Title: The Interplay between Semantic and Phonological Constraints during Spoken-Word
Comprehension

Authors: Angèle Brunellière^{1,2} and Salvador Soto-Faraco³

¹ Univ Lille Nord de France, F-59000 Lille, France

² UDL3, Unité de Recherche en Sciences Cognitives et Affectives, F-59653 Villeneuve d'Ascq, France

³ ICREA & Departament de Tecnologies de la Informació i les Comunicacions, Universitat Pompeu Fabra, Spain

Please address correspondence to:

Angèle Brunellière

Unité de Recherche en Sciences Cognitives et Affectives, Université Charles-de-Gaulle Lille 3, Domaine universitaire du Pont de Bois, BP 149, 59653 Villeneuve d'Ascq Cedex, France

Tel: (+33) 3 20 41 72 04

angele.brunelliere@univ-lille3.fr

Abstract:

The present study addresses how top-down predictions driven by phonological and semantic information interact on spoken-word comprehension. To do so, we measured event-related potentials (ERPs) to words embedded in sentences that varied in the degree of semantic constraint (high or low) and in regional accent (congruent or incongruent) with respect to the target word pronunciation. The data showed a negative amplitude shift following phonological mismatch (target pronunciation incongruent with respect to sentence regional accent). Here, we show that this shift is modulated by sentence-level semantic constraints over latencies encompassing auditory (N100) and lexical (N400) components. These findings suggest a fast influence of top-down predictions and the interplay with bottom-up processes at sub-lexical and lexical levels of analysis.

Count: 846 characters Max: 960 characters

Keywords: Predictive mechanisms, Context effects, spoken-word comprehension, event-related potentials

Introduction

The cortical organization of sensory systems has been proposed to rest upon the principle of hierarchical processing and the mechanism of predictive coding (Baldeweg, 2006; Friston, 2005; Friston & Kiebel, 2009). According to this assumption, incoming sensory input flows up from the initial stages of the hierarchy while top-down predictions constrain the sensory processing by acting at each level of the hierarchy from the level/s above. The particular case of top-down predictions during language comprehension has recently drawn researchers' attention (for reviews, Pickering and Garrod, 2007; Kutas, Delong, Smith, 2011; Hickok, 2012), because numerous studies have reported solid evidence about on-line top-down mechanisms (Brunellière & Soto-Faraco, 2013; Dambacher, Rolfs, Göllner, Kliegl, & Jacobs, 2009; Delong, Urbach, & Kutas, 2005; Molinaro, Barraza, & Carreiras, 2013; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; Wicha, Bates, Moreno, & Kutas, 2003; Wicha, Moreno, & Kutas, 2003, 2004). For example, it has been observed that high-level linguistic knowledge at sentence level, such as the semantic content, can constrain the recognition of upcoming words on-line. In line with some models of word recognition (McClelland & Elman, 1986, Marslen-Wilson & Welsh, 1978, Morton, 1979), the influence of sentence context on word recognition is interpreted as evidence for the pre-activation of candidate word forms, with an impact in lexical processing (e.g., Delong, Urbach, & Kutas, 2005; Dambacher, Rolfs, Göllner, Kliegl, & Jacobs, 2009). Nonetheless, the models of spoken-word recognition claiming for an influence of sentence context can vary about the actual locus of sentence context effects. More particularly, while models such as the Cohort model (Marslen-Wilson & Welsh, 1978) assume that the information flow is exclusively ascending from the feature level to the lexical level (i.e., bottom-up processes), the Trace model (McClelland & Elman, 1986) including three levels of representations (features,

phonemes, words) proposes both an ascending information flow from the feature level to lexical level and top-down feedback connections between the words and the phonemes and between the phonemes and the features. In addition, the Trace model proposes lateral inhibition connections at each level of representation. Regarding the possibility of top-down effects driven by sentence context, the Cohort model described that the top-down effect can act by reducing the activation of cohort candidates at lexical level. In contrast, according to the Trace model, top-down predictions of sentence context could operate at sublexical and lexical levels due its architecture. In fact, the level at which the top-down predictions of sentence context affect the bottom-up processing is still somewhat unknown.

Moreover, the nature and specificity of word form expectations generated by these top-down mechanisms is still far from clear. For instance, in an event-related potential (ERP) study, DeLong et al. (2005) showed that sentence context can lead to the generation of top-down expectations for specific word forms, so that the amplitude of the N400 responses to written words reflected the consequences of contextually based expectations at a lexico-semantic level (for previous studies, Kutas & Federmeier, 2000; Kutas & Hillyard, 1984). In particular, DeLong et al. (2005) reported that the amplitude of the N400 elicited by English indefinite articles, whose word form is only determined by the upcoming noun following a phonological regularity (e.g. “a kite”, “an airplane”), depended on the level of expectancy of highly expected but not yet presented nouns.

Another experimental evidence for top-down predictions of specific word forms driven by sentence context comes from our own recent ERP study in spoken-language comprehension (Brunellière & Soto-Faraco, 2013). We manipulated the regional Catalan accent of the prior sentence context and the phonological form of a sentence-final Catalan target word that was predictable from semantic context. In this way, the ERP effects of

phonological predictions generated by the regional accent of prior context could be measured. When the carrier context sentence was spoken in listener's native accent (Eastern Catalan), the ERPs revealed evidence for early detection of phonological mismatch (negative shift beginning around 250 ms, with respect to the expected regional word form). This early ERP modulation did not appear, when the carrier context sentence was spoken in the non-native regional accent of the listeners (Western Catalan). These results suggest that the phonological context (i.e., native vs. non-native regional accent) tuned the sensitivity to phonological mismatch and thus that the context can induce the generation of phonologically precise predictions. Based on our results, we contended that these predictions of phonologically specific word forms were driven by sentence-level context giving accent and semantic information, and not based on accent mode only, because of the use of strong semantically constraining sentences.

