Gender and number processing in Chinese learners of Spanish – Evidence from Event Related Potentials

Margaret Gillon Dowens a,b,d,*, Taomei Guo c, Jingjing Guo c, Horacio Barber d, Manuel Carreiras b,e

a Centre for Research in Applied Linguistics, University of Nottingham Ningbo, China  
b Basque Center on Cognition, Brain and Language, Donostia-San Sebastián, Spain  
c State Key Laboratory of Cognitive Neuroscience and Learning, BNU, China  
d University of La Laguna, Tenerife, Spain  
e IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

A R T I C L E   I N F O

Article history:  
Received 2 September 2010  
Received in revised form 11 February 2011  
Accepted 16 February 2011  
Available online 22 February 2011

Keywords:  
Bilingualism  
Late learners  
Morphosyntax  
ERPs

A B S T R A C T

Traditionally, age of acquisition (AoA) has been considered the single most important factor in second language (L2) acquisition and processing, particularly in the area of syntax processing. However, there is now growing evidence of the importance of other factors, such as the level of proficiency attained and the degree of overlap or similarity between the first language (L1) and L2 structures and possibility of transfer of features and/or processing routines. However, the relative importance of these factors and the nature of L1–L2 transfer are still unclear. To shed light on these issues, we recorded the electrical brain activity of a group of Chinese proficient late learners of Spanish, using the Event Related Potentials technique, while they read Spanish sentences containing violations of number and grammatical gender agreement (adjective–noun agreement and article–noun agreement). Unlike Spanish, Mandarin Chinese is an isolating language in which morphosyntactic features such as gender and number are not computed and so the ERP results from this group can help to clarify the role of L1–L2 transfer in morpho-syntax processing routines. The results included P600 effects for both gender and number agreement violations, with no differences between these disagreement conditions. These results are taken to support second language acquisition models which stress the roles of proficiency and L1–L2 transfer in L2 syntax processing.

A R T I C L E   B O T T O M

1. Introduction

In a shrinking world with an increasingly global economy it is hardly surprising that more people than ever are learning a second language as adults. How second languages (L2) are acquired and processed has thus become an increasingly fertile and important area of research. This is not only because of the insights that second language research can provide about the principles and mechanisms underlying all language processing but also because of the important political, social and educational implications of L2 research in an increasingly multilingual society. However, despite recent progress, the study of how the brain deals with more than one language is still at an early stage and much more research is needed to resolve basic issues.

One of the important unresolved debates in this area concerns whether syntax processing of second languages acquired after early childhood can become qualitatively similar to first language (L1) processing at high levels of proficiency and what factors determine the nature of L2 syntax processing. Traditionally, theories of language acquisition and processing – see, for example, the Failed Functional Feature Hypothesis (Hawkins & Chan, 1997), have assumed that when the L2 is acquired after early childhood, syntax processing will inevitably be qualitatively different from that of the L1. Age of acquisition has thus been considered the determining factor in this question, with posited maturationally determined critical or sensitive periods for second language acquisition beyond which the L2 will not be processed in a native-like manner, even though high levels of functional proficiency are achieved. However, there is at present no general consensus about the existence of critical or sensitive periods for all aspects of second language processing, with many researchers preferring to describe age-related effects in L2 acquisition as limited to specific domains such as phonology, or indeed in terms of general cognitive decline across the age-span differentially affecting areas of L2 acquisition and processing (see Birdsong, 2006; DeKeyser & Larsen-Hall, 2005; Thomas & Johnson, 2008, for discussions of the critical period debate). Thus, there are language acquisition models, such as the Full Access–Full Transfer Model (Schwartz & Sprouse, 1996) or the Competition Model (Bates

* Corresponding author at: Centre for Research in Applied Linguistics, University of Nottingham Ningbo, 199 Taikang East Road, Ningbo, Zhejiang 315100, China.  
Tel.: +86 574 8818 0515; fax: +86 574 8818 0175.  
E-mail address: Margaret.Dowens@nottingham.edu.cn (M.G. Dowens).

0028-3932/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.  
doi:10.1016/j.neuropsychologia.2011.02.034
Those of English late learners of Spanish (Gillon Dowens, Vergara, Carreiras, 2005) and so provide information about the roles of age of acquisition and proficiency in L2 processing. Furthermore, we will compare the results obtained from the Chinese L1 group with different from those reported for native speakers of Spanish (Barber & Carreiras, 2010) have coincided in indicating the need for more online studies of accuracy or from reaction times during grammatical decision and similar tasks. While these have provided very useful insights, at advanced levels of L2 proficiency behavioural results that indicate L1–L2 equivalence can in fact mask underlying processing differences, which has become increasingly evident in recent online studies of L2 syntax processing in late learners (Abutalebi & Green, 2007; Kotz, 2009; Mueller, 2005; van Hell & Tokowicz, 2010).

