The nature of first and second language processing: The role of cognitive control and L2 proficiency during text-level comprehension*

Ana Pérez, Laura Hansen[1] and Teresa Bajo
University of Granada (Spain)

*ACKNOWLEDGEMENTS
This research was supported by a doctoral research grant from the Spanish Ministry of Education, Culture and Sports (FPU AP2010–3434) to Laura Hansen, by a postdoctoral contract funded by Spanish Ministry of Science and Innovation (PSI2012–33625) to Ana Pérez, and by grants of MINECO (PSI2012- 33625; PCIN-2015-132; PSI2015-65502-C2-1-P) and Junta de Andalucía (Excelencia2012-CTS 2369) to Teresa Bajo.

Address for correspondence:
Ana Pérez[2] and Laura Birke Hansen[3]
Mind, Brain and Behavioral Research Centre (CIMCYC)
Department of Experimental Psychology, University of Granada.
C/ Profesor Clavera s/n, CIMCYC, 18011, Granada, Spain.
E-mails: aip27@cam.ac.uk; laurabirke@gmail.com
Phone: +44 (0)7706619219

Abstract

Text comprehension relies on high-level cognitive processes as it is the ability to revise an erroneous inference. Recent models of language processing hold that native language processing is proactive in nature (highly predictive), whereas processing seems to be weaker in the second language. However, if a prediction fails because unexpected information is encountered, reactive processing is needed to revise previous information. Twenty-four highly proficient late bilinguals were presented with narratives in L1-English and L2-Spanish. Each text demanded the revision of an initial predictive inference. Reading times and N400 amplitude suggested inferential revision is less efficient in the L2 compared to the L1. Importantly, these effects were modulated by individual differences in cognitive control and L2 proficiency. More efficient L1 comprehension was related to a balance between proactive and reactive control and lower L2 proficiency, whereas more native-like L2 comprehension was associated with a strong proactive control and higher L2 proficiency.

Keywords: late bilinguals; reading comprehension; inferential revision; cognitive control; L2 proficiency

1. Introduction

Mastering a non-native language can be very challenging, especially if the second language (L2) has been acquired relatively late in life. Adult learners can reach high levels of proficiency in their L2, and with increasing proficiency many aspects of L2 processing can become more and more native-like (see Birdsong & Molis, 2001). However, some studies suggest that L2 processing remains different from native language (L1) processing. Empirical evidence has demonstrated that between-language differences are present in both the syntactic and semantic domains (Dufour & Kroll, 1995; Duyck & De Houwer, 2008; Wartenburger, Heekeren, Abutalebi, Cappa, Villringer, & Perani, 2003; for reviews, see Clahsen & Felser, 2006; Slabakova, 2006), although these differences tend to be quantitative rather than qualitative in the semantic domain (see Slabakova, 2006; Wartenburger et al., 2003).

Studies looking at lexical and semantic processing in L1 and L2 have used procedures involving words or sentences (see e.g., Foucart, Martin, Moreno, & Costa, 2014; Martin et al., 2013), while relatively fewer studies have involved texts and higher order discourse processes. Importantly, different from lexical and sentence processing, text processing requires the construction of a mental representation (i.e., situation model, van Dijk & Kintsch, 1983) by constantly integrating new information, which depends on multiple high-level comprehension processes (derived from the integration of linguistic and pragmatic information) not present at lower-level processing (exclusively based on linguistic properties). For instance, a crucial aspect of text comprehension is the ability to generate information that has not been explicitly described, referred to as inference making (Cain & Oakhill, 1999). To this end, readers must connect several pieces of information presented in the text and/or activate prior knowledge from long-term memory (see McNamara & Magliano, 2009). An important type of inference in text comprehension is prediction. Predictive inferences help to anticipate upcoming concepts in the story (Beeman, Bowden & Gernsbacher, 2000; see also Pérez, Paolieri, Macizo & Bajo, 2014), they are inherently proactive and tend to be automatically encoded in proficient L1 readers when a) information is quickly and easily available in memory, or b) they are necessary to provide text coherence (McKoon & Ratcliff, 1980). Moreover, as the text unfolds, initial predictions can become outdated, in which case they have to be replaced with new inferences, a process known as inferential revision (see Pérez, Cain, Castellanos & Bajo, 2015). Although research into these high-level text comprehension processes in the L2 is scarce, preliminary evidence suggests that readers with advanced L2 proficiency are able to generate inferences in their L1 and L2 (Horiba, 1996), and with increased proficiency, L2 readers seem to show more efficient integration compared to less proficient readers (Yang, 2002). However, no previous study has directly compared inferential revision in the L1 and the L2. Thus, the main goal of the present study was to understand if the ability to revise inferential information during text comprehension is less efficient in the L2 compared to the L1.

A current theoretical hypothesis states that whereas native language processing tends to be highly anticipatory, with comprehenders continuously predicting upcoming information, non-native speakers show a “Reduced Ability to Generate Expectations” (RAGE hypothesis, Grüter & Rohde, 2013; Grüter, Rohde & Schafer, 2014). According to this hypothesis, rather than relying on predictive processing as occurs in L1 comprehension, L2 comprehension primarily relies on a posteriori integration and thus, tends to be relatively less proactive than L1 processing. Studies looking at brain activity during L1 and L2 sentence have provided evidence for this hypothesis.

A sensitive marker related to integration and prediction is the N400 component (Kutas & Hillyard, 1980). The N400 reflects the ease with which the meaning of a word can be integrated into the current mental representation, with larger amplitude for words that are unexpected (e.g., “It was raining so he grabbed his… *coat*”) compared to expected words (e.g., “*umbrella*”). Interestingly, this effect has also been interpreted as indicating the costs of revising an active prediction (scalar inference) in underinformative sentences like "*some people have lungs*" (Nieuwland, Ditman, & Kuperberg, 2010; Nieuwland & Kuperberg, 2008). Evidence using highly constrained sentences indicates that L2 readers are less likely to make lexical predictions (i.e., whether the gender of an article is expected with respect to the predicted word) than L1 readers (Martin et al., 2013; but see Foucart et al., 2014), suggesting L2 comprehension is performed by passive integration of encountered words, rather than by active lexical prediction (see also Lau, Holcomb, & Kuperberg, 2013). Furthermore, a recent study investigating causal inferences during text comprehension (Foucart, Romero-Rivas, Gort, & Costa, 2016) has shown that, in contrast to native speakers, L2 speakers did not manifest significant N400 differences in texts that were causally unrelated compared to causally related texts (instead, they found an early and late positivity). Moreover, when comparing intermediately related and causally unrelated texts, there were no significant differences either in native or L2 speakers, and only a tendency to significance was generally found, showing more negativity in intermediately related than causally related texts. Although the authors argued that this marginally significant result, along with a late negative component (620-750ms), is evidence for the ability of L2 speakers to generate causal inferences during online comprehension, we believe their results are, at least, ambiguous. Overall, then, the previously discussed findings indicate that quantitative differences in the N400 might, in fact, reflect qualitatively different comprehension processes.

The less predictive nature of L2 processing has been interpreted as being due to limited availability of working memory (WM) capacity (Hopp, 2013). L2 processing requires more WM resources compared to L1 processing (Dussias & Piñar, 2010; Ransdell, Arecco, & Levy, 2001), and even simple linguistic processes such as lexical access increase activation in brain areas associated with cognitive control when executed in the L2 (Ma et al., 2014). As more cognitive resources need to be allocated to these lower-level linguistic processes, less resources are available for higher-level semantic-pragmatic processes, where less proficient

L2 comprehenders in particular often experience difficulties (Horiba, 1996; Horiba & Fukaya, 2015; Yang, 2002). Similarly, the depletion of WM resources by lower-level processes can explain why L2 comprehenders might revert from a proactive (active prediction) processing to a more reactive (passive integration) one. However, L1 processing relies heavily on domain-general cognitive control as well, especially when it comes to linguistic processes beyond the single word or sentence level (Arrington, Kulesz, Francis, Fletcher, & Barnes, 2014; Borella, Caretti, & Pelegrina, 2010; Pérez et al., 2015).

