Brain and Language xxx (2016) xxx-xxx

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com

# Discourse comprehension in L2: Making sense of what is not explicitly said

Alice Foucart[a,*], Carlos Romero-Rivas[a], Bernharda Lottie Gort[b], Albert Costa[a,c]

[a] Center for Brain and Cognition, University Pompeu Fabra, Roc Boronat, 138, 08018 Barcelona, Spain
[b] University of Groningen, Oude Kijk in 't Jatstraat 26, 9712 EK Groningen, The Netherlands
[c] ICREA, Institució Catalana de Recerca I Estudis Avançats, Barcelona, Spain

ARTICLE INFO

*Article history:*
Received 3 May 2016
Received in revised form 25 August 2016
Accepted 6 September 2016
Available online xxx

*Keywords:*
Bilingualism
Discourse comprehension
Causal inferences
ERPs

ABSTRACT

Using ERPs, we tested whether L2 speakers can integrate multiple sources of information (e.g., semantic, pragmatic information) during discourse comprehension. We presented native speakers and L2 speakers with three-sentence scenarios in which the final sentence was highly causally related, intermediately related, or causally unrelated to its context; its interpretation therefore required simple or complex inferences. Native speakers revealed a gradual N400-like effect, larger in the causally unrelated condition than in the highly related condition, and falling in-between in the intermediately related condition, replicating previous results. In the crucial intermediately related condition, L2 speakers behaved like native speakers, however, showing extra processing in a later time-window. Overall, the results show that, when reading, L2 speakers are able to process information from the local context and prior information (e.g., world knowledge) to build global coherence, suggesting that they process different sources of information to make inferences online during discourse comprehension, like native speakers.



## 1. Introduction

'The clown took his shoes off and pretended to eat them. The kids laughed a lot when he spread the warm bread with socks'. In this example, the word 'socks' would definitely be easier to integrate preceded by this discourse context than in the classic example from Kutas and Hillyard (1980) where no information was given about the situation ('He spread the warm bread with socks'). This is because semantic integration of a word in the local discourse context is influenced by the global discourse context (Hald, Steenbeek-Planting, & Hagoort, 2007; van Berkum, Hagoort, & Brown, 1999). The interpretation of an utterance therefore necessitates the simultaneous processing of various types of information such as semantics, syntax, world knowledge and the involvement of both working memory (WM) and long-term memory. Recent studies in second language (L2) have suggested that late L2 speakers sometimes have difficulties integrating multiple sources of information during online comprehension (Foucart, Moreno, Martin, & Costa, 2015; Foucart, Garcia, et al., 2015; Roberts, Gullberg, & Indefrey, 2008; Romero-Rivas et al., 2016; but see, Martin, Garcia, Breton, Thierry, & Costa, 2015). Most of these studies investigated processing at sentence level; in this paper we test whether these difficulties impair semantic processing during discourse comprehension.

Text comprehension has been argued to occur at different levels of representation; the *surface code* (exact wording and syntax), the *textbase* (meaning of words within a proposition) and the *situation model* (representation of what the text is about, like events, actions or persons) (Graesser, Singer, & Trabasso, 1994; Kintsch, 1988; Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). The situation model is an interaction between the explicit text (surface code and textbase) and prior information (discourse-pragmatic information). Hence, when reading a text, comprehenders have to retain the information from the explicit text in the WM to make local coherence, while reactivating information from the long-term memory to build global coherence with the incoming information. For example, at textbase level, when presented with a pronoun in a sentence comprehenders have to assign a reference to it, like in the sentence 'Joey lost the game, he got angry', one has to infer that the pronoun "he" refers to Joey. When building the situation model, pragmatic inferences are made using prior world knowledge to interpret the implicit meaning of the explicit text (i.e., the information not directly stated in the text). For instance, when reading the text 'Joey's brother became furiously angry with him. The next day his body was covered in bruises', one has to infer causality, i.e., that Joey's brother probably hit him (Keenan, Baillet, & Brown, 1984; Myers, Shinjo, & Duffy, 1987).

In an eye-tracking experiment, Roberts et al. (2008) showed that L2 speakers have more difficulties integrating multiple sources of information than native speakers during online sentence comprehension. They investigated how German and Turkish L2 learners of Dutch process co-reference between subject and pronoun. Establishing co-reference involves associating a pronoun with a referent in the earlier discourse; there can be one referent, like in 'The workers are in the office. While Peter is working, he is eating a sandwich.', or more, like in 'Peter and Hans are in the office. While Peter is working, he is eating a sandwich.'. Co-reference must be inferred from various sources of information (syntactic and discourse-pragmatic information). While the Turkish subject-pronoun system varies in terms of referent preference, the German system is comparable to that of

* Corresponding author at: Universitat Pompeu Fabra, Department of Technology (room 55116), Roc Boronat, 138, 08018 Barcelona, Spain.
Email address: alfoucart@gmail.com (A. Foucart)

http://dx.doi.org/10.1016/j.bandl.2016.09.001
0093-934/

Dutch. In a prior paper-and-pencil task, German speakers showed the same referent preference as Dutch native speakers, while Turkish speakers followed the preference of their L1. On the other hand, the eye-tracking results revealed that when processing the sentence online, German speakers patterned together with Turkish speakers, and differently from native speakers. The authors concluded that when having to establish co-reference *online*, L2 speakers have more difficulties integrating syntactic and discourse-pragmatic information than native speakers.

This conclusion converges with recent ERP studies suggesting that integrating semantic and pragmatic information is costlier in L2 than in first language (L1). These studies investigated whether L2 speakers take pragmatic information into account when incrementally processing a sentence and whether this affects their online interpretation of the message. More precisely, they looked at whether semantic processing is influenced by world knowledge violations (e.g., 'Mozart composed classic/jazz music'; Martin et al., 2015; Romero-Rivas et al., 2016), the speaker's identity (e.g., 'Every night I drink a glass of wine', said by an adult or by a child; Foucart, Garcia, et al., 2015) or by one's own moral values ('Nowadays, paedophilia should be prohibited/tolerated across the world'; Foucart, Moreno, et al., 2015). In response to world knowledge violations, a similar N400 component was observed in native and L2 speakers (reflecting lexico-semantic integration difficulty; Kutas & Hillyard, 1980). However, the N400 was sometimes longer lasting in the L2 group (Romero-Rivas et al., 2016), suggesting that integrating world knowledge online requires extra processing in L2. Inconsistencies in indexical properties generated a positivity over centro-parietal scalp sites for native speakers and L2 speakers, although earlier for L2 speakers. Regarding moral values, both groups revealed a larger late positivity for immoral statements compared to moral statements, but only native speakers showed an N400 modulation by morality value. Overall, these studies show that L2 speakers process semantic and pragmatic information incrementally during online comprehension but, depending on the source of pragmatic information, online integration occurs at a different time-course in L1 and L2 or requires extra processing.

These studies focused on processing at sentence level; here, we extend this investigation to discourse level by looking at whether L2 speakers can integrate information from the explicit text (surface code and textbase) and prior discourse-pragmatic information (i.e., information previously stated in the text and one's own stored semantic and world knowledge) online to make sense of the message. To do so, we presented L2 speakers with three-sentence scenarios in which the interpretation of the last sentence required establishing simple or complex causal inferences.