Other ERP studies (Van Berkum, van den Brink, Tesink, Kos, & Hagoort, 2008; Hanulíková, van Alphen, van Goch, and Weber, 2012) also provide converging findings suggesting that listeners seem to tune their predictions based on the speaker's characteristics during spoken language comprehension. However, an important question concerns how the predictions based on the speaker's characteristics, such as accent, can interact with the top-down predictions driven by the high levels of sentence context. For example, according to models assuming that multiple phonological variants of a single word are stored (Goldinger, 1998; Connine, Rambon, & Patterson, 2008), it can be hypothesized that strong sentence context could activate a specific phonological variant due to the phonological variability of context. That is, only highly constraining semantic contexts, but not low constraining ones, will give rise to such top-down predictions expressing in phonologically detailed expectations of word forms. In order to address this question, we conducted an ERP study in which, using an approach analogous to Brunellière & Soto-Faraco (2013), we manipulated the degree of

semantic constraint imposed by the carrier sentence context in addition to the phonological form of the target word, which could be consistent or inconsistent with the regional accent of the context. The manipulation of phonological form was based on the rule of vowel reduction that applies in Eastern, but not Western Catalan, to unstressed syllables (Alarcos, 1953; see also Brunellière and Soto-Faraco, 2013). Specifically, in Eastern Catalan (spoken in Barcelona), vowel reduction is applied in unstressed syllables, such that /a/, /ɛ/ and /e/ segments become a schwa /ə/, and /o/ and /ɔ/ are reduced to /u/. These vowel reductions do not apply in the Western Catalan accent. Instead, in Western Catalan accent only /ɛ/ and /ɔ/ segments are reduced to /e/ and /o/, respectively. These regional variations in vowel realization thus lead to clearly distinct phonological word forms between Eastern and Western Catalan. For example, *tercer* (third) is spoken /tərsé/ vs. /tərsé/, respectively. In the present study, native Eastern Catalan speakers listened to target words embedded in strongly or weakly semantically constraining sentence frames. The sentence frame was spoken in Eastern Catalan (the listeners' native accent) whereas the phonological form of target words could be in Eastern (congruent accent) or Western (incongruent) regional accent. Examples of the experimental stimuli are displayed in Table 1.

<Insert Table 1 here>

According to prior ERP literature in spoken-language comprehension (van den Brink & Hagoort, 2004; van den Brink, Brown, & Hagoort, 2001), the N100 component mostly reflects bottom-up processing of the incoming signal, based on the acoustic properties of word onset. According to this view, the top-down processes driven by sentential context only begin to influence word recognition as early as 200 ms from the incoming word, as revealed by differential ERP responses in amplitude between strongly and weakly constraining contexts (Connolly, Phillips, Stewart, & Brake, 1992; Connolly, Stewart, & Phillips, 1990). Some

authors, (Connolly & Phillips, 1994; Connolly, Phillips, Stewart, & Brake, 1992; Connolly, Stewart, & Phillips, 1990; Hagoort & Brown, 2000) proposed that the N200, occurring between 150 and 300 ms, is triggered by the mismatches at a phonological level with the lexical expectancies from the sentence context, whereas the N400, peaking around 400 ms after stimulus onset with a more posterior distribution across the scalp, reflects the consequences of contextually based expectations regarding upcoming words at a lexico-semantic level. Additionally, ERP studies in written language have shown that the degree of semantic constraint does not modulate the N400 amplitude during the processing of semantically plausible words when they are not the most expected word for sentence (Kutas and Hillyard, 1984; Thornhill and Van Petten, 2012; Van Petten and Luka, 2012). However, the dissociation between the N200 and the N400 responses in term of scalp distribution is not always clear (Brunellière and Soto-Faraco, 2013; Diaz and Swaab, 2007; van den Brink and Hagoort, 2004). This suggests that the neural underpinnings are similar in the two time windows, such that the N200 component is not always functionally separable from the N400. This conclusion is in accordance with the study of Van Petten, Coulson, Rubin, Plante, & Parks (1999) showing one unique N400 wave, triggered by the top-down predictions due to the sentence context.

Regarding the effects of semantic constraint, and based on Connolly and colleagues' results (Connolly, Phillips, Stewart, & Brake, 1992; Connolly, Stewart, & Phillips, 1990), target words embedded in low semantically constraining sentence frames should evoke a response with greater negative amplitude in comparison to the same targets when embedded in high constraining sentence frames (hence, more predictable). This shift may begin as early as 250 ms (putatively, the N200 or early N400 effect) and persist over later processing stages around the N400 window. In addition, based on our previous study (Brunellière & Soto-

Faraco, 2013), we predicted an early negative shift beginning around 250 ms for incongruent phonological word forms with respect to the expected regional word forms, at least in highly semantically constraining sentences (the ones tested in the cited study). Yet, the crucial question here is the interplay between the two factors; the degree of semantic constraint (high or low) and the phonological form of target word (congruent vs. incongruent) regarding the regional accent of context. If top-down predictions incorporate detail of the phonological form down to the regional accent and semantic information corresponding to the context, then we should observe interactive effects between the congruency of regional word form and the semantic constraint of sentence context. For instance, the amplitude of brain responses associated with the phonological mismatch should vary as a function of the degree of semantic constraint.

Methods

Participants

Twenty-four dominant Catalan-speaking students from the Pompeu Fabra University, aged between 18 and 27 years (18 female, mean age: 21, SD age: 2.1) were selected. None reported hearing or language impairments. Catalan was the most frequent language spoken in their everyday life and the only language spoken with their family. The participants came from the Barcelona province and were raised in Eastern Catalan speaking families. The participants' parents were native Eastern Catalan speakers as well and the language spoken with their friends was Eastern Catalan. However, participants could be exposed to Western Catalan via the media (TV or radio) or their occasional interactions with speakers from other Catalan-speaking regions. All were right-handed as assessed by the Edinburgh handedness inventory

(Oldfield, 1971). They received monetary compensation for participation (10€/h). Before the beginning of the experiment, participants gave their written informed consent.

Materials

The experimental stimuli consisted of a set of 224 pairs of sentence frames that were either strongly or weakly semantically constraining and spoken in Eastern Catalan accent (the listeners' native accent) up until the penultimate word. Both sentences in a pair ended with the same target word but phonologically differed in terms of target's regional word forms. The manipulation of contextual semantic constraint (high/low) and the congruency of phonological form of the target (Eastern Catalan accent as congruent forms, or Western Catalan, as incongruent forms) made four experimental conditions (see Table 1): High semantic constraints and congruent phonological word forms (HS-CP), High semantic constraints and incongruent phonological word forms (HS-IP), Low semantic constraints and congruent phonological word forms (LS-CP) and Low semantic constraints and incongruent phonological word forms (LS-IP).

The selection of strongly and weakly semantically constraining sentence frames (112 each) was based on the classical cloze procedure, during which a group of Catalan speakers (different from the ones tested in the ERP experiment) were asked to complete the sentence fragments with the first word that came to mind. Four lists of at least 100 sentence frames were constructed and each one was completed by fifteen participants. As already used in Brunellière & Soto-Faraco (2013), the final (target) word in the highly constraining sentence frames had an average cloze probability of 0.78 (range: 0.53 to 1.00) and was the most expected word in that sentence. The final words in the lowly constraining sentence frames had an average cloze probability of 0.08 (range: 0.00 to 0.33), and were always semantically plausible. The length in terms of number of words (mean: 13.5, range: 7-21) and the number

of phonemes marking the regional accent delivered in sentence frames (mean: 6.4, range: 3-15) were matched between the highly and lowly constraining sentence frames (see Table 2). The phonologically congruent and incongruent word form conditions were equated in terms of the target words' cloze probability, the length of the sentence, and the markers of regional accent in sentence frame. More specifically, in the incongruent phonological word form conditions, the phonological form differed from the phonologically expected word only in its second phoneme, which always indicated that the word was produced in the non-native Western accent (see Table 1). The phonological variation based on reduction rule in unstressed syllables was produced by variation in two phonemic categories (/ə/ vs. /e/ or /u/ vs. /o/).

