

World Knowledge integration during second language comprehension

Running Head: World knowledge integration in L2

Clara Martin^{1,2}, Xavier Garcia³, Audrey Breton⁴, Guillaume Thierry^{5,6}
& Albert Costa^{3,7}

1 BCBL. Basque Center on Cognition, Brain and Language, San Sebastian, Spain

2 IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

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

4 Institut des Sciences Cognitives, Université de Lyon – CNRS, Lyon, France

5 School of Psychology, Bangor University, Bangor, UK

6 Economic and Social Research Council Centre for Research on Bilingualism in theory and Practice, Bangor University, Bangor, UK

7 ICREA, Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain

Corresponding Author

Dr. Clara D. Martin

Basque Center on Cognition, Brain and Language

Parque Tecnológico de Miramón

Paseo Mikeletegi 69, 2nd

20009 San Sebastian

Spain

Work Phone: 00 34 943 309 300 (ext: 204)

Fax: 00 34 943 309 052

E-mail: c.martin@bcbl.eu

Abstract

In order to study the difficulties experienced during sentence comprehension in a foreign language (L2), we investigated semantic and world knowledge information retrieval in L2 comprehenders. Event-related potentials (ERP) were collected in late learners of English whose native language is Spanish, performing a sentence reading task in English. We investigated the mean amplitude of the P2 and N400 ERP components elicited by the critical word of sentences in three conditions: (1) correct; (2) semantic violation; (3) world knowledge violation (semantically acceptable but factually untrue). In the N400 window, ERP modulations elicited by semantic and world knowledge violations had similar amplitudes, as previously observed in L1 comprehenders. However, semantic violations failed to modulate P2 mean amplitude as it did in native speakers. These results suggest that, whilst L2 and L1 readers similarly integrate world knowledge and semantic information, L2 readers lack fast semantic access at around 200 ms after the onset of a critical word.

Keywords

Bilingualism; Sentence comprehension; Semantic integration; ERPs

Word count: 6369

Introduction

Speakers of a foreign language often experience comprehension difficulties while listening to or reading in their second language (L2). These difficulties may stem from different sources, revealing imperfect representations at different levels, e.g., phonological, syntactic, or lexico-semantic (Soares & Grosjean, 1984; Weber-Fox & Neville, 1996; Hahne, 2001; Hahne & Friederici, 2001; Weber-Fox et al., 2003; Sanders & Neville, 2003; Moreno & Kutas, 2005; FitzPatrick & Indefrey, 2010). Furthermore, and given the fast rate at which natural language is experienced, problems circumscribed at any of these levels may percolate to other levels, thus having an impact on overall language comprehension. For example, difficulties during speech segmentation can slow down lexical retrieval. Many researchers have explored these difficulties in order to understand how L2 comprehension departs from native processing (see Moreno et al., 2008 for a review). In the present study, we contribute to this aim by assessing how L2 comprehenders retrieve and integrate world knowledge during sentence comprehension.

Thirty years of research have repeatedly reported ERP modulations when reading or listening to sentences in which a critical word violates the semantic integrity of the message (see Kutas & Federmeier, 2011 for a review). It is widely acknowledged that the N400 ERP component, a negative deflection peaking at around 400 ms after stimulus onset, increases in amplitude when the stimulus (e.g., a critical word in a sentence) is semantically incorrect as in “I take my coffee with cream and *dog*” (Kutas & Hillyard, 1980). Thus, the N400 is considered an index of semantic integration difficulty during sentence processing, and the larger the semantic anomaly, the larger the N400 amplitude (see Kutas & Federmeier, 2011, Moreno et al., 2008 for reviews).

Retrieving literal semantic information is crucial for language comprehension, but retrieving factual knowledge about the world and integrating it with the former is obviously critical also. Indeed, comprehenders do not only need to access semantic information, they also need to check whether such information is compatible with world knowledge in order to make sense of the message. Such a process allows us to realise that the sentence “I like to have a glass of wine before sleeping” is unacceptable if it is said by a child (cf. van Berkum et al., 2008), or

to realise that someone saying “In New York the taxis are green” is lying or confused (cf. Hagoort et al., 2004; Martin et al., 2014).

Hagoort and collaborators (2004) investigated the time-course of world knowledge integration using the N400 component commonly associated with semantic integration (Kutas & Hillyard, 1980). In their study, participants were asked to read sentences containing either a world knowledge or a semantic violation. They compared the ERPs elicited by the critical words that completed sentences in three conditions: (1) correct and true sentences as “In New York the taxis are *yellow*” (critical word in italics); (2) sentences with world knowledge violations as “In New York the taxis are *green*”. The sentence is semantically correct in the sense that taxis can be green but it is in conflict of our knowledge of the world; (3) sentences with semantic violations as “In New York the taxis are *sour*” in which it is not only false that in New York taxis are sour, it is impossible. Our knowledge about the words of a language and their meaning makes us realise that this last sentence is incoherent (Marslen-Wilson et al., 1988; Hagoort et al., 2004). Hagoort et al. (2004) found that the N400 component elicited by semantic and world knowledge violations had a similar latency. They interpreted this result as suggesting that semantic and world knowledge information are integrated in parallel and online during language comprehension.

