Chinese Learners of English Are Conceptually Blind to Temporal Differences Conveyed by Tense

Li, Yang; Casaponsa Gali, Aina; Jones, Manon; Thierry, Guillaume

Brain and Language

DOI:
https://doi.org/10.1111/lang.12584

E-pub ahead of print: 14/05/2023

Publisher's PDF, also known as Version of record

LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

EMPIRICAL STUDY

Chinese Learners of English Are Conceptually Blind to Temporal Differences Conveyed by Tense

Yang Li,[a,c] Aina Casaponsa,[b] Manon Jones,[a] and Guillaume Thierry[a,d]

[a]Bangor University [b]Lancaster University [c]Solent University [d]Adam Mickiewicz University

Abstract: Chinese learners of English often experience difficulty with English tense presumably because their native language is tenseless. We showed that this difficulty relates to their incomplete conceptual representations for tense rather than their poor grammatical rule knowledge. Participants made acceptability judgments on sentences describing two-event sequences that were either temporally plausible or misaligned according to verb tense (time clash). Both upper-intermediate Chinese learners of English and native English speakers were able to detect time clashes between events,

---

CRediT author statement – **Yang Li**: conceptualization; methodology; investigation; formal analysis; writing – original draft preparation; writing – review & editing. **Aina Casaponsa**: formal analysis; writing – review & editing. **Manon Jones**: conceptualization; methodology; formal analysis; writing– original draft preparation; writing – review & editing. **Guillaume Thierry**: conceptualization; methodology; formal analysis; writing – original draft preparation; writing – review & editing.

A one-page Accessible Summary of this article in non-technical language is freely available in the Supporting Information online and at https://oasis-database.org

Guillaume Thierry was supported by the Polish National Agency for Academic Exchange (NAWA) under the NAWA Chair Programme (PPN/PRO/2020/1/00006). This article is dedicated to the memory of Professor Vivian Cook, who passed away in December 2021. Yang Li would like to thank Professor Cook for his selfless support, and especially for his role as supervisor and mentor. We also thank Associate Editor Kara Morgan-Short and anonymous reviewers for their helpful and constructive feedback.

Correspondence concerning this article should be addressed to Yang Li, School of Sport, Health, and Social Sciences, Solent University, UK. Email: yang.li@solent.ac.uk

The handling editor for this manuscript was Kara Morgan-Short.

DOI: 10.1111/lang.12584

showing that Chinese participants could apply tense rules explicitly. However, a predicted modulation of the N400 event-related brain potential elicited by time clashes in English-speaking participants was entirely absent in Chinese participants. In contrast, the same Chinese participants could semantically process time information when it was lexically conveyed in both languages. Thus, despite their mastery of English grammar, high-functioning Chinese learners of English failed to process the meaning of tense-conveyed temporal information in real time.

Keywords grammatical tense; bilingualism; semantics; event-related brain potentials; proficiency; N400

Introduction

Chinese–English bilinguals mostly fail to distinguish between past and recent past in English. For example, they struggle to grasp the difference between *I ate* and *I have eaten*—presumably because tense, a grammatical property that expresses event situation in time (Comrie, 1985), does not exist in Mandarin Chinese. One possibility is that they have a slightly different conceptualization of temporal information from that conveyed by grammatical tense in English, leading them to struggle to organize events on a timeline when temporal information is specifically conveyed by tense in the foreign language. Another possibility is that they superficially struggle to apply grammatical rules of their second language (L2) when such rules do not exist in their native language. In this project, we set out to test these alternative accounts using a sentence reading paradigm and the N400 component of event-related brain potentials (ERPs) as an index of semantic processing difficulty caused by inconsistency in the timeline of events described by sentences. We believe that this is the first attempt to give a neurocognitive account of the mechanism underlying tense processing difficulties in Chinese learners of English.

Despite the universal relevance of the human ability to process temporal information, languages vary in the ways that they encode time (Comrie, 1985; Declerck et al., 2006; Li & Thompson, 1989; Quirk et al., 1985). Given that tense does not exist in Chinese, English verbs in the past perfect, past simple, and present perfect tenses share the same translation in Mandarin Chinese; for example, *he had retired*, *he retired*, and *he has retired* all translate into *tuixiu-le*. In English, Declerck et al. (2006) defined the pre-present (grammatically referred to by the present perfect tense) as the portion of the present time sphere that precedes now (see Figure 1). Chinese, on the other hand, encodes only perfectiveness (Li & Thompson, 1989; Wang & Sun, 2015) and thus does not distinguish between the past and the recent past, that is, it does not grammatically distinguish the pre-present (present perfect) from either the past (past

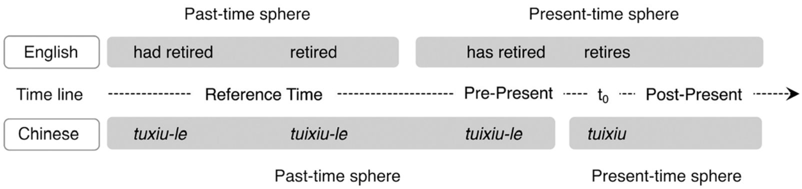

Figure 1 An example of linguistic differences between English and Chinese for the encoding of temporal information in the case of the verb *to retire*. Present time or now is labeled "t_0."

simple) or a time point in the past at which an event was completed (past perfect). In other words, in English, an event described by the present perfect is in the present time sphere, whereas in Chinese, the same event can be construed as belonging in the past time sphere.

Background Literature

Chinese learners of English often struggle with tense. For example, Lardiere (1998a, 1998b) reported the case of Patty, an end-state adult Chinese learner of English who had been studying in America for many years and who used only past tense morphology in 35% of cases where it was required, which is far from the benchmark for acquisition. Similarly, Hawkins and Liszka (2003) found that Chinese learners of English supplied correctly tensed morphemes in a spontaneous oral production task to a significantly lower extent than did German and Japanese learners of English even though all three groups had been matched for English proficiency.

It is now possible to measure the precise timeline of grammatical processes using online methods such as ERPs. ERPs are averaged electrical signals produced by the human brain that are detected at the surface of the scalp, usually in response to an external stimulation such as a picture, a written word, or a sound. The critical manipulation that leads from electroencephalographic data to ERPs is one that computes a mean from a set of time series or epochs that are time-locked to stimulus presentation. Through averaging, all signals that are unrelated to stimulus processing progressively fade out, whereas those that relate to stimulus presentation accumulate. The result is a series of peaks and troughs in brain activity that relate specifically to the set of stimuli presented. These stimuli can be modulated by the cognitive tasks implemented in the experiment. ERPs offer a unique opportunity for researchers to study the processing of tense by both first language (L1) and L2 speakers of a given language. Beyond the classic observation that violations of semantic expectations elicit negative modulations at around 400 ms after the onset of a critical word embedded within a sentence (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980), it has been well established that the human brain responds to detections of grammatical violations (e.g., morphosyntactic errors) through modulations of components such as left anterior negativity (LAN; Guajardo & Wicha, 2014; Hagoort et al., 2003; Hahne & Friederici, 2001; Morris & Holcomb, 2005) and P600 (Allen et al., 2003; Osterhout & Holcomb, 1992; Steinhauer & Ullman, 2002; for a concise overview of P600 in learners of a L2, see Morgan-Short et al., 2022). Violation of grammatical tense has been found to elicit N400 modulations also when the studied syntactic features are interpretatively relevant (Barber & Carreiras, 2005; Choudhary et al., 2009; Guajardo & Wicha, 2014).

Previous research has demonstrated that L2 learners' ERP responses are modulated by several variables such as age of L2 acquisition, L2 proficiency, L1 background, and/or L2 input context (immersion vs. formal education; for a review see, Caffarra et al., 2015; Steinhauer et al., 2009). Some authors have suggested that only bilinguals who receive natural L2 input before puberty can exhibit nativelike ERP responses to syntactic violations, whereas later L2 acquisition tends to result in patterns of neural activity markedly distinct from those associated with L1 processing (Hahne & Friederici, 2001; Hahne et al., 2006; Hernandez & Li, 2007; Wartenburger et al., 2003; Weber-Fox & Neville, 1996). However, Steinhauer et al. (2009) proposed that even late L2 learners can show nativelike neural patterns of activity when they are highly proficient in their L2. Steinhauer et al. also proposed that ERP responses to grammatical violations in L2 are strongly influenced by the degree of grammatical similarity between L1 and L2. Chen et al. (2007) studied ERP responses that indexed Chinese English learners' processing of subject–verb (S–V) agreement in English, a syntactic feature that does not exist in Chinese. They found that, although their Chinese–English learners could behaviorally detect S–V agreement violations, their ERP patterns still differed from those collected from the English native speakers in their study. In addition, Sabourin and Stowe (2008) demonstrated that in order to exhibit nativelike ERP patterns, L2 learners need to deal with a structure that not only exists in their L1 but also operates in a similar manner in their L1 and L2.

