaluate the Match effect in an ANOVA independently for each task (Word MT and Picture MT) considering clusters of three electrodes and the following within factors: Match (two levels: match vs. mismatch), Transparency (two levels: transparent vs. opaque), Longitude (three levels: anterior, medial and posterior cluster position), Hemisphere (three levels: left, central and right cluster position). A further between-subject factor considered was Group (two levels: Spanish natives, Basque natives).

Results

Behavioral: WordMT

Error rates (mean percentage of incorrect responses) and reaction times (RT) from accurate trials were analyzed.

Spanish natives had a general average error rate of 4.50 % (6.54% for the opaque words, and 2.46% for the transparent words). In Basque natives, the average error rate was 3.29 % (5.03% for opaque words, and 1.54% for the transparent words).

In the Spanish group, the mean reaction time for the WordMT was 490 ms (SD: 0.37), while for the Basque groups it was 500 ms (SD: 0.30).

To analyze the reaction time values of the accurate trials we used two mixed-design ANOVAs, with

Transparency and Condition as within factors, and with Group as between factors. No significant differences emerged from the analysis: there were neither main effects, nor interactions among variables (Table 2).

-- Please insert Table 2 around here --

Behavioral: PictureMT

For the PictureMT, there was 3.11% error rate in Spanish natives (4.20% for the opaque words, and 2.02% for the transparent words). Basque natives gave 6.98% of incorrect answers (8.01% for the opaque, and 5.95 for the transparent).

The reaction time for the Spanish natives was 490 ms (SD: 0.50), while for the Basque natives it was 470 ms (SD: 0.24).

Here, again, we analyzed the reaction time values of the accurate trials using two mixed-design ANOVAs, with Transparency and Condition as within factors, and with Group as between factor. No significant differences emerged from the analysis and we did not find main effects, nor interactions (Table 2).

ERP data

In the early time interval (250-350 ms) an interaction between Group, Task, Match, Transparency and Laterality emerged [$F(2, 368) = 3.31$, $p < 0.05$, ges = 0.001]. In a later time interval considered (400-500 ms), the theoretically relevant four-way interaction (among Group, Task, Match, Transparency) emerged [$F(1, 92) = 3.98$, $p < 0.05$, ges = 0.002]. This initial overall analysis was then followed by separate point-by-point comparison for each task, gender type and language group.

WordMT: Determiner

In order to better highlight the mismatch effect time-locked to the determiner presentation, we report data in the time window until 1 second. In the Figures below (2-5), we draw the ERPs for the electrodes in the left hemisphere scalp region in which significant effects mainly emerged; the point-by-point analyses showing activation timing across all the electrodes; and the scalp topography relative to the 100-ms time windows showing theoretically relevant effects.

Opaque items (Figure 2): For the *Spanish* group, determiners preceding the opaque nouns display a more negative effect for Mismatch items: this long-lasting effect starts at about 420 ms post-determiner onset, with a widespread distribution in all the electrodes [400-500 ms, across all electrodes: Match = 0.24 μV, sem (standard error of the mean) = 0.20; Mismatch = −0.71 μV, sem = 0.23]. *Basque* natives showed a similar short-living negative effect starting 430 ms [400-500 ms, in the left posterior electrodes: Match = −1.41 μV, sem = 0.25; Mismatch = −2.21 μV, sem = 0.25]. These timings suggest that the lexical reanalysis effect emerges around 400 ms post-determiner onset and can be identified as a N400 (Kutas & Hillyard, 1984).

-- Please insert Figure 2 around here --

Transparent items (Figure 3): The determiners preceding transparent nouns showed different results. *Spanish* natives display a strong effect of Prediction in the time interval starting around 380 ms post-determiner onset: the effect is more negative for Mismatch compared to Match condition and it is evident in frontal and parietal electrodes [400-500 ms: Match = 0.34 μV, sem = 0.18; Mismatch = −0.21 μV, sem = 0.19]. Crucially, in the same condition *Basque* natives have an earlier effect in the time window from 240 ms to 350 ms. The effect is more positive for the Mismatch condition and particularly evident on the left parietal electrodes [250-350 ms: Match = 0.01 μV, sem = 0.14; Mismatch = 0.45 μV, sem = 0.13]. We identified it as a P200 modulation. The early effect is followed by a negative effect starting at about 430 ms with a wider scalp distribution [400-500 ms, across all electrodes: Match = 1.31 μV, sem = 0.22; Mismatch = 0.66 μV, sem = 0.22].

-- Please insert Figure 3 around here --

Post-hoc: The three-way ANOVA (Match by Transparency by Electrode) was run in the early time interval

(250-350 ms) to evaluate if the early positive mismatch effect depended on the type of stimuli and the type of bilinguals. The analysis performed in the electrodes of interest selected through the point-by-point analysis (P7, P3, and CP5) showed the interaction between Match and Transparency only for Basque natives [Spanish natives: $F (1, 20) = 0.63$, $p > 0.4$, ges = 0.002; Basque natives: $F (1, 20) = 5.01$, $p < 0.05$, ges = 0.006], confirming that the mismatch effect for Basque speakers depends on the transparency of the pre-activated noun.

PictureMT: Determiner

Opaque items (Figure 4): The results of the PictureMT display a different pattern. For determiners preceding opaque nouns, in both *Spanish* and *Basque* natives there is a robust effect in an early time window, starting around 250 ms post-determiner onset and emerges in central-posterior electrodes [Spanish: 250-350 ms, across all electrodes: Match = 1.69 μV, sem = 0.14; Mismatch = 2.77 μV, sem = 0.13; Basque: 250-350 ms, across all electrodes: Match = 2.85 μV, sem = 0.19; Mismatch = 3.49 μV, sem = 0.18]. The effect is more positive for mismatching items compared to matching items in both groups. This indicates that the lexical reanalysis effect in this task emerges around 250 ms, earlier than in the WordMT. Consistent with previous evidence on picture naming (Strijkers & Costa, 2011) we identified this as a P200 component.

-- Please insert Figure 4 around here --

Transparent items (Figure 5): The effect on transparent nouns emerged later. In *Spanish* natives, determiners preceding transparent nouns elicit an increased negative effect for mismatching items at about 420 ms post-determiner onset. The effect is frontally distributed [400-500 ms: Match = 3.01 μV, sem = 0.20; Mismatch = 1.86 μV, sem = 0.19]. A more left-posteriorly distributed negative effect for mismatching items is also present in *Basque* natives [500-600 ms: Match = −2.11 μV, sem = 0.22; Mismatch = −3.23 μV, sem = 0.25]: it onsets much later though, at 550 ms, i.e., around 120 ms later than the effect for the Spanish natives.

-- Please insert Figure 5 around here --

Post-hoc: The post-hoc ANOVA (Match by Transparency by Electrode) was run to evaluate if the early effect (250-350 ms) only emerged for opaque items in the two groups. Indeed, in the Spanish group, the interaction between Match and Transparency emerged in the electrodes of interest [$F(1, 20) = 10.51$, $p < 0.01$, ges = 0.011]. The Basque natives displayed the same interaction [$F(1, 20) = 4.65$, $p < 0.05$, ges = 0.006].