Looking at the L2 literature, it seems that L2 speakers have difficulties making causal inferences. For example, in a recent study, Morishima (2013) showed that L2 speakers failed to detect inconsistency in global context and could only detect it when sentences were adjacent. In a probe-verification task, they showed greater difficulty re-activating prior text information. The author interpreted the results as evidence for limited resource allocation for discourse-level processes in L2 comprehension. This conclusion converges with previous claims that since processing the surface code and textbase is not as automatic as in L1, L2 speakers devote more resources than native speakers at processing lower level information, and therefore have fewer resources left to build the global context (Raney, Obeidallah, & Miura, 2002, for a review). In that sense, difficulties at discourse level in L2 has often been associated with reduced WM (Juffs & Harrington, 2011; Service, Maury, & Luotoniemi, 2007; Van Den Noort, Bosch, & Hugdahl, 2006).

However, once L2 proficiency increases and processing at lower levels becomes less demanding, more resources should be available for L2 comprehenders to process higher levels of information. This claim is consistent with studies showing that language fluency affects situation model development (Horiba, 2000; Tang, 1997; Zwaan & Brown, 1996) and that capacity to built complex inferences (e.g., pragmatic inferences vs. simple bridging inferences) increases with proficiency and higher WM capacities (Horiba, 1996; Rai, Loschky, Harris, Peck, & Cook, 2011).

To better understand the origin of the difficulties L2 speakers have when processing discourse, as revealed by behavioural studies, we use the event-related brain potential (ERP) technique to compare the nature and time-course of the processes native and L2 speakers engage when reading a text whose interpretation requires complex inferences.

### *1.1. The present study*

Using the same paradigm as Kuperberg, Paczynski, and Ditman (2011), we compared the online processes engaged by English native speakers and Spanish native speakers highly proficient in English when reading a text whose interpretation requires complex inferences. In their study, Kuperberg et al. presented native English speakers with three-sentence scenarios in which the final sentence was highly causally related, intermediately related, or causally unrelated to its context (see Table 2 for an example). In the highly causally related condition, the last sentence logically followed the first two, and required a simple inference based on world-knowledge. In the intermediately related condition, the second sentence was modified so that a more complex inference was required to link the second sentence to the third. Finally, in the causally unrelated condition, the third sentence was not a logical continuation of the first two. Information from both the discourse context and world knowledge had to be processed to make sense of the third sentence. In order to observe the timing of establishing casual coherence at situation model level, the authors matched the lexico-semantic relationship between content words across all three conditions (using a Latent Semantic Analysis; Landauer & Dumais, 1997). The results revealed a gradual modulation of the N400 component for critical words in the three conditions. The N400 is an ERP component that reflects the ease with which a word is semantically integrated within its preceding context; the more difficult the integration, the larger the effect (Kutas & Hillyard, 1980). Kuperberg et al. (2011) reported a larger N400 component for critical words in the causally unrelated condition compared to the other two conditions. Its amplitude for critical words in the intermediately related condition fell in-between that of the other two conditions (except at midline scalp sites, where it was reduced to the same level as in the highly related condition). The same gradual pattern was found in the behavioural answers, that is, the scenarios were assessed as easier to connect in the highly related condition than in the intermediately related condition, and easier to connect in the intermediately related condition than in the causally unrelated condition. The authors concluded that early stages of semantic processing of an incoming word are influenced by simple and complex inferences. They also argued that, at situation level, word-by-word discourse comprehension is influenced by causal coherence. These results support models of language comprehension that propose that word-by-word processing is influenced by multiple sources of information such as semantics, syntax and world knowledge (Kintsch, 1988, 1998; MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Van Dijk & Kintsch, 1983).

#### 1.1.1. Hypotheses

If L2 speakers have difficulties integrating information from the explicit text and prior discourse-pragmatic information online, the N400 might be delayed or substituted by a later component revealing that L2 speakers engage other processes than native speakers when establishing causal inferences. On the other hand, if the difficulties to establish causal inferences reported in L2 behavioural studies come from L2 speakers' inability to re-activate prior text information, no ERP difference should be observed between the three conditions, since the interpretation of the last sentence only varies according to the preceding context.

## 2. Method

### 2.1. Participants

Twenty-two English native speakers and 28 Spanish-English L2 speakers took part in the experiment (see Table 1 for participants' details). The data from six participants were excluded because of too many ERP artifacts (N = 2), because they did not complete the task properly (N = 1) or because their language test score was too low (N = 3), resulting in 20 participants in the native group and 24 in the L2 speaker group. They all had normal or corrected-to-normal vision and had not suffered any head trauma. Prior to the experiment, oral and written information about the procedure was given, and written consent was obtained from all participants.

#### 2.1.1. English native speakers

Nine were from the US, 7 from the UK, 4 from Ireland, one from Canada, one from New Zealand. They all lived in Spain at the time of the experiment (mean time lived in Spain: 2.5 years, SD: 2.1, range from 1 month to 7 years). Therefore, they all knew some Spanish (from A1 to C2 on the Common European Framework of Reference for Languages) but none of them had learnt a second language before the age of 11.

#### 2.1.2. Spanish-English L2 speakers

They had learned English at school. Thirteen of them also spoke Catalan. They were asked to assess their proficiency in English (from 1: very poor to 7: excellent) for written/oral comprehension/production. They all had an advanced level in English (at least C1 level of the Common European Framework) as assessed by a language test. Details are reported in Table 1.

**Table 1**
Participants' details. Standard deviations are presented in parentheses and ranges into brackets.

| | Native English speakers (N = 20) | Spanish-English L2 speakers (N = 24) |
|---|---|---|
| Number of females | 9 | 16 |
| Mean age (in years) | 27.4 (4.3) [18–35] | 23.6 (4.9); [19–35] |
| Mean age of English acquisition (in years) | | 7.8 (1.8) [7–14] |
| Months Immersed in an English speaking country | | 8.6 (12.4) [1–48] |
| *Self-rated proficiency in English (1 = least fluent, 7 = most fluent)* | | |
| Written comprehension | | 6.6 |
| Written production | | 5.9 |
| Oral production | | 5.8 |
| Oral comprehension | | 6.4 |
| Language test (in %) | | 77.4 (9.2) [65–96] |

### 2.2. Materials and design

We adapted the original materials used in Kuperberg et al.'s (2011) study; 159 scenarios in English (set of sentence triplets) in which the last sentence was highly causally related, intermediately related or causally unrelated with the first two (see example in Table 2). To verify the causal relatedness of each scenario Kuperberg et al. conducted two pre-tests. The details of the construction of the original scenarios can be found in Kuperberg et al. (2011).

To ensure all the scenarios were understood by non-native speakers, we ran a pre-test on the original materials with 9 Spanish-English L2 speakers (6 females, mean age 29.7 years, SD = 4.5) who did not take part in the ERP experiment. We presented each of them with a list of scenarios (3 lists with 53 scenarios in each condition: highly related, intermediately related, and causally unrelated), and asked them to rate the causal relatedness between the third sentence and the two previous ones from 1 (weak causal relationship) to 7 (strong causal relationship). In addition, to evaluate the difficulty of the scenarios in relation to English, we asked them to indicate their general understanding of the sentences from 1 ('I do not understand') to 7 ('I perfectly understand'). We analysed the results following similar criteria as Kuperberg et al. (2011). Scenarios were rejected if (1) the average rating of the highly related scenario was lower than that of either the intermediately related or the causally unrelated scenario; (2) the average rating of the intermediately related scenario was the same or lower than that of the causally unrelated scenario; (3) the highly related scenario was rated below 4; (4) the intermediately related scenario was rated below 3; and (5) the average understanding was below 6. A total of 21 scenarios were rejected, resulting in 138 experimental scenarios. To further ensure good understanding of the scenarios we slightly adapted the vocabulary (e.g., we replaced "quarters" with "coins", "elated" with "euphoric"), keeping modifications to the minimum and intending to apply them to the three conditions to respect semantic similarity values of content words across conditions. We then created 3 lists containing 46 scenarios in each condition (highly related, intermediately related and causally unrelated). The final sentence was seen only once per list, but in every condition across lists. Half of the critical words appeared mid-sentence; the other half appeared sentence-final. Critical words in mid-sentence position were never followed by a coma, but sentence-final critical words were followed by a period. This was done to check whether inferences were more likely to be made at the end of a sentence (when a cue is provided).