Tense agreement is one of the most investigated morphosyntactic features in neurolinguistics. Studies of tense processing have demonstrated that native speakers of a given language generate a biphasic LAN–P600 complex when they encounter a stimulus that makes a time adverbial and a tense form incompatible, that is, a tense violation, for example, "*Yesterday, we eat Peter's cake in the kitchen” (Steinhauer & Ullman, 2002, p. 63; see also Newman et al., 2007). Employing the same paradigm, White et al. (2012) investigated English tense processing in intermediate Chinese and Korean learners of English, learners whose native language completely lacks a tense system or expresses tense differently from the way English expresses tense. After nine weeks of intensive English exposure, both Chinese and Korean learners of English generated nativelike P600 effects that had initially been absent. White et al. interpreted the presence of P600 effects as indicating that Chinese and Korean learners had engaged the same neurocognitive processes as English native speakers do when presented with tense violations. As to the absence of the LAN effects, White et al. argued that this was due to their participants in the experiments having received only nine weeks of intensive exposure to English. White et al. argued that LAN effects are elicited only when L2 learners can apply English tense knowledge automatically and if they have achieved near-native proficiency in English. However, considering the point put forward by Sabourin and Stowe (2008), it could be that a nativelike LAN modulation is never observed in L2 Chinese and Korean learners of English because the grammatical feature does not exist (Chinese) or functions very differently (Korean) in the native language.

To test the hypothesis that grammatical features that do not exist in the native language cannot elicit nativelike responses in L2 learners, Li et al. (2018) tested how upper-intermediate Chinese–English bilinguals deal with temporal information at a semantic level when it is conveyed by tense, a grammatical feature that does not exist in Chinese. Chinese–English bilinguals and native English speakers were required to make acceptability judgements on sentences in which the adjunct clause started with “after” and was either temporally acceptable or not according to the tensed verb of the main clause (present perfect → past clash: **After the director of the school **has resigned** from the university, he **worked** for a multinational*; future → past clash: **After the director of the school **will have resigned** from the university, he **worked** for a multinational*). Although native English speakers failed to explicitly detect time clashes, they exhibited the expected N400 modulations for verbs that semantically violated the timeline. In contrast, Chinese–English bilinguals showed no such N400 modulation, indicating that they had difficulties keeping track of the conceptual timeline when temporal information was conveyed by tense. Critically, both English native speakers and Chinese–English bilinguals showed normal sentence comprehension as indexed by a standard N400 modulation elicited by incongruent endings (e.g., **After the director of the school resigned from the university, he worked for a **meter***). This being said, the focus on semantic anomalies in Li et al.'s (2018) study may have distracted participants from extended processing of tense and temporal agreement between clauses.

It has remained unclear, therefore, whether difficulties relating to processing temporal information conveyed by tense in Chinese learners of English observed by Li et al. (2018) were due to imprecision in their semantic representations of time or incorrect grammatical mapping across languages. Results from previous research (e.g., White et al., 2012) have suggested that Chinese learners of English can generate P600s like other L2 learners of English whose native language has tense (Korean speakers). Li et al. (2018) found that the most difficult condition for Chinese–English bilinguals to process was the present perfect → past clash condition (e.g., **After the director of the school* ***has resigned*** *from the university, he* ***worked*** *for a multinational*). In our study, we aimed to test whether Chinese–English bilinguals with verified proficiency in English, that is, above B1 according to the Common European Framework Reference for Languages (CEFR), conceptualized the period denoted by the present perfect as belonging to the past or the present time sphere when their attention was actively directed at temporal information. It is noteworthy that in our study, as in Li et al.'s (2018) study, we investigated semantic violations introduced by incorrect tense use across different sentences rather than within-sentence tense violations (e.g., **Tomorrow, I ate an apple*). While tense is likely to create additional grammatical processing demands in Chinese–English bilinguals because it is absent in their Chinese L1, we hypothesized that semantic representations of time elicited by tensed forms are underspecified at a conceptual level, thus resulting in problems with semantic integration of temporal information at the suprasentential level.

We thus aimed to address two limitations of Li et al.'s (2018) study:

- Although most of the Chinese–English bilingual participants provided International English Language Testing System (IELTS) scores in Li et al.'s (2018) study, the times at which they had taken the test varied. Therefore, the IELTS scores may not have reflected the participants' English proficiency at the time of testing. In our study, we recruited Chinese–English bilinguals with a IELTS score of 6.5 or higher, and, in addition, we asked them to take the Oxford Placement Test and achieve results showing that they had an upper-intermediate level of proficiency.
- The semantic violation condition in Li et al.'s study (2018) was construed as a coarse baseline condition in which the violation was elicited by an aberrant semantic concept rather than a temporal clash. The

participants in that study were thus likely to have focused on word-based semantic violation detection at the expense of their detecting time clashes. Therefore, in contrast to Li et al.'s (2018) study, we put the emphasis on temporal information processing in this study to ensure that our participants' result did not derive from a lack of attention to temporal information.

The Present Study

We asked Chinese learners of English to make temporal order acceptability judgements for complex sentences in the form of *After Event 1, Event 2*, for which they had to determine whether the Event 1 → Event 2 time sequence was acceptable or not depending on temporal information conveyed by tense. In a first set of four conditions, we compared present time → past time clashes with past time → past time acceptable sequences to evaluate processing abilities in our bilingual participants using a group of native English speakers as control participants (see Table 1). Li et al. (2018) used a past perfect construction as the control condition (correct: *After the director of the school* ***had resigned*** *from the university, he* ***worked*** *for a multinational*). However, the past perfect is less frequently used in English to refer to a past event than the past simple. Therefore, in our study, we doubled the number of acceptable sentences, with half featuring a past perfect → past sequence and the other half, a past simple → past sequence. Keeping in mind that the present perfect was the most difficult condition in Li et al.'s (2018) study, we set out to compare a time clash condition involving a present perfect → past sequence condition with the two acceptable conditions of past perfect → past and past simple → past. In addition, we also wanted to compare the present perfect → past clash condition to a more obvious temporal clash involving the present simple, namely, present simple → past clash condition (see Table 1). Taken together, we expected the comparison of the present perfect → past clash condition with the past perfect → past and past simple → past acceptable conditions and the present simple → past clash condition, respectively, to shed light on whether Chinese–English bilinguals conceptualize the time sphere denoted by the present perfect as belonging to the present (i.e., clashing with an adjunct clause in the past perfect or past simple tense) or the past (clashing with an adjunct clause in the present simple tense).

Li et al. (2018) tested only time clashes. In this study, we also tested how Chinese learners of English deal with past → future time gaps in comparison to present → future acceptable sequences (see Table 2), assuming that they could process time gaps comparably to English native speakers since the future in

Table 1 Time clash examples of adjunct and main clauses experimental conditions describing Event 1 and Event 2

Tense sequence	Adjunct clause describing Event 1	Main clause describing Event 2	Sequence labels
Past perfect → past	After the director of the school ***had resigned*** from the University,	he ***worked*** for a multinational.	Acceptable
Past simple → past	After the director of the school ***resigned*** from the University,		Acceptable
*Present perfect → past	*After the director of the school ***has resigned*** from the university,		Clash
*Present simple → past	*After the director of the school ***resigns*** [a] from the university,		Clash
Chinese translation	院长从大学辞职(**了**)以后 [a] yuanzhang cong daxue cizhi(le) [b]yihou,	他去**了**–家跨国公司工作。 ta qu-le yijia kuaguo gongsi gongzuo.	

Note. [a]The present form (e.g., *resigns*) in the adjunct clause is translated without the perfective marker *le* in Chinese. [b]In Chinese, the perfective aspect marker *le* is optional when the main clause is in the past. Thus, even if the marker *le* is omitted, the adjunct clause can be interpreted as referring to the past once the past form of the main clause is encountered.

English and Chinese is conveyed by a modal auxiliary (*jiang* "will"). To counterbalance the experimental design (and, by the same token, make the critical verb form of the main clause unpredictable), we implemented two time-gap conditions involving a main clause in the future: a past perfect → future gap condition and a past simple → future gap condition. The acceptable conditions that did not feature a time gap were present perfect → future and present simple → future conditions. Importantly, the acceptable present perfect → future condition allowed another test of Chinese–English bilinguals' ability to conceptualize the present perfect as belonging to the present or past time sphere: If they considered the present perfect as belonging in the past, the acceptable present perfect → future condition would be misconstrued as a time gap condition instead of being acceptable.

We used the N400 as an index of conceptual processing difficulty in relation to resolving temporal relations between two parts of a sentence (adjunct

Table 2 Time gap examples of adjunct and main clauses experimental conditions describing Event 1 and Event 2

Tense sequence	Adjunct clause describing Event 1	Main clause describing Event 2	Sequence labels
Present perfect → future	After the director of the school ***has resigned*** from the university,	he ***will*** [b] ***work*** for a multinational.	Acceptable
Present simple → future	After the director of the school ***resigns*** [a] from the university,		Acceptable
*Past perfect → future	*After the director of the school ***had resigned*** from the University,		Gap
*Past simple → future	*After the director of the school ***resigned*** from the University,		Gap
Chinese translation	院长从大学辞职(了)以后 yuanzhang cong daxue cizhi(le) yihou,	他将要去–家跨国公司工作。 ta jiangyao [b] qu yijia guaguo gongsi gongzuo	

Note. [a]The present form (e.g., *resigns*) in the adjunct clause is translated without the perfective marker *le* in Chinese. [b]Similar to the case of English, the future form is marked by *jiang* (future modal auxiliary) in Chinese.

and main clause). In other words, the experimental paradigm did not involve syntactic violations (e.g., S–V agreement or tense violations) since each clause considered separately was always grammatically correct. This was a deliberate choice, because we aimed at measuring conceptual integration of temporal information conveyed by tense in English in individuals who do not have this device in their native language (e.g., Comrie, 1985). We therefore predicted a modulation of the N400 component of ERPs, which is well known to index semantic processing difficulty (Li et al., 2018; Kutas & Federmeier, 2011; Kutas & Hillyard, 1984a, 1984b; Niewland, 2015; Van Petten & Kutas, 1990).

