

World knowledge in L2

**WORLD KNOWLEDGE AND NOVEL INFORMATION INTEGRATION  
DURING L2 SPEECH COMPREHENSION \***

**Carlos Romero-Rivas <sup>a</sup>, Joanna D. Corey <sup>a</sup>, Xavier Garcia <sup>a</sup>, Guillaume Thierry <sup>b</sup>,  
Clara D. Martin <sup>c,d</sup>, Albert Costa <sup>a,e</sup>**

<sup>a</sup> Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain

<sup>b</sup> School of Psychology, Bangor University, Bangor, UK

<sup>c</sup> Basque Center on Cognition, Brain and Language (BCBL), San Sebastian, Spain

<sup>d</sup> IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

<sup>e</sup> Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Address for correspondence: Center for Brain and Cognition, Universitat Pompeu Fabra. Carrer de Tànger, 122, 08018 Barcelona, Spain. Phone number: +34 93 542 29 63. *E-mail address:* romeriv@gmail.com (Carlos Romero-Rivas)

\* This research was approved by the ethics committee of the Spanish Ministry of Economy and Finance, which funded this study. We thank Silvia Blanch and Xavier Mayoral for their technical support, and Meritxell Ayguasanosa for assistance in testing participants. This research was funded by an FPI grant (BES-2012-056668) and two project grants (PSI2011-23033 and Consolider INGENIO CSD2007-00012) awarded by the Spanish Government; by one grant from the Catalan Government (SGR 2009-1521); and by one grant from the European Research Council under the European Community's Seventh Framework (FP7/2007-2013 Cooperation grant agreement 613465-AThEME). C.D.M. is supported by the IKERBASQUE institution and the Basque Center on Cognition, Brain and Language. A.C. is supported by the ICREA institution and the Center for Brain and Cognition.

## **ABSTRACT**

In this study we explore whether world knowledge (WK) processing differs between individuals listening to their native (L1) or their non-native (L2) language. We recorded event-related brain potentials in L1 and L2 speakers of Spanish while they listened to sentences uttered by native speakers of Spanish. Sentences were either congruent or incongruent with participants' WK. In addition, participants also listened to sentences in which upcoming words could not be anticipated on the basis of WK. WK violations elicited a late negativity of greater magnitude and duration in the L2 than the L1 group. However, sentences in which WK was not helpful regarding word anticipation elicited similar N400 modulations in both groups. These results suggest that WK processing requires a deeper lexical search in L2 comprehension compared to in the L1.

**Keywords:** World Knowledge; L2 processing; Speech comprehension; N400; Bilingualism

## 1. Introduction

During a conversation, we use semantic and factual (world knowledge) information in order to understand a message. That is, in order to communicate with others, we use our knowledge about the words of a language and the way they are related to each other, as well as information about actual facts about the world we live in. Going a step further, we are even able to use this information in order to anticipate upcoming words, so that they will be easier to integrate in the previous context as compared to words that are semantically or factually incongruent (Hagoort, Hald, Bastiaansen, & Petersson, 2004).

Semantic correctness is constrained by grammar and by the way words are related to each other. For instance, a sentence like “*Barack Obama is the president of France*” is perfectly correct at the semantic level, since Obama is a person, a person can be the president of a country, and France is a country. However, the statement is factually incorrect. If a statement is false, rejecting wrong information probably requires extra time and processing. This idea is supported by behavioral results showing that it takes longer to read a false statement than a true one (Singer, 2006; Rapp, 2008). During second language comprehension, however, retrieval and integration processes seem to be less automatic as compared to native language comprehension (Ardal, Donald, Meuter, Muldrew, & Luce, 1990) and, therefore, world knowledge processing might be slower and even compromised.

In this study, we aimed at exploring whether the processing of world knowledge information is affected during second language comprehension, as has been observed for semantic information (see below). Interestingly, while semantic processing in a

second language informs us regarding how bilinguals use linguistic information (the semantic constraint of a sentence) to anticipate and integrate upcoming words in their second language, world knowledge processing would illuminate how bilinguals use their general knowledge about the world to carry out these very same processes. To explore this question, we investigated whether the integration of words at the end of auditory sentences were affected by a native (L1) or non-native (L2) context of presentation. We explored this issue by assessing the modulation of the N400 event-related brain potential (ERP) elicited by world knowledge violations as compared to correct statements from a world knowledge point of view, as well as a third condition in which we used sentences that were difficult to anticipate because they were neutral from a world knowledge perspective. We used ERPs because of their high temporal resolution, allowing us to disentangle sub-stages of language processing, and because of their well-known sensitivity to spontaneous semantic processing, even when participants are not explicitly asked to evaluate statement correctness.

### *1.1 World knowledge vs. semantic information processing*

In linguistics, the “dissociation theory” posits that semantic processing precedes pragmatic integration, at which stage world knowledge is supposedly incorporated (Forster, 1979). However, a number of authors (Clark, 1996; Jackendoff, 2002) have argued for the opposite view, pointing out that, because of polysemy, the meaning of a word can only be fully established after disambiguation of the context on the basis of world knowledge. This latter perspective is referred to as the “simultaneous theory”.