### 2.3. Procedure

Participants were sitting in an armchair in front of a computer screen in a dimly lit sound proof room, separate from the experimenter. After receiving visual and oral instructions, they completed a

**Table 2**
Example of an experimental scenario.

| Conditions | Example of scenario |
|---|---|
| highly related | Jill had very fair skin<br>She forgot to put sunscreen on<br>She had <u>sunburn</u> on Monday |
| Intermediately related | Jill had very fair skin<br>She usually remembered to wear sunscreen<br>She had <u>sunburn</u> on Monday |
| Causally unrelated | Jill's skin always tanned well<br>She always put on sunscreen<br>She had <u>sunburn</u> on Monday |

short practice of 6 trials. Each trial was a scenario composed of 3 sentences. Participants' task was to rate how easy or difficult it was to connect the third sentence to the first two sentences in each scenario by pressing a button (1 = difficult to connect, 2 = in-between, 3 = easy to connect). Sentences were presented visually using the software E-prime 2.0 (Schneider, Eschman, & Zuccolotto, 2012a, 2012b). Participants were presented with one of the three experimental lists, randomised for each participant. Each scenario was preceded by a white fixation cross displayed on the screen for 500 ms. The first two sentences were presented in their entirety until the participant pressed the space bar. The third sentence was presented word by word with each word displayed for 500 ms with a 100 ms inter-stimulus interval. After each scenario, the text "1, 2 or 3?" was displayed on the screen until participants pressed the button corresponding to their choice. Their answer triggered the following trial. In addition, to make sure participants were paying attention, a comprehension question about the scenario they had just read was asked on 11 random occasions during the experiment (e.g., 'Did Jill have sunburn on Tuesday? asked after the scenario presented in Table 2; 91% correct answers for the native group and 89.2% for the L2 speaker group, results were not further analysed). Participants were instructed to minimize blinking, eye movements and to stay still during the experiment. The experiment per se lasted for about 30 min.

### 2.4. EEG recording

Electrophysiological data were recorded from 64 active electrodes attached to an elastic cap (International 10–20 System, electrode impedance set below 10 kΩ) and referenced to the left mastoid (re-referenced offline to the average of the two mastoid electrodes). EEG activity was amplified with a band-pass between 0.1 Hz and 125 Hz (BrainAmps DC amplifier, Brain Products GmbH, Munich, Germany), continuously digitised at a sampling rate of 500 Hz, and re-filtered offline at 40 Hz low pass. Epochs ranged from −100 ms to 1000 ms after the onset of the critical word. Artifacts were automatically rejected using the procedure implemented in Brain Analyzer 2.0 (differences in values 200 µv in 200 ms intervals, and amplitudes of ±100 µv), and resulted in the rejection of trials in the highly related condition (native: 4.9%; L2 speakers: 5%), in the intermediately related condition (native: 3.3%; L2 speakers: 4.8%) and in the causally unrelated condition (native: 4.7%; L2 speakers: 4.7%). Baseline correction was performed in reference to pre-stimulus activity (−100; 0 ms).

### 2.5. ERP data analyses

#### 2.5.1. PLS analysis

Although based on Kuperberg et al.'s (2011) study we were originally particularly interested in the N400 component, the variability visually observed in the grand averages of the two groups (Fig. 2) suggested that relevant differences across conditions and groups also occurred outside the typical time-window of the N400 (300–500 ms). Therefore, we performed Partial Least Squares (PLS; Lobaugh, West, & McIntosh, 2001; McIntosh, Bookstein, Haxby, & Grady, 1996) analysis to statistically define the time-windows and locations of the effects. PLS, which is similar to principal components analysis as it uses singular value decomposition, allows identifying differences in ERP amplitudes between experimental conditions across time and scalp location. Information is extracted from the full time course (from critical word onset) and all electrodes sites, leading to a set of latent variables (LVs) representing differences across experimental conditions. The analysis was run on all electrode sites (except FP1/FP2 and TP9/TP10) for 750 ms (i.e., from 0 ms to 750 ms from critical word onset[1]) using the ERP module of the PLSGUI (http://www.rotman-baycrest.on.ca) implemented in Matlab (Mathworks Inc). Significance of the LVs was statistically assessed calculating permutation tests (200 permutations); an LV was considered significant at $p < 0.05$. Estimation of standard errors for the electrode saliences was done using bootstrap re-samplings (1000 replications). The analysis was done for the 3 conditions (highly related, intermediately related and causally unrelated) including both groups (native and L2 speakers). One LV was significant ($p < 0.03$), representing 31.79% of cross-block covariance, and reflecting differences across conditions and across groups (see Fig. 1). Given the different patterns for both groups in the design contrasts, we looked at the ERP saliences for each group individually, contrasting each level of causal relatedness to each other (causally unrelated vs. highly related, causally unrelated vs. intermediately related, intermediately related vs. highly related). This allowed us to identify the timing and location of the effects for each group. The specific time-windows defined for each contrast had a time range comprised within 3 main time-windows corresponding to early effects (100–200 ms), the N400 (250–500 ms) and later effects (600–750 ms).

#### 2.5.2. ANOVAs

Subsequently, we conducted classic analyses of variances (ANOVAs) on each time-window with Group (native vs. L2 speaker) as a between-subject factor and Causality (2 levels), Position (critical word appearing at mid-sentence or sentence-final) and Hemisphere (left/right hemisphere) as within-subject factors. Analyses were run for each region previously defined by the PLS analysis (Frontal: AF3/AF4, AF7/AF8, F3/F4, F7/F8, FC3/FC4, FC5/FC6, FT7/FT8, FT9/FT10, Central: T7/T8, C1/C2, C3/C4, C5/C6, CP1/CP2, CP3/CP4, CP5/CP6, TP7/TP8, Parietal: P1/P2, P3/P4, P5/P6, P7/P8, PO3/PO4, O1/O2). The same analyses were run at Midline (Fz, Cz, Pz) sites without the factor Hemisphere. Interactions were further analysed in post hoc (Bonferroni) analyses. The Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied to all repeated measures with greater than one degree of freedom; in this case, the corrected $p$-value is reported.

### 2.6. ERP results

#### 2.6.1. Causally unrelated vs. highly causally related conditions

Time-window 250–400 ms. There was a significant interaction of Causality × Group at all but Parietal sites (Frontal: $F(1, 42) = 6.11$, $p < 0.01$, $\eta^2 = 0.12$, Midline: $F(1, 42) = 5.01$, $p < 0.01$, $\eta^2 = 0.11$, Central: $F(1, 42) = 7.21$, $p < 0.01$, $\eta^2 = 0.15$, Parietal: $F(1, 42) = 0.38$, $p < 1$, $\eta^2 = 0.01$). Post-hoc analyses revealed that the effect of Causality was significant only for L2 speakers at Frontal sites ($p = 0.03$), with the causally unrelated condition being more positive than the highly causally related condition.