The same analysis was performed in the later time interval showing the effect only for Basque natives. In the 550-650 ms time interval (considering the electrodes showing the significant difference for transparent items in the Basque group) we observed a significant interaction between the same two factors [$F(1, 20) = 4.60$, $p < 0.05$, ges = 0.005]. However, it should be noted that no similar interaction was observed in the 400-500 ms where the effect emerged for transparent items in the Spanish group [$F(1, 20) = 1.66$, $p > 0.2$, ges = 0.001]. We can thus conclude that Spanish natives showed a statistically reliable Match effect in the 400-500 ms that was not strong enough to be reliable for Basques. In the later 550-650 ms time interval, a reliable Match effect was observed in the Basque native group but not in the Spanish native group.

WordMT and PictureMT: Noun

ERP results time-locked to the onset of the noun are clear-cut (Figures 2 to 5). In both groups, both WordMT and PictureMT produced a strong negative mismatch effect (N400) starting at about 300 ms and ending at about 800 ms. No differences could be observed in the visual inspection of the ERPs time-locked to the noun. The five-way ANOVA for the WordMT showed main effects of Match [$F(1, 40) = 403.90$, $p < 0.001$, ges = 0.927] and Transparency [$F(1, 40) = 192.96$, $p < 0.001$, ges = 0.761]. The same analysis for the PictureMT showed main effects of Match [$F(1, 40) = 211.77$, $p < 0.001$, ges = 0.769] and Transparency [$F(1, 40) = 103.43$, $p < 0.001$, ges = 0.462].

Discussion

The retracing hypothesis

In the present study, we evaluated the top-down processes through which bilinguals can on-line adjust their expectations concerning a target linguistic stimulus. The findings confirm that language pre-activation can be efficient even in a second language, when proficiency is high enough. Importantly, however, the linguistic experience produces an increased sensitivity (bias) to cues in a second language, as evidenced by the timing of the mismatch ERP effect recorded in the two experiments. In fact, consistent with the syntactic relevance that suffixes play in their L1, Basque natives show more sensitivity (compared to Spanish natives) to the formal grammatical gender cues (noun endings) of the pre-activated target word in contrast with the input gender information provided by the intermediate determiner.

We reached this conclusion based on the following interpretation of the dynamic cognitive processes occurring in the present task (what we termed the "retracing hypothesis"). At the moment the prime is presented, a series of hierarchical processes are triggered in order to pre-activate the (task-relevant) representation of the target auditory word: participants expect to hear a word that matches the prime they just perceived, and the intervening determiner provides a gender cue that either supports or contradicts this expectation. The gender feature of the prime is included in the expectation generated. In terms of representational levels, the word prime activates a sub-lexical (orthographic) representation that constrains lexical retrieval in the WordMT[1], while the picture prime directly activates a lexical representation from which the sub-lexical units are derived in the PictureMT. Then, 350 ms after the presentation of the visual stimuli, a determiner is presented auditorily: the perception of the determiner containing a gender cue, calls for an update and leads to an evaluation of the dynamics generated by the first stimulus. This process retraces the processing steps generated by the prime stimulus to determine how the gender information perceived matches with the gender feature of the prime, so as to produce a task-relevant update (either boosting or inhibiting the activation of the pre-activated representation). Thus, the visual information of the first stimuli is retrieved from working memory (where it is still available, given that only 350 ms have passed), and the consequent different representations (sub-lexical and lexical in the PictureMT; lexical and sub-lexical in the

[1]Given the type of the paradigm we used, participants should pre-activate a phonological word representation to compare it with. In the WordMT this phonological representation can be active either simultaneously to the lexical representation (as proposed in McClelland & Rumelheart, 1981) or later in time (than the lexical representation, as suggested in Coltheart et al., 2001) depending on the theoretical framework of reference.

WordMT) are “hierarchically re-processed” in light of the new gender evidence provided by the determiner. In our experiment, the electrophysiological potentials time-locked to the presentation of the intervening article represent the time-course of the hierarchical-processing of the prime stimulus to trace the source of the gender incongruence and possibly operate a task-relevant update. Note that the all of this retracing process happens before the perception and lexical integration of the noun.

Critically, the present study not only manipulated gender, but also transparency: the expected nouns could be either transparent or opaque. Consequently, the “retracing” mismatch effects emerge at different times depending on the representation level at which grammatical gender is better stored. Below we revise in detail the findings for each experiment.

Let us consider the WordMT first. For the unexpected article preceding opaque words, we found the same negativity effect for both Spanish and Basque natives. The effect onsets at ~430 ms in both groups and can be identified as a N400 modulation. The opaque conditions here provide the reference timing for lexical reanalysis. The result is in line with our expectations: participants have to rely on their (later) lexical knowledge (as reflected by the ERP lexical effect) for the gender prediction of opaque words, where no formal cues are available.

The effects on determiners preceding unexpected transparent words showed different results. The ERPs of Spanish natives revealed a negativity starting at ~390 ms roughly similar to the opaque words. Crucially, Basque natives displayed a much earlier mismatch effect at ~240 ms (larger positivity for mismatching items, that here can be interpreted as a P200 effect). This outcome confirmed our predictions: we had assumed that Basque natives rely more on (earlier available) sub-lexical, word-form related information for gender mismatch detection of transparent words and we were expecting this process to be reflected by the modulation of an early ERP effect, similar to the ~240 ms effect found here. On the other hand, in Spanish natives pre-activating transparent words, we were expecting to see a (later) lexical ERP effect (closer to the one for opaque words) as they would rely more on the lexical information for gender mismatch detection.

Let us now turn to the PictureMT results. Here we assumed the lexical information to be traceable earlier compared to the WordMT, as lexical access triggered by the picture would not be mediated by the sub-lexical properties of the word (but they could be available later in time). For the unexpected determiners preceding opaque words we found a similar effect at ~250 ms for both Spanish natives and Basque natives (an effect recognizable as a P200). Since opaque gender can be only identified based on lexical processing, this effect provides the reference timing about lexical analysis in the PictureMT (if we accept the prediction/production parallelism, lexical effects are typically reported with this time course in picture naming studies: Indefrey & Levelt, 2004; Strijkers & Costa, 2011; 2016). Here our expectations were confirmed, as the gender of opaque words can only be extracted from the lexical representation.

The ERPs recorded on determiners preceding transparent nouns displayed a later negative effect at ~430 ms in the Spanish group, and an even later effect (~550 ms) in the Basque group. Both groups presumably performed additional sub-lexical analysis after the lexical one to extract gender information, resulting in later timing. Importantly though, the effect for Basques emerged later in time compared to the Spanish natives. We will discuss this last point in detail next.