In order to test these contrasting views, Hagoort et al. (2004) conducted an EEG study. During the experiment, Dutch speakers had to read sentences such as “the Dutch

trains are *yellow/white/sour* and very crowded” (the critical words are in italics) written in their native language. Dutch people know that Dutch trains are yellow, and therefore the first version of this sentence is correct at the semantic level and, at the same time, true based on world knowledge. However, although the second version of the sentence remains semantically correct, it is a world knowledge violation. Finally, the third version of the sentence is incorrect both in terms of semantic validity and world knowledge, since ‘*sour*’ is a quality that trains cannot possess.

Hagoort et al. (2004) explored the differences between semantic and world knowledge violations by assessing the modulations of the N400 ERP component. The N400 is a negative-going wave, which appears roughly between 250 and 500 ms after the presentation of a meaningful event, and is associated with lexical anticipation and semantic integration processing (Lau, Phillips, & Poeppel, 2008; Federmeier & Laszlo, 2009; Kutas & Federmeier, 2011). It was first observed by Kutas and Hillyard (1980) in response to semantically anomalous words in the context of a written sentence (words that are difficult to integrate in the previous semantic context; e.g., ‘He spread the warm bread with *socks*’). In addition, Kutas and Hillyard (1984) observed that the N400 component is also sensitive to a listener’s expectancy for a specific word during sentence comprehension. More specifically, they found that semantically congruent, terminal words in highly constrained sentences elicited a less negative N400 mean amplitude as compared to terminal words in medium and lowly constrained sentences (e.g., ‘He mailed the letter without a *stamp*’ vs ‘There was nothing wrong with the *car*’).

Hagoort et al. (2004) observed the classical N400 effect for semantic violations (‘*sour*’). Importantly, the N400 effect for world knowledge violations (‘*white*’) was identical in onset and peak latency, and very similar in amplitude and topographic

distribution, to the classic semantic N400 effect. These results seemed to suggest that the brain retrieves and integrates semantic and factual information at the same time (see also Martin, Garcia, Breton, Thierry, & Costa, 2014).

Along the same lines as Hagoort et al.'s (2004) study, Hald, Steenbeck-Planting, and Hagoort (2007) presented Dutch listeners with written sentences that could be semantically coherent and correct based on world knowledge, or semantically coherent but false based on world knowledge. In addition, preceding the presentation of the critical sentences, one of two discourse contexts was presented: The compatible discourse gave supporting context to the information described in the sentences whilst the incompatible discourse changed the relevant focus, so that sentences that were false based on world knowledge seemed more acceptable. Hald et al. (2007) observed the largest N400 effect for sentences containing a world knowledge violation that was incongruent with the previous discourse. However, when world knowledge violations were preceded by a discourse context which provided a more acceptable interpretation of the violation, the N400 effect was reduced. Furthermore, when true sentences were preceded by a discourse context that provided a more acceptable interpretation of world knowledge violations, the N400 amplitude was more negative compared to the condition in which true sentences were preceded by a compatible and congruent discourse context. These results suggest that local discourse context, semantic information and world knowledge information interact during the N400 time window.

Taken together, these findings suggest that semantic and world knowledge processing are to some extent independent, although they interact in several ways

during the N400 time window, when the meaning of words is retrieved and integrated into the previous context.

In the present study, we will investigate world knowledge processing in L2 auditory sentence comprehension. Although we know little about world knowledge processing during L2 comprehension, semantic processing during L2 written comprehension has been extensively studied in recent years. All this information could serve as an anchor point for interpreting new data on world knowledge integration during L2 comprehension.

### *1.2 Semantic information processing during L2 comprehension*

Neurolinguistic research conducted with bilingual speakers has shown that semantic processing is affected when bilinguals are presented with sentences written in their L2. The most common finding is that bilinguals reading in their L2 show a later onset and peak latency of the N400 effect for semantic violations as compared to when they are reading in their L1 or to monolinguals reading in their native language (Ardal et al., 1990; Weber-Fox & Neville, 1996, 2001; Moreno & Kutas, 2005; Ojima, Nakata, & Kakiji, 2005; Braunstein, Ischebeck, Brunner, Grabner, Stamenov, & Neuper, 2012; Newman, Tremblay, Nichols, Neville, & Ullman, 2012). This finding has been interpreted as reflecting an extended lexical search and a lower degree of automaticity of L2 processing as compared to that of the L1 (Ardal et al., 1990). In other studies, in addition to a later onset and peak latency, it has also been observed that bilinguals show a reduced magnitude of the N400 effect for semantic violations in their L2 compared to their L1 (Moreno, Rodríguez-Fornells, & Laine, 2008; Newman et al., 2012). Finally, Ojima et al. (2005) observed (along with the general finding that bilinguals have a later

onset and peak latency for the N400 effect when reading sentences in their second language) that Japanese speakers with low reading proficiency in English (J-Low) exhibit a longer N400 effect (extended in time) than Japanese speakers highly proficient in reading English (J-High). However, the amplitude of the N400 effects for both J-Low and J-High groups were similar.

