

#### **Bangor University**

#### DOCTOR OF PHILOSOPHY

A three-dimensional approach to time conceptualization in Chinese-English bilinguals

Li, Yang

Award date: 2019

Awarding institution: Bangor University

Link to publication

General rights Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
You may not further distribute the material or use it for any profit-making activity or commercial gain
You may freely distribute the URL identifying the publication in the public portal ?

Take down policy If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Yr wyf drwy hyn yn datgan mai canlyniad fy ymchwil fy hun yw'r thesis hwn, ac eithrio lle nodir yn wahanol. Caiff ffynonellau eraill eu cydnabod gan droednodiadau yn rhoi cyfeiriadau eglur. Nid yw sylwedd y gwaith hwn wedi cael ei dderbyn o'r blaen ar gyfer unrhyw radd, ac nid yw'n cael ei gyflwyno ar yr un pryd mewn ymgeisiaeth am unrhyw radd oni bai ei fod, fel y cytunwyd gan y Brifysgol, am gymwysterau deuol cymeradwy.

I hereby declare that this thesis is the results of my own investigations, except where otherwise stated. All other sources are acknowledged by bibliographic references. This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree unless, as agreed by the University, for approved dual awards.

Yang LI

2019-09-20



# A three-dimensional approach to time conceptualization in Chinese-English bilinguals

Yang Li

School of Psychology

Bangor University

2019

This thesis is submitted in partial fulfilment for the degree of Doctor of Philosophy.

#### Acknowledgements

First and foremost, I would like to thank my supervisor, Guillaume Thierry. I especially want to thank him for accepting me as his PhD student at a time when I was under enormous pressure having decided to quit studying at York. He also offered me a scholarship from his own research funding after only one Skype session together. I would also like to thank Guillaume for sparking my interest in science, and also for fully supporting me over the last four years. I do not have enough words to express how grateful I am to Guillaume for his excellent supervision, from how to pronounce the word "fine" to how to write an academic paper. I am especially grateful to Guillaume for being very tolerant with my "bone-headedness", childish streak, and even my small rebellions. In addition, I would like to thank my Master's supervisor, Vivian Cook, for his help in developing my first psycholinguistic experiment, and for suggesting me to contact Guillaume when I decided to leave York.

I would also like to express my sincere thanks to Manon Jones, for her valuable help on my work and Gary Oppenheim for his patient explanation of statistical analyses. In addition, I would like to thank past and present members of the BULET ARIAN lab for their assistance with data collection. In particular, I would like to especially thank Cécile Barbet, for teaching me programming, data analysis skills, hands on experience with data acquisition, and for feeding me with delicious food. To Pippa Beston, I would like to show my appreciation for her encouragement and for being such a wonderful office mate. I have had immense pleasure collaborating with Aina Casaponsa, due to her wonderful suggestions for formulating papers, excellent statistical analysis skills, and her hospitality. Further, I would like to express my appreciation for my interns, my fellow students, and all the participants who kindly agreed to take part in my experiments.

Finally, I wish to express my gratitude and love to my parents, Shumin Li and Luzhen Fan. Without their help, support, and sacrifices, I would not have been able to get such an excellent education. In addition, I would like to thank my cousins for taking care of my parents when I was in UK. I would finally like to thank all my friends for their support.

#### Table of contents

ABSTRACT			
GENERAL INTRODUCTION AND OVERVIEW	4		
CHAPTER 1 REVIEW OF THE LITERATURE	6		
1.1 On linguistic relativity	7		
<ul><li>1.1.1 The pre-linguistic relativity hypothesis period</li><li>1.1.2 The Linguistic Relativity Hypothesis</li><li>1.1.3 Linguistic variations and possible implications</li></ul>	7 8 12		
1.2 On Bilingualism	25		
<ul> <li>1.2.1 The Revised Hierarchical Model</li> <li>1.2.2 The Sense Model</li> <li>1.2.3 The Distributed Representation Model</li> <li>1.2.4 BIA and BIA+ models</li> <li>1.2.5 Conclusion of section</li> </ul>	25 27 28 29 32		
1.3 EEG and Event-related potentials	32		
<ul><li>1.3.1 Principles of EEG</li><li>1.3.2 From EEG to ERPs</li><li>1.3.3 ERPs and linguistic relativity</li><li>1.3.4 Advantages and Disadvantages of ERPs</li></ul>	33 34 35 41		
1.4 Thesis Aims	42		
CHAPTER 2 TENSE AND LINGUISTIC RELATIVITY 2.1 Study 1	45		
Timeline blurring in fluent Chinese-English bilinguals	46		
<ul><li>2.1.1 Introduction</li><li>2.1.2 Materials and Methods</li><li>2.1.3 Results</li></ul>	47 50 54		
2.1.4 Discussion 2.1.5 Conclusion	60 64		

#### 2.2 Study 2

No r	eal-tin	ne semantic processing of tense in fluent Chinese-English bilinguals	65	
	2.2.1	Introduction	66	
	2.2.2	Materials and Methods	68	
	2.2.3	Results	72	
	2.2.4	Discussion	77	
	2.2.5	Conclusion	80	
СН	APTER	3 SPATIOTEMPORAL METAPHOR AND LINGUISTIC RELATIVITY	81	
3.1	Study	y 3		
Bac	k to th	e future? How Chinese-English bilinguals switch		
betv	ween f	ront and back orientation for time	82	
	3.1.1	Introduction	83	
	3.1.2	Materials and Methods	86	
	3.1.3	Results	91	
	3.1.4	Discussion	97	
	3.1.5	Conclusion	100	
3.2	Study	y 4		
Time	e flows	vertically in Chinese	101	
	3.2.1	Introduction	102	
	3.2.2	Materials and Methods	104	
	3.2.3	Results - Vertical experiment	106	
	3.2.4	Results - Horizontal experiment	108	
	3.2.5	Discussion	109	
СН	APTER	4 GENERAL DISCUSSION	112	
4.1	Gran	nmatical variations and time: The case of English tense	113	
42	lexic	cal variations and Time. The case of spatiotemporal metaphors	119	
4.3	Limit	ations and directions for future studies	124	
7.0	121		104	
	4.3.1 1 2 0	On metaphors	124	
	4.J.Z		120	
4.4	Final	Kemarks	129	
Refe	References			
App	Appendices			

#### Abstract

The linguistic relativity hypothesis proposes that the language we speak affects the way we perceive the world (Whorf, 1956). This hypothesis is at the origin of a vigorous debate amongst philosophers, linguists, psychologists, and more recently cognitive neuroscientists: Does language shape thought? In the work presented here, we have chosen to investigate language-time conceptualisation relationships in bilinguals whose two languages refer to time differently.

After reviewing key aspects of the literature and summarising recent findings of experiments that have tested linguistic relativity predictions, I present four experimental studies using event-related brain potentials to test cross-language grammatical and lexical influences on time conceptualisation in highly proficient Chinese-English bilinguals and native English controls.

In Study 1, we set out to characterise how native speakers of Mandarin Chinese –a tenseless language– deal with reference time misalignments conveyed by tense in English. Participants made acceptability judgements of sentences in which the adjunct clause started with the connective 'after' and was either temporally aligned or not with the tensed verb of the main clause. Chinese participants' data lacked an N400 ERP effect that we had observed in their English native peers, suggesting that they experience difficulties locating events on a timeline when temporal information is conveyed by tense.

In Study 2, we asked Chinese-English bilinguals to pay direct attention to temporal information conveyed by tense in order to test whether they have a different conceptualisation of the recent past when reading in English. For instance, the sentence "\*After she has worked in this hospital for ten years, she retired to Spain" features a time clash because an event situated in the past cannot follow a just-completed event. Although bilinguals could detect such clashes just like native English controls, the expected N400 modulation was entirely absent in their data. Despite their mastery of English grammatical rules, Chinese-English bilinguals thus failed to conceptually distinguish past and recent past on the basis of tense.

In a further two studies, we turned to vocabulary differences between Chinese and English to see whether specific expressions of Chinese can selectively shape time conceptualisation.

In Study 3, we focussed on Chinese spatiotemporal metaphors that are inconsistent with the future-infront convention of time representation. We tested temporal conceptualisation along the sagittal axis using stimuli –days of the week and years– played through loudspeakers that were physically located in the front or the back of participants. Brain activity revealed interference when space-time combinations clashed with spatiotemporal metaphors, but, surprisingly, only when participants operated in English, showing that time conceptualisation can selectively affect the second language.

In Study 4, we tested the controversial claim that Chinese organises time along the vertical axis. We adapted the arrow flanker task by replacing the middle arrow with spatiotemporal metaphors for months (e.g.,  $\pm \uparrow \exists$ -'last month', literally 'up month') and tested whether expected N400 modulations would generalise to purely temporal expressions (e.g.,  $\pm \mp$  –'last year', literally 'gone year'). N400 modulations elicited by spatial words (i.e., up / down) and spatiotemporal metaphors surprisingly generalised to non-directional temporal expressions of Chinese, providing strong evidence in favour of the claim that Chinese native speakers conceptualise time along the vertical axis.

Taken together, these results suggest that Chinese second language speakers of English do not process English tense at a semantic level as they read and that temporal expressions idiosyncratic to Chinese have a language-dependent impact on time conceptualisation in bilinguals. We conclude that an intuitive proposition put forward by Whorf in the 1940s may have been correct and possibly had an even stronger basis than he, himself, was prepared to admit.

#### **General introduction**

The linguistic relativity hypothesis, also known as the Sapir-Whorf hypothesis, proposes that the language we speak can affect the way we perceive the world (Whorf, 1956). Since the concept was developed by scholars such as Franz Boas (1858-1942), and later, Edward Sapir (1884-1939) and his mentee Benjamin Lee Whorf (1897-1941), linguistic relativity has attracted attention from various fields, such as psychology, linguistics, archaeology, and more recently cognitive neuroscience. Recent behavioural and neuroscientific studies have provided evidence that language affects several aspects of human perception and cognition, such as colour perception, object categorization, and event conceptualization. One domain of human cognition where language is likely to have a profound impact is that of temporal information processing. The ability to process time (e.g., relative time: future/past or time duration) is a vital human faculty. This thesis aims to provide novel scientific evidence regarding language-time cognition relationships in speakers of two languages, namely highly proficient Chinese-English bilinguals.

The first chapter provides an overview of three main components underpinning this PhD thesis: (i) a concise review of the field of research, namely linguistic relativity theory; (ii) a selective review of models of language representations in bilinguals; and (iii) a brief survey of the main method employed here, event-related potentials (ERPs), with a particular focus on the N400 ERP component, as the index of choice to quantify conceptual processing. Since it is at the core of the current doctoral research presented in this thesis, the literature review presented in this first chapter will particularly focus on temporal cognition.

The second chapter is the first experimental chapter of the thesis. There, I present two studies (1 & 2) designed to investigate how tense -a grammatical device- influences Chinese-English bilinguals' perception of time. In the following chapter, Chapter 3, I then turn to the question of how lexical variations between languages in terms of temporal expressions influence bilinguals' conceptualization of time. In Study 3, I focus on native spatiotemporal metaphors in an attempt to show how such language-specific spatial references for time can influence time conceptualization along the front-back axis. This is followed by an investigation of how native temporal expression can influence time conceptualization along the vertical and horizontal axis (Study 4). In the final chapter, Chapter 4, I provide a general discussion of the findings combining observations from the four experiments reported in this

PhD thesis and the overall implications for linguistic relativity theory. I also discuss limitations and ideas for future research directions.

Overall, I contend that this thesis provides novel, decisive evidence for the influence of cross-language grammatical and lexical diversity on human temporal cognition. Note that the focus of this work is not primarily to understand the relationship between the two languages of different script bilinguals (here Mandarin Chinese and English) but rather to understand the consequence at a conceptual level of having different verbal representations within the same mind, to communicate and manipulate time.

 $\infty$ 

# Chapter 1

**Review of the literature** 

# 1.1 On linguistic relativity

The research presented in this thesis is primarily concerned with interactions between language and conceptual representations in the human mind. Ever since classic philosophy in ancient Greece, with thinkers like Herodotus (~450 BC) and Aristotle (~350 BC), humans have been interrogating the relationship between language and thought. The consideration of this fundamental question was vigorously revived in the age of enlightenment with scholars like Wilhelm von Humboldt. In the 20<sup>th</sup> century, provocative theories were put forward by linguists like Edward Sapir and his student Benjamin Lee Whorf, leading to the posthumous crystallisation of the linguistic relativity hypothesis, namely that specific language forms (lexical, grammatical, semantic) influence or even determine the thought process in a specific fashion. In this thesis, I am fundamentally interested in idiosyncrasies of Chinese Mandarin compared to English, particularly those concerning time conceptualization, and how the grammatical and lexical features of each language differentially shape and modulate time conceptualization in individuals that master both these languages.

#### 1.1.1 The pre-linguistic relativity hypothesis period

Wilhelm von Humboldt (1767-1835) was the first scholar to suggest that languages reveal the scale of any difference between cultures. The main principle of Humboldt's *Weltanschauung* ('worldview') hypothesis is that extreme differences in the internal structure of language contribute to individual variance in worldview (von Humboldt, 1963). This hypothesis is regarded as an extremely strong version of the linguistic relativity hypothesis, since the *Weltanschauung* hypothesis is that language does not just constrain, but actually creates thought.

Another relatively rigid view of the relationship between language and thought was proposed by Edward Sapir (1884-1939), who suggested that language determines how speakers perceive and conceptualise the world (Sapir, 1949<sup>1</sup>).

"Language is a guide to 'social reality'. Though language is not ordinarily thought of as of essential interest to the students of social science, it powerfully conditions all our thinking about social problems and processes. Human beings

<sup>&</sup>lt;sup>1</sup> Edward Sapir died in 1939. The date stated here is the date of publication of the works from which the citation is extracted. This comment also applied to Benjamin Lee Whorf.

do not live in the objective world alone, nor alone in the world of social activity as ordinarily understood, but are very much at the mercy of the particular language which has become the medium of expression for their society. It is quite an illusion to imagine that one adjusts to reality essentially without the use of language and that language is merely an incidental means of solving specific problems of communication or reflection. The fact of the matter is that the 'real world' is to a large extent unconsciously built up on the language habits of the group. No two languages are ever sufficiently similar to be considered as representing the same social reality. The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached."

(Sapir, 1949, pp.162)

Sapir's perspective thus directly contradicted that espoused by his teacher, Boas, for whom language merely represented a constraint on thought that directs individual attention.

"It seems very questionable in how far the restriction of the use of certain grammatical forms can really be conceived as a hindrance in the formulation of generalised ideas. It seems much more likely that the lack of these forms is due to the lack of their need."

(Boas, 1911, pp.60)

#### 1.1.2 The Linguistic Relativity Hypothesis

The first scholarly observations of correspondences between specific structures of language and particular patterns of thinking were reported by Benjamin Lee Whorf in the early 1940s (1940/1956). Whorf studied Hopi, a Native American language, observing that it *"is seen to contain no words, grammatical forms, constructions or expressions that refer directly to what we call 'time', or to past, present or future, or to enduring or lasting"* (Whorf, 1956, pp. 57-58). Although Whorf admitted that native speakers of Hopi could express temporal situations, he suggested that linguistic constructions to express time were different to what Whorf called "Standard Average European" languages (e.g., *"I stayed five days*" in Standard Average European languages equates to *"I left on the fifth day"* in Hopi; Carroll, 1957, pp. 216). Like his mentor, Edward Sapir, Whorf thus argued for the existence of a strong link between language and thought, leading to the concept of linguistic determinism:

"From this fact proceeds what I have called the "linguistic relativity principles," which means, in informal terms, that users of markedly different grammars are pointed by their grammars toward different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers but must arrive at somewhat different views of the world."

(Whorf, 1956, pp. 221)

This strong Whorfian view was met with widespread scepticism. First, experimental support for linguistic relativity had produced contradictory findings. The most convincing evidence for the strong version of the linguistic relativity hypothesis is English native speakers' ability to process and detect colour, largely driven by the codability of colours in English terms even when native speakers of English are not verbalising (e.g., Brown & Lenneberg, 1954; Lantz & Stefflre, 1964, as cited in Gentner and Goldin-Meadow, 2003). However, Heider (1972, as cited in Gentner and Goldin-Meadow, 2003) found that the Dani people of New Guinea tended to having better recognition memory for focal colours than non-focal colours as compared to native speakers of English despite having only two colour terms in the Dani language. Therefore, Heider (1972) concluded that colour perception is driven by neurobiology rather than language.

Pinker (1995, pp. 63), a psychologist, attacked linguistic determinism, arguing that *"no matter how influential language may be, it would seem preposterous to a physiologist that it could reach down into the retina and rewire the ganglion cells"*. However, such an extreme claim was never made by proponents of linguistic relativity and the fact that Pinker's claim is likely true does little to progress the debate on linguistic relativity. In other words, Pinker's approach of the problem can be considered a strawman strategy (Thierry, 2016).

In the domain of theoretical linguistics, Chomsky's Universal Grammar (UG; 1986; 1995) theory famously postulated the existence of a universal basis for all human languages. If this is true, of course, the premises of linguistic relativity become immediately questionable, because a universal basis to language implies a universal basis to thinking. Thus, moving away from the linguistic determinism question, relativity is not very meaningful if one assumes that languages all have common grammatical representations. This being said, the proposition that all languages share a grammatical foundation does not exclude variations deriving from lexical diversity, for instance, and particularly the fact that some labels exist in some languages and not others (for a discussion, see Thierry, 2016).

One other fundamental assumption of UG that stands in direct contradiction with the principles of linguistic relativity, is that the language function is distinct from other aspects of human cognition, a theoretical framework which has become known as the modularity of mind theory (Fodor, 1983). Modularity assumes that the human mind can be decomposed into a set of innate and somewhat self-sufficient modules, language being one of them. If we take this perspective, language as a module is unlikely to ever influence other 'cognitive modules' to a great extent because it is functionally distinct from other cognitive processes

9

such as attention, memory, or perception. However, the modular view of the human mind has largely been abandoned by cognitive scientists given the extraordinary level of interactivity observed between different cognitive systems and the brain structures that underpin them. For example, verbal interference has been shown to affect memory and attention (and also perception, see Athanasopoulos & Bylund, 2013), and neuropsychological cases of complete dissociation (i.e., double dissociations) between two cognitive systems such as language and memory remain a very rare and questionable occurrence (Wang & Bellugi, 1994).

Actually, Whorf also provided a more interactive explanation of the relationship between language and thought. However, this account has mostly been ignored by critics.

"Any activations [of the] processes and linkages [which constitute] the structure of a particular language... Once incorporated into the brain [are] all linguistic patterning operations, and all entitled to be called thinking."

(Whorf, 1937, as cited in Lee, 1996, pp. 57-58)

*A priori*, the strong version of linguistic relativity may be seen as unconvincing, since research has shown that infants and animals can exhibit high-level cognitive function before any language skills are acquired or, in the case of animals, when they cannot be acquired (e.g., Gallistel, 1989; Hare, Call, & Tomasello, 2001; Phillips & Santos, 2007). However, the weak version of linguistic relativity, namely that language influences thought only in some circumstances, has been supported by the findings of some empirical experiments conducted in the late 90s (e.g., Davidoff, Davies, & Roberson, 1999; Heider & Olivier, 1972; Hermer-Vazquez, Spelke, & Katsnelson, 1999; Kay & Kempton, 1984).

In the 1990s, the advancement of theoretical debates and a new emphasis on experimental approaches, have led to a resurgence of interest in the linguistic relativity hypothesis (e.g., Lucy, 1996; Slobin, 1996). Lucy (1996), in particular, proposed that language and thought interact on three levels: (a) the semiotic level, i.e., the general impact of using any natural language; (b) the structure-centred level, i.e., how different morphosyntactic configurations in natural language can influence perceptions of reality; and (c) the functional level, which entails the influence of particular ways of speaking on thought.

Proponents of the '*thinking for speaking*' hypothesis, a more processing-oriented view of the relationship between language and thought, have suggested that "*there is a special kind of thinking intimately tied to language, namely, the thinking that is carried out, on-line, in the process of speaking*" (Slobin, 1996, pp. 75). During language production, speakers must pay

attention to language-specific features. To some degree, the thinking for speaking hypothesis and the weak version of linguistic relativity have some similarities: Language influences thought but only under specific conditions (Malt, Sloman, & Gennari, 2003). The key distinction between these two theories is that, in the thinking for speaking hypothesis, language-specific patterns shape thought only during on-line processing, whilst in the linguistic relativity hypothesis, language continuously influences thought, whether it is active or not (Lai, Rodriguez, & Narasimhan, 2014).

Recently, with the advent of neurobiological methods, such as event-related potentials (ERPs), and functional magnetic resonance imaging (fMRI), research has found that specific language representations affect higher level conceptual processing (i.e., semantics) and even basic perceptual processing (e.g., Boutonnet, Dering, Viñas-Guasch, & Thierry, 2013; Mo, Xu, Kay, & Tan, 2011; Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009). However, other research has suggested that this 'deep' effect is unstable and possibly illusory since it has been shown to be easily eliminated by verbal interference (e.g., Roberson & Davidoff, 2000; Winawer et al., 2007; Witzel & Gegenfurtner, 2011).

In order to resolve these opposite views, Lupyan (2012) proposed a language-thought interaction hypothesis, the 'Label-feedback' hypothesis, according to which, "*language produces transient modulation of ongoing perceptual (and higher-level) processing.*" (Lupyan, 2012, pp 4). For example, learning the word *ble* (dark blue in Greek) links distinctive perceptual and semantic features of the corresponding region of the colour space with the label. When the mental representation of the word *ble* is activated, perceptual features of the corresponding colour and objects associated with this colour are also activated (e.g., the sea; see **Fig. 1**). This co-activation would increase what Lupyan refers to as "*hybrid visuo-linguistic experience*": mastery of the colour word *ble* means that individual experience will be affected, and visual representations will in turn become more categorical. Lupyan has argued that this modulation process could be either increased by hearing the label ("*ble*") prior to seeing a visual display, or down-regulated, for instance through verbal interference.



**Figure 1.** Schematic of a neural network architecture to account for on-line effects of labels on perceptual representations. Adapted from Lupyan (2012) pp. 4.

#### 1.1.3 Linguistic variations and possible implications

Languages obviously differ from one another in many ways. Beyond the fact that words naturally differ from one language to another in terms of phonological and/or orthographic form, some words only exist in one language, and thus do not have a single word translation equivalent in another (e.g., *Torschlusspanik* in German; Thierry, 2019<sup>2</sup>). Such differences exist at all representational levels of language, i.e., lexical, grammatical, and semantic-conceptual. Here, I review a number of studies which have investigated differences in perception and cognition relating to such linguistic diversity. Special attention is given to studies in bilinguals since the question is not solely to understand how differences between languages result in cognitive differences between speakers of these languages but also how patterns of cognitive processing can change with second language learning.

#### 1.1.3.1 Lexical variations – Colour and object categories

*Colour perception.* One of the most-studied Whorfian effects in the lexical domain relates to colour terminology and how different labels for colour categories may influence colour perception. Whilst studies of colour categorisation have suggested that variations in colour terminology have a profound effect on basic aspects of colour perception (e.g., Kay & Kempton, 1984; Roberson & Davidoff, 2000; Roberson, Davidoff, Davies, & Shapiro, 2005), others have failed to find a significant effect (e.g., Heider, 1972; Heider & Olivier, 1972).

<sup>&</sup>lt;sup>2</sup> Cf. https://theconversation.com/the-power-of-language-we-translate-our-thoughts-into-words-but-words-also-affect-the-way-we-think-111801

More recently, Winawer et al. (2007) examined whether language could influence basic perceptual colour discrimination in native speakers of English and Russian. Unlike English, the Russian lexicon distinguishes lighter (*goluboj*) and darker (*sinij*) shades of blue in its lexicon. In the study by Winaver et al. (2007), participants presented with triads of blue squares varying in terms of lightness had to judge as quickly as possible which of the two bottom squares had the same colour to the top square. If the colours of the two bottom squares could be labelled differently in Russian, RTs were significantly faster than when both the shades of blue presented shared the same label. However, English native speakers did not display such categorical advantage, arguably because all 20 shades of blue used in the experiment fell within the same linguistic label in English, namely 'blue'. Interestingly, when the same task was accompanied by verbal interference, the categorical advantage in the Russian group disappeared. Winawer et al. (2007) argued that these results support the claim that lexical distinctions in colour terminology affects on-line colour perception.

Gilbert, Regier, Kay and Ivry (2006) provided further evidence for linguistic effects on colour perception using a well-controlled within-subjects design based on a lateralised colour discrimination task. On a given trial, participants were presented with a circular array of coloured squares in which one square had a slightly different colour (see **Fig. 2**).



Figure 2. Rings of colour patches adapted from Gilbert et al. (2006) pp. 490.

The task was to indicate whether the differently coloured square (target) was on the left or the right of the display as quickly as possible by pressing buttons set under participants' left and right index fingers. The colour of the target square was one of two shades of blue or two shades of green whilst the other squares were either within-category stimuli (same colour, different shade) or cross-category stimuli (different colour). Critically, the distance in hue between the colour of the target square and that of the within- and crosscategory stimuli was always the same. Participants were significantly faster in the betweencategory condition than the within category condition, but, only when the target square was presented in the right visual field (RVF). The authors interpreted this result as showing that variation in terminology influences colour perception since stimuli perceived in the RVF are primarily processed in the left hemisphere, which is dominant for language processing in a majority of right-handed individuals. Furthermore, this laterality effect was attenuated or even eliminated when participants were engaged in a verbal working memory task. This pattern of results was overall considered as strong evidence for the claim that lexical variation influences colour perception, even though RT benefits could also be attributed to post-perceptual processing. Indeed, participants' RTs are collected at the end of the processing stream, presumably well after perceptual analysis has finished. In any case, Gilbert et al. (2006) argued that their results provide support for the linguistic relativity hypothesis, regardless of the stage of neurocomputation concerned.

Using ERPs as a more direct, implicit index of visual perception, Thierry et al. (2009) tested the extent to which native speakers of Greek and native English controls automatically perceive the distinction between light and dark shades of blue compared to the same contrast in the green domain. Indeed, Greek has different labels for light blue (ghalazio) and dark blue (ble), whereas in English there is only one basic colour term for blue. Participants completed 4 blocks (two involving light and dark shades of blue and two involving light and dark shades of green) of an oddball shape discrimination task. Targets were square shapes, irrespective of colour (in fact colour was not mentioned at any point during the experimental session) and circle were to be ignored, although they afforded the critical colour change supposed to elicit a visual mismatch negativity (vMMN), an early pre-conscious modulation of the N1-P1 complex elicited by low probability visual events. Whereas vMMN effects were of comparable magnitude for blue and green contrasts in English controls, Thierry et al. (2009) found a predicted 3-way interaction between group, colour and deviancy, unpinned by the fact that the magnitude of the vMMN modulation for blues was greater in Greek than English participants whilst the green contrast elicited similar modulations between groups. Thus, Greek participants appear to consider colour contrasts crossing the *ghalazio-ble* boundary as more salient than their English peers, leading the authors to propose that variations in colour terminology not only influence behaviourally observable colour discrimination, but also early, pre-conscious phases of visual processing.

Athanasopoulos, Dering, Wiggett, Kuipers and Thierry (2010) expanded upon the conclusions of Thierry et al. (2009), by showing that length of immersion in the foreign

culture (and thus language) could shift pre-attentive colour perception. Greek participants from the study by Thierry et al. (2009) who stayed in the UK for less than one year showed greater vMMN for the blue than for the green contrast but not Greek participants who stayed longer than 18 months in the UK. Interestingly, the long-stay Greek-English bilinguals also distinguished less between- and within-category contrasts behaviourally. Overall, these results provided convincing evidence that exposure to another culture and language allows neural restructuring of core components of perception.

Mo et al. (2011) further proposed that Whorfian effects on early perceptual processing are confined to interactions within the left hemisphere of the human brain. In their experiment, participants were presented with an oddball stream of coloured squares presented in pairs, side-by-side, surrounding a fixation cross. Participants were asked to press a button if the fixation cross changed into a circle and were not instructed about colour. The squares had either a neutral green, a bluish green, a greenish blue, or a neutral blue colour. Mo et al. (2011) found a significantly greater vMMN in response to the between-category deviants (i.e., the low probability neutral green and greenish blue squares) compared with the within-category deviants (i.e., the low probability neutral green and bluish green squares) only when stimuli were presented in the RVF. These results bridged together findings from Gilbert et al. (2006) and Thierry et al. (2009) by showing that the left hemisphere of the brain, which is dominant for language processing, can modulate the activity of the visual system in early and pre-attentive stages of perceptual processing.