Following Hagoort and collaborator's (2004) seminal study, we recently investigated the time-course of world knowledge integration in English native speakers (Martin et al., 2014). We compared the ERPs elicited by critical words in three conditions: (1) correct and true, e.g., “The football player Maradona was a *forward* in the Argentinean team” (critical word in italics); (2) sentences with world knowledge violations as “The football player Maradona was a *goalkeeper* in the Argentinean team”; (3) sentences with semantic violations as “The football player Maradona was a *dress* in the Argentinean team”. Similarly to Hagoort and collaborators (2004), we observed that the words “goalkeeper” and “dress” in the example above are detected as violations with a similar latency. Thus, native comprehenders appear to retrieve lexico-semantic information (indexed by semantic violations) and integrate such information with world knowledge (indexed by world knowledge violations) simultaneously, at around 400 ms after the

critical word onset (i.e., the time range of the N400 ERP component; Martin et al., 2014; Hagoort et al., 2004). Importantly, our study also revealed effects of literal semantic access as early as 200 ms after the onset of the critical word. That is, semantic violations (but not world knowledge violations) modulated the P2 as compared to correct sentences. P2 mean amplitude was significantly smaller for semantic violations compared to both world knowledge violations and correct sentences (those two conditions did not differ). Therefore, native comprehenders appear to benefit from semantic access and contextual integration before they take into account world knowledge (see also Martín-Loeches et al., 2004; Landi and Perfetti, 2007; Penolazzi et al., 2007; Pinheiro et al., 2010; Regel et al., 2010). Such P2 event modulation by semantic violation has been reported previously in several studies. In a sentence reading task, Pinheiro and colleagues (2010) reported smaller P2 amplitude for semantically incongruent as compared to congruent critical words. Penolazzi and collaborators (2007) also observed effects of semantic context integration around 200 ms after the critical word onset during sentence reading (see also Landi and Perfetti, 2007). Using eye-tracking, Rayner and colleagues (2004) measured eye movements in participants reading (a) correct sentences (e.g., *John used a knife to chop the large carrots for dinner*), (b) sentences with world knowledge violations (e.g., *John used an axe to chop the large carrots for dinner*) and (c) sentences with literal semantic violations (e.g., *John used a pump to inflate the large carrots for dinner*). They observed that semantic violations disrupted eye movements earlier than world knowledge violations in the course of reading (see also Warren & McConnell, 2007; Warren et al., 2008).

Here, we explore the presence of such effects in a foreign language (English). P2 was used as an index of fast (early) literal semantic access. N400 was used as an index of simultaneous integration of semantic and world knowledge information integration.

Previous research suggests that L2 sentence comprehension significantly departs from native sentence comprehension. For instance, N400 modulations elicited by semantic violations tend to be smaller in L2 than L1. Hahne (2001) found a reduced N400 effect in participants reading in L2 as compared to matched controls reading in L1. In a within-group comparison,

Kutas and Kluender (1994) found that N400 effect amplitudes were smaller in the lesser proficient language of a bilingual, an effect also reported in highly proficient bilinguals (Proverbio et al., 2002). Semantic processing (as indexed by semantic violation) may thus proceed differently in a first and second language (see Moreno et al., 2008 for a review). However, several studies failed to show differences in N400 amplitudes between L1 and L2 semantic processing (see for instance Moreno & Kutas, 2005, Weber-Fox & Neville, 1996). As suggested by Moreno et al. (2008), it is important to start investigating other aspects of semantic processing than those characterised by violations, especially in the field of bilingualism. Our study makes a first step in that direction by exploring world knowledge retrieval and integration. The current study has two main aims: First, there already is evidence that L2 readers do not use contextual information as L1 readers do when sentences do not contain outright literal semantic violations (see for instance Kaan, 2014; Martin et al., 2013). With this study, we propose to go a step further by investigating world knowledge violations. Second, we set out to characterise early semantic access in L2 sentence comprehension. More specifically, we test whether L2 readers benefit from early semantic access in the P2 time-window before stages of integration between word meaning, paralinguistic information, and long-term memory representations (N400 time-window), as it was shown in L1 readers (Martín-Loeches et al., 2004; Landi and Perfetti, 2007; Penolazzi et al., 2007; Pinheiro et al., 2010; Regel et al., 2010; Martin et al., 2014).

Our main hypothesis was that L2 comprehenders might not be able to integrate semantic and world knowledge information online as has been observed in L1 comprehenders. If so, one would expect ERP modulations associated to world knowledge violations to appear in a later time-window than literal semantic violations (expected to elicit N400 modulations). This means that significant differences in ERP mean amplitudes between semantic violations and correct sentences should be found in the N400 time-window, whereas amplitude differences between world knowledge violations and correct sentences would appear in a later time-window. Moreover, we anticipate lesser efficient semantic integration in L2 readers, and

therefore ERP modulations by literal semantic violations to be limited to the N400 time-window, with no measurable P2 modulation.

Material and Methods

Participants.

Eighteen late Spanish-English bilinguals (11 females; 21.7 years \pm 2.8) who learned English mainly at school (mean age of first exposure to English = 7.4 years \pm 2.4) took part in the experiment. They were administered the Essex L2 test to assess their English vocabulary knowledge (mean proportion of correct responses = .72 \pm .16) and they self-rated their English reading proficiency on a five level Likert scale (from 1 = “*low proficient*” to 5 = “*high proficient*”, Median = 4). All participants gave written consent to take part in the study that was approved by the ethics committee of Bangor University, Wales, UK.

Task and procedure.

Stimuli consisted of 120 sentences with three different versions such as “Before the age of eight, children start to *read/smoke/bark* and to write.” (critical words are in italics; see Table 1 for more examples). The first version of this sentence is semantically acceptable and true. The second version is literally correct but factually false (world knowledge violation). The third version elicits a semantic violation. Three lists of 120 sentences were created, each of them containing 40 correct sentences, 40 sentences with world knowledge violations and 40 sentences with semantic violations. Each sentence context was used only once per list. Each of the three versions of a sentence was used once among the three lists with participants randomly assigned to each list. The 120 sentences were mixed with 120 filler sentences, which were not analyzed. Filler sentences were semantically and syntactically congruent and did not refer to common or general knowledge (e.g., “Peter waited for Anna because he wanted to speak to her”). The order of sentences was pseudo-randomised across participant.

Table 1: Examples of sentences used as experimental material.