For the time clash manipulation, we predicted that, while Chinese learners of English should be able to behaviorally detect clashes conveyed by grammatical tense in a similar fashion to a control group of native English speakers, crucial differences would emerge at the level of semantic processing indexed by ERPs: In the English participants, N400 amplitude increases should have been observed for both time clashes and time gaps compared to acceptable event

Table 3 Planned comparisons and prediction of N400 modulation between conditions

		Prediction	
Test	Planned comparisons	English group	Chinese group
Time clash	Present perfect clash vs. Past perfect acceptable	Yes	No
	Present perfect clash vs. Past simple acceptable	Yes	No
	Present perfect clash vs. Present simple clash	No	Yes
Time gap	Past perfect gap vs. Present perfect acceptable	Yes	Yes
	Past simple gap vs. Present perfect acceptable	Yes	Yes
	Present simple acceptable vs. Present perfect acceptable	No	Yes

sequences (see Table 3). In contrast, we expected that Chinese–English bilingual participants would present poor discrimination ability in the N400 range when we compared present perfect → past clashes to both acceptable past perfect → past and past simple → past sequences because the Chinese–English bilingual participants might construe the present perfect as belonging to the past time sphere. Conversely, for the time gap manipulation, we anticipated past perfect → future and past simple → future time gaps to elicit N400 modulations in all the participants since the future is conveyed by a modal auxiliary in both English and Chinese. In addition, we expected that Chinese learners of English would show abnormal N400 modulation in the acceptable present perfect → future condition, should they have perceived the present perfect as referring to the past (see Table 3).

Given that we recruited the bilingual participants for their having achieved a good mastery of English grammar, we expected them to perform well in the explicit temporal sequence acceptability judgement task. In other words, we expected a dissociation between online, conceptual processing indexed by the N400 and Chinese participants' explicit ability to detect problems in sentences that contravened a learned rule of English grammar.

Method

Participants

Twenty Chinese learners of English[1] (15 females: $M_{\text{age}} = 24.80$ years, $SD = 5.03$, range = 19–35) and 24 native English speakers (13 females: $M_{\text{age}} = 19.71$ years, $SD = 1.81$, range = 18–26) took part in the study. We discarded data from five English native speakers due to low accuracy (below 55%) in the explicit judgement task during the ERP session ($M_{\text{accuracy}} = 54\%$, $SD = 1$, 95% CI [52, 56]). We applied the same threshold to the bilingual participants without leading to data exclusion. In addition, we removed one English and one Chinese participant from analyses due to heavy blinking, excessive alpha wave contamination, or other muscle artefact contamination. In order to systematically control the bilingual participants' English proficiency, we recruited only bilingual participants who had achieved 6.5 or above in the IELTS. In addition, we required the bilingual participants to take the Oxford Quick Placement Test (OQPT, Allen, 1992; e.g., see Lemhöfer & Broersma, 2012, for use of this test in empirical research). We further excluded one Chinese–English bilingual's dataset as this participant failed to achieve above 67% in this test, which corresponds to B1 (intermediate) proficiency according to the CEFR. Mean English proficiency of the bilingual participants was 6.8 out of maximum score of 9 ($SD = 0.5$, range = 6.5–8.0) according to the IELTS. Accuracy in the OQPT was 79% on average ($SD = 5$, range = 70–92), meaning that the participants had a level of at least B2 (upper intermediate) according to the CEFR. Finally, we also collected self-rated proficiency of the participants for reading, writing, speaking, and listening in both Chinese and English, as well as age of acquisition, length of L2 learning, and daily language exposure (see Table 4).

All the participants in the native English group declared having beginner's knowledge of languages other than English or no knowledge of another language, and no English native participant reported any knowledge of Chinese. All the native English participants were studying at Bangor University in the United Kingdom and received either payment or five course credits for their time. The ethics committee of the School of Psychology at Bangor University approved the study.

Stimuli

We used a total of 70 English sentence sets, each containing eight complex sentences built upon the same template (stimuli are available at https://osf.io/yn63x). For four sentences in each set, the main clause ended in the past simple, with two sentences featuring adjunct clauses with a temporally

Table 4 Chinese participants' language background in the final sample ($n = 18$)

Measure	*M*	*SD*	95% CI
Age of L2 acquisition	10.1	3.0	[8.56, 11.55]
Length of L2 learning (years)	12.9	3.1	[11.36, 14.41]
Daily Chinese usage (%)	58	18	[49, 67]
Daily English usage (%)	42	18	[33, 51]
Self-rated proficiency in Chinese (0–10)			
Reading	8.9	1.1	[8.42, 9.47]
Writing	8.3	1.2	[7.72, 8.95]
Speaking	9.1	1	[8.56, 9.55]
Listening	8.9	1.1	[8.39, 9.50]
Self-rated proficiency in English (0–10)			
Reading	7.2	1.3	[6.59, 7.85]
Writing	6.2	1.6	[5.38, 6.95]
Speaking	6.6	1.5	[5.83, 7.28]
Listening	6.4	1.2	[5.85, 7.04]

acceptable timeline and the other two featuring a time clash (i.e., present perfect → past and present simple → past clash conditions; see Table 1). Half of the verbs that we used were regular (suffix *–d/ed*) and half were irregular, but we did not systematically analyze regularity due to a lack of statistical power.[2] For the other four sentences in a set, the main clause ended in the future. Two of these sentences featured adjunct clauses with a temporally acceptable timeline and the other two, a time gap (i.e., past perfect → future and past simple → future gap conditions; see Table 3). The location of the conceptual timeline violation always coincided with the verb of the main clause. For each sentence set, half of the participants saw the clash versions and the other half saw the gap versions. There were thus two presentation lists, such that when the main clause of the sentence frame was in the past simple tense (temporal clash testing) in one list, the main clause was in the future in the other list (temporal gap testing). Each presentation list contained 280 stimuli with 35 stimuli per condition and was split across four blocks, ensuring that conditions were equally represented across blocks and that a given sentence frame was never repeated within the same block. Finally, sentence stimuli were randomized within each block.

Procedure

All the participants first filled out a language background questionnaire. The bilingual participants also completed the OQPT. The time for completion was set to 40 mins. During the ERP session, the participants read the first clause

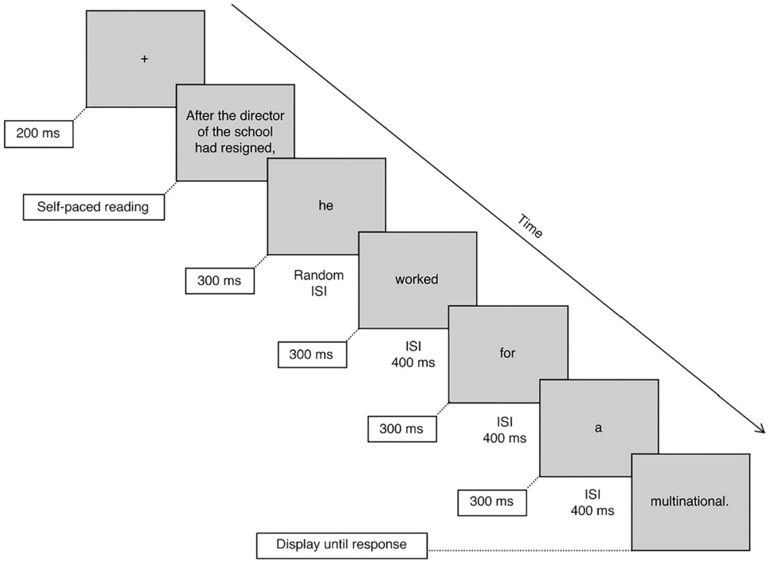

Figure 2 Trial structure and timing of stimulus presentation. ISI = interstimulus interval.

of each sentence all at once at their own pace and then pressed any button of a response box to trigger a step-by-step presentation of the main clause in the center of the screen. Each step involved the presentation of one or two words for a duration of 300 ms, with an interstimulus interval of 400 ms, except for the interval preceding the main clause verb or auxiliary-verb combination before which the interval was selected between 220–400 ms in 20 ms gap (random interstimulus interval in Figure 2). For the main clause, each presentation step involved either one or two words, such that nominal phrases (e.g., "the lawyer") and auxiliary–verb combinations (e.g., "will retire") were presented all at once, whereas all other words (e.g., pronouns, nouns, individual verbs) were presented individually. As a result, the target verb of the main clause was always displayed in second position within the stream (whether or not together with an auxiliary), regardless of the experimental condition. Once the participants had finished reading the whole sentence, they were asked to judge whether or not the temporal order of the events described in the whole sentence was correct with the instruction, "Using the two designated keys, indicate whether or not the two events described by each sentence happened in an acceptable temporal order." The ERP session lasted about 40 minutes. After the end of the ERP session, we asked the participants to rate the acceptability of the temporal order of the events described in each sentence on a 5-point scale (only for the half of the stimuli presented during the ERP session). For this posttest, the participants freely read the sentences presented all at once and there was no time limit.