Recently, two factors have come to light as being particularly relevant to this issue. The first is L2 proficiency, as recent electrophysiological results for high proficiency late learners (Hahne, Mueller, & Clahsen, 2006; Rossi, Gugler, Friederici, & Hahne, 2006) are very different from earlier studies of lower proficiency participants (Weber-Fox & Neville, 1996). The other important factor that is increasingly seen as relevant is that of L1–L2 typological similarity/difference and the possibility of L1 transfer of language features and processing routines (Frenck-Mestre, Osterhout, McLaughlin, & Foucart, 2008; Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005). However, although language transfer has been widely discussed in second language research and examined using offline measures, to date, there has been surprisingly little systematic research of this question using electrophysiological and neuroimaging techniques and the few results obtained have not been consistent. While some authors have interpreted their findings in terms of L1 transfer effects (Guo, Guo, Yan, Jiang, & Peng, 2009; Jeong et al., 2007; Kotz, Holcomb, & Osterhout, 2008; Sabourin & Stowe, 2008), others found no evidence of transfer (Clahsen & Felser, 2006; Marinis, Roberts, Felser, & Clahsen, 2005; Papadopoulou & Clahsen, 2003). It is hardly surprising then, given these contrasting results, that several recent reviews of L2 syntax research (Kotz, 2009; Mueller, 2005; Van Hell & Tokowicz, 2010) have coincided in indicating the need for more online studies of highly proficient late learners and for studying different language combinations, to tease out the roles of AoA, proficiency and L1–L2 interaction in L2 processing. The aim of this study is to respond to this need. We will examine morphosyntax processing in proficient Chinese late learners of Spanish to see whether the electrophysiological correlates of this processing are similar or different from those reported for native speakers of Spanish (Barber & Carreiras, 2005) and so provide information about the roles of age of acquisition and proficiency in L2 processing. Furthermore, we will compare the results obtained from the Chinese L1 group with those of English late learners of Spanish (Gillon Dowens, Vergara, Barber, & Carreiras, 2010) to shed light on the question of L1–L2 transfer effects.

One online technique which is particularly suited to measuring the fast complex brain processes involved in language processing is that of Event Related Potentials (ERPs), due to its excellent temporal resolution in milliseconds and the multidimensional data points it provides in terms of latency, amplitude, polarity and scalp distribution. Another important advantage of using ERPs to study morphosyntactic processing is the considerable amount of first language research already carried out in this domain, which has revealed a consistent pattern of findings against which to compare late L2 learners. Studies of morphosyntactic agreement violations in L1 regularly result in the same ERP components. The first is a negative deflection in the waveform with a maximum between 300 and 500 ms after stimulus onset and usually present over left anterior scalp sites (Barber & Carreiras, 2005; Gunter, Friederici, & Schriefers, 2000; Hahne & Friederici, 1999), although other scalp distributions have been reported (Hagoort, Wassenaar, & Brown, 2003; Silva-Pereyra & Carreiras, 2007). This component is thus usually referred to as the Left Anterior Negativity (LAN). The second component is a longer-lasting positive deflection between 500 ms and 900 ms after stimulus onset with a central–parietal scalp distribution, usually referred to as the P600 (Friederici, Hahne, & Saddy, 2002; Hagoort & Brown, 1999; Osterhout & Holcomb, 1992). These effects have been reported in many languages including English (Coulson, King, & Kutas, 1998), Dutch (Gunter et al., 2000), Spanish (Barber & Carreiras, 2005), Finnish (Palolahti, Leino, Jokela, Kopra, & Paavilainen, 2005) and Hebrew (Deutsch & Bentin, 2001). Although Chinese lacks grammatical inflections marking relations such as gender, number, person and case, LAN and P600 effects have been found in Chinese in response to local syntactic violations (Jiang & Zhou, 2009; Ye, Luo, Friederici, & Zhou, 2006). In the former study, a LAN effect was elicited by violations of local phrase structure constraints (misapplication of structural auxiliaries marking adjective or adverb category). As these structural auxiliaries in Chinese function to specify a particular syntactic category to the lexical items with which they are connected, the authors interpreted the LAN result as indicating that a morphosyntactic process was involved in this type of violation.