Interaction among hierarchical levels

This study provides further insights regarding how bilinguals update their lexical expectation due to informative cues in the on-going speech signal. We hypothesized that, within this top-down process, different representational formats of the expected word are activated immediately after the visual stimuli presentation, and then "retraced" after the presentation of the determiner looking for possible incongruences. In the WordMT, the neurophysiological results highlight two representational stages during language pre-activation. At an earlier one, participants activate the sub-lexical information of the expected noun (mainly evident from the Basque group). At a later stage, the lexical information they have about the expected noun is available (as stored in the mental lexicon). The two levels are differently represented in terms of neural timing, with the first operating earlier, at about 240 ms (reflecting a P200 modulation) after the gender value update triggered by the determiner, and the second operating later (about 400 ms, reflecting a N400 effect), as clearly emerged from the WordMT. This result supports what we recently observed during sentence reading (Molinaro et al., 2017), where similarly composed groups of bilinguals showed a lexical N400-like effect for lexical prediction during sentence reading (irrespective of gender transparency of the expected noun), while the Basque natives only showed an earlier N200-like effect only for the pre-activation of transparent words[2]. It is interesting to note that the results from the WordMT emerged using a completely different task, as here we did not use highly constraining contexts in order to elicit top-down language processing, like in the previous experiment, but instead we used a WordMT where the expectation was evoked by a single word on a screen. This is evidence that the kind of representation pre-activated in both Molinaro et al. (2017) and in the WordMT is strictly word-related and does not reflect language-independent semantic knowledge. In addition, the anticipation effects that we found are observable independently of the kind of experimental setting, supporting the idea that expectation-related top-down mechanisms are able to quickly and flexibly adjust to the different tasks a language user has to go through. Furthermore, in the present study, the target words were not visually presented (as in Molinaro et al., 2017) but auditorily, providing critical evidence that this type of predictive processing is modality independent.

Given the specifics of the PictureMT, we decided to approach and discuss its results taking advantage of the parallel flow of processing that has been suggested to take place for language production and prediction (Dell & Chang, 2014; Federmeier, 2007; Pickering & Garrod, 2007). In our experiment, participants initially perceive the image stimuli and to perform the task (and ultimately pre-activate the auditory word that is going to come with its relative gender), they activate the conceptual information relative to that image. This semantic information leads to lexical selection and consequently to the pre-activation of the sub-lexical information. Picture naming studies typically report earlier lexical-related responses compared to word recognition studies, with the more challenging lexical recognition condition showing larger P200 around 200-300 ms (Strijkers & Costa, 2011). The positive effect around 200 ms in the PictureMT could thus reflect lexical-related processing as the later negative effect affecting the N400 emerging in the WordMT. Importantly, the modulation in the time-course of these ERP effects reflects the interaction of distinct pre-activation processes with the perceived determiner information. Indeed, Basque native speakers show "mirror time-courses" in the two tasks (early effect for transparent and late for opaque word in the WordMT, and the opposite for the PictureMT). This is coherent with the representational level (sub-lexical or lexical)

[2] It is interesting that the effect in the reading experiment (Molinaro et al., 2017) was more negative for the mismatching determiner, while in this listening paradigm it is more positive. In our opinion this is due to the large difference between the two experimental paradigms and deserves more attention in future research.

at which the mismatch between the pre-activation and the determiner gender cue is retraced. Instead, Spanish natives show similar time courses for both types of gender in the WordMT and a noticeable difference in the onset of the ERP effect for opaque (early P200, ~250 ms) and transparent (later negative effect, ~430 ms) items in the PictureMT. We would have expected similar time-courses for transparent and opaque items in both tasks (late effect in the WordMT and early in the PictureMT) for this latter group.

This discrepancy in the onset difference that emerged in the PictureMT between the ERP mismatch effects for transparent and opaque words might be due a different amount of interaction between the lexical level and sub-lexical level for the two groups. The P200 effect found on the determiner preceding the opaque words is a fully lexical effect. It is not possible to be anything else because the gender of opaque words can only be retraced at the lexical representational level. In contrast, the later negative effect emerging at 430 ms in Spanish natives, and at 550 ms in Basque natives for articles preceding transparent words, could be due to a different recruitment of sub-lexical resources in the two groups for proper pre-activation reanalysis. Basque natives could rely exclusively on the later, sub-lexical analysis for the retracing of gender information. The result of this process is reflected by an effect happening ~130 ms later than that of the Spanish natives. On the other hand, Spanish natives instead could jointly take advantage of both the formal (sub-lexical) and the lexical information to perform the prediction. This would probably be the reason why the ERP mismatch effect for transparent target nouns happens later than for opaque nouns, but earlier than the effect that Basque natives show for transparent nouns.

This interaction between hierarchical levels might also be partially visible in the WordMT. In this task, sub-lexical information is available earlier than the lexical information. As indicated above, Spanish natives display a lexical effect in the N400 time window at ~390 ms on determiners preceding transparent words; this effect is also present for the determiner preceding opaque words in both groups, but it happens slightly later, at ~430 ms. It is possible that this earlier ERP effect found for Spanish natives expecting transparent words in the WordMT emerges from the joint evaluation of lexical and formal, sub-lexical, gender cues available from an earlier processing stage. Crucially, Spanish natives do not fully “trust” formal gender cues in Spanish because they are diagnostic of gender value only for two thirds of the cases, but they can still take advantage of these formal cues for fast update of their language expectations. This inter-level interaction does not take place for the pre-activation of opaque words because their gender cannot be extracted from sub-lexical information, and it does not happen in Basque natives pre-activating transparent words, as they weight much more (than Spanish natives) sub-lexical information for gender value extraction. This explanation is debatable and requires further experimental validation. However, it does not affect the retracing hypothesis as a mechanism through which speakers (independently from their language background) reanalyse and update their linguistic expectations.

The role of native experience

This experiment provides further evidence that lexical pre-activation does not only involves anticipating the semantic traits of a word (Federmeier, 2007, Federmeier et al., 2002), but also its grammatical features. These features can be pre-activated even when they are not present in the bilinguals' L1, thus supporting the idea that there are no separate language domains for anticipatory processing, and instead, that it is a unified mechanism looking for informative cues independently of the language processed.

As outlined above, the Shallow Structure hypothesis (Clahsen & Felser, 2006) claims that even highly proficient second-language speakers are not able to construct detailed representations of the L2, because the L2 processing is fundamentally different. Our results instead suggest that predictive processing in L2 is not inherently different from predictive processing in L1, but it could be modulated by factors related to non-native comprehension. In agreement with the connectionist models reviewed in the first part of the Introduction, we think that the main difference between L2 speakers and native speakers lies in the frequency of their exposure to specific combinations of words and structures. Less exposure leads to altered statistical estimates concerning a target language, therefore differences in predictive processing. A weaker representation of the distributional properties of a language can either trigger weaker predictions, or predictions that are different from those native speakers make (Gollan et al., 2008). Furthermore, bilingual speakers have to deal with two languages at the same time. Therefore, when they process words in one language, they are also likely to activate connected words in their other language (e.g., Dijkstra, Grainger, & van Heuven, 1999; Van Hell & Dijkstra, 2002). Of course, this significant activation of competing information may damage language pre-activation, because anticipation is more difficult if several almost equally possible continuations are probable. Therefore, L2 speakers potentially activate more information during processing than native speakers, so they are likely to suffer more from competition: this may obstruct or slow down the generation of expectations. Nonetheless, our results clearly indicate that when L2 proficiency is high enough, L2 pre-activation mechanisms can be very much like the ones used in the L1, because the efficiency of this predictive mechanism mainly depends on the proficiency in each language. Consequently, there is no need to assume qualitative differences between native and non-native speakers; instead, the same mechanisms that affect anticipatory processing in native speakers can account for the anticipation mechanisms used by language learners (Kaan, 2014).