Regarding speech comprehension, Hahne & Friederici (2001) observed that semantically incorrect sentences showed similar N400 ERP patterns in L1 and L2 listeners. However, during L2 speech comprehension, semantic violations also elicited a late negativity, an effect that was not observed for native listeners. The authors interpreted this late negativity as an effect more closely related to conceptual-semantic processing while the N400 would be more related to lexical-semantic processing (Hahne & Friederici, 2001). In addition, Hahne (2001) observed a delayed peak latency for semantic violations in the L2 group compared to the L1 group. Finally, FitzPatrick and Indefrey (2010) also observed a delayed N400 effect for semantic violations during L2 speech comprehension, mostly explained by L2 words, but not by L1 translation equivalents that were initially congruent with the sentence context (see also FitzPatrick & Indefrey (2014) for further research on this phenomenon).

These findings denote that semantic information processing is present (although slower) early on during exposure to an L2, and that it becomes faster as proficiency increases (Ojima et al., 2005; see also Oliver, Gullberg, Hellwig, Mitterer; & Indefrey, 2012), and as target words are more similar to L1 words (FitzPatrick & Indefrey, 2010, 2014).

Nevertheless, as previously mentioned, the extent to which L2 comprehension affects the processing of world knowledge information remains essentially unknown. In order to explore this issue, we presented Spanish native speakers and French, Italian and Portuguese non-native speakers of Spanish with auditory sentences recorded in Spanish. These sentences were either factually correct or incorrect from a world knowledge viewpoint. In addition, we also introduced a third condition (unknown sentences) in which world knowledge could not be used to anticipate the final and critical word. Thus, correct sentences will shed light on world knowledge integration. Sentences including a world knowledge violation (WKv) will show what happens when world knowledge is contradicted. Unknown sentences (US) will show what happens when listeners integrate novel information (pure lexical integration in the absence of anticipation). Comparing correct and WKv conditions will reveal the simple effect of world knowledge violation. Comparing correct and US will characterize differences in terms of anticipation based on world knowledge. And comparing WKv and US will characterize differences in processing unexpected words/concepts when expectations based on world knowledge are contradicted vs. lack of anticipation. ERPs were time-locked to the onset of critical words presented in a word-final position, in order to avoid an overlap with upcoming words in the sentence (see **Table 1** for examples). Our design thus allowed us to explore whether world knowledge processing is modulated by the listeners' language mastery (either native or non-native).

<Please insert Table 1 about here>

We hypothesized that during L2 speech comprehension, the integration of world knowledge information is somewhat compromised. We based this assumption on

previous results showing that bilinguals reading or listening to their second language have extended lexical searches and manifest a lower degree of automaticity compared to when they are processing their L1 (Ardal et al., 1990; Weber-Fox & Neville, 1996, 2001; Hahne, 2001; Hahne & Friederici, 2001; Moreno & Kutas, 2005; Ojima et al., 2005; Braunstein et al., 2012; Newman et al., 2012). This could be manifested in two ways: a reduced N400 effect magnitude, or an extended N400 effect/a late negativity for world knowledge violations during L2 as compared to L1 speech comprehension.

## **2. Methods**

### *2.1 Participants*

Twenty-two native speakers of Spanish, as well as twenty-two L2 speakers of Spanish (native languages: French, Italian and Portuguese), participated in this study in return for monetary compensation (10€/h). None of them reported any hearing or neurological impairments. Before the beginning of the experiment, subjects gave their written informed consent.

Due to a noisy signal (based on visual inspection of raw recordings) or a small number of epochs per condition (< 60%), after outlier rejection seventeen participants remained per group (L1: 10 female, all right handed, mean age = 23.56 years, range = 19-35 years; L2: 12 female, all right handed, mean age = 25.95 years, range = 21-31 years).

The L2 group was formed by 8 Italians, 5 French, and 4 Portuguese L2 speakers of Spanish, who were living in Barcelona at the time. Twelve of them learned Spanish

both in school and by immersion, and the rest acquired Spanish by immersion only. They were first exposed to Spanish at a mean age of 19.23 years (range = 10-28 years).

Before the experiment, all participants had to rate their proficiency in Spanish on a 7-point Likert scale for reading, writing, speaking, and auditory comprehension. As expected, we found differences between the two groups in all four measures (reading L1 mean: 6.94, L2 mean: 6.23; writing L1 mean: 6.71, L2 mean: 5.53; speaking L1 mean: 6.94, L2 mean: 5.76; listening L1 mean: 7, L2 mean: 6.23). However, although there were differences between groups, the results from the L2 group indicated high levels of proficiency in Spanish comprehension.

In addition, participants had to respond to a Spanish vocabulary knowledge test. The mean proportion of correct responses for the L1 group was 96% (SD = .04), while the mean proportion of correct responses for the L2 group was 87% (SD = .05). This further supports the L2 group being proficient in Spanish.

## 2.2 *Materials*

The stimuli consisted of a total of 360 Spanish sentences divided into 120 sets of 3 triplets (for examples see **Table 1**). Each version within a triplet corresponded to one of 3 conditions (critical word in italics): (1) sentences that are known to be true were Known Sentences (KS), ex “El color de los taxis en Nueva York es el *amarillo*” (“the color of taxis in New York is *yellow*”). (2) Sentences that are known to be false were World Knowledge Violations (WKv), ex “El color de los taxis en Nueva York es el *verde*” (“the color of taxis in New York is *green*”). (3) True sentences whose content is not part of common knowledge were Unknown Sentences (US), ex “El color de los taxis en Pekín es el *verde*” (“the color of taxis in Beijing is *green*”). By changing only a

few key words within each triplet, we ensured that the conditions were matched for average number of words per sentence. Three lists were created with each containing 40 sentences in each condition. Each version of the 120 sentences was used in each list. Each participant was randomly assigned to one list.