Most L1 studies of morphosyntax (e.g. case, number and gender violations) have reported either a biphasic pattern of LAN followed by P600 (Barber & Carreiras, 2005; Gunter et al., 2000; Molinaro, Vespiognani, & Job, 2008) or only P600 effects (Hagoort, 2003). However, despite the ubiquitousness of these ERP responses to L1 morphosyntactic processing, there is still a great deal of debate about their functional significance and whether they are domain-specific to language. Thus, some authors consider the LAN to be indexing general rule-based transformation processing (Hoen, Michel, & Dominey, 2000) or working memory constraints (King & Kutas, 1995; Klunder & Kutas, 1993) and the P600 to be a reflection of anomaly detection or general monitoring processes (Hagoort, 2003; van Herten, Koik, & Chwilla, 2005). However, most studies have reported these components within the context of language processing. One popular interpretation of these ERP results within a general model of syntax processing is that of Friederici (2002). In this author’s view, the LAN is a marker of an early stage of morphosyntactic processing, while the P600 is indicative of a later, more controlled processing stage of realanalysis and repair of the syntactic structure.

This consistent pattern of ERP results for L1 morphosyntactic agreement violations is, then, a useful one against which to compare data from L2 late learners. In the ERP studies carried out so far, however, the results for late learners are not as consistent as those for native speakers. While some studies reported no evidence of these components, or only P600 effects which were different from those of native speakers in terms of delayed latency and reduced amplitude (Chen, Shu, Liu, Zhao, & Li, 2007; Guo et al., 2009; Hahne & Friederici, 2001; Hahne, 2001), other stu-
ies have sometimes yielded more native-like P600 effects or the LAN/P600 pattern, depending on the experimental manipulation and the L1–L2 combinations studied (Hahne et al., 2006; Iset, 2007; Ojima et al., 2005; Rossi et al., 2006). Some authors, therefore, have argued for the importance of L1–L2 transfer effects in determining the nature of L2 processing. In a recent electrophysiological study (Gillon Dowens et al., 2010), we examined the ERP responses of a group of high proficiency English L1 late learners of Spanish while they read Spanish sentences containing grammatical gender and number agreement violations. The results showed the native-like LAN–P600 pattern for gender and number violations in the within-phrase (determiner–noun agreement) condition and P600 effects for both types of violation in both the within-phrase and the across-phrase (noun–postmodifying adjective agreement) conditions. Interestingly, however, in contrast to Spanish native speakers, there were significant latency and amplitude differences between number and grammatical gender processing, with number – the morphosyntactic feature that is present in English – showing “stronger” effects (earlier onset and increased amplitude) than grammatical gender, which is a syntactic feature that is not present in English. These data were interpreted as evidence for transfer effects, with the representations and/or processing routines already present in the L1 facilitating L2 feature processing. However, it could also be argued that there are differences between L2 gender and number processing per se that resulted in these different effects, so to explore this L1 feature transfer hypothesis, it is important to look at what happens with different types of L1–L2 language combination. Spanish is a language which is rich in morphosyntactic agreement. In English, further along the agglutinative-isolating spectrum, some inflectional agreement features are also computed, although considerably less than Spanish. Therefore, to clarify this issue of L1 syntactic feature transfer, it is interesting to look at gender and number agreement processing in participants whose L1 is an analytic language where features such as case, gender and number are not marked and so no such inflectional agreement is computed. One of the clearest examples of this is Mandarin Chinese (for descriptions of Chinese, see Grosjean, Li, Münthe, & Rodríguez-Fornells, 2003; Li, Bates, Tan, & Tzeng, 2006; Zhou, Ye, Cheung, & Chen, 2009). If the Chinese L1 late learners of Spanish show an ERP pattern which is similar for gender and number processing, in contrast to English late learners, this would provide evidence for the role played by L1 transfer in the morphosyntax processing of English L1 learners of Spanish.

Our predictions for this study, therefore, could include the following: firstly, if age of acquisition is the only important factor affecting L2 morphosyntax processing, then there may be no evidence of the LAN and/or P600 components usually reported for native speakers. This has in fact been the finding from recent studies of grammatical agreement processing in Chinese late L2 learners, which have yielded no evidence of the LAN–P600 pattern (Chen et al., 2007; Guo et al., 2009). However, it is in contrast to recent studies from European language combinations, showing comparable ERP results for native speakers and late L2 learners (Gillon Dowens et al., 2010; Rossi et al., 2006). So if, as these studies would seem to indicate, achieved proficiency in the L2 is a factor which can offset age of acquisition effects, then there may be evidence of these syntax-related components.