Data Analysis

Behavioral Data

We analyzed the participants’ accuracy ratings with logit mixed effects models with condition and group as fixed effects predictors. We fitted the maximal random structure including random intercepts for participants and sentences and random slopes for all within-participant and within-sentence predictors including main effects and interactions (Barr, 2013; Barr et al., 2013). We fitted models using R (R Core Team, 2012) with the lme4 package (Bates et al., 2008). We reduced the random structure of the model for the time clash analysis to arrive at a parsimonious model since the data did not support the execution of the maximal model random structure. To do so, we computed principal component analyses of the random structure (see Bates et al., 2015) and dropped the components that did not significantly contribute to the cumulative variance (see step-by-step analyses at https://osf.io/yn63x). We applied effect coding to all categorical predictors before model fitting by subtracting the mean from each dummy-coded level. We computed Type III analyses of variance (ANOVA) for main effects and interactions using the car package (Fox & Weisberg, 2019). We computed estimate effects for main effects with the emmeans package (Lenth, 2020) as well as z values for planned comparisons (see Appendix S1 in the Supporting Information online for a full summary of the results, including asymptotic lower and upper 95% confidence intervals for differences between conditions). We analyzed only accuracy data given that we collected reaction times at the end of sentence presentation, and thus they do not reflect online processing of acceptability based on the critical word of the main clause.

Event-Related Potential Data Collection and Preprocessing

We recorded the electrophysiological data at a rate of 1 kHz from 64 Ag/AgCl electrodes according to the extended 10–20 convention. The reference electrode was Cz, and we kept the impedances below 5 kΩ. We filtered the electroencephalogram online using a bandpass filter with cutoff values of 0.05 Hz low pass, 200 Hz high pass, and an accuracy of 0.15 nV/LSB. For analysis, we used the eeglab Toolbox (Delorme & Makeig, 2004) and the erplab Toolbox (Lopez-Calderon & Luck, 2014) for offline data processing. We down-sampled the signals offline to 250 Hz and high-pass filtered them using a zero-phase shift filter with a cutoff of 0.1 Hz (slope 24 dB/oct). We visually inspected the data for abnormalities and removed segments containing major artefacts (i.e., muscle artefacts). We then rereferenced the data to the global average reference (excluding electrooculogram channels) before applying independent component analysis correction for blinks, eye movements, and muscle activity. We then applied a low-pass filter with a cutoff frequency of 30 Hz (slope 24 dB/oct). We extracted epochs ranging from −200 to 1,000 ms after stimulus onset from continuous data and applied baseline correction relative to prestimulus activity (−200 to 0 ms). We discarded epochs with activity exceeding $\pm 100\ \mu V$ at any electrode site except the electrooculogram channels. Each participant had more than 30 trials included in the final analysis in each experimental condition. Overall, we excluded 1.5% of trials on average.

Event-Related Potential Data Analysis

We extracted N400 mean amplitudes in the time window where N400 maximal sensitivity is usually observed, between 300–500 ms after onset of the critical word (i.e., the second stimulus of the main clause) at electrodes of predicted maximal sensitivity in the case of visual word presentation (C1, Cz, C2, CP1, CPz, and CP2; Kutas & Federmeier, 2011; Kutas & Hillyard, 1984a, 1984b; Van Petten & Kutas, 1990). In line with our predictions, we conducted two analyses of N400 amplitudes. Both analyses entailed a 2 (group) × 4 (conditions) repeated-measures ANOVA. Group was the between-subjects variable with two levels: Chinese–English bilingual, English native speaker. Condition was the within-subjects variable with four levels: past perfect, past simple, present perfect, present simple. Importantly, the conditions did not have the same status with respect to acceptability in the two analyses: Whereas the past perfect and past simple conditions were acceptable in the time clash analysis; present perfect and present simple were acceptable in the gap analysis (see Table 3). Finally, we ran planned comparisons testing the specific differences predicted in our hypotheses, namely, present perfect versus past simple, past perfect, and present simple, bearing in mind that our predictions were different for each of the two analyses (see Appendix S2 in the Supporting Information online for a full summary of the results).

Results

Behavioral Results

Time Clashes

The logit mixed effects models that we conducted on the accuracy data revealed a main effect of condition, $\chi^2(3) = 43.18, p < .001$, such that overall accuracy

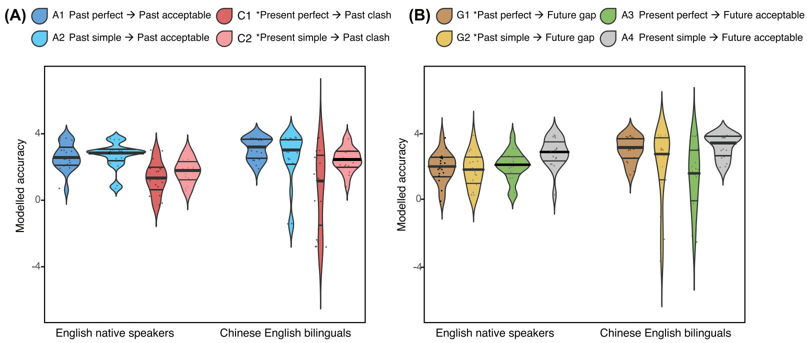

Figure 3 English native speakers and Chinese–English bilingual participants' response accuracy (logit transformation): modeled accuracy data for the time clash manipulation (Panel A) and modeled accuracy data for the time gap manipulation (Panel B). Individual datapoints were plotted as dots within a violin plot for each experimental condition. Thick lines depict the median and upper and lower thin lines depict 25 and 75 percentiles, respectively. A1 = past perfect → past acceptable; A2 = past simple → past acceptable; C1 = present perfect → past clash; C2 = present simple → past clash. A3 = present perfect → future acceptable; A4 = present simple → future acceptable; G1 = past perfect → future gap; G2 = past simple → future gap.

ratings were lower for the present perfect → past clash condition ($M = 69\%$) compared to the present simple → past clash ($M = 89\%$), the past perfect → past acceptable ($M = 95\%$), and past simple → past acceptable ($M = 90\%$) conditions: present perfect → past clash versus present simple → past clash, $b = -1.14$, $SE = 0.32$, 95% CI = [−1.97, 0.31], $z = -3.51$, $p < .005$; present perfect → past clash versus past perfect → past acceptable, $b = -2.28$, $SE = 0.38$, 95% CI = [−3.25, −1.3], $z = -6.02$, $p < .001$; present perfect → past clash versus past simple → past acceptable, $b = -2.28$, $SE = 0.47$, 95% CI = [−3.49, −1.08], $z = -4.87$, $p < .001$ (see Figure 3, Panel A). The main effect of group was not significant, $\chi^2(1) = 0.46$, $p = .50$. The interaction of group and condition also did not reach significance levels, $\chi^2(3) = 7.32$, $p = .06$. Post hoc comparisons showed that the two groups did not significantly differ one from the other in their acceptability ratings for the present perfect → past clash condition, $b = -0.67$, $SE = 0.71$, 95% CI = [−2.06, 0.73], $z = -0.94$, $p = .35$, but Chinese learners of English were more accurate than the native English speakers in the present simple → past clash and in the past perfect → past acceptable conditions, respectively: present simple → past clash condition, $b = 0.86$, $SE = 0.31$, 95% CI = [0.25, 1.48], $z = 2.75$, $p = .006$; past perfect →

past acceptable condition, $b = 0.93$, $SE = 0.43$, 95% CI = [0.09, 1.78], $z = 2.17$, $p = .03$.

Time Gaps

We found a main effect of condition, $\chi^2(3) = 31.74$, $p < .001$(see Figure 3, Panel B), such that accuracy ratings were overall lower in the present perfect → future acceptable condition ($M = 79\%$) than in the present simple → future acceptable condition ($M = 96\%$), but not in the past perfect → future gap ($M = 91\%$) and past simple → future gap ($M = 82\%$) conditions: present perfect → future acceptable versus present simple → future acceptable, $b = -1.90$, $SE = 0.34$, 95% CI = [−2.78, −1.01], $z = -5.51$, $p < .001$; present perfect → future acceptable versus past perfect → future gap, $b = -0.97$, $SE = 0.39$, 95% CI = [−1,97, 0.04], $z = -2.48$, $p = .06$; present perfect → future acceptable versus past simple → future gap, $b = -0.59$, $SE = 0.52$, 95% CI = [−1.93, 0.75], $z = -1.14$, $p = .67$. We also found a marginal main effect of group, $\chi^2(1) = 3.52$, $p = .06$, and an interaction of group and condition, $\chi^2(3) = 13.48$, $p < .005$. Post hoc comparisons revealed a similar pattern of results as for the pattern of the time clash analyses. The two groups did not significantly differ in their acceptability ratings in the present perfect acceptable condition: present perfect → future acceptable, $b = -0.71$, $SE = 0.61$, 95% CI = [−1.90, 0.48], $z = -1.17$, $p = .24$, but Chinese–English bilinguals were significantly more accurate than native English speakers in the present simple acceptable and the past perfect gap conditions: present simple → future acceptable, $b = 1.14$, $SE = 0.57$, 95% CI = [0.02, 2.26], $z = 1.99$, $p = .05$; past perfect → future gap, $b = 1.73$, $SE = 0.44$, 95% CI = [0.86, 2.60], $z = 3.91$, $p < .001$.