All critical words were sentence final so as to permit comparisons between the conditions not only for integration but also for the truth evaluation of the entire sentence (“wrap-up processes”, Hagoort, 2003), and to avoid confounding ongoing integration effects. The critical words were matched across conditions using the B-Pal Corpus (David & Perea, 2005) with regards to the following criteria: mean number of syllables, number of phonemes, onset, frequency, familiarity and imageability. In addition, the sentences were also matched across conditions for the number of words and the duration of the critical word (see **Table 2**). The only differences between conditions before the critical words were the familiarity with the concepts and their association (e.g. “New York”, “taxi”, “yellow”) or lack thereof.

<Please insert Table 2 about here>

The sentences were recorded by various native Spanish speakers in a soundproof room (7 male, 6 female; half of the sentences were recorded by males and half by females). The audio files were edited using Audacity (Version 2.0.2; <http://audacity.sourceforge.net>) such that the conditions were matched for the acoustic length (duration) of the critical word (see **Table 2**).

A Cloze probability task was administered to 20 native Spanish speaking participants who did not take part in the EEG experiment. Two lists were created using the 120 KS sentences and 120 US sentences, each with 60 KS sentences and 60 US sentences. The sentences were randomized within each list. Each participant was randomly assigned to one list. The participants were given a list with each sentence typed out with the critical (and last) word missing, and were then asked to complete the sentence such that the contents of the sentence were true according to their knowledge. If they did not know the appropriate ending, they were asked to guess. This task also included a confidence rating in which the participants were asked to rate how positive they were that the meaning of the completed sentence was true. The rating scale was a 5-point Likert scale (1=not sure; 5=completely sure). This measure was included as an explicit measure of common knowledge in order to determine to what extent the critical words in each condition were known, and therefore could have been anticipated. The critical words of known sentences had an average cloze probability of 53.2% ( $SD = 0.35$ ). The critical words of unknown sentences had an average cloze probability of 16.4% ( $SD = 0.21$ ). The cloze probability of world knowledge violations was calculated by the percentage of instances in which the critical word constituting a world knowledge violation appeared in the KS condition. The critical words of world knowledge violations had an average cloze probability of 0.02% ( $SD = 0.1$ ). The average confidence rating for KS sentences was 3.43 ( $SD = 1.64$ ), and for US sentences it was 1.47 ( $SD = 1.59$ ). Therefore, participants were fairly certain of things that they knew to be true whereas they guessed at those whose truth-value they did not know.

It is of note that a substantial part of the experimental sentences could be completed with several words (e.g., “*In 1992, Olympic Games were held in Spain*”; in this particular case, participants responded either “Spain” or “Barcelona” with similar

frequency). Thus, even if the particular word was not the most frequent response in the cloze probability test, the general concept was anticipated in most cases. The behavioral results of the rating task conducted after the ERP experiment are consistent with this view (see **Figure 3** below).

## 2.3 *Procedure*

### 2.3.1 *EEG task*

After having completed the linguistic background questionnaire and given written consent, each participant was seated in front of a computer in a soundproof room, at a distance of approximately two feet.

We used E-Prime 2.0 (Psychology Software Tools, Inc.) to run the experiment. Participants were presented with stimuli via loudspeakers, and were asked simply to pay attention and comprehend. They were asked to look at the fixation cross as long as it appeared on the screen, and to avoid blinking while the fixation cross was on the screen. The fixation cross appeared for 1000 ms before the onset of each sentence, during, and 1500 ms after the sentence offset. The screen remained blank for 3500 ms in between each sentence. In 25% of the trials the sentence was followed by a comprehension question displayed on the screen to which participants were prompted to respond “yes” or “no” to ensure that they were paying attention and understanding the stimuli. Sentences were presented in 6 blocks of 20 trials with rest periods in between each block. The listening task took approximately 25 minutes.

### 2.3.2 Rating task

After the experiment, participants were asked to perform a rating task (also run using E-Prime 2.0). Each sentence was again presented randomly via loudspeakers. After each presentation, participants were prompted via the computer screen to evaluate the truth-value of the sentence, and how confident they were, by typing in a number using the keyboard. The response options presented on screen after every sentence were as follows: (1) Confident True; (2) Guess True; (3) Confident False; (4) Guess False; (5) Cannot Guess. The rating task took approximately 20 minutes.

## 2.4 EEG recording and processing

The EEG signal was recorded from 32 electrodes (impedances were kept below 5 k $\Omega$ ) mounted in an elastic cap, at standard 10-20 locations. The on-line reference electrode was attached to the left mastoid, and the signal was re-referenced off-line to the mastoid average. Lateral eye movements were recorded with an electrode beside the right eye, and eye blinks were recorded with another electrode below the right eye. Data were amplified by the BrainAmp Standard amplifier ([www.brainproducts.com](http://www.brainproducts.com); resolution: .1 $\mu$ V per bit), and EEG signal was filtered on-line (software filter) with a 0.1-100 Hz bandpass filter and digitized at 500 Hz. As for the criteria used for artifact rejection, we set the maximal allowed voltage steps at 30  $\mu$ V/ms, maximal allowed differences of values in intervals at 100  $\mu$ V (interval lengths: 200 ms), minimal and maximal allowed amplitudes at -100 and 100  $\mu$ V respectively, and the lowest allowed activity in intervals was 0.5  $\mu$ V (interval lengths: 100 ms).