An alternative approach to assess the role of cognitive control within-language is to quantify *how* cognitive control is implemented during task performance. According to the general dual-mechanisms framework within the executive control field (Braver, 2012), individuals may employ two different control modes (proactive or reactive) to exert attentional control during ongoing task performance. Which control mode is implemented depends on individual tendencies as well as situational demands. Proactive control is implemented pre-emptively, through sustained goal maintenance and anticipatory monitoring throughout performance of a cognitive task. This type of control is highly dependent on WM capacity (Braver, 2012). Reactive control, on the other hand, consists of the momentary and transient activation of the task goal in the light of conflict or interference, which has been associated with inhibition, both empirically and theoretically (Morales, Gómez-Ariza, & Bajo, 2013). Since reactive control is resource-economic, it is usually employed when cognitive resources are limited either due to individual capacity limits or when the task demands are particularly high. From this perspective, rather than relying on domain-general control to some degree, it might be the case that cognitive control is implemented in a different way when bilinguals are processing in their L2 vs. their L1. For instance, the RAGE hypothesis suggests L1 comprehenders mostly remain in a proactive control mode, but shift towards the less demanding reactive control mode when processing in the L2.

Although many bilingual studies have explored the relationship between language and cognitive control, most of them have focused on how control processes are engaged during language selection (Levy, McVeigh, Marful, & Anderson, 2007; Martín, Macizo, & Bajo, 2010; Morales, Paolieri, & Bajo, 2011; Blumenfeld & Marian, 2013), or on how repeated practice at language selection may enhance attentional control (Bialystok, Craik, & Luk, 2008; Costa, Hernández, & Sebastián-Gallés, 2008; Morales, Yudes, Gómez-Ariza & Bajo, 2015). However, very few studies have directly addressed how cognitive control mechanisms are recruited for the resolution of semantic difficulties during L1 and L2 sentence and/or text comprehension (Moreno, Bialystok, Wodniecka, & Alain, 2010), and none has investigated the role of cognitive control in inferential revision in both languages. Crucially, given that exclusive reliance on active predictions based on previous information (proactive control) would counteract the revision of inferences, inferential revision also requires the flexibility of passive integration (reactive control) to accommodate unexpected upcoming information. Thus, both proactive control, which facilitates the generation of predictive inferences, and reactive control, which enables revision in case prior predictions are not fulfilled, are requisite for successful text comprehension. Accordingly, an additional goal of our study was to explore whether the ability to revise inferential information in L1 and L2 was related to differences in proactive/reactive cognitive control.

1.1. The present study

A paradigm developed to investigate inferential revision is the situation model revision task (Pérez et al., 2015). In this task, participants are presented with short narrative texts (see Table 1). The first three sentences of each text present a constrained Context that facilitates a predictive inference ("*guitar*"). In the following sentence, readers are presented with one of three conditions: neutral, where the sentence does not refer back to the inference ("*…at the prestigious national concert hall.*"); non-update, consistent with the inference primed by the context ("*...with a beautiful curved body.*"); and update, mismatching the inference primed in the context and facilitating the generation of a new inference ("*...with a matching bow.*"). This latter condition primes the replacement of the previous inference with a new one (revision) and the integration of this new information into the situation model. Reading times are measured in the fourth sentence (RT sentence). The final sentence presents the disambiguating word ("*violin*"), that is always inconsistent with the inference primed by the context ("*guitar*"), but consistent with the inferential information facilitated in the update condition ("*matching bow*"). Event-related potentials are recorded during presentation of this disambiguating word (ERP word). Importantly, the status of the ERP word depends on the condition presented in the RT sentence, being a) "expected" when coming from the update condition ("*matching bow*" → "*violin*"), because previous inferential information is coherent; b) "unexpected" when coming from the non-update condition ("*curved body*" related to the idea of guitar → "*violin*"), because previous inferential information is improbable but still plausible; and c) "uncertain" when coming from the neutral condition ("*concert hall*" → "*violin*"), because previous information is not related to the critical words.

<Insert Table 1 about here>

In line with previous literature, reading times for RT sentences allow us to draw conclusions about inference making when reading the three sentences presented as context. Coming from a specific situation (idea of "*guitar*"), longer RTs for the update ("*matching bow*") compared to the non-update ("*curved body*") and neutral ("*concert hall*") conditions would indicate that readers are able to generate the predictive inference facilitated in the context, and subsequently detect a mismatch when new information is presented. In addition, the N400 elicited by the ERP word reflects the processing cost of revising and integrating this word ("*violin*"). Therefore, a reduced N400 elicited in the expected condition (coming from *"matching bow"*) compared to the unexpected (coming from *"curved body"*) and uncertain (coming from *"concert hall"*) conditions, would indicate that comprehenders have been able to take advantage of the prior update information to successfully revise their initial inference, and integrate the new prediction into their situation model. Based on previous literature, we anticipate inference making and revision-integration to be less efficient in the L2 compared to the L1. In addition, given that the only previous study on inferential revision found WM differences to be associated with the revision-integration processes but not with inference making (Pérez et al., 2015), we expect between-language processing differences to be more pronounced for inferential revision than for inference making. Furthermore, RT and N400 data is expected to be modulated by individual differences in both cognitive control and L2 proficiency.

Moreover, to understand whether differences in cognitive control entail processing differences in L1 and L2 text comprehension, we evaluated proactive/reactive control by means of the *Behavioural Shift Index* (BSI) of the AX-CPT task (see Method). This index reflects the individual tendency towards a strong proactive control (BSI near 1) or a balance between proactive and reactive control (BSI near 0). Although previous studies have not explored the effect of cognitive control on high-level comprehension processes in terms of proactive/reactive control, we tentatively propose that efficient inferential revision processes might be supported by a good balance between proactive control (necessary for predictions) and reactive control (required in revision), in both the L1 and L2. Finally, because processing differences due to language status can often be explained in terms of linguistic proficiency alone (Horiba & Fukaya, 2015; Kaan, 2014; Newman, Tremblay, Nichols, Neville & Ullman, 2012; Yang, 2002), we also expect text comprehension processes will be more native-like in more L2 proficient readers.

2. Method

2.1. Participants

Participants were 24 native speakers of English (15 females, 9 males, Age: *M* = 28.58, *SE* = 1.25), who were highly proficient in Spanish, their L2. All participants had learned Spanish after the age of 11 (*M* = 17.21, *SE* = 0.63, range 11-28), and at the time of testing, had been living in a Spanish speaking country for a minimum of 1 year (*M* = 4.05, *SE* = 0.63, range 1-13; see Table 2 for more details). Participants gave their informed consent prior to testing and received a monetary compensation for their participation.

2.2. Materials

Our materials are divided in three different sections: a) language background and cognitive control measures, b) individual differences indices and c) high-level comprehension processes. First, a series of measures were employed to assess participants' language background as well as cognitive control. Secondly, two individual differences indices were extracted from previous measures to evaluate the contribution of L2 proficiency and cognitive control to high-level comprehension processes. Finally, high-level comprehension processes were assessed by means of the situation model revision task.

Language background and cognitive control measures

Language questionnaires. Adapted versions of the Language History Questionnaire (LHQ, Li, Sepanski, & Zhao, 2006) and the Bilingual Switching Questionnaire (BSWQ, Rodríguez-Fornells, Krämer, Lorenzo-Seva, Festman, & Münte, 2012) were used to measure language background, self-assessed L2 abilities in reading, writing, speaking and listening, daily language use and frequency of exposure to friends who were English or Spanish speakers (see Table 2), as well as language switching habits. Differences between L1 and L2 in the two language exposure variables were non-significant: a) percentage of daily language use, $t(23) = 0.70$, $p = .49$, and b) percentage of friends who were native English or Spanish speakers, $t(23) = 0.12$, $p = .99$. These results suggest that bilinguals used their L2 as frequently as their L1 throughout the day, and did so in a similar language environment (i.e., same percentage of friends native speakers of both languages). However, it is important to acknowledge that there was a large variability in the amount of years our bilinguals had been immersed in a Spanish speaking country (see 2.1. Participants).

<Insert Table 2 about here>

Vocabulary tests. A first vocabulary test was created comprising critical words from the situation model revision task. This test served to measure participants' L2 proficiency and was preferred over a more general test to identify any words in the main experimental task that participants may not understand. Words were presented individually in the L2 in the center of a computer screen in randomized order, and participants were asked to translate them into the L1 by typing their response on a blank screen. Participants were given as much time as needed to respond and, in case they did not know the correct translation of a word, were asked to make a guess or try to deduce the meaning. Two versions of this list were created to coincide with the counterbalanced version of the situation model revision task. The accuracy proportion was calculated for each participant to yield individual vocabulary scores (M = 87%, SE = 2%).