Secondly, if the results yield ERP effects for gender and number agreement violations, with significant differences between these conditions, as was the case for English L1 late learners of Spanish, this would argue against the role of L1 transfer, as the Chinese participants cannot transfer either of these features from the L1. On the other hand, if differences in the ERP correlates between the agreement and the agreement violation conditions are of the same size for gender and number, this would provide support for the role of L1 transfer in the English late learners’ processing of Spanish.

Thus, examining on-line how proficient Chinese late learners process sentences containing grammatical gender and number agreement violations in Spanish can help to clarify the question of L1–L2 transfer effects, as well as shedding further light on the roles of AoA and proficiency in L2 morphosyntax processing.

2. Methods

2.1. Participants

Twenty-six Chinese L1 late learners of Spanish (14 women, age range 20–24 years) participated in the experiment in exchange for a small sum. Two participants were excluded from the statistical analyses due to too many musculature artifacts in the EEG. The participants were recruited from the modern language faculties of different universities in China and the Cervantes Institute of Beijing and all of them were at the end of third or fourth year of a degree in Spanish studies. The participants were L1 speakers of Chinese, with no history of neurological or psychiatric impairment and with normal or corrected-to-normal vision. All of the participants started to learn Spanish after the age of 18. Most of the participants were also taking an intensive course to be Spanish–Chinese interpreters during the Olympic Games. They had been selected for this course due to their high levels of proficiency in Spanish as measured by a Cervantes Institute aptitude test equivalent to C1 (Effective Operational Proficiency) in the Common European Framework of Reference for Languages. All participants were right-handed, as assessed by an abridged version of the Edinburgh Handedness Inventory (Oldfield, 1971). They filled in a self-rating questionnaire about their language background, daily use of Chinese and Spanish and their proficiency in both languages. The results indicated that the participants used both Chinese and Spanish on a daily basis and rated their competence in Spanish as being good to excellent (on a scale of 1–4, the average for speaking was 3, listening 3, reading 3 and writing 2). After the experiment, they also filled in a series of short grammatical tests, with a section on adjective agreement which included the experimental items. The results from this confirmed that off-line, the participants were able to choose the correct form of the adjective in terms of gender and number agreement, at 96% accuracy.

2.2. Materials

The materials were the same as those used by Barber and Carreiras (2005) with Spanish native speakers and by Gillon Dowens et al. (2010) with English–Spanish bilinguals. The details of these materials are as follows: two lists of 120 experimental sentences each were generated in order to manipulate agreement at the beginning and in the middle of the sentence. The first list manipulated the article–noun agreement at the beginning of the sentence, while the second list manipulated the agreement relationship between the noun and its post-modifying adjective which appeared after the verb, in the middle of the sentence. Examples of the sentences used can be seen in Table 1.

Between two and four filler words were included after the target adjective in order to avoid wrap-up effects on the target words. Three versions of each list were created in order to counterbalance the experimental conditions. Assignment of sentences to conditions in each version was counterbalanced across participants. Thus, each sentence occurred three times across subjects, once in each condition, so that each subject only saw one form of each sentence. In addition, a list of 80 well-formed sentences was included. Fillers included nouns with opaque or irregular gender and adjectives with neuter gender, to avoid strategies based on purely orthographic features. All the target items were medium to high frequency words and contained between 3 and 7 letters. The mean length and frequency of the words in each experimental condition were equivalent in each counterbalanced group. Word frequencies were obtained from the LEXESP database (Sebastián-Gallés, Martí, Carreiras, & Cueto, 2000).

2.3. Procedure

Participants were seated comfortably in a sound-attenuated chamber. All stimuli were presented on a high-resolution monitor that was positioned at eye level 80–90 cm in front of the participant. The words were displayed in light-gray lowercase Arial 36 against a black background. Participants performed a grammatical decision task: they were instructed to press one of two keys (L and S) on a keyboard to indicate whether the sentence was grammatically correct or not. For half of the participants the right key (L) was used to signal the “Yes” response and left key (S) was assigned the “No” response. For the remaining participants the order was reversed. The sequence of events in each trial is described as follows: First, a fixation point (•) appeared in the centre of the screen and remained there for 700 ms. This fixation point was followed by a blank screen interval of 300 ms, then the sentence was displayed word by word. Each word appeared for 300 ms and was followed by a 300 ms blank interval. At the end of the sentence, a question mark was presented on a scale of 1–4, the average for speaking was 3, listening 3, reading 3 and writing 2. After the experiment, they also filled in a series of short grammatical tests, with a section on adjective agreement which included the experimental items. The results from this confirmed that off-line, the participants were able to choose the correct form of the adjective in terms of gender and number agreement, at 96% accuracy.
and were instructed to favor accuracy over speed in their responses. A practice session of 5 trials was given. The sentences were presented in 3 blocks of approximately 14 min each, with two short rest breaks. The whole experiment lasted approximately 90 min, including set-up time.