Time-window 350–420 ms. The main factor of Causality tended towards significance at Central sites only ($F(1, 42) = 3.51$, $p = 0.07$, $\eta^2 = 0.08$). A significant interaction of Causality × Group was found at Frontal ($F(1, 42) = 6.89$, $p < 0.01$, $\eta^2 = 0.14$), Midline ($F(1, 42) = 4.87$, $p < 0.03$, $\eta^2 = 0.10$) and Central ($F(1, 42) = 7.59$, $p < 0.001$, $\eta^2 = 0.15$) sites but not at Parietal sites ($F(1, 42) = 0.34$, $p < 1$, $\eta^2 = 0.01$). Post-hoc analyses revealed a significant effect of

[1] Although epochs ranged from −100 ms to 1000 ms after the onset of the critical word, we conducted the PLS analyses only up to 750 ms as later effects might reflect processing of the following word in the sentence.

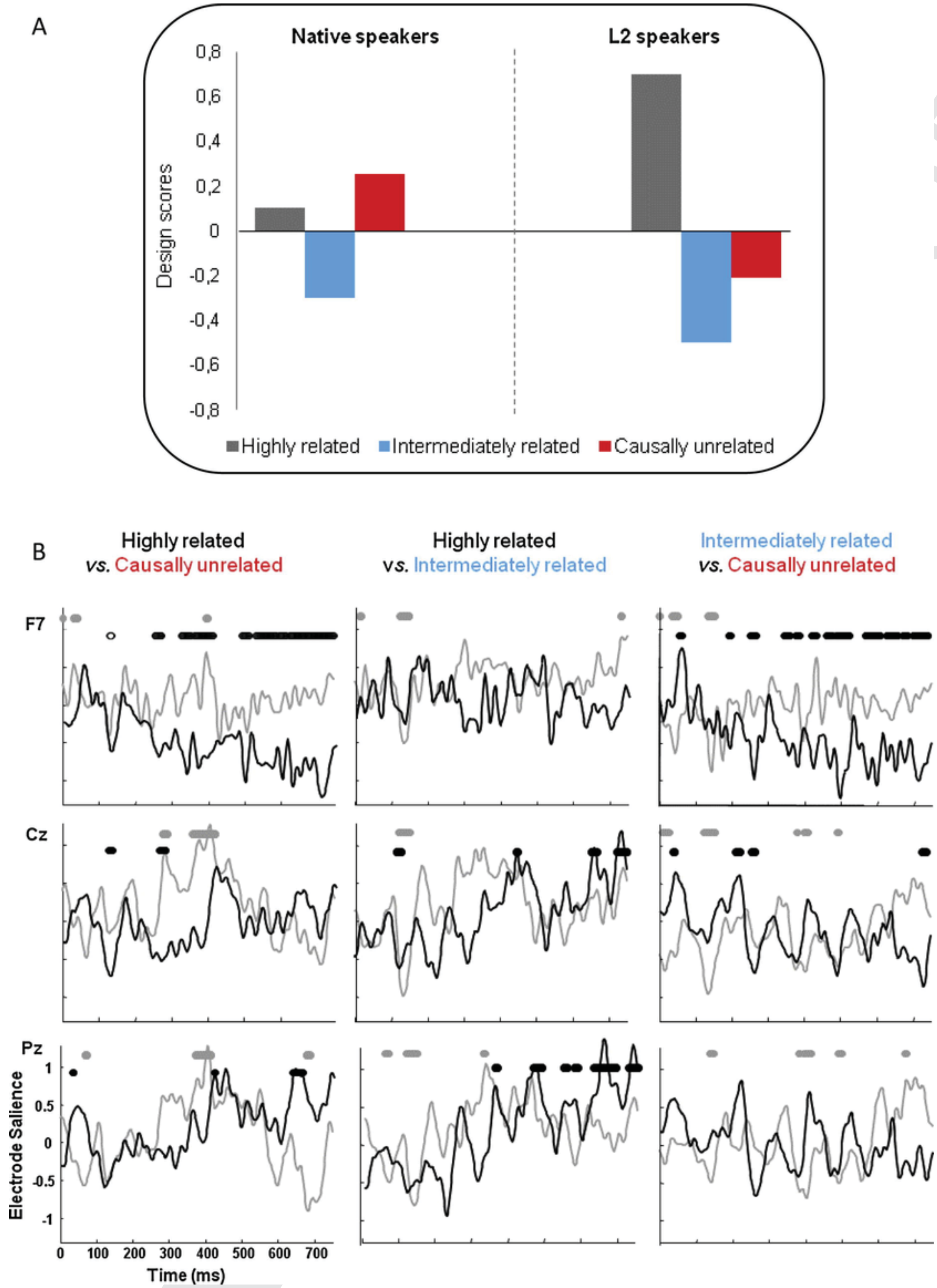


**Fig. 1.** Panel A: Design scores demonstrating the contrast between highly causally related (black), intermediately related (blue), and causally unrelated (red) scenarios for the native speakers and the L2 speakers. Panel B: ERP saliences contrasting each condition against another over F7, Cz and Pz for both groups (native speakers: grey lines; L2 speakers: black lines). Markers on top of each channel represent stable differences between conditions (native speakers: grey dots; L2 speakers: black dots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Causality only for native speakers at Central sites ($p = 0.02$), the causally unrelated condition being more negative than the highly causally related condition.

Time-window 600–720 ms. A significant interaction Causality × Hemisphere × Group was observed at Frontal sites only ($F(1, 42) = 4.87$, $p < 0.05$, $\eta^2 = 0.10$), revealing a larger positivity in