ERP Results

N400 Analyses: Time Clashes

The repeated measures ANOVA that we conducted on mean N400 amplitudes revealed a significant main effect of group, $F(1, 34) = 6.31$, $p = .02$, $\eta_p^2 = .16$ (large effect, according to Cohen, 1969, as cited in Richardson, 2011), such that N400 amplitude was overall more negative in the English native speakers than in the Chinese–English bilinguals. We also found a significant main effect of condition, $F(3, 102) = 8.97$, $p < .001$, $\eta_p^2 = .21$ (large effect). As shown in Figure 4, the interaction of condition and group was also significant, $F(3, 102) = 6.66$, $p < .001$, $\eta_p^2 = .16$ (large effect).

Planned comparisons showed that N400 amplitudes in the English group were significantly more negative in the present perfect clash condition compared to both the past perfect acceptable and past simple acceptable conditions,

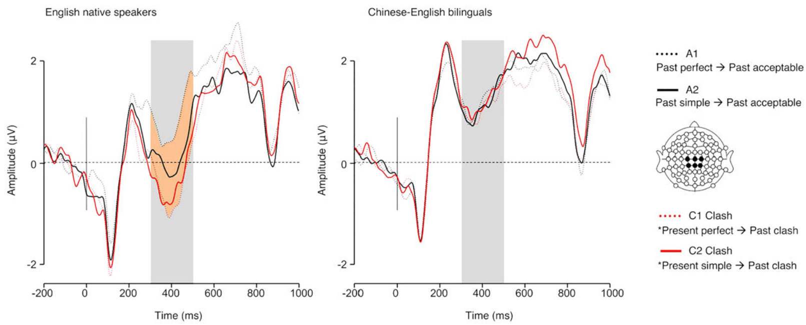

Figure 4 Event-related brain potentials elicited in the past perfect → past acceptable (A1), past simple → past acceptable (A2), present perfect → past clash (C1), and present simple → past clash (C2) conditions in English native speakers and Chinese–English bilingual participants. Waveforms depict the linear derivation of signals recorded at six electrode sites (C1, Cz, C2, CP1, CPz, CP2). The time-window highlighted in light grey is the predicted window of maximal N400 sensitivity: 300–500 ms. The orange shading highlights differences among the conditions found in the N400 time window.

respectively: present perfect → past clash versus past perfect → past acceptable, $t(17) = -4.98$, $p < .001$, $d = -1.17$ (strong effect, according to Cohen, 1988); present perfect → past clash versus past simple → past acceptable, $t(17) = -2.54$, $p = .01$, $d = -0.60$ (medium effect), whereas we found no significant differences between the two time clash conditions: present perfect → past clash versus present simple → past clash, $t(17) = -0.37$, $p = .36$, $d = -0.09$ (small effect). The Chinese participants did not show significant N400 amplitude differences between the present perfect clash and any of the other experimental conditions: present perfect → past clash versus past perfect → past acceptable, $t(17) = -1.03$, $p = .16$, $d = -0.24$ (small effect); present perfect → past clash versus past simple → past acceptable, $t(17) = -0.25$, $p = .40$, $d = -0.06$ (small effect); present perfect → past clash versus present simple → past clash, $t(17) = -0.53$, $p = .30$, $d = -0.12$ (small effect).[3] Figure 5 depicts the topography of the N400 differences elicited by each of the ad hoc contrasts in the case of main clauses describing an event situated in the past.

N400 Analyses: Time Gaps

N400 mean amplitudes differed significantly between the two groups when the main clause event was situated in the future, $F(1, 34) = 4.98$, $p = .03$, $\eta_p^2 = .13$ (medium effect), such that the English native speakers showed overall more negative amplitudes than did the bilingual participants. As shown in

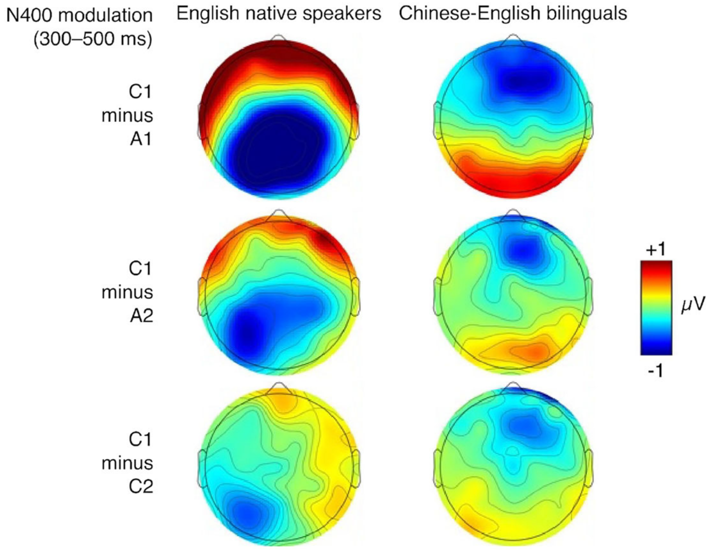

Figure 5 Topographical maps of even-related brain potential activity across the 64-channel array between 300–500 ms after the onset of the critical word in the present perfect clash condition compared to the past perfect acceptable (top row), past simple acceptable (middle row), and present simple clash (bottom row). N400 modulations were elicited by all contrasts in English native speakers but were not observed for any of the three contrasts in Chinese–English bilinguals. A1 = past perfect → past acceptable; A2 = past simple → past acceptable; C1 = present perfect → past clash; C2 = present simple → past clash.

Figure 6, we also found a main effect of condition, $F(3, 102) = 4.45$, $p = .01$, $\eta_p^2 = .12$ (medium effect), and the interaction of condition and group was also significant, $F(3, 102) = 3.61$, $p = .02$, $\eta_p^2 = .10$ (medium effect).

Planned comparisons in the English native speaker group showed that N400 mean amplitudes were more negative in both the two time-gap conditions compared to the present perfect acceptable condition: past perfect → future gap versus present perfect → future acceptable, $t(17) = -2.35$, $p = .02$, $d = 0.55$ (medium effect); past simple → future gap versus present perfect → future acceptable, $t(17) = -1.79$, $p = .05$, $d = 0.42$ (small effect). In contrast, the Chinese participants showed no significant difference between the past simple gap and the present perfect acceptable condition: past simple → future gap versus present perfect → future acceptable, $t(17) = 0.44$, $p = .33$, $d = -0.10$ (small effect). As for the past perfect gap versus present perfect acceptable comparison, the Chinese participants had significantly more positive

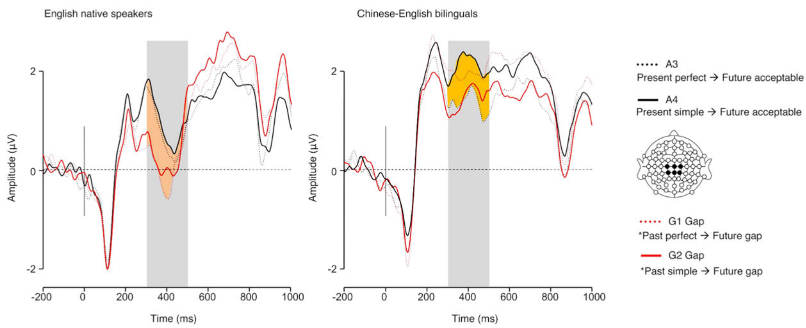

Figure 6 Event-related brain potentials in the present perfect → future acceptable (A3), present simple → future acceptable (A4), past perfect → future gap (G1), and past simple → future gap (G2) conditions in English native speakers and Chinese–English bilingual participants. Waveforms depict the linear derivation of signals recorded at six electrode sites (C1, Cz, C2, CP1, CPz, CP2). The time-window highlighted in light grey is the predicted window of maximal N400 sensitivity: 300–500 ms. The orange shading highlights differences among the conditions found in the N400 time window. The yellow shading highlights the difference between conditions present perfect → future acceptable and present simple → future acceptable found in Chinese–English bilinguals.

N400 amplitudes in response to the past perfect gap condition: past perfect → future gap versus present perfect → future acceptable, $t(17) = 1.89$, $p = .04$, $d = 0.45$ (small effect). In addition, they also showed a difference in N400 mean amplitudes between present simple acceptable and present perfect acceptable conditions: present simple → future acceptable versus present perfect → future acceptable, $t(17) = 3.34$, $p = .002$, $d = -0.79$ (large effect), which was not present in the English native speakers, $t(17) = 0.97$, $p = .17$, $d = -0.23$ (small effect). Figure 7 depicts the topography of the N400 differences elicited by each of the ad hoc contrasts in the case of main clauses describing an event situated in the future.

P600 Analyses

Upon invitation from our reviewers, we analyzed P600 mean amplitudes by means of repeated measures ANOVAs with group (English native speakers, Chinese–English bilinguals) as a between-subjects variable and condition (past perfect, past simple, present perfect, present simple) as a within-subjects variable. We extracted mean amplitudes in a time window classically associated with P600 modulations, 600–900 ms after onset of the critical word, at electrodes CP1, CP2, CPz, P1, P2, and Pz.