2.4. EEG recording and analyses

A 64 Channel QuickCap (NeuroScan) with Sintered Ag/AgCl electrodes was used to collect the EEG data (see Fig. 1 for recording sites). The vertical electrooculogram (VEOG) was recorded from electrodes placed above and below the right eye and the horizontal electrooculogram (HEOG) was recorded from electrodes placed at the outer canthus of each eye. The EEG recording was referenced online to the left mastoid and re-referenced off-line to the averaged mastoids. All impedances were kept below 5 kΩ. The EEG signal was digitized with 500 Hz and amplified within a band pass from 0.05 to 100 Hz. The signals were sampled continuously throughout the experiment with a sampling rate of 250 Hz.

Epochs of the EEG corresponding to 1500 ms after word onset presentation were averaged and analyzed. Baseline correction was performed using the average EEG activity in the 200 ms preceding the onset of the target stimulus as a reference signal value. Following baseline correction, epochs with artifacts were rejected. In addition, trials that were not responded to correctly were not included in the analysis. This resulted in the exclusion of approximately 16% of the trials. Separate ERP were formed for each of the experimental conditions, each of the subjects and each of the electrode sites. Nine regions of interest were computed out of the 58 electrodes, each containing the mean of a group of electrodes. These regions were (see electrode numbers in Fig. 1): left-anterior (F1, F3, F5, FC1, FC3, FC5), left-medial (C1, C3, C5, CP1, CP3, CP5), left-posterior (P1, P3, P5, P03, P05, P07), right-anterior (F2, F4, F6, FC2, FC4, FC6), right-medial (C2, C4, C6, CP2, CP4, CP6), right-posterior (P2, P4, P6, P04, P06, P08), midline anterior (Fz, FCz), midline-medial (Cz, CPz) and midline-posterior (Pz, POz).

Mean amplitudes were obtained for different time windows. For each window, two repeated-measures ANOVAs were performed, one for the lateral electrode groups and one for the midline electrode groups. The first included grammatical agreement (agreement, gender violation and number violation), sentence position (beginning/middle), hemisphere (left/right) and electrode region (anterior, medial and posterior) as factors. Where appropriate, critical values were adjusted using the Greenhouse-Geisser (1959) correction for violation of the assumption of sphericity. In addition, post hoc Sidak contrasts were performed after interactions or main effects of grammatical agreement to control for type I error in multiple comparisons. Effects for the sentence position factor, the hemisphere factor, and for the electrode region factor are only reported when they interact with the experimental manipulations. Reaction times were not analyzed, as the response in the grammatical judgment task was delayed and participants were instructed to favor accuracy over speed. The overall mean percentage of correct answers was 90% (range 84–96%).

3. Results

The ERP grand averages time-locked to the onset of the target words are represented in Figs. 2 and 3 over twelve representative recording sites, showing the three conditions: Agreement, Gender violation and Number violation. Fig. 2 shows the grand averaged waveforms corresponding to the agreement manipulations in the first sentence position (beginning of the sentence/within-phase violation) and Fig. 3 the second sentence position (middle of the sentence/ across-phase violation). In the first time window, between 300 and 500 ms, there appear to be no differences between the agreement and the disagreement conditions. Then, between 500 and 800 ms there is a widely distributed positivity consisting in larger positive amplitudes for both disagreement conditions compared to the agreement condition. Finally, a sustained negativity in frontal scalp areas can be observed from 1000 to 1300 ms, with more negative amplitudes for both disagreement conditions. These effects seem to be similar for the first and second sentence positions.

Following visual inspection of the grand averaged waveforms, omnibus ANOVAs were performed separately for lateral and midline electrode groups, in 3 different temporal windows, on the basis of calculations of mean amplitudes: between 300–500 ms for the LAN effect, 500–800 ms for the P600 effect and 1000–1300 ms for the sustained negativity effect.