Sentences	Conditions		
	Correct	WK violation	Semantic violation
Before the age of eight, children start to ... and to write.	read	smoke	bark
People go to parks when they want to ... and have a walk.	rest	buy	bite
When it is rainy, people cannot ... as though it's sunny.	tan	speak	meow
Mines are ... and dangerous.	dark	crowded	happy
During summer, many women wear ... and dresses.	sandals	boots	carrots
During underwater diving sessions it is common to see ... and starfish.	jellyfish	eagles	smells
The Beatles were ... in the 60's.	popstars	lawyers	horses
The Egyptian pyramids are very ... buildings.	old	small	savory
Santa Claus is very ... and famous.	friendly	young	bumpy
The football player Maradona was a ... in the Argentinean team.	forward	goalkeeper	dress
Everest is a ... and tall mountain.	snowy	tropical	studious
Pope Benedict XVI is ... and lives in the Vatican.	German	Asian	pollinated

The critical word in correct sentences was neither the only possible candidate nor the most expected candidate. For instance, the sentence “Before the age of eight, children start to...” can be completed with the verbs *speak*, *walk*, *go to school*, *read*, etc. A Cloze probability¹ rating test was administered to 22 Spanish-English bilingual participants who did not participate in the experiment, and who had similar levels of English as the L2 participants of the ERP experiment. The critical word of correct sentences had an averaged cloze probability of 8.9% \pm 1 and was, on average, the third most expected word (Average cloze probability of the first and second best completions: 35.2% \pm 1.3 and 15.6% \pm .6 respectively). The critical words of sentences with world knowledge violations and semantic violations had an averaged cloze probability of 0%. In addition, the critical word was never the last word of the sentence (averaged number of words following the critical word = 2.5 \pm 1.0). The critical words were matched across conditions on the following criteria: average length in characters ($p = 0.90$) and syllables ($p = 0.62$), log word frequency ($p = 0.17$), concreteness ($p = 0.23$), imageability ($p = 0.20$) and word class (equated within each pair; see Table 2 for numerical values). Finally, working memory requirements were balanced between semantic and world knowledge violations: The distance between the violation and the word in the sentence that revealed the violation did not significantly differ

¹ Cloze probability of a word in a particular sentence is defined as the percentage of time it is produced by a group of participants asked to complete the sentence.

between conditions (3.6 ± 1.6 words in the WK violation condition; 3.4 ± 1.5 words in the semantic violation condition; t-test: $p = 0.18$).

Table 2: Critical word criteria controlled across conditions.

	CS	WK	SV
Length in characters	6.4 (2.1)	6.3 (2.1)	6.3 (2.0)
Syllable number	1.8 (.7)	1.7 (.8)	1.8 (.8)
Log- word frequency	1.6 (.7)	1.6 (.6)	1.4 (.5)
Concreteness	510 (115)	472 (110)	507 (108)
Imageability	521 (99)	514 (78)	546 (85)

CS = correct sentences; WK = sentences with world knowledge violations; SV = sentences with semantic violations. Standard deviations are reported into bracket.

Each sentence was presented one word at a time (200 ms duration and 500 ms Stimulus Onset Asynchrony) in the centre of a CRT monitor. Sentences were separated by a fixation cross displayed for 800 ms. The instruction was to read each sentence silently and answer yes or no to the comprehension question, when applicable, i.e., in $\frac{1}{4}$ of the trials, by pressing designated buttons on a response pad. The latter quiz test ensured that participants processed sentence meaning during silent reading.

In order to take into account individual world knowledge in data processing, a surprise follow-up test was administered at the end of the experiment. The goal of the test was to determine which sentences were true or false for each individual (since semantic violations were obvious violations, sentences featuring such violations were not rated in the follow-up test). ERP data analyses were conducted based on individual truth-value assessment (rather than that pre-defined by the experimenter when creating the material). The 40 true and 40 false sentences used during the experiment were presented on the screen, one at a time, along with a rating scale (in a randomized order). Participants had to rate each sentence as true or false by pressing “1” or “2”. They had to press “3” if they did not know if the sentence was true or false and “4” if they could not decide because the sentence was meaningless.

Electrophysiological recording and data analyses.

Electrophysiological data (EEG) were recorded using a Brain Vision Recorder (version 1.10; Brain Products, inc., München) in reference to an electrode placed on the participant's nose at a frequency rate of 500 Hz from 31 tin electrodes placed according to the 10-20 convention (Jasper, 1958). Vertical and horizontal electrooculograms were recorded simultaneously with EEG. Impedances were kept below 5 kOhm. EEG activity was filtered off-line with a 30 Hz (48 dB) low-pass filter and a 0.1 Hz (12 dB) high-pass filter. Eye-blink artifacts were mathematically corrected using the procedure recommended by Gratton and Coles (1989), implemented in Brain Vision Analyzer 2.0 (Brain Products, inc.) and any remaining artifacts were manually dismissed. Epochs ranged from -100 to 700 ms after the onset of the critical word. Baseline correction was performed in reference to pre-stimulus activity (from -100 to 0 ms) and individual averages were digitally re-referenced offline to the mean of left and right mastoid channels. P2 and N400 ERP components were analyzed over a subset of 18 electrode sites (Left Frontocentral scalp: F3, FC1, FC5; Right Frontocentral: F4, FC2, FC6; Left Centroparietal: CP1, CP5, C3; Right Centroparietal: CP2, CP6, C4; Left Parietooccipital scalp: P3, P5, PO1; Right Parietooccipital scalp: P4, P6, PO2). P2 mean amplitude was measured as the average of the ERP amplitude in the [150–200] ms time-window and N400 mean amplitude was measured as the average of the ERP amplitude in the [350–550] ms time-window (based on the time-windows defined by Martin et al., 2014). Mean amplitudes of the P2 and N400 peaks were analyzed using a 3 x 3 x 2 repeated measure analysis of variance (ANOVA). The ANOVA factors were Condition (Correct sentence (CS) *versus* World Knowledge violation (WK) *versus* Semantic violation (SV)), Region (Frontal *versus* Central *versus* Parietal) and Hemisphere (Left *versus* Right) as within subject factors.