A second customized paper and pencil vocabulary test containing any remaining vocabulary participants had not been able to correctly translate in session 1, was administered upon completion of the main task (session 2). The second vocabulary test was also a translation task, but this time critical words were presented within a sentence (taken from the situation model revision task), to assess if participants were able to comprehend the word by using the context while reading the text. Any text containing target words that a participant had failed to correctly translate was excluded from statistical analyses of the situation model revision task for that participant. In addition, participants were asked to assess the perceived difficulty of the texts, to be indicated on a 5-point scale (from 1 = Very easy, to 5 = Very difficult). The average score for perceived difficulty was 2.20 (*SD* = 0.17), indicating that participants found the texts easy to understand.

Verbal fluency. Participants were given a category name, and had to name exemplars from this category. There were two blocks: one in L1-English and the other in L2-Spanish, with two categories in each. Participants were given 30 seconds to name as many exemplars from the current category as possible. Subsequently, participants heard a tone and the word "STOP" appeared on the screen for 1500ms. There was one practice category at the beginning of the task ("*furniture*"). The order of blocks and the two categories included in them (i.e., body parts/professions, colours/fruits and vegetables, and animals/clothes) were counterbalanced across participants. Verbal fluency scores were calculated as the average of correctly named exemplars in the two categories, for the L1 and the L2. In addition, to see the relationship between the two languages, we divided the L2 by the L1 score, where less than 1 means better verbal fluency in the L1 compared to the L2, and more than 1 means the opposite. A mean ratio of 0.48 (range = 0.29-0.61) demonstrated all bilinguals had better verbal fluency in L1 than in L2, suggesting better vocabulary size and faster lexical access in L1. Interestingly, the variability found in these scores also indicated different verbal fluency levels in both L1 and L2.

Working Memory. An operational span task identical to the one described by Morales and colleagues (2015) was used to measure WM capacity. This task was selected to understand how bilinguals maintained and manipulated lexical information in their two languages. Trials consisted of the presentation of an arithmetical equation that may be either correct or incorrect, followed by a single word (either in English or Spanish). Participants were instructed to indicate whether or not each equation was correct by pressing a "Yes" or a "No" button, and memorize the subsequently presented words. Trials were organized into sets of varying size. Each set ended in a recall phase, where participants were asked to recall the words that were presented in the current set. Set size ranged from 2 to 6, and after 3 sets of the same size, set size increased. Again we tested two blocks: one in the L1-English and the other in the L2-Spanish, with 15 sets each. The order of the blocks was counterbalanced across participants. WM scores were calculated by summing up all the correctly recalled words across the 15 sets, for the L1 and the L2. A recalled word was not added to the sum if the corresponding equation had not been solved correctly, or if the word was the last one presented in a set but was typed first during the recall phase. A paired *t*-test comparison in the scores obtained with this WM task showed no significant differences between languages, $t(23) = 1.16$, $p = .27$ ($Ms = 31.19$ and 28.11, for L1 and L2, respectively), discarding the possibility that participants' general WM capacity could explain L1 vs L2 differences in our main experimental task. Similarly, although the operational span task is not a pure measure of verbal WM, the lack of differences between the span scores when performed in the participants' L1 and L2 suggests that language differences in the inferential revision task are not due to differences in WM capacity across the two languages.

Cognitive control. The AX-Continuous Performance Task (AX-CPT) as described by Morales and colleagues (2013) was used to measure cognitive control. Participants saw red and white capital letters presented over a black background. A trial consisted of a sequence of five letters, a red cue letter, three-white distractor letters, followed by a red target letter. Participants were instructed to respond to the final target letter by pressing one of two keys. Specifically, they were instructed to press a "Yes" key only if the target letter was the letter "X", and if the preceding cue had been the letter "A" (AX-trials). In any other case, they were to respond by pressing a "No" key. In addition, participants were asked to always press the designated "No" key in response to each of the white letters presented between cue and target. The task comprised 100 trials, 70% of which were AX-trials. The other trials could comprise a valid cue (A) but invalid target (any other letter than X; AY-trials), an invalid cue (any letter other than A) followed by a valid target (X; BX-trials) or an invalid cue followed by an invalid target (BY-trials), each of which occurred 10% of the time. Cue letters appeared on the screen for a duration of 300ms, distractors and target letters for a duration of 300ms or until participants pressed the “Yes” or “No” key. Between each letter and at the end of a trial a blank screen appeared for the duration of 1000ms. There was a practice phase comprising 10 trials (seven AX-trials, one AY-trial, BX-trial and BY-trial each), during which participants received feedback regarding speed and accuracy of their response. High error rates in the AY-condition reflect reliance on proactive control and/or failure to engage reactive control, whereas high error rates in the BX-condition reflect reliance on reactive control and/or absence of proactive control. Due to the large percentage of AY-trials, this version of the task prompts predominant reliance on proactive control.

Individual differences indices

From the language background and cognitive control tasks described above, we calculated two indices: a) cognitive control, and b) L2 proficiency, to be entered as individual differences variables in the upcoming analysis of the situation model revision task.

Behavioral Shift Index. To measure individual tendencies towards proactive vs. reactive control, we calculated the Behavioral Shift Index (BSI; Braver, Paxton, Locke, & Barch, 2009; see also Chiew & Braver, 2014). More errors/slower RTs in AY trials reflect better proactive control, whereas more errors/slower RTs in BX trials signal better reactive control. The BSI is based on the formula (AY-BX)/(AY+BX), computed for errors, RTs and then averaged. Trials where errors were equal to 0 were corrected as [(errors + 0.5)/ frequency of trials + 1]. Accordingly, a higher BSI (above 0) indicates a preference for proactive control, whereas a smaller BSI (below 0) signals a tendency towards reactive control. However, it is important to note the distribution of our sample in regards to properly interpret individual BSI scores. Young healthy adults do not typically show a preference for reactive control in the AX-CPT task, and accordingly, BSI scores in our sample accumulate in the “proactive range” from around 0 to 1 with very few negative scores at all. Therefore, a high BSI *within our sample* reflects a strong proactive control whereas a small BSI reflects more balanced reliance on proactive and reactive control.

L2 proficiency. To create a comprehensive measure of L2 proficiency, we carried out a Principal Component Analysis (PCA) on three variables: mean scores for L2 reading, writing, speaking and listening abilities, scores obtained in the first vocabulary test, and verbal fluency in the L2 divided by L1 fluency. Only one principal component with an eigenvalue of > 1.0 emerged accounting for 62.03% of overall variance. Preliminary testing revealed a Kaiser-Meyer-Olkin (KMO) value of .51, and significance of Bartlett's test for sphericity (p = .001), justifying the use of PCA. Factor loadings were .91 for vocabulary, .82 for mean L2 ability according to self-assessment, and .59 for L2 verbal fluency. The resulting factor scores, representing variance shared by the three base variables, were extracted and submitted to linear mixed models as proficiency scores (see below).

High-level comprehension processes

Situation model revision task. We used the paradigm developed by Pérez et al. (2015). To assess performance in both languages, all texts were translated from English to Spanish. Each participant was presented with a total of 90 (45 English and 45 Spanish) experimental texts that were five sentences long each (see Table 1), in addition to 3 practice texts at the beginning of each language block. In each text, the first three sentences presented a context that biased a predictive inference ("*guitar*"). Then, the RT sentence could bring a) information not related to the previous prediction ("*…concert hall*"; neutral), b) inferential information consistent with the previous prediction ("*…curved body*"*;* non-update), and c) new inferential information that mismatched the previous prediction and facilitated the generation of a new inference ("*...matching bow*"; update). RTs were measured for this sentence. Word length of the RT sentence did not differ between conditions in English, $F(2, 178) = 1.74$, $p = .18$ (Ms =

11.70, 11.46, and 11.81, for the neutral, non-update and update conditions, respectively), or Spanish, $F(2, 178) = 1.79$, $p = .17$ (*Ms* = 12.00, 12.13, and 12.49, respectively).

The fifth sentence ended with the ERP word ("*violin*") which was always inconsistent with the predictive inference biased by the context ("*guitar*") and consistent with the prediction supported by the RT sentence in the update condition. Consequently, the ERP word was inconsistent in the uncertain and unexpected conditions, prompting inferential revision, and consistent in the expected condition, assuming revision-integration to have taken place already. EEG was recorded at the onset of this word. At the end of each text, a comprehension sentence requiring a true or false judgment was presented to ensure that participants read for understanding.