3.1. 300–500 ms time window: LAN

Lateral electrode groups. There were no significant effects of Grammatical Agreement at the beginning or at the middle of the sentence in this time window in lateral or midline electrode groups.

Midline electrode groups. There were no significant effects of Grammatical Agreement at the beginning or at the middle of the sentence in this time window in lateral or midline electrode groups.

3.2. 500–800 ms time window: P600

Lateral electrode groups. The ANOVA revealed two interactions in this time window. The first was an interaction of grammatical agreement × hemisphere (F2,92 = 6.05; p < .01; ε = .436). Post hoc tests revealed that in both hemispheres both disagreement conditions were significantly more positive than the agreement condition. The second interaction was between grammatical agreement and electrode region (F4,92 = 8.25; p < .005; ε = .770). Paired comparisons revealed significant differences between the agreement and the disagreement conditions only in the medial and posterior regions, with both disagreement conditions more positive than the agreement condition. The statistical results of the paired contrasts can be found in Table 2.

Midline electrode groups. The ANOVA revealed an interaction of grammatical agreement and electrode region (F4,92 = 4.89; p < .05; ε = .452). Paired comparisons revealed significant differences between the agreement and the disagreement conditions in all three regions, with both disagreement conditions more positive than the agreement condition. The statistical results of the paired contrasts can be found in Table 2.

3.3. 1000–1300 ms: late frontal negativity

Lateral electrode groups. In this time window the ANOVA revealed an interaction between the factors grammatical agreement and electrode group (F4,92 = 9.31; p < .001; ε = .568). Paired comparisons indicated significant differences only in the anterior region, with the gender disagreement condition marginally more negative than the agreement condition and the number disagreement condi-

Table 1

<table>
<thead>
<tr>
<th>Position of manipulation</th>
<th>Conditions</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning of sentence</td>
<td>Agreement</td>
<td>El suelo está plano y bien acabado</td>
</tr>
<tr>
<td></td>
<td>Gender non-agreement</td>
<td>La suelo está plano y bien acabado</td>
</tr>
<tr>
<td></td>
<td>Number non-agreement</td>
<td>Los suelo está plano y bien acabado</td>
</tr>
<tr>
<td>Middle of sentence</td>
<td>Gender disagreement</td>
<td>Themas-sex.-singuflor mas.-singuflor was long</td>
</tr>
<tr>
<td></td>
<td>Number disagreement</td>
<td>Themas-sex.-singuflor mas.-singuflor was long</td>
</tr>
</tbody>
</table>

Themas.-sing. flight was long (mas-.-, masc.-sing) but pleasant
tion more negative than the agreement condition. Statistical results of the paired contrasts can be found in Table 2.

Midline electrode groups. The midline ANOVA in this window again revealed an interaction between the factors grammatical agreement and electrode group \((F_{4,92} = 7.57; p < .01; \varepsilon = .551)\). Paired contrasts revealed that the gender disagreement condition was marginally more negative than the agreement condition and the number disagreement condition was more negative than the agreement condition. Statistical results of the paired contrasts can be found in Table 2.

Table 2
Paired contrasts.

<table>
<thead>
<tr>
<th>Time windows</th>
<th>df</th>
<th>Gender</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lateral analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300–450 ms – no sig. effects</td>
<td>1</td>
<td>13.03</td>
<td>22.91</td>
</tr>
<tr>
<td>500–800 ms</td>
<td>1</td>
<td>18.99</td>
<td>27.26</td>
</tr>
<tr>
<td>Agreement × hemisphere</td>
<td>1</td>
<td>20.66</td>
<td>30.11</td>
</tr>
<tr>
<td>Agreement × electrode group</td>
<td>1</td>
<td>25.90</td>
<td>39.33</td>
</tr>
<tr>
<td>1000–1300 ms</td>
<td>1</td>
<td>5.99</td>
<td>10.21</td>
</tr>
<tr>
<td>Agreement × electrode group</td>
<td>1</td>
<td>7.78</td>
<td>11.86</td>
</tr>
<tr>
<td>Midline analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300–450 ms – no sig. effects</td>
<td>1</td>
<td>7.78</td>
<td>11.86</td>
</tr>
<tr>
<td>500–800 ms</td>
<td>1</td>
<td>25.76</td>
<td>34.06</td>
</tr>
<tr>
<td>Agreement × electrode group</td>
<td>1</td>
<td>33.18</td>
<td>46.08</td>
</tr>
<tr>
<td>1000–1300 ms</td>
<td>1</td>
<td>6.35</td>
<td>12.01</td>
</tr>
<tr>
<td>Agreement × electrode group</td>
<td>1</td>
<td>4.05</td>
<td>6.01</td>
</tr>
</tbody>
</table>

Fig. 1. Electrode groups.
Fig. 2. First sentence position.