## 2.5 *ERP analyses*

EEG epochs were set for the last word of each sentence. We extracted the epochs from 200 ms before to 1200 ms after the onset of the critical (final) word of each sentence. EEG waveforms were baseline corrected in reference to a 200 ms pre-stimulus onset, and averaged per participant and condition. For measurement, we applied a 0.1-40 Hz bandpass filter. Mean amplitudes in specific time windows were analyzed with repeated measures ANOVAs, analyzing three regions: frontal (F3, Fz, F4, FC1 and FC2), central (C3, Cz, C4, CP1 and CP2), and posterior (P3, Pz, P4, PO1 and PO2).

Statistical analyses were performed on two main time-windows. First, we established a time-window for the auditory N400 (250-500 ms) based on previous literature (e.g., FitzPatrick & Indefrey, 2014). Second, we established a later time-window (500-750 ms) in order to explore for further effects of world knowledge processing beyond the N400 time window. We based this late time-window on Hahne & Friederici's (2001) observation of a late negativity for semantic violations during L2 speech comprehension.

All effects and interactions including a variable with three levels were corrected for sphericity using the Greenhouse-Geisser correction.

## 3. Results

### 3.1 *EEG task*

### 3.1.1 N400 mean amplitude (250-500 ms)

As argued in the introduction, modulations of the N400 component could be taken as an index of different processing strategies during the integration of world knowledge.

The repeated measures ANOVA for the N400 effect (250-500 ms) included the within-subjects factors Topography (frontal, central, posterior), and Condition (known sentences – KS, world knowledge violations – WKv, unknown sentences – US), as well as the between-subjects factor language (L1 vs. L2). We obtained significant effects of Condition ( $F(2,64) = 9.08; p < .001$ ) and Topography ( $F(2,64) = 6.60; p < .01$ ).

Planned comparisons for the Condition effect revealed that the mean amplitude of the N400 ERP component for KS was significantly less negative than those of both WKv ( $F(1,32) = 13.98; p < .001$ ) and US ( $F(1,32) = 13.40; p < .001$ ). However, there were no significant differences between WKv and US ( $F(1,32) = 0.20; p = .66$ ).

Planned comparisons for the Topography effect showed that the mean amplitude over the central region was significantly more negative than over the posterior region ( $F(1,32) = 27.13; p < .001$ ). There were no more significant differences between topographic regions.

<Please insert Figure 1 about here>

### 3.1.2 Late negativity mean amplitude (500-750 ms)

The repeated measures ANOVA for this time window included the within-subject factors Topography (frontal, central, posterior), and Condition (KS, WKv, US), as well as the between-subjects factor language (L1 vs. L2). We obtained significant effects of Condition ( $F(2,64) = 8.85; p < .001$ ) and Topography ( $F(2,64) = 55.60; p < .001$ ), and a significant interaction between Language, Condition and Topography ( $F(4,128) = 3.04; p < .05$ ).

Planned comparisons for the three-way interaction are described in **Table 3** ( $F(1,32)$ ). Most importantly, they revealed that the only difference between language groups was the mean amplitude of WKv, in the posterior region: The late negativity was larger for WKv over the back of the head during L2 speech comprehension compared to that of the L1. In addition, we also observed that L1 speakers showed no significant differences between the three experimental conditions in this time window. For L2 speakers, the late negativity was significantly more negative for WKv than US, and for US than KS sentences.

<Please insert Table 3 about here>

<Please insert Figure 2 about here>

### 3.2 *Rating task*

As stated in the procedure section, the rating task was run after the EEG experiment. During this task, participants had to judge the truth-value of the experimental sentences, which were presented again through loudspeakers.

Repeated measures ANOVA for this analysis included the within-subjects factors Condition (KS, WKv, US) and Response Key (Confident True, Guess True, Confident False, Guess False, Cannot Guess), as well as the between-subjects factor Language (L1 vs. L2). The dependent variable was the number of times a specific response was given. We obtained a significant effect of Response Key ( $F(4,128) = 7.68$ ;  $p < .001$ ), and a significant interaction between Condition and Response Key ( $F(8,256) = 236.18$ ;  $p < .001$ ).

Planned comparisons for the interaction between Condition and Response Key showed that the number of times a given response was given differed significantly across conditions (see **Figure 3**).

Importantly, there were no significant differences involving language groups, meaning that both groups correctly interpreted and judged KS and WKv conditions. In addition, the most common response by both groups for the US condition was “Cannot Guess”, as expected, followed by “Guess True” and “Guess False” in a similar amount.

<Please insert Figure 3 about here>

#### 4. Discussion

This study aimed at exploring whether the processing of world knowledge information differs during L1 and L2 speech comprehension. We explored this issue by looking at the modulation of the N400 and the late negativity, two indexes of word meaning integration into previous context. In a nutshell, our results show that:

- a) Language groups do not differ significantly in ERP mean amplitude in the classical N400 time window (250-500 ms). Both the N400 mean amplitudes for world knowledge violations and unknown sentences were more negative as compared to known sentences, regardless of subjects' native language.
- b) However, we observed a larger late negativity for world knowledge violations during L2 speech comprehension compared to L1 speech comprehension in the 500-750 ms time window. Also, during L2 speech comprehension, world knowledge violations and unknown sentences elicited a greater negativity than known sentences in the 500-750 ms time window. In addition, this effect was significantly larger for world knowledge violations. During L1 speech comprehension, there were no differences between conditions in the 500-750 ms time window.