A fundamental point emerging from the study is that language anticipatory processing is tuned to the characteristics of the L1 and consequently is extremely sensitive to the native experience. As the most noticeable difference between groups is the language learned first, we can assess that the prediction differences found between Basque and Spanish natives are likely due to the early L1 exposure. During the first three years of their lives (in which they were only exposed to the L1) Basque speakers are mainly tuned to the distributional regularities of Basque language. This knowledge probably still influences Basque natives' anticipatory processes even if they have been exposed to the statistical regularities of Spanish on a daily basis afterwards. On the other hand, speakers originally tuned to Spanish process transparent and opaque items more similarly during language pre-activation, as a result of the more robust mapping of grammatical gender from lexical information in the Spanish language. Results along these lines have already emerged in experiments with words in isolation (Caffarra et al., 2014), and also in studies investigating the effect of transparency during determiner-noun agreement (Caffarra & Barber, 2015). It is also worth noting that the difference in predictive processing between the two groups is very unlikely to be due to a reduced proficiency for Basque natives in Spanish, since in the general proficiency assessment the two groups were fairly balanced in both languages, and also the behavioral results in both tasks did not reveal any significant differences.

Results on the Noun

Finally, the ERPs recorded on the incongruent noun showed a robust N400 effect. The effect was similar in both groups, in both tasks, in all the conditions. There was no significant difference and no interactions among the independent factors manipulated in this experiment. This lexical effect is coherent with several previous studies on bilingual processes (Hahne, 2001; Hahne & Friederici 2001; Weber-Fox & Neville, 1996), and confirms that integration semantic processes are not influenced by the native language experience, especially when proficiency is very high. Also, this outcome is in line with the behavioural results we report, where no difference between groups was found. Given the capacity of speakers to mediate between native language experiences, new language environment cues and task requirements, we propose that anticipatory language processing is extremely plastic and flexible in order to be able to "absorb" the possible processing costs of integration.

Conclusion

With the present set of data, we emphasise the plasticity of language anticipatory mechanisms. This study not only shows that language expectations can rapidly adapt to new contexts, taking advantage of all the available cues in order to construct an internal representation of the new environment, but also that early experience modulates the way in which different cues are weighted to obtain optimal predictions. The whole process is extremely flexible, ready to adapt to the environmental demands and to mediate between native experience settings and new language feature requirements.

Acknowledgments

This research was supported by the Spanish Ministry of Economy and Competitiveness (MINECO), Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER) [PSI2015-65694-P to N.M.]. The authors acknowledge financial support from the Spanish Ministry of Economy and Competitiveness, through the "Severo Ochoa" Programme for Centres/Units of Excellence in R&D"

(SEV-2015-490).

We wish to express our gratitude to the BCBL lab-staff and the research-assistants who helped to recruit the participants and collect the data.

Finally, we would like to thank Margaret Gillon-Dowens, James McQuilken, Ruth de Diego Balaguer, Irene Monsalve, 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. doi: 10.1016/S0010-0277(99)00059-1

Altmann, G. T., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science, 33(4)*, 583-609. doi: 10.1111/j.1551-6709.2009.01022.x

Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences*, *11(7)*, 280–289. doi: 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 feed forward and feedback influences through distinct frequency channels. *Neuron, 85(2),* 390–401. doi: 10.1016/j.neuron.2014.12.018

Bates, E., & MacWhinney, B. (1989). Functionalism and the Competition Model. In B. MacWhinney & E. Bates (Eds.), The crosslinguistic study of sentence processing (pp. 3—73). NewYork: Cambridge University Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International 5:9/10*, 341-345.

Boersma, P. & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 6.0.36, retrieved 11 November 2017 from http://www.praat.org/

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. doi: 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. doi: 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. doi:10.1016/j.neuropsychologia.2014.08.016

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology,14,* 177–208. doi: 10.1080/026432997381664

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

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

Clahsen, H. and Felser, C. (2006b). How native-like is non-native language processing? *Trends in Cognitive Sciences 10*, 564-570. doi: 10.1016/j.tics.2006.10.002

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

Coltheart, M., Rastle, K., Perry, C., Langdon, R., Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204–256.

DeLong, K., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8(8),* 1117–112. doi: 10.1038/nn1504

Dehaene, S. Duhamel, J., Hauser, M., and Rizzolatti, G. (2005) *From Monkey Brain to Human Brain.* A Fyssen Foundation Symposium. MIT Press.

Dell, G. S., & Brown, P. M. (1991). Mechanisms for listener-adaptation in language production: Limiting the role of the "model of the listener". In D. J. Napoli & J. A. Kegl (Eds.), *Bridges between psychology and linguistics: A Swarthmore Festschrift for Lila Gleitman* (Vol. 105, pp. 105–129). Psychology 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 of London.* Series B, Biological Sciences, 369(20120394), 1–9.

Dikker, S., & Pylkkänen, L. (2013). Predicting language: MEG evidence for lexical preactivation. *Brain and language*, *127*(1), 55-64. doi: 10.1016/j.bandl.2012.08.004

Dikker, S., Rabagliati, H., Farmer, T. A., & Pylkkänen, L. (2010). Early occipital sensitivity to syntactic category is based on form typicality. *Psychological Science, 21(5),* 629-634.

Dijkstra, A., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language, 41,* 496-518.

Duchon, A., Perea, M., Sebastian-Galles, N., Marti, A., Carreiras, M. (2013). EsPal: one-stop shopping for Spanish word properties. *Behavioral Research Methods, 45(4),* 1246-1258. doi: 10.3758/s13428-013-0326-1.

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44,* 491-505. doi: 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. doi: 10.3758/BF03193082

Federmeier, K. D., McLennan, D. B., De Ochoa, E., Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study. *Psychophysiology, 39,* 133-146. doi: 10.1017/S0048577202001373

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. doi:10.1037/a0036756

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

Friston, K. J. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11(2),* 127-138. doi: 10.1038/nrn2787

Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and Linguistic Compass, 3(1),* 128-156.
doi: 10.1111/j.1749818X.2008.00121.x

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition. A multiple read-out model. *Psychological Review, 103,* 518 –565.

Gollan, T., & Frost, R. (2001). Two Routes to Grammatical Gender: Evidence from Hebrew. *Journal of Psycholinguistic Research, 30,* 627–651.

Gollan T., Montoya R., Cera C, Sandoval T. (2008). More use almost always a means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language, 58(3)*:787-814. doi: 10.1016/j.jml.2007.07.001

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research, 30,* 251–266.