As a consequence, words beginning with an unstressed syllable, on which regional variation in Catalan occurs, were selected from the Catalan Dictionary of Frequencies database (Rafel i Fontanals, 1998) as target words. The target words were 2 or 3 syllables long and were not subject to regional phonological variations on the second or third syllable. The target words had a mean frequency of 873 tokens per million, a mean number of phonemes of 5.9, and belonged to various lexical categories such as noun, verb, adverb, or adjective. None of the target words was repeated in sentence frames. All target words began with a plosive segment (/p/, /t/, /b/, /d/) to provide a clear physical marker on the spectrogram, which made it possible to easily align the onset of targets for the ERP recordings.

We avoided exposing participants to repeated presentations of same sentence frame or target word. For each participant, only one version of each target word (either its congruent or incongruent phonological form) embedded in one of the two various sentence frames (either high or low semantically constraining) was presented. To this end, we constructed four equivalent experimental lists of 28 trials per condition and all experimental conditions (HS-

CP, HS-IP, LS-CP and LS-IP) were equally represented in each list. In addition to the experimental stimuli, fifty-six filler sentences were created to avoid the development of strategies based on the two first phonemes of words and to reduce the proportion of incongruent phonological word form trials (around 30%, overall). In these fillers, the target words began with any other phonemes except for /p/, /t/, /b/ and /d/ (the initial phonemes of experimental stimuli), and the phonological forms were congruent with the sentence accent.

For recording, all stimuli were produced several times by a dominantly Catalan-speaking female with a native Eastern Catalan accent and were digitized at a sampling rate of 44 kHz with 16-bit. The speaker was asked to pronounce the sentences with natural prosody at normal speaking rate. To make sure that intonation and speaking rate were kept constant between experimental conditions, the four experimental conditions within one target word were recorded one after the other. The order of experimental stimuli within each target word was counterbalanced to avoid an effect of first reading in a particular experimental condition. The incongruent phonological word form was created such that the speaker was instructed to apply the phonological features of Western Catalan during the production of the final target word. We selected the auditory sentences as a function of the natural intonation and speaking rate, and correct pronunciations of phonological features of Western Catalan in the incongruent phonological word form conditions. The selection of correct pronunciations for the word forms in each experimental condition was determined by a native Eastern Catalan speaker. The selection of auditory sentences, the determination of sentence onsets and offsets and the extraction of acoustic values were performed by using the speech editing software, Praat (version 5.3; Boersma and Weenink, 2011). The total duration of sentence context (up to the onset of the final word) and that of the target word, was similar across the experimental conditions (see, Table 2). Acoustic measures (duration, maximum intensity, standard deviation of fundamental frequency and mean fundamental frequency) of the critical vowel

(used to create the phonological manipulation) indicated no statistical differences in these acoustic parameters between the experimental conditions (see, Table 3).

<Insert Tables 2 and 3 here>

Experimental procedure

After 2000 ms inter-trial interval, a fixation cross appeared at the centre of the monitor 500 ms before the onset of one auditory sentence and remained there until 1000 ms after the end of sentence. The sentences were presented binaurally at a constant, comfortable sound pressure level via headphones. To minimize artefacts, participants were asked to maintain their gaze on the fixation cross and try to keep their eyes as still as possible. They were encouraged to try and make any movements for comfort when the fixation cross disappeared (after every trial). Participants listened to 12 practice sentences prior to the set of four 8-min blocks of 42 trials, each containing sentences from all experimental conditions and fillers presented in random order. During the experiment, participants were instructed to listen to the sentences attentively for comprehension, without any further tasks (for similar approaches, Hagoort and Brown, 2000, van den Brink and Hagoort, 2004, van den Brink, Brown and Hagoort, 2001).

EEG recording and analyses

The EEG signal was recorded using a 31-electrode setup mounted on an elastic cap and referenced to the tip of the nose. The passive channels were distributed over the head cap according to the 10% standard system of the American Electroencephalographic Society (see Figure 1). Two electrodes placed close to the right eye were used to record the eye movements. The activity of electrodes over the right and left mastoids was also measured. Electrode impedance was kept below 5 kOhm. The electrophysiological signal was digitized

at 500 Hz and filtered on-line with a 0.1-100 Hz bandpass filter. EEG epochs started 100 ms before and lasted until 700 ms after the onset of target words. Each epoch was corrected to a 100-ms pre-baseline and filtered off-line with a 1-30 Hz bandpass filter and a 50 Hz notch filter. ERP waveforms were calculated for each participant, experimental condition and electrode. Trials containing artefacts were removed under a rejection criterion of $\pm 70 \mu\text{V}$ at any channel within the time window of epochs. All ERP waveforms were time-locked to the onset of auditory target word and were composed of at least 25 epochs for a particular participant and condition. The number of accepted EEG epochs (calculated from the sum of accepted EEG epochs per participant and condition) was equal across experimental conditions (mean and rate values for each condition, HS-CP, 26.7, 95.4%; HS-IP, 26.5, 94.6%; LS-CP, 26.6, 95% and LS-IP, 26.7, 95.4%). For each participant, bad channels were interpolated (Perrin, Pernier, Bertrand, Giard & Echallier, 1987) and an average mastoid reference (left and right) was applied off-line. The overall proportion of interpolated data is 8.3% and a maximum of five channels was interpolated per participant.

The ERP analyses were conducted on three negative ERP components known to be associated with auditory word processing, the N100, N200 and N400. Based on the prior literature and visual inspection, we quantified the mean amplitude of each putative ERP component across participants and experimental conditions in the following time windows: 110-160 ms (N100), 220-300 (N200) and 350-600 (N400). For example, for the N400 time window, the segment contains the maximum peak amplitude at Pz and large standard time range (van den Brink, Brown and Hagoort, 2001; van den Brink and Hagoort, 2004). A four-way repeated measures ANOVA was computed on the mean ERP amplitude over each time window with the following factors: Semantic constraint (2: High vs. Low), Congruency of phonological word forms (2: Congruent with prior context accent vs. Incongruent), Anterior/Posterior (2:

Anterior vs. Posterior) and Laterality (3: Left, Midline, Right). The anterior/posterior and laterality topographical factors¹ defined eighteen scalp sites for the statistical analysis: Left Anterior: F3, FC5, C3; Midline Anterior: Fz, FC1, FC2; Right Anterior: F4, FC6, C4; Left Posterior: P3, CP5, PO1; Midline Posterior: CP1, CP2, Pz; Right Posterior: P4, CP6, PO2. The two topographical factors were chosen to provide appropriate scalp coverage for the components of interest (Kutas and Federmeier, 2000; Van den Brink & Hagoort, 2004). When there was more than one degree of freedom in the numerator, the Greenhouse-Geisser correction was applied to adjust for violations of sphericity (Greenhouse and Geisser, 1959); the corrected *p*-values are reported. When a significant interaction was found, pairwise *t*-tests were performed to interpret the significance of effects. To ensure that possible shifts over the 100-ms pre-stimulus baseline did not affect the pattern of results measured from the onset of the target words, we conducted a four-way repeated measures ANOVA (same factors as previously described) over the -100 to 0 ms time window. This statistical analysis did not reveal any main effects or interactions. To examine further potential pre-stimulus effects caused by acoustic differences due to co-articulation, we conducted paired sample-by-sample *t*-tests comparing the ERPs to congruent and incongruent phonological forms of the target words over the 500-ms time window before target onset, at each level of semantic constraint. The *t*-tests did not reveal any significant difference at any electrode over the scalp (under the significance criterion of $p < 0.05$ for consecutive 50 ms or longer). Permutation tests (obtained by randomly classifying our ERPs as congruent vs. incongruent phonological forms of the target words within each participant) revealed the same results as paired sample-by-sample *t*-tests at the alpha-level 0.05 (for similar approach, Maris and Oostenveld, 2007).

Results

<Insert Figures 1, 2 and 3 here>

Grand-average waveforms corresponding to the phonologically congruent and incongruent target word forms are shown separately for each level of semantic constraint (see, Figures 2 and 3). The topographical effects of phonological mismatch are displayed in Figure 6 at each level of semantic constraint. The target words in all conditions elicited the typical auditory N100 responses to onset, albeit reduced in amplitude. This is not surprising since we are measuring ERPs to words embedded in continuous speech. Indeed, previous ERP studies in the auditory domain (e.g., Connolly, Phillips, Stewart, & Brake, 1992; Hagoort and Brown, 2000; Näätänen and Picton, 1987) showed that the lack of a sufficiently long pause interval before target words to span the refractory period of the N100 caused its small amplitude.

In the N100 time window, the ANOVA revealed a significant interaction between the Semantic constraint, Congruency of phonological word form and Anterior/Posterior factor [$F(1,23)=8.49$, $p<.01$]. We thus investigated the critical interaction between the semantic constraint and the congruency of phonological word form at each level of the Anterior/Posterior factor² (see, Table 4). Interestingly, in low semantic constraining sentence, both the anterior and posterior sites showed a larger N100 for the incongruent phonological form of target words in comparison to the congruent forms (anterior sites, $t(23)=2.49$, $p<.05$; posterior sites, $t(23)=2.9$, $p<.05$). In contrast, no significant difference was found between the incongruent and congruent phonological form of target words in high semantic constraints at any sites (anterior sites, $t(23)=0.97$; posterior sites, $t(23)=0.83$). The significant interaction with the Anterior/Posterior factor seemed to emerge because over anterior sites we did not observe an effect of semantic constraint for incongruent and congruent phonological forms,

while over posterior sites, an effect of semantic constraint was found only for incongruent phonological forms (see, Table 4).

In order to establish the onset of the effects of phonological mismatch over the N100 time window, we performed paired *t*-tests between the ERPs to congruent and incongruent phonological target forms for low semantic constraints. Each time frame was tested separately for each level of the Anterior/Posterior factor under the 0.05 alpha criterion for at least consecutive 50 ms. This revealed an effect of phonological mismatch started as soon as 80 ms after target onset over posterior sites and somehow later (around 140 ms after target onset) over anterior sites. When the same analysis was done for high semantic constraints, no significant difference was found over the N100 time window. Nonetheless, phonological mismatch effects were observed from 240 ms after target onset over both anterior and posterior sites. Permutation tests revealed the same latency of phonological mismatch effects in strongly and weakly constraining contexts.

<Insert Table 4 here>

In regard to the N200 time window, there were significant main effects of Semantic constraint [$F(1,23)=11.83$, $p<.01$] and Congruency of phonological word form [$F(1,23)=11.40$, $p<.01$]. As is commonly reported in prior literature, target words embedded in low semantically constraining sentence frames presented larger N200 amplitude with respect to those embedded in high semantically constraining sentence frames. Interestingly, whatever the level of semantic constraints, the incongruent phonological form of target words elicited an increase in the N200 amplitude in comparison to the congruent phonological form. Therefore, there was no significant interaction between phonological congruency and semantic constraint [$F(1,23)=0.25$] in this time window. No other interactions were significant in this analysis.

Similarly to the N200 time window, the N400 time window revealed a main effect of Semantic constraint, with larger amplitude for targets in the low constraining sentences compared to targets in high constraining ones [$F(1,23)=14.47$, $p<.001$]. This effect was stronger over posterior sites, as revealed by a significant Semantic constraint $\times$ Anterior/Posterior interaction [$F(1,23)=19.42$, $p<.001$]. The N400 amplitude associated with the effect of semantic constraint was larger over posterior sites than anterior sites ($t(23)=5.42$, $p<.05$). Finally, a significant Semantic constraint $\times$ Congruency of phonological word form interaction was observed [$F(1,23)=7.97$, $p<.01$]. Pairwise tests revealed that, for this time window, the effect of Phonological Congruency was present when targets appeared under low but not under high semantic constraints (see, Table 5, low semantic constraints, $t(23)=2.71$, $p<.05$, high semantic constraints, $t(23)=0.71$). Indeed, the phonologically incongruent word forms produced larger N400 amplitude as compared to their respective phonologically congruent forms, but only when they were embedded in low semantically constraining sentence frames. The phonologically incongruent and congruent word forms in high semantically constraining sentence frames presented the same size of N400 amplitude.

<Insert Table 5 here>

<Insert Figures 4, 5 and 6 here>

Topographical analyses

To sum-up, as commonly observed in prior studies, the semantic constraint affected the processing of spoken words as early as 250 ms, in the form of a negative shift, and this effect carried over to later stages. Grand-average waveforms corresponding to the low and high semantic constraints are shown separately for each phonological form of target words (see, Figures 4 and 5). Topographical analyses of semantic constraint between the time window of

the (putative) N200 and the traditional large time window of the N400 were conducted. To examine whether the effect of semantic constraint elicited different scalp topographies between these two time windows, we calculated ERP difference waves between the low and high constraining conditions for each participant when the phonological forms were congruent in the time windows of the N200 and the N400. To avoid differential amplitude effects between the two time windows, we extracted the mean amplitude of the difference waves from each scalp electrode after normalization to the global field power (GFP) (see, Murray, Brunet and Michel, 2008) within the N200 and N400 time windows. An ANOVA was conducted on mean amplitude with the following factors: time window (N200 vs. N400) and electrode (31 electrodes). Similar to Van den Brink and Hagoort (2004), no topographical differences from all electrodes across the scalp were found in these two time windows [$F(1,23)=0.05$]). Therefore, like in prior studies, this experiment does not provide evidence for two separable ERP components and because of this, these two time windows will thus be referred to as a single component that is triggered by sentence-level semantic constraint and carries over various stages. We will label this component hereafter early and late segments of the N400 (see, Figures 4 and 5).