**Object categorisation**. To my knowledge, Gilbert, Regier, Kay and Ivry (2008) were the first to hypothesise that Whorfian effects found in the domain of colour perception should in theory also apply to higher level object categorisation. Using a visual search task, they investigated whether linguistic labels for object categories (here animals) would affect the detection of target animal silhouettes (e.g., of cats) presented amongst foils (e.g., silhouettes of dogs) in the RVF compared to the LVF. Circular arrays of 12 animal silhouettes surrounding a fixation point were presented on each trial. Participants were asked to report whether the target animal (i.e., a cat amongst dogs, or a dog amongst cats) was on the left or the right of the display by pressing designated buttons (as in Gilbert et al., 2006). Remarkably, participants needed less time to detect between- than within-category targets, only when the target appeared in the RVF, suggesting that effects observed in the domain of colour perception generalise to the domain of object categorisation. Moreover, the RT

difference was eliminated when participants performed a secondary verbal working-memory task in parallel, but not when they performed a nonverbal task. These results suggest that Whorfian effects linking left-hemispheric lexical representations to categorical perception in the domain of colour generalise to the case of object categories.

Boutonnet et al. (2013) were the first to attempt measuring implicit Whorfian effects in object perception using ERPs. In English, the objects cup and mug each have their own label, whereas both these objects are called *taza* in Spanish. In two block of trials, native speakers of Spanish and English native controls were invited to detect the occurrence of a bowl within a stream of images of cups and mugs, in which cups and mugs occurred either frequently (standards) or rarely (deviants). In one block, cups served as standards and mugs as deviants and the reverse scenario applied to the other block, whilst bowl were systematically the target stimulus. Deviant stimuli elicited a larger deviant-related negativity (DRN, equivalent to the vMMN when stimuli are presented at fixation) in the English native participants as compared to the Spanish participants who failed to discriminate between cups and mugs in the DRN range. These findings support the claim that lexical labels affect unconscious object discrimination during visual perception.

As in the domain of colour perception, studies based on sociolinguistic observations have also shown the effect of language on object perception is malleable, and that L2 exposure can progressively shift object categorisation (e.g., Ameel, Malt, Storms, & Van Assche, 2009; Pavlenko & Malt, 2011). Results from the above studies on colour and object categorisation provide a compelling account of linguistic relativity because they are close to early, automatic, and unconscious phases of perceptual processing, thereby offering support for a strong account of Whorfian effects.

#### 1.1.3.2 Grammatical variations – Aspect and Gender

Aspect. Whorf (1940/1956, pp.221) argued that "users of markedly different grammars are pointed by their grammars toward different types of observations and different evaluations of externally similar acts of observation." In studies examining grammar, one of the most frequent questions is whether variation across languages in aspect marker use influences event conceptualisation. Most previous experiments have established that aspect marking influences individuals' description of an event: The speakers of languages that mark aspect (e.g., English) are less likely to mention the goal of an event, while speakers of languages

that have no aspect marking (e.g., German) are more likely to refer directly to the goal of action (e.g., Athanasopoulos & Bylund, 2013; Athanasopoulos et al., 2015; Schmiedtová, von Stutterheim, & Carroll, 2011).

Athanasopoulos and Bylund (2013) investigated whether the existence of aspect marking can affect nonverbal cognitive processing of events. In four behavioural experiments, the authors tested whether aspect would contribute to differences between speakers of English and Swedish in terms of memory and cognition. English tends to mark aspect in its grammar (as in, "a person is walking", i.e., the progressive), while Swedish does not mark grammatical aspect and thus "a person is walking" and "a person walks" translate in the same form in Swedish. This cross-linguistic difference in aspect marking has been related to a differential focus on endpoints when individuals consider motion event. For instance, because English has the progressive, speakers of English tend to focus more on the ongoingness of an event rather than its endpoint. In contrast, since Swedish does not mark the progressive, there is a greater tendency to consider the event in a holistic fashion and thus put more emphasis on the endpoint of a motion event (as in, *en person går till ett hus* – 'a person walks to a house').

Both English and Swedish participants were first asked to describe video clips depicting certain actions in their native language. As expected, Swedish speakers were more likely to mention event endpoints than their English native peers. Then, participants took part in a memory-based ABX task in which they had to judge whether a target video clip (X) depicting a motion event was more similar to video clip A or to video clip B, presented previously. In all trials either video clip A or B depicted an ongoing event without clear completion, and the other depicted an event in which the action reached its endpoint or the movement was completed. In other words, video clips A and B differed strongly as to whether they were ongoing or completed. Critically, video clip X was ambiguous in terms of its completion status, that is, it was just half-way between unfolding and completion, so as to reveal participants preference for ongoingness or completion that might have been driven by grammar.

Results from the sequential ABX task showed that Swedish native speakers had a greater preference for clips mentioning endpoints than English native speakers. However, if video clip A, B and X appeared together, the effect was eliminated, showing that the effect did not survive when no memorization was required. In other words, linguistic encoding strategies might affect motion event conceptualisation but effects of grammar might not be

17

found during on-line processing. In addition, the effect was also eliminated if participants engaged in a verbal interference task (e.g., counting backwards silently from 100 every 3 digits). These findings are consistent with claims that suggest language can shape other cognitive processes, here memory. Together the results from the four experiments reported in Athanasopoulos and Bylund (2013) provide support for the basic claim underlying dual coding theory, namely that bilinguals have two separate language systems, one for each language.

A later study by Athanasopoulos et al. (2015) provided evidence for the claim that Whorfian effects on motion event processing are context-bound and transient. Employing the same memory-based task as in the study by Athanasopoulos and Bylund (2013), the authors investigated whether German-English bilinguals show different responses depending on language context and verbal interference. As expected, bilinguals had a higher preference toward motion-completion choice, similarly to German monolinguals, when they were tested in a German context. However, when they were tested in an English context, they behaved more similarly to English monolinguals. In addition, when bilinguals received verbal interference in German, their motion categorisation preference dropped to a pattern more consistent with that of English monolinguals. When the verbal interference was in English, however, their preference patterned more with that of German monolinguals. This suggests that verbal interference blocks access to conceptual representation of the active language (the language in which the verbal interference task is executed) rather to the language faculty overall. In addition, these results also indicate that conceptual properties of the non-active language can be activated during linguistic encoding.

To my knowledge, the first study investigating the effects of grammatical aspect on motion event perception using ERPs was conducted by Flecken, Athanasopoulos, Kuipers and Thierry (2015). In that study, both English native speakers and German native speakers were asked to take part in an event-matching task in which they viewed one-second video animations depicting a dot moving along a trajectory towards a geometrical shape, followed by a target picture. There were four different animations: matching the target picture both in terms of the trajectory and endpoint (full match condition), matching the target picture in terms of trajectory but not endpoint (trajectory match condition), matching the target picture in terms of endpoint but not trajectory (endpoint match condition), and matching neither properties (mismatch condition; see **Fig. 3**).

18



Figure 3. Examples of target pictures following animated primes in Flecken et al. (2015. pp.45).

German participants generated a greater P3 in the endpoint match than the trajectory match condition, whereas English native speakers showed no difference between these conditions. This result suggests that German speakers pay greater attention to endpoint than the trajectory, as predicted by their grammar (Athanasopoulos et al., 2015) while English natives tend to pay equal attention to both properties. It is noteworthy that German and English participants differ in terms of response time in a behavioural motion matching task using the same stimuli as those used in the ERP task. Therefore, differences observed in the ERP task could not be merely due to motion event verbalisation. Together with the findings from Athanasopoulos and Bylund (2013) and Athanasopoulos et al. (2015), these ERP results suggest that aspect marking of the native language can unconsciously influence individuals' motion event perception, conceptualisation and memorization.

*Gender*. Gender is a grammatical feature that mandatorily assigns nouns to two or three classes. For example, Italian, Spanish and French have a two-gender system, masculine and feminine. Some languages do not mark gender at all except in relation to semantic gender, like Chinese or English<sup>3</sup>, whilst other have three levels of grammatical gender distinction like German. Gender marking thus provides a natural test bed for the premises of linguistic relativity given the ease of comparing highly contrastive situations in language pairs that are relatively common. Experimental studies have shown that grammatical gender can affect mental representations at both the semantic and conceptual levels (e.g, Boroditsky, 2000; Boroditsky, Schmidt, & Phillips, 2003; Konishi, 1993; Phillips & Boroditsky, 2003). However, other research has suggested that the effect of gender markings in cognition is very limited (e.g., Kousta, Vinson, & Vigliocco, 2008; Sera et al., 2002; Vigliocco, Vinson, Paganelli, & Dworzynski, 2005). For instance, Kousta et al. (2008) showed that Whorfian

<sup>&</sup>lt;sup>3</sup> There are few notable exceptions in English for historical and cultural reasons, such as car, ship, or plane all of whom are optionally feminine in English.

effects are found in some experiments when participants resort to using grammatical gender strategically to complete particular tasks, but not when access to grammatical gender is optional.

The most convincing evidence relating to conceptual interference of grammatical gender was perhaps provided by Cubelli, Paolieri, Lotto and Job (2011), who detected Whorfian effects even when minimising the likelihood of using grammatical gender strategically. In this experiment, both English native speakers and Italian native speakers were presented with two pictures and invited to judge whether they were from the same semantic category. Picture pairs where equally distributed between semantic relatedness and gender congruency contrasts: 50% were semantic-related and 50% unrelated, 50% had a congruent gender, and 50% had incongruent genders. Although the task did not explicitly mention grammatical gender, Italian participants' performance was influenced both by semantic and grammatical gender manipulations. When grammatical gender was congruent between picture labels, participants had shorter reaction time (RTs) for both semantically related and unrelated pairs. In contrast, English native speakers, whose mother tongue does not mark gender in its grammar, only showed a semantic effect. These results suggest that grammatical gender interferes with semantic processing, and either speeds-up semantic processing for gender congruent pairs or hinders semantic processing for incongruent pairs. In a second experiment, Italian participants were tested alongside Spanish native speakers whose mother tongue also marks gender in its grammar, and gender congruency was manipulated in picture pairs to be overlapping or mismatched across languages. For example, the picture pair "nose-eye" is congruent in terms of gender in Italian [naso-occhio - 'nostrileye', both masculine] but incongruent in Spanish [nariz-ojo, nariz is feminine and ojo is masculine]. Both Italian and Spanish speakers demonstrated a significant effect of gender congruency and semantic relatedness but the two groups differed based on the gender of object labels in their respective language. Remarkably, when Spanish participants engaged in verbal interference, the gender effect was eliminated, but semantic relatedness effects remain measurable. Therefore, Cubelli et al. (2011) argued that grammatical gender does not affect conceptual object representations but facilitates the processing of object labels from congruent gender categories.

Boutonnet et al. (2012) further examined the effect of grammatical gender on object categorisation using ERPs. Spanish-English bilinguals and English native speakers were asked to complete a semantic categorisation task, indicating whether the last picture in a

20

triplet presented sequentially belonged to the same semantic category as the first two. Picture series belonged to four conditions: (a) gender congruent and semantically related (e.g., tomato, celery, asparagus); (b) gender incongruent and semantically related (e.g., tomato, celery, carrot); (c) gender congruent and semantically unrelated (e.g., tomato, celery, truck); and (d) gender incongruent and semantically unrelated (e.g., tomato, celery, bike). Unlike Cubelli et al. (2011), Boutonnet et al. (2012) found no behavioural difference between conditions. However, both English and Spanish participants showed a semantic priming effect on N400 amplitudes. In addition, Spanish participants had a more negative Left Anterior Negativity (LAN) in the gender incongruent condition suggesting that they automatically activated grammatical gender representations of the picture names even when such information was not required to perform the task.

In this section, we have reviewed studies showing that aspect and grammatical gender convincingly yield linguistic relativity effects. Whilst studies have shown similar effects with grammatical number marking (e.g., the plural; Athanasopoulos, 2006; Lucy, 1992), other syntactic properties of language also presumably affect non-language cognition, whether at an overt, explicit level of at a more unconscious, implicit level of mental representation.

#### 1.1.3.3 Semantic variations – The case of spatiotemporal metaphors

Despite the universal relevance of temporal cognition in humans, cultures vary considerably in how they verbalise and express temporal information. Humans often use spatial metaphors to convey time duration or relative temporal references (Alverson, 1994; Clark, 1973; Traugott, 1978), i.e., they employ spatiotemporal metaphors (e.g., 'let's bring the meeting *forward* to Tuesday' or 'I look *forward* to defending my PhD viva in a month's time'). Furthermore, it has been proposed that spatiotemporal metaphor diversity across languages has an impact on time conceptualisation.

In most languages, time is mostly represented along the *sagittal* axis (front-back). Whereas speakers of English generally conceptualise the future as orientated towards the front space with the past behind them (e.g., Boroditsky, 2001; Boroditsky & Ramscar, 2002; Casasanto & Jasmin, 2012; Lai & Boroditsky, 2013), speakers of Aymara use the same word *nayra* for 'past' and 'front' and the word *qhipa* for both 'future' and 'back'. Accordingly, speakers of Aymara conceptualise the future as being orientated toward the back space and the past in the front (Núñez & Sweetser, 2006).

In a seminal study, Boroditsky (2001) used pictures depicting spatial arrangements of objects as primes for temporal judgements. The spatial arrangements presented in the picture primes could be either sagittal (e.g., "The black worm is ahead of the white worm.") or vertical (e.g., "The black ball is above the white ball."). After viewing each picture prime, participants were required to make an acceptability judgment regarding a temporal statement (e.g., "March comes before April" or "March comes earlier than April"). In this experiment, English native speakers needed less time to respond to the statement "March comes before April" after horizontal than vertical primes. In contrast, Chinese native speakers were quicker to respond to "March comes earlier than April" after vertical primes. Thus, participants appeared to conceptualise time depending on their native language, namely sagitally in English and vertically in Chinese. Beyond issues with replication of these results (Chen, 2007; January & Kako, 2007; Tse & Altarriba, 2008), it must be noted that the behavioural evidence from priming obtained is limited to overt, conscious stimulus evaluation and thus provides little insights regarding the conceptual basis for temporal representation.

Ng, Goh, Yap, Tse and So (2017) presented Chinese and English native participants with videoclips depicting a gesture (someone pointing to the front or the back space) followed by an auditorily presented word and asked them to make a temporal judgement: "Does the word refer to the future or the past?". English native speakers required relatively less time to respond to a temporal word congruent with gesture orientation; i.e., when future-related words were presented with forward-pointing gestures, and past-related words were coupled with backward-pointing gestures. However, Chinese participants did not show any congruency effect, possibly because the future and past are not systematically mapped with front and back space in these individuals, as is implied by the existence of spatiotemporal metaphors of Chinese that are incompatible with the general future-in-front orientation of the timeline (**Table 1**).

Table 1 - Examples of spatiotemporal metaphors in Mandarin Chinese that formed the core of the manipulation	on
in Study 3 of the present thesis (see Chapter 3.1).	

	Pin Yin	Literal translation	English translation
前年	qian-nian	front year	the year before last
后年	hou-nian	back year	the year after next
前天	qian-tian	front day	the day before yesterday
后天	hou-tian	back day	the day after tomorrow

Spatial representation of time in the horizontal plane also involves the *transversal* (left-right) axis in languages such as English, German, French or Chinese, but the vertical (up-down) axis is much less frequently used. Fuhrman et al. (2011) conducted research in which an experimenter stood next to a participant and pointed to a spot directly in front of them, asking "If this here is today, where would you put yesterday?". Findings of this study showed that Mandarin speakers not only arrange time from left-to-right on the transversal axis but also from top-to-bottom along the vertical axis, whereas English native speakers predominantly arrange time from left-to-right. Other experiments have employed pictures presented left-to-right or top-to-bottom and depicting the unfolding of time (e.g., a person growing from childhood to old age or a fruit being progressively cut into pieces) to examine whether participants tend to represent time along the transversal or the vertical axis (e.g., Chen, 2007; January & Kako, 2007; Miles, Tan, Noble, Lumsden, & Macrae, 2011; Tse & Altarriba, 2008; Yang & Sun, 2016). Although Mandarin Chinese features vertical metaphor to express time (e.g.,  $\pm \uparrow \beta$ -'last month'[literally 'up month'],  $\neg \uparrow \gamma \beta$ -'next month', [literally 'down month']), native speakers of Chinese performed similarly to their native English peers in these tasks, displaying congruency effects only along the traversal axis. Based on slightly different results, Miles et al. (2011) and Yang and Sun (2016) argued that Chinese speakers can represent time along the vertical axis but also along the transverse axis.

Previous research has found that the conceptualization of time is influenced by language-specific spatial-temporal metaphors (Boroditsky, 2001; Casasanto et al., 2004; Fuhrman et al., 2011; Lai & Boroditsky, 2013; Núñez & Sweetser, 2006); direction of writing and reading (Bergen & Lau, 2012; Ouellet, Santiago, Israeli, & Gabay, 2010; Tversky, Kugelmass, & Winter, 1991) and the task at hand (Boroditsky, 2001; Boroditsky & Ramscar, 2002; Lai & Boroditsky, 2013). In line with similar considerations regarding the linguistic relativity hypothesis in general, Boroditsky (2001) proposed that spatial-temporal metaphors can be considered to affect time conceptualisation to a lesser or greater extent, depending on spatial schema activation. The effect would be strong if spatial schemas (concrete domain) are systematically activated to conceptualise time (abstract domain). Alternatively, it would be weak if spatial metaphors influence time conceptualisation but do not determine it. Casasanto and Boroditsky (2008) further proposed an intermediate view, the Integrated Metaphoric Structuring View, and pointed out that both experiential correlation and linguistic metaphors are involved in the development of conceptual mappings of time. To my knowledge only one experimental study on linguistic relativity has considered effects of language change on time conceptualisation in bilinguals, looking at duration. Bylund and Athanasopoulos (2017) investigated duration estimation ability in speakers of Swedish and Spanish. In Swedish, duration is mostly expressed in terms of length ("long/short time") whereas in Spanish it is mostly expressed as quantity ("much/small time"). Accordingly, the authors observed that native speakers of Swedish made larger errors of duration estimation when stimuli varied in length, whilst Spanish participants made greater errors in duration estimation when stimuli varied in terms of volume. When Spanish-Swedish bilinguals engaged in the same tasks their behaviour changed based on a language cue presented at the beginning of each trial: They behaved in a 'Swedish' length-based way when presented with a Swedish cue, and in a 'Spanish' quantity-based way when presented with a Spanish cue. Bylund and Athanasopoulos (2017) concluded that time conceptualisation can be modulated online by language context.

#### 1.1.3.4 Conclusion of section

Overall, I have shown that a number of behavioural and electrophysiological studies investigating grammatical, lexical, and semantic diversity across languages hint at perceptual and cognitive effects. Beyond the fact that the evidence is peicemeal and sometimes unconvincing because it is unstable and mostly derived from behaviorual measures, many oustanding questions remain: How specific are the effects of language on time conceptualisation? Is it the case that speakers of different languages conceptualise time differently just because they are from different cultures, have different history, live in a different languages deeply driven by *linguistic* differences? How flexible is the cognition of bilinguals and how much change in cognitive processing should we expect when a bilingual switches between language contexts? In other words, in this PhD thesis, I aim to understand the links between specific constructs of particular languages (here Mandarin Chinese and British English) and thinking about time in the broad sense, rather than studying a constellation of human differences between cultures on the same process, and I aim to better characterise cognitive flexibility in Chinese-English bilinguals.

## 1.2 On Bilingualism

I chose to focus my study of language-thought relationships in Chinese-English bilinguals because it is interesting to understand not only how one language can shape thought patterns, but even more so to see whether a change of language leads to changes in conceptualisation. Testing bilinguals and intepreting the resulting observations in relation to linguistic relativity requires a good understanding of how the bilinguals' two languages interact with one another at lexical and semantic levels. Over the past 25 years, several models of word processing in bilinguals have been put forward to account for effects reported in the litterature. Here, I selectively review some aspects of main stream models of bilingual word processing in order to evaluate the extent to which results from our empirical studies are compatible with their assumptions.

It is now well-stablished that lexical access in bilinguals is overwhelmingly language non-selective (Dimitropoulou, Duñabeitia, & Carreiras, 2011a, 2011b; Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Duyck & Warlop, 2009; Hoshino & Thierry, 2012; Ng & Wicha, 2013) and operates within an integrated lexicon (Dijkstra & van Heuven, 2002; Duyck et al., 2007; Grainger & Dijkstra, 1992; Grainger, Midgley, & Holcomb, 2010; Kroll, van Hell, Tokowicz, & Green, 2010). In other words, there are automatic competition effects within and across languages at the lexical level, even when bilinguals operate in a monolingual language context (Spalek, Hoshino, Wu, Damian, & Thierry, 2014; Thierry & Wu, 2004, 2007; Wu & Thierry, 2010, 2012). It is worth noting however, that evidence in favour or against a unified semantic system where semantic and conceptual representations linked to a given language also compete for selection in a language non-specific fashion is sparse. Here, I review the most relevant theoretical models underpinning bilingual language processing, primarily focussing on word comprehension, including some of their limitations that offer insights into the workings of the bilingual mind.

#### 1.2.1 The Revised Hierarchical Model

The revised hierarchical model (also known as RHM; Kroll & Stewart, 1994) was developed in the early 90's to account for asymmetric patterns observed in translation production. It is also a well-known attempt to capture the way in which varying degrees of L2 proficiency modulate access to conceptual representations. The RHM proposes the existence of two separate lexicons interconnected via lexical links, the strength of which varies with increasing levels of L2 proficiency. Kroll and Stewart (1994) suggested that at lower levels of proficiency, the strength of the connections of L2 lexical representations with concepts is weaker than with L1 translation equivalents. Hence, they argued that access to conceptual representations when proficiency is low is mediated by L1 lexical representations. As proficiency increases, the strength of links between L2 lexical representations and concepts increases, granting direct access to meaning from L2 words. RHM accounts for low proficiency bilinguals' longer translation time from L1 to L2 than from L2 to L1 (e.g., Kroll, Michael, Tokowicz, & Dufour, 2002). Under these circumstances, L1 is the strongest route to access meaning (see **Fig. 4**).



Figure 4. The Revised Hierarchical Model adapted from Kroll & Stewart (1994) pp.158.

The RHM specifically represents L1 and L2 lexicons as separate, and a semantic store accessible to both the languages, albeit not with the same efficiency. In the model, lexical associations from L2 to L1 (solid line) are stronger than those from L1 to L2 (dotted line), as late bilinguals initially tend to acquire the meaning of L2 words through translation. Similarly, the link between L1 and conceptual memory (solid line) is hypothesized as stronger than the link between L2 and conceptual memory (dotted line). In the initial stages of L2 acquisition, bilinguals establish a lexical link between second and first language words, thus building a strong association with conceptual memory via the native language (L2 to L1 to concepts). As proficiency increases, they establish a more direct link between L2 words and conceptual links.

The RHM has been criticised on several counts: First, for assuming two separate lexicons for L1 and L2 (Brysbaert & Duyck, 2010), which is incompatible with the now large body of evidence in favour of language non-selective access in bilinguals (e.g., Dijkstra, 2005; Schwartz & Kroll, 2006; Thierry & Wu, 2007; van Hell & de Groot, 2008). Whilst Brysbaert and Duyck (2010) have argued that parallel access within an integrated lexicon (see also, Dijkstra & van Heuven, 2002) can provide a better explanation of the current data regarding bilingual word recognition, Kroll et al. (2010) suggested that parallel lexical access might not require an integrated lexicon and still accounts well for data from the production literature. Second, RHM failed to account for the existence of translation ambiguity (e.g., Eddington & Tokowicz, 2013; Tokowicz & Kroll, 2007) as well as empirical findings from visual word recognition using the masked translation priming paradigm (e.g., Duñabeitia, Perea, & Carreiras, 2009; Perea, Dunabeitia, & Carreiras, 2008), suggesting stronger lexical access from L1 to L2 than L2 to L1.

A third criticism, which is important in the context of the present thesis, is whether bilinguals' semantic store should be considered a unitary store, with unified representations for both languages. Indeed, the RHM does not make a distinction between languagedependent and language-independent concepts (Brysbaert & Duyck, 2010).

#### 1.2.2 The Sense Model

In contrast with the RHM, the Sense model of bilingual word comprehension (Finkbeiner, Forster, Nicol, & Nakamura, 2004) proposes a semantic system in which the sharing of representations is based on overlap between word senses. This proposal provides an account for the observation that the proportion of shared word senses across languages determines the level of priming between semantically related words. As an example, given by Finkbeiner et al. (2004), *black* in English and *kuroi* in Japanese are translation equivalents when referring to colour. However, "black" has another 20 meanings in English, which are not shared with *kuroi* in Japanese, and *kuroi*, reciprocally, has several meanings that are not shared with "black" (see Fig. 5).

Since bilinguals are generally more proficient in their native language than their second language, the Sense model assumes that they store more semantic senses associated with L1 words than L2 words. Accordingly, the Sense model predicts that L1 can prime a large proportion of L2 concepts in the L1-L2 direction, while the amount of L1 concepts

primed by the L2 is limited. Such predictions are supported by empirical findings from a large number of studies (e.g., de Groot & Nas, 1991; Jiang, 1999; Jiang & Forster, 2001; Keatley, Spinks, & de Gelder, 1994). In order to accommodate the result that L2 primes can effectively prime L1 targets in a semantic categorization task, the Sense model assumes that the task at hand serves as a filter to screen out category-irrelevant senses. Thus, even when L2-L1 priming is weak, the meaning that is relevant for the task is still accessed by L2 primes. It must be kept in mind, however, that the Sense model was constructed to account for results from masked priming experiments and that patterns of semantic priming are modulated by the task at hand (i.e., whether it focuses on semantic or lexical priming).



Figure 5. The Sense Model adapted from Finkbeiner et al. (2004), pp.9

#### 1.2.3 The Distributed Representation Model

The Distributed Representation Model (DRM, Duyck & Brysbaert, 2004) is an attempt to account for the classic asymmetry in semantic and translation priming in bilingual word comprehension from a connectionist perspective, namely the fact that L1 to L2 priming is usually much stronger than the reverse. The DRM proposes that L2 words activate only some semantic features as compared to L1 translation equivalents (Schoonbaert, Duyck, Brysbaert, & Hartsuiker, 2009).

One of the key assumptions of the DRM is that the greater the number of semantic features shared between L1 and L2, the shorter the translation times and the larger the priming effects. Interestingly, this assumption is consistent with those of the Sense model. Schoonbaert et al. (2009) further extended the DRM by proposing that semantic

representations are much richer in L1 than L2 in unbalanced bilinguals. L2 primes can thus only activate a subset of L1 conceptual nodes, and therefore L2 primes cannot activate more semantic nodes than L1 primes (see **Fig. 6**). Moreover, a translation equivalent prime can activate more conceptual nodes than a semantically related prime and thus accesses richer conceptual representations, leading to translation priming being stronger than semantic priming. Finally, concrete primes are considered to activate more shared conceptual nodes than abstract primes. In other words, the DRM proposes non-overlapping semantic representations between a bilinguals' languages based on semantic feature distributions and accounts for lexical priming effect through semantic mediation.



Figure 6. The Distributed Representation Model adapted from Schoonbaert et al. (2009), pp.581

#### 1.2.4 BIA and BIA+ models

The Bilingual Interactive Activation (BIA) model (van Heuven, Dijkstra, & Grainger, 1998) and its successor, the Bilingual Interactive Activation Plus (BIA+) model (Dijkstra & van Heuven, 2002; van Heuven & Dijkstra, 2010) were designed to account for bilinguals' language non-selective word recognition. Critically, these models share the assumptions that lexical access (a) is language non-selective and (b) takes places within an integrated lexicon. Furthermore, BIA assumes that lexical access can be modulated in a top-down fashion by inhibitory connections from language nodes (see **Fig. 7**), while BIA+ contends that language node activation is bottom-up and does not feedback to the lexical level (see **Fig. 8**).

The BIA model is an extension of the monolingual Interactive Activation model originally proposed by McClelland & Rumelhart (1981). In BIA, there are four hierarchical levels: Feature, Letter, Word, and Language (see **Fig. 7**). When a word is visually presented, particular features at each letter position activate letters containing these features, while

constraining the access to letters without these features at the same time (thus increasing contrast between target and nontarget letters). The letters then, via feed forward connections, activate word forms which contain these letters at the required positions, while inhibiting other word forms, irrespective of language. In addition, the activated word forms are associated with a corresponding language node, which then inhibits activation of the other language.



Figure 7. The BIA model adapted from van Heuven, Dijkstra, and Grainger (1998), pp. 475

Some empirical findings have suggested that BIA fails to explain consecutive item effects and observations relating to language switching (e.g., Thomas & Allport, 2000). From a BIA perspective, when a Chinese-English bilingual reads a word in English, this word would activate English language nodes and, on the next trial, the activated English node would inhibit Chinese lexical representations. In addition, BIA does not make predictions regarding phonological and semantic representations (for a summary, see van Heuven & Dijkstra, 2010), which has led the authors to the proposal of BIA+ (see **Fig. 8**).