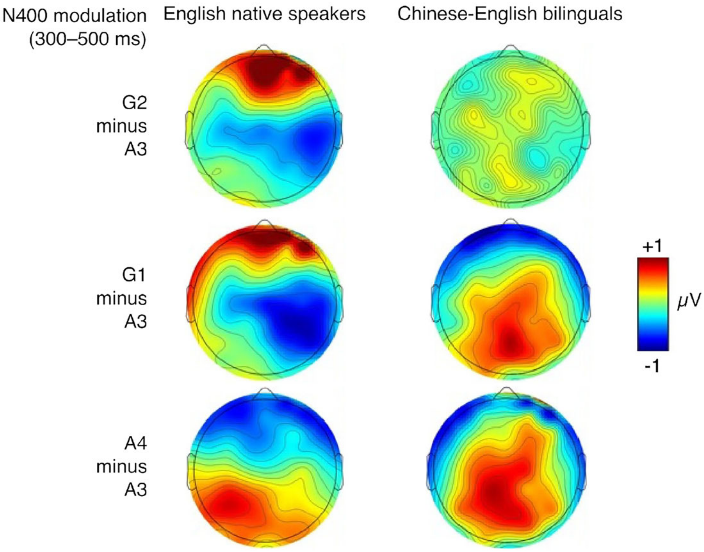

Figure 7 Topographical maps of event-related brain potential activity across the 64-channel array between 300–500 ms after the onset of the critical word in the present perfect acceptable compared to the past perfect gap (top row), past simple gap (middle row), and present simple acceptable (bottom row) conditions. N400 modulations were elicited by both time gaps in English native speakers (top two rows). In Chinese learners of English, time gap conditions failed to elicit N400 modulations relative to the present perfect acceptable condition, but, critically, the present perfect acceptable condition elicited an N400 relative to the present simple acceptable condition, showing up as a positive scalp distribution in the bottom right row, given the direction of the comparison. A3 = present perfect → future acceptable; A4 = present simple → future acceptable; G1 = past perfect → future gap; G2 = past simple → future gap.

Time Clashes

The main effect of condition was marginally significant, $F(3, 102) = 2.58$, $p = .06$, $\eta_p{}^2 = .07$ (medium effect). As illustrated in Figure 8, there was no main effect of group, $F(1, 34) = 0.03$, $p = .87$, $\eta_p{}^2 = .001$ (small effect), and no interaction of group and condition, $F(3, 102) = 2.07$, $p = .11$, $\eta_p{}^2 = .06$ (medium effect).

In an exploration of the main effect of condition, pairwise comparisons, with p values corrected using the Bonferroni procedure for multiple comparisons, showed that in both the English native speaker and the Chinese groups, the participants had significantly more positive P600 amplitudes in the present simple clash condition compared to the past simple acceptable

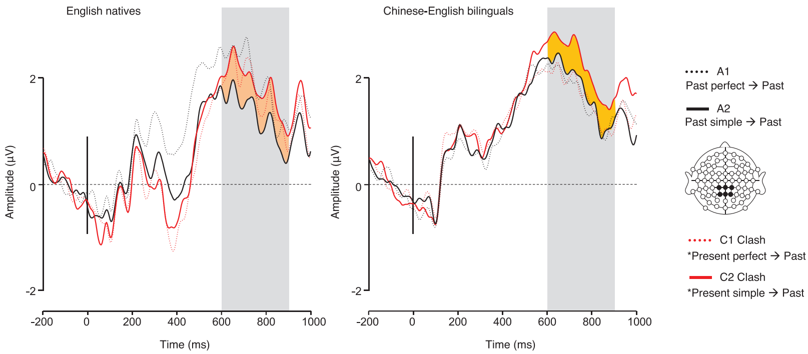

Figure 8 Event-related brain potentials in the past perfect → past acceptable (A1), past simple → past acceptable (A2), present perfect → past clash (C1), and present simple → past clash (C2) conditions in English native speakers and Chinese–English bilingual participants. Waveforms depict the linear derivation of signals recorded at six electrode sites (P1, P2, Pz, CP1, CPz, CP2). The highlighted time window is the predicted window of maximal P600 sensitivity: 600–900 ms.

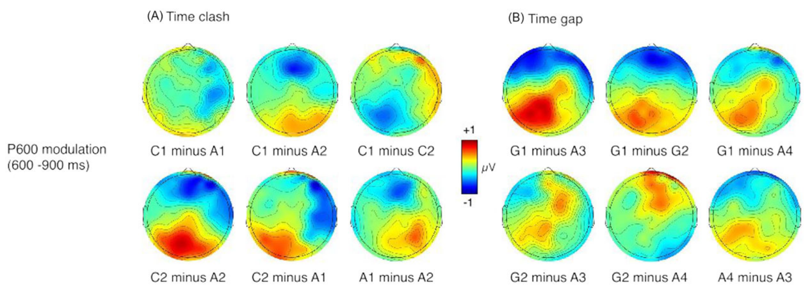

Figure 9 Topographical maps of event-related brain potential activity across the 64-channel array between 600–900 ms after the onset of the critical word in the four conditions for (Panel A) the time clash manipulation and (Panel B) the time gap manipulation. A1 = past perfect → past acceptable; A2 = past simple → past acceptable; C1 = present perfect → past clash; C2 = present simple → past clash; A3 = present perfect → future acceptable; A4 = present simple → future acceptable; G1 = past perfect → future gap; G2 = past simple → future gap.

condition: present simple → past clash versus past simple → past acceptable, $p = .05$ (see Figure 9 and Appendix S2).

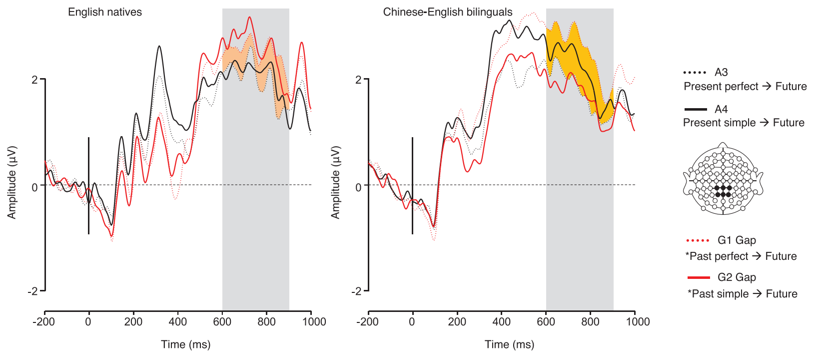

Figure 10 Event-related brain potentials in the present perfect → future acceptable (A3), present simple → future acceptable (A4), past perfect → future gap (G1), and past simple → future gap (G2) conditions in English native speakers and Chinese–English bilingual participants. Waveforms depict the linear derivation of signals recorded at six electrode sites (P1, P2, Pz, CP1, CPz, CP2). The highlighted time-window is 600–900 ms.

Time Gaps

For the analysis of time gaps, the main effect of condition was significant, $F(3, 102) = 3.72$, $p = .01$, $\eta_p{}^2 = .10$ (medium effect). As shown in Figure 10, there was no main effect of group, $F(1, 34) = 0.52$, $p = .48$, $\eta_p{}^2 = .02$ (small effect), but the interaction of group and condition was significant, $F(3, 102) = 2.70$, $p = .05$, $\eta_p{}^2 = .07$ (medium effect).[4] We further explored the main effect of condition via pairwise comparisons, with the p values corrected for multiple comparisons using the Bonferroni procedure (see Figure 9 and Appendix S2). Both the English native speakers and the Chinese–English bilinguals showed significant P600 amplitude differences between the past perfect gap condition and the present perfect acceptable condition ($ps \leq .05$). Finally, the Chinese–English bilinguals elicited more positive P600 amplitudes in the past perfect gap than in the past simple gap condition: past perfect → future gap versus past simple → future gap, $p = .02$.

Discussion

This study investigated timeline processing in Chinese learners of English reading English sentences in which tense conveyed temporal information. More specifically, we tested whether performance and patterns of N400 modulations elicited by semantic processing of event sequences involving either a time clash or a time gap would differ in native speakers of English and upper intermediate Chinese–English bilinguals.

Behaviorally, despite more variable performance within their group, the Chinese–English bilinguals performed overall similarly to their English peers and in fact out-performed them when dealing with present simple time clashes (present simple → past clash) and past perfect temporal gaps (past perfect → future gap). In stark contrast, electrophysiological results showed that the Chinese–English bilinguals failed to show the expected pattern of N400 modulation by time clashes found in English native speakers. In the case of time gaps, we found the reverse pattern, with the present perfect condition (present perfect → future acceptable) eliciting an N400 modulation in bilinguals when it should not have done so. However, in the P600 window, we found main effects of condition affecting both participants groups. In the case of clashes, this effect was driven by a P600 amplitude difference between the present simple clash (present simple → past clash) and the past simple acceptable condition (past simple → past acceptable). In the case of gaps, it was driven by a difference between the past perfect gap (past perfect → future gap) and the present perfect acceptable (present perfect → future acceptable) condition. Finally, we found that the Chinese–English bilinguals selectively elicited greater P600 amplitudes in the past perfect gap condition (past perfect → future gap) than the past simple gap condition (past simple → future gap).

In this study, we tested whether time clashes triggered by the critical verb would be accurately detected and whether this would be accompanied by a disruption of semantic processing. Even though they performed on a par with their English native speaker peers in terms of explicitly detecting temporal clashes and gaps, the Chinese learners of English failed to elicit the expected N400 modulations when responding to time clashes. Our findings present the first compelling evidence that upper intermediate Chinese learners of English are not able to extract temporal information online as they read English tensed forms, even though they can acquire the rules governing English tense and apply them explicitly.