Fig. 3. Second sentence position.
3.4. Summary of results

The violations of grammatical agreement in the sentences presented resulted in a clear pattern of results with no significant differences between the two sentence positions. In the 300–500 ms time window there were no reliable effects for either agreement violation condition. However, beginning around 500 ms, there was a widely distributed sustained positive deflection in both violation conditions compared to the control condition, with no statistically significant differences in onset or amplitude between the gender and the number violation conditions. Finally, in both violation conditions compared to the control condition there was a sustained negative deflection from 1000 to 1300 ms which was limited to the anterior scalp region and again showed no statistically significant differences in onset or amplitude between the gender and the number violation conditions.

4. Discussion

In this study, we used ERP recording and a grammatical decision task to examine on-line processing of grammatical gender and number violations in Spanish sentences by proficient Chinese late learners of Spanish, to shed light on the question of L1–L2 transfer as well as proficiency and age of acquisition in late L2 learners. The results included a late positivity in both sentence positions for grammatical gender and for number disagreement compared to the agreement condition, an effect which corresponds in latency and distribution to the P600 widely reported for syntactic agreement errors. This was followed by a very late sustained negativity at frontal electrodes for both disagreement conditions.

The P600 has been widely interpreted as a family of effects related to syntactic integration difficulties and/or less domain-specific processes associated to anomaly detection or general monitoring processes (Carreiras, Salillas, & Barber, 2004; Coulson et al., 1998; Friederici, Hahne, et al., 2002; Hagoort, 2003; van Herten, Kolk, & Chwilla, 2005). One interpretation of the P600 commonly reported for this type of morphosyntactic agreement manipulation is that of Friederici's (2002) neurocognitive model of sentence processing. The three stages of this model consist of initial syntactic structure assignment on the basis of word category information (reflected in an early left frontal negativity (ELAN)), followed by morphosyntactic and/or semantic integration processes (LAN and/or N400 effects) and finally a third stage of syntactic reanalysis and/or repair of anomalous sentences, indexed by P600 effects. Other authors (Bornkessel & Schlesewsky, 2006; Kuperberg, 2007) have described the P600 in terms of a late stage of integration of both semantic and syntactic information, as P600 effects have been reported for not only syntactic, but for some types of semantic and indeed discourse anomalies (Ferretti, Singer, & Patterson, 2008). Our study shows that L1 Chinese proficient second language learners of Spanish show a P600 effect in response to violation of a morphosyntactic agreement feature that is not present in their L1 language. Thus, syntactic integration difficulties or and reanalyses processes may be at work in response to morphosyntactic violation in a second language. The P600 response reported in this study has been previously reported in studies of proficient second language learners in both natural languages (Frenck-Mestre, Foucart, Carrasco, & Herschenson, 2009; Hahne et al., 2006; Osterhout et al., 2008; McLaughlin et al., 2010) and in artificial languages (Mueller, Grgisdies, & Friederici, 2008; Mueller, Hahne, Fujii, & Friederici, 2005). However, it is in contrast to previous findings from syntax processing studies of Chinese L1 late learners of English (Chen et al., 2007; Guo et al., 2009). In these studies, although the English native-speaker control groups yielded the classic LAN–P600 pattern in response to violations of subject–verb agreement and P600 effects to violations of verb-subcategorisation, no evidence of P600 effects were found for the Chinese late learners of English.

The difference in our results could be due to the characteristics of the L2s studied: in our case Spanish, a richly and regularly inflected language as opposed to English, which has comparatively little and irregular morphosyntactic inflexion. The greater constraints of Spanish morphosyntax compared to English could contribute to explaining the difference between our results and previous findings with Chinese L1 participants. According to the Competition Model (Bates & MacWhinney, 1987; MacWhinney, 2002; MacWhinney, 2008) the “learnability” of a language feature is related to the strength of the cue validity and reliability of the feature, as well as to the presence or absence of competing cues from the L1. Therefore, according to this view, the greater reliability and validity of the morphosyntactic cues for gender and number agreement in Spanish would result in more successful acquisition of these features than is the case for subject–verb agreement in English, where the reliability and validity of the cues is much lower.