The two critical words ("*guitar*" and "*violin*") were controlled in character and syllable length, neighbourhood size, age of acquisition, concreteness, frequency, familiarity and imageability for both languages (see Table 3). The words did not differ in most of these measures in the L1-English: number of characters, $t(89) = -0.74$, $p = .47$; number of syllables, $t(89) = -0.47$, $p = .64$; neighbourhood size, $t(89) = -0.48$, $p = .63$; age of acquisition, $t(89) = -2.14$, $p < .05$; concreteness, $t(89) = 2.06$, $p < .04$; frequency, $t(89) = 0.58$, $p = .56$; familiarity, $t(89) = 0.73$, $p = .47$; and imageability, $t(89) = 0.81$, $p = .42$. In addition, none of these measures differed in the L2-Spanish: number of characters, $t(89) = -0.16$, $p = .87$; number of syllables, $t(89) = 0.41$, $p = .69$; neighbourhood size, $t(89) = 0.62$, $p = .54$; age of acquisition, $t(89) = -0.60$, $p = .55$; concreteness, $t(89) = 0.90$, $p = .37$; frequency, $t(89) = 1.19$, $p = .24$; familiarity, $t(89) = 0.67$, $p = .51$; and imageability, $t(89) = 1.34$, $p = .19$.

<Insert Table 3 about here>

A prior norming study suggested that native English speakers activated the non-update concept after reading the context in their L1, but not the update concept (see Pérez et al., 2015). Accordingly, we run a second norming study to provide evidence of concept preferences in native Spanish speakers. Twenty participants (*M* = 30.6 years old; range: 26-33) read the context of each text (sentences 1-3) and were then presented with a single word. Their task was to rate from 1 (Improbable) to 5 (Very probable) how probable was the word in the context of the story. The word was either the non-update concept, which was most strongly supported by the context ("*guitar*"), or the update concept, which was a plausible but not probable alternative ("*violin*"). Two versions of the same questionnaire were created to ensure participants saw only one of the two concepts for each text. A linear mixed model with Participants and Items as random factors and Concept type as fixed factor, was performed on the mean rate. Results demonstrated a main effect of concept type, $F(1) = 199.29$, $p < .001$, $dv = .37$, where the non-update concept was highly probable ($M = 4.40$, $SE = 0.09$) compared to the update concept ($M = 2.65$, $SE = 0.13$), which was considered almost "neutral" (score of 3). As intended, this difference suggests that, after reading the context, native Spanish speakers were significantly more likely to activate the non-update concept than the update concept.

2.3. Procedure

There were two experimental sessions. Participants who volunteered for participation were contacted by e-mail or telephone and asked a few screening questions to ensure whether they met L2 proficiency requirements. Only those participants who had lived in a Spanish-speaking country for at least one year, had learned their L2 after the age of 10 and considered themselves to have a high level of proficiency were invited to session 1. In this session, participants completed both language questionnaires (LHQ and BSWQ) and all behavioral tasks (vocabulary, verbal fluency, WM and AX-CPT). Only participants who reached an accuracy cut-off value of 60% on the vocabulary test were invited to session 2.

In session 2, EEG recordings took place while participants completed the situation model revision task (approx. 90 minutes). This task was administered in two blocks, one in

English and the other in Spanish. Each trial started with a fixation cross ("+") that remained on the screen until the participant pressed the "Yes" or "No" key on the keyboard to start reading. The first four sentences were presented one sentence at a time, and participants were asked to read each sentence at their own pace, pressing the same key to display the next sentence. RTs for the fourth sentence were registered. Subsequently, the fifth sentence was presented word by word with a fixed stimulus-onset asynchrony (SOA) of 300ms per word. In addition, there was a delay of 700ms after the ERP word to ensure the recording of activity during a sufficiently long time window (SOA = 1000ms). Participants were instructed to try not to blink during this final sentence, in order to prevent excessive noise in the EEG data. Finally, a comprehension sentence was presented, and participants were instructed to press the "Yes" key if they thought the sentence was true, or "No" if they thought it was false. Each of the 90 experimental texts was presented to each participant only once, in one of the two languages and the three conditions (six cross conditions). The assignment of language and condition to text was counterbalanced across participants, so that each participant read 15 texts within each factor level combination of condition and language. The order of language block was also counterbalanced. A practice of three trials ensured that instructions were understood. In addition, at the end of the second session, participants were also asked to complete the second vocabulary test (5-10 minutes).

2.4. Apparatus

Tasks were presented by the E-prime software (Schneider, Eschman, & Zuccolotto, 2002), administered on a 19" inch. CRT video monitor (refresh rate = 75 Hz). For the situation model revision task, we recorded scalp voltages using a SynAmps2 64 channels Quik-Cap, plugged into a Neuroscan SynAmps RT amplifier with a continuous sample rate of 250 Hz. The ground (FCZ) and reference (FPZ) electrodes worked as referential signals. The vertical and horizontal electrooculogram (VEOG and HEOG, respectively) was registered supra- and infraorbitally to the left eye and at the outer canthi. Impedances were kept below 5 kΩ. Subsequently, the electrical signal was amplified with a 1-30 Hz band-pass filter. Blinks and ocular movements were corrected by using singular value decomposition. Trials with artifacts (2.94%) were rejected, and recordings from electrodes with high level of artifacts (>1%) were substituted by the average value of the group of nearest electrodes. Epochs from -200 and 800ms with respect to the presentation of the ERP word were averaged and analyzed. We applied a baseline correction, using the average EEG activity in the 200ms previous to target onset as a reference. ERPs were averaged for each factor level combination by participant, text, and region of interest. Individual averages were re-referenced off-line to the average of left and right mastoids (M1 and M2).

2.5. Data analysis

Reading times index. RTs (in milliseconds) were measured for the RT sentence of the situation model revision task. To factor out differences between the L1 and the L2 in baseline reading speed, we divided the RT sentence (fourth sentence) by averaged RTs for the first three sentences (context) of each text. In addition, this also helped to control for differences in text content due to the linguistic properties of each language (e.g., passive vs. active voice in English and Spanish respectively), although it is still possible that cultural aspects may have influenced how our bilinguals were comprehending in both languages.

Event-related potentials. To analyze ERPs, we used the same six regions of interest (ROI) referenced by Pérez et al. (2015): left frontal (F1, F3, F5, FC3, and FC5), right frontal (F2, F4, F6, FC4, and FC6), central (C1, C2, CZ, FCZ, and CPZ), left parietal (P1, P3, P5, CP3, and CP5), right parietal (P2, P4, P6, CP4, and CP6), and occipital (O1, O2, POZ, PO3, and PO4). The N400 component was measured as the mean amplitude (in microvolts) in the time window from 300 to 500ms, averaged for each ROI, and ROI was included as a predictor variable in the N400 analysis. Outliers, defined as amplitude values 2.5 standard deviations above or below the mean by language, condition, and ROI (0.86%), were replaced with corresponding mean values (see Pérez et al., 2015).

Linear mixed models. LMMs were conducted using the lmer function of the lme4 R package, version 1.1-7 (Bates, Maechler, Bolker, & Walker, 2015), with participants and items as random factors, and language, condition, ROI, and centered values for both the BSI of cognitive control and L2 proficiency as fixed factors (see Schielzeth, 2010). Separate models were conducted for each dependent variable (RTs and N400). Thus, the full fixed structure run with RTs contained two three-way interactions (language x condition x BSI + language x condition x L2 proficiency), whereas the full fixed structure of the N400 model contained two four-way interactions (including ROI) as well as all their lower level interactions and main effects. Texts containing target words that participants did not know (11%) and texts to which the comprehension sentence was answered incorrectly (8%), were eliminated from analyses. The mean final number of trials used per participant was 37.79 texts in the L1-English (neutral = 12.79, non-update = 12.33 and update = 12.67), and 35.67 texts in the L2-Spanish (neutral = 11.50, non-update = 11.83 and update = 12.33). Accordingly, the relatively small amount of items used for the electrophysiological analyses entails the need to interpret our N400 results with caution.

