University of Reading

# *Double-number marking matters for both L1 and L2 processing of nonlocal agreement similarly: an ERP investigation*

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Cheng, Y., Cunnings, I. ORCID: https://orcid.org/0000-0002-5318-0186, Miller, D. and Rothman, J. (2022) Double-number marking matters for both L1 and L2 processing of nonlocal agreement similarly: an ERP investigation. Studies in Second Language Acquisition, 44 (5). pp. 1309-1329. ISSN 1470-1545 doi: 10.1017/S0272263121000772 Available at https://centaur.reading.ac.uk/101639/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1017/S0272263121000772

Publisher: Cambridge University Press

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

CrossMark

*Studies in Second Language Acquisition* (2021), 1–21
doi:10.1017/S0272263121000772

# *Research Article*

## DOUBLE-NUMBER MARKING MATTERS FOR BOTH L1 AND L2 PROCESSING OF NONLOCAL AGREEMENT SIMILARLY

*AN ERP INVESTIGATION*

***Yesi Cheng***

*University of Reading, UK*

***Ian Cunnings****

*University of Reading, UK*

***David Miller***

*University of Illinois at Chicago, USA*

***Jason Rothman***

*UiT, The Arctic University of Norway, Norway, and Universidad Nebrija, Spain*

**Abstract**

The present study uses event-related potentials (ERPs) to examine nonlocal agreement processing between native (L1) English speakers and Chinese–English second language (L2) learners, whose L1 lacks number agreement. We manipulated number marking with determiners (*the* vs. *that/these*) to see how determiner-specification influences both native and nonnative processing downstream for verbal number agreement. Behavioral and ERP results suggest both groups detected nonlocal agreement violations, indexed by a P600 effect. Moreover, the manipulation of determiner-number specification revealed a facilitation effect across the board in both grammaticality judgment and ERP responses for both groups: increased judgment accuracy and a larger P600 effect amplitude for sentences containing violations with demonstratives rather than bare determiners. Contrary to some claims regarding the potential for nonnative processing, the present data suggest that L1 and L2 speakers show similar ERP responses when processing agreement, even when the L1 lacks the relevant distinction.

* Corresponding author. E-mail: i.cunnings@reading.ac.uk

*© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.*

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

## INTRODUCTION

Whether adult second language (L2) processing is fundamentally different from native language (L1) processing remains a topic of debate. A broad range of theories, such as the Failed Functional Feature Hypothesis (Hawkins & Chan, 1997; Hawkins & Liszka, 2003) and Shallow Structure Hypothesis (Clahsen & Felser, 2006, 2018), predict qualitative differences between L1 and L2 acquisition and processing, such that L2 speakers are predicted to have difficulty when acquiring a novel L2 grammatical feature absent in their L1 grammar or underutilize syntactic information during real-time processing. Alternatively, other approaches maintain that L2 speakers can acquire novel L2 features or construct syntactic representations in a similar way to L1 speakers (e.g., Cunnings, 2017; Hopp, 2014; Kaan, 2014; Lardiere, 2009; Lim & Christianson, 2015; McDonald, 2006; Schwartz & Sprouse, 1996). In sufficiently proficient L2 speakers, when quantitative differences between L1 and L2 arise, such theories predict that differences might be best explained by issues pertaining to task effects (e.g., Lim & Christianson, 2015), difficulties in anticipating upcoming information (Grüter et al., 2014) or with lexical processing efficiency (e.g., Hopp, 2014), memory retrieval interference (e.g., Cunnings, 2017), or participant-level individual differences (e.g., Lempert, 2016; Sagarra & Herchensohn, 2010; Tanner et al., 2014).

Situated within the preceding discussion, a growing body of research has investigated processing of subject-verb agreement as in (1) and (2) during comprehension in L2 populations with an L1 that does not license overt agreement (e.g., Chinese, Japanese, Korean) to inform these long-standing theoretical debates in L2 acquisition and processing (e.g., Clahsen & Felser, 2006; Cunnings, 2017; Hawkins & Chan, 1997; Schwartz & Sprouse, 1996).

(1) *Turtles moves slowly. (local agreement violation)
(2) *The key to the cabinet/s are rusty. (nonlocal agreement violation)

In English, present-tense verbs agree with their subjects in number and person, which is seen reliably in the third-person singular -s marking or with greater distinction in the copula *to be*. Both (1) and (2) contain an agreement violation as the subject does not match the verb in number. Contexts like (1) are local as the sentence subject and verb are directly adjacent to each other. However, contexts as in (2) are more complicated as they contain a nonlocal linguistic dependency where an intervening noun phrase (intervening NP; "the cabinets") is embedded between the subject "the key" and the verb "are." Research has tested how L2 speakers whose L1 does not have agreement process local and nonlocal agreement (violations) in juxtaposition to L1 speakers during real-time comprehension, using a variety of techniques including event-related potentials (ERPs), although results have been mixed (e.g., Armstrong et al., 2018; Chen et al., 2007; Jiang, 2004; Lim & Christianson, 2015; Ojima et al., 2005).

Recent ERP research has also examined how double-number marking on a subject NP ("Many cookies"), as in (3b), influences sensitivity to agreement violations in comparison to sentences like (3a), without additional marking ("The cookies") in both L1 and L2 speakers (Armstrong et al., 2018; Tanner & Bulkes, 2015). To date, however, how double-number marking influences agreement processing has been tested in local

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

agreement contexts only, and how double-number marking regulates processing of nonlocal number agreement violations is yet to be explored.

(3a) *The cookies tastes the best when dipped in milk.
(3b) *Many cookies tastes the best when dipped in milk.

Extending previous research, we aim to gauge: (a) ERP responses to nonlocal agreement violations like (2) in English L1 speakers and Chinese speakers of L2 English in an immersion setting and (b) how double-number marking influences nonlocal agreement violation processing in L1 and L2 comprehension. Findings will weigh in on different L1/L2 processing accounts, providing evidence to inform theoretical debates regarding potential L2 computational capacity for nonlocal linguistic dependencies and whether linguistic features absent in an L2 speaker's L1 are indeed particularly problematic.

## *AGREEMENT PROCESSING IN L1*

A significant amount of L1 research on language comprehension has tested processing of local and nonlocal agreement violations using behavioral and ERP experimentations (e.g., Alemán Bañón & Rothman, 2019; Dillon et al., 2013; Osterhout & Mobley, 1995; Pearlmutter et al., 1999; Shen et al., 2013; Tanner, 2011; Tanner & Bulkes, 2015; Tanner et al., 2017; Tanner et al., 2014; Tanner et al., 2012; Wagers et al., 2009). It has been consistently shown that agreement violations are detected by L1 speakers, reflected by either longer reading times or a P600 effect for sentences containing agreement violations like (1) and (2) compared to grammatical controls (e.g., Chen et al., 2007; Dillon et al., 2013; Lim & Christianson, 2015; Osterhout & Mobley, 1995; Osterhout et al., 1996; Tanner et al., 2012; Tanner & Bulkes, 2015), although other ERP responses have also been reported in some studies (e.g., Osterhout, 1997; Tanner, 2019; Tanner et al., 2018; Tanner & Van Hell, 2014). With nonlocal violations like (2), English natives show shorter reading times or reduced P600 effects when the intervening NP matches the number properties of the verb compared to when it does not (e.g., Dillon et al., 2013; Lim & Christianson, 2015; Pearlmutter et al., 1999; Shen et al., 2013; Tanner et al., 2012; Tanner et al., 2017; Tanner, Nicol et al., 2014). This is hypothesized to occur due to a so-called attraction effect: matching agreement between the intervening NP and verb in number, even though the head noun of the complex NP is the actual agreement controller (e.g., Dillon et al., 2013; Lim & Christianson, 2015; Pearlmutter et al., 1999; Shen et al., 2013; Tanner et al., 2012; Tanner et al., 2017).