Results

Behavioural results

Accuracy in the quiz test was of 87.0% \pm 8.5. In the follow-up test, participants rated 79% \pm 3 of correct sentences as true (8% \pm 2 as false and 13% \pm 2 as “Don’t know”). They rated 71% \pm 4 of WK sentences as false (10% \pm 2 as true and 19% \pm 3 as “Don’t know”). In order to take into account individual world knowledge, four ERP conditions were computed: (1) correct sentences: average of all sentences individually rated as true in the follow-up test; (2) world knowledge violations (WK): average of all sentences individually rated as false in the follow-up test; (3) “don’t know” sentences (DK) for which participants had insufficient knowledge to make a decision; and (4) semantic violations (SV). Overall, 30% \pm 4 of the sentences were considered as correct, 26% \pm 5 as WK, 10% \pm 6 as DK, and 33% \pm 0 as SV. Among the 30% of sentences considered as correct, 89% \pm 2 were originally true and 11% \pm 2 were false. Among the 26% of sentences considered as world knowledge violations, 90% \pm 2 were originally false and 10% \pm 2 were true. Among the 10% of sentences of the Don’t know condition, 43% \pm 5 were originally true and 57% \pm 5 were false. Individual assessment varied across sentences. In 13% of the cases, experimenter and individual assessments converged maximally for correct sentences and world knowledge violations (rated as True or False, respectively, by all except one participants). In 8% of the cases, correct sentences were rated as True and world knowledge violations were rated as False by less than half of the participants (mostly “Don’t know” answers). These observations lead to computing averages based on individual assessment rather than experimenter knowledge. Note that sentences were re-categorized based on individual participants’ ratings (rather than average rating from all participants).

When each condition was computed taking into account individual assessment, the critical word of true sentences had an averaged cloze probability of 7.9% \pm .4. The critical words of sentences with world knowledge violations and semantic violations had an averaged cloze probability of 1.2% \pm .3 and 0% respectively. The critical words of “Don’t know” sentences had an averaged cloze probability of 5.4% \pm 1.6.

ERP results

ERP were computed on the basis of individual assessment and after dismissal of trials with artifacts. The average percentage of data excluded due to artifact rejection was $4.4\% \pm 6.6$ for the True condition, $4.1\% \pm 6.0$ for the False condition and $4.1\% \pm 6.3$ for the SV condition. Because of a small number of epochs and large inter-individual variability, the “Don’t know” sentences were not included in the ERP analysis. Table 3 shows the ANOVA results on P2 and N400 mean amplitudes. Figure 1 depicts the ERPs elicited by correct sentences, semantic violations and world knowledge violations.

Table 3: General ANOVA for CS versus WK versus SV comparison.

	dF	P2 component		N400 component	
		F value	p value	F value	p value
Condition	2, 34	.43	.65	<i>4.36</i>	<i>.02</i>
Hemisphere	1, 17	.49	.50	.08	.78
Region	2, 34	1.02	.37	.89	.42
Condition x Hemisphere	2, 34	.83	.45	1.23	.31
Condition x Region	4, 68	.61	.66	.76	.55
Hemisphere x Region	2, 34	.07	.94	.33	.72
Condition x Hemisphere x Region	4, 68	.14	.97	.91	.46

dF = degree of freedom; Significant effects are labelled in bold italic.