First, keeping the full fixed structure, we fitted each model with the maximal random effects structure by participants and items using restricted Maximum Likelihood (Barr, Levy, Scheepers, & Tily, 2013). Convergence problems were solved by removing one by one the effects for which less variance was observed when the summary function was applied to the partially converged solution (for participants or items), until the model converged[4]. Secondly, keeping the maximal random effects structure possible, we conducted stepwise model comparisons starting from the most complex model using Maximum Likelihood (ML) and removing effects that did not account for significant variance in the data, as determined by χ^2

Log-likelihood tests. Finally, for models with significant fixed effects, *p* values were provided by the *anova* function of the lmerTest R package (Kuznetsova, Brockhoff, Christensen, 2015), using ML. Explained deviance was calculated using the *pamer.fnc* function of the LMERConvenienceFunctions R package (Tremblay & Ransijn, 2015). This statistic assesses the overall goodness of fit and serves as a generalization of R^2 by measuring the marginal improvement or reduction in unexplained variability in the fixed component after accounting for a given predictor effect (see Pérez, Joseph, Bajo & Nation, 2016). To follow-up on three-way interactions, we divided the data into subsets according to the levels of language and/or condition and fitted adjusted LMM for these subsets. To qualify two-way interactions, we ran pairwise comparisons within each factor level combination by using the *test Interactions* function of the phia R package (De Rosario-Martínez, 2013).

3. Results

Our results are organized into two sections. We first analyzed the RT index extracted from the RT sentence, and then we examined the N400 amplitude recorded in response to the ERP word (see Table 4 for means and standard errors of RTs and ERP amplitude). Taking into account the large number of results presented in this study, we focused on the fixed effects of each LMM. Summary details (lmerTest package) regarding model fit and random effects of each model are provided in the Appendix.

<Insert Table 4 about here>

3.1. RT sentence: Inference making

To address the question whether comprehenders had previously generated the predictive inference and were able to detect a mismatch in the update condition (see Introduction for hypotheses), we performed a LMM with language (L1 vs. L2), condition (neutral vs. non-update vs. update), and both individual differences indices (BSI and L2 proficiency) as fixed factors, and RT index (RT sentence/context, in milliseconds) as the dependent variable.

Main factors. The final model (Model 1, Appendix) demonstrated significant main effects of language, $F(1) = 5.34$, $p < .05$, $dv = .01$, where readers took longer in the L1 ($M = 0.929$, $SE = 0.01$) compared to the L2 ($M = 0.924$, $SE = 0.02$); and condition, $F(2) = 5.52$, $p < .01$, $dv = 1.38$, where RTs were longer in the update ($M = 1.00$, $SE = 0.02$) compared to the non-update ($M = 0.94$, $SE = 0.02$), $t(89) = 5.93$, $p < .001$, and the neutral condition ($M = 0.92$, $SE = 0.01$), $t(85) = 8.04$, $p < .001$; and the non-update was marginally longer than the neutral condition, $t(84) = 2.26$, $p = .07$. In addition, there was a significant two-way interaction of language x condition, $F(2) = 5.21$, $p < .01$, $dv = .20$ (see Figure 1). According to the interaction, although the effect of condition was significant in both languages [L1-English: $\chi^2(2) = 64.98$, $p < .001$, and L2-Spanish: $\chi^2(2) = 22.96$, $p < .001$], pairwise comparisons within language revealed different patterns. In the L1, comprehenders took longer to read the update compared to the non-update, $t(289) = 6.59$, $p < .001$, and the neutral condition, $t(251) = 7.21$, $p < .001$; with no differences between the non-update and the neutral condition, $t(237) = 0.67$, $p = .98$. These effects indicate that reading in the L1, comprehenders had generated the predictive inference prompted by the context, and then detected new inconsistent information (longer RTs in the update condition). In the L2, on the other hand, the pattern was somewhat different. That is, although comprehenders took longer in the update than in the neutral condition, $t(272) = 4.79$, $p < .001$, the difference between the update and the non-update condition was not significant, $t(290) = 2.11$, $p = .28$; and there was a marginal difference between the non-update and the neutral condition, $t(259) = 2.69$, $p = .08$. These results suggest that reading in the L2, comprehenders performed some level of inferential processing (longer RTs in the update compared to the neutral), but they had difficulties to fully generate the predictive inference prompted by the context (lack of difference between the update and non-update), and therefore, required more information to make it (marginally longer RTs in the non-update compared to the neutral).

<Insert Figure 1 about here>

BSI of cognitive control. The main effect of BSI and its interaction terms with the other variables were dropped from the final model during the backwards stepwise procedure, as neither of them made a significant contribution to the model (all *ps* >.05). Thus, cognitive control did not explain language differences in the ability to predict and subsequently detect a mismatch with that prediction.

L2 proficiency. Participants' L2 proficiency significantly interacted with language, $F(1) = 5.00$, $p <.05$, $dv = .19$, and condition, $F(2) = 6.96$, $p <.001$, $dv = .27$. No other effect was significant, (*ps* >.05). Both two-way interactions with L2 proficiency resulted from opposite regression slopes: a) for language, χ^2 (1) = 5.00, $p <.05$, lower L2 proficiency was related to faster RTs in the L1 compared to the L2 (see Figure 2a); and b) for condition, χ^2 (2) = 13.92, $p <.001$, higher L2 proficiency was associated with longer RTs in the update compared to the non-update and neutral conditions (larger differences between conditions, see Figure 2b). No other pairwise comparison was significant (all *ps* >.36). Therefore, linguistic proficiency signaled a differential tendency between languages, where comprehenders with lower L2 proficiency were faster in their L1, and between conditions, where more L2 proficient comprehenders approached a more native-like inference making.

<Insert Figure 2 about here>

3.2. ERP word: Revision-Integration

To assess the question whether comprehenders had revised their previous prediction and integrated the newly inferred concept into their situation model (see Introduction for hypotheses), we conducted an LMM with language, (L1 vs. L2), condition (uncertain vs. unexpected vs. expected), ROI (left frontal vs. right frontal vs. central vs. left parietal vs. right parietal vs. occipital) and both individual differences indices (BSI and L2 proficiency) as fixed factors, and N400 component (mean amplitude, in microvolts) for the ERP word as the dependent variable.

Main factors. The final model (Model 2, Appendix) showed a main effect of condition, $F(2) = 8.97$, $p < .001$, $dv = .07$, where as predicted, the expected condition manifested less negativity than the unexpected, $t(89) = 3.81$, $p < .001$, and the uncertain, $t(90) = 3.50$, $p < .01$; no differences were observed between the unexpected and the uncertain condition, $t(88) = 0.12$, $p = .99$. Importantly, the two-way interaction between language and condition was also significant, $F(2) = 18.41$, $p < .001$, $dv = 0.08$. No other effects reached significance (all *ps* >.05). Pairwise comparisons within language in the interaction demonstrated that, although the effect of condition was significant in both languages [L1-English: χ^2 (2) = 29.26, $p < .001$, and L2-Spanish: χ^2 (2) = 7.73, $p < .05$], again they revealed different patterns (see Figure 3). In the L1, as hypothesized, the expected condition showed less negativity compared to the unexpected, $t(121) = 4.25$, $p < .001$, and the uncertain condition, $t(118) = 4.96$, $p < .001$; with no differences between the unexpected and uncertain conditions, $t(116) = 0.88$, $p = .95$. This pattern suggests that when reading in their L1, comprehenders revised/integrated their situation model by replacing a misleading predictive inference ("*guitar*") with a more plausible one ("*violin*") in the previous update condition (less negative amplitude in the expected compared to the unexpected and uncertain conditions). In the L2, by contrast, the expected condition was only marginally less negative than the unexpected, $t(125) = 2.77$, $p = .07$; and no differences were found between the expected and the uncertain, $t(124) = 1.54$, $p = .64$, or between the unexpected and the uncertain condition, $t(122) = 1.09$, $p = .89$. Thus, although comprehenders carried out some level of revision when reading in their L2, this process seemed quantitatively less efficient than in the L1 (only marginal tendency of less negativity in the expected than in the unexpected condition). Interestingly, pairwise comparisons within condition demonstrated significant differences between languages only in the expected condition, χ^2 (1) = 9.30, p <.01, with more negativity in the L2 compared to the L1. Language differences were not significant in either the unexpected, χ^2 (1) = 2.19, p =.42, or the uncertain condition, χ^2 (1) = 0.53, p =1.00. Thus, when reading in the L2, comprehenders did not fully replace their initial prediction with the new inference (larger negativity in the expected condition for the L2 compared to the L1), confirming greater difficulties to revise their situation model in L2 processing.