In BIA+, visual input first activates sublexical orthographic representations, which, in turn, activates whole-word orthographic representations and sublexical phonological representations, before activation spreads to semantic representations and language nodes. However, the language nodes do not act on the activation level of representations held in the word identification system.

#### Task/Decision System

- Specific processing steps for task at hand
- Receives continuous input from the word identification system
- Decision criteria determine when a response is made based on relevant codes



Figure 8. The BIA+ Model adapted from van Heuven and Dijkstra (2010), pp. 114

Both BIA (van Heuven et al., 1998) and BIA+ (Dijkstra & van Heuven, 2002; van Heuven & Dijkstra, 2010) assume that differences in word activation depend on resting activation levels of representations which are highly correlated with word frequency. High frequency words are associated with higher resting levels of activation and shorter reaction times, whereas low frequency words are associated with lower resting levels of activation, and longer reaction times. Furthermore, difference in proficiency between L1 and L2 would determine the weight of resting activation levels for L1 and L2 word representations.

Although some assumptions of the BIA+ model have been challenged by results from priming studies in bilinguals (e.g., Duñabeitia, Kinoshita, Carreiras, & Norris, 2011; Grainger & Holcomb, 2009; Duyck and Warlop, 2009), BIA+ best accounts for the effects of cross-language activation reported in the literature, and is supported by a large number of behavioural and neuroimaging studies (for a summary, please see van Heuven & Dijkstra, 2010).

#### 1.2.5 Conclusion of section

Overall the different models of bilingual lexical processing reviewed here provide a framework to conceptualise the stages of information processing involved when bilinguals encounter words in the L1 or the L2. The structural properties of these models constrain how overlap between L1 and L2 word forms –or lack thereof– can produce selective interference affecting various levels of language representation all the way to semantic-conceptual levels. For instance, according the BIA model, a word in L2 should lead to inhibition of L1 word representations via activation of the L2 language node but this assumption is not made by BIA+, since language nodes do not inhibit the word representation level in return. As to semantic representations, all previously reviewed models suggested that bilinguals have a shared semantic representation level that is accessible in their L1 and their L2. What is unclear is how these models can accommodate linguistic relativity effects, that is, how language forms idiosyncratic to a given language may restructure perceptual and conceptual representations, against the predictions of the current models.

### 1.3 EEG and Event-related potentials

The great majority of the processes described above, whether in relation to linguistic relativity or models of language processing in bilinguals have been described mostly on the basis of behavioural data analysis. However, it is widely acknowledged that the basis of these mechanisms is mostly unconscious and implicit in nature. This bring up the question of validity, insofar as the source data may only apply to situations where participants are fully aware of the processes at work and metacognitively engaged in language tasks. In order to gain new insights into these processes in conditions where participants' strategies are unlikely to distort the findings, we need to use methods that measure responses from the brain more directly, that is measures that are less susceptible to overt task manipulations and participants explicit knowledge.
# 1.3.1 Principles of EEG

The electrical activity generated by the human brain was first measured by German psychiatrist Hans Berger (1873-1941) in 1924. The electroencephalogram or EEG is measurable when neurotransmission occurs in hundreds of thousands of (mostly) pyramidal cells organised perpendicularly to the cortical surface (Luck, 2014). The signal measured at the scalp is the difference in electrical potential between any given electrode set on the scalp and a reference electrode. This signal is amplified in the connection box for each electrode and digitalized before being stored on a computer disk. Experiments presented in this thesis all used 64 electrodes sites located according to the 10-20 convention (see Fig. 9).



**Figure 9.** Electrode array used in experiments presented in this thesis. FPz = ground; Cz = Online reference; VEOG = Vertical electroocculogram (+ up, - down); HEOG = Horizontal electroocculogram.

Four electrodes are used to measure vertical and horizontal eye movements (two for each) as the muscles involved in eye movements (e.g., eyelid movements and ocular rotation) and the equivalent dipole corresponding to the retina generate ocular artefacts distributed over the scalp. Therefore, eye movements need to be recorded in order to compute a mathematical model of eye blink artefacts for correction based on the regression method proposed by Gratton, Coles and Donchin (1983).

After local degreasing of the scalp using 70 degree rubbing alcohol, the electrodes are bridged with the scalp using a conductive gel (mix of water, electrolytes, sand, and jellifying agent) and impedances are progressively brought down to below 5 k $\Omega$  by gentle abrasion of the scalp using a cotton bud. Once impedances have been stabilised evenly below 5 k $\Omega$  across the scalp, the voltage can be amplified, filtered online in AC mode between 0.05–200 Hz and sampled at a rate of 1 kHz (one measure every millisecond).

Once EEG data have been recorded, they are filtered using a zero-phrase shift bandpass digital filter with a high cut off frequency of 20 Hz (slope: 48 dB/oct) to reduce contamination from exogenous electrical noise (e.g., line noise and EMG noise) and a low cut off frequency of 0.1 Hz (slope: 24 dB/oct) to reduce skin potentials and other slow voltage changes. Extreme modulations (exceeding +/- 75  $\mu$ V) are then marked for rejection before ocular artefact correction is performed.

EEG data are then re-referenced to the global average reference or the bi-mastoidic reference, TP7 and TP8, in case comparison with previous publication requires it (see **Fig. 9**) to recover signal at key strategic electrode Cz and topographical information. The re-referencing method adopted in this thesis was the global average reference.

# 1.3.2 From EEG to ERPs

Continuous EEG recordings were then cut into epochs ranging from 200 ms before to 1000 ms after the onset of the critical stimulus. *Event-related potentials* (ERPs) are the average time-series computed from a collection of EEG epochs time-locked to specific events, such as sensory, cognitive, or motor events (Luck, 2014). The epochs are averaged across trials belonging to a given experimental condition and then averaged across participants to obtained a grand-average ERP, to further increase average-driven smoothing and normalisation of data. Mean peak amplitudes are then measured around the peak latency of components of interest (e.g., the N400, the P300, or the P600), preferably in a predictive fashion, that is, based on the time windows and electrode collections determined in previous research (Picton et al., 2000).

The ERP waveforms comprises a series of positive-going and negative-going voltage deflections called ERP *components* (**Fig. 10**). P is employed to label positive-going peaks and N is employed to label negative-going peaks. Numbers either indicate a peak's ordinal

position in the series (e.g., N2 is the second negative peak) or the canonical latency of the peak in milliseconds (e.g., N200 refers to a negative component with a usual peaking latency of about 200 ms).



Figure 10. Illustration of some ERP component.

## 1.3.3 ERPs and linguistic relativity

Here I present a highly selective description of the main ERP component that have been used as indices of choices in studies of linguistic relativity and language processing in bilinguals. I put particular emphasis on the N400, since it is the main component used in the studies present in this thesis.

#### 1.3.3.1 The vMMN and the P300

The Mismatch Negativity (MMN) is a modulation of the N1 peak elicited by auditory stimuli indexing sudden changes in a repetitive auditory stream of stimuli. The MMN peaks at around 100-250 ms after the onset of a deviant stimulus (low probability) in a stream of standard stimuli (high probability; Sams, Paavilainen, Alho, & Näätänen, 1985). The latency of the MMN is influenced by sound frequency, duration, intensity, and inter-stimulus interval amongst other parameters (for a summary, see Näätänen et al., 1997). The equivalent of the MMN in the visual domain is the visual mismatch negativity (vMMN; e.g., Astikainen, Ruusuvirta, Wikgren, & Korhonen, 2004; Czigler, Balázs, & Pató, 2004; Pazo-Alvarez, Cadaveira, & Amenedo, 2003), widely regarded as an index of automatic and pre-attentive processing of a deviant visual stimulus presented within a stream of highly repeated standard stimuli (Kimura, Schröger, Czigler, & Ohira, 2010). The vMMN peaks in the same time range as the MMN but over posterior electrode sites. In the study of linguistic relativity, it has been used to test the influence of language terminology on colour perception (e.g. Thierry et al., 2009; Athanasopoulos et al., 2010) and object categorisation (e.g., Boutonnet et al., 2012; Boutonnet et al., 2013).

The P300, sometimes called the late positive component (LPC) when its peaking time is delayed beyond 450 ms, is arguably the first ERP component reported in scientific history by Sutton, Braren, Zubin and John (1965). Over the decade following its discovery, two alternative forms of P300 have been distinguished: The P3a, which tends to peak early and has a frontal to frontocentral topography, and the P3b, which tends to peak later and has a centroparietal to parietal distribution (Squires, Squires, & Hillyard, 1975). Both P3a and P3b components can be elicited by infrequent, unpredictable changes in the standard stimulus string but the P3b tends to be generated when stimulus deviancy coincides with task targets (Luck, 2014). The P300 usually peaks over midline electrodes (Fz, Cz, Pz) and increases at frontal and parietal recording sites for target stimuli (Johnson, 1993). The neural source for P3a has been identified in the frontal lobe (Knight, 1984, 1990, 1996; Knight, Grabowecky, & Scabini, 1995). The sources of the P3b, on the other hand, have been localised in the temporal lobe (Halgren et al., 1980; McCarthy, Wood, Williamson, & Spencer, 1989).

Although a large number of experiments have investigated the neural process indexed by the P300, there is still no clear consensus regarding its functional interpretation. The most frequently cited process in relation to the P300 is "context updating", i.e., the revision of a current mental representation (updating) induced by incoming stimuli (Donchin, 1981), with the P3a tending to index such updating in response to a novel, surprising stimuli, and the P3b tending to index target detection in conditions where target characteristics have been explicitly defined. Based on the context-updating theory, when a new stimulus is detected, updating of representations elicits a P300 but if no new stimulus is presented (i.e., in the case of standards), the context is maintained and only sensory potentials are evoked. Various factors can modulate P300 amplitudes. Target and deviant stimulus within-experiment probabilities have a strong and highly reliable impact on P300 amplitudes. When the target probability gets smaller, P300 amplitude increases (Duncan-Johnson & Donchin, 1977). The P300 is also sensitive to the amount of attentional resources available: Increasing memory load reduces P300 amplitude since less attentional resources are available to support updating mechanisms (e.g., Kok, 2001; Wijers, Mulder, Okita, Mulder, & Scheffers, 1989), as does dual-tasking (Polich, 2012). P300 amplitudes elicited by stimuli relevant for a secondary task

are large when the primary task is easy, while P300 amplitudes are smaller if the primary task is difficult (Kramer, Wickens, & Donchin, 1985).

Together, (v)MMN and P300 index two facets of change detection by the human brain: The former indexes pre-attentive, unconscious and spontaneous change detection without any requirement for over processing and the latter indexes conscious, contextupdating mechanisms, maximal when a response is required, but in any case, requiring explicit acknowledgement by the participant of an infrequent (oddball) event. The vMMN has been used as the index of choice to study colour and object categorisation effects in relation to terminology (e.g., Thierry et al. 2009; Boutonnet et al., 2013) and the P300 has been used to investigate preference toward endpoints over trajectory in German and English native speakers viewing motion event animations (e.g., Flecken et al, 2015).

#### 1.3.3.2 LAN and P600

The left anterior negativity (LAN) is a negative going wave reflecting first pass parsing processes during sentence comprehension (Friederici, 1995). The LAN is most obvious over left anterior recording sites (but see Hagoort, Wassenaar, & Brown (2003) who showed that LAN scalp distribution is not consistent across studies). Two types of LAN have been described in the literature: an early LAN (ELAN) peaking between 100–200 ms after critical word onset and the standard LAN, peaking between 300–500 ms after critical word onset.

The ELAN is supposed to reflect a highly automatic process of initial structure building (Friederici, 1995). Word category or phrase structure violations due to word omissions, word category violations, and word order violations have been reported to elicit ELANs. However, more recent studies by Steinhauer and colleagues have convincingly shown that the ELAN is likely to be an artefact generated by anticipatory mechanisms when participants can predict the occurrence of a violation in non-rotated designs (Steinhauer & Drury, 2012).

The LAN, the existence of which has not been contested, typically peaks between 300–500 ms over left frontocentral and frontal electrodes. The LAN is classically elicited by two types of grammatical violations: Agreement violation (e.g., Gunter, Stowe, & Mulder, 1997; Hagoort, Brown, & Groothusen, 1993) and verb-argument structure violation (e.g., Coulson, King, & Kutas, 1998; Rösler, Pütz, Friederici, & Hahne, 1993). The neural generators of the LAN are thought to be the left anterior frontal cortex and the left anterior

temporal lobe (Friederici et al., 1999). In bilinguals, the duration of L2 immersion appears to be a dominant factor modulating LAN effects (Caffarra et al., 2015). A LAN modulation has been reported by Boutonnet et al. (2012) in an experiment involving picture triads (see section 1.1.3.2 above) showing that gender information is automatically access by individuals asked to make semantic decisions on pictures, even when they have no requirement to access lexical information.

The P600 (a.k.a. P6 or Syntactic Positive Shift, SPS) is a positive-going deflection elicited by morphosyntactic or word order violations in sentence processing. The P600 onsets at around 500 ms after the presentation of the critical word and peaks at around 600 ms. The P600 and the P300, especially the P3b, have a lot in common: Their distribution is centroparietal, their peak latency is late and highly correlated with reaction time (Sassenhagen & Fiebach, 2018). In terms of neural origins, both the P600 and the P300 have been linked to generators in the basal ganglia based on the study of neuropsychological patients with focal brain lesions (e.g., Friederici, Von Cramon, & Kotz, 1999; Kotz, Frisch, Von Cramon, & Friederici, 2003).

The P600 was first reported by Osterhout and Holcomb (1992), who recorded ERPs in participants reading grammatically correct sentences and sentences featuring grammatically ambiguous words (i.e., so-called "garden-path" sentences). An example of a pair of such sentences is: "The broker hoped to sell the stock". / \*"The broker persuaded to sell the stock". The latter sentence is ambiguous since it could go on (as in: "The broker persuaded to sell the stock ended up in jail"). Since its discovery, the P600 has been repeatedly observed in response to a range of grammatical errors, such as tense agreement violation (e.g., Allen, Badecker, & Osterhout, 2003; Steinhauer & Ullman, 2002), number and gender agreement violation (e.g., Barber & Carreiras, 2005; Hagoort, 2003; Molinaro, Barber, & Carreiras, 2011), subcategorization violation (e.g., Ainsworth-Darnell, Shulman, & Boland, 1998; Osterhout, Holcomb, & Swinney, 1994), long-distance dependencies (e.g., Kluender & Kutas, 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991), etc. However, the scalp distribution of the P600 elicited by grammatical violations is more posterior than that elicited by garden path sentences of the kind used in the original 1992 study by Osterhout and Holcomb. And within the realm of grammatical violations, P600 scalp distribution can vary substantially (e.g., Neville et al., 1991; Osterhout & Holcomb, 1992). Sassenhagen and Fiebach (2018) recently argued that the P600 is a language-bound process from the P300

family. However, Gouvea, Phillips, Kazanina and Poeppel (2010) have previously argued that the P600 is not language-specific and functionally distinct from the P300.

Just like the N400 (see below), the P600 is not exclusively elicited by linguistic stimuli. For example, Martín-Loeches, Casado, Gonzalo, de Heras and Fernández-Frías (2006) reported a P600 elicited by mathematical symbol violations in participants tasked with solving mathematical problems. Cohn, Paczynski, Jackendoff, Holcomb and Kuperberg (2012) also showed that cartoon clips presented within an incorrect time sequence also elicit a P600. And Patel (2003) found out-of-key chords within a music phrase can also elicit a P600, to give only a few examples.

A mainstream interpretation of P600 functional significance relates to structural reanalysis, which can be triggered by rule violations or interruptions in either the linguistic or the non-linguistic domain. Contrary to the N400, however, the P600 peak latency varies based on the difficulty of syntactic structure recovery (Friederici & Mecklinger, 1996). In bilinguals, L2 proficiency has been proposed as a key modulating factor of P600 amplitude. Bilinguals who achieved higher than 75% accuracy rate in terms of explicitly judgements elicit P600 of greater amplitudes and more often (Caffarra et al., 2015). Age of acquisition is another key factor modulating P600 amplitude: Bilingual participants who have learnt their L2 before the age of 11 years can generate native-like P600, whereas participants who began learning their L2 after the age of 11 exhibit a more anterior negativity that is bilaterally distributed (Weber-Fox & Neville, 1996) but see Steinhauer, White and Drury (2009) for arguments against the critical period hypothesis.

#### 1.3.3.3 The N400

In 1980, Martha Kutas and Steve Hillyard famously reported a negative going component peaking at around 400 ms after the onset of a word semantically difficult to integrate within the preceding context (e.g., \*"He spread the warm bread with socks"). This N400 component usually has a central or centroparietal, and often slightly right-lateralised, topography, which can vary depending on the presentation modality (e.g., visual or auditory) and the nature of the stimulus (e.g., word, picture, music, environmental sound). Several cortical territories have been proposed to be at the origin of the N400 wave (for a review, see Van Petten & Luka, 2006) with the most likely neural sources located in the anterior temporal lobes (Nobre & McCarthy, 1995).

Previous research has established that the N400 is highly correlated with a word's expectancy within the context (for a review, see Kutas & Federmeier, 2011), word imaginability (Swaab, Baynes, & Knight, 2002), and word class (Brown, Hagoort, & Keurs, 1999), to name only a few. Contextual modulation of the N400 can be observed at various linguistic levels, ranging from single (preceding) word to sentence or even discourse level (for a review, please see van Berkum, Hagoort, & Brown, 1999). Furthermore, the N400 elicited by a single word-elicited tends to be smaller in amplitude than that elicited by a sentence, but their distribution and latency are similar if stimuli are presented in the same modality (van Petten, 1993). N400 modulation is not limited to linguistic constructions since it can be elicited by non-verbal stimuli, such as pictures and music (for a review, see Kutas & van Petten, 1988; Kutas & van Petten, 1994).

From a developmental perspective, pre-verbal children as young as 9 months old can generate N400s to unexpected events in action sequences (e.g., Reid et al., 2009, as cited in Kutas and Federmeier, 2011). In the language domain, children aged between 12-14 months can display picture/word congruency effect modulating N400 amplitude (Friedrich & Friederici, 2010). As for the case of bilinguals, Age of acquisition (AoA) has been suggested as a major factor of N400 modulation compared with L1-L2 similarity, L2 proficiency, and length of L2 immersion: L2 speakers who learnt a second language before the age of 10 years display greater N400 modulations (Caffarra et al., 2015).

All-in-all, the N400 component of ERPs is highly reliable index of online semantic and conceptual processing. Modulations of this component can be systematically engineered to serve as effect of reference and simultaneously tested in conditions in which modulations are yet undetermined (e.g., form repetition, emotional valence sensitivity, cross-language interference, etc.; Thierry & Wu, 2007; Wu & Thierry, 2010, 2012).

Together, vMMN, P300, LAN, P600, and N400 offer a diverse and functionally informative set of tools to study information processing in bilinguals whose languages are fundamentally different in nature. In the studies presented here, we have chosen to focus particularly on the N400 to index semantic processing (and the P600 to index re-evaluation / repair) in relation to language-perception integration mechanisms in Chinese-English bilinguals.

# 1.3.4 Advantages and Disadvantages of ERPs

The advantage of ERPs over behavioural methods is that they can provide an on-line measure of processing while behavioural measures only offer insight into the end point of the chain of events having led to a response (Luck, 2014). For example, Kousta, Vinson and Vigliocco (2008), in line with arguments put forward by scholars like Steven Pinker, have argued that it is unreliable to determine the strength of Whorfian effects based solely on behavioural measures, because participants in these experiments could explicitly resort to access language knowledge in order to solve the tasks. ERPs, on the other hand, have a very high temporal resolution, which allows differences between experimental conditions to be tested continuously from stimulus onset to behavioural response and beyond. Apart from that, participants do not need to behaviourally respond in an ERP experiment, meaning that ERP measures can provide information about cognitive processing even when no response computation is at play. This aspect of ERPs is also extremely helpful when investigating populations that cannot physically respond such as infants or neurologically impaired individuals (for a review, see Coch & Gullick, 2012). ERPs are non-invasive and are inexpensive compared with all mainstream neuroimaging techniques such as functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET) or Magnetoencephalography (MEG). In addition, ERP components in specific conditions allow determining the kind of mental process involved and indeed can differentiate between perception, categorisation, working memory updating, semantic processing, implicit and explicit target detection, etc, whereas fMRI, for instance, does not allow isolating a particular mental process given that the same neurons may underpin different processes (Kappenman & Luck, 2012; Luck, 2014).

Nevertheless, ERPs also have substantial limitations. Beyond the issue of the very poor spatial resolution, the most obvious challenge of ERP research is the *superposition* problem (Luck, 2014). The voltage recorded on the scalp represents the sum of underlying components and it is difficult to extract single meaningful components from the combination observed at the surface. The poor spatial resolution of ERPs can be somewhat addressed by combining EEG data collection with e.g., Transcranial Magnetic Stimulation (TMS) or transcranial Direct-Current Stimulation (tDCS; Ritter & Villringer, 2006). Combination of ERPs with fMRI, on the other hand, has proven difficult, not necessarily because of the difficulty of measuring weak currents using conductive leads in a high-powered electromagnetic environment, but rather because formulating strong hypotheses

simultaneously for the two methods is challenging. Finally, a large (potentially infinite) number of mental processes do not have characteristic ERP signatures, although many different ERP components have characterised (Luck, 2014). This means that ERPs are limited in that they can only be used to investigate those neural processed indexed by a distinct, reliable, and predictable ERP component.

# 1.4. Thesis Aims

In this Chapter, I reviewed the premises of *Linguistic Relativity (LR)* and a string of experimental results that have provided some evidence in favour of a systematic effect of language on perception and cognition, in monolingual speakers as well as bilingual individuals. In the present thesis, I am concerned with time conceptualisation in Chinese-English bilinguals. The current state of the literature is highly inconsistent and there is little objective, neuroscientific evidence for effects of language on time conceptualisation. Chinese differs from English regarding how it verbally conveys temporal information. These differences are manifest at a grammatical level (English has tense, Chinese does not) and at a lexical level (some spatiotemporal metaphors of Chinese have no equivalent in English).

The bilingual mind accommodates two language systems and whilst *models of word comprehension (and production) in bilinguals* assume different possible configurations of lexical and sublexical processes, most of these models make little or no assumptions about possible subdivisions within the bilingual semantic system. The premises of LR are thus far from being readily transposable to psycholinguistic models of word recognition.

To understand how language diversity can affect conceptual representations of time in Chinese-English bilinguals, I have thus chosen to use *ERPs as a predictive tool* to investigate the consequences of grammatical and lexical variations between Mandarin Chinese and British English, and investigate how the bilingual mind copes with such diversity when the second language (here English) is mastered to a high level of proficiency. The tool of choice is *the N400*, since it is the most reliable index of semantic-conceptual processing available in neurolinguistics to date.

First, in two studies, we investigated the consequences at a semantic-conceptual level of having no tense in the native language when English is the second language. These studies attempt to explain the difficulty experienced by Chinese native speakers when they read English and possible confusion arising regarding the recent past in particular, since the distinction between past and recent past relies heavily on tense marking in English, whereas it is conveyed by time adverbials, contextual information, and aspect markers in Chinese. In *Study 1*, I have tested the ability of highly fluent Chinese-English to detect reference time misalignments between clauses making up a sentence, as participants were asked to make acceptability judgments regarding sentence meaning.

In *Study 2*, to test that results from Study 1 were not driven by lack of proficiency in our bilingual participants, or by the focus on semantic violations detracting participants' attention away from temporal information, we tested another group of highly proficient bilinguals using a normalised grammatical proficiency test and direct their attention directly onto temporal information rather than coarse semantic violations. Together Study 1 and 2 offer radically new insights into the sparsely investigated phenomenon of tense processing in native speakers of Chinese who have otherwise achieved very high levels of proficiency in English.

Second, in two further studies, we examined how spatiotemporal metaphors of Chinese are accessed depending on language of operation and context of testing and how such language expressions idiosyncratic to Mandarin Chinese construe the way in which native speakers of the language conceptualise the passing of time. In *Study 3*, I focussed on the sagittal (front-back) axis for spatial representations of time and tested the potential impact of having spatiotemporal metaphors (e.g., *hou tian*– 'the day after tomorrow', literally translating as 'back-day') inconsistent with the future-in-front convention shared by both Chinese and English.

In *Study 4*, I turned my attention to the vertical (up-down) axis, since several prominent publications have previously reported that Chinese individuals "think of time vertically". In this last study, I capitalised on the fact that Mandarin Chinese features vertical spatiotemporal metaphors for months and weeks that directly refer to orientation in space (e.g., *shang ge yue*–'last month', literally translating as 'up-month') and others that do not use spatial references (e.g., *qu nian*–'last year', literally translating as 'gone year').

Together, these studies offer important new insights into time conceptualisation (a) derived from tense in English (and a mechanism explaining difficulties Chinese-English bilinguals experience with the concept of tense) and (b) along three different axes of time representations: (i) the sagittal axis (front-back) in relation to the future in front convention of time mapping, (ii) the transverse axis (left-right), which is a representational axis for the passing of time, and (iii) the vertical axis (up-down) which serves as a core spatial reference for time in Chinese but not in English.

Finally, in a fourth Chapter, I aim to discuss the results obtained along this PhD journey, by considering the validity of my findings in light of the broader literature, and discussing limitations as well as opportunities for future research arising from my studies.

 $\infty$ 

# Chapter 2

Tense and linguistic Relativity

# 2.1 Study 1

# **Timeline blurring in fluent Chinese-English bilinguals**

# Abstract<sup>4</sup>

Linguistic relativity effects arising from differences in terminology and syntax between languages have now been established in various domains of human cognition. Although metaphors have been shown to affect time conceptualisation, there is little evidence to date that the presence or absence of tense within a given language can affect how one processes temporal sequences of events. Here, we set out to characterise how native speakers of Mandarin Chinese-a tenseless language- deal with reference time misalignment using event-related brain potentials. Fluent Chinese-English participants and native speakers of English made acceptability judgements on sentences in which the adjunct clause started with the connective 'after' and was either temporally aligned or not with the main clause in terms of reference time conveyed by the verb. Native speakers of English failed to overtly report such reference time misalignments between clauses, but significant N400 modulations showed that they nevertheless required additional semantic processing effort. Chinese speakers, however, showed no such N400 modulation suggesting that they did not covertly detect reference time misalignments between clauses in real time. Critically, all participants manifested normal sentence comprehension as shown by a standard N400 semantic violation elicited by incongruent endings. We conclude that Chinese speakers of English experience difficulties locating events on a timeline in relation to one another when temporal information is conveyed by tense.

*Keywords:* Reference time; tenselessness; event-related potentials; syntax-semantics interference; linguistic relativity; sentence processing.

<sup>&</sup>lt;sup>4</sup> Section 2.1 appears in print as Li, Y., Jones, M., Thierry, G., (2018). Timeline blurring in fluent Chinese-English bilinguals. *Brain Research*, 1701, 93-102.

## 2.1.1 Introduction

Recent research has provided evidence that language influences cognitive functioning (Athanasopoulos, 2009; Boroditsky, 2001; Boutonnet et al., 2012; Choi & Bowerman, 1991; Gentner & Goldin-Meadow, 2003; Lantz & Stefflre, 1964; Lucy, 1992; Lupyan & Ward, 2013; 1996; Whorf, 1956). Such effects have been demonstrated at the level of elementary visual perception (Thierry et al., 2009) and object categorisation (Boutonnet et al., 2012; Cubelli et al., 2011; Phillips & Boroditsky, 2003), through to high-level, abstract meaning processing such as event conceptualization (Flecken et al., 2015) and cultural semantics (Ellis et al., 2015).

Grammatical variations between languages can influence event conceptualization as shown by studies of motion event categorization both in language tasks (Flecken, 2011; von Stutterheim & Carroll, 2006) and non-verbal tasks (Athanasopoulos, 2009; Athanasopoulos & Bylund, 2013; Flecken et al., 2015; Flecken, von Stutterheim, & Carroll, 2014; Papafragou, Hulbert, & Trueswell, 2008). For example, both English and Arabic speakers – whose native languages have aspect markers– spontaneously mention the temporal properties of motion events (e.g., "Two women are walking down a path"). In contrast, native speakers of German –whose mother tongue lacks aspect categorization altogether– describe the same events in more holistic terms, including the mention of a possible endpoint (von Stutterheim & Carroll, 2006).