As we proposed in the introduction, differences between L1 and L2 processing in the retrieval and integration of world knowledge information might be indexed by modulations of the N400 component and/or late negativity across the language groups. This hypothesis was based on previous observations (Ardal et al., 1990; Weber-Fox & Neville, 1996, 2001; Hahne, 2001; Moreno & Kutas, 2005; Ojima et al., 2005; Braunstein et al., 2012; Newman et al., 2012) showing that semantic processing (as

indexed by the N400 and late negativity) during L2 comprehension is less automatic and requires a more extended lexical search than L1 comprehension.

The observation of a greater negativity for world knowledge violations during L2 speech comprehension (compared to L1 comprehension) in the 500-750 ms time window is congruent with our hypothesis that L2 speakers manifest extended lexical search and a lower degree of automaticity compared to L1 speakers. This result goes in the same direction as Hahne and Friederici's (2001) and Ojima et al. (2005) observations of a late negativity for semantic violations processing during L2 speech comprehension. In addition, and also consistent with our view, during L2 speech comprehension, world knowledge violations elicited a larger late negativity amplitude than both known sentences and unknown sentences in the 500-750 ms time window.

A tentative explanation for these observations is a more durable N400 effect (meaning extended in time) for world knowledge violations during L2 than L1 speech comprehension. Thus, L2 speech processing would need additional resources to complete lexical searches when world knowledge information is contradicted. In other words, this process would be less automatic during L2 speech comprehension than L1 comprehension. This interpretation is consistent with observations showing that semantic processing is also less automatic during L2 comprehension compared to L1 comprehension. For instance, several studies (Ardal et al., 1990; Weber-Fox & Neville, 1996, 2001; Hahne, 2001; Moreno & Kutas, 2005; Ojima et al., 2005; Braunstein et al., 2012; Newman et al., 2012) found a delayed N400 latency for semantic violations during L2 comprehension compared to L1 comprehension. In addition, Ojima et al. (2005) also observed a more durable N400 effect for written semantic violations in L2 as compared to L1 comprehension. Such results have been interpreted as a sign of slower although robust semantic processing during L2 comprehension, that becomes

faster when people become more proficient, with native speakers being the most proficient.

Alternatively, and as pointed out above, Hahne & Friederici (2001) observed a late negativity for semantic violations, and combined syntactic and semantic violations only during L2 speech comprehension. These authors interpreted this late negativity as an effect more closely related to conceptual-semantic processing, while the N400 would be more related to lexical-semantic concepts. Therefore, during L2 comprehension, accessing unexpected words would be as difficult as during L1 comprehension (similar N400 effects for unknown and world knowledge violation sentences in L1 and L2 comprehension). However, later on, processing the conceptual-semantic features of words listeners' world knowledge would be more difficult during L2 than L1 comprehension (larger late negativity for world knowledge violations during L2 than L1 comprehension). In this way, our results could also be read as a more difficult processing of world knowledge conceptual-semantic information during L2 compared to L1 comprehension.

Thus, L2 speakers would benefit from robust but slower world knowledge processing during language comprehension compared to L1 speakers, as is also the case for semantic processing (Ojima et al., 2005).

Importantly, the inclusion of unknown sentences in our design allowed us to tease apart the integration of information that is not congruent with the previous context based on world knowledge compared to the integration of novel information. We observed that the N400 ERP was more negative for both WKv and US than correct sentences (either because of contradicted expectations (WKv) or unknown information

(US)). This was the case both in L1 and L2. Furthermore, our results are congruent with those reported by Hagoort et al. (2004) regarding L1 comprehension, since world knowledge violations and sentences with low semantic expectations elicited similar responses.

However, in a subsequent time-window, unknown sentences and world knowledge violations required different processing strategies during L2 speech comprehension, as indexed by the reduced late negativity effect for unknown sentences as compared to world knowledge violations in the 500-750 ms time window. This suggests that the larger effect in this time window for WKv than US for the L2 group is due to the listeners not only having to integrate unexpected words/concepts, but also having to integrate critical words that violate world knowledge representations in their long-term memory.

Future research including semantic and world knowledge violations during L2 comprehension may shed some light on the differences between semantic and world knowledge processing in a second language. Our results on the processing of world knowledge violations are in line with the previous literature on semantic violations processing during L2 comprehension (Ardal et al., 1990; Weber-Fox & Neville, 1996, 2001; Hahne, 2001; Hahne & Friederici, 2001; Moreno & Kutas, 2005; Ojima et al., 2005; Braunstein et al., 2012; Newman et al., 2012). This might suggest the use of similar neural mechanisms for semantic and world knowledge processing while comprehending a second language. However, some experimental paradigms (such as word-by-word sentence reading; Martin et al., 2014) and methods (such as EEG power analysis; Hagoort et al., 2004) have lead to the identification of differences in the

processing of semantic and world knowledge violations during L1 comprehension. The combined use of these paradigms and methods, as well as manipulating both semantic and world knowledge congruity in the same experiment, is crucial for clarifying this issue. A first attempt was recently made by Martin, Garcia, Breton, Thierry, and Costa (2015), who reported that L2 and L1 readers similarly integrate world knowledge and semantic information, although L2 readers lack fast semantic access.