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition, 4(2),* 123-141. doi: 10.1017/S1366728901000232

Hasson, U., Chen, J., & Honey, C. J. (2015). Hierarchical process memory: memory as an integral component of information processing. *Trends in Cognitive Science, 19* (6), 304-313. doi: 10.1016/j.tics.2015.04.006

Hawelka, S., Schuster, S., Gagl, B., & Hutzler, F. (2015). On forward inferences of fast and slow readers. An eye movement study. *Scientific reports, 5.* doi: 10.1038/srep08432

Holcomb P.J., Grainger J. (2006). On the time-course of visual word recognition: An ERP investigation using masked repetition priming. *Journal of Cognitive Neuroscience*, *18*, 1632–1643. doi: 10.1162/jocn.2006.18.10.1631

Indefrey P., Levelt W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition 92,* 101–144. doi: 10.1016/j.cognition.2002.06.001

Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43(4),* 635-652. doi:10.1037/xlm0000315.

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

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press. p. 477.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: syntactic priming is affected by the prime’s prediction error given both prior and recent experience. *Cognition, 127(1)*, 57-83. doi: 10.1016/j.cognition.2012.10.013

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

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49,* 133–156

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. doi: 10.1162/jocn_a_00148

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

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307(5947),* 161-163.

Laka, I. (1996). *A Brief Grammar of Euskara, the Basque Language*. Euskal Herriko Unibertsitatea: Leioa-Donostia (Spain)

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behaviour Research Methods, 44 (2),* 325- 343 doi: 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. doi: 10.1016/j.cognition.2007.05.006

Luck S.J., Hillyard S.A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology*, *31*, 291–308

Luke, S. G., & Christianson, K. (2016). Limits on lexical prediction during reading. *Cognitive Psychology*, *88,* 22-60. doi: 10.1016/j.cogpsych.2016.06.002.

MacDonald, M. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*. doi: 10.3389/fpsyg.2013.00226

MacWhinney, B. (1987). Applying the competition model to bilingualism. *Applied Psycholinguistics,8,* 315-327. doi: 10.1017/S0142716400000357

MacWhinney, B. (1987b). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249-308). Hillsdale, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (1992). Transfer and competition in second language learning. In R. Harris (Ed.), *Cognitive processing in bilinguals* (pp. 371-390). Amsterdam: Elsevier.

MacWhinney, B. (1997). Second Language Acquisition and the Competition Model. In A. De Groot and J. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives (pp. 113-142)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Marslen-Wilson W.D. (1987). Functional parallelism in spoken word-recognition. *Cognition 25(1-2),* 71-102.

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. doi: 10.1016/j.jml.2013.08.001

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: part I: an account of basic findings. *Psychological Review, 88,* 375. doi:10.1037/0033-295X.88.5.375

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

Molinaro, N., Barber, H.A., Perez, A., Parkkonen, L., & Carreiras, M. (2013). Left fronto-temporal dynamics during agreement processing: Evidence for feature specific computations. *Neuroimage, 78,* 339-352. doi: 10.1016/j.neuroimage.2013.04.025

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. doi: 10.1016/j.neuroimage.2013.01.031

Molinaro, N., Giannelli, F., Caffarra, S., & Martin, C.D., (2017). Hierarchical levels of representation in language prediction: The influence of first language acquisition in highly proficient bilinguals. *Cognition, 164,* 61-73. doi: 10.1016/j.cognition.2017.03.012

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.

Morris, R. K. (2006). Lexical processing and sentence context effects. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed., pp. 377–402).

Peirce, JW (2007) PsychoPy-Psychophysics software in Python. *Journal of Neuroscience Methods, 162(1-2)*, 8-13.

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

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

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2(1),* 79-87. doi: 10.1038/4580

Rijk, R.P.G. (2008). *Standard Basque, a Progressive Grammar*. Cambridge MA: MIT Press.

Real Academia Española. (2009). Nueva gramática de la lengua española. Asociaciòn de academias de la lengua Española.

Roelofs, A., (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition 42,* 107-142

Strijkers K, Costa A (2011). Riding the lexical speedway: A critical review on the time course of lexical selection in speech production. *Frontiers in Psychology, 2,* 356.

Strijkers K, Costa A (2016). On words and brains: linking psycholinguistics with neural dynamics in speech production. *Language, Cognition and Neuroscience, 31(4)*, 524-35.

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin and Review, 9(4)*, 780-789.

Walter, W. G., Cooper, R., Aldridge, V. J., McCallum, W. C., & Winter, A. L. (1964). Contingent negative variation: an electric sign of sensori-motor association and expectancy in the human brain. *Nature*, *203*(4943), 380.

Weerts, T. C., & Lang, P. J. (1973). The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV). *Biological psychology*, *1*(1), 1-19.

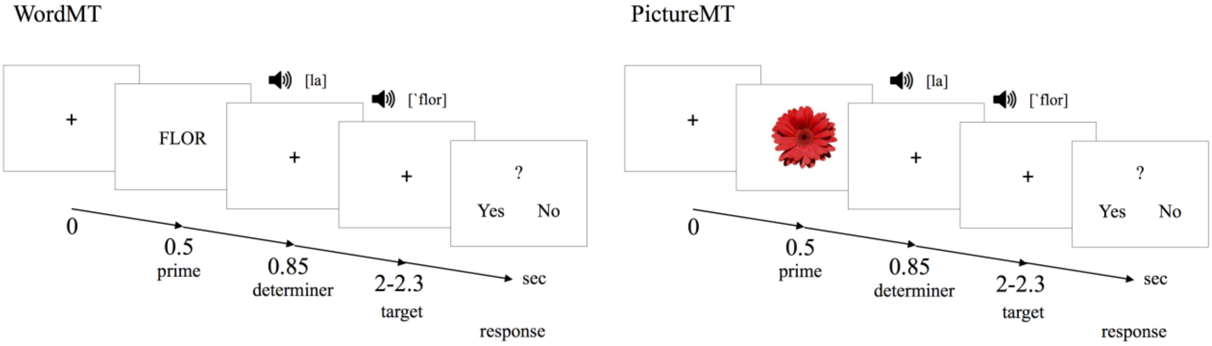
Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience, 16(7),* 1272– 1288. doi: 10.1162/0898929041920487

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

Measure	Spanish natives (N=21)	Basque natives (N=21)
Self-evaluation (0-10)		
Spanish	9.37 (0.29)	9.39 (0.24)
Basque	8.02 (0.49)	9.04 (0.16)
LexTALE		
Spanish (0-60)	52.76 (5.21)	54.09 (4.13)
Basque (0-50)	34.38 (5.88)	46.04 (2.67)
Picture naming (0-65)		
Spanish	64.47 (0.92)	63.38 (1.62)
Basque	50.09 (9.63)	64.19 (1.47)
Interview (0-5)		
Spanish	5 (0)	5 (0)
Basque	4.14(0.47)	4.95(1.33)