Here, we set out to examine whether linguistic differences in *tense* marking can affect the representation of temporal relationships between events. Tense is a linguistic device that locates a given situation in time (Declerck, Reed, & Cappell, 2006). It is accepted that English is a tensed language although there is a controversy over the existence of the future tense (Comrie, 1985; Declerck et al., 2006; Quirk, Greenbaum, Leech, & Svartvik, 1985). Other languages lack absolute tense altogether. That is, they do not mark either present-future or past-non-past distinctions in their grammar. In Mandarin Chinese, for example, specifying the temporal location of an event is not compulsory (Comrie, 1985; Li & Thompson, 1989; 1991). Instead, temporal information is optionally expressed through time adverbials (e.g. *zuó tiān* – 'yesterday'; *ming tiān* – 'tomorrow'), modal auxiliaries (e.g. yao – 'will'; *jiāng* – 'will'), or through context (Duff & Li, 2002; Smith, 1991), the default position being that the event unfolds in the present (Smith, 2008). Qiu and Zhou (2012), for instance, found that native speakers of Mandarin Chinese are sensitive to the disagreement between a modal

auxiliary (e.g., *jiāng yào* – 'will') or an aspectual particle (e.g., the marker of perfect aspect –  $gu\partial$ ) and a temporal noun phrase (e.g., *shàng gè yuè* – 'last month'), as in the following sentence:

\* 上个月 联合国 将要 派出 特别调查组。<sup>5</sup> Last month UN jiangyao Dispatch Investigation unit Last month, the United Nations will dispatch a special investigation unit.

(From Qiu and Zhou, 2012, pp. 94)

To investigate whether cross-linguistic differences in tense marking can influence readers' perception of time, we tested fluent Mandarin-English bilinguals reading English sentences. According to the approach proposed by Reichenbach (1947), the timeline corresponding to a situation described by an utterance involves three time points: (a) *Speech time* (the time at which the utterance is produced), (b) *reference time* (the perspective from which a situation is perceived), and (c) *event time* (the time at which the event happens). In order to understand the temporal order of events in a given sentence, and therefore its overall meaning, it is necessary to encode on the one hand the relationship between Speech Time and Reference Time (theoretically encoded by tense), and on the other hand the relationship between Grae of a tensed language, the three time points and their relationships are coded directly by inflection (Smith, 2008). However, in the case of tenseless Mandarin Chinese, the relationship between Speech Time and Reference Time and Reference Time can remain unspecified because it is not encoded by an inflectional morpheme within the verb (Smith, 2008) and specifying temporal information is not compulsory (Smith, 1991).

We thus created complex English sentences featuring a reference time misalignment (RTM) between their adjunct and main clauses. In all cases, adjunct clauses began with the connective '*after*' and systematically described a first event with perfect aspect – a grammatical category that exists in both English and Chinese. In the RTM conditions (see **Fig. 11B** and **11C**) the adjunct clause was in the present or the future tense, whereas the main clause was in the absolute past tense (simple past). Note that such RTM is different from tense violation, since the latter entails grammatically incorrect tense forms within a given clause, as in "Yesterday, I sail Diane's Boat to Boston" (from Steinhauer & Ullman, 2002, pp

<sup>&</sup>lt;sup>5</sup> Pin yin version of sentence: Shàng-gè yuè lián hé guó jiāng yào pài chū tèbié diào chá zǔ.

63). We also created a semantic violation condition in which the statement was made meaningless by the presence of an incongruent word ending designed to serve as a semantic control, to ensure that bilingual participants could understand the meaning of the materials presented (see **Table 2**).

Conditions	Stimuli	Chinese translation (pin-yin)
Control Correct	After the director of the school <b>had</b> <b>resigned</b> from the university, he worked for a multinational.	院长从大学辞职后, 他去了一家跨国公司工作。
Present-Past Misalignment	*After the director of the school <b>has</b> <b>resigned</b> from the university, he <b>worked</b> for a multinational.	yuàn zhăng cóng dà xué cí zhí hòu, tā qù-le yī jiā kuà guó gōng sī gong zuò
Future-Past Misalignment	*After the director of the school <b>will have</b> <b>resigned</b> from the university, he <b>worked</b> for a multinational.	院长从大学辞职后, 他去了一家米工作。
Semantic violation	*After the director of the school had resigned from the university, he worked for <b>a meter</b> .	yuàn zhăng cóng dà xué cí zhí hòu, tā qù-le yī jiā mǐ gōng zuò

Table 2.	Examples	of sentences	in each of	f the four	experimental	conditions	and Fillers
----------	----------	--------------	------------	------------	--------------	------------	-------------

In control (correct) sentences, the adjunct clause was in the past perfect tense (see **Fig. 11A**). It shared its reference time and speech time with the main clause in the simple past tense, and thus was correct according to the rule of temporal connectives (Hornstein, 1990). The RTM conditions were of two kinds: (1) a Present-Past Misalignment (PPM; **Fig. 11B**) and a Future-Past Misalignment (FPM; **Fig. 11C**), in which the tense of adjunct clauses does not share speech time and reference time with the main clause.



**Figure 11.** Timeline representation of the critical stimuli. E: Event time; R: Reference Time; S: Speech Time. Index number identifies the clause (1 for the adjunct clause and 2 for the main clause).

We tested monolingual speakers of English as controls and Mandarin-English speakers, who were fluent in both Mandarin and English, that is, able to hold a normal, fluid conversation in either of their languages (Grosjean, 1989; Macnamara, 1967; Titone, 1990). We did not expect marked differences between fluent Mandarin-English bilinguals and native speakers of English in a sentence acceptability task. However, we expected event-related brain potentials (ERPs) recorded simultaneously to index differences in online processing of temporal information. For native English speakers, we predicted that RTM conditions would elicit greater negativity in the N400 range compared with correct sentences (see Liang et al., 2016; Nieuwland, 2015), owing to semantic difficulties in reconciling the incongruous timelines presented in the adjunct and main clauses (Kutas & Federmeier, 2011; Kutas & Hillyard, 1984a; Kutas, Lindamood, & Hillyard, 1984; Van Petten & Kutas, 1990). However, because tense encodes the relationship between reference time and speech time, and since Chinese does not encode tense directly, we predicted that native speakers of Mandarin Chinese would be less sensitive to RTMs than their native English peers. This should translate into a relatively weaker N400 modulation in bilinguals and we thus predicted a group x condition interaction in the N400 range.

## 2.1.2 Materials and Methods

#### 2.1.2.1 Participants

Thirty Mandarin-English bilinguals and 25 native speakers of English took part in this study. Data from three bilingual participants and 6 native speakers of English were discarded due to poor ERP data quality, such as heavy blinking and excessive alpha elicitation. Of the remaining 27 bilingual participants, 10 were males and 17 females, with a mean age of 22.3 (SD=2.7) and were all right-handed. In the English native group, 19 participants' dataset were kept (8 males and 11 females; *Mean* age= 22.4, SD=9.3; one left-hander and 18 right-handers). All participants were students at Bangor University, UK, and received either payment or course credits for their participation.

The average age at which Mandarin-English bilinguals started to learn English was 6.9 years (*SD*=3.2), and all were living in the UK at the time of testing. The average IELTS score for the bilingual group was 6.9 (*SD*=0.5, from 6.5 to 8). Five further bilinguals did not

provide IELTS scores, since they received English medium instruction since high school. **Table 3** summarizes Mandarin-English participants' language background.

Measure	Mean	(SD)
Age of L2 acquisition	6.9	(3.2)
Length of L2 learning (years)	15.2	(3.5)
Length of staying in an English Speaking country (Months)	44.7	(61.4)
Daily Chinese usage (%)	50.7	(20.4)
Daily English usage (%)	49.3	(20.4)

Table 3. Chinese-English bilingual participants' language background

Bilingual participants also self-reported their proficiency in both English and Mandarin Chinese (see **Fig. 12**). All participants had normal or corrected to normal vision. The study was approved by the School of psychology, Bangor University ethics committee.



**Figure 12.** Chinese-English bilingual participants' self-estimation of their Chinese and English level (10 point- scale). Error bar represents stand error.

#### 2.1.2.2 Stimuli

The materials consisted of 70 sentence sets, each containing 8 sentences. Four were experimental sentences featuring either a (i) correctly tensed verb, (ii) PPM, (iii) FPM, or (iv) semantic violation (see **Table 2 and Appendix B**) and 4 sentences served as fillers. The locus of the reference time misalignment coincided always with the second word of the main clause. For the main analyses, we compared the control condition (i) with the two RTM conditions (ii) and (iii). An additional analysis comprised (i) and (iv), in order to ascertain that the Mandarin-English bilinguals comprehended the overall meaning of the sentences.

In order to dilute the critical experimental manipulations, filler sentences were included, in which the matrix sentences used the simple future tense. There were two presentation lists, which alternated so as to present experimental items and fillers in a fully counterbalanced fashion. Each presentation list featured 4 blocks and a given sentence from a given condition was only presented once per block. Stimuli from the same set were never presented together in the same block. In addition, verb regularity was systematically manipulated such that half were regular and the other half irregular. There was no significant difference in lexical frequency between regular and irregular lists even though there was a trend for irregular verbs to be more frequent (U = 451.5; p = 0.06).

#### 2.1.2.3 Procedure

Bilingual participants first filled out a language background questionnaire (see **Appendix A**). All participants were seated 100 cm away from a 19-inch computer monitor and responded by pressing button on a reaction time box. The first clause of each sentence was presented at once and participants were instructed to press any button when they had finished reading. The rest of the sentence then comprised individually presented words, in the centre of the screen, for a duration of 300 ms (ISI 400 ms), in order to minimise eye movements. Once the whole sentence had been read, participants were required to judge whether or not it made sense (see **Fig. 13**).



Figure 13. Structure of an experimental trial.

#### 2.1.2.4 Design and behavioural data analysis

In this Study, we compared two groups (English native speakers, Mandarin-English bilinguals) and, within-subject, three reference time alignment conditions (correct, PPM, FPM). In addition, participants understanding of the sentences was assessed by analysing effects of semantic violations in sentence completions (final word). Accuracy and reaction times (RT) were modeled as a function of one between-groups factor: Native language (English, Mandarin Chinese), and one within-subject factor: RTM (correct, PPM, FPM). Accuracy was analyzed using a binomial logistic regression. Reaction time data were log transformed and analyzed based on linear mixed effects modeling using R (R Development Core Team, 2008) and the lme4 library (Bates, Maechler, & Dai, 2008).  $\beta$ -values are reported and tested at p < 0.05. As recommended by Barr, Levy, Scheepers and Tily (2013), we modeled the maximal random effect values of participant and items intercepts and slopes across groups and condition in both models (when models successfully converged).

#### 2.1.2.5 ERP recording and Analysis

Electrophysiological data were recorded at a rate of 1 kHz from 64 Ag/AgCl electrodes according to the extended 10-20 convention using an online (0.05 - 200 Hz) bandpass filter. Two additional electrodes were used to monitor eye movements, one below and one above the right eye. Electrode Cz was the reference electrode and impedances were kept below 5 kΩ. EEG data was filtered bandpass using zero-phrase shift digital filtering (0.1 Hz, 24 dB/oct- 20 Hz, 48 dB/oct). Periods of EEG instability corresponding to experiment pauses were removed manually as well as major artefacts through visual inspection of the data and then we adopted the procedure proposed by Gratton, Coles and Donchin (1983) to mathematically correct eye-blink artefacts. ERPs were computed from epochs ranging from -200 ms to 1500 ms after the onset of critical word, always in second position within the main clause. For the semantic violation condition, epochs ranged from -200 ms to 1500 ms, so as to coincide with onset of the sentence-final word. Epochs with any activity exceeding  $\pm 100$ µV at any electrode site except electroocculogram channels were eliminated. More than 30 trials in each participant and condition were included in the averaging procedure. Baseline correction was performed in reference to pre-stimulus activity and individual averages were digitally re-referenced to the global average reference. All analyses were conducted again using the average of the mastoid electrodes as reference and all effects reported based on the global average reference were qualitatively replicated in this analysis.

For RTM analyses, we measured ERP amplitudes over 6 centroparietal electrodes, CP1, CPz, CP2, Cz, C1, C2 at which the N400 is usually maximal (M. Kutas & Hillyard, 1980a, 1980b, 1984b). We identified two time-windows for analysis, the usual N400 time-window between 350–500 ms after the onset of the critical word (the verb in the main clause: e.g., *worked*) and a window between 1200–1350 ms corresponding to the N400 window of the post-critical word. For semantic violation analyses, N400 modulations were analysed between 350–500 ms after the onset of the final word.

# 2.1.3 Results

#### 2.1.3.1 Behavioural data

*Reference Time Misalignment.* Collinearity was not an issue in the models used for analysis: Fixed-effects correlations (|r|) were less than 0.7 for all predictors. For both accuracy and RT models, the intercept represents the average likelihood that English participants were accurate in the control condition. Each coefficient compares the average for a different combination of fixed factor levels against this intercept, and *p* values are derived from the normal approximation method (Barr et al., 2013, see **Table 4** and **5**).

Model 1 (English at int.)	Accuracy		Reaction Time			
	b	SE	Z	b	SE	t
Int. English / Control	4.59	0.51	8.99***	6.40	0.09	71.17
FF1. English / PPM	-8.64	0.74	-11.65***	-0.02	0.04	-0.52
FF2. English / FPM	-8.27	0.74	-11.12***	0.03	0.04	-0.59
FF3. Chinese / Control	-0.60	0.57	-1.05	-0.03	0.11	-0.27
I1. Chinese / PPM	0.46	0.86	0.54	0.11	0.06	2.00*
12. Chinese / FPM	0.84	0.88	0.96	0.06	0.06	1.04
Model 2 (Chinese at int.)		Accura	cy	Reaction Time		
-						
	b	SE	Ζ	b	SE	t
Int. Chinese / Control	3.91	0.38	10.41***	6.37	0.08	80.54***
FF1. Chinese / PPM	-8.07	0.62	-12.92***	0.09	0.04	2.36*
FF2. Chinese/ FPM	-7.33	0.59	-12.52***	0.03	0.04	0.9
FF3. English/ Control	0.78	0.58	1.34	0.02	0.11	0.27
I1. English / PPM	-0.71	0.86	-0.83	-0.11	0.06	-2*
I2. English /FPM	-1.06	0.87	-1.22	-0.06	0.06	-1.04

Table 4. LMM analyses for Reference Time Misalignment behavioral data

\* p < .05; \*\* p < .01; \*\*\* p < .001; FIXED EFFECTS (FF): Int. is the intercept (baseline) condition against which all other contrasts are compared. FF1 & FF2 consider only the data from the group

used for intercept, and examine whether performance differs between the specified conditions and the intercept. FF3 examines differences between groups in the control condition. INTERACTION EFFECTS: I1 & I2 examine whether comparison group participants' performance on the PPM condition (I1) or FPM condition (I2) differs from their own performance in the control condition, and how the magnitude of this difference compares to that shown by the English group. Thus, whilst a significant interaction signals group differences as a function of condition, no significant interaction shows that both groups behaved similarly across conditions.

Both PPM and FPM yielded significantly lower accuracy compared with the past perfect control condition (**Fig. 14** and **Table 4**, *FF1* and *FF2*). There was no fixed effect of Native language (*FF3*), indicating that English and Mandarin participants had similar performance for baseline control sentences. No interaction was found (*I1* and *I2*), suggesting that bilingual Mandarin participants were similarly *inaccurate* in detecting either type of RTM.





As for RTs, English monolingual participants showed no processing time costs for RTM conditions compared with the intercept condition, and Mandarin-English bilinguals did not differ from English monolinguals at baseline (**Table 4**, *FF1*, *FF2*, *FF3*). However, a significant interaction in the analysis showed that Mandarin bilinguals had longer RTs in the PPM condition relative to the additive contribution to the model of their RTs in the control condition, and the RTs of the English natives in the control and PPM conditions (*I1*). No such interaction emerged for the FPM, however (*I2*).

*Semantic Violation.* Accuracy and reaction times (RT) were modeled as a function of native language (English, Mandarin Chinese) as between-group factor, and semantic violation

(control, semantic violation) as within-participant factor. In all other respects, our models were implemented similarly to those described in the previous section: Intercept values comprised the average likelihood that English monolingual participants were accurate in the baseline control condition. For accuracy data (see **Fig. 15**), all participants performed at ceiling (> 97% accuracy on average).



**Figure 15**. Mean accuracy and Reaction time of semantic judgement for control condition and semantic violation condition for both English native participants and Chinese English bilingual participants. Error bars represent SE of the mean.

	Model 1 (English at int.)		Accura	асу		Reaction	Time
		b	SE	Z	b	SE	t
	Int. English / Control	4.45	0.52	8.52***	6.40	0.09	70.87***
	FF1. English / PSV	0.67	0.97	0.69	-0.03	0.06	-0.48
	FF2. Chinese / Control	-0.63	0.59	-1.07	-0.03	0.11	-0.26
	I1. Chinese /PSV	0.65	1.14	0.57	0.23	0.09	2.63**
	Model 2 (Chinese at int.)		Accur	acy		Reaction	Time
		b	SE	Z	b	SE	t
70.87***	Int. Chinese / Control	3.84	0.37	10.36***	6.37	0.08	80.35***
-	FF1. Chinese / PSV	1.11	0.73	1.53	0.20	0.06	3.18**
-	FF2. English/ Control	0.63	0.59	1.06	0.03	0.11	0.27
2.63**	I1. English /PSV	-0.04	1.33	-0.03	-0.23	0.08	-2.71**

0.06 0.11 0.09 Table 5. LMM analyses for Semantic Violation behavioral data

\* \*p < .01; \*\*\* p < .001; FIXED EFFECTS (FF): Int. is the intercept (baseline) condition against which all other contrasts are compared. FF1 consider only English participants' data, and examine whether performance differs on the PSV vs. Control condition (FF1). FF2 examines differences between Chinese and English participants on the Past Perfect control condition. INTERACTION EFFECTS: 11 examine whether Chinese participants' performance on the PSV condition (11) differs from their own performance on the control condition, and how the magnitude of this difference compares to that shown by the English group. Thus, whilst a significant interaction signals group differences as a function of condition, no significant interaction infers that both groups behave similarly across conditions.

Thus, no significant differences emerged in the model (**Table 5**, *FF1*, *FF2 and 11*). For RT data, English monolinguals took the same amount of time to respond to the control as the semantic violation sentences (*FF1*), and Mandarin-English bilinguals did not differ from English controls at baseline (*FF2*). However, a significant interaction emerged, such that Mandarin-English bilinguals were significantly slower responding to semantic violations than the additive contribution to the model of their own performance on correct trials and English monolinguals' performance in correct and semantic violation trials (*II*).

#### 2.1.3.2 Electrophysiological data

**Reference Time Misalignment.** ERP mean amplitudes were analysed with repeatedmeasures analyses of variance (ANOVA). Mean amplitudes were modelled as a function of native language (English, Mandarin Chinese) and condition (past perfect control, PPM and FPM). The N400 effect was examined after the onset of the critical verb (henceforth N4-1) and after the onset of the next word in line (henceforth N4-2).

N4-1: There was a significant main effect of native language (F (1, 44) = 7.35, p = 0.01,  $\eta_p^2 = 0.14$ ) on N4-1 mean amplitude, and a significant interaction between native language and condition (F (2, 88) = 4.84, p = 0.01,  $\eta_p^2 = 0.1$ ). The condition main effect was not significant (F (2, 88) = 0.97, p = 0.38,  $\eta_p^2 = 0.02$ ). In English controls, N400 negativity was significantly greater in the PPM than in the baseline condition (t (18) =1.39, p = 0.09; one-tailed t-test). In Chinese-English bilinguals, however, the difference between PPM and baseline condition did not attain statistical significance (t (26) = -0.55; p = 0.29; the one-tailed t-test; **Fig. 16a; Appendix C1**). As for the FPM versus baseline comparison, native speakers of English had significantly more negative N400 amplitudes in response to FPM (t (18) = 2.64, p = 0.01; one-tailed t-test, **Appendix C2**) but we found no such difference in the Chinese-English bilinguals (t (26) = -1.62, p = 0.06; one-tailed t-test). In fact, the difference tended to go in the opposite direction in Chinese participants (**Fig. 16b**).



**Figure 16.** ERPs elicited by reference time alignment manipulations. (a) ERPs elicited by the critical word and post-critical word in the past perfect control condition (black lines) and the PPM condition (blue line); (b) ERPs elicited by the critical word and post-critical word in the past perfect control condition (black lines) and the FPM condition (red line). Left, English natives; Right, Chinese-English bilinguals. ERPs depict variations of a linear derivation of channels C1, C2, Cz, CP1, CP2, and CPz.

N4-2: There was a significant main effect of native language on N4-2 mean amplitudes (F (1,44) = 42.50, p < 0.001,  $\eta_p^2 = 0.25$ ), and a significant interaction between native language and condition (F (2,88) = 3.48, p = 0.035,  $\eta_p^2 = 0.07$ ). However, the main

effect of condition was not significant (F (2,88) = 2.66, p = 0.08,  $\eta_p^2 = 0.06$ ). In English native controls, N400 mean amplitudes elicited by the post-critical word (N4-2) differed statistically between PPM and baseline conditions (t (18) = 2.54, p = 0.01; one-tailed t-test) but this difference was not significant in the Chinese-English bilingual group (t (26) = -0.3, p = 0.38; one-tailed t-test; **Fig. 15a; Supplementary Fig. B1**). FPM and baseline conditions also difference significantly in native speakers of English (t (18) = 2.74, p < 0.01; one-tailed t-test; **Supplementary Fig. B2**) and again, this difference was not significant in the Chinese participants (t (26) = -0.29, p = 0.39; one-tailed t-test; **Fig. 16b**).

Semantic Violation. The N400 elicited by the sentence-final word was analysed using a repeated-measures ANOVA with native language (Mandarin Chinese, English) as betweengroup factor and semantic violation (control condition, semantic condition) as within-subject factor. Only the main effect of semantic violation proved statistically different (F (1,44) =  $20.58, p < 0.001, \eta_p^2 = 0.32$ ; Fig. 17); there was no significant effect of native language (F (1,44) =  $0.75, p = 0.75, \eta_p^2 = 0.002$ ), nor was there an interaction between native language and semantic violation (F (1,44) =  $3.43, p = 0.07, \eta_p^2 = 0.07$ ). Both native speakers of English and Chinese-English bilinguals showed greater negativity in the N400 range for semantic violations as compared to control sentences (English: t (18) = 3.39, p = 0.002; one-tailed t-test; Mandarin Chinese: t (26) = 2.6, p = 0.008; one-tailed t-test).



**Figure 17.** ERPs elicited by the last word in the past perfect control condition (black line) and the semantic violation condition (yellow line). Left, English native speakers; Right, Chinese English bilinguals. ERPs depict variations of a linear derivation of channels C1, C2, Cz, CP1, CP2, and CPz.

### 2.1.4 Discussion

Here we examined whether linguistic differences between the Mandarin Chinese and the English language in the domain of temporal encoding influences speakers' sensitivity to temporal sequence violations. Despite showing a lack of metacognitive awareness regarding the semantic acceptability of sentences featuring a reference time misalignment, native speakers of English produced a significant N400 modulation in response to the verb in critical position as well as the post-critical word. As expected, N400 mean amplitude was increased for verbs inducing an RTM as compared to verbs that did not conflict with the reference time of the adjunct clause. Critically, no such effect was found in fluent Mandarin-English bilinguals with a good command of English grammar: Not only were they indistinguishable from their monolingual peers in terms of acceptability judgement, but in addition, there was no measurable shift in N400 amplitude in any of the conditions. Nevertheless, lexical-semantic violations elicited a classical N400 modulation in both groups, suggesting good levels of overall sentence meaning integration.

Most previous research investigating tense processing in L2 learners, healthy L1 speakers, or language impaired L1 speakers have tended to use simple constructs in which a time adverbial and the tense form are incompatible (i.e., the tense violation paradigm; see Newman, Ullman, Pancheva, Waligura, & Neville, 2007; Steinhauer & Ullman, 2002; White, Genesee, & Steinhauer, 2012). These studies focused on ERP amplitude modulations of either the P600, the left anterior negativity (LAN), or a biphasic LAN-P600 complex. However, in the current study, we did not investigate tense violations occurring within a given clause, but rather reference time misalignment *between* two clauses, each of them being grammatically correct when considered independently. Tense, thus, had global rather than local temporal relevance in our study. It was therefore only when participants encountered the verb of the main clause that they were in a position to perceive a reference time misalignment, bearing in mind that the tense used in the main clause did not constitute a tense violation until they were able to recall the temporal information encoded in the first clause. And indeed, an explorative analysis revealed no significant modulation of P600 mean amplitudes in either group of participants and for any experimental contrast.

We expected that native speakers of English would identify RTMs or –at the very least– that they would detect them more often than their Mandarin-English bilingual peers. This is because an RTM produces a semantic incongruence at the level of the entire statement

and results in a content that effectively does not 'make sense'. The absence of RTM detection in the behavioural data suggests that the information conveyed by tense can be subtle, especially when the misalignment depends on long-range integration of information across two clauses. This may be explained by automatic repair mechanisms in reading, especially in the context of this experiment in which we used word-by-word presentation and given that RTM differences are rather difficult to identify in general. Indeed, word-by-word presentation (Kaiser, 2014; Marinis, 2010; VanPatten, 2014) is very unnatural (even though it is often imperative in ERP studies of reading) and it is likely to tap into working memory more than natural reading, which may have contributed to blurring the events' timeline. Also, the task used in the experiment likely biased the participants to make basic semantic adequacy judgements because of the presence of a clearly aberrant word in the semantic violation condition. In a recent study by Nieuwland (2015), participants were required to either explicitly assess stimulus plausibility or simply read the same statements for comprehension. In both case, participants displayed larger N400 amplitudes for stimuli which were inconsistent with real-world knowledge. In addition, our data is consistent with recent findings from the language comprehension literature, in which language processing is construed as "good enough" (characterized by underspecified grammatical representations). For the purposes of our offline task, participants may have been using a simple heuristic to interpret these sentences according to existing schemata; avoiding full linguistic computation since this was task-irrelevant (e.g., Ferreira & Karimi, 2015; Ferreira & Patson, 2007). Thus, the subtle between-clause violations in FPM and PPM conditions may have passed participants' initial scrutiny in terms of the degree to which these sentences "made sense".

Critically, however, English speakers did process the tense configuration of the matrix clauses as indicated by a modulation of the N400 elicited by the post-critical word following the locus of a reference time misalignment in the case of PPMs, and both the critical verb and the following word in the case of FPMs. We interpret this result as showing that the temporal representation of events was successfully extracted on the basis of tense information by native speakers of English, even though this did not translate into behavioural effects. Note that the RTM resulted in an N400 modulation as early as the critical verb for FPM but only at the post-critical word in the case of PPM. Even though we did not predict such a difference, we could have anticipated this on the basis of the magnitude of the misalignment. Indeed, an FPM is arguably more salient than a PPM, due to the time gap being wider. In addition, recall that it is a matter of debate whether or not the future form in English qualifies as tense, due to

the mandatory use of the auxiliary 'will'. In other words, it could be that the auxiliary produced a strong expectation for a shift of the reference time into the future, leading to more salient incongruence than in the case of the PPM.

It must be noted that although the reference time of the present perfect is the present, it is mostly used to describe events that have happened in the recent past, that is, in the prepresent zone (Declerck et al., 2006). We propose that this contributes to making the PPM condition relatively less contrastive than the FPM condition. In this case, one could reasonably expect N400 modulations to appear later for PPM than FPM conditions, an effect akin to a spill-over, which is commonly observed in eye-tracking studies (Kaiser, 2014; Keating, 2014; Reichle, Pollatsek, & Rayner, 2007). Beyond the fact that spill-over effects have previously been identified in studies of tense violation in L1 (e.g. Qiu & Zhou, 2012), it is unsurprising that the N400 modulation elicited by RTM should carry over to the post-critical word because of the requirement for cross-clause integration in order to retrieve the temporal relationship between the two events described.

One may wonder if the language proficiency of our Mandarin-English bilingual participants was high enough to detect RTMs. First, the native speakers of Mandarin Chinese involved in this study performed with a very high level of accuracy in the semantic violation conditions, on a par with their English native peers, indicating that their level of comprehension was indeed excellent. Second, most of the bilingual participants obtained a high score at the IELTS, a standard test of English proficiency. Although 5 participants did not provide a score, their English proficiency level was expected to be high since they received instruction exclusively through the medium of English from high-school onwards. We also conducted a split-group analysis excluding the participants without a numerical score to test for potential differences in RTM sensitivity in relation to IELTS score and found that the latter failed to relate to the former (see **Supplementary Analysis; Appendix C3**).

Importantly, all the bilingual participants involved in this study reported having high English proficiency (see **Fig. 12**). Based on an extensive review of the literature (e.g. LeBlanc & Painchaud, 1985; Palmer & Bachman, 1981; Rea, 1981; von Elek, 1981; 1982), Blanche and Merino (1989) concluded that self-reports provide "good or very good" measures of proficiency, and such measures are often used in ERP experiments involving bilingual participants since they are very quick to obtain (e.g., Dowens, Guo, Guo, Barber, & Carreiras, 2011; Lehtonen et al., 2012). To further assess the role of proficiency in the results, we tested an additional group of 21 Chinese-English bilinguals closely matched in IELTS score with the participants tested here on an overt time alignment judgment task along with a new group of native English controls. This new group of Chinese-English bilingual performed similarly to their English native peers (see **Supplementary Analyses; Appendix C3**). Therefore, we assume that low proficiency in English is not the reason why Chinese participants failed to detect PPMs and FPMs.