In regard to the topography of effects triggered by the phonological mismatch, the negative shift elicited by the phonological manipulation presented similar scalp topography across the various time windows and regardless of the level of semantic constraint (see, Figure 6). To perform the topographical analyses on the phonological mismatch, we calculated ERP difference waves between the incongruent and congruent phonological forms at each level of semantic constraint for every participant. Then, we extracted the mean amplitude of ERP difference waves from all electrodes across the scalp after normalization to the GFP for each time window of interest. Based on the mean amplitude in low semantically

constraining contexts, an ANOVA was computed including the factors, time window (N100, early and late stages of N400) and all scalp electrodes. This analysis did not reveal any significant topographical difference across the three time windows [$F(2,46)=0.28$]. An additional ANOVA was performed over the early stage of N400 (220-300 ms), with the factors, semantic constraint (low vs. high) and all scalp electrodes. This topographical analysis also indicated no significant difference between the low and high semantically constraining contexts [$F(1,23)=0.16$]. Moreover, the two ANOVAs (i.e. the ANOVA based on the three time windows in low semantically constraining contexts and that over the early stage of N400) showed an effect of electrode factor (respectively, $F(30,690)=2.75$, $p<0.05$; $F(30,690)=2.95$, $p<0.05$). As suggested by the topographical maps in Figure 6, this indicates that the amplitude of negative shift was stronger over Cz, Pz, CP1 and CP2 electrodes. Consequently, the negative shift associated with the phonological mismatch (i.e., the incongruence of regional word forms) presented quite similar spatial distribution to the classical effect of semantic constraint known to be enhanced over centroparietal sites. Finally, following the same logical approach as the semantic topographical effects, we could describe one single component N400 associated with the phonological mismatch over various time windows that we labelled very early, early and late segments, respectively (see, Figures 2, 3, 4 and 5).

Discussion

The present study focused on the neural signature of the interplay between predictive mechanisms based on two different levels of linguistic analysis: phonological (manipulated by producing a mismatch in regional accent) and high-level semantic constraints imposed by sentence context. Based on models assuming storage of multiple phonological variants for a single word (Goldinger, 1998; Connine, Rambon, & Patterson, 2008), we hypothesized that if top-down predictions express in terms of specific phonological word forms, brain responses to regional phonological mismatch should be influenced by the word's expectancy defined by the semantic constraint. We reported that brain responses to phonological mismatch (in the form of a negative shift, like the N400 component) depended on expectancy arising from semantic context. When the sentence context was less constraining, the phonological mismatch brain response was stronger such that the response started sooner with an early expression around 100 ms, and persisted later over a time window encompassing the classical N400 response. Instead, when the sentence context was highly constraining, the phonological mismatch brain response was of briefer duration, present in the 220-300 ms latency window only. Consequently, our results are in line with past ERP studies reporting insensitivity of the N400 amplitude to the strength of high-level semantic constraint during the processing of plausible word forms but when these forms were not the most expected (Kutas and Hillyard, 1984; Thornhill and Van Petten, 2012; Van Petten and Luka, 2012). Although stronger semantic constraint may have led to stronger predictions about the upcoming word, this did not lead to a larger phonological mismatch effect, but instead a later and shorter-lived effect. One potentially plausible interpretation of this result is that an unexpected change in regional pronunciation made it difficult to identify the final word when only weak sentence context was available. Phonological mismatch was more rapid when the word benefits from strong

semantically predictive context. Since the phonological mismatch elicited a single continuous brain response but occurring at various timing based on the input, we discuss below our findings according to past literature describing two stages of word processing.

Sub-lexical level: Early stages of word processing

One particular novel and somewhat unexpected aspect of the present results is the finding of top-down prediction effects at early stages of input analysis. Indeed, we found phonological mismatch effects within a time window around 100 ms after the presentation of target words only for low semantically constraining contexts. This is surprising given that this time window is known to be the classical time window of N100 component, reflecting exclusively the bottom-up processes operating at sub-lexical level. In particular, the N100 has been shown to be an index of perceptual processing of auditory input and phonological processing (Krumbholz, Patterson, Seither-Preisler, Lammertmann, Lütkenhöner, 2003; Näätänen, 2001; Obleser, Scott, Eulitz, 2006). We could envisage that this early finding could have been facilitated by co-articulatory cues³ before the presentation of target words. Moreover, crucially to our study, the critical vowel that we used to create the phonological mismatch did not differ in duration, intensity, fundamental frequency (F0) between low and high semantic constraining contexts. In addition, the euclidean distance in the F1-F2 plane between the matching and the mismatching vowel in each sentence at the acoustic mid-point of the vowel did not vary as a function of the level of semantic constraint. Consequently, the earlier detection of the phonological mismatch in low semantic constraining contexts cannot

result from differential acoustic realization of the critical vowels in low semantic constraining contexts compared to high semantic constraining contexts.

Nonetheless, the finding of early effects of the preceding context is not absolutely unheard-of, since previous magnetoencephalographic (MEG) studies (Bonte, Parviainen, Hytönen, & Salmelin, 2006; Uusvuori, Parviainen, Inkinen, & Salmelin, 2008) exploring the latency of the top-down influence of context on phonological forms have found N100 amplitude modulations after the presentation of isolated spoken words or syllables. Bonte and colleagues (2006) reported that syllables preceded by linguistic context (words or sentences) elicited weaker N100 amplitude in the left hemisphere than syllables not preceded by a context. Uusvuori and colleagues (2008) presented a semantic or phonological context composed of three words and followed by a final word in accordance or not with this context. In particular, they observed reduced N100 amplitude after a phonologically-related final word compared to an unrelated word when the context was phonologically constraining. These MEG studies thus did not investigate how the top-down predictions driven by phonological and semantic constraints interacted during spoken-word comprehension. However, note that the analysis of the MEG signal at the source level offers a good opportunity to probe the influence of top-down processes at early stages (Uusvuori, Parviainen, Inkinen, & Salmelin, 2008).