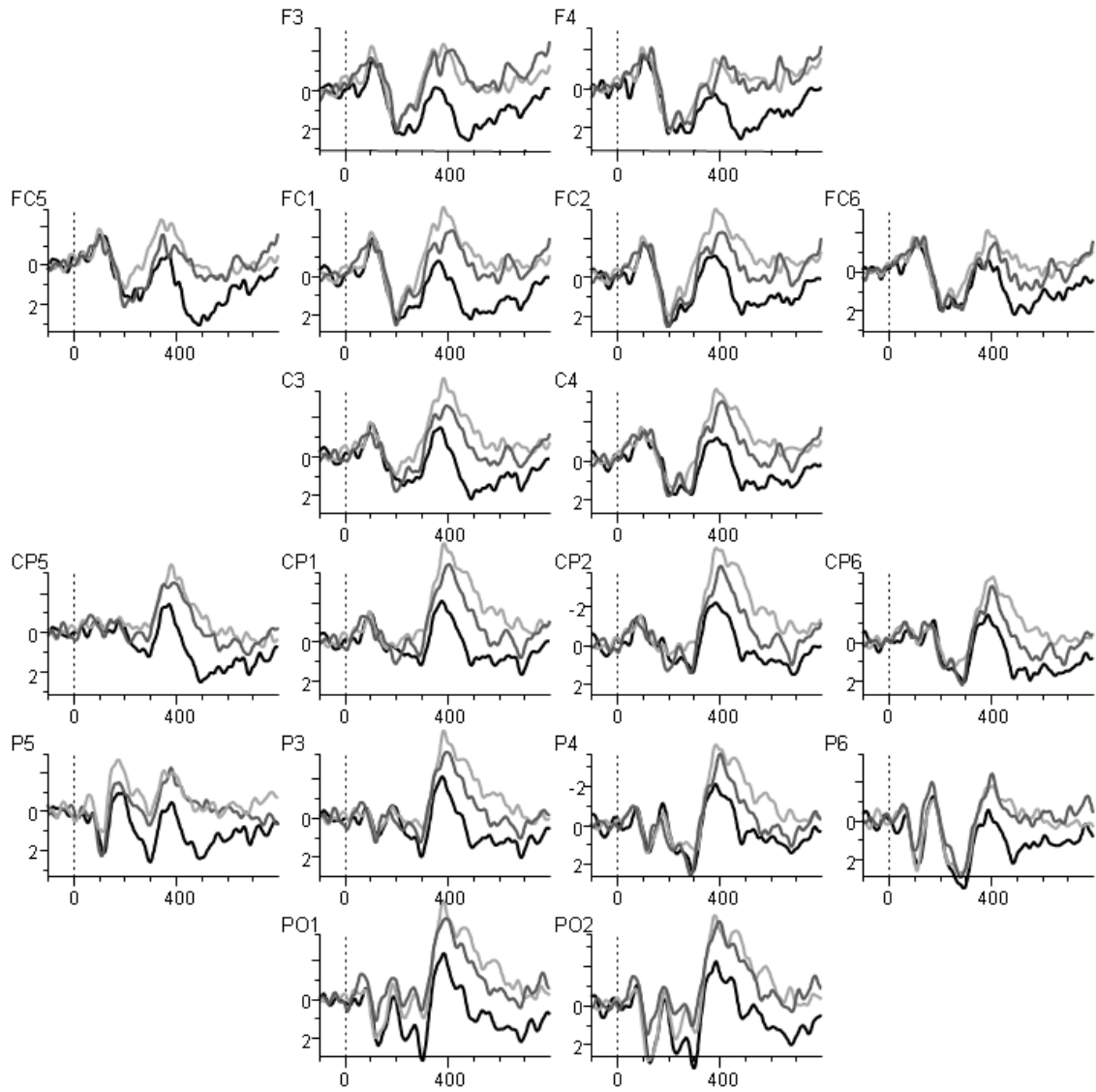


Figure 1: Grand averaged ERPs in the [-100; 700] ms time-window over the 18 electrodes included in the analyses (F3, F4, FC1, FC2, FC5, FC6, CP1, CP2, CP5, CP6, C3, C4, P3, P4, P5, P6, PO1, PO2). Grand averages for L2 comprehenders exposed to correct sentences (black lines), sentences with world knowledge violations (dark grey lines) and sentences with semantic

violations (light grey lines). Time zero corresponds to the onset of the critical word of the sentence. Negativity is plotted up.

P2 time-window

The ANOVA performed on the P2 mean amplitude did not reveal any significant effect or interaction (see Table 3 for statistical results and Figure 1). The mean amplitude of the P2 component was neither modulated by semantic or world knowledge violations².

N400 time-window

The ANOVA performed on the N400 mean amplitude revealed a significant effect of condition. No other effect or interaction reached significance (see Table 3 for statistical results and Figure 1). Post-hoc analysis of the condition effect (Bonferroni test) revealed that both conditions with violations differed from the correct sentence condition. The N400 mean amplitude was larger for sentences with world knowledge violations than correct sentences ($p = .03$). The N400 was also larger for sentences with semantic violations than correct sentences ($p = .009$). The two conditions with violations did not significantly differ ($p = .58$).

Between-group comparison

For the sake of completeness, we performed an ANOVA with the same within-subject factors and with group as a between-subject factor. This allowed us to make a direct comparison between the ERP results obtained in the present group of L2 comprehenders and the ones previously obtained in a group of L1 comprehenders. As reported in Martin et al. (2014), this

² We ran another ANOVA to ensure that the absence of effect in the P2 time-window was not due to individual variability. In fact, given the large variability in L2 populations, we assumed that the P2 component might be sensitive to semantic violations in some participants and not in others. We looked at individual data and split the 18 participants in two groups of nine: one sub-group of participants in which the mean amplitude of the P2 component tended to be smaller for semantic violations than for correct sentences and one sub-group with no tendency for any modulation of the P2 component. Data from the first sub-group were submitted to a similar ANOVA, which did not reveal, as for the entire group, any modulation of the P2 component. The effect of condition was still far from being significant ($F[2, 16] = 2.12, p = .15$).

control group was made of 18 native English speakers who performed the exact same task, following the same procedure.

The results of the ANOVA performed on P2 mean amplitudes are reported in table 4 (see also Figure 2 for a comparison of ERP waves in L2 and L1 comprehenders from Martin et al., 2014).

Table 4a: General ANOVA for group comparison in the P2 time-window

	P2 component		
	dF	F value	p value
Group	1, 34	9.99	.004
Condition	2, 68	1.38	.26
Hemisphere	1, 34	1.39	.25
Region	2, 68	1.26	.29
Condition x Group	2, 68	3.61	.033
Condition x Hemisphere	2, 68	1.09	.34
Condition x Region	4, 136	1.43	.23
Hemisphere x Group	1, 34	.15	.70
Hemisphere x Region	2, 68	.13	.88
Region x Group	2, 68	5.12	.009
Condition x Hemisphere x Group	2, 68	.97	.38
Condition x Region x Group	4, 136	.62	.65
Hemisphere x Region x Group	2, 68	.002	1.00
Condition x Hemisphere x Region	4, 136	.08	.99
Condition x Hemisphere x Region x Group	4, 136	.12	.97

Table 4b: P2 Post-hoc analysis – Bonferroni test of the Condition x Group interaction

	L1 comprehenders	L2 comprehenders
SV versus WK	.017	.92
SV versus CS	.015	.24
WK versus CS	.94	.28

Table 4c: P2 Post-hoc analysis – Bonferroni test of the Region x Group interaction

	L1 comprehenders	L2 comprehenders
Frontal versus Central	.18	1.00
Frontal versus Parietal	.048	1.00

CS = Correct sentences; WK = World knowledge violations; SV = Semantic violations; dF = degree of freedom; Significant effects and interactions are labelled in bold italic.