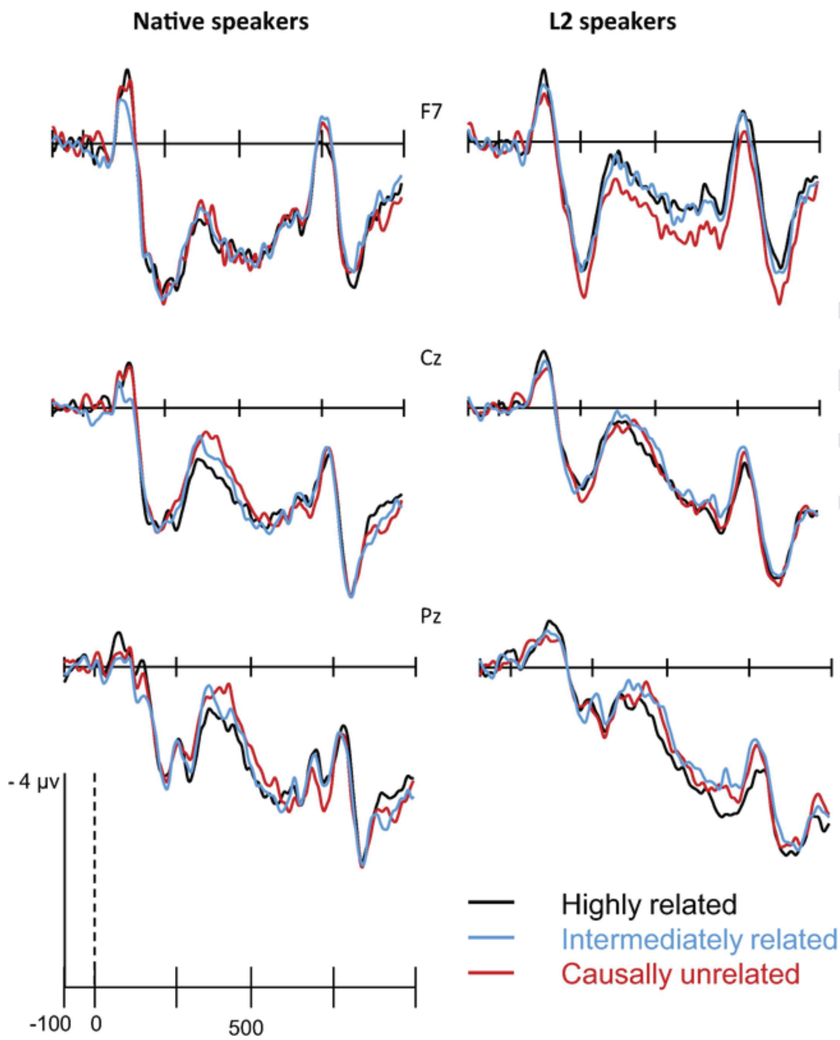


**Fig. 2.** Event-related potential grand average at F7, Cz and Pz for the native speakers and L2 speakers. ERP results to critical words in highly causally related (black), intermediately related (blue), and causally unrelated (red) scenarios. Negativity is plotted up. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the left hemisphere for the causally unrelated condition than the highly causally related condition for L2 speakers only ($p < 0.001$).

In sum, in the native speaker group the causally unrelated condition was significantly more negative than the highly causally related condition in the N400 time-window. In the L2 speaker group, a reverse effect was observed at early and late time-windows, the causally unrelated condition being more positive than the other condition.

*2.6.2. Causally unrelated vs. intermediately related conditions*

The factor Causality did not reach significance nor did it interact with other factors.

*2.6.3. Intermediately related vs. highly related conditions*

Time-window 110–170 ms. The factor Causality was significant, indicating that the intermediately related condition was more positive than the highly related condition (Frontal: $F(1, 42) = 4.46$, $p < 0.05$, $\eta^2 = 0.10$, Midline: $F(1, 42) = 6.87$, $p < 0.01$, $\eta^2 = 0.14$, Central: $F(1, 42) = 9.45$, $p < 0.001$, $\eta^2 = 0.18$, Parietal: $F(1, 42) = 8.09$, $p < 0.001$, $\eta^2 = 0.16$). The factor Group also reached significance, showing that both conditions were overall more negative for L2 speakers than for native speakers (Frontal: $F(1, 42) = 5.59$, $p < 0.01$, $\eta^2 = 0.11$, Midline: $F(1, 42) = 7.58$, $p < 0.001$, $\eta^2 = 0.15$, Central: $F(1, 42) = 6.58$, $p < 0.01$, $\eta^2 = 0.13$, Parietal: $F(1, 42) = 2.92$, $p = 0.09$, $\eta^2 = 0.06$). The interaction Causality × Position was significant at Midline ($F(1, 42) = 6.44$, $p < 0.01$, $\eta^2 = 0.13$) and at Frontal sites ($F(1, 42) = 4.15$, $p < 0.05$, $\eta^2 = 0.09$). Post-hoc analyses showed that the difference between the two conditions was significant only at mid-sentence position (Frontal: $p = 0.07$; Midline: $p = 0.01$). The interaction Causality × Position × Group (Midline: $F(1, 42) = 6.07$, $p < 0.01$, $\eta^2 = 0.13$, Central: $F(1, 42) = 8.02$, $p < 0.001$, $\eta^2 = 0.16$, Parietal: $F(1, 42) = 3.47$, $p = 0.07$, $\eta^2 = 0.07$) indicated that the effect was significant only for native speakers at mid-sentence position (Midline: $p = 0.01$, Central: $p = 0.02$).

Time-window 400–500 ms. The factor Group was significant at Frontal sites ($F(1, 42) = 5.74$, $p < 0.01$, $\eta^2 = 0.12$) and tended towards significance at Midline ($F(1, 42) = 3.21$, $p = 0.08$, $\eta^2 = 0.07$), indicating that the two conditions were overall more negative for L2 speakers than for native speakers. The factor Causality tended towards significance (Midline: $F(1, 42) = 3.59$, $p = 0.06$, $\eta^2 = 0.08$, Central: $F(1, 42) = 3.29$, $p = 0.08$, $\eta^2 = 0.07$; Parietal:

$F(1, 42) = 2.99$, $p = 0.09$, $\eta^2 = 0.07$), suggesting that the intermediately related condition was slightly more negative than the highly causally related condition. No significant interaction involving the factors Causality and Group was observed.

Time-window 620–750 ms. A significant interaction Causality × Group was found at Parietal sites only ($F(1, F(1, 42) = 5.44$, $p < 0.01$, $\eta^2 = 0.11$), revealing a significantly larger negativity for the intermediately related condition than the highly related condition for L2 speakers only ($p = 0.03$).

In sum, in the native group, the intermediately related condition was more positive than the highly related condition in an early time-window (110–170 ms) only when the critical word appeared at mid-sentence position. In the bilingual group, the contrast between the two conditions revealed a significant negativity in a later time-window (620–750 ms). In the N400 time-window (400–500 ms), although the intermediately related condition had slightly larger negative amplitude than the highly related condition, this difference was only marginally significant.

## . 2.7Behavioural results

ANOVAs with Group (native vs. L2 speakers) as a between-subject factor and Causality (3 levels) and Position (critical word appearing at mid-sentence or sentence-final) as within-subject factors were conducted on the means of participants' assessment of the scenarios (1 = difficult to connect; 2 = in between; 3 = easy to connect). Means for each level of causality are reported in Table 3. ANOVAs showed a significant main effect of Causality ($F(2, 82) = 350.45$, $p < 0.001$, $\eta^2 = 0.89$) and an interaction Group × Causality ($F(2, 82) = 9.07$, $p < 0.001$, $\eta^2 = 0.18$). Post-hoc analyses revealed a significant gradual level of Causality, scenarios being assessed as easier to connect in the highly related condition than in the intermediately related condition ($p < 0.001$), and easier to connect in the intermediately related condition than in the causally unrelated condition ($p < 0.001$). This pattern was true for both groups with, however, a slightly smaller difference between the highly related condition and the intermediately related condition for native speakers (0.3) than for L2 speakers (0.5). The interaction Causality × Position did not reach significance ($F(2, 82) = 1.39$, $p = 0.25$, $\eta^2 = 0.03$).