How number is marked can also influence agreement violations in L1 processing. In English, number can be marked morphologically (e.g., "cookies") and also lexically, with words such as "many," as in "many cookies." With a quantifier like "many," which indicates plurality, the following nominal head must also be overtly marked by the plural marker "-s" ("cookies"). This combination forms a case of double-number marking. Tanner and Bulkes (2015) manipulated this factor in a design that tested sentences like (4) to investigate whether double marking using quantifiers, as in (4c/d), would facilitate perception of local agreement violations compared to cases like (4a/b), without double marking.

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

(4a) The cookies taste the best when dipped in milk. (Grammatical, Unquantified)
(4b) *The cookies tastes the best when dipped in milk. (Ungrammatical, Unquantified)
(4c) Many cookies taste the best when dipped in milk. (Grammatical, Quantified)
(4d) *Many cookies tastes the best when dipped in milk. (Ungrammatical, Quantified)

Indeed, they found the difference in the amplitude of the P600 effect was larger between (4c) and (4d), where the plural subject NP was preceded by a number-marked quantifier, compared to between (4a) and (4b), where it was preceded by a number-unspecified determiner. This suggests that double-number marking from the quantifier makes agreement errors more salient in L1 processing.

### *AGREEMENT PROCESSING IN L2*

While mixed findings have been reported in L2 processing, most research on L1-L2 pairs of typologically similar languages reveals that L2 speakers can be nativelike when processing local and nonlocal agreement violations (e.g., Alemán Bañón et al., 2017; Frenck-Mestre et al., 2008; Sagarra & Herchensohn, 2010; Tanner et al., 2013; Tanner, Inoue et al., 2014). Conversely, with language pairs where morphological agreement is not present in the L1, the picture is less clear, with some prior behavioral studies suggesting similarity to L1 speakers and others differences (e.g., Jiang, 2004; Jiang et al., 2011; Lempert, 2016; Lim & Christianson, 2015). The existing evidence from ERP studies, both in and outside of a native-English immersion context, also paints an ambiguous picture, with studies reporting either similar or different neural responses to relevant agreement violations from L2 speakers of these languages compared to L1 speakers' (e.g., Armstrong et al., 2018; Chen et al., 2007; Ojima et al., 2005).

Cross-study divergence seems, at least in part, to be driven by methodological differences (e.g., materials, whether participants were tested in an immersion setting or not). Processing of local agreement violations like (1) was tested in Japanese speakers of English (Ojima et al., 2005) and Chinese speakers of English (Armstrong et al., 2018) and contradictory results were found. Ojima et al (2005) found that whilst L1 English controls demonstrated both a left-lateralized negativity and a P600 component at the verb for sentences containing local violations (e.g., "Turtles move slowly" vs. "*Turtles moves slowly"), the highest proficiency L2 individuals tested who were living in Japan only showed the left-lateralized negativity and absence of the P600 effect. Hence, they claimed qualitative differences between L1 and L2 processing. Conversely, Armstrong et al. (2018) showed that both English controls and Chinese L2-English participants living in the United States exhibited a P600 effect to local violations (e.g., "The cookies taste…" vs. "*The cookies tastes.."), demonstrating that L2 speakers whose L1 lacks the relevant morphological agreement, at least under certain conditions such as in immersion, can demonstrate nativelike neural responses to agreement violations.

Armstrong et al. (2018) employed the design and materials as in (4) from Tanner and Bulkes (2015), testing whether a stronger response to local agreement violations would be evoked following double-number marking in Chinese speakers of L2 English. Recall that unlike English, number is not morphologically marked on nouns in Chinese but can be

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

marked on determiners using quantifiers (e.g., "Many cookie") and demonstratives (e.g., "Those cookie"). As such, double-number marking is not possible in Chinese. The results found that, unlike English L1 speakers who showed an enhanced P600 effect for violations following double marking, L2 participants showed a reduced P600 effect for (4c)–(4d) relative to (4a)–(4b), suggesting that double marking decreased sensitivity to local violations. Armstrong et al. proposed that the L2 speakers' failure to utilize double-number marking in a nativelike way was due to an L1 processing strategy that arises from the overlap in quantification between Chinese and English. Specifically, Armstrong et al. argued that once the Chinese L2 speakers parsed the number marked quantifier "many," the way number marking happens in their L1, they paid less attention to the morpho-syntactic cues on the noun. However, as they only tested one group of L2 speakers, Armstrong et al. acknowledged this could also be a general L2 processing strategy. Nevertheless, there is also a potential confound in their materials, as some of the quantifiers used (e.g., "some") are number-ambiguous. Given that "some" can also occur with singular nouns (e.g., "Some bread is on the table"), this could have contributed to the L2 speakers' apparent reduced sensitivity.

As to nonlocal agreement, Chen et al. (2007) tested a group of Chinese speakers of English in China using a design as in (5), which manipulated sentence grammaticality and the number properties of the intervening noun (car/s).

(5a). The price of the car was too high. (Grammatical, Singular Intervening Noun)
(5b). The price of the cars was too high. (Grammatical, Plural Intervening Noun)
(5c). *The price of the car were too high. (Ungrammatical, Singular Intervening Noun)
(5d). *The price of the cars were too high. (Ungrammatical, Plural Intervening Noun)

For grammatical sentences, a P600 effect was elicited for (5b), where the intervening NP does not match the verb in number, compared to (5a), where it does, in the Chinese speakers of English. Chen et al. interpreted this as indicating that the L2 speakers focused on the incongruency between the local noun and verb. For ungrammatical sentences, even though the L2 speakers detected nonlocal agreement violations in both (5c) and (5d), irrespective of the intervening NP's number, they showed a distinct neural response, a late negative shift, from the L1 speakers who showed a P600 component. Hence, Chen et al. (2007) sustained the claims of Ojima et al. (2005), concluding the neural underpinnings of L2 processing are qualitatively different from L1 processing when the processed features are absent in the L1. However, the Chen et al. (2007) study, like Ojima, was conducted outside of an immersion setting. Whether immersed L2 speakers' neural responses to nonlocal violations can be nativelike and how double marking regulates nonlocal agreement processing has not been examined. The present study aims to address these questions.

### *THE PRESENT STUDY*

In summary, to our knowledge, no existing published studies have used ERPs to examine the processing of nonlocal agreement in Chinese speakers of English in an immersion setting where both quantity and quality of native input exposure is increased. Related work has shown an association between naturalistic or immersion-like L2 exposure and

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

nativelike grammatical processing (e.g., Dussias, 2003; Morgan-Short et al., 2010; Morgan-Short et al., 2012; Pliatsikas & Marinis, 2013). Thus, by testing the same domain of grammar, nonlocal agreement, as in Chen et al. (2007) in the context of immersion, we will be able to test for further evidence of this inference. Whilst prior research has tested local agreement and double-number marking in immersion, no existing published studies have examined the case of nonlocal agreement and double marking. Also, the relevant previous research has examined double marking using quantifiers, some of which can be number-ambiguous (e.g., *Some* bread is on the table; *Some* breads are made of corn). Thus, the present study employed demonstratives (e.g., these, those) that more clearly mark number unambiguously. Moreover, by testing cases of double-number marking using demonstratives, as opposed to quantifiers as in Tanner and Bulkes (2015) and Armstrong et al. (2018), the data will shed light on whether the previously reported effect was from quantification or more generalizable to all instances of double marking.