Finally, it is also important to note that we used spoken sentences as stimuli. Although most of the previous studies on L2 sentence comprehension have been carried out by using written sentences (Ardal et al., 1990; Weber-Fox & Neville, 1996, 2001; Moreno & Kutas, 2005; Ojima et al., 2005; Moreno et al., 2008; Braunstein et al., 2012; Newman et al., 2012; Martin, Thierry, Kuipers, Boutonnet, Foucart, & Costa, 2013; Martin et al., 2015; Foucart, Moreno, Martin, & Costa, 2015), the present results, along with previous studies on L2 speech comprehension (Hahne, 2001; Hahne & Friederici, 2001; FitzPatrick & Indefrey, 2010, 2014) suggest that the most frequently/reliably observed effects for violations during L2 comprehension (such as an extended N400 effect/late negativity) are modality-independent, and shared for semantic and world knowledge processing.

## **Conclusion**

The present study showed that the processing of world knowledge information during sentence comprehension is dependent on the listener's status (L1 vs. L2) in the target language. Relative to the case of L1 comprehension, world knowledge violations in the L2 elicit an extended late negative effect. This effect seems to be explained

mainly, if not exclusively, by the effort of integrating critical words that are incompatible with world knowledge stored in long-term memory, rather than a generic deficit of word anticipation in the L2.

## References

Ardal, S., Donald, M. W., Meuter, R., Muldrew, S., Luce, M. (1990). Brain responses to semantic incongruity in bilinguals. *Brain and Language*, 39, 187-205.

Braunstein, V., Ischebek, A., Brunner, C., Grabner, R. H., Stamenov, M., & Neuper, C. (2012). Investigating the influence of proficiency and semantic processing in bilinguals: An ERP and ERD/S analysis. *Acta Neurobiologiae Experimentalis*, 72, 421-438.

Clark, H. H. (1996). *Using language*. Cambridge University Press, Cambridge.

Davis, C. J., & Perea, M. (2005). BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37, 665-671.

Federmeier, K. D., & Lazslo, S. (2009). Time for meaning: electrophysiology provides insights into the dynamics of representation and processing in semantic memory. In *The Psychology of Learning and Motivation*, ed. B. H. Ross, pg. 1-44. San Diego, CA: Elsevier.

FitzPatrick, I., & Indefrey, P. (2010). Lexical competition in non-native speech comprehension. *Journal of Cognitive Neuroscience*, *22*, 1165-1178.

FitzPatrick, I., & Indefrey, P. (2014). Head start for target language in bilingual listening. *Brain Research*, *1542*, 111-130.

Forster, K. I. (1979). Levels of processing and the structure of the language processor. In *Sentence Processing: Psycholinguistic Essays Presented to Merrill Garrett*, ed. W. E. Cooper & C. T. Walker, pg. 27-85. Hillsdale, NJ: Erlbaum.

Foucart, A., Moreno, E., Martin, C. D., & Costa, A. (In press). Integration of moral values during L2 sentence processing. *Acta Psychologica*.

Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*, 883-899.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension. *Science*, *304*, 438-441.

Hahne, A. (2001). What's different in second-language processing? Evidence from event-related 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, 123-141.

Hald, L. A., Steenbeck-Planting, E. G., & Hagoort, P. (2007). The interaction of discourse context and world knowledge in online sentence comprehension. Evidence from the N400. *Brain Research*, 1146, 210-218.

Jackendoff, R. (2002). *Foundations of language: brain, meaning, grammar, evolution*. New York, NY: Oxford University Press.

Kutas, M., & Federmeier, K. D. (2001). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Science*, 4, 463-470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Reviews in Psychology*, 62, 621-647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*, 203-205.

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

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, *9*, 920-933.

Martin, C. D., 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*, 574-588.

Martin, C. D., Garcia, X., Breton, A., Thierry, G., & Costa, A. (2014). From literal meaning to veracity in two hundred milliseconds. *Frontiers in Human Neuroscience*, *8* (40), 1-12.

Martin, C. D., Garcia, X., Breton, A., Thierry, G., & Costa, A. (In press). World knowledge integration during second language comprehension. *Language, Cognition, and Neuroscience*. DOI: 10.1080/23273798.2015.1084012

Moreno, E. M., & Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish-English bilinguals. *Cognitive Brain Research*, 22, 205-220.

Moreno, E. M., Rodríguez-Fornells, A., & Laine, M. (2008). Event-related Potentials (ERPs) in the study of bilingual language processing. *Journal of Neurolinguistics*, 21, 477-508.

Newman, A. J., Tremblay, A., Nichols, E. S., Neville, H. J., & Ullman, M. T. (2012). The influence of language proficiency on lexical semantic processing in native and late learners of English. *Journal of Cognitive Neuroscience*, 24, 1205-1223.

Oliver, G., Gullberg, M., Hellwig, F., Mitterer, H., & Indefrey, P. (2012). Acquiring L2 sentence comprehension: A longitudinal study of word monitoring in noise. *Bilingualism: Language and Cognition*, 15, 841-857.