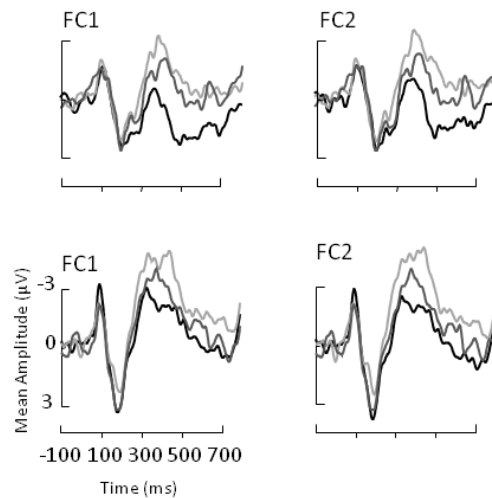


Figure 2: Grand averaged ERPs in the [-100; 700] ms time-window over FC1 and FC2 electrodes. **Up:** Grand averages for L2 comprehenders exposed to correct sentences (black lines), sentences with world knowledge violations (dark grey lines) and sentences with semantic violations (light grey lines). **Down:** Equivalent grand averages for L1 comprehenders exposed to the same sentences (data from Martin et al., 2014). Time zero corresponds to the onset of the critical word of the sentence. Negativity is plotted up.

There was a main group effect showing that the P2 component was significantly larger in L1 than in L2 comprehenders. Furthermore, we observed the expected condition x group interaction: The P2 was significantly smaller for semantic violations than the two other conditions in L1 comprehenders (see Martin et al., 2014) whereas its mean amplitude was

neither modulated by semantic nor world knowledge violations in L2 comprehenders. There was also a significant region x group interaction showing that P2 was larger over frontal than parietal regions in L1 but not L2 comprehenders.

Results of the ANOVA performed on N400 mean amplitudes are reported in table 5 (see also Figure 2). There was a main group effect showing that the N400 was significantly larger in L1 than in L2 comprehenders. Furthermore, there was a significant effect of condition and a significant condition x hemisphere x group interaction. Post-hoc analyses (Bonferroni test) confirmed previous results reported in L1 comprehenders: The three conditions significantly differed over both hemispheres. Furthermore, the N400 component was larger over the right hemisphere in semantic violations but not in the two other conditions (see Martin et al., 2014). In L2 comprehenders, correct sentences significantly differed from the two other conditions over both hemispheres. Semantic and world knowledge violation conditions did not differ neither over the right nor the left hemisphere. Moreover, the N400 mean amplitude did not differ between the two hemispheres in none of the three conditions.

Table 5a: General ANOVA for group comparison in the N400 time-window

	N400 component		
	dF	F value	p value
Group	1, 34	4.47	.042
Condition	2, 68	10.23	<.001
Hemisphere	1, 34	3.69	.063
Region	2, 68	.31	.73
Condition x Group	2, 68	1.04	.36
Condition x Hemisphere	2, 68	.93	.40
Condition x Region	4, 136	.61	.65
Hemisphere x Group	1, 34	2.49	.12
Hemisphere x Region	2, 68	.46	.63
Region x Group	2, 68	1.36	.26
Condition x Hemisphere x Group	2, 68	3.24	.045
Condition x Region x Group	4, 136	.97	.43
Hemisphere x Region x Group	2, 68	1.34	.27
Condition x Hemisphere x Region	4, 136	.75	.56
Condition x Hemisphere x Region x Group	4, 136	1.64	.17

Table 5b: P2 Post-hoc analysis – Bonferroni test of the Condition x Hemisphere x Group interaction

		Left	Right		SV	W K	CS
L1 comprehenders	SV versus WK	<.00 1	<.001	Left versus Right	<.001	.32	.09
	SV versus CS	<.00 1	<.001				
	WK versus CS	<.00 1	.02				
		Left	Right		SV	W K	CS
L2 comprehenders	SV versus WK	.73	.08	Left versus Right	.28	.73	.14
	SV versus CS	<.00 1	<.001				
	WK versus CS	<.00 1	<.001				

CS = Correct sentences; WK = World knowledge violations; SV = Semantic violations; dF = degree of freedom; Significant effects and interactions are labelled in bold italic.

Discussion

We assessed semantic integration in bilingual participants reading in their L2 to characterise the interplay between world knowledge and word meaning integration. In particular, we tested whether or not these two types of information are integrated in parallel as in the case of L1 comprehension (Martin et al., 2014; Hagoort et al., 2004). To answer that question, we investigated the ERP N400 elicited by the critical word of sentences in three conditions: (1) correct sentences; (2) sentences with semantic violations; (3) sentences with world knowledge violations. We also investigated P2 modulations since they have been shown sensitive to fast literal semantic access in L1 comprehension (Martin et al., 2014; Landi and Perfetti, 2007; Penolazzi et al., 2007; Pinheiro et al., 2010).

The N400 was significantly larger for semantic violations and world knowledge violations as compared to correct sentences and we found no difference in the P2 range. Whilst the N400 result is similar to that previously reported for L1 comprehension (Martin et al., 2014; Hagoort et al., 2004), the absence of a P2 modulation contrasts with the previously reported

differences between semantic violations and both world knowledge violations and correct sentences (Martin et al., 2014; see Figure 2).

N400 time-window

Perhaps the most interesting result of our study is the observation that the time-relationship between world knowledge and semantic integration is qualitatively similar in L2 and native sentence comprehension. This result suggests that L2 comprehenders –just like native comprehenders (Hagoort et al., 2004; Martin et al., 2014)– simultaneously integrate semantic information about words and world knowledge information with the preceding sentence context in the course of language processing. Thus, we conclude that comprehenders of a second language use information from different sources in parallel to make sense of the message (see Kelly et al., 2004; van Berkum et al., 2008; van den Brink et al., 2012; Hagoort et al., 2004; Martin et al., 2014 for similar argument in native speakers). Thus, difficulties in L2 comprehension cannot merely result from relatively slower integration of world knowledge during sentence comprehension.