In a previous study by Li et al. (2018), Chinese learners of English failed to detect time misalignments (i.e., time clashes similar to those tested here). However, the participants were instructed to focus on semantic violations induced by incongruous sentence endings, and their attention was not explicitly directed at temporal information. This may be why the English native speakers in that study also failed to behaviorally detect temporal clashes as semantic violations, even though these violations elicited expected N400 modulations. In this study, we explicitly instructed the participants to process the temporal ordering of events and thus pay direct attention to temporal information. While our English native speaker participants exhibited N400 modulations akin to those observed in Li et al.'s (2018) study, our Chinese–English bilinguals still showed no N400 modulation. This replication of modulation in native speakers across studies is consistent with the findings reported in Niewland's (2015) study in which participants generated more negative N400 amplitudes to false statements, regardless of whether they had been instructed to verify a statement explicitly or implicitly.

If our Chinese individuals considered the present perfect as belonging in the past time sphere, we could have expected that they would not have perceived the present perfect clash (present perfect → past clash) as a clash. In contrast, we would have expected N400 modulations in the present simple clash (present simple → past clash) condition. However, we did not observe such differences, even though the contrast between present perfect and present simple can be morphologically marked in Chinese (Figure 1). We contend that either: (a) the Chinese participants failed to understand the English material sufficiently or (b) they failed to integrate temporal information conveyed by tense as they read, that is, they experienced some kind of conceptual tense blindness. Account (a) is unlikely because the bilingual participants tested in our study were not only proficient in English but also performed almost on a par with their English L1 peers. Account (b) is more likely since tense does not exist in Chinese, and the Chinese participants did show a difference in N400 modulations between the present simple → future acceptable and the present perfect → future acceptable conditions.

While bilingual participants were insensitive to time clashes in the N400 range, P600 results suggest that they reevaluated present simple clashes (present simple → past clash) to a greater extent than past simple acceptable (past simple → past acceptable) sequences, like their native English peers. This may be a sign that the bilingual participants could use their grammatical knowledge of tense in order to correctly detect errors in the timeline (see behavioral results). Indeed, P600 or late positive component modulations have often been associated with overt/explicit detection of grammatical errors and reevaluation (Fu et al., 2017; Morgan-Short et al., 2022; Thierry et al., 2003; Voss & Paller, 2009) and controlled processing involving conscious awareness (Batterink & Neville, 2013; Hahne & Friederici, 1999). We thus speculate that, while they failed to implicitly detect time clashes in the window associated with semantic processing, the bilingual participants were able to detect them at a later stage on the basis of controlled reevaluation and explicit monitoring.

We also investigated how the same participants processed temporal gaps between past and future, bearing in mind that Chinese has a modal auxiliary *jiang* equivalent to *will* in English. As was the case for time clashes, the bilingual participants performed on a par with or better than English native speakers in behavioral judgements regarding time gaps. As we had expected, N400 amplitudes elicited in the time gap conditions (past perfect → future gap and past simple → future gap) were more negative than in the present perfect acceptable (present perfect → future acceptable) condition in the English native speaker participants. However, in the Chinese–English bilinguals, the difference between past perfect → future gap and present perfect → future acceptable was in the opposite direction to that observed in the English native speakers, and the present simple → future acceptable and present perfect → future acceptable conditions differed significantly, with more positive N400 amplitude in the present simple → future acceptable condition than in the present perfect → future acceptable condition. Such results are consistent with the idea that Chinese learners of English conceptualize the present perfect as belonging to the past rather than to the present time sphere when reading in English.

Exploratory P600 results for the time gap manipulation revealed significant amplitude modulation between the past perfect → future gap and present perfect → future acceptable conditions but this time in the same direction across groups. These results echo what was found for time clashes: The Chinese–English bilinguals were sensitive to the past perfect → future gap condition relative to the present perfect → future acceptable condition similarly to their native English peers. We speculate that the lack of sensitivity observed in the N400 range was temporary and gave way to late detection probably involving explicit monitoring mechanisms. This speculation is also consistent with the fact that the Chinese–English bilinguals tended to reevaluate the past perfect → future gap condition to a greater extent than the past simple → future gap condition, possibly because they assessed the time gap as being greater in the past perfect → future gap than in the past simple → future gap.

Limitations and Future Directions

One limitation of our study is that we could not recruit Chinese–English bilinguals with nativelike proficiency in English. This being said, we believe that our results cannot be attributed merely to a lack of proficiency in English in our bilingual participants, given:

- their upper-intermediate IELTS scores;

- their high self-reported proficiency;
- the fact that we selected them on the basis of their performance in the OQPT above 70%, $M = 79\%$, $SD = 0.05$, 95% CI [77, 82], which confirmed their upper-intermediate level;
- their performance in the online acceptability judgement task that was on a par or better than that of their English native peers;
- and their performance in a post hoc task conducted with a half of the experimental stimuli, in which they were asked to make temporal sequence acceptability judgements without a time constraint: bilinguals, $M = 85\%$, $SD = 0.1$, 95% CI [80, 91]; English native speakers, $M = 87\%$, $SD = 0.13$, 95% CI [80, 93]; $t(34) = 0.29$, $p = .78$, $d = 0.10$ (small effect).

Considering that our bilingual participants were highly accurate in their judgement of tense usage in context, it would be wrong to claim that they were entirely unaware of time sequence violations in the clash and gap conditions. And indeed, the results of the P600 analysis are consistent with late, controlled access to grammatical rules akin to that observed in English native speakers. We conclude, therefore, that the Chinese–English bilinguals were unable to extract semantic information conveyed by tense in real time when reading English sentences, which is not to say that they could not apply grammatical rules that they had learned and mastered in English. The latter processes, however, are likely to involve mechanisms different from semantic integration, occurring in a time window beyond that of verb form integration, that is, beyond 600 ms after stimulus onset (P600 range). There remains thus a possibility that Chinese–English bilinguals with very high (i.e., nativelike) proficiency in English could develop a semantically grounded ability to process English tense, but this will require a dedicated study.

It is also noteworthy that our study contained no filler trials. This means that the outcome might have been slightly different if only a subsample of the sentences presented had contained temporal clashes and gaps. However, given that the participants' attention was deliberately directed to temporal information and that we asked them to make explicit judgements, the inclusion of filler items would probably not have changed the outcome because the critical timeline manipulation was not hidden.

Overall, we can infer that Chinese learners of English conceptualize the prepresent zone operationalized by the present perfect in English as part of the past rather than the present time sphere. Indeed, a key temporal reference tool in Chinese is perfectiveness: When a perfective form indicates that an event is complete, it belongs to the past. The timeline in English is further blurred by the fact that temporal information can be conveyed flexibly in Chinese, using time adverbials, aspect markers, or context.

In this study, we found clear evidence that Chinese learners of English do not process the meaning of temporal information conveyed by tense online, that is, as they encounter a tensed verb during reading, despite their excellent command of English grammatical rules, presumably because tense does not exist in Chinese. This suggests that differences between English and Chinese regarding the encoding of temporal information, and particularly the distinction between recent past and past, do not conceptually overlap between speakers of the two languages, consistent with linguistic relativity effects observed on the basis of cross-language lexical and syntactic differences (e.g., Athanasopoulos & Bylund, 2013; Fausey & Boroditsky, 2011; Flecken et al., 2015; Thierry, 2016).

Conclusion

This study demonstrated that Chinese learners of English with an upper-intermediate level of proficiency in English and an excellent mastery of English grammar are conceptually tense-blind, that is, they are not able to process the meaning of temporal information conveyed by tense online as they read English. This is not to say that Chinese–English bilinguals are unable to detect tense misuse after they have encountered a verb, but such neurocognitive processes are likely to take place beyond the timeframe in which they process the tensed verb. We suggest that this is not due to insufficient proficiency since the Chinese learners of English made accurate explicit decisions about tense. Moreover, we provide evidence that shows, for the first time, that Chinese learners of English conceptualize the present perfect as part of the past time sphere, whereas English native speakers consider it as referring to the present time. We conclude that English tense represents a conceptual challenge for Chinese learners of English and that they are unlikely to adopt the same temporal framework as their English peers when resolving temporal information. Further studies are required to determine whether Chinese learners of English with high proficiency can access temporal information conveyed by tense in real time. According to our findings, access to temporal information in upper intermediate learners is likely to occur beyond the time frame of verb form processing. Another key question will be to determine whether there exists a critical period in the acquisition of English by Chinese learners after which they cannot process tense in a nativelike fashion.

Final revised version accepted 6 March 2023

Open Research Badges

This article has earned Open Data and Open Materials badges for making publicly available the digitally-shareable data and the components of the research methods needed to reproduce the reported procedure and results. All data and materials that the authors have used and have the right to share are available at https://osf.io/yn63x. All proprietary materials have been precisely identified in the manuscript.

Notes

1 After peer review of this article, we removed one participant's data upon that participant's request. This removal only resulted in one change in the pattern of statistical significance reported in the Results section (see Note 4).

2 The results of a mass univariate analysis of regular and irregular items failed to detect any difference in the English participant group while there were minor differences in the bilingual group at isolated electrode sites and in time windows that did not match any predicted variation (FT9: 372–592 ms; FT10: 408–620 ms; P1: 408–672 ms). We could not investigate potential differences between regular and irregular verbs between each individual condition because of a lack of power (with only 17 or 18 trials per condition).

3 As a reviewer requested, we also conducted correlation analyses of bilingual participants' proficiency and N400 modulations, on the one hand, and P600, on the other hand. We did not find any significant correlations (see Appendix S3 in the Supporting Information online).

4 Before we had to remove one participant's data (see Note 1), the interaction of group and condition did not reach significance.