Therefore, our study fills a number of gaps in the literature with three interrelated goals: (a) reconciling some of the inconsistencies found across the preceding reviewed studies, (b) understanding more specifically what role lexical and morphological cues play in L2 parsing through examining the case of double marking, and (c) interpreting what our results can add to debates within L2 acquisition and processing more generally. With this in mind, we addressed the following research questions:

(1) Will a P600 effect be elicited by nonlocal violations in Chinese speakers of English living in an immersion setting where increased exposure to native English is afforded?
(2) Does double-number marking from determiner-number specification using demonstratives elicit a larger or smaller P600 effect in English L1 speakers and Chinese speakers of L2 English?

Based on previous L1/L2 findings, the following hypotheses are proposed. In this immersion setting, if nativelike processing is attainable, Chinese speakers of L2 English and native L1 speakers alike should exhibit a P600 effect to agreement violations in nonlocal dependencies. Furthermore, double marking should enhance neural sensitivity to nonlocal violations in L1 speakers, leading to a larger P600 effect for violations following double marking (Tanner & Bulkes, 2015). If we replicate Armstrong et al.'s (2018) results for L2 speakers, we should observe a reduced sensitivity to nonlocal violations following double marking in Chinese speakers of English, that is, a smaller P600 effect. Alternatively, if L2 speakers are able to utilize double-number marking like L1 speakers, the P600 effect should be larger in cases of double marking in both groups.

## METHOD

### *PARTICIPANTS*

The experiment was conducted in an English immersion setting with 32 English L1 speakers (mean age = 21.4) and 32 Chinese–English L2 speakers who learned English in school settings in China (mean age = 25.3). All participants were recruited from the University of Reading and were enrolled in either an undergraduate or postgraduate course. They received a small payment or course credit upon completion of the study. The L2 speakers were born and raised in China and came to the United Kingdom for higher

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

education. They were living in the United Kingdom at the time of testing and reported their lengths of immersion experience, which ranged from 2 to 48 months (mean = 17.7 months, SD = 13.18). Their English proficiency was measured by a short version of the Oxford Quick Placement Test (Oxford University Press, 2004). The proficiency scores ranged from 24–54 out of 60 (mean = 40, SD = 7.87). All participants were right-handed and had normal or corrected to normal vision.

### *MATERIALS*

We recorded EEG with ERP time locking concurrent with a grammaticality judgment task (GJT) to test participants' online processing and comprehension of nonlocal subject-verb agreement. Following that, we also administered a whole sentence GJT, which was slightly different from the EEG concurrent GJT in terms of stimuli presentation, as described in the following text. For the EEG task, 160 critical items like (6) were created, with four target conditions (40 trials per condition) that were distributed across four separate lists so that participants only saw one condition of each item. Each experimental sentence contained a critical verb (either "is" or "has") and manipulated sentence grammaticality (grammatical vs. ungrammatical). The subject was either singular or plural such that half the sentences were grammatical, as in (6a/c), and half were ungrammatical, as in (6b/d). The intervening noun was always singular so that it matched the number properties of the verb. Number specification on the determiner (number-specified vs. number-unspecified) was also manipulated using demonstratives. Conditions (6a/b) had a number-unspecified determiner ("The") while conditions (6c/d) had a demonstrative that specified number ("This/These" or "That/Those"). Across items, these two sets of demonstratives were used an equal number of times. Within the critical sentences, half had "is" as the verb and half had "has." Another 160 fillers were created with half being grammatical and half being ungrammatical. Some of the fillers contained a similar structure to the critical items but had a plural verb (i.e., The biscuits on the table are tasty.) to minimize the possibility of participants expecting that the verb would always be singular given that all critical items contained singular verbs. All the sentences were displayed word by word.

(6a) The window of the house is really clean. (Grammatical, Number-Unspecified [NU])
(6b) The windows of the house is really clean. (Ungrammatical, Number-Unspecified [NU])
(6c) That window of the house is really clean. (Grammatical, Number-Specified [NS])
(6d) Those windows of the house is really clean. (Ungrammatical, Number-Specified [NS])

The whole sentence GJT task consisted of a different set of 24 experimental items, that manipulated the same four conditions as in (6), and 30 fillers, using a slightly different procedure from the EEG concurrent GJT. Instead of showing one word at a time, a whole sentence was presented at once, during which the participants made their response. The items were pseudorandomized in a Latin-square design so that each participant saw a different list. Participants only saw one condition of each item and, therefore, read six sentences for each condition. Participants were asked to judge whether the sentence they

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

read was grammatical or not by pressing 1 (grammatical) and 2 (ungrammatical) on the keyboard. Correct answers were coded as 1 and incorrect answers were coded as 0. As such, a value closer to 1 indicates higher accuracy. The materials for the EEG and whole sentence GJT experiments can be found in the Online Supplementary Materials.

### *PROCEDURE*

The study was conducted in one session. All participants were first asked to provide information on their language experiences by completing a participant form, followed by the main EEG experiment presented in rapid serial visual presentation (RSVP) while participants' EEG activity was recorded. Participants were told to read as naturally as possible and to make sure they understood the sentences. Before each sentence, a fixation marker appeared in the middle of the screen. Following that, the words of each sentence were displayed one at a time for 450 ms with interstimulus intervals of 200 ms. After each sentence, a happy face and a sad face that represented "grammatical" and "ungrammatical," respectively, appeared onscreen. Even though it was untimed, participants were asked to make a judgment as quickly and accurately as possible about whether the sentence they read was grammatical or not using the mouse clicking with their right hands. After that, a 1000 ms blank screen appeared before the presentation of the next sentence. Participants familiarized themselves with the procedure by first completing some practice trials before the experiment. After the EEG task, all participants completed the whole sentence GJT. Finally, the L2 speakers completed the proficiency test.

### *DATA ACQUISITION AND ANALYSIS*

The EEG activity was recorded by a 64-channel active cap system using Brain Vision Recorder and a BrainAmpDC amplifier system (Brain Products, Germany). Eye movements were monitored with Fp1 and Fp2. The data were recorded with a reference to FCz and rereferenced offline to the average of the mastoids. Impedances were maintained below 5 Ω for all channels. The EEG signals were digitized at a sampling rate of 1000 Hz with a bandpass filter of 0.016 to 200 Hz. Data preprocessing was done by Brain Vision Analyzer (Brain Products, Germany). The data were filtered offline at 0.1–30 Hz. Epochs of 1500 ms were segmented around the critical verb with 300 ms before the onset of the critical stimulus and 1200 ms postonset. The baseline (300 ms prestimulus) was corrected for all epochs. Using similar parameters found in Spychalska et al. (2016), semiautomatic artifact rejection was applied to help spot any trials with the absolute amplitude difference more than 200 mV /200 ms, or with the amplitude lower than –130 mV or higher than 130 mV, or with the activity lower than 0.5 mV in intervals of 100 ms, or with a voltage step higher than 50 mV/ms. Trials with blinks, eye movements, excessive amplifier drift, or noisy electrodes were removed, which kept at least 63% of the trials in any of the four experimental conditions for each participant in the L1 group and 75% in the L2 group. After the preprocessing procedure, 7% and 5% of the total data were excluded in the L1 and L2 groups, respectively, prior to averaging and grand averaging. ERPs were time-locked to the onset of the critical verb and averaged offline for each condition at each electrode for each participant. For each participant, mean amplitudes were computed in the 500–1000 ms poststimulus window that covers the P600 time window.