Nevertheless, N400 modulation was greater for semantic than world knowledge violations in previous studies of L1 comprehenders (Hagoort et al., 2004; Martin et al., 2014). One can speculate regarding the similar N400 magnitude observed for the two types of violations in the present study. First, it could be that L2 comprehenders are less sensitive to semantic violations than L1 comprehenders because of their relatively lower proficiency in the language. This would be reflected in a smaller N400 effect for semantic violations, making the effects of semantic and world knowledge violations more similar. However, we did not find significant correlations between L2 proficiency and N400 effect magnitude (all $p > .46$ for correlations with Essex scores, English reading proficiency and mean age of first exposure to English). One can also speculate that L1/L2 N400 effect differences relate to a different organisation of –or access to– L1 and L2 long-term semantic memory representations (see Federmeier and Kutas, 1999a,b; Federmeier et al., 2002; Martin et al., 2014). Because of the organisation of long-term semantic memory, a specific sentence context primes various entries

of a semantic field. When the critical word is encountered, the closer it is to the activated semantic field, the lower the N400 deflection elicited. In other words, the N400 is mostly sensitive to semantic relatedness between the critical word and previous words in the sentence context (that is, lexico-semantic priming). Indeed, N400 effects are usually smaller for within- than between-category violations (Federmeier and Kutas, 1999a,b; Federmeier et al., 2002). Here, world knowledge violations can be compared to within-category violations (relatively strong lexico-semantic priming since world knowledge violations concern a given semantic domain). On the other hand, semantic violations can be assimilated to between-category violations (relatively weak lexico-semantic priming since semantic violations require no relationship between critical word and sentence context). Long-term memory organisation could thus explain the relatively smaller N400 effect observed for world knowledge violations in native speakers (Martin et al., 2014). Assuming that sentence context does not similarly prime semantic fields in L2 comprehenders, within- and between-category violations become more comparable, yielding similar N400 modulations. Thus, a critical difference between L1 and L2 sentence comprehension would be that L2 sentence context does not lead to semantic field priming, and thus does not lead to facilitated within-category word integration. This assumption will need testing further in studies of within- and between-category violations in second language comprehension. We note that the absence of a differential N400 effect between semantic and world knowledge violations in L2 is a null result that could be due to a lack of statistical power.

P2 time-window

Martin et al. (2014) reported an early effect of semantic violations in the P2 time window (see Landi and Perfetti, 2007; Penolazzi et al., 2007; Pinheiro et al., 2010 for similar early semantic effects) which they interpreted as a sign of early discrimination between information that is semantically interpretable or contextually meaningless, before world knowledge is taken into account. Based on the framework of long-term memory organisation presented above, this would mean that semantic (between-category) violations would be quickly

recognized as contextually meaningless and integrated differently from within-category violations and correct critical words (which are both semantically interpretable) as early as 200 ms after critical word onset. Assuming that semantic field priming from sentence context is absent in L2 comprehension, no difference should be observed between the different conditions in the P2 time-window, which is what we observed. Nevertheless, one cannot exclude the possibility that early discrimination between semantically interpretable and contextually meaningless information exists in L2 comprehension but is delayed. In that case, the process would take place later and resulting ERP modulations would be spill over to N400 time-window. Therefore, word integration during L1 sentence comprehension seems to unfold in two phases: First, early discrimination between what is contextually meaningful and what is not, leading to between-category (semantic) violation detection. Second, word meaning, paralinguistic information, and long-term memory representations are integrated so that the truth-value of sentences is appreciated (see Forster, 1979). In this framework, L2 comprehension difficulties may thus be accounted for by absent (or slower) primary semantic integration in L2. We would expect eye-tracking studies to validate this hypothesis by showing that fast and early disruptive effects of semantic violations on eye movements are absent in L2 (Rayner et al., 2004; Warren & McConnell, 2007; Warren et al., 2008).

Hemispheric differences between L1 and L2 comprehenders

The N400 recorded in L2 comprehenders had a bilateral distribution in the three conditions (we found no effect of hemisphere and no hemisphere x condition interaction). However, Martin et al. (2014) showed that N400 elicited by semantic violations was larger over the right hemisphere. A possible interpretation of this difference may be that L1 comprehenders benefit from higher levels of hemispheric specialisation for semantic information retrieval and integration as compared to L2 comprehenders (cf. predictive versus integrative processing theory; Federmeier and Kutas, 1999b; Kutas and Federmeier, 2000; Wlotko and Federmeier, 2007; Federmeier et al., 2008). Possible differences between L1 and L2 comprehenders in terms

of the neural organisation underlying semantic integration awaits further investigation (see Ardal et al., 1990; Proverbio et al., 2002; Moreno et al., 2008).

Conclusion

L2 comprehenders –like L1 comprehenders– retrieve and integrate semantic and world knowledge information in parallel during sentence comprehension. Comprehension difficulties in L2 cannot be simply attributed to a delay in world knowledge integration but rather the absence of fast literal semantic access, such as that observed in L1 comprehenders.

Acknowledgements

This research was funded by grants from the Spanish Government (PSI2011-23033 and Consolider Ingenio 2010 CSD2007-00048), from the Catalan Government (SGR 2009-1521) and from the European Research Council under the European Community's Seventh Framework (FP7/2007-2013 Cooperation grant agreement 613465-AThEME). CM was supported by the Basque Foundation for Science (IKERBASQUE) and the BCBL institution.