<Insert Figure 3 about here>

BSI of cognitive control. There was a significant three-way interaction between language, condition, and the BSI, $F(2) = 9.17$, p <.001, dv = .07 (see Figure 4). No other effects reached significance (all ps >.05). To follow up on this three-way interaction, we first divided the data by language. Significant interactions between condition and BSI were found in both languages [L1-English, $F(2) = 13.57$, p <.001, and L2-Spanish, $F(2) = 3.61$, p <.03], but once more, analyses within language signaled different patterns. In the L1, a smaller BSI (more balanced reliance on proactive and reactive control) predicted less negativity in the expected condition, χ^2 (1) = 13.32, p <.001; and no differences in the unexpected, χ^2 (1) = 0.22, p =1.00, and the uncertain condition, χ^2 (1) = 0.24, p =1.00. Therefore, when reading in the L1, comprehenders with a more proactive-reactive balance were better at revising the no longer relevant predictive inference and integrating the new inferential information into their situation model. In the L2, on the other hand, the BSI did not manifest differences in any of the three conditions: expected, χ^2 (1) = 0.66, p =1.00, unexpected, χ^2 (1) = 0.04, p =1.00, and uncertain, χ^2 (1) = 2.13, p =.43. The interaction between condition and BSI in the L2 came from opposite regressions slopes between the three conditions, χ^2 (2) = 7.22, p <.05, where a higher BSI (strong proactive control) was associated with more pronounced differences between conditions.

<Insert Figure 4 about here>

Because the interactions between condition and BSI seemed to be associated with different effects of condition within languages, we decided to further explored the three-way interaction by dividing the analysis by condition. A significant interaction between language and BSI appeared only in the expected condition, $F(1) = 10.25$, p <.001, where a smaller BSI (balanced cognitive control) predicted less negativity in the L1, χ^2 (1) = 7.51, p <.05, but no differences in the L2, χ^2 (1) = 1.04, p =.61. The same interaction was not significant either in the unexpected, $F(1) = 0.82$, p =.37, or the uncertain condition, $F(1) = 2.44$, p =.12.

Overall, these findings indicated that differences in cognitive control predict brain activity in the L1 and the L2. In the L1, more balanced reliance on proactive and reactive control showed expectancy effects with less negativity for the expected condition, suggesting the misleading predictive inference was revised and replaced with the new prediction during the previous sentence, being now fully expected. In contrast, in the L2, a strong proactive control caused a general differentiation between conditions suggesting that, although less efficient than in the L1, a strong proactive tendency demonstrated more native-like revision in the L2.

L2 proficiency. The inclusion of L2 proficiency showed a significant two-way interaction between language and L2 proficiency, $F(1) = 6.43$, p <.05, dv = .08, a marginal interaction between condition and L2 proficiency, $F(1) = 2.97$, p =.05, dv = .02, and a significant three-way interaction between language, condition and L2 proficiency, $F(1) = 30.80$, p <.001, dv = .23 (see Figure 5). No other effects were significant (all ps > .05). Once more, to follow up on the three-way interaction, we divided the analysis by language. Similar to BSI, the interaction between condition and L2 proficiency was significant in both languages [L1-English, $F(2) = 13.59$, $p <.001$, and L2-Spanish, $F(2) = 21.23$, $p <.001$], but in different ways. In the L1, higher L2 proficiency was not associated with differences in any of the three conditions: expected, χ^2 (1) = 1.36, p =.73, unexpected, χ^2 (1) = 2.97, p =.25, and uncertain, χ^2 (1) = 2.60, p =.32. The interaction between condition and L2 proficiency in the L1 came from opposite regressions slopes between the three conditions, χ^2 (2) = 17.18, p <.001, where lower L2 proficiency was related to less negativity in the expected compared to the unexpected and uncertain conditions. Therefore, lower linguistic proficiency in the second language was associated with better ability to revise inferential information (a more pronounced difference between conditions) in L1 comprehension. In the L2, on the other hand, higher L2 proficiency showed less negativity in the expected condition, χ^2 (1) = 15.94, p <.001; and no differences were found in the unexpected, χ^2 (1) = 0.002, p =1.00, or the uncertain condition, χ^2 (1) = 1.84, p =.53. Thus, a higher linguistic proficiency in the second language helped readers revising and integrating inferential information into their situation model during L2 processing, but not during L1 processing.

<Insert Figure 5 about here>

4. Discussion

The aim of the present study was twofold. The vast majority of previous research into native vs. non-native semantic processing has been focused on single word or sentence level (e.g., Foucart et al., 2014; Martin et al., 2013), placing constraints on ecological validity. Thus, our first aim was to explore high-level text comprehension processes in late bilinguals' L1 and L2. More concretely, we predicted less efficient inference making and inferential revision-integration in the L2 compared to the L1. Secondly, we investigated to what extent language status differences were modulated by individual differences in cognitive control and

L2 proficiency. A more efficient inferential revision was expected to be explained by a good balance between proactive and reactive control, as well as higher proficiency in the second language, in both the L1 and the L2.

4.1. Inference making in L1 and L2 text processing

In line with previous studies on sentence-level semantic processing, participants' performance on the situation model revision task revealed a number of L1 vs. L2 differences. The RT index for the RT sentence allowed us to assess whether comprehenders had generated the predictive inference facilitated by the context ("*guitar*"), and subsequently detected an inconsistency between this prediction and new inferential information. Accordingly, RTs in the L1 reflected efficient inference making, in that comprehenders took longer to read information that was inconsistent ("*matching bow*") compared to information that was consistent ("*curved body*") or unrelated ("*concert hall*") to the initial inference. This finding replicates the results found in the original study investigating L1 comprehension (Pérez et al., 2015), where both high and low WM readers were able to generate the inference biased in the context and detect new inconsistent information. In the L2, on the other hand, comprehenders also took longer to read the inconsistent (update) compared to the unrelated (neutral) condition, demonstrating they had performed some level of inferential prediction while reading the context and subsequently detected new information in their L2. However, unlike in the L1, comprehenders showed no differences between the consistent (non-update) and the inconsistent (update) condition, suggesting they had problems to fully generate the initial prediction in the L2. In fact, a marginal tendency with comprehenders taking longer in the consistent (non-update) compared to the unrelated (neutral) condition in the L2 also suggested the generation of the predictive inference prompted by the context continued throughout the reading of the RT sentence, and thus, was slower and less efficient in the L2 than in the L1. More specifically, it seems that when reading in their L2, comprehenders need the additional, consistent information given in the non-update condition ("*...with a beautiful curved body"*) to actively predict the inferential concept ("*guitar"*), whereas they were able to form this prediction faster and with less input in their L1. This result is coherent with evidence suggesting that compared to the L1, L2 comprehension relies on passive integration rather than on active lexical prediction (Martin et al., 2013). Indeed, rather than signalling just better mismatch detection in L1 comprehension when encountering inconsistent information ("*matching bow*"), the longer RTs found in the L1 compared to the L2 (main effect of language) may also indicate more efficient processing to generate the second inference ("violin") during online reading comprehension.

4.2. Revision-Integration in L1 and L2 text processing

The next question, then, was whether comprehenders were able to use the new information they received in the update condition of the RT sentence to revise their initial prediction and integrate the new inference into their situation model. To answer this question, we analyzed the N400 component elicited by the ERP word ("*violin*"). When reading in their L1, comprehenders showed less negativity in the expected condition (coming from "*matching bow*") compared to the unexpected (coming from "*curved body*") and the uncertain condition (coming from "*concert hall*"), and no differences between the unexpected and uncertain conditions. These effects signaled greater ease of integration in the expected condition, demonstrating that in their L1, comprehenders were able to revise their initial prediction when previous alternative inferential information had been provided. Importantly, this result also replicates the findings of the original study in L1 comprehension (Pérez et al., 2015), where higher WM readers showed less negativity in the expected compared to the other two conditions. In the L2, by contrast, the expected condition was only marginally less negative than the unexpected, and no differences were found between the uncertain and the other two conditions. Therefore, although comprehenders carried out some level of revision/integration in their L2, they did so quantitatively less efficiently than in their L1. This interpretation was also corroborated by a significant difference between languages in the expected condition, where L2 comprehension manifested more negativity (less revision) than L1 comprehension. Nonetheless, it is also possible that the marginal tendency found in the L2 would have been significant with a larger number of items (approximately 12 texts per condition in the present study after data cleaning), which would indicate similar inferential revision in L1 and L2 comprehension. Future research should try to clarify this question.