Beyond electrophysiological studies in spoken language, the modulation effect of top-down predictions driven by high levels on bottom-up processing at putatively sub-lexical levels is in agreement with ERP studies in written sentence contexts, showed that top-down predictions can operate quite early, at orthographic low-level forms (Dambacher, Rolfs, Göllner, Kliegl, & Jacobs, 2009; Kim & Lai, 2012; Molinaro, Barraza, Carreiras, 2013). For instance, ERP effects of top-down predictions driven by semantic constraint of sentence

context have been found in word reading at latencies of 100 to 200 ms, known to be the moment at which the initial processing of low-level visual features arises. Interestingly, in our study using auditory presentation, top-down predictions imposed by semantically constraining contexts could have prevented the emergence of early phonological mismatch brain responses. One potential explanation of this pattern could reside on the overall reduction of electrophysiological response for the incongruent phonological forms upon the recognition of expected initial phoneme due to the high constraining contexts. In contrast, when the context is less semantically constraining, the bottom-up processing of perceived word onset interacts only with the phonological predictions due to the regional accent of context at sublexical levels. In consequence, the detection of phonological mismatch in the case of incongruent forms is not prevented by the semantic information. In a fully interactive framework, the top-down predictions could influence the processing of sublexical levels in spoken language, when the context is constraining enough. This account is compatible with the theoretical assumption that the high levels in context information by top-down predictions can exert an influence at early stages of word recognition, such as the perceptual analysis (McClelland and Rumelhart, 1981).

Lexical level: later stage of word processing

Concerning processing stages subsequent to 100 ms after presentation of target words, prior studies have argued that after the recognition of initial phonemes (reflected from 200 ms of word onset), follows the activation of lexical candidates based on bottom-up process (e.g., Van den Brink & Hagoort, 2004; Van den Brink, Brown, & Hagoort, 2001). In ERP phonological priming experiments (Desroches, Newman, Joanisse, 2009; Praamstra, Meyer, & Levelt, 1994; Radeau, Besson, Fonteneau, & Castro, 1998), it has been proposed that the

moment from which there is sufficient of incoming signal to register a mismatch between the activated lexical candidates from the bottom-up processing and the lexical top-down predictions due to the context elicit the onset of enhanced negative brain response, usually called N400 effect. For example, during a picture-word matching task (Desroches, Newman, Joanisse, 2009), the N400 effect began later for a final-phoneme mismatch with the picture context (e.g. CONE-comb) than for an initial phoneme mismatch (e.g. CONE-bone). In an analogous way and converging with our previous study (Brunellière and Soto-Faraco, 2013), regional phonological mismatching based only on second phoneme induced an N400 effect from 200 ms to 300 ms after the onset of target word and did not persist in time when both semantic and phonological information are highly constraining. When the influence of top-down processes driven by sentence context was weak, the phonological mismatch effect persisted for longer than in the high-constraint condition. It is possible that the unfamiliar accent made it difficult for listeners to identify the phonologically incongruent words without the benefit of a strong semantic context.

Beyond the interplay between context-driven expectancy and stimulus-driven processes, we can consider whether our findings can shed some light on the question of how words are represented in the mental lexicon (Goldinger, 1996, 1998; Marslen-Wilson, 1984; Morton, 1979; Norris, McQueen & Cutler, 2003). On the one side, according to abstract representations models (Marslen-Wilson, 1984; Norris, McQueen & Cutler, 2003), variability in the speech input (e.g., inter-speaker variability in voice, speech rate, dialect-dependent phonological and phonetic realizations) is removed early on applying a filter that can match the incoming signal to abstract representations in the lexicon (a normalization process). For instance, psycholinguistic models of spoken word recognition including the Trace and Cohort

models assume a normalization process of acoustic variability rather than multiple exemplars of a single word (Goldinger, 1998 ; Connine, Rambon, & Patterson, 2008). Indeed, the Trace model contains two intermediate sublexical representations, features and phonemes and the Cohort model presents only feature representations at sublexical level. Under the view of abstract representations, we should expect that the variation from the incongruent phoneme would be normalized, producing an enhanced brain response with respect to the congruent regional phoneme. Then, after this process of normalization, no evidence of differential brain response should emerge (for similar findings, Goslin, Duffy, & Floccia, 2012). Whereas the phonological mismatch effect could result from a normalization process in strongly constraining contexts, this was not the case in less constraining contexts, since phonological mismatch effects persisted at later stages. Hence, despite our findings seem in principle inconsistent with the hypothesis of abstract lexical representations, our study cannot provide a conclusive answer concerning how words are represented in the mental lexicon. To disentangle the abstract and exemplar representations, the experimental design should introduce words with phonetic and phonological variations in pronunciation embedded in sentence context such as the manipulation of regional accent of the context (for similar approaches, e.g. Brunellière and Soto-Faraco, 2012; Goslin, Duffy, & Floccia, 2012)

In conclusion, our findings provide evidence that the two different linguistic constraints, phonological and high-level semantic constraints imposed by sentence context interacted together at sublexical and lexical levels. In spoken-language comprehension, we could envisage that the sentence context constrains at the lexical level during the lexical activation of phonological word forms and also at lower levels such as the sub-lexical levels. This hypothesis would be in favor of psycholinguistic models of spoken word recognition

(e.g., Trace model, McClelland & Elman, 1986) claiming for top-down feedback between the lexical and sublexical levels.

References:

Alarcos, E. (1953). “Sistema fonemático del catalán”. *Archivum*, III, pp. 135-146. Reeditado en *Estudis de lingüística catalana*. pp. 11-32. Barcelona: Ariel.

Baldeweg, T. (2006). Repetition effects to sounds: evidence for predictive coding in the auditory system. *Trends in Cognitive Sciences*, 10, 93-94. doi :10.1016/j.tics.2006.01.010

Boersma, P. & Weenink, D. (2011). Praat: doing phonetics by computer [Computer program]. Version 3.4, retrieved 2 Jan 2011 from <http://www.praat.org/>

Bonte, M., Parviainen, T., Hytönen, K., & Salmelin, R. (2006). Time course of top-down and bottom-up influences on syllable processing in the auditory cortex. *Cerebral Cortex*, 16, 115-123.

Brunellière, A., & Soto-Faraco, S. (2013). The speakers' accent shapes the listeners' phonological predictions during speech perception. *Brain and Language*, 125, 82-93. doi :10.1016/j.bandl.2013.01.007

Connine, C.M., Rambon, L.J., & Patterson, D.J. (2008). Processing variant forms in spoken word recognition: the role of variant frequency. *Perception and Psychophysics*, 70, 403-411. doi:10.3758/PP.70.3.403

Connolly, J.F., & Phillips, N.A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6, 256–266. doi : 10.1162/jocn.1994.6.3.256

Connolly, J.F., Phillips, N.A., Stewart, S.H., & Brake, W.G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, 43, 1-18. doi : 10.1016/0093-934X(92)90018-A

Connolly, J.F., Stewart, S.H., & Phillips, N.A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language*, 39, 302-318. doi : 10.1016/0093-934X(90)90016-A

Dambacher, M., Rolfs, M., Göllner, K., Kliegl, R., & Jacobs, A.M. (2009). Event-related potentials reveal rapid verification of predicted visual input. *PloS One*, 4, e5047. doi :10.1371/journal.pone.0005047

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

Desroches, A.S., Newman, R.L., & Joanisse, M.F. (2009). Investigating the time course of spoken recognition: electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience*, 21, 1893-1906. doi : 10.1162/jocn.2008.21142