Note that Mandarin-English bilingual speakers needed a longer time to judge whether PPM sentences were acceptable as compared to control ones. It may be that re-evaluation mechanisms taking place over the processing of the entire sentence were longer in this condition because the sentences were in fact perceived as 'strange', but this effect could simply stem from the ambiguity of the present perfect form itself: is it a past form or a present form? Independently of whether there is an RTM between clauses, the delay in the PPM condition would then be due to internal processing issues rather than RTM resolution.

It is thus likely that the lack of detection of RTMs in the bilingual participants relates in some way to cross-linguistic differences between Mandarin Chinese and English, and more particularly, to the way temporal information about events is conveyed by language. Although Mandarin Chinese, just like English, features the perfective aspect, it has no direct equivalent for tense. This means that Mandarin-English bilinguals reading a perfect form in English will know that the particular event described in the adjunct clause is completed but will have difficulty figuring out when completion occurs: past, present, or post-present.

The relative inability of Mandarin-English bilinguals to perceive RTMs in English may have implications well beyond the domain of second language sentence comprehension and indeed concern time conceptualisation more generally. Given that native speakers of Mandarin Chinese tend to culturally care about the past more than their Canadian counterparts (Ji, Guo, Zhang, & Messervey, 2009), difficulties in identifying temporal relations in English may lead to significant misunderstandings in everyday language use. In other words, Mandarin-English bilinguals may be expected to experience a blurred relationship between past and present when interacting in English, which would stand in stark contrast with their experience of the same relationship in their native language.

# 2.1.5 Conclusion

Mandarin-English bilinguals do not register reference time misalignments between clauses when event time is encoded by tense in English, and probably over-rely on the adverbial form "after" to figure out temporal order, since the same is used in their native language to specify temporal sequencing. Despite such strategy, they fail to accurately position two events in relation to one another on the timeline, which becomes blurred as a result. Future studies will determine whether such effects remain when participants are directly instructed to process temporal sequences.

 $\infty$ 

# 2.2 Study 2

# No real-time semantic processing of tense in fluent Chinese-English bilinguals

#### Abstract

Chinese-English bilinguals find it difficult to distinguish between past and recent past in English, e.g., 'I ate' vs. 'I have eaten', presumably because both phrases translate into the same form in Mandarin Chinese, a tenseless language. However, it is unknown whether Chinese individuals actually have a different conceptualisation of the recent past, or whether they superficially struggle with a grammatical representation that does not exist in their native language. Here, we show that highly proficient Chinese-English bilinguals confuse past and just-completed events whilst making judgements about time in English. We recorded event-related brain potentials in participants making acceptability judgments on sentences that could either feature a time clash between clauses or not. For instance, the sentence "\*After she has worked in this hospital for ten years, she retired to Spain" features a time clash because an event situated in the past cannot follow a just-completed event. Behaviourally, both bilinguals and native English controls were able to detect time clashes, showing that Chinese participants could apply tense rules explicitly. However, the N400 elicited by time clashes in English controls was entirely absent in Chinese participants. Thus, Chinese-English bilinguals do not process the meaning of tense-conveyed temporal information online. In contrast, when time information is lexically-conveyed by a modal auxiliary (e.g., *jiang* - 'will'), the same participants process it online as they read English. We conclude that despite their mastery of English grammatical rules, high-functioning Chinese-English bilinguals fail to conceptually distinguish past and recent past on the basis of English tense.

*Keywords:* Grammatical tense, Bilingualism, Semantics, Event-related brain potentials, Proficiency, N400

#### Significance statement

Presumably because grammatical tense is absent in their native language, Chinese-English bilinguals have difficulty distinguishing between different time periods situated in the past when reading in English. Using measures of brain activity, we aimed to establish whether or not Chinese-English bilinguals make a conceptual distinction between past and recent past when this information is conveyed by tense. Participants were able to apply explicit knowledge of English tense when detecting time clashes between clauses describing events, but they did not show the modulation of brain activity that indexes conceptual processing of time. These results reveal for the first time the striking difficulty for processing the meaning of tense experienced by Chinese individuals who are otherwise highly functional in English.

# 2.2.1 Introduction

Temporal cognition is a core feature of the human mind. Despite the universal relevance of time processing, languages vary in the ways 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 tense share the same translation in Mandarin Chinese, e.g., 'he had retired', 'he retired', and 'he has retired' all translate into *tuixiu-le*. In English, Declerck, Reed, and Cappell (2006) defined the prepresent (grammatically referred to by the present perfect tense) as the portion of the present time sphere that precedes now. Chinese, on the other hand, only encodes perfectiveness (Li & Thompson, 1989; Wang & Sun, 2015) and thus does not distinguish the past and recent past, that is, does not grammatically distinguish the pre-present (present perfect) from either the past (past simple) or a time point in the past by 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 is in the past time sphere (see Fig. 18).

	Past time sphere			ent time sphere	Future time sphere
English	had retired	retired	has retired	retires	will retire
Time line		Past	- Pre-Present	t <sub>0</sub> Post-Present	> Future>
Chinese	tuxiu-le	tuixiu-le	tuixiu-le	tuixiu	jian-yao tuixiu
		Past time sphere		Present time sphere	Future time sphere

**Figure 18.** Linguistic differences between English and Chinese for the encoding of temporal information.

There is evidence that Chinese learners of English struggle with tense (Hawkins & Chan, 1997; Hawkins & Liszka, 2003; Lardiere, 1998). It is unknown, however, whether these difficulties are due to bilinguals' imprecise semantic representations of time when they read in English, or whether tense creates additional grammatical processing demands because it is absent in the native language. Event-related brain potentials offer a compelling index of semantic processing which can be used to detect difficulties with conceptual processing of tense: The N400 (Kutas & Federmeier, 2011; Kutas & Hillyard, 1984a, 1984b; Kutas et al., 1984; Van Petten & Kutas, 1990).

Here, we set out to determine why highly fluent Chinese-English bilinguals experience difficulty with tense in a task requiring temporal order acceptability judgements regarding complex sentences in the form of "After *event 1, event 2*". In these sentences, the event 1 – event 2 time sequence was acceptable or not depending on temporal information conveyed by tense. In a first set of 4 conditions, we compared present–past time clashes with past–past acceptable sequences to evaluate tense processing abilities in our bilingual participants, using a group of native English speakers as controls (**Table 6**). Given that Chinese does not distinguish between present perfect, past simple, and past perfect, we tested both perfect and simple forms to thoroughly evaluate tense sensitivity. To ascertain the specific role of tense, we also compared processing of past–future time gaps with present– future acceptable sequences to determine whether a distinction not conveyed by tense could be processed by Chinese-English bilinguals comparably to English controls. Indeed, the future in English and Chinese is conveyed by a modal auxiliary (*jiang*–'will') and should thus not yield the same difficulties as temporal information conveyed by tense.

Test	Tense sequence	Adjunct clause describing event 1	Main clause describing event 2	Condition labels
Time clash	Past perfect – past	After the director of the school <i>had resigned</i> from the University,	he <i>worked</i> for a multinational.	Past perfect acceptable
	Past – past	After the director of the school <b>resigned</b> from the University,		Past simple acceptable
	*Present perfect – past	*After the director of the school <i>has resigned</i> from the university,		Present perfect clash
	*Present – past	*After the director of the school <b>resigns</b> from the university,		Present simple clash

 Table 6. Examples of adjunct and main clauses describing event 1 and 2 and corresponding experimental conditions

Test	Tense sequence	Adjunct clause describing event 1	Main clause describing event 2	Condition labels
Time gap	Present perfect – future	After the director of the school <i>has resigned</i> from the university,		Present perfect acceptable
	Present – future	After the director of the school <i>resigns</i> from the university,	he <i>will work</i> for a multinational.	Present simple acceptable
	*Past perfect-future	*After the director of the school <i>had resigned</i> from the University,		Past perfect gap
	*Past-future	*After the director of the school <b><i>resigned</i></b> from the University,		Past simple gap

For the time clash manipulation, we predicted that whilst Chinese-English bilinguals would be able to detect clashes in a similar fashion to a control group of native English speakers, crucial differences would emerge at the level of semantic processing: In English participants, N400 amplitude increases should be observed for both time clashes and time gaps as compared to acceptable event sequences. In contrast, we expected that Chinese-English bilingual participants would confuse present perfect, past simple, and past perfect and thus show poor discrimination ability in the N400 range when comparing the present perfect condition to the past conditions. For the time gap manipulation, we anticipated past–future time gaps to elicit N400 modulations in all participants, since the future is conveyed by a modal auxiliary in both English and Chinese, but we again expected that Chinese-English bilinguals would perceive the present perfect as referring to the past, triggering an abnormal N400 modulation in the present-perfect acceptable condition (**Table 7**). Given that bilingual participants were recruited for having achieved a good mastery of English grammar, we expected them to perform well in the temporal sequence acceptability judgement task.

Table 7. N400 differences expected	in the different experimental conditions
------------------------------------	--

Contrasts	English Group	Chinese Group
Present perfect clash vs Past perfect acceptable	yes	?
Present perfect clash vs Past simple acceptable	yes	?
Present perfect clash vs Present simple clash	no	?
Present perfect acceptable vs Past perfect gap	yes	?
Present perfect acceptable vs Past simple gap	yes	?
Present perfect acceptable vs present simple acceptable	no	yes
#### 2.2.2 Materials and Methods

#### 2.2.2.1 Participants

Twenty-one Chinese-English bilinguals (Female:16; Male: 5; Mean age = 25, range from = 19-35; 20 right-hand and 1 left-hand) and twenty-four native English speakers (Female: 13; Male: 11; Mean age = 20, range from 18 -26<sup>6</sup>; 21 right-hand and 3 left-hand) took part in the study. Data from five English native speakers were discarded due to their low accuracy in explicit judgement. In addition, one English and one Chinese participant were removed from analysis due to heavy eye-blinking, alpha wave, and artefact contamination. The average self-reported English proficiency of the Chinese-English bilingual participants according to the International English Language Testing System was 6.9 out of 9 (SD = 0.5, range from 6.5-8; see **Fig. 19**). One Chinese-English bilingual was further removed as he failed to achieve above 67% in the Oxford Quick Placement Test (19). The English proficiency for the rest of the group according to the Common European Framework of Reference (CEFR) was above B2. The average accuracy achieved by the Chinese-English bilingual participants in the Oxford Quick Placement Test (SD = 5%, range from 70%-90%).



**Figure 19.** Chinese-English bilingual participants' self-estimation of their English and Chinese level (10-point scale). Error bar represents standard error.

<sup>&</sup>lt;sup>6</sup> T-test indicated that Chinese-English group were older than English group.

The average age at which the bilingual participants had started learning English was 9.7 years old (SD = 3.2; see **Table 8**), and they had learnt English for an average of approximately 13.5 years (SD = 3.7). The average length of stay in an English-speaking country was 21 months (SD = 20).

Measure	Mean	SD
Age of L2 acquisition	9.95	2.95
Length of L2 learning (years)	13.47	3.69
Daily Chinese usage (%)	58	20.57
Daily English usage (%)	42	21.29

Table 8. Chinese-English bilingual participants' language backgrounds

All the participants in the native English group declared having a beginner's knowledge of other languages than English or none, and no English native participant reported any knowledge of Chinese. All the native English participants were studying at Bangor University, UK, and received either **£15** or five course credits for their time. The study was approved by the ethics committee of the School of Psychology at Bangor University.

#### 2.2.2.2 Stimuli

The investigation used a total of 70 English sentence sets, each containing 8 complex sentences (**Appendix D**). For four of these sentences, the main clause ended in the simple past, with two featuring adjunct clauses with a temporally acceptable timeline, and the other two featuring a time clash (i.e., present perfect – past and present simple – past). For the other four sentences, the main clause ended in the future. Two of these sentences features adjunct clauses with temporally acceptable timeline, whilst the remaining two sentences feature a time gap (i.e., past perfect – future and past simple – future, **Table 6**). The location of the time violation was always situated in the second word of the main clause. For each sentence set as defined above, we ensured between-subjects counterbalancing of clash and gap violations. Stimuli from the same set were not moreover presented in the same block.

Otherwise, half of the stimuli were past tense regular verbs (suffix -d/ed) and the other half were irregular verbs.

#### 2.2.2.3 Procedure

All participants first filled out a language background questionnaire (**Appendix A**). Bilingual participants were required to complete an English language proficiency test, the Oxford Quick Placement Test, and the time for completion was set to 40 mins. During the ERP session, participants read the first clause of each sentence all at once at their own pace, and then pressed any button in a reaction time box to trigger a word-by-word presentation of the main clause in the centre of the screen for a duration of 300ms (ISI 400ms). Once the participants had finished reading the whole sentence, they were asked to judge whether the temporal order of the whole sentence was correct or not (see **Fig. 20**). After they had finished the ERP session, the participants were asked to judge whether half of the ERP stimuli were temporally reasonable or not using a 5-point scale.



Figure 20. Trial structure and timing of stimulus presentation.

#### 2.2.2.4 Behavioural data analyses

Participants' accuracy ratings were analysed with logit mixed effects models with acceptability, form, 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, Levy, Scheepers, & Tily, 2013). Models were fitted using R (R Core Team, 2012) with *lme4* package (Bates et al., 2008). Effects coding was applied to all predictors before model fitting by subtracting the mean from each dummy coded level. Type III Anovas for main effects and interactions were computed using *car* package (Fox &Weisberg, 2014). Estimate effects for main effects were computed with lsmeans (Lenth, 2016), as well as z-values for planned comparisons.

#### 2.2.2.5 ERP data collection and processing.

Electrophysiological data were recorded at a rate of 1 kHz from 64 Ag/AgCl electrodes according to the extended 10-20 convention. The reference electrode was Cz and the Impedances were kept below 5 k $\Omega$ . The electroencephalogram (EEG) was filtered online using a bandpass filter (0.05 Hz–200 Hz), and offline using a low-pass, zero phase-shift digital filter (0.1Hz, 24 dB/oct–20 Hz, 28 dB/oct). Eye-blink artefacts were first manually removed via visual inspection of the data. The remaining artefacts were then mathematically corrected based on the procedure advocated by Gratton, Coles, and Donchin (1983). Epochs ranging from -200 to 1200 ms after onset were extracted from continuous EEG recordings. Epochs with activity exceeding  $\pm$  100  $\mu$ V at any electrode site, except the electrooculogram channels, were discarded. Baseline correction was performed in reference to pre-stimulus activity. Individual averages were digitally re-referenced to the global average reference. N400 mean amplitude were extracted in the usual time window, 350–500 ms after onset of the critical word (i.e., the second word of the main clause) at electrodes of predicted maximal sensitivity (C1, Cz, C2, CP1, CPz and CP2 ;10-13).

#### 2.2.3 Results

Only accuracy data is reported, since RTs were collected after the end of the sentence and thus do not reflect online processing of time acceptability based on the critical word of the main clause. RTs were likely contaminated by main clause length variation, since the number of words in the main clauses varied from 4 to 8 in total, that is, between 2 and 6 words after the critical word. This led to a delayed response and thus RTs are not informative. Acceptability judgements were analysed using logit mixed effects models since accuracy is a

binary variable with group, form, acceptability, and their interactions as predictors (*Methods*). As regards ERPs, we report comparisons of N400 mean amplitudes in the classical window between 350–500 ms at electrodes C1, Cz, C2, CP1, CPz, CP2. N400 amplitudes were analysed using two separate 2 (group) x 2 (form) x 2 (acceptability) repeated-measures ANOVAs for time clashes and time gaps. Group was the between-subjects factor (Chinese-English bilinguals, English controls). Form (simple, perfect) and acceptability (acceptable, incorrect) were within-subject factors. Further statistical testing involved planned comparisons for the specific contrasts considered in the predictions above.

#### 2.2.3.1 Behavioural results

**Time Clashes**. Logit mixed effects models conducted on accuracy data revealed a main effect of acceptability,  $\chi^2(1) = 36.28$ , P < 0.001, such that accuracy ratings were higher for acceptable sentences (M = 93%) than for time clashes (M = 79%; effect estimate = 1.67, SE = 0.28). We also found a main effect of form,  $\chi^2(1) = 5.59$ , P = 0.02, such that participants were less accurate in conditions involving perfect forms (M = 83%) than simple forms (M = 90%; effect estimate = -0.55, SE = 0.23). The interaction between these two factors was significant,  $\chi^2(1) = 4.38$ , P = 0.04, but, critically, we also found an interaction between acceptability, form, and group,  $\chi^2(1) = 5.69$ , P = 0.02 (see Fig. 21*A*; Table 9).



**Figure 21.** English controls and Chinese-English bilingual participants' response accuracy. *(A)* Accuracy data for the time clash manipulation. *(B)* modelled data for the time gap manipulation. Error bars represent standard deviation.

Event 2 in the past								
	Past perfec	t acceptable	Present p	perfect clash	Past simple	e acceptable	Present sir	nple clash
Chinese	0.97	(0.01)	0.65	(0.09)	0.87	(0.06)	0.92	(0.02)
English	0.93	(0.02)	0.77	(0.04)	0.94	(0.02)	0.85	(0.02)
Event 2 in the future								
	Past pe	rfect gap	Present per	fect acceptable	Past sin	nple gap	Present simpl	e acceptable
Chinese	0.96	(0.01)	0.78	(0.05)	0.82	(0.08)	0.97	(0.01)
English	0.85	(0.03)	0.87	(0.03)	0.83	(0.03)	0.94	(0.02)

Table 9. Accuracy in each of the experimental conditions and standard error of the mean in parentheses.

Planned comparisons revealed that English participants were less accurate in the present perfect clash condition as compared to both the past perfect acceptable and past simple acceptable conditions, but there was no difference between the present perfect and present simple clash conditions (Past perfect acceptable: effect estimate = -1.50, SE = 0.51, z = -2.95, P < 0.01; Past simple acceptable: effect estimate = -2.02, SE = 0.64, z = -3.17, P < 0.01; Present simple clash: effect estimate = -0.39, SE = 0.45, z = -0.87, P = 0.39). However, Chinese-English bilinguals were less accurate in the present perfect clash as compared to the two acceptable conditions and also the present simple clash condition (Past perfect acceptable: effect estimate = -2.88, SE = 0.53, z = -5.45, P < 0.001; Past simple acceptable: effect estimate = -2.42, SE = 0.64, z = -3.76, P < 0.001; Present simple clash: effect estimate = -1.75, SE = 0.45, z = -3.86, P < 0.001).

**Time Gaps.** We found a main effect of form,  $\chi^2(1) = 8.58$ , P < 0.01, such that accuracy ratings were overall higher for simple forms (M = 89%) than perfect forms (M = 84%; effect estimate = 0.69, SE = 0.32). We also found an interaction between form and condition,  $\chi^2(1) = 19.37$ , P < 0.001. Critically, we found a three-way interaction between group, form, and acceptability,  $\chi^2(1) = 8.220$ , P < 0.01. Planned comparisons revealed that for English participants, accuracy in the present perfect acceptable condition was lower than in the present simple acceptable condition (effect estimate = -0.95, SE = 0.44, z = -2.16, P < 0.001), but not lower than in the past perfect gap (effect estimate = 0.26, SE = 0.57, z = 0.46, P = 0.64), or the past simple gap conditions (effect estimate = 0.21, SE = 0.76, z = 0.28, P = 0.78). However, for Chinese-English bilinguals, accuracy in the present perfect acceptable condition, and both the past perfect and past simple gap conditions (Present simple acceptable: effect estimate = -

2.99, SE = 0.52, z = -5.78, P < 0.001; Past perfect gap: effect estimate = -2.51, SE = 0.62, z = -4.05, P < 0.001; Past simple gap: effect estimate = -1.80, SE = 0.82, z = -2.19, P = 0.03; see **Fig. 21***B*; **Table 9**).

#### 2.2.3.2 ERP results

**Time Clashes.** The repeated measures ANOVA conducted on mean N400 amplitudes revealed a significant main effect of group, F(1, 35) = 5.93, P = 0.02,  $\eta_p^2 = 0.14$ , such that N400 amplitude was overall more negative in bilingual participants than English controls. We also found a significant main effect of acceptability, F(1, 35) = 23.7, P < 0.001,  $\eta_p^2 = 0.4$ , such that N400 amplitudes were more negative in time clash than acceptable conditions. The interaction between group and acceptability was also significant, F(1, 35) = 14.04, P = 0.001,  $\eta_p^2 = 0.29$ , as was the interaction between acceptability and form, F(1, 35) = 4.68, P = 0.04,  $\eta_p^2 = 0.12$ . Critically, we found a significant three-way interaction amongst group, acceptability, and form, F(1, 35) = 4.48, P = 0.04,  $\eta_p^2 = 0.11$  (see Fig. 22).



**Figure 22.** Comparison of event-related potentials in the present perfect clash condition as compared to the past perfect acceptable, past simple acceptable, and present simple clash in English controls and Chinese-English bilingual participants. Waveforms depict the linear derivation of signals recorded at six electrode sites (C1, Cz, C2, CP1, CPz, CP2). The highlighted time-window is the predicted window of maximal N400 sensitivity: 350–500 ms.

Planned comparisons showed that N400 amplitudes in the English control group were significantly more negative in the present perfect clash condition as compared to both the past perfect acceptable and past simple acceptable conditions, respectively [Past perfect acceptable: t(17) = 6.34, P < 0.001; Past simple acceptable: t(17) = 3.03, P < 0.01], whereas no significant differences were found between the two time clash conditions, t(17) = 1.00, P

= 0.33. Chinese-English bilinguals did not show significant N400 amplitude differences between the present perfect clash and any of the other experimental conditions [Past perfect acceptable: t(18) = 0.62, P = 0.54; Past simple acceptable: t(18) = 0.84, P = 0.41; Present simple clash, t(18) = 0.33, P = 0.74]. **Figure 23** 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.



**Figure 23.** Topographical maps of ERP activity across the 64-channel array between 350–500 ms after the onset of the critical word in the present perfect clash condition as compared to the past perfect acceptable (top row), past simple acceptable (middle row), and present simple clash (bottom row). Note that N400 modulations were elicited by all contrasts in English native controls but were not observed for any of the three contrasts in Chinese-English bilinguals.

**Time Gaps**. N400 mean amplitudes significantly varied as a function of group when the main clause event was situated in the future, F(1, 35) = 5.52, P = 0.03,  $\eta_p^2 = 0.14$ , such that English controls showed overall more negative amplitudes than bilingual participants. We also found a main effect of acceptability, F(1, 35) = 9.72, P < 0.01,  $\eta_p^2 = 0.22$ , such that N400 amplitude elicited in the time gap conditions were greater than in the acceptable conditions. In addition, there was a significant main effect of form, F(1, 35) = 7.49, P = 0.01,  $\eta_p^2 = 0.18$ , with conditions involving a perfect form eliciting more negative amplitudes than those in the simple form. Again, a three-way interaction amongst group, acceptability, and form was found, F(1, 35) = 6.93, P = 0.01,  $\eta_p^2 = 0.16$  (see Fig. 24).



**Figure 24.** Comparison of event-related potentials in the present perfect acceptable as compared to the past perfect gap, past simple gap and present simple acceptable condition in English controls and Chinese-English bilingual participants. Waveforms depict the linear derivation of signals recorded at six electrode sites (C1, Cz, C2, CP1, CPz, CP2). The highlighted time-window is 350–500ms.

Planned comparisons in the English control group showed that N400 mean amplitudes were more negative in both the two time gap conditions compared to the present perfect acceptable condition [Past simple gap: t(17) = 2.27, P = 0.04; Past perfect gap: t(17) = 2.49, P = 0.02]. In contrast, the Chinese-English bilinguals showed no significant differences in the same contrasts [Past simple gap: t(18) = -1.48, P = 0.16; Past perfect gap: t(18) = -1.39, P = 0.18]. However, they did show a difference in N400 mean amplitudes between present perfect acceptable and present simple acceptable, t(18) = -4.1, P < 0.01, which was not present in the English controls, t(17) = -1.37, P = 0.19. Figure 25 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.



**Figure 25.** Topographical maps of ERP activity across the 64-channel array between 350–500 ms after the onset of the critical word in the present perfect acceptable as compared to the past perfect gap (top row), past simple gap (middle row) and present simple acceptable (bottom row) conditions. Note that N400 modulations were elicited by both time gaps in English native controls (top two rows). In Chinese-English bilinguals time gaps 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.

#### 2.2.4 Discussion

This study investigated timeline processing in Chinese-English bilinguals reading in English when temporal information is conveyed by tense. More specifically, we sought to determine whether the lack of differentiation between past perfect, past simple and present perfect in Chinese would yield predictable patterns of semantic processing indexed by the N400 when event sequences involved either a time clash or a time gap.

First, 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 peers in terms of explicitly detecting temporal clashes, Chinese-English bilinguals failed to elicit the expected N400 modulations when responding to time clashes. Our findings present the first compelling evidence that although highly fluent Chinese-English bilinguals can acquire the rules governing English tense, they are not able to extract temporal information on-line when reading English tensed forms.

In a previous study by our group (Li, Jones, & Thierry, 2018), we showed that Chinese-English bilinguals struggle with online processing of English tense when encountering time misalignments. Bilingual participants failed to notice that events presented in immediate succession were temporally misaligned when temporal information was conveyed by tense. However, since the task involved detection of semantic violations induced by incongruous sentences endings, participants' attention was not actively directed at temporal information. And indeed, in Li et al. (2018), English controls also failed to behaviourally detect temporal misalignments, but nevertheless showed modulation of the N400 wave. In the present study, critically, participants had to pay direct attention to temporal information, since their instruction concerned the temporal ordering of events, and

78

we tested potential variations elicited by differences in form (perfect vs. simple) to characterise how Chinese-English bilinguals conceptualise the present perfect. Put simply, we found no difference between present perfect, past perfect, and past simple conditions.

If the present perfect, as we hypothesised, is perceived by Chinese individuals as being in the past time sphere, we would expect the present perfect and the present simple conditions to elicit N400s of different amplitude in the bilingual group. However, we failed to observe such a difference in the case of time clashes, even though this contrast can be morphologically marked in Chinese (see Fig.18). There are two possible explanations for this: (a) Chinese participants fail to understand the English material sufficiently, (b) the critical verb of the main clause being a tensed verb, Chinese participants fail to integrate the temporal information associated with the corresponding event in real-time. Explanation (a) is unlikely, because the bilingual participants tested here were not only highly proficient in English (see below for more detailed discussion of proficiency) but also able to perform almost of a par with their English peers in the task. Explanation (b) is more likely since there is no tense in Chinese. We expected, nevertheless, that Chinese-English bilinguals could make distinctions based on perfectiveness since their native language has a way of marking it. However, the use of the perfective in Chinese can be optional in the case of the subordinate clause of a complex sentence, since temporal information is generally disambiguated in the main clause. Thus, it is possible and even likely that tense information of the subordinate clause was entirely overlooked.

We also contrasted performance and N400 modulations in the same participants making acceptability judgements on time gaps, when the critical verb of the main clause was a future form involving a modal auxiliary in either English or Chinese (*jiang* - 'will'). More specifically, we tested whether processing difficulties induced by past–future time gaps would elicit the expected N400 modulation. Again, the pattern of response accuracy was overall similar in bilingual participants and English controls, and N400 amplitudes in the present-perfect acceptable condition were distinctive from time gap conditions (i.e., past perfect gap and past simple gap) in English participants but not in Chinese-English bilinguals.

According to our hypotheses, we also found a significant N400 modulation in the present-perfect acceptable condition compared to the present simple acceptable condition in Chinese participants, although the former condition did not feature a time gap (and indeed this contrast did not elicit an N400 difference in English controls). This result further

79

indicates that Chinese-English bilinguals conceptualise the present perfect as belonging to the past rather than the present time sphere when reading in English.

It must be noted here, that these results could not be merely attributed to a lack of proficiency in English in our bilingual participant sample, given (i) their high IELTS scores; (ii) their high level of self-reported proficiency ; (iii) the fact that they were selected based on a performance in the Oxford quick placement test above 70% (M = 79%), which is advanced or very advanced; (iv) 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 = 85.94%, SD = 0.09; English: M = 86.53%, SD = 0.13; t-test: t (35) = 0.16, P = 0.87).