**Table 3**
Mean (Standard Deviation) causal relatedness judgement score (1 = difficult to connect, 2 = in-between, 3 = easy to connect) for native speakers and L2 speakers.

| Causal relatedness | Mean Relatedness Judgment Score: All Scenarios | Mean Relatedness Judgment Score: Scenarios with Mid-sentence Critical Words | Mean Relatedness Judgment Score: Scenarios with Sentence-final Critical Words |
|---|---|---|---|
| *Native speakers* | | | |
| Highly related | 2.8 (0.2) | 2.8 (0.2) | 2.9 (0.2) |
| Intermediately related | 2.5 (0.2) | 2.5 (0.2) | 2.4 (0.3) |
| Causally unrelated | 1.6 (0.3) | 1.6 (0.3) | 1.5 (0.3) |
| *L2 speakers* | | | |
| Highly related | 2.8 (0.1) | 2.8 (0.1) | 2.8 (0.2) |
| Intermediately related | 2.3 (0.2) | 2.3 (0.3) | 2.3 (0.3) |
| Causally unrelated | 1.8 (0.3) | 1.7 (0.3) | 1.8 (0.4) |

## 3Discussion

The present study compared a group of native speakers and a group of L2 speakers to investigate whether the latter can integrate information from the explicit text (surface code and textbase) and prior discourse-pragmatic information online to interpret a message during discourse comprehension. This question originated from studies that have shown that L2 speakers have difficulties integrating multiple sources of information at sentence or local discourse level (Foucart, Moreno, et al., 2015; Foucart, Garcia, et al., 2015; Roberts et al., 2008; Romero-Rivas et al., 2016; but see, Martin et al., 2015). Here, we tested whether these difficulties impair semantic processing during discourse comprehension. We compared the online processes engaged by native speakers and L2 speakers when reading a text whose interpretation requires establishing complex causal inferences. The behavioural responses indicated that both native speakers and L2 speakers establish causal coherence across sentences; however, ERP results revealed some differences in the online processing.

In the native group, a larger negativity was found for critical words in the causally unrelated condition than for those in the highly causally related condition in the N400 time-window. In this same time-window the amplitude of the effect for critical words in the intermediately related condition fell in-between that provoked by the other two conditions. This pattern was independent of the position of the critical word (mid-sentence or sentence final). No later effect was observed.[2] Crucially, the results of interest for our purpose, i.e., the ERP modulation in the N400 time-window, replicated those reported in the original study of Kuperberg et al. (2011), and thus confirmed once again that early semantic processing of incoming words is influenced by causal coherence across sentences at situation level. These results converge with previous studies showing that readers establish causal inferences online during text comprehension (Burkhardt, 2006; George, Mannes, & Hoffman, 1997; Keenan et al., 1984; Myers & Duffy, 1990; Myers et al., 1987; Yang, Perfetti, & Schmalhofer, 2007), and that they are immediately sensitive to a break in global text coherence (Hald et al., 2007; van Berkum et al., 1999).

Of central interest here, the ERP data in the L2 speaker group did not show exactly the same pattern. When contrasting the causally unrelated condition and the highly related conditions, differences were observed in an early (250–400 ms) and a late (600–720 ms) time-window at frontal sites, the former condition being more positive than the latter. Although significance was reached in these two specific time-windows, the inspection of the grand averages suggests that this positive ERP deflection might be a single effect starting from around 250 and lasting until around 700 ms. This positive modulation diverges from the negativity observed in the native group, and its interpretation is not clear. Indeed, positivities, originally taken to reflect syntactic anomalies and ambiguities, have more recently been reported to be modulated also by lexical and discourse factors (Kuperberg, 2007; Nieuwland & Van Berkum, 2005, for a review). However, they usually emerge in a later time-window than here, corresponding more or less to the P600 component (500–800 ms or

[2] An additional difference was found in a very early time-window (110–170 ms), the intermediately related condition being more positive than the highly related condition, only when the critical word appeared at mid-sentence position. We do not have a clear explanation for this observation. Note that very early differences seemed to appear as well in Kuperberg et al.'s (2011) original study (see Fig. 2) but the authors do not specify whether it reached significance or not. However, unlike here where the difference was evoked by critical words in the intermediately related condition, the difference in their study seemed to be provoked by critical words in the causally unrelated condition.

later). Moreover, here, the positivity was observed at frontal sites; in the literature, frontal positivities have been observed in response to expectancy violations (DeLong, Urbach, Groppe, & Kutas, 2011), but the scenarios we used were not constrained to trigger strong expectancy. Thus, we will refrain from making strong claims about the specific processes reflected by the present positivity, and will interpret it as an indication of an active attempt to integrate critical words with the context and information stored within long-term memory.

Of most interest here are the results obtained in the intermediately related condition, when a complex inference was required to link the second sentence to the third. Similarly as in the native group, the amplitude of the N400 effect in the L2 group tended to be slightly larger than in the highly related condition, but the difference was only marginally significant. Note that, in the non-native group, the difference between these two conditions became significant in a later time-window (620–750 ms). The present results are in line with previous studies reporting a delayed and sometimes longer lasting N400 effect in L2 compared to L1 (Ardal, Donald, Meuter, Muldrew, & Luce, 1990; Braunstein et al., 2012; Hahne, 2001; Moreno & Kutas, 2005; Newman, Tremblay, Nichols, Neville, & Ullman, 2012; Ojima, Nakata, & Kakigi, 2005; Weber-Fox & Neville, 1996). These results show that L2 speakers made inferences that facilitated integration of incoming words. Had they not been able to make inferences online, a similar ERP pattern would have been found for the intermediately related condition and the causally unrelated condition. The late negativity suggests that semantic integration might have been, however, slightly more costly than in the highly related condition and required extra processing. Note that a similar argument was put forward to account for a negativity of longer duration for L2 speakers than for native speakers in response to world knowledge violations (Romero-Rivas et al., 2016).

Overall, these results show that L2 speakers make causal inferences online during text comprehension and that they are immediately sensitive to coherence break. These results show once again that L2 speakers process words incrementally as they read (Dussias & Cramer Scaltz, 2008; Frenck-Mestre & Pynte, 1997; Roberts, 2013, for a review). Moreover, they extend the results reported in studies investigating L2 processing at sentence level, in that they show that incremental processing does not occur only in local context (i.e., at sentence level) but also in global context (i.e., at text level). Our results suggest that the mechanisms involved during L2 discourse comprehension are similar as in L1, even though the qualitative differences in the ERP data imply that extra processing is required in L2. A stronger interpretation of these differences is difficult to provide given the current lack of ERP studies investigating L2 discourse comprehension.

The modulation of the ERP signal across the three conditions indicates L2 speakers' sensitivity to the complexity of the inference required to make sense of the message; the more complex the inference, the more effortful the processing. These results suggest that our participants were able to re-activate information previously stated in the text when processing the third sentence, since the interpretation of the last sentence only varied according to the preceding context. Hence, at least when simple inferences need to be made, L2 speakers can integrate information from the incoming word, the context and their own stored semantic and real-world knowledge online. If there is a good match between these three elements, integration of the critical word is facilitated, as reflected by a reduction of the N400 amplitude (Hagoort, Hald, Bastiaansen, & Petersson, 2004; Kuperberg et al., 2011; Kutas, Van Petten, & Kluender, 2006). If the match is less obvious, L2 speakers are able to establish more complex inferences to integrate new information in the situation model. On the other hand, when incoming information does not logically follow and the discourse model needs to be updated, L2 speakers encounter more difficulty. Updating abilities have been shown to depend on WM capacity (Boudewyn, Long, & Swaab, 2013; Li, Zheng, Zhao, & Xia, 2014). Individuals with higher WM capacity are better at processing different sources of information when integrating incoming information than individuals with lower WM capacity who tend to rely on one prevailing source. This is consistent with recent studies showing that L2 speakers may rely more on pragmatic information when facing integration difficulties in sentence comprehension (Foucart, Garcia, et al., 2015; Roberts, 2013; Roberts & Felser, 2011). Actually, a large positive ERP deflection similar to that observed here in the causally unrelated condition (although larger at central sites) was observed in response to a conflict between the message delivered by the speaker and the speaker's identity (Foucart, Garcia, et al., 2015).