Diaz, M.D., & Swaab, T.Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*, 1146, 85-100. doi : 10.1016/brainres.2006.07.034

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the royal society B*, 360, 815-836. doi : 10.1098/rstb.2005.1622

- Friston, K., & Kiebel, S. (2009). Cortical circuits for perceptual inference. *Neural Networks*, 22, 1093-1104. doi : 10.1016/j.neunet.2009.07.023
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 1166-1183. doi : 10.1037//0278-7393.22.5.1166
- Goldinger, S.D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279. doi : 10.1037//0033-295X.105.2.251
- Greenhouse, S.W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, 24, 95-111. doi : 10.1007/BF02289823
- Hagoort, P., & Brown, C.M. (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia*, 38, 1518-1530. doi : 10.1016/S0028-3932(00)00052-X
- Hanulíková, A., van Alphen, P.M., van Goch, M.M., & Weber, A. (2012). When one person's mistake is another's standard usage. The effect of foreign accent on syntactic processing, *Journal of Cognitive Neuroscience*, 24, 878-887. doi : 10.1162/jocn_a_00103
- Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders*, 45, 393-402. doi : 10.1016/j.jcomdis.2012.06.004
- Holcomb, P. J., & Neville, H. J. (1991). Natural sentence processing: An analysis using event related brain potentials. *Psychobiology*, 19, 286-300.

- 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, 1104-1112. doi : 10.1162/jocn_a_00148
- Krumbholz, K., Patterson, R.D., Seither-Preisler, A., Lammertmann, C., & Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, 13, 765-772. doi : 10.1093/cercor/13.7.765
- Kutas, M., Delong, K.A., & Smith, N.J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (ed), *Predictions in the Brain: Using Our Past to Generate a Future*, pp. 190-207. Oxford University Press.
- Kutas, M., & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 12, 463-470. doi : 10.1016/S1364-6613(00)01560-6
- Kutas, M., & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161-163. doi : 10.1038/307161a0
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164, 177-190. doi : 10.1016/j.jneumeth.2007.03.024
- Marslen-Wilson, W.D. (1984). Function and process in spoken word recognition: a tutorial review. In H. Bouma and D. G. Bouwhuis (eds), *Attention and Performance X: Control of Language Processes*, pp. 125-150. Erlbaum, Hillsdale, NJ.
- Marslen-Wilson, W.D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.

- McClelland, J.L., & Elman, J.L. (1986). The Trace model of speech perception. *Cognitive Psychology*, 18, 1-86. doi : 10.1016/0010-0285(86)90015-0
- McClelland, J.L. & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of Basic Findings. *Psychological Review*, 88, 375-407. doi : 10.1037//0033-295X.88.5.375
- 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
- Morton, J. (1979). Word recognition. In J. Morton and J. C. Marshall (eds), *Structures and Processes*, pp. 108–56. MIT Press, Cambridge, MA.
- Murray, M.M., Brunet, D., & Michel, C.M. (2008). Topographic ERP analyses : A step-by-step tutorial review. *Brain Topography*, 20, 249-264. doi : 10.1007/s10548-008-0054-5
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1-21. doi : 10.1111/1469-8986.3810001
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, 24, 375-425.
- Norris, D., McQueen, J.M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204-238. doi : 10.1016/S0010-0285(03)00006-9

Obleser, J., Scott, S.K., Eulitz, C. (2006). Now you hear it, now you don't: transient traces of consonants and their nonspeech analogues in the human brain. *Cerebral Cortex*, 16, 1069-1076. doi : 10.1093/cercor/bhj047

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, 9, 97-113. doi : 10.1016/0028-3932(71)90067-4

Perrin, F., Pernier, J., Bertrand, O., Giard, M-H., & Echallier, J.F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography and Clinical Neurophysiology*, 66, 75-81. doi : 10.1016/0013-4694(87)90141-6

Pickering, M.J., & Garrod, S. (2007). Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences*, 11, 105-110. doi: 10.1016/j.tics.2006.12.002

Praamstra, P., Meyer, A.S., & Levelt, W.J. (1994). Neurophysiological manifestations of phonological processing: latency variation of a negative ERP component time locked to phonological mismatch. *Journal of Cognitive Neuroscience*, 6, 204-219. doi: 10.1162/jocn.1994.6.3.204.

Radeau, M., Besson, M., Fonteneau, E., & Castro, S.L. (1998). Semantic, repetition, and rime priming between spoken words: behavioral and electrophysiological evidence. *Biological Psychology*, 48, 183-204. doi : 10.1016/S0301-0511(98)00012-X

Rafel i Fontanals, J. (1998). Diccionari de freqüències. 3, dades globals [Frequency dictionary. 3, global data]. Barcelona: Institut d'Estudis Catalans.

Thornhill, D.E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83, 382-392. doi: 10.1016/j.ijpsycho.2011.12.007

Uusvuori, J., Parviainen, T., Inkinen, M., & Salmelin, R. (2008). Spatiotemporal interaction between sound form and meaning during spoken word perception. *Cerebral Cortex*, 18, 456-466. doi : 10.1093/cercor/bhm076

Van Berkum, J.J, Brown, C.M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence From ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443-467. doi : 10.1037/0278-7393.31.3.443

Van Berkum, J.J, van den Brink, D., Tesink, C.M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20, 580-591. doi : 10.1162/jocn.2008.20054

Van den Brink, D., Brown, C.M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, 13, 967-985. doi : 10.1162/089892901753165872

Van den Brink, D., & Hagoort, P. (2004). The influence of semantic and syntactic context constraints on lexical selection and integration in spoken-word comprehension as revealed by ERPs. *Journal of Cognitive Neuroscience*, 16, 1068-1084.

Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 25, 394-417. doi : 10.1037//0278-7393.25.2.394

Van Petten, C., & Luka, B.J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83, 176-190. doi: 10.1016/j.ijpsycho.2011.09.015

Wicha N.Y.Y, Bates, E.A, Moreno, E.M, & Kutas, M. (2003). Potato not Pope: human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, 346, 165-168. doi : 10.1016/S0304-3940(03)00599-8

Wicha, N.Y.Y., Moreno, E.M., & Kutas, M. (2003). Expecting gender: an event related brain potential study on the role of grammatical gender in comprehending a line drawing within a written sentence in Spanish. *Cortex*, 39, 483-508. doi : 10.1016/S0010-9452(08)70260-0

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, 1272-1288. doi : 10.1162/0898929041920487

Author Notes: This research was supported by the Spanish Ministry of Science and Innovation (PSI2010-15426 and Consolider INGENIO CSD2007-00012), Comissionat per a Universitats i Recerca del DIUE-Generalitat de Catalunya (SGR2009-092), and the European Research Council (StG-2010263145). ERP analyses were performed with the Cartool software (supported by the Center for Biomedical Imaging of Geneva and Lausanne). We would like to thank Nara Ikumi for her help in constructing and recording the sentences. We also want to thank three anonymous referees and the editor for their useful comments.