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

Nine regions of interest were included in the analysis: left anterior (F1, F3, F5, FC3, FC5, FC1), left medial (C1, C3, C5, CP1, CP3, CP5), left posterior (P1, P3, P5, P7, PO3, PO7), midline anterior (Fz and FCz), midline medial (Cz and CPz), midline posterior (Pz and POz), right anterior (F2, F4, F6, FC2, FC4, FC6), right medial (C2, C4, C6, CP2, CP4, CP6), and right posterior (P2, P4, P6, P8, PO4, PO8). Similar to some previous ERP studies (e.g., Armstrong et al., 2018; Miller & Rothman, 2020; Tanner & Bulkes, 2015), repeated measures ANOVAs were conducted separately for the midline and lateral electrode sites due to the different numbers of electrodes these sites had, with Group (L1 and L2) as a between-subject variable, Grammaticality (grammatical and ungrammatical), Number Specification (number-specified and number-unspecified), Caudality (anterior, medial and posterior), and Hemisphere (left and right [only for lateral analysis]) as within-subject variables. Following Armstrong et al. (2018), we only report effects and interactions relevant to Grammaticality and Number Specification effects. For any main effects and interactions involving a variable with more than two levels (caudality), we report the results based on the Mauchly's test for sphericity and sphericity corrections. Post hoc analyses were conducted for any further interactions.

The GJT data from the EEG recording were analyzed using generalized (binomial) mixed-effects logistic regression (Jaeger, 2008). A generalized mixed model was conducted including sum coded (–1/1) fixed effects of Group, Grammaticality, and Number Specification and their interactions. One Chinese participant was removed due to the loss of data. The whole sentence GJT data were analyzed using the same methods. The maximal models were computed and fit using the maximal random effects model that converged (Barr, 2013; Barr et al., 2013). Random intercepts for subjects and items were included. By-subject random slopes included grammaticality*number specification and by-item random slopes included group*grammaticality*number specification. When the maximal model failed to converge, we refitted the model by first removing the random correlation parameters. If the model still failed to converge, the random effect that accounted for the least variance was iteratively removed until convergence was achieved.

## RESULTS

### *WHOLE SENTENCE GJT*

The descriptive results from the four conditions in the whole sentence GJT are shown in Table 1. The overall judgment score across all four experimental conditions was 0.91 in the L1 group (range = 0.63 to 1, SD = 0.1) and 0.92 in the L2 group (range = 0.58 to 1, SD = 0.09). The statistical results (all estimates are in logits) revealed neither main effects of Group nor any interactions by Group (all $z < 1.02$, $p > .3$). The main effect of Grammaticality and Grammaticality by Number Specification interaction were significant (Grammaticality: estimate = –0.55, SE = 0.18, z = –3.13, $p = .002$; Grammaticality by Number Specification: estimate = –0.3, SE = 0.13, z = –2.4, $p = .02$). The Grammaticality effect showed both groups made more incorrect judgments for the ungrammatical sentences relative to the grammatical ones. For the two-way interaction, follow-up analyses indicated that while there was no difference between the two grammatical conditions (estimate = 0.22, SE = 0.28, z = 0.79, $p = .428$), participants made

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

TABLE 1. Judgment accuracy in the whole sentence GJT (standard errors in parentheses)

| | L1 Speakers | L2 Speakers |
|---|---|---|
| Grammatical, NU | 0.97 (0.01) | 0.96 (0.01) |
| Ungrammatical, NU | 0.82 (0.03) | 0.87 (0.02) |
| Grammatical, NS | 0.95 (0.02) | 0.94 (0.02) |
| Ungrammatical, NS | 0.90 (0.02) | 0.92 (0.02) |

significantly more correct judgments on the ungrammatical sentences with a number-specified determiner than those with a number-unspecified determiner (estimate = –0.36, SE = 0.13, z = –2.81, $p$ = .005). Also, ungrammatical sentences were judged significantly more poorly than the grammatical ones only when the sentences had a number-unspecified determiner (estimate = –0.86, SE = 0.21, z = –4.14, $p$ < .001), but not for number-specified determiner sentences (estimate = –0.29, SE = 0.25, z = –1.16, $p$ = .25).

### *GJT DURING EEG*

The descriptive results from the four experimental conditions in the EEG concurrent GJT are shown in Table 2. The overall score across all the experimental conditions was 0.86 (range = 0.53 to 0.98, SD = 0.09) in the L1 group and 0.85 (range = 0.51 to 0.99, SD = 0.1) in the L2 group. The results suggest no main effects of Group or any interactions by Group (all z < –1.5, $p$ > .13). There was a significant main effect of Number Specification (estimate = –0.11, SE = 0.03, z = –3.13, $p$ = .002), which was qualified by a significant Grammaticality by Number Specification interaction (estimate = –0.21, SE = 0.03, z = –6.44, $p$ < .001). The follow-up analyses showed within the ungrammatical conditions, number-specified determiners elicited more correct judgments than the number-unspecified ones (estimate = –0.31, SE = 0.05, z = –6.57, $p$ < .001). Also, within the grammatical conditions, the number-unspecified determiners elicited more correct judgments than the number-specified ones (estimate = 0.11, SE = 0.05, z = 2.23, $p$ = .03). Additionally, both groups made better judgments on the grammatical sentences compared to the ungrammatical counterparts for sentences with a number-unspecified determiner (estimate = –0.41, SE = 0.12, z = –3.35, $p$ < .001) but did not exhibit such difference for sentences with a number-specified determiner (estimate = –0.002, SE = 0.12, z = –0.01, $p$ = .989).

TABLE 2. Judgment accuracy in the EEG GJT (standard errors in parentheses)

| | L1 Speakers | L2 Speakers |
|---|---|---|
| Grammatical, NU | 0.89 (0.01) | 0.90 (0.01) |
| Ungrammatical, NU | 0.83 (0.01) | 0.76 (0.01) |
| Grammatical, NS | 0.86 (0.01) | 0.90 (0.01) |
| Ungrammatical, NS | 0.88 (0.01) | 0.84 (0.01) |

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

***ERPs***

Figure 1 illustrates the voltage deflections elicited by (6a–d) at 19 electrodes in both groups.[1]

***LATERAL ANALYSIS RESULTS***

The ANOVA results for the mean voltage measured along the lateral electrodes during the 500–1000 ms time window indicated a significant main effect of Grammaticality showing the ERP responses were more positive for the ungrammatical sentences than the grammatical ones ($F(1, 62) = 29.72$, $p < .001$), which reflects a P600 effect. The Group by Number Specification interaction was significant ($F(1, 62) = 4.39$, $p = .04$). Follow-up tests showed the voltage was more positive for sentences with a number-unspecified determiner than those with a number-specified determiner in the L1 speakers ($t = 3.59$, $p < .001$), but the opposite in the L2 speakers ($t = –7.23$, $p < .001$). The Grammaticality by Caudality interaction was also significant ($F(2, 124) = 56.96$, $p < .001$). Follow-up t-tests demonstrated that the brain responses elicited by the ungrammatical sentences were more positive than the grammatical ones in both medial ($t = 17.05$, $p < .001$) and posterior ($t = 28.31$, $p < .001$) areas but less positive than the grammatical sentences in the anterior region ($t = –4.99$, $p < .001$).