Overall, we can infer that Chinese-English bilinguals conceptualize the pre-present zone operationalised 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. This being said, the timeline in English is further blurred for native speakers of Chinese by the fact that temporal information in Chinese can be conveyed flexibly, in a variety of ways, e.g., using lexical markers, modal auxiliaries, or grammatical markings. In the present study, we found clear evidence that Chinese-English bilinguals cannot process the meaning of temporal information conveyed by tense online despite their excellent command of English grammatical rules. This suggests that the differences between English and Chinese regarding the encoding of temporal information, and particularly the distinction between recent past and distant past, do not conceptually overlap between speakers of the two languages, consistent with linguistic relativity effects observed on the basis of lexical and syntactic differences between languages (Athanasopoulos & Bylund, 2013; Fausey & Boroditsky, 2011; Flecken et al., 2015; Thierry, 2016).

#### 2.2.5 Conclusion

This study demonstrates that Chinese-English bilinguals with a high level of proficiency in English and a mastery of English grammar are tense-blurred, that is, they are not able to process the meaning of temporal information conveyed by tense as they read English. Critically, we have shown that this effect is not due to insufficient proficiency since they are able to make accurate explicit decisions about tense. More specifically, we show for the first time that, whereas native English speakers conceptualize the pre-resent conveyed by the present perfect as part of the present time sphere, Chinese-English bilinguals conceptualize it as part of the past time sphere. We conclude that English tense represents a significant challenge for Chinese learners of English, who are unlikely to adopt the same temporal framework as their English peers when resolving temporal information. Further studies are required to determine whether the differences in temporal cognition revealed by tense use in English generalise to Chinese individuals who don't know English, which would require dedicated non-verbal tasks targeting temporal perception.

 $\infty$ 

# Chapter 3

Spatiotemporal Metaphors and linguistic relativity

## 3.1 Study 3

# Back to the future? How Chinese-English bilinguals switch between front and back orientation for time

#### Abstract

The ability to conceive time is a vital aspect of human cognition. Here, we tested whether specific expressions of Chinese selectively shape time conceptualisation. Counterintuitively by western standards, the Chinese spatiotemporal metaphor for the day after tomorrow is *hou-tian*, literally "back-day", and the day before yesterday is *qian-tian*, literally "front-day". We engaged Chinese-English bilinguals in a simple interval reporting task: "Is the date you hear one or two days away from today?", irrespective of whether it referred to past or future. For the first time, we actually tested temporal conceptualisation along the sagittal axis since stimuli were played through loudspeakers physically located in the front or in the back of participants. Brain activity revealed interference when space-time combinations clashed with Chinese metaphors, but, surprisingly, only when participants operated in English. This pattern, replicated with year stimuli, shows that native language expressions can affect time conceptualisation selectively in the second language.

*Key words:* Bilingualism, spatiotemporal metaphors, semantics, event-related brain potentials, unconscious processing

#### Significance

Time conceptualisation is a vital aspect of human cognition. We provide the first demonstration that specific expressions in the native language can shift time conceptualisation in bilinguals operating in their second language. We showed this by engaging Chinese-English bilinguals in a simple interval reporting task: "Is the date you hear one or two days away from today?" and measuring interference effects from expressions specific to Chinese that use space to represent time. Brain activity measures revealed interference when space-time stimulus configurations were incompatible with corresponding expressions in Chinese, but only when participants operated in English. Thus, bilinguals can access irrelevant, native language specific information when operating in their second language and this affects the way in which they conceptualise time.

83

#### 3.1.1 Introduction

Conceptualising the passing of time is a core aptitude of the human mind. One of the most common ways to represent time, an abstract concept, is to use space, a concrete concept. However, linguistic metaphors from different languages use spatial axes in different ways. For instance, spatiotemporal metaphors of Chinese frequently refer to the sagittal (front-back) and vertical (up-down) axes to represent time (e.g., Boroditsky, 2001; Lai & Boroditsky, 2013). Western languages, in contrast, tend to rely more exclusively on the sagittal axis.

Languages even differ in terms of orientation along the same axis. Whereas Aymara, like Moroccan, associates the past with the front space (*nayra*) and the future with the back space (*qhipa*), the majority of languages place the future in front and the past in the back (de la Fuente, Santiago, Román, Dumitrache, & Casasanto, 2014; Núñez & Sweetser, 2006). Variations even exist within languages, as is the case in Chinese, which conforms to a future-in-front convention (e.g., *qian-tu* – 'future prospects' literally translates into "front-path") but features exceptions with a reverse orientation along the same axis (e.g., *hou-tian* – 'the day after tomorrow', which literally translates as "back-day", **Table 10**).

Chinese	Pin Yin	English translation	Time	Literal translation
后天	hou-tian	the day after tomorrow	future	'back day'
前 天	qian-tian	the day before yesterday	past	'front day'
后年	hou-nian	the year after next	future	'back year'
前年	qian-nian	the year before last	past	'front year'

Table 10. Spatiotemporal metaphors of Mandarin Chinese conflicting with the future-in-front convention

One fundamental question, however, is whether such linguistic differences are mirrored by differences at a conceptual level, that is, the question at the centre of the linguistic relativity debate (Lupyan, 2012; Slobin, 1996; Thierry, 2016; Whorf, 1956). In the domain of time representation, Boroditsky (2001) has argued based on behavioural data that native speakers of Chinese predominantly conceptualise time along the vertical axis, whereas English natives predominantly embody time along the horizontal axis. However, Chen (2007) and others (January & Kako, 2007; Tse & Altarriba, 2008) have failed to replicate these results and have argued that Chinese speakers also predominantly think of time horizontally despite the existence of vertically oriented spatiotemporal metaphors in Chinese. In order to assess how specific linguistic expressions such as spatiotemporal metaphors influence how speakers of different languages conceive time, we need an implicit, automatic, and unconscious index of conceptual processing that is resilient to strategic effects and does not rely on verbalisation (Thierry, 2016). A well-established such index is the N400 peak of event-related brain potentials (Kutas & Hillyard, 1980, 1984a; Kutas et al., 1984). Here, we set out to test whether spatiotemporal metaphors specific to Chinese that conflict with the future-in-front convention<sup>7</sup> selectively affect time conceptualisation in fluent Chinese-English bilinguals.

We engineered a conflict between metaphor orientation and stimulus presentation along the front-back axis in the space around the participant. To our knowledge, no previous study has physically presented a stimulus in the back space surrounding participants, since all previous studies involved stimuli presented in the visual domain. In *Experiment 1*, we used days of the week as stimuli. For instance, when a participant was tested on a Wednesday, we presented the auditory stimulus 'Friday' through loudspeakers situated in the front of the participant, potentially clashing with the corresponding spatiotemporal metaphor of Chinese as compared to the same stimulus presented in their back, since the Chinese expression for 'the day after tomorrow' literally translates as "back-day" in English. We asked participants to make interval judgements ('Is the date you hear one or two days away from today?'). Critically, sound origin in space was irrelevant as was the future or past reference afforded by the stimuli, and spatiotemporal metaphors were never presented or mentioned.

We expected that Chinese-English bilinguals would experience interference from conflicting metaphors of Chinese in the case of 2-day gaps, but not in the case of 1-day gaps

<sup>&</sup>lt;sup>7</sup> We chose the sagittal (front–back) axis for three reasons: (i) The sagittal axis is the most frequently used; (ii) It is common to Mandarin Chinese and English, which is critical because we tested Chinese-English bilinguals in the UK; (iii) Exceptional violations of the future-in-front convention only occur in Chinese.

since *ming-tian* – 'tomorrow' and *zuo-tian* – 'yesterday' are not spatiotemporal metaphors in Chinese (**Table 11**).

Chinese	Pin Yin	English translation	Relative time	Literal translation
明天	ming-tian	tomorrow	future	ʻbright day'
昨天	zuo-tian	yesterday	past	'yesterday'
明年	ming-nian	next year	future	'bright year'
去年	qu-nian	last year	past	'gone year'

Table 11. Temporal expressions of Mandarin Chinese neutral vis-à-vis the future-in-front convention

In *Experiment 2*, conducted in late 2017 in the same session as Experiment 1, participants made interval judgements about years instead of days. Our predictions were broadly the same as for Experiment 1 (see Fig. 26).



**Figure 26.** Experimental design. In experiment 1, participants heard days of the week presented through loudspeakers set in front of them and in their back. Stimuli depended on the day of testing (e.g., if the current day was Wednesday, stimuli were Monday, Tuesday, Wednesday, Thursday, and Friday in English and *xing-qi yi, xing-qi er, xing-qi san, xing-qi si,* and *xing-qi wu* in Chinese). Participants were instructed to press one button for stimuli one day away (in the future or the past) and the other button for stimuli two days away from the day of testing. For the current day, they had to press both buttons simultaneously (filler trial). In experiment 2, participants heard year labels: twenty-fifteen, twenty-sixteen, twenty-seventeen, twenty-eighteen, and twenty-nineteen (and *er-ling yi-wu, er-ling yi-liu, er-ling yi-jiu* in Chinese). Instructions were the same as in Experiment 1 but response was based on temporal distance in years, 2017 being the year of testing. Congruency is defined based on alignment between sound origin (front / back), temporal reference (future /past), and spatiotemporal metaphors of Mandarin Chinese.

Overall, we predicted that incongruent stimulus configurations involving 2-day or 2year gaps presented from a location incompatible with the orientation embedded in native spatiotemporal metaphors of Chinese would differentially increase the amplitude of the N400 as compared to congruent configurations. In the case of 1-day or 1-year gaps, configurations violating the future-in-front convention were not expected to elicit semantic interference since no relevant spatial information was available, either in Chinese or in English.

#### 3.1.2 Materials and Method

#### 3.1.2.1 Participants

Twenty-four Chinese-English bilingual participants (8 Male and 16 Female; Mean age = 23, range from 19 to 33; 24 right-handed) and 21 native speakers of English (9 Male and 12 Female; Mean age = 22, range from 19 to 34, 4 left-handed and 17 right-handed) <sup>8</sup> participated in this study. All participant took part in both Experiment 1 and Experiment 2. Data from 5 bilingual participants and 4 native speakers of English were discarded due to poor electrophysiological recording quality, excessively high impedances, excessive blinking, or insufficient number of trials per condition. All Chinese participants reported their International English Language Test System (IELTS) score (Mean = 6.3/9, SD = 0.4) and were resident in the UK at the time of testing. Bilingual participants self-reported their proficiency in both English and Mandarin Chinese (**Fig. 27**) and their language background is summarised in **Table 12**.





<sup>&</sup>lt;sup>8</sup> T-test indicated that the two groups did not differ from each other in terms of age.

All participants had normal or corrected-to-normal vision and self-reported normal audition. Participants either **received £15 or 5 course credits** for their participation in the study that was approved by the ethics committee of the School of Psychology at Bangor University. We expected that a sample size of 18 participants in each of the groups would yield suitable statistical power for this experiment based on previous studies targeting similar effects in ERPs and spanning 9 years of research (e.g., Kuipers & Thierry, 2010; Kuipers, Jones, & Thierry, 2018).

Measure	Mean	SD
Age of L2 acquisition	9.42	2.36
Length of L2 learning	14.26	3.57
Daily Chinese usage (%)	67	16.7
Daily English usage (%)	33	14.1

Table 12. Chinese-English bilingual participants' language background

#### 3.1.2.2 Materials

Stimuli consisted of digital audio files of days of the week and year numbers in Mandarin Chinese and English. All stimuli were recorded once in English by a native speaker of English and once in Chinese by a native speaker of Chinese. A cross-splicing procedure (using Adobe Audition<sup>TM</sup>) was employed to ensure that participants could not guess the particular day or year stimulus presented in each trial on the basis of stimulus beginning alone<sup>9</sup>. Cross-splicing offered a good baseline and optimal accuracy in marking the onset of the critical information in the sound stream (see **Fig. 28**).

<sup>&</sup>lt;sup>9</sup> Note that for year stimuli in Chinese, we elected not to cross-splice between the decade digit ( $y_i$  – 'one' in Chinese) and the final digit (5, 6, 7, 8, or 9) because of co-articulation in the case of  $y_i$ - $w_u$  – 'fifteen', which would have created an artefact for that sound file. On average the duration of  $y_i$  was 250 ms (range 230-272 ms), and thus RTs were artificially extended by the same duration in the corresponding condition.



Figure 28. Cross-splicing procedure and stimuli presented. (A) Exp 1 (days). (B) Exp 2 (years). In Experiment 1, stimuli consisted of the names for the 7 days of the week. For any participant, only 5 days of the week were presented in order to cover a time interval of two days before to two days after the day of testing. Average stimulus duration was 900 ±75 ms for days in Chinese and 845 ±66 ms for days in English day. Average auditory stimulus intensity was 48 dB (range 46–55 dB).

In experiment 2, stimuli were 4-digit numbers referring to 5 years surrounding the year of testing (2017), i.e., 2015, 2016, 2017, 2018, and 2019. Average stimulus duration was  $1076 \pm 23$  ms for years in Chinese and  $1163 \pm 66$  ms for years in English day. Average auditory stimulus intensity was 48 dB (range 47–52 dB).

#### 3.1.2.3 Procedure

Participants first completed a language background and reading habits questionnaire whilst being fitted with the cap for electrophysiological recording. They were seated in the centre of a sound-attenuated testing booth, with two speakers located in the front and two speakers located behind them, set at a distance of between 1.4–1.6 meters from their ears. A 19-inch CRT monitor was placed 100 cm in front of their eyes and displayed a black fixation cross on a white background throughout the recording session. In experiment 1, participants were asked to judge whether each stimulus referring to a day of the week corresponded to a period of time situated one or two days away from the current day. In experiment 2, participants made the same judgements for stimuli referring to years. Responses were given by pressing

designated left and right buttons on a response box. Response sides were counterbalanced between participants. Half of the stimuli were presented through the speakers located in front of the participant's chair, and the other half were presented in their back. When participants heard the current day or the current year, they were instructed to press both left and right buttons simultaneously. They heard 30 pseudo-randomly intermixed iterations of each individual stimulus condition. Apart from present day (one fifth of trials), half of the stimuli were one day away from the time of testing and the other half were two days away from the time of testing. Similarly, half of the stimuli referred to the future and half to the past, making a total of 300 trials per block in each experiment. Control native speakers of English performed the task in English only (600 trials in total) and Chinese-English participants performed the task once in English and once in Chinese (1200 trials in total) with order counterbalanced between languages. Every individual trial started with a pink fixation cross displayed in the centre of the screen for 300 ms. The fixation then turned to black after a pseudorandom inter-stimulus interval of 300-500 ms. The target auditory stimulus was then presented through loudspeakers either to the front or the back of the participant's chair whilst the black fixation stayed on the screen until participant's response with a maximum duration of three seconds from the onset of the sound stimulus. Participant's response immediately triggered a 200 ms inter-trial interval before the next pink fixation. Every 7 trials, the pink fixation lasted for four seconds, during which participants were encouraged to blink if they needed to, in order to minimise the occurrence of eye blink artefacts during the interval of time between auditory stimulation and response.

#### 3.1.2.4 ERP recording and processing

Electrophysiological data were recorded at a rate of 1 kHz from 64 Ag/AgCl electrodes according to the extended 10-20 convention and referenced to electrode Cz. Impedances were kept below 5 k $\Omega$ . The electroencephalogram (EEG) was filtered using an online bandpass filter (0.05–200 Hz), and offline using a low-pass, zero phase-shift digital filter (0.1 Hz, 24 dB/oct–20 Hz, 28 dB/oct). Eye-blink artefacts were first manually removed through visual inspection of the data and the remaining artefacts were then mathematically corrected using the procedure advocated by Gratton, Coles and Donchin (1983). Epochs ranging from -200 to 1200 ms after stimulus onset were extracted from continuous EEG recordings. Epochs with activity exceeding  $\pm$  100 µV at any electrode site, except the vertical electroocculogram channels, were discarded. Baseline correction was performed in reference to pre-stimulus activity, and individual averages were digitally re-referenced to the global average reference.

#### 3.1.2.5 Behavioural data analysis

Stimulus onsets were corrected to the onset of the critical information in the sound stream (see Fig. 28). Reaction times (RTs) below 200 ms were removed from the analysis (0.05%). Trials with RTs that deviated 2.5 interguartile range below the 1<sup>st</sup> and above the 3<sup>rd</sup> quartile of each participant in each intra-subject variable were considered outliers and discarded from data analyses (1.49%). Accuracy data and RTs of correct answers were then analysed with logit and linear mixed-effect models respectively [*lme4* (Bates, Maechler, & Dai, 2008) package in R (R core Team, 2012)]. Collinearity was not an issue in the models: variance inflation factor (VIF) ranged from 1 to 1.5. All models included random intercepts for subjects and items and maximal random slopes for each within-subjects and within-items predictor respectively. Following Barr et al. (2013) and Barr (2013) when models with maximal random structure failed to converge, maximal within-items and within-subject interactions for random slopes were used. All fixed effects were contrast coded before analyses using sum coding so that each model's intercept represented the mean value of each predictor. Significance P-values and Type III F-statistics for main effects and interactions for continuous variables (RTs) were calculated using Satterthwaite approximations to denominator degrees of freedom as implemented in the LmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) package, and planned comparisons and  $\beta$  estimates were calculated using difflemeans and lsmeans as implemented in the lmerTest package. Binary outcomes (accuracy data) were analysed using logit mixed-effects models (Jaeger, 2008). Type III Wald  $\chi^2$ statistics, *P*-values, planned comparisons and  $\beta$  estimates for main effects and interactions were calculated using car (Fox & Weisberg, 2014) and incorporated lsmeans packages (Lenth, 2016).

#### 3.1.2.6 EEG data analysis

ERP amplitudes were measured at 6 centroparietal electrodes (C1, Cz, C2, CP1, CPz, and CP2) where the N400 is usually maximal (Kutas & Hillyard, 1980, 1984; Kutas et al., 1984). In experiment 1, for the English day block, mean N400 amplitude were computed between

91

350–500 ms, determined predictively based on previous literature (Kutas & Hillyard, 1980; 1984; Kutas et al., 1984). For the Chinese Day block, the N400 window was 813–963 ms (since *xing-qi* lasted 463 ms, **Fig. 28**). In experiment 2, for the English year block, the predicted time-window of the N400 was 630–780 ms after stimulus onset, given that the 'twenty-' portion of the auditory stream lasted 280 ms (**Fig. 28**). In the Chinese year block, the N400 time window was 869–1019 ms (since *er-ling* lasted for 519 ms).

#### 3.1.3 Results

To analyse our results, we proceeded in four steps. First, we analysed behavioural measures and ERP results from Experiment 1 (days), starting with 2-day gaps, where spatiotemporal metaphor effects were anticipated. We then analysed data for the 1-day gaps where only effects of conventionality could be expected. Third, we analysed data collected in Experiment 2 (years), to establish whether the pattern of results obtained for days would also obtain for years (replication). Starting with 2-year gaps, we tested for spatiotemporal metaphor congruency and then for conventionality effects in the case of 1-year gaps. Reaction times, accuracy data, and ERP's time-windows were corrected to the onset of the critical information in the sound stream.

#### Chinese spatiotemporal metaphors for days affect time conceptualisation

In experiment 1, we tested whether a change of language would affect congruency between spatiotemporal metaphors of Chinese and spatiotemporal configuration of the stimuli in Chinese-English bilinguals in the case of two-day intervals. Accuracy was at ceiling in the interval calculation task whether bilinguals heard day stimuli in Chinese or in English (**Fig. 29***A*). We found no significant main effect of language (English, Chinese) or congruency (congruent, incongruent) on accuracy and no interaction. As for Reaction Times (RTs), we found a main effect of language (F(1,19.53) = 24.66, P < 0.001) so that bilingual participants were slower responding to English ( $\beta = 1057$ , SE = 54) than Chinese stimuli ( $\beta = 861$ , SE = 37). There was no other effect.



**Figure 29.** Behavioural results. (*A*) Two-days gap. (*B*) One-day gap. Bars represent reaction times and bullets represent accuracy. Error bars depict s.e.m.

We then analysed mean N400 amplitudes in the same Chinese-English bilinguals to determine whether spatiotemporal metaphors interfered with time conceptualisation during the task. A repeated measure ANOVA with language (Chinese, English) and congruency (congruent, incongruent) as within-subject factors revealed a significant effect of congruency  $(F(1,18) = 21.83, P < 0.001, \eta_p^2 = 0.55)$ . The effect of language was marginally significant ( $F(1,18) = 4.14, P = 0.06, \eta_p^2 = 0.2$ ) and the interaction between congruency and language was also significant ( $F(1,18) = 7.06, P = 0.02, \eta_p^2 = 0.28$ ). Planned comparisons showed that incongruent stimulus configurations elicited significantly more negative N400 amplitudes than congruent ones when bilingual participants operated in English (t(18) = 4.66, P < 0.001; see **Fig. 30**). No such effect was found when participants responded to Chinese stimuli (t(18) = -0.53, P = 0.3).



**Figure 30.** Event-related brain potentials elicited in experiment 1 (days). ERPs depict the linear derivation of 6 electrodes (C1, Cz, C2, CP1, CPz, CP2). Topographical maps show ERP activity across the 64-channel array in the following predictively determined time-windows. N400 amplitudes were computed between 350–500 ms based on previous literature, from the onset of the unique sound streams, irrespective of language or stimulus. In the case of Chinese stimuli, the interval of N400 amplitude extraction was 813–963 ms (since *xing-qi* – 'week' lasted 463 ms, see Methods) and in the case of English stimuli, the interval of N400 extraction was 350-500 ms, since day stimuli differed from one another from their onset. Topographies depict differences between incongruent and congruent conditions in all cases.

In order to further investigate the congruency effect found in bilinguals operating in English, we compared their results with that of English native participants. Accuracy was at ceiling in English native controls. No significant main effect of congruency (congruent, incongruent) or group (English, Chinese-English bilingual) was found on accuracy and there was no interaction (see **Fig. 29***A*). Regarding RTs, Chinese-English bilinguals operating in English were significantly slower ( $\beta = 1056$ , SE = 48) than their English native peers ( $\beta = 855$ , SE = 50), as reflected by a significant main effect of group (F(1, 34.21) = 8.45, P < 0.001).

A between-subjects repeated measures ANOVA, with congruency as within-subject factor and group (English, Chinese-English bilingual) as between-subject factor conducted on N400 mean amplitude revealed a significant main effect of group (F(1, 34) = 7.95, P = 0.01,  $\eta_p^2 = 0.19$ ) and a significant effect of congruency (F(1, 34) = 5.54, P = 0.02,  $\eta_p^2 = 0.14$ ). The interaction was also significant (F(1, 34) = 5.99, P = 0.02,  $\eta_p^2 = 0.15$ ). Planned comparisons showed that incongruent stimulus configurations elicited more negative N400 amplitudes than congruent configurations in bilingual participants (t(18) = 4.66, P < 0.001; see **Fig. 30**), but not in their English peers (t(16) = -0.05, P = 0.48).

#### Conventionality effects for one-day gaps affect behaviour but not ERP amplitudes

We first tested for effects of conventionality in Chinese-English bilinguals' mind. With regard to accuracy, there was a significant interaction between language (Chinese, English) and conventionality (conventional, unconventional;  $\chi^2_1 = 3.88$ , P = 0.05). However, *post hoc* comparisons failed to show effects of conventionality in either Chinese ( $\beta = -0.68$ , SE = 0.45, z = -1.52, P = 0.13) or English ( $\beta = 0.49$ , SE = .40, z = 1.23, P = 0.22) considered separately. The effect of language in the conventional (z = -1.11, p = 0.27) and unconventional (z = 1.02, p = 0.31) conditions were not significant either. Regarding RTs, a significant main

effect of language (F(1, 20.89) = 7.82, P = 0.01) showed that Chinese-English bilinguals were slower responding to English stimuli ( $\beta = 1043$ , SE = 60) than Chinese stimuli ( $\beta = 880$ , SE = 46; see Fig. 4B). The effect of conventionality was just significant (F(1, 66.96) = 3.88, P = 0.05), bilinguals being slower responding to unconventional ( $\beta = 972$ , SE = 45) than conventional stimuli ( $\beta = 951$ , SE = 45). Amplitude analysis revealed no main effect of conventionality or group on N400 amplitude and no interaction (Ps > 0.1; see Fig. 31).

As was the case in the bilingual group, English participants' accuracy was at ceiling in the one-day gap condition. Analysis comparing the Chinese-English bilinguals in English with the native English controls revealed no main effect of conventionality or group on accuracy and no interaction. As regards RTs, a main effect of group (F(1, 35.01) = 6.29, P =0.02) showed that Chinese-English bilinguals were slower responding to English stimuli ( $\beta =$ 1051, *SE* = 66) than their English native peers ( $\beta = 888, SE = 56$ ). Amplitude analysis only revealed a significant main effect of group ( $F(1, 34) = 6.75, P = 0.01, \eta_p^2 = 0.17$ ) on N400 amplitude and no effect of conventionality (see **Fig. 31**).



**Figure 31.** Event-related brain potentials elicited in experiment 1. ERPs depict the linear derivation of 6 electrodes (C1, Cz, C2, CP1, CPz, CP2). Topographical maps show ERP activity across the 64-channel array in the following predictively determined time windows: 813–963 ms after Chinese stimulus onset and between 350–500 ms after English stimulus onset. Topographies depict differences between incongruent and congruent conditions in all cases.

#### Replication of the spatiotemporal metaphor effect with year stimuli

In experiment 2, as was the case for days, Chinese-English bilinguals were at ceiling in the interval calculation task with two-year gap stimuli in both the congruent and the incongruent conditions and in both their languages (**Fig. 32***A*).



**Figure 32.** Behavioural results. (*A*) Two-years gap. (*B*) One-year gap. Bars represent reaction times and bullets represent accuracy. Error bars depict s.e.m.

Results revealed no significant main effect of language (Chinese, English) or congruency (congruent, incongruent) on accuracy and no interaction (all Ps > 0.1). We found no effect of language of operation or congruency on RTs and no interaction (all Ps > 0.1).

The within-subject repeated measures ANOVA of ERP data revealed a main effect of congruency on mean N400 amplitude in bilingual participants (F(1,18) = 6.96, P = 0.02,  $\eta_p^2 = 0.28$ ) and a significant interaction between congruency and language (F(1,18) = 4.6, P = 0.05,  $\eta_p^2 = 0.2$ ). Replicating the pattern found for 2-day gap calculations, planned comparisons showed that N400 amplitude was significantly greater for incongruent than congruent stimulus configurations when bilinguals operated in English (t(18) = 3.89, P < 0.001; **Fig. 33**) but not when they operated in Chinese (t(18) = 0.31, P = 0.38).



**Figure 33.** Event-related potentials elicited in experiment 2. ERPs depict the linear derivation of 6 electrodes (C1, Cz, C2, CP1, CPz, CP2). Topographical maps show ERP activity across the 64-channel array between 869-1019 ms after Chinese stimulus onset and 630-780 ms after English stimulus onset. The predicted time-window of the N400 for Chinese stimuli was between 869–1019 ms after stimulus onset, given that the er-ling – 'twenty' portion of the auditory stream lasted for 519 ms. In the case of English stimuli, the N400 time analysis window was 630–780 ms (since 'twenty' lasted 280 ms). Topographies depict differences between incongruent and congruent conditions.

As in Experiment 1, we sought to further characterise the congruency effect found for the English condition in bilinguals by comparing their results with that of native English speakers. English participants' accuracy was at ceiling. No significant main effects or interaction between congruency and group was detected. Regarding RTs, Chinese-English bilinguals operating in English were significantly slower ( $\beta = 1002$ , SE = 74) than English native participants ( $\beta = 819$ , SE = 61), as shown by a main effect of group (F(1, 28.13) = 6.96, P = 0.01). No other effects were found (all P > 0.6).

A between-subject repeated measures ANOVA on N400 mean amplitudes showed a significant main effect of group (F(1, 34) = 4.13, P = 0.05,  $\eta_p^2 = 0.11$ ) and a significant main effect of congruency (F(1, 34) = 7.21, P = 0.01,  $\eta_p^2 = 0.18$ ). The interaction between group and congruency was also significant (F(1, 34) = 4.51, P = 0.04,  $\eta_p^2 = 0.12$ ). Planned comparisons showed that incongruent stimulus configurations elicited greater N400 amplitudes than congruent ones in bilingual participants (t(18) = 3.89, P < 0.001; **Fig. 33**), but not in English controls (t(16) = 0.35, P = 0.37).

#### No measurable effect of conventionality in the case of 1-year gaps

As previously, we first compared bilingual participants' performance in English and Chinese using within-subject analyses. No significant main effect of language or conventionality on either accuracy or RT and no interaction were detected (Ps > 0.1; **Fig. 32***B*). The analysis conducted on mean N400 amplitude showed no significant main effect or interaction (Ps > 0.1, **Fig. 34**).



**Figure 34.** Event-related potentials elicited in experiment 2. ERPs depict the linear derivation of 6 electrodes (C1, Cz, C2, CP1, CPz, CP2). Topographical maps show ERP activity across the 64-channel array between 869-1019 ms after Chinese stimulus onset and 630-780 ms after English stimulus onset. Topographies depict differences between incongruent and congruent conditions.