Overall, the results of the situation model revision task suggest that the efficiency of predictive inference making and especially inferential revision-integration is reduced in late bilinguals' L2, relative to their L1. Once more, this is consistent with previous evidence showing smaller N400 effects in L2 compared to L1 observed at the sentence level (Martin et al., 2013; Newman et al., 2012). Importantly, these language differences were observed despite the fact that participants in our sample were highly proficient. In line with others, we propose that processing differences by language status might be due to the fact that in L2 comprehension, cognitive resources might be depleted to ensure lower-level processing (such as lexical processing), leaving fewer resources available for conceptual processes (see Horiba & Fukaya, 2015; Segalowitz, Watson, & Segalowitz, 1995; Yang, 2002). Inferential revision in particular is a resource demanding process, as it requires high verbal WM capacity to be efficiently performed (see Pérez et al., 2015). Furthermore, the process of revising outdated inferential information is assumed to demand inhibition (see Pérez et al., 2016), and there is evidence that the inhibition of irrelevant meaning is less efficient in L2 compared to L1 reading (Frey, 2005). Thus, processing differences between the L1 and the L2 might be mediated by the availability of cognitive resources. We address this possibility in section 4.3.

4.3. Cognitive control in L1 and L2 text processing

The second aim of our study was to understand the role of cognitive control in the context of L1 vs. L2 text comprehension. Interestingly, the results of our study suggest that different control styles support text-level processing in L1 and L2 comprehension. In the L1, individual differences in cognitive control explained revision-integration but not inference making (no effect on RTs), which is in line with previous results found with individual differences in WM capacity (see Pérez et al., 2015)[5]. Specifically, inferential revision was most efficient in participants whose performance on the AX-CPT reflected a balance between proactive and reactive control (less negativity in the expected condition for people with smaller BSI). We believe this balance was necessary to first anticipate information by actively generating the predictive inference biased by the context (proactive control), and subsequently disengage from that prediction to accommodate a new inference (reactive control). Thus, very strong reliance on proactive control seems to reduce flexibility to adapt to a new interpretation when new information is encountered. Furthermore, reactive control as measured by the AX-CPT has been related to inhibition (Morales et al., 2013), a process that as we mentioned, is likely involved in overriding the initial inference during revision (see Pérez et al., 2016). Therefore, the process to revise a previous prediction by generating and integrating into the situation model a new inference seems to require a proper balance between proactive and reactive control in L1 comprehension.

Conversely, in L2 comprehension, the BSI did not manifest significant differences on any condition, and a balance between proactive and reactive control was associated with more negativity (less integration) in the expected condition for the L2 compared to the L1. Nonetheless, participants whose AX-CPT performance reflected strong proactive control (higher BSI) tended to show more pronounced differences between conditions, indicating a more native-like revision process. Given that proactive control is implemented through continuous maintenance of the task set (here the situation model) and is resource-costly (Braver, 2012), it seems plausible that a strong tendency towards proactive control helped comprehenders to reduce the costs of lower-level linguistic processes in the L2. Overall, the L2 pattern suggests that the ability to generate a prediction and then, replace it with new information is costlier in the L2, where lower-level linguistic processes tend to be more resource-consuming (e.g. Horiba, 1996; Yang, 2002).

These results contribute to our understanding of the nature of L1 and L2 processing. Recent theoretical proposals hold that L2 processing tends to be less proactive than L1 processing, and that this factor may account for many differences observed between the L1 and the L2 across linguistic domains (RAGE hypothesis, Grüter & Rohde, 2013; Grüter et al., 2014; Grüter, Lew-Williams & Fernald, 2012; Hopp, 2013; Martin et al., 2013). Underlying this notion is the belief that typical L1 comprehension is highly proactive, in that good comprehenders continuously predict upcoming information on the basis of incrementing lexical, semantic and morphosyntactic cues. Our observations support the view that when comprehension requires high-level cognitive processes like the replacement of a previous misleading prediction with a new one, a balance between proactive and reactive control (that is, a more flexible cognitive system) predicts better understanding. At least this was true for L1 comprehension. In the L2, on the other hand, the ability to revise inferential information was not predicted by more flexible cognitive control, but by strong proactive control, suggesting the processing cost of lower-level linguistic processes had been minimized. Interestingly, it has been suggested that the larger N400 effect found in L1 comprehenders could be reflecting the combination of both active prediction of the upcoming information and passive integration of the encountered word, whereas a smaller N400 effect in L2 comprehenders would indicate less semantic processing, with the exclusive use of posteriori integration (Lau et al., 2013, Martin, et al., 2013). Therefore, in relation to our results, the larger N400 effect associated with a balance between proactive and reactive control showed by good L1 comprehenders points at the use of both active prediction and passive integration mechanisms for efficient comprehension in the L1. In contrast, the larger N400 effect associated with proactive control in good L2 comprehenders suggests that an active prediction process was what distinguished them from poor comprehenders.

4.4. L2 proficiency in L1 and L2 text processing

Our data also align with those of others in suggesting that L1 vs. L2 differences might ultimately be due to proficiency asymmetry between the two languages. In regards to the inference making process, higher L2 proficiency was generally associated with longer RTs in the update compared to the non-update and neutral conditions, suggesting a more native-like inference making process. Similarly, the revision-integration process in the L2 was more native-like with higher L2 proficiency, as indicated by less negativity in the expected condition ("*violin*" coming from "*matching bow*"). Note also that to some extent, higher L2 proficiency and proactive control played a similar role in L2 comprehension (both associated with a greater difference between conditions), indicating that to some degree, cognitive and linguistic abilities can compensate each other.

Finally, some attention should be dedicated to findings demonstrating that L2 proficiency predicted performance not just in the L2 but also in the L1. Lower L2 proficiency was related to faster RTs in the L1 compared to the L2, signaling comprehenders with less proficiency in the L2 were faster comprehending in the L1 than those with more proficiency in the L2. As suggested by the N400 results, contrary to L2 comprehension, lower (rather than higher) L2 proficiency was associated with better discrimination between conditions in the L1, indicating a better inferential revision-integration process. Importantly, this difference between conditions in the L1 was attenuated with higher L2 proficiency, indicating less efficient comprehension in the L1 compared to the L2. This means that L1 vs. L2 processing differences were less marked in participants with higher L2 proficiency not just due to enhanced processing efficiency in the L2, but also to reduced efficiency in the L1. Although surprising, these findings cohere with some recent evidence suggesting that the acquisition of a second language later in life can modulate an already established L1 (Baus, Costa, & Carreiras, 2013; Chang, 2012; Linck, Kroll, & Sunderman, 2009; Malt, Li, Pavlenko, Zhu & Ameel, 2015), an effect called attrition. The observation of a smaller N400 effect in the L1 for the most proficient bilinguals also mirrors the findings of one previous study where semantic effects were generally reduced in bilinguals' compared to monolinguals' L1 (Ardal, Donald, Meuter, Muldrew & Luce, 1990). These data suggest that there is a trade-off between L1 and L2 processing efficiency in active bilinguals: to reach very high levels of proficiency in their L2, late bilinguals might "sacrifice" processing efficiency in their L1, or alternatively, it could be the case that a permeable L1 system that is susceptible to change is requisite for reaching native-like proficiency in a late L2 (see Kroll, Bobb, & Hoshino, 2014). Although further study is needed to explore the mechanisms and temporal dynamics underlying this relationship, these findings speak to an evolving and reciprocal relationship between language systems.

4.5. Conclusions

To sum up, the present study extends previous research into sentence processing by showing that the efficiency of high-level text comprehension processes such as inference making and inferential revision-integration, is reduced in an L2 acquired in adulthood, compared to the L1. Modulatory effects suggest that these processing differences may be ultimately rooted in reduced linguistic proficiency and consequentially, limited availability of cognitive resources to engage control processes in the L2. Thus, to some extent, individual differences in cognitive control can compensate limitations in linguistic proficiency, whereas very high proficiency in the L2 can, in principle, compensate non-native language status. Modulatory effects of L2 proficiency on the native language bear witness of a bidirectional and dynamic relationship between a bilingual's language systems. Further study is needed to fully understand the dynamic interaction between L1 and L2 processing, cognitive control and linguistic proficiency, during online text comprehension, as well as possible differences of this interaction by comparing monolinguals and bilinguals.

5. References

Alonso, M. A., Fernández, A., & Díez, E. (2015). Subjective age-of-acquisition norms for 7,039 Spanish words. *Behavior Research Methods*, *47*, 268-274.