***MIDLINE ANALYSIS RESULTS***

Regarding the results over the midline electrodes during the 500–1000 ms time window, the main effect of Grammaticality ($F(1, 62) = 34.11$, $p < .001$) indicated the ungrammatical sentences elicited more positive-going brain responses than the grammatical ones, which corresponds to the P600 effect. Also, the Grammaticality by Number Specification interaction was significant ($F(1, 62) = 6.71$, $p = .012$). Follow-up t-tests revealed that while both sentences with a number-unspecified determiner and with a number-specified determiner clearly demonstrated the a grammaticality effect (number-unspecified: $t = 5$, $p < .001$; number-specified: $t = 9.99$, $p < .001$), the brain responses to the ungrammatical sentences with a number-specified determiner were more positive than those with a number-unspecified one ($t = 4.84$, $p < .001$) whereas no differences were observed between the grammatical sentences with a number-unspecified determiner and with a number specified one ($t = –0.66$, $p = .512$). This suggests a larger P600 effect elicited by double-number marking in both groups, which is visualized in Figure 1 and Figure 2. Furthermore, the three-way Group by Grammaticality by Caudality interaction was also significant ($F(2, 124) = 3.63$, $p = .038$). Follow-up analyses suggested the two groups differed in terms of the Grammaticality effect in the posterior region ($F(1, 62) = 8.57$, $p = .005$). As shown in Figure 3, although both groups exhibited the P600 effect (L1: $t = 13.07$, $p < .001$; L2: $t = 8$, $p < .001$), the voltage of the ungrammatical sentences was more positive in the L1 speakers in comparison to in the L2 speakers ($t = 3.71$, $p < .001$) when there was no between-group difference regarding the grammatical sentences ($t = –0.75$, $p = .456$). This indicates a larger P600 effect in the posterior area in the L1 group due to its longer duration than in the L2 group, as can be

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

seen in Figure 1. In addition, we found a Group effect in the anterior area ($F(1, 62) = 4.9$, $p = .03$), showing a significant difference between the two groups in terms of voltage polarity across grammaticality, with positive-going brain responses in the L1 speakers and negative-going responses in the L2 speakers, as displayed in Figure 3. However, as this effect did not interact with grammaticality, we do not discuss it further.

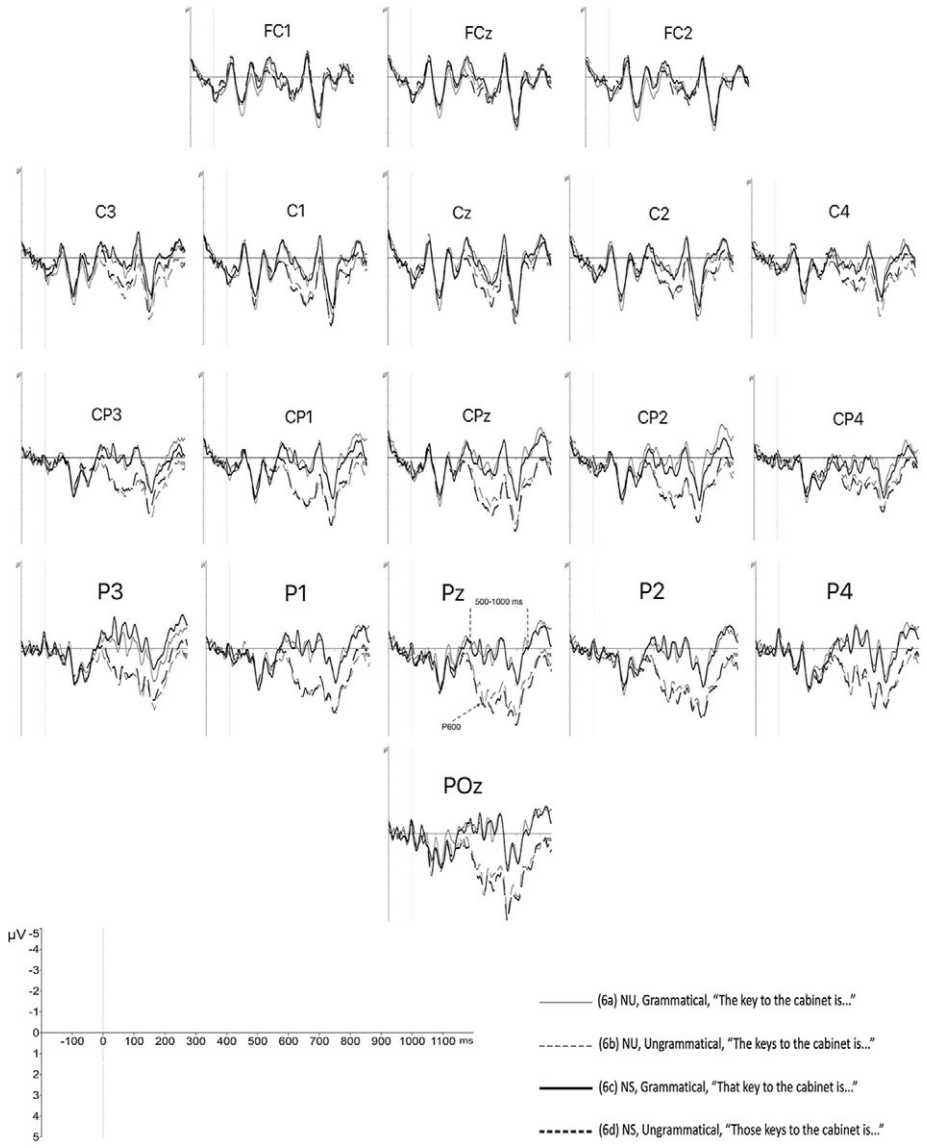


FIGURE 1A. Grand average waveforms for (6a–d) at 19 electrodes in the L1 group (time window analyzed was 500–1000 ms).

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

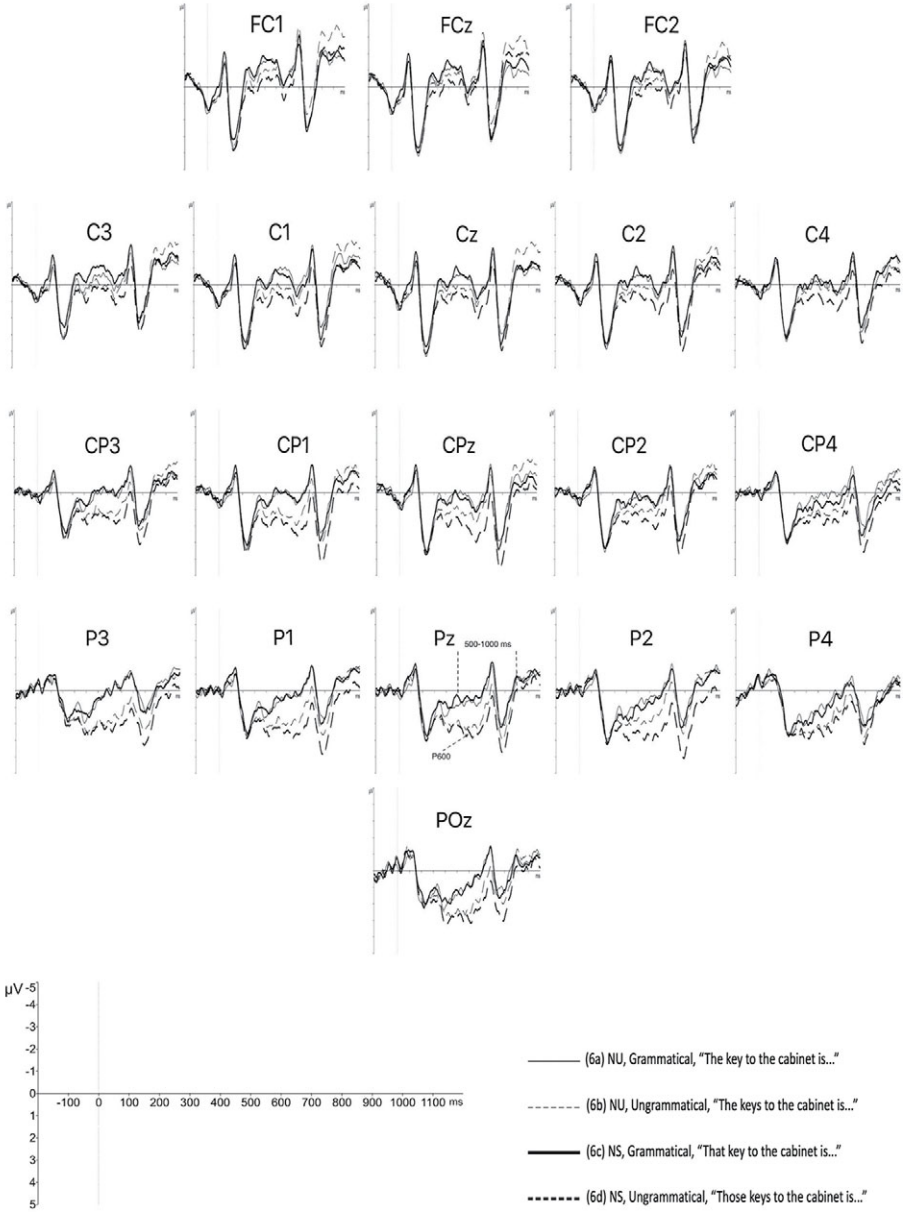


FIGURE 1B. Grand average waveforms for (6a–d) at 19 electrodes in the L2 group (time window analyzed was 500–1000 ms).