Table 2: ANOVA results of WordMT and PictureMT reaction time values

Measure	WordMT	PictureMT
Condition	F (4, 52) = 1.36, p = .251, η_p^2 = .43	F (4, 52) = .62, p = .649, η_p^2 = .16
Transparency	F (1, 38) = .76, p = .389, η_p^2 = 0.20	F (1, 38) = .29, p = .596, η_p^2<.01
Group	F (1, 38) = 1.36, p = .758, η_p^2 = .03	F (1, 38) = .33, p = .567, η_p^2<.01
Condition*Group	F (4, 52) = 1.32, p = .265, η_p^2 = .37	F (4, 52) = .79, p = .531, η_p^2 = .20
Transparency*Group	F (1, 38) = 2.16, p = .150, η_p^2 = .54	F (1, 38) = 89, p = .353, η_p^2 = .23
Condition*Transparency	F (4, 52) = 1.34, p = .260, η_p^2 = .24	F (4, 52) = .54, p = .706, η_p^2 = .14
Condition*Transparency*Group	F (4, 52) = .31, p = .869, η_p^2<.01	F (4, 52) = .60, p = .662, η_p^2 = .01



Stimuli WordMT	**Stimuli PictureMT**	**Features**	**D-match N-match** (26.7 %)	**nD-match N-match** (26.7 %)	**D-mismatch N-mismatch** (26.7 %)	**nD-match N-mismatch** (10 %)	**D-match N-mismatch** (10 %)
BOTELLA		f [+t]	LA BOTELLA	MI BOTELLA	EL CLAVO	SU CLAVO	LA MANZANA
FLOR		f [-t]	LA FLOR	SU FLOR	EL LAPIZ	MI LAPIZ	LA NUEZ
CUCHILLO		m [+t]	EL CUCHILLO	SU CUCHILLO	LA MANZANA	MI MANZANA	EL CLAVO
COCHE		m [-t]	EL COCHE	MI COCHE	LA FLOR	SU FLOR	EL LAPIZ

Figure 1: Upper panels: Example trials for the WordMT and the PictureMT experimental blocks. Lower panel: Experimental conditions of relative percentage of appearance in the experimental blocks.

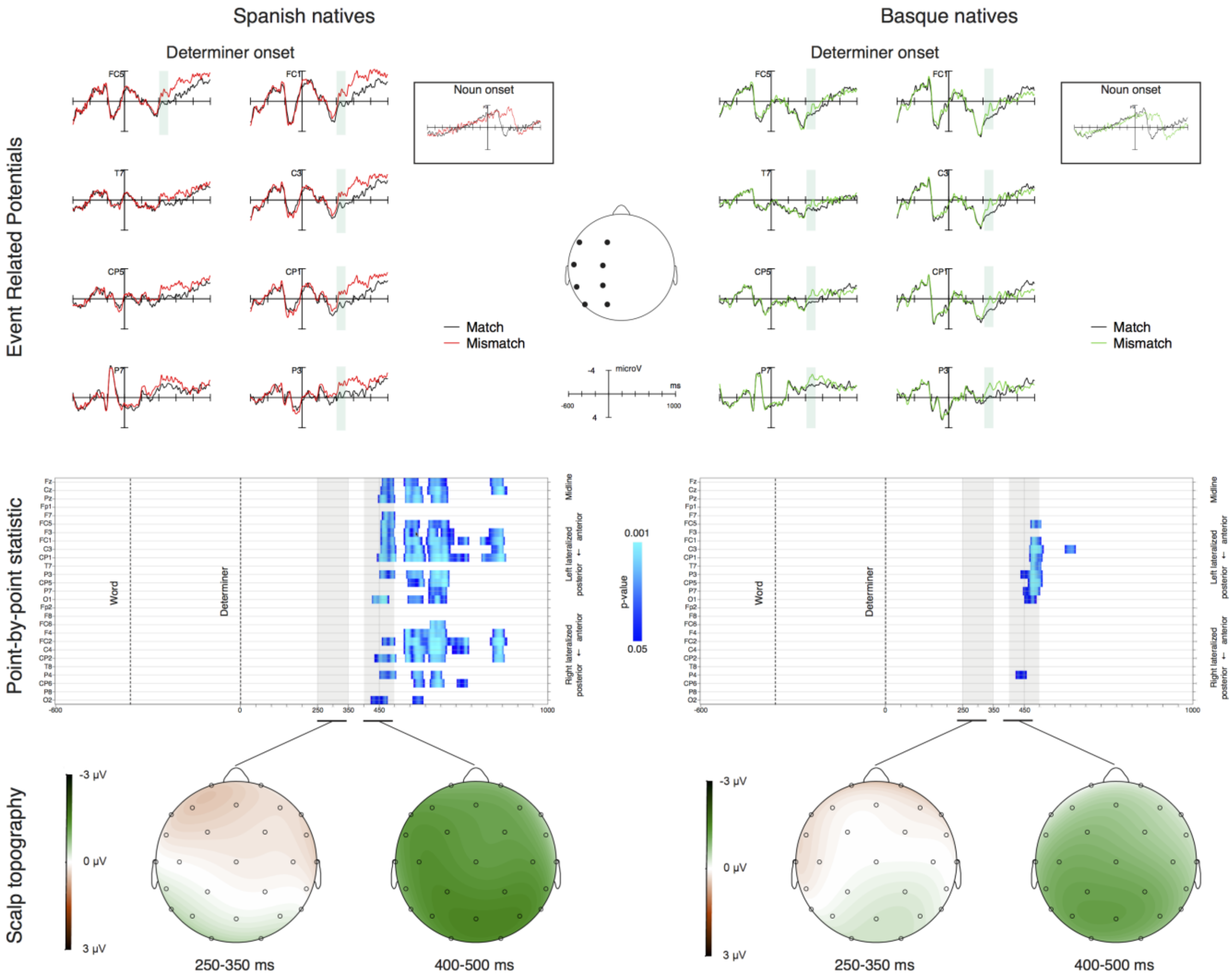


Figure 2: Upper panels: ERPs for the WordMT relative to the match and mismatch condition time-locked to the presentation of the determiner preceding opaque nouns (plotted in –600 to 1000 for better display of the ERP effects). A subset of left-lateralized electrodes showing relevant ERP modulations is reported for the two groups of participants. Shadowed differences show the statistically significant effects emerging from the point-by-point analyses below. Small rectangles on the right side of the graphs present the grand average ERPs relative to the match and mismatch condition time-locked to the presentation of the opaque noun (plotted in –1600 ms to 1500 ms) for a representative electrode.
Middle panels: The point-by-point plot t-test for each electrode (Guthrie and Buchwald, 1991, corrected) comparing determiner match and mismatch conditions. Vertical blue lines indicate the interval in which the effect emerged, with lighter color for statistically more significant effects.
Lower panels: The topographical distribution of the difference effect (mismatch minus match) in the early and late time intervals of interest (250-350 ms; 400-500 ms) corresponding to shadowed intervals in the point-by-point graphs.