As well as P600 effects, our results also included a very late sustained negativity recorded at frontal electrodes. Late frontal negativities have been reported in previous studies and are generally considered to be related to working memory and/or task demands (Kaan, 2002; Ruchkin, Graffman, Cameron, & Berndt, 2003; Sabourin & Stowe, 2004). In this study, the participants were required to maintain the grammatical decision in memory until the end of the sentence and, as this same effect was reported in our previous study for both native Spanish readers and late English readers of Spanish, we consider that the frontal negativity reflects this effort to retain the grammatical decision in memory until the response.

4.1. L1–L2 transfer effects

One of the motivating questions of this study was to clarify the possible L1 transfer effects suggested by our previous study of English learners of Spanish, where the P600 effects for number agreement violations were “stronger” than those for gender agreement violations in terms of latency and amplitude. In the present study, in contrast to the English L1 group, the results of the Chinese learners of Spanish show no significant differences between the P600 effects for gender and the number violations. This would seem to confirm that the differential effects in the English L1 group were due to L1 feature transfer, as English does have number agreement to some extent for example demonstrative pronoun–noun agreement such as “those dogs”, whereas Chinese has no such morphosyntactic agreement and so neither feature can be transferred. Although there have been few ERP studies specifically examining syntactic transfer effects (Kotz et al., 2008; Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005) these have reported positive transfer effects where syntactic features are comparable in the L1 and L2. This fits our pattern of results from the late English learners of Spanish, where the number feature that exists in English (L1) is processed “earlier and more” as reflected by the latency and amplitude of the P600 response in contrast to that to the grammatical gender feature, which is not computed in English, but which the late learners have acquired at this high level of proficiency. In the case of the Chinese L1 late learners of Spanish, neither feature can be transferred as Chinese does not compute this type of morphosyntactic agreement, but as the level of proficiency is high, both features have been acquired to the same degree, as reflected in the similar P600 response to both types of violation.

One major difference, however, between our previous English L1 results (Gillon Dowens et al., 2010) and the present Chinese L1 findings is the absence of LAN effects in the present study. The pattern for native Spanish readers in the previous study was the classic LAN/P600 for both gender and number in both sentence positions.
The English L1 group, in contrast, only showed LAN effects in the first sentence position and we attributed the lack of LAN effects in the second sentence position to the greater cost of processing the L2 across the phrase boundary, in terms of working memory resources. Again, there were differences in this group between the LAN effects for number, which is the feature already present in the L1, and grammatical gender, which is not represented in English. These differences consisted in earlier onset and longer duration of the LAN for number violations than that for gender, suggesting some differences in the “degree” of processing of these two types of agreement violation.

For the L1 Chinese group, the results could have been expected to again include LAN effects in the first sentence position, but this time without differences between number and gender, as neither feature can be transferred from Chinese. However, this was not the case as only P600 effects were found. LAN effects have been argued to reflect an earlier, possibly more automatic stage of morphosyntactic processing (Friederici, Steinhauser, & Pfeifer, 2002; Friederici, 2001) compared to P600 effects which could reflect a more controlled stage at which different types of information (including semantic) can be taken into account. The lack of LAN effects in this group, therefore, would seem to indicate that this earlier, less controlled stage of morphosyntactic processing of violations is not available to the Chinese late learners. This could possibly be due to characteristics of the participants in terms of L2 exposure, as the English L1 participants were living and had learned Spanish in an immersion environment, while the Chinese participants were in an L1 Chinese environment. It is possible that even at high levels of proficiency as measured by off-line tests, automaticity of processing some language features requires more time and L2 exposure.

Taken together, our results seem to be more compatible with approaches such as the Competition Model and Full Access–Full Transfer approach, which argue for the importance of L1 transfer effects in L2 acquisition but also posit that features not present in the L1 can be acquired at higher stages of proficiency. However, the cognitive manipulation of these features may not be as automatic as in the case of native speakers. This would explain why the Chinese late learners in our study produce only a P600, reflecting a less automatic response.

Acknowledgements

This research was partially supported by grant number AP2003-1787 from the Spanish Ministry of Science and Innovation to the first author and by grant number 30600179 from the National Natural Science Foundation of China to Taomei Guo. The authors would like to thank the teachers and students of the Cervantes Institute, Beijing for their help and participation in the study. This research has been partially supported by Grants PS12008-04009/PSIC and CONSOLIDER-INGENIO2010 CSD2008-00048 from the Spanish Government.

References