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

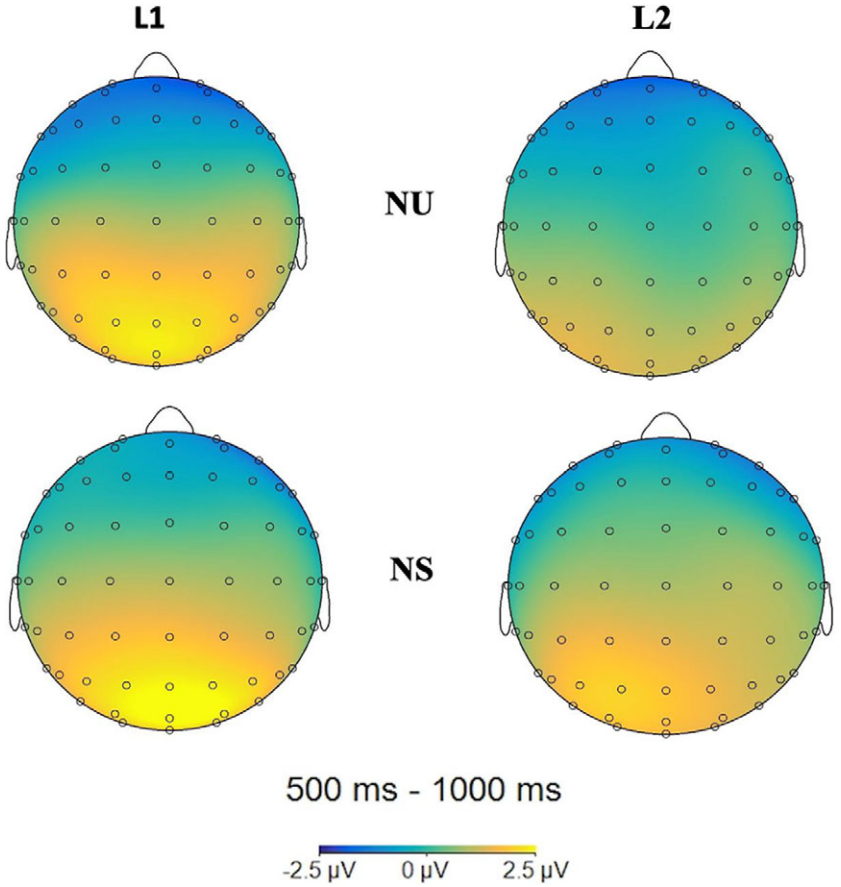


FIGURE 2. Topographic distribution of the P600 effects (ungrammatical minus grammatical difference) observed in the number unspecified (NU) conditions (6b–6a) and number specified (NS) conditions (6d–6c) during the 500–1000 ms window in the L1 and L2 group.

## GENERAL DISCUSSION

This study examined processing of nonlocal agreement violations in English L1 speakers and Chinese L2 speakers and tested whether this process was influenced by double marking from determiner-number specification. The results indicated that despite a relatively poorer judgment performance on the sentences containing violations, both L1 and L2 groups showed in general high accuracy to nonlocal agreement in the whole sentence and EEG GJTs. The EEG data indicated both groups exhibited a P600 during processing of nonlocal agreement violations during incremental comprehension. Also, the effect of determiner-number specification on detection of violations was attested in both behavioral and neurophysiological measures. We discuss our behavioral and EEG results, along with their implications for theories of L2 sentence processing, in turn in the following text.

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

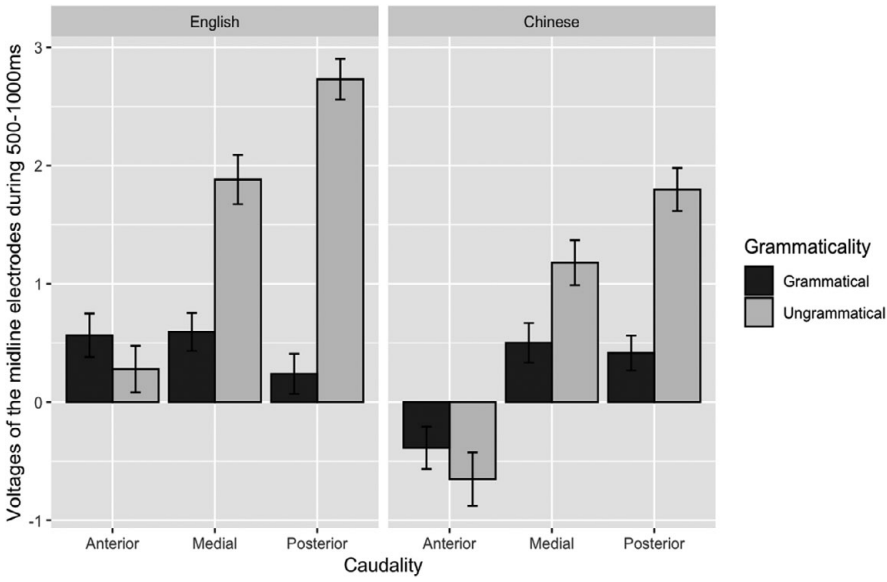


FIGURE 3. The Grammaticality effect in each region along the midline electrodes in each group.

### *WHOLE SENTENCE AND EEG GJTS*

Consistent with previous findings (e.g., Armstrong et al., 2018; Chen et al., 2007), the behavioral results indicated both L1 and L2 speakers were equally able to detect syntactic errors in sentences containing nonlocal agreement violations. However, both groups made more incorrect judgments on ungrammatical sentences than grammatical ones in the whole sentence GJT. The judgment errors here could be either due to response bias that favors grammatical responses (e.g., Hammerly et al., 2019; Tanner & Bulkes, 2015) or attraction from the number match between the intervening NP and verb (e.g., Dillon et al., 2013; Pearlmutter; et al., 1999; Shen et al., 2013). Although attraction is not typically found, or is reduced, in contexts where the intervening NP and verb are singular while the subject is plural (e.g., Bock & Miller, 1991), we do not rule out this possibility. However, it was not our aim to tease apart this issue and our study cannot distinguish between these accounts as we neither manipulated the number of the intervening NP nor neutralized the response bias. Regardless and important for our research questions, our results showed L2 speakers did not significantly differ from L1 speakers in this regard.

The findings also showed that double-number marking led to greater accuracy for sentences containing nonlocal agreement violations in both groups, which was attested in both whole sentence and EEG GJTs. Even though ungrammatical sentences were generally more poorly judged than the grammatical ones, this difference disappeared in sentences with a number-specified determiner, suggesting determiner-number specification facilitates detection of nonlocal violations. It could be that the number representation of the subject NP becomes more salient because of double-number marking and hence number violations more noticeable. Therefore, these findings suggested double-number marking from determiner-number specification increases sensitivity to nonlocal number

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

violations. This effect is not limited to quantification, the domain tested in Tanner and Bulkes (2015), but indeed to demonstratives as well. Additionally, in the EEG GJT, judgment accuracy for grammatical sentences with a demonstrative determiner was slightly lower than those with a number-unspecified determiner. This might be because grammatical sentences were judged to be more felicitous when there was a bare determiner compared to a demonstrative. However, it is also possible that this difference is spurious as it was not found in the whole sentence GJT or EEG data.