Unlike Morishima's (2013) results, our study shows that L2 speakers are able to detect inconsistency in global context and have no difficulty re-activating prior text information. Our findings rather suggest that L2 speakers encounter difficulties mainly when incoming information does not logically follow the situation model, and that these difficulties seem to occur at integration phase and not at activation phase. In that sense, our results imply that L2 speakers behave like native speakers with low WM capacity or poor comprehension abilities (Boudewyn et al., 2013; Li et al., 2014; Long & Chong, 2001). Our study did not specifically manipulate individual measures and therefore the impact of individual factors (e.g., WM) on L2 discourse comprehension processing is not proved by our results. Nevertheless, a parallel with studies that did include individual measures suggest that our findings would support, to some extent, views that claim that differences in L1 and L2 processing originate from factors like proficiency and/or limited cognitive capacity, reducing speed and WM (Dekydtspotter, Schwartz, & Sprouse, 2006; Hopp, 2010), and they diverge from others that claim that processing in L2 is fundamentally different (Clahsen & Felser, 2006a, 2006b). Future research should investigate how individual differences in WM capacity (Juffs & Harrington, 2011) and language skills and experience (MacDonald & Christiansen, 2002) affect discourse processing in L2.

In conclusion, this study showed that highly proficient L2 speakers can process different sources of information to make inferences online. Just like native speakers, when reading a text, they are able to retain the information from the explicit text in their WM to make local coherence, while re-activating information from the long-term memory to build global coherence with the incoming information. Our findings suggest that the mechanisms involved in L1 and L2 discourse comprehension are similar, even though some extra processing seems to be required in L2. This study opens up future directions to better understand discourse processing in L2 speakers relative to native speakers.

## Acknowledgements

This work was supported by grants from the Spanish Government (PSI2011-23033, CONSOLIDER-INGENIO2010 CSD2007-00048, ECO2011-25295, and ECO2010-09555-E), from the Catalan Government (SGR 2009-1521) and from the Grup de Recerca en Neurociència Cognitiva (GRNC) - 2014SGR1210. It has also received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 613465. We would like to thank Dr Gina Kuperberg for sharing her materials and Dr Kalinka Timmer for her help with the analyses.

## 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. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/2224493>.

Boudewyn, M.A., Long, D.L., Swaab, T.Y., 2013. Effects of working memory span on processing of lexical associations and congruence in spoken discourse. Frontiers in Psychology 4, 60. http://dx.doi.org/10.3389/fpsyg.2013.00060.

Braunstein, V., Ischebeck, A., Brunner, C., Grabner, R.H., Stamenov, M., Neuper, C., 2012. Investigating the influence of proficiency on semantic processing in bilinguals: An ERP and ERD/S analysis. Acta Neurobiologiae Experimentalis 72 (4), 421–438. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/23377272>.

Burkhardt, P., 2006. Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. Brain and Language 98 (2), 159–168. http://dx.doi.org/10.1016/j.bandl.2006.04.005.

Clahsen, H., Felser, C., 2006. AUTHORS' RESPONSE Continuity and shallow structures in language processing. Applied Psycholinguistics 27, 107–126.

Clahsen, H., Felser, C., 2006. How native-like is non-native language processing?. Trends in Cognitive Sciences 10 (12), 564–570. http://dx.doi.org/10.1016/j.tics.2006.10.002.

Dekydtspotter, L., Schwartz, B.D., Sprouse, R.A., 2006. The comparative fallacy in L2 processing research. In: Grantham O'Brien, M., Shea, C., Archibald, J. (Eds.), Proceedings of the 8th generative approaches to second language acquisition conference (GASLA 2006). Cascadilla Proceedings Project, Somerville, MA, pp. 33–40.

DeLong, K.A., Urbach, T.P., Groppe, D.M., Kutas, M., 2011. Overlapping dual ERP responses to low cloze probability sentence continuations. Psychophysiology 48 (9), 1203–1207. http://dx.doi.org/10.1111/j.1469-8986.2011.01199.x.

Dussias, P.E., Cramer Scaltz, T.R., 2008. Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. Acta Psychologica 128 (3), 501–513. http://dx.doi.org/10.1016/j.actpsy.2007.09.004.

Foucart, A., Garcia, X., Ayguasanosa, M., Thierry, G., Martin, C.D., Costa, A., 2015. Does the speaker matter? Online processing of semantic and pragmatic information in L2 speech comprehension. Neuropsychologia 75, 291–303. http://dx.doi.org/10.1016/j.neuropsychologia.2015.06.027.

Foucart, A., Moreno, E.M., Martin, C.D., Costa, A., 2015. Integration of moral values during L2 sentence processing. Acta Psychologica 162, 1–12. http://dx.doi.org/10.1016/j.actpsy.2015.09.009.

Frenck-Mestre, C., Pynte, J., 1997. Syntactic ambiguity resolution while reading in second and native languages. Quarterly Journal of Experimental Psychology 50A, 119–148.

George, M.S., Mannes, S., Hoffman, J.E., 1997. Individual differences in inference generation: An ERP analysis. Journal of Cognitive Neuroscience 9 (6), 776–787. http://dx.doi.org/10.1162/jocn.1997.9.6.776.

Graesser, A.C., Singer, M., Trabasso, T., 1994. Constructing inferences during narrative text comprehension. Psychological Review 101 (3), 371–395. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/7938337>.

Greenhouse, S.W., Geisser, S., 1959. On methods in the analysis of profile data. Psychometrika 24, 95–112.

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M., 2004. Integration of word meaning and world knowledge in language comprehension. Science (New York, N.Y.) 304 (5669), 438–441. http://dx.doi.org/10.1126/science.1095455.

Hahne, A., 2001. What's different in second-language processing? Evidence from event-related brain potentials. Journal of Psycholinguistic Research 30 (3), 251–266. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/11523274>.

Hald, L.A., Steenbeek-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. http://dx.doi.org/10.1016/j.brainres.2007.02.054.

Hopp, H., 2010. Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. Lingua 120 (4), 901–931. http://dx.doi.org/10.1016/j.lingua.2009.06.004.

Horiba, Y., 1996. Comprehension processes in L2 reading: Language competence, textual coherence, and inferences. Studies in Second Language Acquisition 18 (04), 433. http://dx.doi.org/10.1017/S0272263100015370.

Horiba, Y., 2000. Reader control in reading: Effects of language competence, text type, and task. Discourse Processes 29 (3), 223–267. http://dx.doi.org/10.1207/S15326950dp2903_3.

Juffs, A., Harrington, M., 2011. Aspects of working memory in L2 learning. Language Teaching 44 (02), 137–166. http://dx.doi.org/10.1017/S0261444810000509.

Keenan, J.M., Baillet, S.D., Brown, P., 1984. The effect of causal cohesion on comprehension and memory. Journal of Verbal Learning and Verbal Behavior 23, 115–126.

Kintsch, W., 1988. The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review 95 (2), 163–182. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/3375398>.

Kintsch, W., 1998. Comprehension: A paradigm for cognition. Cambridge University Press, New York.

Kuperberg, G.R., 2007. Neural mechanisms of language comprehension: Challenges to syntax. Brain Research 1146, 23–49. http://dx.doi.org/10.1016/j.brainres.2006.12.063.

Kuperberg, G.R., Paczynski, M., Ditman, T., 2011. Establishing causal coherence across sentences: An ERP study. Journal of Cognitive Neuroscience 23 (5), 1230–1246. http://dx.doi.org/10.1162/jocn.2010.21452.