Finally, we compared Chinese-English bilinguals in English with their English native peers. We found a significant interaction between group and conventionality on accuracy ( $x_1^2 = 7.14$ , P < 0.001). However, *post hoc* comparisons showed that there was no effect of conventionality in either Chinese-English bilinguals ( $\beta = 0.46$ , SE = 0.27, z = 1.68, P = 0.09) or English natives ( $\beta = -0.39$ , SE = 0.28, z = -1.43, P = 0.15). As regards RTs, there was a significant effect of group (F(1, 32.92) = 6.04, P = 0.02), bilingual participants ( $\beta = 1109$ , SE = 55) being slower responding to English stimuli than English native participants ( $\beta = 911$ , SE = 58; Fig. 7B). As regards the ERP analysis, we only found a significant main effect of group on mean N400 amplitude (F(1, 34) = 4.56, P = 0.04,  $\eta_p^2 = 0.12$ ; **Fig. 34**).

#### 3.1.4 Discussion

Here we investigated a potential effect of native spatiotemporal metaphors on time conceptualization in Chinese-English bilinguals operating in their native or their second language. When tested in Chinese, participants did not display congruency effects predicted by spatiotemporal metaphors. Strikingly, however, when they were presented with English stimuli, native language representations interfered with time conceptualization as indicated by more negative N400 amplitudes in the incongruent conditions. Importantly, this pattern of result was mostly replicated using years instead of days as auditory stimuli. In contrast, conventionality effects only appeared as subtle behavioural variations in the case of 1-day intervals and did not entail any N400 amplitude modulation.

First, our results are consistent with previous studies that have established unconscious language non-selective access in bilinguals, and particularly Chinese-English bilinguals operating in English (Thierry & Wu, 2007). Indeed, and despite recent attempts to provide an alternative account for this mechanism (Costa, Pannunzi, Deco, & Pickering, 2017; Oppenheim, Wu, & Thierry, 2018), Chinese-English bilinguals appear to automatically access Chinese when processing input in English, because otherwise it would be difficult to account for the interference effects observed here. The results thus expand our understanding of language non-selective lexical activation mechanisms in different script bilinguals (Thierry & Wu, 2007; Wu & Thierry, 2010, 2012) by showing unconscious activation of spatiotemporal metaphor representations of Chinese when participants hear English words.

Our findings are partly compatible with results from previous behavioural studies suggesting that spatiotemporal metaphors can influence individuals' conceptualization of time (Boroditsky, 2001; Casasanto et al., 2004; Fuhrman et al., 2011; Lai & Boroditsky, 2013; Núñez & Sweetser, 2006, but see Chen, 2007; January & Kako, 2007; Tse & Altarriba, 2008). Critically, however, our data establish the locus of interference between language specific expression and time representation at a conceptual level in the absence of participants' awareness, since congruency effects were detected in N400 amplitude modulations rather than behavioural measurements and in conditions where time orientation was irrelevant. Indeed, at debriefing, detailed questioning of the participants revealed no explicit knowledge of hidden manipulations relating to spatiotemporal metaphors. All participants reported having interpreted the task as a simple arithmetic problem, that is, computing an interval of 1 or 2 days, or 1 or 2 years, irrespective of future or past temporal reference. Even when directly confronted with the actual construction of the experiment, none of the participants recognised that the future or past reference afforded by the stimuli should conflict with the location of the speakers through which these stimuli were presented, or having resorted consciously to labelling 2-day and 2-year gaps as "front/backday" or "front/back-year" in Chinese.

In addition, the results shed light on a fundamental question in embodied cognition (see Barsalou, Niedenthal, Barbey & Ruppert, 1999; Casasanto, 2009; Glendberg & Kaschak, 2002), namely whether spatiotemporal metaphors translating temporal concepts into spatial ones are based on bodily experience or are purely linguistic. Note that in previous research (e.g., Boroditsky, 2000; Miles et al., 2011; Yang and Sun, 2016), contrary to the present study, participants were tested based on representations rather than direct embodiment of spatial orientation (e.g., left-right orientation on a figure representing front-to-back axis in Boroditsky, 2000). In the current experiment, the stimuli were presented either from a location in the front or the back space relative to participants' location in real space. Chinese participants systematically exhibited unconscious congruency effects predicted by corresponding native spatiotemporal metaphors in relation to their physical position in space rather than a relative position inferred on the basis of a representation. Critically, the task never required explicit processing of spatiotemporal configurations. Thus, we can conclude that abstract temporal information is implicitly mapped onto the physical space around the participant spontaneously, which provide strong support for the premise of embodied cognition.

It may be considered a surprise, however, that bilingual participants experienced the spatiotemporal metaphor interference effect when performing the task in English rather than Chinese, given that the metaphors belong to Chinese. But this result is in fact compatible with the frequent observation that verbal interference tends to cancel effects of language on conceptualisation (Drivonikou et al., 2007; Gilbert et al., 2006; Roberson & Davidoff, 2000). When stimuli are presented in Chinese, participants suffer within-language competition, such that they cannot verbally recode information because accessing the labels for days and years and engaging in arithmetic computations in Chinese directly compete for selection with metaphoric lexical representations. However, this is arguably not the case when participants operate in English, since no direct within-language competition applies: Metaphors in Chinese can be accessed through cross-language activation. Then, and only then, can interference take place. This mechanistic explanation is consistent with selective interference effects previously shown in bilinguals switching back and forth between their first and second language, whilst making non-verbal decisions on motion events (Athanasopoulos et al., 2015).

100

In other words, we contend that only when participants heard temporal references in English, they accessed conceptually related expressions specific to their native language. For instance, when a participant tested on a Wednesday heard the English word "Monday", they would have activated *qian-tian* (literally translated as "front–day"), given that Monday was the day before yesterday relative to the day of testing. This would arguably not have happened when the same participant was tested in the native language Chinese because of the within-language competition effects described above. Alternatively, this would not have happened because days and years in Chinese contain a digit enabling direct gap calculation (with the exception of Sunday). For instance, *xing-qi yi* – 'Monday' literally translates into "week-1" in English and *er-ling yi-wu* – '2015' literally translates into "two-zero-one-five". Thus, calculating intervals is straightforward in Chinese but not in English, given the previously noted difficulty of bilinguals to compute operation in the second language (Salillas & Wicha, 2012).

As expected, we found a difference between conventional and unconventional control conditions in the case of 1-day gaps in the absence of any metaphorical interference, presumably due to there being no spatiotemporal metaphor for tomorrow and yesterday in either English or Chinese. Indeed, in Chinese, tomorrow is ming-tian (literally, "bright-day"), yesterday is zuo-tian ("past-day"), next year is ming-nian (bright-year), and last year is qunian ("gone-year"), thus any effect of orientation for one day/year gaps could only relate to effects of spatial orientation conventions for time. Conventionality had an effect in experiment 1 (days) but not experiment 2 (years). We contend that this was the case because time conventionality effects weaken as the size of time chunks increases, i.e., it is more difficult to conceptualise the year ahead as in front than the day ahead as in front (Hellström & Rammsayer, 2004; Lewis & Miall, 2003). Furthermore, conventionality did not affect ERP amplitude as metaphor congruency did. Here the argument would be that interference between convention and time representation does not occur at a semantic level but rather in terms of direct mapping between stimulus and response. Spatiotemporal metaphors rely exclusively on language and thus result in a semantic interference effect to start with (here resulting in a measurable N400 modulation). In other words, spatiotemporal metaphors are resolved at a pre-response, semantic level, whereas conventionality effects do not come into play during semantic access but rather interfere directly with the task at hand (particularly in the case of days).

#### 3.1.5 Conclusion

The present study provides the first electrophysiological evidence for a deep, unconscious, and pervasive influence of native spatiotemporal metaphors on time conceptualization in bilinguals. These findings not only bridge unconscious language non-selective access in bilinguals with predictions from linguistic relativity theory but also demonstrate the staggering level of interactivity involved. After all, our Chinese-English bilingual participants suffered semantic interference when the English label of the day after tomorrow was played through loudspeakers located in front of them, as compared to when the same label was played in their back. Given that this did not happen when they listened to the label of tomorrow, or any label in Chinese, and that it generalised to year labels, our study demonstrates that abstract concepts such as that of time are highly permeable to linguistic representations specific of the native language even when bilinguals operate in their second language.

 $\infty$ 

## 3.2 Study 4

#### Time flows vertically in Chinese

#### Abstract

Time is systematically represented along a horizontal axis, whether oriented front-back or left-right, in a majority of western languages such as English. Previous studies have shown a strong tendency in native speakers of Mandarin Chinese to organise the timeline along the vertical axis, but the behavioural evidence obtained so far has been repeatedly contested. Here, we adapted the arrow flanker task by replacing the middle arrow with Chinese spatiotemporal metaphors for months, i.e., 上个月--'last month'(literally 'up month'), and 下个 月--'next month' (literally 'down month') and evaluated metaphor-spatial orientation congruency based on N400 brain potential modulations in Chinese participants asked to report the direction of the flanker arrows. Critically, we tested whether such modulations would generalise to temporal expressions that are not spatiotemporal metaphors of Chinese, namely, 去年 -- 'last year' (literally 'gone year') and 明年-- 'next year' (literally 'bright year'). Whilst behavioural measures displayed differences that were marked for spatial words, attenuated for spatiotemporal metaphors, and absent for temporal expressions, N400 amplitude increased when spatiotemporal metaphor orientation conflicted with flanker arrow direction along the vertical, but not the horizontal axis. As expected, we found significant N400 modulations on the vertical axis for both spatial words (i.e., up / down) and spatiotemporal metaphors but also, crucially, for non-directional temporal Chinese expressions. On the basis of direct brain measurements indexing implicit semantic processing, and in the absence of contrastive behavioural patterns, we thus demonstrate that native speakers of Chinese conceptualise time along the vertical axis.

*Keywords:* Spatiotemporal metaphor; Chinese-English bilinguals; Vertical axis; Horizontal axis; event-related brain potentials

#### 3.2.1 Introduction

Certain gestures and linguistic forms are frequently used to anchor the temporality of events in the spatial domain (e..g, Boroditsky, Fuhrman, & McCormick, 2011; Gu, Zheng, & Swerts; Lai & Boroditsky, 2013). In English-speaking cultures, for example, people often point to the front space when stating that they *look forward* to an event set in the future, or they will *look back* on a memorable life event with nostalgia. Such spatiotemporal metaphors serve as concrete reference in order to conceptualise time, an abstract concept.

Whilst most cultures and languages represent time using space, there is considerable variation in how this relationship manifests. For example, speakers of Aymara and Moroccan will *look back* (rather than forward) to a future event (de la Fuente, Santiago, Román, Dumitrache, & Casasanto, 2014; Núñez & Sweetser, 2006), and in some languages, such as Chinese, the same orientation (e.g., front) can refer either to the future (e.g., *zhan-wang–* 'looking into the future', literally, 'unfold-to gaze') or the past (e.g., *qian-tian–*'the day before yesterday', literally, 'front day'). Moreover, time can be represented along different axes within the same language: English represents time on the horizontal (left-right) axis but also refers to it on the sagittal (front-back) axis, whilst other languages such as Mandarin Chinese, can refer to time along the vertical axis (up-down, e.g., Boroditsky, 2001; Boroditsky et al., 2011; Gu et al., 2018; Lai & Boroditsky, 2013).

Thus, an interesting question is whether the way in which language refers to time relates to time conceptualization in the spatial domain. Several renowned studies have suggested that Chinese-English bilinguals conceptualize time vertically (e.g. Boroditsky, 2001; Boroditsky et al., 2011; Chen, 2007; Fuhrman et al., 2011; Lai & Boroditsky, 2013; Miles et al., 2011; Yang & Sun, 2016). For example, using a spatial priming paradigm, Boroditsky (2001) showed that English native speakers were faster to respond to temporal relationships following a horizontal prime (left/right), whereas Chinese native speakers were faster after experiencing a vertical prime. While this result was consistent with observations from follow-up experiments (Miles et al., 2011; Yang & Sun, 2016), others have failed to replicate these effects (Chen, 2007; January & Kako, 2007; Tse & Altarriba, 2008), their authors arguing instead that Chinese-English bilinguals predominantly think of time horizontally and that the use of the vertical axis for the representation of time is semantically redundant.
Here, we examined whether using the vertical axis as temporal reference is indeed semantically redundant or whether it is a core feature of time conceptualisation in Chinese, that is, a preferred spatial representation against horizontal reference axes. Importantly, we moved away from measuring behavioural responses in participants asked to verbalise mental representations of time and space, and instead used event-related brain potentials (ERPs) to measure congruency between minimal temporal and spatial information delivered in a highly controlled fashion. This is an important new step since ERP measures index wholly implicit and pre-conscious information processing, in the absence of participants' explicit appraisal or metacognitive awareness of the process at work. More specifically, we presented participants with stimuli involving both spatial and temporal information in the form of written words flanked by arrows (see Fig. 35).



**Figure 35.** Sample of stimulus stream in the vertical (a) and the horizontal (b) experiments. Participants reported the direction the flanking arrows as quickly as possible when words in Chinese did not depict food items.

To quantify the congruency between verbal and spatial information conveyed by the word-arrow composite stimuli, we quantified the mean amplitude of the N400 component, a well-known index of semantic integration (Kutas & Federmeier, 2011; Kutas & Hillyard, 1984a, b; Kutas et al., 1984). Chinese-English bilinguals indicated the direction of the arrows flanking a verbal expression in three experimental conditions (see **Table 13**): *spatial words* (e.g., *L*, *shang* – 'up' and  $\overline{r}$ , <u>xia</u> – 'down'), *spatiotemporal metaphors* (i.e.,  $L \uparrow \overline{A}$ , *shang* – *ge-yue* – 'last month', literally 'up month', and  $\overline{r} \uparrow \overline{A}$ , *xia-ge-yue* – 'next month', literally

'down month'), or *temporal expressions* (i.e., 去年, *qu-nian* – 'last year', literally 'gone year' and 明年, *ming-nian* – 'next year', literally 'bright year').

Spatial words		Spatiotemporal metaphors		Temporal words			
Chinese	English	Chinese	English	Literally	Chinese	English	Literally
上 下	up down	上个月	last month	[up month]	去年	last year	[gone year]
左	Left	下个月	next month	[down month]	明年	Next year	[bright year]
右	Right						

 Table 13. Spatial temporal metaphors in Chinese.

Critical Chinese characters and English literal translations noted in bold.

We hypothesised that Chinese individuals conceptualise time vertically and thus that we would observe verbal-spatial interference between temporal expression and arrow direction in the vertical experiment. Assuming that time is conceptualized along the vertical axis in Chinese, we predicted larger N400 modulation would emerge in response to incongruent word-arrow configurations along the vertical axis. Whilst we expected to find significant congruency effects for last month and next month, since Chinese translations contain the characters for 'up' and 'down', any such modulation found for last year and next year would demonstrate conceptual access to a vertically oriented representation of time, given that Chinese translations for these expressions make no reference to spatial orientation. As for the horizontal experiment, we expected broadly similar results if Chinese participants would resort to mapping temporal information consistently along the left-right axis.

## 3.2.2 Materials and Methods

#### 3.2.2.1 Participants

Twenty-three Chinese English bilingual participants (5 male and 18 female; mean age = 23, range from 20 to 28; 22 right-handed and 1 left-handed) took part in the experiment. Data from five Chinese-English bilinguals were removed, due to poor electrophysiological recording quality, e.g., excessively high impedances, excessive blinking, and /or alpha

contamination. All participants reported having achieved scores of 6 or above in their International English Language Test System (IELTS), were resident in the UK at the time of testing. Participants received £15 or 5 course credits. The experiment was approved by the ethics committee at the School of Psychology, Bangor University.

### 3.2.2.2 Design & Stimuli

Arrow direction was blocked according to horizontal (left/right) and vertical (up/down) direction. In a horizontal block, the lexical stimuli were spatial words ( $\pounds$  – 'left',  $\hbar$  – 'right'), spatiotemporal metaphors ( $\pounds$   $\uparrow$   $\beta$ , *shang-ge-yue*, 'up month', last month;  $\top$   $\uparrow$   $\beta$ , *xia-ge-yue*, 'down month', next month), and temporal words ( $\pm$   $\mp$ , *qu-nian*, 'gone year', last year;  $\Re$   $\mp$ , *ming-nian*, 'bright-year', next year; see **Table 13**). In a vertical block, the two spatial words were  $\pounds$  – 'up' and  $\top$  – 'down'. Spatiotemporal metaphors and the temporal words were same as those in the horizontal block.

#### 3.2.2.3 Procedure

Participants filled in a language background questionnaire, and a reading/writing habits questionnaire whilst the 64-channel Ag/AgCl electrophysiological cap was being fitted. They sat in front of a 19-inch CRT monitor, placed at a distance of 100 cm in a sound-proofed room. Stimuli were presented 35 times in the horizontal condition (4 blocks) and 35 times in the vertical condition (4 blocks). The order of the blocks was pseudo-randomised between participants and counterbalanced. Each trial started with a fixation screen, worked as a pseudorandom inter-stimulus interval of 200-300 ms, followed by a verbal stimulus (one or two words) flanked by two arrows to the left and right of the written stimulus or above and below it for a maximum duration of 800 ms. Participants were asked to indicate the direction of the arrows when the word(s) presented at fixation did not refer to food or drinks. Filler trials featured 36 food or drinks words presented 8 times throughout the experiment, flanked by arrows pointing in the 4 directions. In the horizontal blocks, participants responded using the left or right index finger. In the vertical blocks, keys to indicate arrows point up or down were set under the left and right finger of each participant, but side was counterbalanced between participants: half of the participants were required to use their right hand to indicate

arrows pointing up and half of the participants were required to use their right hand to indicate arrows pointing down (see Fig. 35).

### 3.2.2.4 EEG data recording and processing

Electrophysiological data were recorded at a rate of 1 kHz from 64 Ag/AgCl electrodes according to the extended 10-20 convention. It was referenced to electrode Cz and the impedances were kept below 5 k $\Omega$ . The electroencephalogram (EEG) was filtered online using a bandpass filter (0.05 Hz–200 Hz), and offline using a low-pass, zero phase-shift digital filter (0.1Hz, 24 dB/oct–20 Hz, 28 dB/oct). Eye-blink artefacts were first manually removed through visual inspection of the data. The remaining artefacts were then mathematically corrected based on the procedure advocated by Gratton, Coles, and Donchin (1983). Epochs ranging from -200 to 1000 ms after onset were extracted from continuous EEG recordings. Epochs with activity exceeding  $\pm$  100  $\mu$ V at any electrode site, except the electrooculogram channels, were discarded. Baseline correction was performed in reference to pre-stimulus activity and individual averages were digitally re-referenced to the global average reference. N400 mean amplitudes were extracted in the usual time window, 300–500 ms after onset of stimuli at electrodes of predicted maximal sensitivity (i.e., FC1, FCz, FC2, C1, Cz, C2; Kutas and Hillyard, 1980a, b, 1984; Kutas et al., 1984).

#### 3.2.2.5 Behavioural data analysis

Reaction times (RTs) below 200 ms were removed from analysis. Then, RTs for correct responses were inverse transformed based on the Box-cox procedure to correct a deviation from normality. Accuracy data and RTs of correct answers were then analysed with logit and linear mixed-effect models, respectively, as implemented in the afex package (Singmann, Bolker, Westfall, & Aust, 2017). All fixed effects were contrast-coded before analyses using sum coding so that each model's intercept represented the mean value of each predictor.

## 3.2.3 Results - Vertical experiment

*Behavioural Results.* We found a main effect of congruency ( $\chi^2_1 = 17.34$ , P < 0.001) and a main effect of condition ( $\chi^2_1 = 6.62$ , P = 0.04) on RTs. Critically, there was a significant

interaction between congruency and condition ( $\chi^2_1 = 32.16$ , P < 0.001; see **Fig. 36**). *Post hoc* comparisons show that participants were faster responding in the congruent condition (M = 484, SE = 21) than the incongruent condition (M = 555, SE = 34;  $\chi^2_1 = 16.77$ , P < 0.001) in the case of spatial words. The difference between congruent (M = 490, SE = 21) and incongruent (M = 537, SE = 30;  $\chi^2_1 = 17.33$ , P < 0.001) conditions in the spatiotemporal metaphor condition also reached a statistically significant level. However, the effect of congruency for temporal words did not reach significance (Congruent: M = 504, SE = 24; Incongruent: M = 504; SE = 25;  $\chi^2_1 = 0.008$ , P = 0.93).



**Figure 36**. Reaction time (left) and mean accuracy (right) of vertical words, spatiotemporal metaphors and temporal words. Error bars represent s.e.m.

As for accuracy, we found a main effect of congruency ( $\chi^2_1 = 20.45$ , P < 0.001) and a main effect of condition ( $\chi^2_1 = 6.33$ , P = 0.04). A two-way interaction between congruency and condition also reached significance ( $\chi^2_1 = 28.63$ , P < 0.001). *Post hoc* comparisons show that participants obtained higher accuracy in congruent (M = 99%, SE = 0.02) than incongruent trials (M = 88%, SE = 0.08;  $\chi^2_1 = 25.34$ , P < 0.001) in the spatial word condition. In addition, participants also achieved higher accuracy in congruent (M = 99%, SE = 0.02) than incongruent trials in the spatiotemporal metaphor condition (M = 91%, SE = 0.07;  $\chi^2_1 = 24.47$ , P < 0.001). However, such congruency effect was absent in the case of temporal words (Congruent: M = 99%, SE = 0.03; Incongruent: M = 98%; SE = 0.03;  $\chi^2_1 = 0.008$ , P = 0.93).

*ERP Results.* The same repeated measures ANOVA with vertical stimuli revealed a significant main effect of congruency, F(1, 34) = 68.61, P < 0.001,  $\eta_p^2 = 0.8$ . Pairwise

comparisons suggested that incongruent conditions along the vertical axis generated more negative N400 amplitudes than congruent ones (see **Fig. 37**).



**Figure 37.** ERPs depict the signal from a linear deviation of six electrodes in Chinese-English bilinguals. The highlighted time window shaded in light grey is 300–500 ms.

## 3.2.4 Results - Horizontal experiment

*Behavioural Results.* No significant main effect of condition or congruency on either accuracy or RT and no interaction were detected (Ps > 0.1; see Fig. 38).



**Figure 38**. Reaction time and mean accuracy of horizontal words, spatiotemporal metaphors and temporal words. Error bars represent s.e.m.

*ERP Results.* We ran a repeated measure ANOVA with condition (spatial, spatiotemporal, temporal) and congruency (congruent, incongruent) on N400 amplitudes. No significant main effect of condition, F(2, 34) = 2.19, P = 0.13,  $\eta_p^2 = 0.11$ , or congruency, F(1, 17) = 2.23, P =

0.15,  $\eta_p^2 = 0.12$ , was found and there was no interaction, F(2, 34) = 0.91, P = 0.41,  $\eta_p^2 = 0.05$ (see Fig. 39).



**Figure 39.** ERPs depict the linear deviation of six electrodes for Chinese English bilinguals in Chinese. Time Window is 300-500 ms.

## 3.2.5 Discussion

This study investigated whether native speakers of Chinese tend to conceptualize time along the vertical axis, and how such tendency compares with conceptualisation along the horizontal (transversal) axis. Beyond the anticipated congruency effects for up and down ( $\pm$ and  $\overline{r}$ ), we found N400 modulations for spatiotemporal metaphors, but also for purely temporal expressions that do not imply any directional information in space. As for the second experiment, along the horizontal axis, we found no modulation.

Firstly, our results were consistent with studies suggesting that Chinese-English bilinguals represent time along the vertical axis (e.g., Boroditsky, 2001; Boroditsky et al., 2011; Lai & Boroditsky, 2013) when spatiotemporal metaphors are directly mentioned (last / next month). Furthermore, participants displayed congruency effects similar to that observed for spatial words and spatiotemporal metaphors when processing temporal metaphors that do not specify a spatial orientation (e.g., last / next year). The only explanation for this is that Chinese native speakers conceptualize time along on the vertical axis, that is, they generalise the axis used for space and spatiotemporal metaphors of Chinese to metaphors that are not spatial.

In several published studies, authors have argued that native speakers of Mandarin Chinese may organise time along the vertical axis because of the writing direction of Chinese (e.g., Bergen & Lau, 2012; Orly Fuhrman & Boroditsky, 2010). However, this writing direction was abandoned by modern China in 1956. And indeed, data from questionnaires showed that our participants predominantly wrote and read Chinese horizontally. Therefore, our results do not shed light on the issue of writing direction – temporal orientation relationships (e.g., Miles et al. 2011; Yang & Sun, 2016).

It is interesting to consider why we found no congruency effect along the horizontal axis. First, it must be noted that a direct comparison between horizontal and vertical blocks was not possible in this study given that the experimental conditions varied in three different ways across vertical and horizontal experiments: (a) the spatial words  $\underline{/}$  – 'up' and  $\overline{/}$  – 'down' were used in the vertical experiment whereas  $\pounds$  – 'left' and  $\hbar$  – 'right' were used in the horizontal experiment; (b) flanking arrows fell within the nasal field of view in the vertical experiment, since they were presented along the midline of the screen, immediately above and below the word(s), whereas they were presented in the peripheral field of view in the horizontal experiment, to the left and right of each word stimulus; (c) In the vertical experiment, participants needed to map response side with arrow direction (e.g., left hand with arrow pointing up) whereas such mapping was arguably straightforward and simple in the case of arrows point left (left hand) or right (right hand). These three differences between vertical and horizontal axis conditions made it impossible to directly compare the effects observed in one condition with those observed in the other, hence our presentation of the data in two experiments. The drawback of this approach is that we cannot assess any interaction between the two experimental contexts and therefore we are not in a position to state that, everything being equal, Chinese participants conceptualise time along the vertical and not along the horizontal axis, since is possible that in different circumstances, interference could be demonstrated along the horizontal axis also.

In summary, the findings of the current study are consistent with the general assumption of Linguistic Relativity (Whorf, 1956), suggesting that language structure (in this case spatiotemporal metaphors of Chinese) can shape individuals' conceptualisations (in this case the representation of time along the vertical axis). The experiment revealed that Chinese-English bilinguals can conceptualize time vertically, even when terms contains no

reference to direction in space. However, the evidence afforded by this study regarding conceptualisation of time along the horizontal axis is insufficient. This being said, it is also unknown what the results would look like if the experiment did not feature any spatial term at all, or any spatiotemporal metaphor, either. Indeed, one could imagine that the effects found for non-directional temporal metaphors such as  $\pm \mp$  – 'last year' and  $\# \mp$  – 'next year' depend on presentation of spatially oriented terms within the same block.

Future studies will need to address some of the key limitation of the present experiments, namely, the full comparability of stimuli and response mappings across axis orientations as well as test for contextual effects of spatial expression on temporal expression processing. One suggestion would be to implement the same paradigm without any mention of metaphors or spatial terms but rather based on content words that imply temporal orientation without calling upon the metaphors themselves as in Experiment 3 reported in this thesis (see section 3.1 above, and general discussion, section 4.3).

 $\infty$ 

# Chapter 4

**General Discussion** 

# 4.1 Grammatical variations and time The case of English tense

The ERP studies reported in Chapter 2 sought to expand the study of linguistic relativity solely from a grammatical perspective (e.g., Athanasopoulos et al., 2015; Flecken et al., 2015) to investigate the case of grammatical tense. In Study 1, we examined how highly fluent Chinese-English bilinguals process temporal information as conveyed by tense, a grammatical device that does not exist in their native language. Both Chinese-English bilinguals and English native controls were required to make acceptability judgements upon exposure to English sentences featuring a reference time misalignment (RTM) between the adjunct and the main clause:

(a). \*After the director of the school has resigned from the university, he worked for a multinational.<sup>10</sup> (Present Past Misalignment)

(b). \* After the director of the school will have resigned from the university, he worked for a multinational. (Future past misalignment)

In this first Study, only the English controls showed significant N400 modulations in response to the two types of RTMs. We interpreted these results as suggesting that tense information is not processed in real time by Chinese-English bilinguals, even when they have a relatively high level of proficiency in English.

However, we identified two limitations of Study 1, namely the bilingual group sampling criteria and the experiment task, that preclude a straightforward interpretation of the results as demonstrating difficulty at a semantic level of processing. First, Chinese-English bilinguals who took part in Study 1 could have been screened for grammatical proficiency in English with greater accuracy. The recruitment of bilingual participants relied on the IELTS

<sup>&</sup>lt;sup>10</sup> Even though both these examples feature time misalignments, example (a) is more subtle and only incorrect insofar as the present perfect refers to an event in the recent past having consequence for the present time.

scores provided by the participants. Although the participants were required to have achieved high scores in the IELTS to participate, the time at which they sat the test varied between them. Therefore, the scores recorded may not have reflect their current level of English. In addition, IELTS features four components: reading, writing, speaking and listening, which only examine grammatical skills indirectly. So, there is a possibility that some participants might have attained a very high IELTS score, whilst having a limited knowledge of English tense. Although self-report ratings have proven to be accurate to estimate participants' proficiency level in general, Chinese individuals tend to be culturally humble and are thus likely to underestimate their reported proficiency level.