References

Allen, D. (1992). *Oxford Quick Placement Test* (Oxford University Press).

Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: an ERP study. *Language and Cognitive Processes*, *18*, 405–430. https://doi.org/10.1080/01690960244000054

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science*, *37*(2), 286–309. https://doi.org/10.1111/cogs.12006

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137–153. https://doi.org/10.1162/0898929052880101

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, Article 328. https://doi.org/10.3389/fpsyg.2013.00328

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*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv. https://doi.org/10.48550/arXiv.1506.04967

Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed effects models using s4 classes (R package, Version 0.999375-28) [Computer software]. http://cran.r-project.org/web/packages/lme4/index.html

Batterink, L., & Neville, H. J. (2013). The human brain processes syntax in the absence of conscious awareness. *Journal of Neuroscience*, *33*(19), 8528–8533. https://doi.org/10.1523/JNEUROSCI.0618-13.2013

Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: An empirical review. *Neuroscience & Biobehavioral Review*, *51*, 31–47. https://doi.org/10.1016/j.neubiorev.2015.01.010

Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject–verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*(2), 161–174. https://doi.org/10.1017/S136672890700291X

Choudhary, K. K., Schlesewsky, M., Roehm, D., & Bornkessel-Schlesewsky, I. (2009). The N400 as a correlate of interpretively relevant linguistic rules: Evidence from Hindi. *Neuropsychologia*, *47*(13), 3012–3022. https://doi.org/10.1016/j.neuropsychologia.2009.05.009

Cohen, J. (1969). *Statistical power analysis for the behavioural sciences*. Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Erlbaum.

Comrie, B. (1985). *Tense*. Cambridge University Press. https://doi.org/10.1017/CBO9781139165815

Declerck, R., Reed, S., & Cappell, B. (2006). *The grammar of the English verb, I: The grammar of the English tense system: A comprehensive analysis*. Mouton de Gruyter. https://doi.org/10.1515/9783110199888

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eye-witness memory. *Psychonomic Bulletin & Review*, *18*, 150–157. https://doi.org/10.3758/s13423-010-0021-5

Flecken, M., Athanasopoulos, P., Kuipers, J. R., & Thierry, G. (2015). On the road to somewhere: Brain potentials reflect language effects on motion event perception. *Cognition*, *141*, 41–51. https://doi.org/10.1016/j.cognition.2015.04.006

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion

Fu, Q., Liu, Y.-J., Dienes, Z., Wu, J., Chen, W., & Fu, X. (2017). Neural correlates of subjective awareness for natural scene categorization of color photographs and line-drawings. *Frontiers in Psychology*, *8*, Article 210. https://doi.org/10.3389/fpsyg.2017.00210

Guajardo, L. F., & Wicha, N. Y. (2014). Morphosyntax can modulate the N400 component: Event related potentials to gender-marked post-nominal adjectives. *Neuroimage*, *91*, 262–272. https://doi.org/10.1016/j.neuroimage.2013.09.077

Hagoort, P., Wassenaar, M. E. D., & Brown, C. M. (2003). Syntax-related ERP effects in Dutch. *Cognitive Brain Research*, *16*(1), 38–50. https://doi.org/10.1016/S0926-6410(02)00208-2

Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience*, *11*(2), 194–205. https://doi.org/10.1162/089892999563328

Hahne, A., Mueller, J., & Clahsen, H. (2006). Morphological processing in a second language: Behavioral and event-related potential evidence for storage and decomposition. *Journal of Cognitive Neuroscience*, *18*(1), 121–134. https://doi.org/10.1162/089892906775250067

Hahne, A., & Friederici, A. D. (2001). Processing a second language: Late learners' comprehension mechanisms as revealed by event-related brain potential. *Bilingualism: Language and Cognition*, *4*(2), 123–141. https://doi.org/10.1017/S1366728901000232

Hawkins, R., & Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In R. van Hout, A. Hulk, F. Kuiken, & R. Towell (Eds.) *The lexicon-syntax interface in second language acquisition* (pp. 21–44). John Benjamins. https://doi.org/10.1075/lald.30

Hernandez, A., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, *133*(4), 1–13. https://doi.org/10.1037/0033-2909.133.4.638

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kutas, M., & Hillyard, S. A. (1980).Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, *207*, 203–205. https://doi.org/10.1126/science.7350657

Kutas, M., & Hillyard, S. A. (1984a). Brain potentials during reading reflect word expectancy and semantic assoication. *Nature*, *307*, 161–163. https://doi.org/10.1038/307161a0

Kutas, M., & Hillyard, S. A. (1984b). Event-related brain potentials (ERPs) elicited by novel stimuli during sentence processing. *Annals of the New York Academy of Sciences*, *425*, 236–241. https://doi.org/10.1111/j.1749-6632.1984.tb23540.x

Lardiere, D. (1998a). Case and tense in the ‘fossilized’ steady state. *Second Language Research*, *14*(1), 1–26. https://doi.org/10.1191/026765898674105303

Lardiere, D. (1998b). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research*, *14*(4), 359–375. https://doi.org/10.1191/026765898672500216

Lemhöfer, K., & Broersma, M., (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lenth, R. V. (2020). *emmeans: Estimated marginal means, aka least-squares means* (R package, Version 1.4.7) [Computer software]. https://CRAN.R-project.org/package=emmeans

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. University of California Press. https://doi.org/10.2307/603476

Li, Y., Jones, M., & Thierry, G. (2018). Timeline blurring in fluent Chinese–English bilinguals. *Brain Research*, *1701*, 93–102. https://doi.org/10.1016/j.brainres.2018.07.008

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, *8*, Article 213. https://doi.org/10.3389/fnhum.2014.00213

Morgan-Short, K., Finestrat, I., Luque, A., & Abugaber, D. (2022). Exploring new insights into explicit and implicit second language processing: Event-related potentials analyzed by source attribution. *Language Learning*, *72*(2), 365–411. https://doi.org/10.1111/lang.12492

Morris, J., & Holcomb, P. (2005). Event-related potentials to violations of inflectional verb morphology in English. *Cognitive Brain Research*, *25*(3), 963–981. https://doi.org/10.1016/j.cogbrainres.2005.09.021

Newman, A. J., Ullman, M. T., Pancheva, R., Waligura, D. L., & Neville, H. J. (2007). An ERP study of regular and irregular English past tense inflection. *Neuroimage*, *34*(1), 435–445. https://doi.org/10.1016/j.neuroimage.2006.09.007

Nieuwland, M. S. (2015). The truth before and after: Brain potentials reveal automatic activation of event knowledge during sentence comprehension. *Journal of Cognitive Neuroscience*, *27*(11), 2215–2228. https://doi.org/10.1162/jocn_a_00856

Osterhout, L., & Holcomb, P. J. (1992). Event-related potentials and syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785–804. https://doi.org/10.1016/0749-596X(92)90039-Z

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.

R Core Team. (2012). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org

Richardson, J.T.E. (2011). Eta squared and partial eta squared as measurements of effect size in educational research. *Educational Research Review*, *6*, 135–147. https://doi.org/10.1016/j.edurev.2010.12.001

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–430. https://doi.org/10.1177/0267658308090186

Steinhauer, K., & Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language*, *83*(1), 62–65. https://doi.org/10.1016/S0093-934X(02)00059-7

Steinhauer, K., White, E. J., & Drury, J. E. (2009). Temporal dynamics of late second language acquisition: Evidence from event-related brain potentials. *Second Language Research*, *25*(1), 13–41. https://doi.org/10.1177/0267658308098995

Thierry, G., Cardebat, D., & Démonet, J. (2003). Electrophysiological comparison of grammatical processing and semantic processing of single spoken nouns. *Cognitive Brain Research*, *17*(3), 535–547. https://doi.org/10.1016/S0926-6410(03)00168-X

Thierry, G. (2016). Neurolinguistic relativity: How language flexes human perception and cognition. *Language Learning*, *66*(3), 690–713. https://doi.org/10.1111/lang.12186

Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word-frequency in event-related brain potentials. *Memory & Cognition*, *18*, 380–393. https://doi.org/10.3758/BF03197127

Voss, J. L., & Paller, K. A. (2009). An electrophysiological signature of unconscious recognition memory. *Nature Neuroscience*, *12*, 349–355. https://doi.org/10.1038/nn.2260

Wang, W. S. Y., & Sun, C. (2015). *The Oxford handbook of Chinese linguistics*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199856336.001.0001

Wartenburger, I., Heekeren, H. R., Abutalebi, J., Cappa, S. F., Villringer, A., & Perani, D. (2003). Early setting of grammatical processing in bilingual brain. *Neuron*, *37*(1), 159–170. https://doi.org/10.1016/s0896-6273(02)01150-9

Weber-Fox, C., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, *8*(3), 231–256. https://doi.org/10.1162/jocn.1996.8.3.231

White, E. J., Genesee, F., & Steinhauer, K. (2012). Brain responses before and after intensive second language learning: Proficiency based changes and first language background effects in adult learners. *Plos One*, *7*(12), Article e52318. https://doi.org/10.1371/journal.pone.0052318

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Accessible Summary

Appendix S1. Analysis of Accuracy in R With Logistic Mixed-Effects Models.

Appendix S2. Event-Related Brain Potential Amplitude Analyses Using Repeated-Measures Analyses of Variance.

Appendix S3. Correlation Analyses Between Proficiency and Event-Related Brain Potential Mean Amplitudes.