References

- Ardal, S., Donald, M.W., Meuter, R., Muldrew, S., & Luce, M. (1990) Brain responses to semantic incongruity in bilinguals. *Brain and Language*, 39(2), 187–205.
- Federmeier, K.D., and Kutas, M. (1999a) Right words and left words: electrophysiological evidence for hemispheric differences in meaning processing. *Brain Res. Cogn. Brain Res.* 8, 373–392.
- Federmeier, K.D., and Kutas, M. (1999b) A rose by any other name: long-term memory structure and sentence processing. *J. Mem. Lang.* 41, 469–495.
- Federmeier, K.D., Mc Lennan, D.B., De Ochoa, E., and Kutas, M. (2002) The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: an ERP study. *Psychophysiology* 39, 133–146.

- Federmeier, K.D., Wlotko, E.W., Meyer, A.M. (2008) What's "right" in language comprehension: Event-related potentials reveal right-hemisphere language capabilities. *Lang. Linguist. Compass* 2, 1–17.
- FitzPatrick, I., & Indefrey, P. (2010). Lexical competition in non native speech comprehension. *Journal of Cognitive Neuroscience*, 22, 1165–1178.
- Forster, K. I. (1979). Levels of processing and the structure of the language processor. *Sentence Processing: Psycholinguistic essays presented to Merrill Garrett*. W. E. Cooper and E. Walker. Erlbaum, Hillsdale, N.J.: 27-85.
- Gratton, G., and Coles, M.G.H. (1989) Generalization and evaluation of eye-movement correction procedures. *J. Psychophysiol.* 3, 14–16.
- 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., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potentials. *Bilingualism: Language and Cognition*, 4, 123–142.
- Hahne, A. (2001). What's different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, 30, 251–266.
- Jasper, H.H. (1958) The ten-twenty electrode system of the International Federation. *Electroencephalogr. Clin. Neurophysiol.*, 1958, 10: 370-375.
- Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, 257–282.
- Kelly, S.D., Kravitz, C., and Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain Lang.* 89, 253–260.
- Kutas, M., and Hillyard, S.A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207, 203–205.
- Kutas, M., & Kluender, R. (1994). What is who violating? Are consideration of linguistic violations in light of event-related brain potentials. In H.J. Heinze, T.F. Munte, & G.R.

- Mangun (Eds.), *Cognitive electrophysiology: Basic and clinical applications* (pp. 183–210). Boston: Birkhauser.
- Kutas, M., and Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn. Sci.* 4, 463–470.
- Kutas, M., and Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–47.
- Landi, N., and Perfetti, C.A. (2007). An electrophysiological investigation of semantic and phonological processing in skilled and less-skilled comprehenders. *Brain Lang.* 102, 30–45.
- Marslen-Wilson, W.D., Brown, C., Tyler, L.K. (1988). Lexical representations and language comprehension. *Lang Cognitive Proc* 3: 1-16
- Martin, C. D., Thierry, G., Kuipers, J-R, Boutonnet, B., Foucart, A, Costa, A. (2013). Bilinguals reading in their second language do not anticipate upcoming words as native readers do. *Journal of Memory & 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. doi: 10.3389/fnhum.2014.00040.
- Martín-Loeches, M., Hinojosa, J.A., Casado, P., Muñoz, F., and Fernández- Frías, C. (2004). Electrophysiological evidence of an early effect of sentence context in reading. *Biol. Psychol.* 65, 265–280.
- 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., and Laine, M. (2008). Event-related potentials (ERPs) in the study of bilingual language processing. *J. Neurolinguistics* 21, 477–508.

- Penolazzi, B., Hauk, O., and Pulvermüller, F. (2007). Early semantic context integration and lexical access as revealed by event-related brain potentials. *Biol. Psychol.* 74, 374–388.
- Pinheiro, A.P., Galdo-Alvarez, S., Sampaio, A., Niznikiewicz, M., and Gonçalves, O.F. (2010). Electrophysiological correlates of semantic processing in Williams syndrome. *Res. Dev. Disabil.* 31, 1412–1425.
- Proverbio, A.M., Cok, B., & Zani, A. (2002). Electrophysiological measures of language processing in bilinguals. *Journal of Cognitive Neuroscience*, 14(7), 994–1017.
- Rayner K, Warren T, Juhasz BJ, Liversedge SP. (2004). The effect of plausibility on eye-movements in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30:1290-1301.
- Regel, S., Gunter, T.C., and Friederici, A.D. (2010). Isn't it ironic? An electrophysiological exploration of figurative language processing. *J. Cogn. Neurosci.* 23, 277–293.
- Sanders, L. D., & Neville, H. J. (2003). An ERP study of continuous speech processing II: Segmentation, semantics and syntax in non-native speakers. *Cognitive Brain Research*, 15, 214–227.
- Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and abilingual speech mode – The effect on lexical access. *Memory & Cognition*, 12, 380–386.
- van Berkum, J.A., van den Brink, D., Tesink, C.M., Kos, M., and Hagoort, P. (2008). The neural integration of speaker and message. *J. Cogn. Neurosci.* 20, 580–591.
- van den Brink, D., van Berkum, J.A., Bastiaansen, M.C.M., Tesink, C.M.J.Y., Kos, M., Buitelaar, J.K., et al. (2012). Empathy matters: ERP evidence for inter-individual differences in social language processing. *Soc. Cogn. Affect. Neurosci.* 7, 173–183.
- Warren T, McConnell K. (2007). Investigating effects of selectional restriction violations and plausibility violation severity on eye-movements in reading. *Psychonomic Bulletin & Review*, 14:770-775.

- Warren T, McConnell K, Rayner K. (2008). Effects of context on eye movements when reading about plausible and impossible events. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34:1001-1010.
- Weber-Fox, C., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8, 231–256.
- Weber-Fox, C., Davis, L. J., & Cuadrado, E. (2003). Event-related brain potential markers of high-language proficiency in adults. *Brain and Language*, 85, 231–244.
- Wlotko, E.W., and Federmeier, K.D. (2007). Finding the right word: hemispheric asymmetries in the use of sentence context information. *Neuropsychologia* 45, 3001–3014.