Footnote:

¹ When the one of two topographical factors is not significant, the amplitude of ERP components was averaged over sites within each level of the significant topographical factor during follow-up *t*-tests.

² Since the laterality factor was not significant, the amplitude across a set of nine anterior sites was averaged into a single number just as the amplitude across the nine posterior sites in the follow-up *t*-tests.

³ Previous studies in spoken sentences have also shown early semantic congruity effects as soon as 150 ms after the onset of target words in spoken sentences that provide co-articulatory cues even before the presentation of target words (e.g., Hagoort and Brown, 2000; Holcomb and Neville, 1991). These semantic congruence effects arise later (around 200 ms) when pauses disrupted the co-articulatory cues (Holcomb and Neville, 1991; Van Petten, Coulson, Rubin, Plante, & Parks, 1999).

Figure Captions

Figure 1. Distribution of electrodes on the scalp. In a dotted line, the groupings of electrodes used as factors in the analysis of variance (see main text).

Figure 2. Grand-average waveforms for the target words embedded in strongly constraining sentence context according to the phonological form (congruent vs. incongruent phonological form). Stars (*) denote significant effects.

Figure 3. Grand-average waveforms for the target words embedded in weakly constraining sentence context according to the phonological form (congruent vs. incongruent phonological form). Stars (*) denote significant effects.

Figure 4. Grand-average waveforms for the congruent phonological words according to the degree of semantic constraint (low vs. high). Stars (*) denote significant effects.

Figure 5. Grand-average waveforms for the incongruent phonological words according to the degree of semantic constraint (low vs. high). Stars (*) denote significant effects.

Figure 6. Topographical effects of phonological mismatch at each level of sentence constraints (at the top, low constraints, at the bottom, high constraints). Subtraction 2D-maps illustrating the time windows during which significant differences between congruent and incongruent phonological forms were found.

Table 1. Examples of experimental conditions

Conditions	Semantic constraints	Phonological word forms	Examples
HS-CP	High	Congruent	<p>En veure el nen tranquil al seu llit, la mare s'acostà a la seva galta per fer-li un /pətó/.</p> <p><i>When seeing the child quiet in his bed, the mother approached his cheek to make him a kiss.</i></p>
HS-IP	High	Incongruent	<p>En veure el nen tranquil al seu llit, la mare s'acostà a la seva galta per fer-li un /petó/.</p> <p><i>When seeing the child quiet in his bed, the mother approached his cheek to make him a kiss.</i></p>
LS-CP	Low	Congruent	<p>En Joan estava convençut de que al final aconseguiria un /pətó/.</p> <p><i>John was convinced that at this end, he would get a kiss.</i></p>
LS-IP	Low	Incongruent	<p>En Joan estava convençut de que al final aconseguiria un /petó/.</p> <p><i>John was convinced that at this end, he would get a kiss.</i></p>

Table 2. Properties of experimental conditions

Conditions	Semantic constraints	Phonological word forms	Length	Context markers of regional accent	Sentence context duration (ms)	Target word duration (ms)	Critical vowel duration (ms)
HS-CP	High	Congruent	13.5	6.4	3794	482	63
HS-IP	High	Incongruent	13.5	6.4	3792	484	64
LS-CP	Low	Congruent	13.1	6.4	3707	481	66
LS-IP	Low	Incongruent	13.1	6.4	3724	483	64

Conditions varying in the degree of semantic constraints and the congruency of phonological word forms (High semantic constraints and congruent phonological word forms, HS-CP, High semantic constraints and incongruent phonological word forms, HS-IP, Low semantic constraints and congruent phonological word forms, LS-CP and Low semantic constraints and incongruent phonological word forms, LS-IP), **Length of sentence frames** in terms of number of words, **Context markers of regional accent** in terms of the number of phonemes marking the regional accent delivered in sentence frames, **Sentence context duration** (the duration in milliseconds before the onset of target word), **Target word duration** (the duration in milliseconds between the onset and offset of target word), **Critical vowel duration** (the duration in milliseconds between the onset and offset of target word's first vowel)

Table 3. Statistical results of acoustic properties on the critical vowels

Main effects of semantic constraint		Main effects of congruency of phonological word forms	
Duration	F(1,111)=2.17	Duration	F(1,111)=0.12
Maximum intensity	F(1,111)=1.01	Maximum intensity	F(1,111)=0.89
Standard deviation F0	F(1,111)=1.66	Standard deviation F0	F(1,111)=2.5
Mean F0	F(1,111)=0.57	Mean F0	F(1,111)=1.42
Interaction between semantic constraint and congruency of phonological word forms			
Duration		F(1,111)=1.68	
Maximum intensity		F(1,111)=0.96	
Standard deviation F0		F(1,111)=1.89	
Mean F0		F(1,111)=0.09	
Fundamental frequency, F0 , All probability levels were equal or greater than 0.17			

Table 4. Paired comparisons on mean amplitude in the N100 time window of 110-160 ms

Comparisons	Anterior Sites	Posterior Sites
HS-CP vs. HS-IP	t(23)=0.97	t(23)=0.83
LS-CP vs. LS-IP	t(23)=2.49, p<.05	t(23)=2.9, p<.01
HS-CP vs. LS-CP	t(23)=0.58	t(23)=0.34
HS-IP vs. LS-IP	t(23)=1.21	t(23)=2.83, p<.01
HS-CP vs. LS-IP	t(23)=2.17, p<.05	t(23)=2.70, p<.05
HS-IP vs. LS-CP	t(23)=0.74	t(23)=0.64

High semantic constraints and congruent phonological word forms, **HS-CP**, High semantic constraints and incongruent phonological word forms, **HS-IP**, Low semantic constraints and congruent phonological word forms, **LS-CP** and Low semantic constraints and incongruent phonological word forms, **LS-IP**

Table 5. Paired comparisons on mean amplitude in the late N400 time window of 350- 600 ms

Comparisons	Statistical results
HS-CP vs. HS-IP	t(23)=0.39
LS-CP vs. LS-IP	t(23)=2.71, p<.05
HS-CP vs. LS-CP	t(23)=2.09, p<.05
HS-IP vs. LS-IP	t(23)=4.92, p<.0001
HS-CP vs. LS-IP	t(23)=3.86, p<.001
HS-IP vs. LS-CP	t(23)=2.21, p<.05

High semantic constraints and congruent phonological word forms, **HS-CP**, High semantic constraints and incongruent phonological word forms, **HS-IP**, Low semantic constraints and congruent phonological word forms, **LS-CP** and Low semantic constraints and incongruent phonological word forms, **LS-IP**

Figure 1

Figure 2

Figure 3

Figure 4

Figure 5

Figure 6