In summary, L1 and L2 speakers were sensitive to number violations in nonlocal agreement in the two judgment tasks. Number cues from determiner-number specification were similarly processed by L1 and L2 speakers as double-number marking facilitated detection of nonlocal agreement violations in both groups.

### *ERP EFFECTS IN L1 AND L2*

The ERP results during the 500–1000 ms time window from both lateral and midline electrodes showed a typical P600 effect elicited by sentences containing nonlocal agreement violations irrespective of number specification in the L1 and L2 groups, suggesting both L1 and L2 speakers detected the nonlocal violations during incremental processing. Also, the P600 effect was mainly distributed in the medial and posterior areas of the scalp across the board, which confirms that the P600 effect is largely displayed in the centro-parietal region. However, some between-group differences were also observed, as the posterior P600 effect was larger in the L1 speakers than the L2 speakers.

Our findings are consistent with previous L1 literature (e.g., Osterhout & Mobley, 1995; Osterhout et al., 1996; Tanner et al., 2012; Tanner & Bulkes, 2015) and some existing L2 studies (e.g., Armstrong et al., 2018; Lim & Christianson, 2015) that suggest L2 processing of nonlocal dependencies is not fundamentally different from L1 processing in an immersion setting, even when it comes to processing of a linguistic feature absent in the L1. In comparison to L2 studies that indicated agreement computation in non-immersed L2 speakers whose L1 does not have subject-verb agreement is qualitatively different from that in L1 speakers (e.g., Chen et al., 2007; Ojima et al., 2005), our results provided further neurocognitive evidence demonstrating that neural responses to nonlocal agreement computation in Chinese speakers of English are not destined to remain distinct from L1 speakers', at least when the L2 speakers have ample experience in a native immersion setting. Therefore, our findings suggest it is likely that immersive input is at least partially deterministic in explaining differences between studies conducted in an immersion setting and those with L2 speakers who lack this relevant experience. In other words, the boost in quality input and opportunity to use the L2 that immersion provides could be responsible for the neurocognitive substrates underlying nativelike grammatical processing in our L2 learners in juxtaposition to other similar studies reviewed herein. As such, as opposed to claims that L2 speakers cannot acquire features, in this case number, that are not instantiated in their L1 (Hawkins & Chan, 1997), our findings suggest that it is possible for Chinese speakers of English to process nonlocal linguistic dependencies similarly to L1 speakers, even when the relevant feature is not realized in their L1.

Despite the L1/L2 similarities discussed in the preceding text, it was not the case that our study provides evidence that L1 and L2 processing are exactly the same. Within the

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

500–1000 ms time window, the posterior P600 effect was longer in the L1 group than in the L2 group, as shown in Figure 1. The P600 effect extended beyond 1000 ms in the L1 group but ended around 800 ms in the L2 group.[2] We argue that the fact that both groups reliably showed the P600 effect in the same time window with no significant distributional differences indicates quantitative rather than qualitative differences in neural responses to nonlocal agreement violations between L1 and (immersed) L2 speakers. This quantitative difference might indicate that agreement violations were detected online more consistently by the L1 group than the L2 group. Given the nature of this L1/L2 difference, it is compatible with theories that predict quantitative differences between L1 and L2 processing (e.g., Grüter et al., 2014; Hopp, 2014). Although our findings are not compatible with a strong view of "shallow" L2 processing that would predict L2 speakers cannot construct well-specified syntactic representations, the possibility that L2 speakers may not compute agreement as consistently as L1 speakers might be compatible with a weaker version of the Shallow Structure Hypothesis (Clahsen & Felser, 2006, 2018). Our results may also fall in line with the RAGE hypothesis (Grüter et al., 2014), which proposes weaker anticipatory processing in L2 speakers, in that the smaller P600 effect in the L2 group may have reflected reduced ability to predict upcoming verb features during subject-verb agreement processing. Note also, that given that in all our experimental sentences the noun that intervened between the verb and sentence subject was singular and thus matched the number properties of the verb, the smaller P600 for L2 speakers might also be suggestive of L2 speakers being more sensitive to interference from intervening constituents during online processing (Cunnings, 2017). While some existing research has investigated interference/attraction in L2 processing (Lago & Felser, 2018; Lim & Christianson, 2015; Tanner et al., 2012), further ERP research that manipulates the number properties of the intervening noun is required here to tease these accounts apart.

### *DOUBLE-NUMBER MARKING AND ERP EFFECTS IN L1 AND L2 PROCESSING*

The processing of agreement violations was also found to be modulated by double-number marking over the midline electrodes across the board as the P600 effect was larger when the sentences had a number-specified determiner compared to when they had a number-unspecified determiner for both groups, indicating double marking enhanced the neural signal to nonlocal agreement violations in L1 and L2 processing. Tanner and Bulkes (2015) argued that readers start predicting the number of an upcoming verb based on the number features of the subject NP. They argued that double-number marking has a higher degree of predictability and allows readers to make earlier anticipations as a quantifier, or demonstrative determiner in our case, clearly indicates the number features of the verb before encountering the subject NP. Previous studies have suggested stronger brain responses are associated with increased predictability and stronger predictions in lexical and syntactic processing (e.g., Brothers et al., 2015; DeLong et al., 2005; Wlotko & Federmeier, 2012). Therefore, the larger P600 effect we observed for double marking is compatible with the hypothesis that double-number marking leads to a stronger prediction being made in both groups.

Our results are consistent with Tanner and Bulkes (2015) for the L1 speakers but contrast in ways with Armstrong et al. (2018) for the L2 speakers. Similar to

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

Armstrong et al., our L2 speakers demonstrated P600 effects to number violations, but our results differ to Armstrong et al. in relation to the effect of double-number marking. Recall that Armstrong et al. observed smaller, rather than larger, P600 effects in their Chinese L2 speakers in sentences with double marking. They hypothesized that although Chinese speakers of English could acquire the underlying syntactic features of plurality marking in English there was an L1-influence effect for double marking. In other words, Armstrong et al. claimed that because Chinese exclusively marks plurality using quantifiers/demonstratives alone (there is no double marking), when Chinese speakers encounter plurality marked in the way that seemingly overlaps with Chinese, that is in a prenominal position using a determiner/quantifier in English, they allocate processing resources to this shared cross-linguistic cue, and consequently, less cognitive resources are allocated to the processing of morphosyntactic agreement cues. As noted previously, one difference between our study and Armstrong et al. is that while we used number-marked demonstratives (that/those), they used quantifiers (many/some). From the perspective of Armstrong et al.'s transfer-based account however, it is not clear that the type of double-number marking should matter, as Chinese has both quantifiers and demonstratives. As such, it does not immediately follow from their account that the type of prenominal (double) marking element should matter. However, as mentioned previously, Armstrong et al. (2018) used quantifiers, some of which (e.g., *some*) can occur with both plural and singular nouns with appropriate verbal agreement. This fact alone could possibly lead to a reduced P600 effect in the L2 speakers. As a result, in our study we avoided this issue by employing demonstratives that are strictly confined to either singular ("this/that") or plural ("these/those") nouns and thus obligatorily either singular (e.g., "is") or plural verbs (e.g., "are"). In doing so, we found the effect of determiner-number specification modulated L1 and L2 processing in the same direction without the need for further consideration. Another methodological difference between the two studies is the structures tested. While Armstrong et al. (2018) used local agreement, we adopted nonlocal agreement. Thus, future research is required to determine whether different findings between the two studies are related to structural complexity. Regardless, our data suggest that Chinese speakers do not merely rely on or prioritise the lexical cue from the determiner (when available) for number encoding, but also utilize morphological cues for number agreement computation.