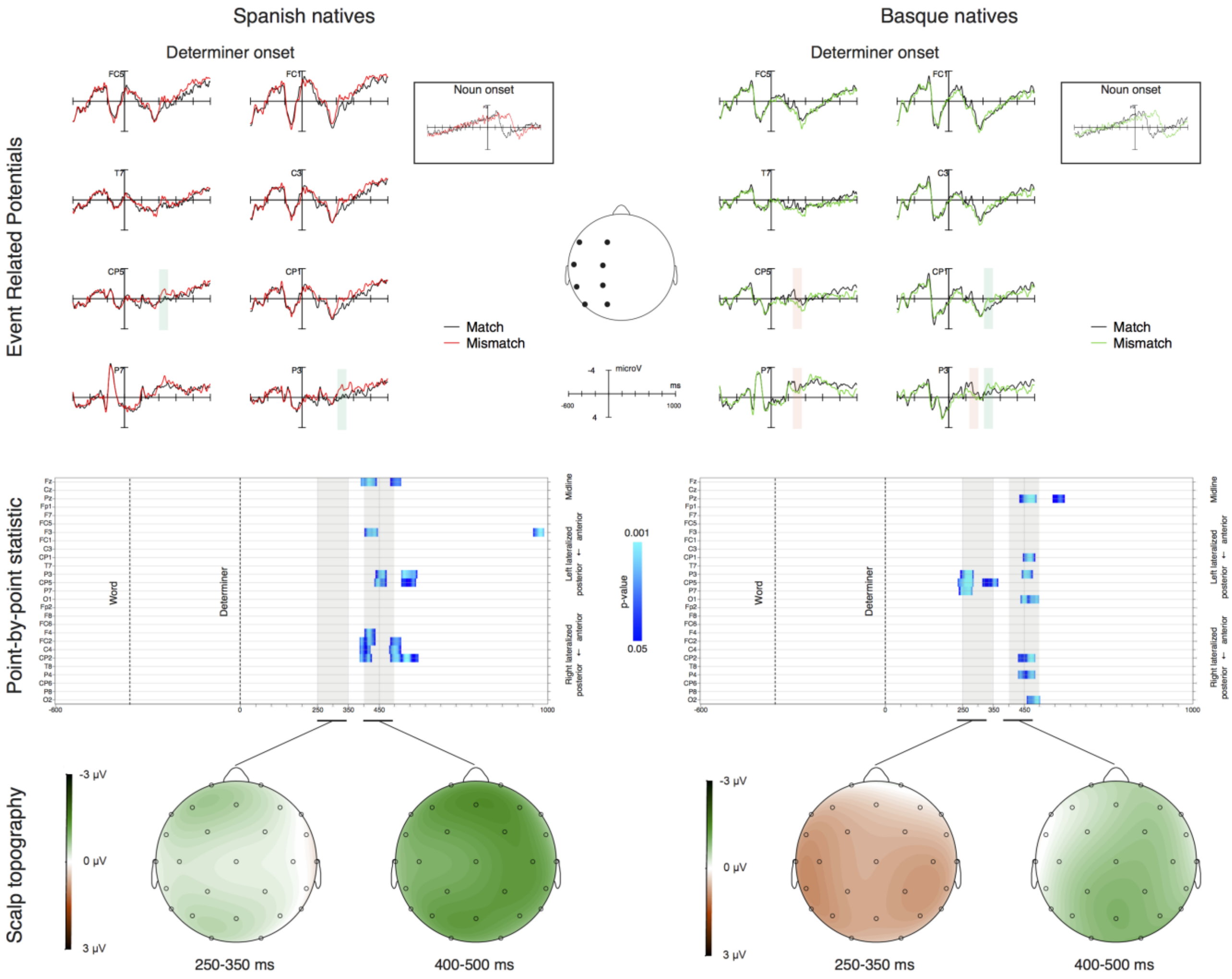


Figure 3: Upper panels: ERPs for the WordMT relative to the match and mismatch condition time-locked to the presentation of the determiner preceding transparent nouns (plotted in –600 to 1000 for better display of the ERP effects). A subset of left-lateralized electrodes showing relevant ERP modulations is reported for the two groups of participants. Shadowed differences show the statistically significant effects emerging from the point-by-point analyses below. Small rectangles on the right side of the graphs present the grand average ERPs relative to the match and mismatch condition time-locked to the presentation of the opaque noun (plotted in –1600 ms to 1500 ms) for a representative electrode.
Middle panels: The point-by-point plot t-test for each electrode (Guthrie and Buchwald, 1991, corrected) comparing determiner match and mismatch conditions. Vertical blue lines indicate the interval in which the effect emerged, with lighter color for statistically more significant effects.
Lower panels: The topographical distribution of the difference effect (mismatch minus match) in the early and late time intervals of interest (250-350 ms; 400-500 ms) corresponding to shadowed intervals in the point-by-point graphs.