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

Arrington, C. N., Kulesz, P. A., Francis, D. J., Fletcher, J. M., & Barnes, M. A. (2014). The contribution of attentional control and working memory to reading comprehension and decoding. *Scientific Studies of Reading, 18*, 325-346.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255-278.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). *lme4* package. Retrieved from https://cran.r-project.org/

Baus, C., Costa, A., & Carreiras, M. (2013). On the effects of second language immersion on first language production. *Acta Psychologica*, *142*, 402-409.

Beeman, M. J., Bowden, E. M., & Gernsbacher, M. A. (2000). Right and left hemisphere cooperation for drawing predictive and coherence inferences during normal story comprehension. *Brain and language*, *71*, 310-336.

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*, 240-250.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*, 235-249.

Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology, 25*, 547-567.

Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning Disabilities, 43,* 541-552.

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences, 106,* 7351-7356.

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends in Cognitive Sciences*, *16*, 106-113.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior research methods*, *46*, 904-911.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11*, 489-503.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, *40*, 249-268.

Chiew, K. S., & Braver, T. S. (2014). Dissociable influences of reward motivation and positive emotion on cognitive control. *Cognitive, Affective, & Behavioral Neuroscience, 14*, 509-529.

Clahsen, H., & Felser, C. (2006). How native-like is non-native language processing? *Trends in Cognitive Sciences, 10*, 564-570.

Costa, A., Hernández, M., & Sebastián-Gallés, N. (2008). Bilingualism aids conflict resolution: Evidence from the ANT task. *Cognition*, *106*, 59-86.

Davis, C. J. (2005). N-Watch: A program for deriving neighborhood size and other psycholinguistic statistics. *Behavior research methods*, *37*, 65-70.

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

De Rosario-Martínez, H. (2015). *Phia* package. Retrieved from https://cran.r-project.org/

Dufour, R., & Kroll, J. F. (1995). Matching words to concepts in two languages: A test of the concept mediation model of bilingual representation. *Memory & Cognition*, *23*, 166-180.

Dussias, P. E., & Piñar, P. (2010). Effects of reading span and plausibility in the reanalysis of wh-gaps by Chinese-English second language speakers. *Second Language Research*, *26*, 443-472.

Duyck, W., & De Houwer, J. (2008). Semantic access in second-language visual word processing: Evidence from the semantic Simon paradigm. *Psychonomic Bulletin & Review*, *15*, 961-966.

Foucart, A., Martín, C. D., Moreno, E. M., & Costa, A. (2014). Can bilinguals see it coming? Word anticipation in L2 sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1461.

Foucart, A., Romero-Rivas, C., Gort, B. L., & Costa, A. (2016). Discourse comprehension in L2: Making sense of what is not explicitly said. *Brain and language, 163*, 32-41.

Frey, L. (2005, January). The nature of the suppression mechanism in reading: Insights from an L1-L2 comparison. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, *27*, 714-719.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*, 191-215.

Grüter, T., & Rohde, H. (2013). L2 processing is affected by RAGE: Evidence from reference resolution. In *the 12th conference on Generative Approaches to Second Language Acquisition (GASLA)*.

Grüter, T., Rohde, H., & Schafer, A. (2014, May). The role of discourse-level expectations in non-native speakers' referential choices. In *Proceedings of the annual Boston university conference on Language Development*.

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*, 33-56.

Horiba, Y. (1996). Comprehension processes in L2 reading: Language competence, textual coherence, and inferences. *Studies in Second Language Acquisition, 18*, 433-473.

Horiba, Y., & Fukaya, K. (2015). Reading and learning from L2 text: Effects of reading goal, topic familiarity, and language proficiency. *Reading in a Foreign Language, 27*, 22-46.

Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*, 257-282.

Kroll, J. F., Bobb, S. C., Misra, M. M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica, 128*, 416-430.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978-990.

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). *lmerTest* package. Retrieved from https://cran.r-project.org/

Lau, E. F., Holcomb, P. J., & Kuperberg, G. R. (2013). Dissociating N400 effects of prediction from association in single-word contexts. *Journal of cognitive neuroscience*, *25*, 484-502.

Levy, B.J., McVeigh, N.D., Marful, A., & Anderson, M.C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science, 18*, 29-34.

Li, P., Sepanski, S., & ZhAo, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods, 38*, 202-210.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*, 1507-1515.

Ma, H., Hu, J., Xi, J., Shen, W., Ge, J., Geng, F., & Yao, D. (2014). Bilingual cognitive control in language switching: An fMRI study of English-Chinese late bilinguals. *PloS one*, *9*, e106468.

Malt, B. C., Li, P., Pavlenko, A., Zhu, H., & Ameel, E. (2015). Bidirectional lexical interaction in late immersed Mandarin-English bilinguals. *Journal of Memory and Language*, *82,* 86-104.

Martín, M.C., Macizo, P., & Bajo, M.T. (2010). Time course of inhibitory processes in bilingual language processing. *British Journal of Psychology, 101*, 679-693.

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language, 69*, 574-588.

McKoon, G., & Ratcliff, R. (1980). The comprehension processes and memory structures involved in anaphoric reference. *Journal of Verbal Learning and Verbal Behavior*, *19*, 668-682.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation, 51*, 297-384.

Morales, J., Yudes, C., Gómez-Ariza, C. J., & Bajo, M. T. (2015). Bilingualism modulates dual mechanisms of cognitive control: Evidence from ERPs. *Neuropsychologia, 66*, 157-169.

Morales, J., Gómez-Ariza, C. J., & Bajo, M. T. (2013). Dual mechanisms of cognitive control in bilinguals and monolinguals. *Journal of Cognitive Psychology, 25*, 531-546.

Morales, L., Paolieri, D., & Bajo, M.T. (2011). Grammatical gender inhibition in bilinguals. Frontiers in Psychology, *2*, 284.

Moreno, S., Bialystok, E., Wodniecka, Z Alain, C. (2010) Conflict resolution in sentence processing by bilinguals. *Journal of Neurolinguistics*, *23*, 564-579.

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

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of memory and language*, *63*, 324-346.

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, *19,* 1213-1218.

Pérez, A., Cain, K., Castellanos, M. C., & Bajo, T. (2015). Inferential revision in narrative texts: An ERP study. *Memory & Cognition, 43,* 1105-1135.

Pérez, A., Joseph, H. S., Bajo, T., & Nation, K. (2016). Evaluation and revision of inferential comprehension in narrative texts: an eye movement study. *Language, Cognition and Neuroscience*, *31*, 549-566.

Pérez, A. I., Paolieri, D., Macizo, P., & Bajo, T. (2014). The role of working memory in inferential sentence comprehension. *Cognitive processing*, *15,* 405-413.

Ransdell, S., Arecco, M. R., & Levy, C. M. (2001). Bilingual long-term working memory: The effects of working memory loads on writing quality and fluency. *Applied Psycholinguistics*, *22*, 113-128.

Rodríguez-Fornells, A., Krämer, U. M., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology, 2,* 1-15.

Schielzeth, H. (2010). Simple means to improve the interpretability of regression coefficients. *Methods in Ecology and Evolution, 1,* 103-113.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Incorporated.

Segalowitz, N., Watson, V., & Segalowitz, S. (1995). Vocabulary skill: Single-case assessment of automaticity of word recognition in a timed lexical decision task. *Second Language Research, 11,* 121-136.

Slabakova, R. (2006). Is there a critical period for semantics? *Second Language Research, 22,* 302-338.

Tremblay, A., & Ransijn, J. (2015). *LMERConvenienceFunctions* package. Retrieved from https://cran.r-project.org/

Van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension (pp. 11-12). New York: Academic Press.

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in the bilingual brain. *Neuron, 37,* 159-170.

Yang, Y. F. (2002). Reassessing readers’ comprehension monitoring. *Reading in a Foreign Language, 14,* 18-42.

Footnotes

[1] First authorship is shared between the first two authors of the manuscript.

[2] Present affiliation/address: Theoretical and Applied Linguistics (TAL), University of Cambridge, English Faculty Building, 9 West Road, Cambridge, CB3 9DP, United Kingdom.

[3] Present affiliation/address: Mühlenbergklinik Holsteinische Schweiz, Frahmsallee 1-7, 23714 Bad Malente, Germany.

[4] Removing the intercepts of participants and items or the correlation between those intercepts and the random slopes did not solve convergence problems.

[5] Please, note that in this study we controlled for general WM differences across languages, however it is still possible that a more demanding verbal WM task would have explained (at least partly) the between-language differences found in our situation model revision task.