## CONCLUSION

We observed sensitivity to nonlocal agreement violations and its interaction with double-number marking in both L1 and L2 groups across three tasks. Therefore, we suggest that, despite some observed quantitative differences, the Chinese speakers of English we tested in an immersion setting were, like English L1 speakers, able to compute agreement violations in nonlocal dependencies, and that double marking from determiner-number specification facilitates detection of number violations in both L1 and L2 processing. The P600 effects we observed suggest that Chinese speakers of English, at least in an immersion setting, have similar neural responses to L1 speakers when processing a novel agreement feature absent in the L1.

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

## SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0272263121000772.

## NOTES

[1]Following comments from two reviewers, we also conducted an analysis using a 500–800ms time window, which is the same time window as in Armstrong et al. (2018). Although other aspects of our findings stayed the same in this time window, the between-groups difference in the size of the P600 effect was no longer significant. This is in line with our claims that the P600 effect showed a longer duration in the L1 than the L2 group.

[2]Following Armstrong et al. (2018), we also conducted individual differences analyses with L2 proficiency and length of immersion, but did not find any significant effects. Given our L2 sample size, we do not draw any conclusions here about individual differences.

## REFERENCES

Alemán Bañón, J., Miller, D. & Rothman, J. (2017). Morphological variability in second language learners: An examination of electrophysiological and production data. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *43*, 1509–1536, https://doi.org/10.1037/xlm0000394

Alemán Bañón, J., & Rothman, J. (2019). Being a participant matters: Event-related potentials show that markedness modulates person agreement in Spanish. *Frontiers in Psychology*, *10*, 746.

Armstrong, A., Bulkes, N., & Tanner, D. (2018). Quantificational cues modulate the processing of English subject-verb agreement by native Chinese speakers: An ERP study. *Studies in Second Language Acquisition*, *4*, 731–754.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328.

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.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*, 45–93.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, *136*, 135–149.

Chen, L., Shu, H. U. A., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*, 161–174.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, *40*, 693–706.

Cunnings, I. (2017). Parsing and working memory in bilingual sentence processing. *Bilingualism: Language and Cognition*, *20*, 659–678.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*, 1117–1121.

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*, 85–103.

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition*, *25*, 529–557.

Frenck-Mestre, C., Osterhout, L., McLaughlin, J., & Foucart, A. (2008). The effect of phonological realization of inflectional morphology on verbal agreement in French: Evidence from ERPs. *Acta Psychologica*, *128*, 528–536.

Grüter, T., Rohde, H., & Schafer, A. J. (2014). The role of discourse-level expectations in non-native speakers' referential choices. In W. Orman & M. J. Valleau (Eds.), *Proceedings of the 38th Boston University Conference on Language Development* (pp. 179–191). Cascadilla Press.

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

Hammerly, C., Staub, A., & Dillon, B. (2019). The grammaticality asymmetry in agreement attraction reflects response bias: Experimental and modeling evidence. *Cognitive Psychology*, *110*, 70–104.
Hawkins, R., & Chan, C. Y. H. (1997). The partial availability of Universal Grammar in second language acquisition: The "failed functional features hypothesis." *Second Language Research*, *13*, 187–226.
Hawkins, R., & Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. *Language Acquisition and Language Disorders*, *30*, 21–44.
Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition*, *21*, 250–278.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.
Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*, 603–634.
Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, *61*, 940–967.
Kaan, E. (2014). Predictive sentence processing in L2 and L1: What is different? *Linguistic Approaches to Bilingualism*, *4*, 257–282.
Lago, S., & Felser, C. (2018). Agreement attraction in native and non-native speakers of German. *Applied Psycholinguistics*, *39*, 619–647.
Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research*, *25*, 173–227.
Lempert, H. (2016). Late L2ers can acquire grammatical features that do not occur in their L1: Evidence from the effect of animacy on verb agreement in L1 Chinese. *Memory & Cognition*, *44*, 538–553.
Lim, J. H., & Christianson, K. (2015). Second language sensitivity to agreement errors: Evidence from eye movements during comprehension and translation. *Applied Psycholinguistics*, *36*, 1283–1315.
McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*, 381–401.
Miller, D., & Rothman, J. (2020). You win some, you lose some: Comprehension and event-related potential evidence for L1 attrition. *Bilingualism: Language and Cognition*, *23*, 869–883.
Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*, 154–193.
Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*, 933–947.
Ojima, S., Nakata, H., & Kakigi, R. (2005). An ERP study of second language learning after childhood: Effects of proficiency. *Journal of Cognitive Neuroscience*, *17*, 1212–1228.
Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*, 494–522.
Osterhout, L., McKinnon, R., Bersick, M., & Corey, V. (1996). On the language specificity of the brain response to syntactic anomalies: Is the syntactic positive shift a member of the P300 family? *Journal of Cognitive Neuroscience*, *8*, 507–526.
Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*, 739–773.
Oxford University Press (2004). *Quick Placement Test: Version 1.* Oxford University Press.
Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*, 427–456.
Pliatsikas, C., & Marinis, T. (2013). Processing empty categories in a second language: When naturalistic exposure fills the (intermediate) gap. *Bilingualism: Language and Cognition*, *16*, 167–182.
Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua*, *120*, 2022–2039.
Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*, 40–72.
Shen, E. Y., Staub, A., & Sanders, L. D. (2013). Event-related brain potential evidence that local nouns affect subject–verb agreement processing. *Language and Cognitive Processes*, *28*, 498–524.

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772

Spychalska, M., Kontinen, J., & Werning, M. (2016). Investigating scalar implicatures in a truth-value judgement task: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, *31*, 817–840.

Tanner, D. (2011). *Agreement mechanisms in native and nonnative language processing: Electrophysiological correlates of complexity and interference.* University of Washington.

Tanner, D. (2019). Robust neurocognitive individual differences in grammatical agreement processing: A latent variable approach. *Cortex*, *111*, 210–237.

Tanner, D., & Bulkes, N. Z. (2015). Cues, quantification, and agreement in language comprehension. *Psychonomic Bulletin & Review*, *22*, 1753–1763.

Tanner, D., Goldshtein, M., & Weissman, B. (2018). Individual differences in the real-time neural dynamics of language comprehension. In *Psychology of Learning and Motivation* (Vol. *68*, pp. 299–335). Academic Press.

Tanner, D., Grey, S., & van Hell, J. G. (2017). Dissociating retrieval interference and reanalysis in the P600 during sentence comprehension. *Psychophysiology*, *54*, 248–259.

Tanner, D., Inoue, K., & Osterhout, L. (2014). Brain-based individual differences in online L2 grammatical comprehension. *Bilingualism: Language and Cognition*, *17*, 277–293.

Tanner, D., McLaughlin, J., Herschensohn, J., & Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, *16*, 367–382.

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, *76*, 195–215.

Tanner, D., Nicol, J., Herschensohn, J., & Osterhout, L. (2012). Electrophysiological markers of interference and structural facilitation in native and nonnative agreement processing. In *Proceedings of the 36th Boston University Conference on Language Development* (pp. 594–606). Cascadilla Press.

Tanner, D., & Van Hell, J. G. (2014). ERPs reveal individual differences in morphosyntactic processing. *Neuropsychologia*, *56*, 289–301.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*, 206–237.

Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, *62*, 356–366.

Downloaded from https://www.cambridge.org/core. IP address: 82.20.212.126, on 14 Dec 2021 at 12:32:44, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0272263121000772