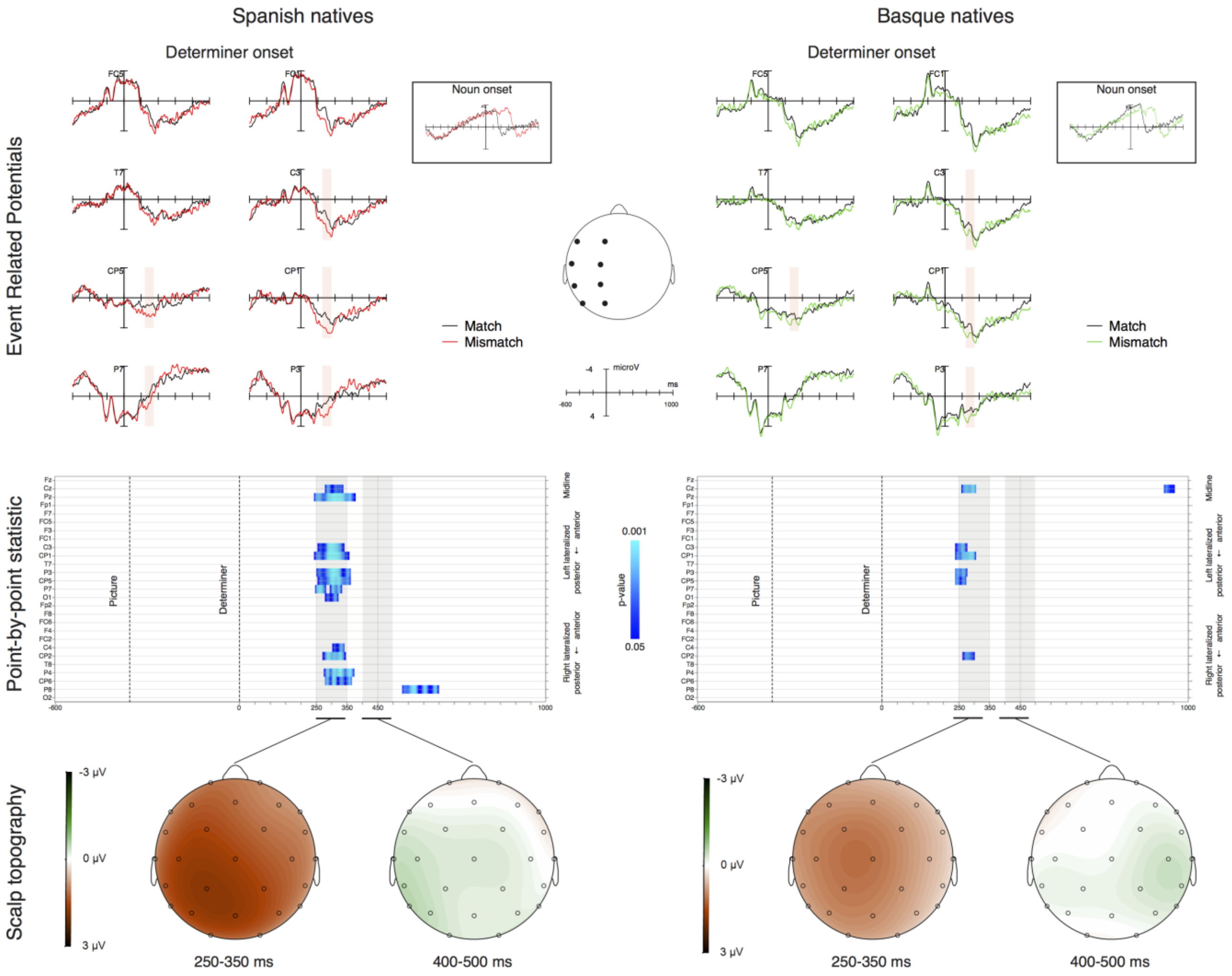


Figure 4: Upper panels: ERPs for the PictureMT relative to the match and mismatch condition time-locked to the presentation of the determiner preceding opaque nouns (plotted in –600 to 1000 for better display of the ERP effects). A subset of left-lateralized electrodes showing relevant ERP modulations is reported for the two groups of participants. Shadowed differences show the statistically significant effects emerging from the point-by-point analyses below. Small rectangles on the right side of the graphs present the grand average ERPs relative to the match and mismatch condition time-locked to the presentation of the opaque noun (plotted in –1600 ms to 1500 ms) for a representative electrode.
Middle panels: The point-by-point plot t-test for each electrode (Guthrie and Buchwald, 1991, corrected) comparing determiner match and mismatch conditions. Vertical blue lines indicate the interval in which the effect emerged, with lighter color for statistically more significant effects.
Lower panels: The topographical distribution of the difference effect (mismatch minus match) in the early and late time intervals of interest (250-350 ms; 400-500 ms) corresponding to shadowed intervals in the point-by-point graphs.

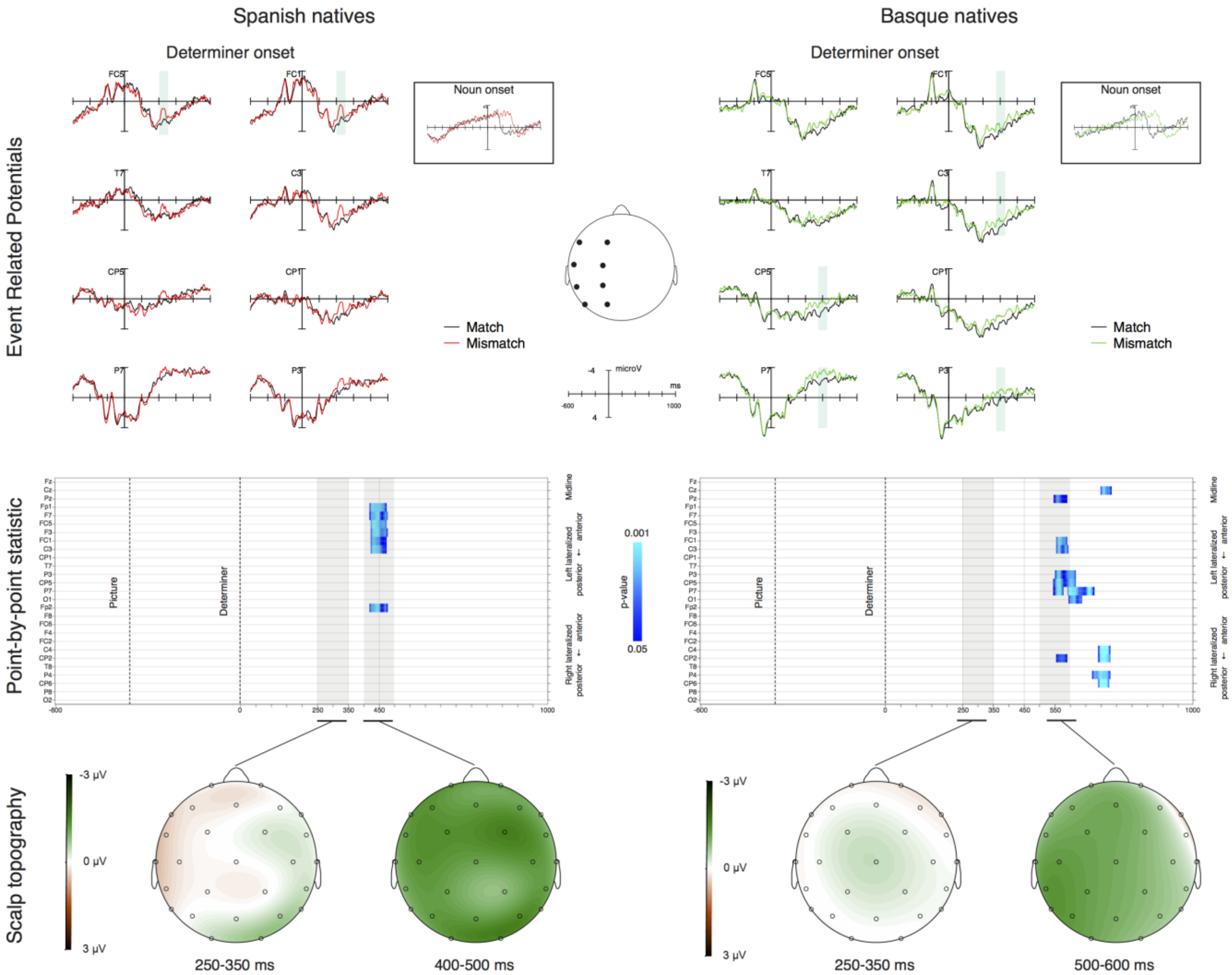


Figure 5: Upper panels: ERPs for the PictureMT relative to the match and mismatch condition time-locked to the presentation of the determiner preceding transparent nouns (plotted in –600 to 1000 for better display of the ERP effects). A subset of left-lateralized electrodes showing relevant ERP modulations is reported for the two groups of participants. Shadowed differences show the statistically significant effects emerging from the point-by-point analyses below. Small rectangles on the right side of the graphs present the grand average ERPs relative to the match and mismatch condition time-locked to the presentation of the opaque noun (plotted in –1600 ms to 1500 ms) for a representative electrode.
Middle panels: The point-by-point plot t-test for each electrode (Guthrie and Buchwald, 1991, corrected) comparing determiner match and mismatch conditions. Vertical blue lines indicate the interval in which the effect emerged, with lighter color for statistically more significant effects.
Lower panels: The topographical distribution of the difference effect (mismatch minus match) in the early and late time intervals of interest (250-350 ms; 400-500 ms for Spanish natives and 500-600 ms for Basque natives) corresponding to shadowed intervals in the point-by-point graphs.