Kutas, M., Hillyard, S., 1980. Reading senseless sentences: Brain potentials reflect semantic incongruity. Science 207, 203–205.

Kutas, M., Van Petten, C., Kluender, R., 2006. Psycholinguistics electrified II: 1994–2005. In: Traxler, M., Gernsbache, M.A. (Eds.), Handbook of psycholinguistics, 2nd ed. Elsevier, New York, pp. 659–724.

Landauer, T.K., Dumais, S.T., 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review 104, 211–240.

Li, X.-Q., Zheng, Y.-Y., Zhao, H.-Y., Xia, J.-Y., 2014. How the speed of working memory updating influences the on-line thematic processing of simple sentences in Mandarin Chinese. Cognitive Neurodynamics 8 (6), 447–464. http://dx.doi.org/10.1007/s11571-014-9292-2.

Lobaugh, N.J., West, R., McIntosh, A.R., 2001. Spatiotemporal analysis of experimental differences in event-related potential data with partial least squares. Psychophysiology 38 (3), 517–530. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/11352141>.

Long, D.L., Chong, J.L., 2001. Comprehension skill and global coherence: A paradoxical picture of poor comprehenders' abilities. Journal of Experimental Psychology. Learning, Memory, and Cognition 27 (6), 1424–1429. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/11713877>.

MacDonald, M.C., Christiansen, M.H., 2002. Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). Psychological Review 109 (1), 35–54. discussion 55–74. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/11863041>.

MacDonald, M.C., Pearlmutter, N.J., Seidenberg, M.S., 1994. The lexical nature of syntactic ambiguity resolution [corrected]. Psychological Review 101 (4), 676–703. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/7984711>.

Martin, C.D., Garcia, X., Breton, A., Thierry, G., Costa, A., 2015. World knowledge integration during second language comprehension. Language, Cognition and Neuroscience Retrieved from. <http://www.tandfonline.com/doi/abs/10.1080/23273798.2015.1084012>.

McIntosh, A.R., Bookstein, F.L., Haxby, J.V., Grady, C.L., 1996. Spatial pattern analysis of functional brain images using partial least squares. NeuroImage 3 (3 Pt 1), 143–157. http://dx.doi.org/10.1006/nimg.1996.0016.

Moreno, E.M., Kutas, M., 2005. Processing semantic anomalies in two languages: An electrophysiological exploration in both languages of Spanish-English bilinguals. Brain Research. Cognitive Brain Research 22 (2), 205–220. http://dx.doi.org/10.1016/j.cogbrainres.2004.08.010.

Morishima, Y., 2013. Allocation of limited cognitive resources during text comprehension in a second language. Discourse Processes 50 (8), 577–597. http://dx.doi.org/10.1080/0163853X.2013.846964.

Myers, J.L., Duffy, S.A., 1990. Causal inferences and text memory. In: Graesser, A.C., Bower, G.H. (Eds.), The psychology of learning and motivation: Inferences and text comprehension. Academic Press, San Diego, CA, pp. 159–173.

Myers, J.L., Shinjo, M., Duffy, S.A., 1987. Degree of and, causal relatedness and memory. Journal of Memory and Language 26, 453–465.

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 (5), 1205–1223. http://dx.doi.org/10.1162/jocn_a_00143.

Nieuwland, M.S., Van Berkum, J.J.A., 2005. Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. Brain Research. Cognitive Brain Research 24 (3), 691–701. http://dx.doi.org/10.1016/j.cogbrainres.2005.04.003.

Ojima, S., Nakata, H., Kakigi, R., 2005. An ERP study of second language learning after childhood: Effects of proficiency. Journal of Cognitive Neuroscience 17 (8), 1212–1228. http://dx.doi.org/10.1162/0898929055002436.

Rai, M.K., Loschky, L.C., Harris, R.J., Peck, N.R., Cook, L.G., 2011. Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension. Language Learning 61 (1), 187–218. http://dx.doi.org/10.1111/j.1467-9922.2010.00592.x.

Raney, G.E., Obeidallah, S., Miura, T., 2002. Text comprehension in bilinguals: Integrating perspectives on language representation and text processing. In: Heredia, R., Altarriba, J. (Eds.), Bilingual sentence processing. Elsevier Sciences, Amsterdam, pp. 165–186.

Roberts, L., Felser, C., 2011. Plausibility and recovery from garden paths in second language sentence processing. Applied Psycholinguistics 32 (02), 299–331. http://dx.doi.org/10.1017/S0142716410000421.

Roberts, L., Gullberg, M., Indefrey, P., 2008. Online pronoun resolution in L2 discourse. Studies in Second Language Acquisition 30 (3), 333–357.

Roberts, L., 2013. Sentence processing in bilinguals. In: van Gompel, R. (Ed.), Sentence processing: Current issues in language. Psychology Press, London, pp. 221–246.
Romero-Rivas, C., Corey, J.D., Garcia, X., Thierry, G., Martin, C.D., Costa, A., 2016. World knowledge and novel information integration during L2 speech comprehension. Bilingualism: Language and Cognition 1–12. http://dx.doi.org/10.1017/S1366728915000905.
Schneider, W., Eschman, A., Zuccolotto, A., 2012. E-prime reference guide, Psychology Software Tools Inc., Pittsburgh.
Schneider, W., Eschman, A., Zuccolotto, A., 2012. E-prime user's guide, Psychology Software Tools Inc., Pittsburgh.
Service, E., Maury, S., Luotoniemi, E., 2007. Individual differences in phonological learning and verbal STM span. Memory & Cognition 35 (5), 1122–1135. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/17910194>.
Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.C., 1995. Integration of visual and linguistic information in spoken language comprehension. Science (New York, N.Y.) 268 (5217), 1632–1634. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/7777863>.
Tang, H., 1997. The relationship between reading comprehension processes in L1 and L2. Reading Psychology: An International Quarterly 18, 249–301.
van Berkum, J.J.A., Hagoort, P., Brown, C.M., 1999. Semantic integration in sentences and discourse: Evidence from the N400. Journal of Cognitive Neuroscience 11 (6), 657–671. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/10601747>.
Van Den Noort, M.W., Bosch, P., Hugdahl, K., 2006. Foreign language proficiency and working memory capacity. European Psychologist 11 (4), 289–296.
Van Dijk, T.A., Kintsch, W., 1983. Strategies of discourse comprehension. Academic Press, New York.
Weber-Fox, C.M., Neville, H.J., 1996. Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. Journal of Cognitive Neuroscience 8 (3), 231–256. http://dx.doi.org/10.1162/jocn.1996.8.3.231.
Yang, C.L., Perfetti, C.A., Schmalhofer, F., 2007. Event-related potential indicators of text integration across sentence boundaries. Journal of Experimental Psychology. Learning, Memory, and Cognition 33 (1), 55–89. http://dx.doi.org/10.1037/0278-7393.33.1.55.
Zwaan, R.A., Brown, C.M., 1996. The influence of language proficiency and comprehension skill on situation-model construction. Discourse Processes 21 (3), 289–327. http://dx.doi.org/10.1080/01638539609544960.
Zwaan, R.A., Radvansky, G.A., 1998. Situation models in language comprehension and memory. Psychological Bulletin 123 (2), 162–185. Retrieved from. <http://www.ncbi.nlm.nih.gov/pubmed/9522683>.