Rapp, D. N. (2008). How do readers handle incorrect information during reading? *Memory & Cognition*, 36, 688-701.

Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, 54, 574-591.

Strauß, A., Kotz, S. A., & Obleser, J. (2013). Narrowed expectancies under degraded speech: revisiting the N400. *Journal of Cognitive Neuroscience*, 25, 1383-1395.

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specialization for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8, 231-256.

Weber-Fox, C. M., & Neville, H. J. (2001). Sensitive periods differentiate processing of open- and closed-class words: An ERP study of bilinguals. *Journal of Speech, Language, and Hearing Research*, 44, 1338-1353.

**Table 1.** Examples of sentences with English translation.

1	<b><u>KS</u></b>	La religión mayoritaria en Italia es el <u>catolicismo</u> . <i>The majority religion in Italy is <u>Catholicism</u>.</i>
	<b><u>WKv</u></b>	La religión mayoritaria en Italia es el <u>judáismo</u> . <i>The majority religion in Italy is <u>Judaism</u>.</i>
	<b><u>US</u></b>	La religión mayoritaria en Etiopía es el <u>catolicismo</u> . <i>The majority religion in Ethiopia is <u>Catholicism</u>.</i>
2	<b><u>KS</u></b>	El ataque de las Torres Gemelas se produjo en <u>septiembre</u> . <i>The World Trade Center attack occurred in <u>September</u>.</i>
	<b><u>WKv</u></b>	El ataque de las Torres Gemelas se produjo en <u>diciembre</u> . <i>The World Trade Center attack occurred in <u>December</u>.</i>
	<b><u>US</u></b>	El ataque de Pearl Harbor se produjo en <u>diciembre</u> . <i>The Pearl Harbor attack occurred in <u>December</u>.</i>
3	<b><u>KS</u></b>	El número de jugadores en un equipo de fútbol es de <u>once</u> . <i>The number of players on a football team is <u>eleven</u>.</i>
	<b><u>WKv</u></b>	El número de jugadores en un equipo de fútbol es de <u>cuatro</u> . <i>The number of players on a football team is <u>four</u>.</i>
	<b><u>US</u></b>	El número de jugadores en un equipo de curling es de <u>cuatro</u> . <i>The number of players on a curling team is <u>four</u>.</i>

KS = Known sentences; WKv = World Knowledge violations; US = Unknown sentences. Critical words are underlined.

**Table 2.** Sentence and critical words properties.

	<b>KS</b>		<b>WKv</b>		<b>US</b>	
	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>	<i>Mean</i>	<i>SD</i>
<b>Number of words per sentence</b>	9.18	2.18	9.25	2.24	9.29	2.18
<b>Number of phonemes per CW</b>	5.24	0.99	5.30	1.01	5.30	0.96
<b>Number of syllables per CW</b>	5.95	2.13	6.30	2.10	6.27	2.06
<b>Onset of CW (ms)</b>	2509.20	743.73	2501.55	768.22	2598.55	756.23
<b>Duration of CW (ms)</b>	561.11	171.44	564.73	142.71	594.25	250.52
<b>Frequency of CW (log)</b>	1.28	0.94	1.05	0.90	1.08	0.86
<b>Familiarity of CW</b>	2.57	0.91	2.63	1.02	2.57	1.01
<b>Imageability of CW</b>	5.32	1.05	5.41	0.95	5.41	0.90

CW = Critical Words; KS = Known sentences; WKv = World Knowledge violations;

US = Unknown sentences; SD = standard deviation.

**Table 3.** Planned comparisons for the three-way interaction Language x Condition x Topography in the ANOVA carried out for the 500-750 ms time window.

<u>Language</u>	<u>Condition</u>	<u>Region</u>	<u>F</u>
L1 vs. L2	KS	Frontral	1.32
L1 vs. L2	KS	Central	0.08
L1 vs. L2	KS	Posterior	0.03
L1 vs. L2	WKv	Frontral	0.61
L1 vs. L2	WKv	Central	2.21
L1 vs. L2	WKv	Posterior	<b>4.18 *</b>
L1 vs. L2	US	Frontral	0.87
L1 vs. L2	US	Central	0.25
L1 vs. L2	US	Posterior	0.26
L1	KS vs. WKv	Frontral	3.56
L1	KS vs. WKv	Central	3.79
L1	KS vs. WKv	Posterior	2.00
L1	KS vs. US	Frontral	1.18
L1	KS vs. US	Central	3.59
L1	KS vs. US	Posterior	2.46
L1	WKv vs. US	Frontral	0.77
L1	WKv vs. US	Central	0.13
L1	WKv vs. US	Posterior	0.12
L2	KS vs. WKv	Frontral	1.83
L2	KS vs. WKv	Central	<b>17.86 **</b>
L2	KS vs. WKv	Posterior	<b>21.91 **</b>
L2	KS vs. US	Frontral	0.36
L2	KS vs. US	Central	<b>5.12 *</b>
L2	KS vs. US	Posterior	<b>6.09 *</b>
L2	WKv vs. US	Frontral	0.63
L2	WKv vs. US	Central	<b>4.42 *</b>
L2	WKv vs. US	Posterior	<b>4.44 *</b>

\* =  $p$  value < .05. \*\* =  $p$  value < .001.