In addition, the existence of a semantic violation condition positioned the task in Study 1 within the category of simple aberration detection. Compared with RTMs, semantic violations were trivial and thus particularly noticeable, leading Chinese participants to focus on that condition and assume that their task was simply to detect semantic violations. In contrast, English participants likely realized that RTMs were also difficult to integrate, given their sensitivity to temporal information conveyed by tense. Therefore, even though they failed to explicitly detect RTMs, they generated more negative N400 amplitudes in response to misalignments, given that N400 modulation is wholly automatic and does not require conscious evaluation (Kutas & Hillyard, 1984a, b; Van Petten & Kutas 1990).

In Study 2, we made sure that the two limitations of Study 1 highlighted above, namely the absence of direct grammatical proficiency testing and the focus on semantic violations, were addressed by (a) measuring proficiency using the Oxford Placement Test during the experimental session and (b) shifting the attention of participants towards the timeline of events. Furthermore, in Study 1, the most difficult condition to process for Chinese-English bilinguals was the Present–Past misalignment condition. In Study 2, we aimed to go beyond simple testing of sensitivity to temporal information conveyed by tense and make predictions regarding N400 modulations based on the correspondence between English expressions and their translation equivalents in Mandarin Chinese. More specifically, this enabled us to test whether fluent Chinese-English bilinguals have a different conceptualization of the recent past. That is, we wanted to know how Chinese-English bilinguals conceptualize the period of time referred to by the present perfect: Is it part of the past time sphere or the present time

sphere? To answer this question, we replaced the future–past misalignment condition with a present simple clash condition in Study 2 (see **Table 14**).

Study	Conditions Adjunct clause describing eve		Main clause describing event 2	
Study 1	Past perfect control <sup>11</sup>	After the director of the school <b>had resigned</b> from the university,		
	Present past misalignment	* After the director of the school <b>has resigned</b> from the university,	he <b>worked</b> for a multinational.	
	Future past misalignment	* After the director of the school <b>will</b> have resigned from the university,	×	
	Semantic violation	* After the director of the school <b>had resigned</b> from the university,	he <b>worked</b> for a <b>meter</b> .	
Study 2	Past perfect acceptable	After the director of the school <b>had resigned</b> from the university,		
	Past simple acceptable	After the director of the school <b>resigned</b> from the university,	he <b>worked</b> for a multinational.	
	Present perfect clash	* After the director of the school <b>has resigned</b> from the university,		
	Present simple clash	* After the director of the school <b>resigns</b> from the university,		

Table 14. Comparison of experimental conditions of Study 1 and the clash conditions of Study 2.

In addition to providing their IELTS scores prior to the experiment, participants in Study 2 completed the Oxford quick placement task, to assess their current proficiency level in English. It is noteworthy that the Oxford quick placement task particularly focusses on grammatical knowledge (e.g., Hawkins & Chan, 1997; Hawkins & Liszka, 2003). One participant was removed from the study, due to poor performance in that test even though his IELTS was over 6.5. In addition, participants completed an off-line task after the ERP recording session, in which they were asked to rate whether or not half of the stimuli from the ERP session were accurate in terms of timeline. In order to orient the participants' attention

<sup>&</sup>lt;sup>11</sup> The conditions in grey are same conditions across the two experiments.

onto temporal information, they were instructed to rate the acceptability of the succession of events depicted in the two clauses.

Study 2 showed that bilingual participants were able to achieve high accuracy in terms of explicitly detecting time clashes and time gaps, on a par with their English native peers. Nevertheless, they failed to generate more negative N400 amplitudes in the present perfect clash condition (i.e., the present–past misalignment condition of Study 1), relative to the past perfect acceptable condition or the past simple acceptable condition, whereas the same contrasts yielded significant N400 modulations in the English control group. In addition, whilst this was expected in the English native controls, the present perfect clash condition in the bilingual group either (see **Table 15**).

			Chinese English bilingual
Experiment	Contrast	English group	group
	Past perfect control		
Experiment 1	vs Present past misalignment	yes	No
	Past perfect control		
	vs Future past misalignment	yes	No
	Present perfect clash		
	vs Past perfect acceptable	yes	No
	Present perfect clash		
	vs Past simple acceptable	yes	No
	Present perfect clash		
Experiment 2	vs Present simple clash	No	No
	Present perfect acceptable		
	vs Present simple acceptable	No	yes
	Present perfect acceptable		
	vs Past perfect gap	yes	No
	Present perfect acceptable		
	vs Past simple gap	yes	No

#### Table 15. Summary of the N400 modulations in experiment 1 and 2.

One explanation for the lack of difference between present perfect clash and present simple clash conditions is that the critical words were tensed verbs. And thus, even though they could in theory distinguish past simple from present simple on the basis of lexical distinctions in Chinese, the lack of sensitivity to this clash could originate in their inability to process temporal information conveyed by the main clause verb, since it was tensed. Another possibility is that bilingual participants processed the present simple clash based on the equivalent expression in Chinese. In Chinese, the use of the perfective marker is optional in an adjunct clause, since any ambiguity will be resolved in the main clause which is always marked, as in the example below:

张三 去 英国 以后,他参观了 白金汉 宫。 Zhangsan go UK after, he visit-*le* Buckingham Palace. After Zhangsan went to the UK, he visited Buckingham Palace.

Indeed, "go" in the adjunct clause above is not marked, since the main clause verb "visit" is in the perfective form, marked by '*-le*' in Chinese. When bilingual participants process the present simple clash condition in English, they thus might rely on the verb of the main clause to determine the temporal sequence of the event in the adjunct clause. Therefore, they are likely to miss the temporal sequence error in the present simple clash condition given that the same sentence in Chinese features no error.

Furthermore, bilingual participants in Study 2 generated more negative N400 amplitudes in the present perfect correct condition involving a future form in the main clause (that is, the present perfect gap condition) as compared to the present simple correct condition (see **Table 15** and **Table 16**), whereas their English native peers showed no statistical differences for the same contrast. Conversely, differences between the present perfect correct and either the past perfect or past simple gap conditions failed to reach a statistically significant level in the bilingual group whereas the same comparison yielded significant differences in the English controls. These results suggest that Chinese-English bilinguals conceptualize the time period referred to by the present perfect as part of the past time sphere, whereas English natives conceptualize it as part of present time sphere.

Overall, Study 2 showed that grammatical features of the L2 that do not have an equivalent in the L1 are not fully processed conceptually. Furthermore, it suggests that the properties of L1 translation equivalents are heavily relied upon by bilinguals, showing that effects of language non-selective access observed previously at a lexical level (Thierry and Wu, 2007; Wu and Thierry, 2010, 2012) also obtain at a grammatical level (Sanoudaki & Thierry, 2015; Vaughan-Evans, Kuipers, Thierry, & Jones, 2014), even when syntactic features do not actually exist in the L1. This means that bilingual individuals may engage on

an approximation strategy (i.e., calling upon 'best possible translations') when native forms corresponding to second language constructs do not exist, a process akin to the 'good enough' translation concept (e.g., Ferreira & Karimi, 2015; Ferreira & Patson, 2007).

Study	Conditions Adjunct clause describing event 1		Main clause describing event 2	
Study 2	Past perfect gap	* After the director of the school had resigned from the university,		
	Past simple gap	* After the director of the school <b>resigned</b> from the university,	he <b>will work</b> for a multinational.	
	Present perfect acceptable	After the director of the school <b>have</b> resigned from the university,		
	Present simple acceptable	After the director of the school <b>resigns</b> from the university,		

**Table 16.** Time gap conditions and stimulus examples in Study 2

Chinese marks temporal aspect using perfective markers (Chao, 1968; Li & Thompson, 1989). Therefore, when Chinese-English bilinguals read a sentence in the present perfect tense in English, they detect the fact that the event is completed, and thus conceptualise the event as belonging to the past time sphere, just as they would conceptualise it in Chinese. In other words, they rely on the Chinese timeline to process the temporal information of English. However, the results of Study 1 and Study 2 do not mean that Chinese-English bilinguals are unable to acquire English tense, since they could make correct, explicit decisions regarding its use. It is thus implicit, online semantic integration aspects of tense meaning that are not developed hand-in-hand with their explicit knowledge, effectively blurring their representation of tense as they read English.

In the early 1940s, Benjamin Lee Whorf was the first to make observations linking specific structures of language and particular ways of thinking. In the proposal, Whorf focused on both the grammatical and lexical differences between Hopi and "Standard Average European" languages when talking about temporal information. In the last fifty years, a substantial number of behavioural studies have been conducted to study the effects of linguistic relativity at both grammatical and lexical levels. Most of these studies have failed to establish strong Whorfian effects, either because these effects are too subtle (and implicit) to be detected based on overt participants' response, or because such effects are liable to

interference from conscious monitoring strategies and metacognitive evaluation, leaving deep processing untouched (Pinker, 2007). To the best of my knowledge, no previous research before the two studies discussed above has shown such marked differences between speakers of languages with different grammars using direct indices of brain activity. This is perhaps the clearest demonstration to date that an intuitive proposition put forward by Whorf in the 1940s was actually correct and possibly had an even stronger basis than he himself was prepared to admit.

# 4.2 Lexical variations and Time

# The case of spatiotemporal metaphors

The ERP studies reported in Chapter 3 sought to test the linguistic relativity hypothesis in the domain of time in relation to expressions idiosyncratic to one language in bilinguals. In Study 3, we evaluated how spatiotemporal metaphors of Chinese shape time conceptualization along the sagittal (front-back) axis. In a majority of languages, such as English, Chinese, or French, the future is conceptualised as orientated towards the front, and the past towards the back of the speaker.

Capitalizing on the fact that Mandarin Chinese features lexical exceptions that conflict with the future-in-front orientation (i.e., *hou-tian*, 'back-day', the day after tomorrow; *qian-tian*, 'front-day', the day before yesterday; *hou-nian*, 'back-year', the year after; *qian-nian*, 'front-year', the year before tomorrow), we tested whether stimuli best described by one of these spatiotemporal metaphors would elicit processing interference when presented from a location incongruent with their literal meaning. We found that stimulus configurations incongruent with the orientation implied by spatiotemporal metaphors in Chinese (e.g., "Friday" presented in the front of the participant during a session held on a Wednesday, since this refers to the day after tomorrow, literally translating as 'back-day' in Chinese) proved conceptually more challenging to process than congruent configurations (e.g., the same stimulus presented in their back). Importantly, as was the case in experiments 1 and 2, processing difficulty was indexed by N400 modulations, which are wholly automatic, unconscious, and driven by implicit mechanisms. Strikingly, the interference effect

originating in spatiotemporal metaphors only arose when bilingual participants processed the stimuli in their L2 English and not in their L1 Chinese, from which the spatiotemporal metaphors were sourced.

According to mainstream models of bilingual processing such as DRM, BIA+ or RHM, semantic representations are shared between languages, suggesting that conceptual information accessed from L1- or L2-specific input should be essentially the same. In the case of the Sense model, lexicosemantic representations are not entirely shared between L1 and L2, and the semantic field of L1 is supposedly more extended that that of L2 in imbalanced bilinguals. The effects observed in Study 3 cannot be readily be accounted for by any of the models above, because access to spatiotemporal metaphors of L1 was only observed when participants were tested in L2. Indeed, whilst it is relatively easy to account for the fact that L1 representations are accessed when participants are tested in the L2 (see below), it is highly surprising that the same representations are not accessed when participants are tested in their L1. In other words, L1 words should be associated with more semantic features than L2 words, not fewer.

One possible explanation discussed in Chapter 3.1 is that the task used in that study was too easy when participants performed it in their L1 Chinese. Aside from the case of "Sunday", participants only needed to identify the digit of the day or the final digit of the year number and determine whether the difference between that digit and the reference digit (present day or present year) was 1 or 2. For example, if the day of testing was Wednesday ('weekday-3' in Chinese) and the stimulus was Friday (i.e., 'weekday-5' in Chinese) the answer was 2. In contrast, when bilinguals performed the task in English, days of the week did not provide immediate numerical information to resolve the task. Even in the case of years in English, interval calculation required (1) number translation and (2) calculation from two digit numbers as compared to single digits in Chinese. Furthermore, it is a well-established fact that bilinguals systematically resort to first language operations when they do mental operations, irrespective of the language of input (e.g., Clarkson, 2007; Clarkson & Galbraith, 1992). Beyond providing a ground for the asymmetry observed, this imbalance in difficulty across languages provides a hint as to the source of the differential effect in L1 and L2.

122

The results of Study 3 are consistent with findings from previous ERP studies of lexical access in bilinguals (e.g., Thierry and Wu, 2007; Wu and Thierry, 2010, 2012) showing that bilingual participants automatically access first language lexical representations, when they are presented with second language words, whether auditorily or visually. Despite recent controversy around this theoretical standpoint involving modelling of alternative accounts (e.g., Costa et al., 2017; Oppenheim et al., 2018), the mainstream interpretation in the literature remains in favour of online language non-selective lexical activation, particularly in the case of Chinese-English bilinguals.

Thus, one can anticipate that when bilinguals heard the stimulus "Friday", they automatically activated the Chinese translation equivalent "xing-qi wu" and also "hou tian (back-day)", when the session was held on a Wednesday since Friday is the day after tomorrow in relation to Wednesday. This interpretation of the results is overall compatible with the view that a bilingual speaker's lexicon is shared between L1 and L2 (see BIA, BIA+ and DRM model descriptions in Chapter 1.2). This being said a non-selective lexical access account does not provide a simple explanation for the lack of effect when the Study was conducted in Chinese, which possibly involves lexical competition (see discussion of Chapter 3.1). In addition, the current findings are consistent with the claim made by the DRM; i.e., that not only translation equivalents, but also semantically related words are activated in bilingual word processing. However, according to DRM, L1 word stimuli should activate more senses than L2 word stimuli. And thus, congruency effects predicted by native language spatiotemporal metaphors should be more likely when bilingual participants operate in their first than their second language. Surprisingly, we observed a congruency effect only when bilinguals operated in their second language, which is difficult to square with the premises of the DRM.

In Study 4, we decided to test how Chinese-English bilinguals conceptualize time along the vertical (top-down) and transverse (left-right) axes, but this time we only tested participants in their native language, Chinese, and we exposed them directly to the critical spatiotemporal metaphors. Indeed, in Study 3 participants were never presented with the spatiotemporal metaphors at the origin of the congruency effects measured in the N400 range, and thus it was not possible to directly evaluate how these spatiotemporal metaphors affect time conceptualisation (See next section for future experimental developments). Furthermore, the main aim of this last study was to determine whether vertical thinking about

123

time not only can be triggered by spatiotemporal metaphor use but also generalise to the case of metaphors without a spatial orientation.

In an adapted version of the arrow flanker task, in which the middle arrow was replaced by a Chinese word, participants were asked to report the flanking arrows' direction, only when the middle word did not refer to a food or drinks item<sup>12</sup>. Behaviourally, both accuracy and reaction times showed effect of congruency between word meaning and arrow direction only in the vertical conditions and only for expressions that contained a spatial reference. We also observed significant N400 modulations along the vertical axis for space words (up-down); for spatiotemporal words, namely,  $\pm \uparrow \exists$ -'last month'(literally 'up month'),  $\not{\top} \uparrow \exists$ -'next month', (literally 'down month'); and, critically, for temporal metaphors without spatial orientation, namely,  $\pm \not{\mp} =$ -'last year' (literally 'gone year'),  $\varPi \not{\mp} =$ - 'next year' (literally 'bright year'). Strikingly, we found no modulation of either behavioural measures or ERP amplitudes for the horizontal axis manipulations.

One possible explanation for the absence of a congruency effect along the horizontal axis is that the task was likely less demanding than along the vertical axis. Indeed, in the horizontal blocks, participants simply pressed the left key with their left hand to indicate arrows pointing left and the right key with their right hand to report arrows pointing right. In contrast, in the vertical blocks, participants needed to remap the direction of the arrows in relation to upper and lower keys on the response pad, given that up and down directions still had to be reported with left and right hands<sup>13</sup>. It is reasonable to assume that the full alignment between arrow direction, side of response key, and response hand in the horizontal experiment enabled participants to better control interference from the word meaning, be it spatial or temporal. When these factors were not aligned, the required remapping process likely increased processing difficulty along the vertical axis, leading to measurable interference effects.

<sup>&</sup>lt;sup>12</sup> This was required to ensure that participants processed the meaning of the word in every trial rather than focussing exclusively on the direction of the arrows

<sup>&</sup>lt;sup>13</sup> In addition, recall that the mapping was rotated between participants so as to avoid lateralisation effects due to hand dominance.

It is interesting to consider that authors of previous experiments have argued that Chinese native speakers predominantly conceptualize time on the horizontal axis when we found not sign of interference along this axis for temporal expressions (e.g. Miles et al., 2011; Yang and Sun, 2016). One could argue that the lack of an effect on the horizontal axis is not surprising for 'last month' and 'next month', given that the corresponding spatiotemporal metaphor of Chinese refers to the vertical axis directly, but it is surprising as regards 'last year' and 'next year' since the corresponding metaphors of Chinese are not bound to any direction in space. Thus, even though they are not unequivocal and although a generalisation of the findings would require follow-up experiments (see next section for future developments), our results suggest that Chinese native speakers have a preference for vertical orientation when conceptualising the passing of time.

Boroditsky (2001) famously showed that Chinese-English bilinguals are faster to respond to a temporal question after a stimulus priming vertical orientation than a stimulus priming a horizontal orientation. However, this effect failed to replicate in later experiments (e.g., Chen, 2007; January & Kako, 2007). Importantly, all the studies conducted so far in this domain have been based on behavioural measures, and mostly reaction times. In Study 4 of the present thesis, the participants showed behavioural effects only when the target words were spatial words or spatiotemporal metaphors, but not when they were non-spatial metaphors about years. Thus, if we were to draw conclusions solely based on behavioural measures, we could consider that only spatiotemporal metaphors of Chinese lead to conceptualising time along the vertical axis, an effect akin to basic linguistic relativity, i.e., when the label is available, conceptual representations are affected but when the label is not, no effect is measured. However, when ERP results are considered, the story changes noticeably. Given congruency effects on N400 modulation in the case of non-spatial temporal words ('gone / bright year'), a more accurate conclusion should be that Chinese-English bilinguals tend to conceptualize time in the vertical domain, even in the absence of spatially transparent temporal metaphors.

Overall, the results of Studies 3 and 4 are broadly congruent with the core assumptions of the Label-Feedback Hypothesis put forward by Lupyan (2012), namely, that language representations can modulate on-line perceptual and higher-level cognitive processing. When, in Study 3, a participant heard the English stimulus '*Friday*' played from a speaker situated in their front during a session held on a Wednesday, the corresponding native spatiotemporal metaphor of Chinese, *hou-tian* – 'back day' was accessed unconsciously and this interfered with the source of the physical stimulus in the space surrounding the participant, even though both the temporal reference (here the future) and the spatial origin of the stimulus (here the front space) were irrelevant regarding the simple interval counting task at hand. Presumably, a critical step in this stream of mental representation activations was the retrieval of the label for the day after tomorrow in Chinese, which is the only link available to spatial information. In Study 4, when the participants saw the word  $\pounds \uparrow \varPi$  - 'last month', together with downward flanking arrows, arrow direction and word meaning interfered because of the presence of the character  $\pounds$  meaning 'up'. Again, this can be construed as an LFH-compatible effect since the presence of the character for upward orientation conflicts with arrow direction, even though that character was part of a temporal metaphor meaning 'last month', rather than the word 'up'. Going well beyond the assumptions of the LFH, however, we also found signs of generalisation of spatial orientation for time, even when no reinforcing label is available, which questions the source of that effect. Future studies (see next section) will determine the basis for such generalisation.

# 4.3 Limitations and directions for future studies

Naturally, the experiments reported in this thesis have limitations relating to the choice and selection of participants, the particular linguistic devices studied, the methodological parameters selected, such as recording system, ERP indices of cognitive operations (i.e., the N400), modality of stimulus presentation, etc. In this section, I selectively consider some of the limitations that seem critical and propose experimental developments that may address these limitations in the future.

## 4.3.1 On tense

In both *Study 1 and 2*, Chinese-English bilingual participants appeared to behave in a similar way to English controls in terms of response accuracy. Whilst the lack of difference could

have been due to the focus on semantic violations in Study 1 (or even a lack of proficiency, but see discussion of Chapter 2.1), the lack of differences in Study 2 suggest that participants were indeed able to resolve temporal information conveyed by tense at an explicit level. However, in either study and particularly in Study 2, bilingual participants failed to generate more negative N400 amplitudes in the present perfect time clash condition when compared to control conditions. This bring up the question of *when* bilingual participants will have detected an issue with tense information embedded in the sentences.

Unfortunately, the design of both experiments did not allow us to behaviourally measure temporal information resolution since reaction times were postponed until the end of the sentence, that is between 2-5 words after the critical word (verb of the main clause). An exploratory analysis of the P600 time window in Study 2 revealed that bilingual participants in fact generated larger P600 amplitudes in the present simple clash condition as compared to the two correct conditions, showing that despite the absence of online semantic detection of tense conflict between clauses, bilingual participants were able to explicitly register the temporal misalignments (see **Fig. 40**). However, they did not generate more positive P600 amplitudes in the *present perfect clash condition* when contrasted with either of the two correct conditions. This suggests that the past simple clash condition was easier to spot as compared to the present perfect clash condition, but it does not resolve the issue of when the present perfect clash was detected by bilingual participants.



**Figure 40.** Comparison of event-related potentials in the present perfect clash condition as compared to the past perfect acceptable, past simple acceptable, and present simple clash in English controls and Chinese-English bilingual participants. Waveforms depict the linear derivation of signals recorded at six electrode sites (CP1, CPz, CP2, P1, Pz, P2). The highlighted time window is 500-700 ms.

Future studies implementing a similar paradigm will need to tackle the ERP correlates of the decision making process driving behavioural measures. Detecting the late phase of processing during which bilingual participants resolved temporal misalignments will likely require modifying the paradigm and/or the procedure. For instance, response times could be recorded as soon as the critical information is delivered rather than postponed until the end of the sentence. This would have a dual advantage over the current experimental design: (a) RTs would become directly informative and (b) P600 amplitude would likely correlate with RT measures if the process of misalignment detection is indeed explicit in bilinguals. There would be a cost however, since response planning and execution are known to produce ERP modulations that will contaminate signals in the N400 window and thus make control native English participant data less interpretable. Alternatively, the Study could rerun with a response only required for a subset of the critical words (say when they are displayed in green font).

Another future development concerns differences between languages that are closer typologically than Chinese and English, like English and French or English and Spanish. Indeed, some grammatical features of French or Spanish can be invisible in English, such as the conditional. Consider the following example for instance: "If we go to the cinema, I will buy you an ice-cream"; "go" is in the conditional form even though it appears exactly like the present simple form. In French this sentence translates into "Si on allait au cinema, je t'acheterais une glace"; "allait" is in the conditional form (the present simple form being "va"), and thus the verb form directly codes the fact that the action described in the main clause will take place only if the first event is realised. Would English-French bilinguals ignore the conditional nature of the main clause event and consider it more certain than French native participants?

## 4.3.2 On metaphors

A limitation of *Study 3* is the absence of control measurement for metaphor processing. Participants in that Study did not encounter spatiotemporal metaphors (e.g., *hou-tian*, 'back-day', the day after tomorrow; *qian-tian*, 'front-day', the day before yesterday) directly in the Study. Even though this can be considered a strength of the design, since predicted interference effects were found without the need to present metaphors, it would be highly informative to see the interference effects causes by metaphors when directly presented in the experiment. In addition, in Study 3, auditory stimuli were presented via two loudspeakers located in front of the participant's chair and two speakers behind it. This may have blurred the effects of sound origin localisation, along with room configuration (size, height, etc.).

Future research is needed to assess how Chinese-English bilinguals would deal with spatiotemporal metaphors along the front-back axis directly. One possible approach would be to ask bilingual participants to press a given key when they hear the day before yesterday ('front day') and another designated key when they hear the day after tomorrow ('back day'). This would ensure direct exposure to the spatiotemporal metaphors in the Study and may prompt congruency effect in the native language, Chinese. Participants would not necessarily need to determine the spatial origin (front or back) or the relative time (future or past) of the stimuli. Another approach would be to present the spatiotemporal metaphors as visual primes and a date (day or year) as auditory target played from the front of the back, and ask participants to assess whether prime and target stimuli are congruent. Hopefully, these Study designs would demonstrate clear congruency effect in the native language since the spatiotemporal metaphors would be directly available for processing.

Another development in the spirit of Study 3 but testing temporal conceptualisation along the vertical axis would be to play auditory stimuli from speakers located above or below the chair of the participant. This experimental set up, combined with adequate spatiotemporal metaphors (last month/year and next month/year in Chinese) or dates calling upon such metaphors (month/year labels in the future or the past of the day of testing), would further shed light onto the mechanisms involved in resolving the timeline in Chinese-English bilinguals and disentangle the role of metaphors from that of physical spatial reference in relation to time in bilinguals.

One limitation of Study 4 concerns the spatial words used in the horizontal and vertical blocks, which were not the same, i.e., "left" and "right" in the horizontal experiment and "up" and "down" in the vertical experiment. Arrow positions were not compatible between experiments either. Arrows in the horizontal experiment will have involved the participants' two visual fields, while arrows in the vertical experiment were presented in the

nasal field of view, i.e., shared by the two hemispheres. Finally, as mentioned in the discussion above, participants needed to remap the direction of arrows in the vertical block, since up or down arrows needed reporting using left and right hands, whereas such remapping was not required for horizontal arrow direction reporting. Taken together, these differences between axes orientations made a direct comparison between horizontal and vertical experiment data impossible.

In future research that might follow on from the design developed for Study 4, it would be advisable to make data compatible across axes to allow direct comparison. One way to do this would be to simply add "up" and "down" in the horizontal experiment (thus becoming an experimental block) and "left" and "right" in the vertical block as spatial word stimuli. However, we already know from the results of the Study 4 that left and right as word stimuli failed to elicit interference effects measurable in behavioural data or the N400 amplitude data. Thus, it would be better to consider alternative ways of making the blocks comparable. One option would be to use arrows that point in all four directions, because incompatible arrow line-ups are known to elicit P300/N400 modulations. Another interesting development would be to present all arrows in the nasal visual field to avoid a potential contribution from hemispheric asymmetries in visual processing (see **Fig. 41**). Also, in the spirit of Study 3, metaphors could be implicitly suggested rather than overtly presented in the same paradigm (see **Fig. 41**, rightmost item) in order to test for unconscious access to spatiotemporal representations along the vertical axis (this would require setting a temporal reference time as in Study 3).



**Figure 41.** Example of stimuli options to control for spatial mapping, visual field of presentation and metaphor visibility.

Another obvious development of Study 4 would be to test Chinese-English bilinguals and English controls in English using the same paradigm. In fact, I collected these data but chose not to report them here due to an interpretational problem relating to the fact that stimuli in English are made up of two words, which can be processed separately and indeed without consideration for the temporal meaning of the expressions. For instance, "last month" and "next month" in English can prime different spatial locations without requiring processing of their temporal meaning, and this can happen in the same way for Chinese or English participants, leading to inconclusive results. Indeed, in the data collected, both the Chinese-English bilinguals and English native controls elicited a congruent effect along the horizontal axis. However, these effects could be caused by a preferential association between last and next with the left and right space, or the upper and lower space, respectively.

# **4.4 Final Remarks**

Although results from the four studies reported in this thesis suggest that native language can influence Chinese-English bilinguals' perception of time, these results were derived from experimental designs that involved language. In other words, in these studies, participants were invited to access language representations at any moment, whether consciously or unconsciously. Therefore, future research targeting temporal perception and conceptualisation in Chinese individuals would need to involve non-verbal tasks, in which access to verbal representations is either minimised or excluded. One could consider in the extreme, that conclusions in experiments 3 and 4 are in fact limited to the labels used and that such effects could not necessarily generalise to temporal processing in the absence of verbal cues. However, it is difficult to see how the absence of tense in Chinese could be studied as a phenomenon without resorting to using language stimuli...

Since bilingual participants in all four studies were Chinese-English bilinguals, English proficiency had to be assessed, especially in Studies 1 and 2. Whilst participants in all experiments had a high level of English proficiency, the majority of them learned English after puberty, and thus were dominant in Chinese. Beyond the fact that proficiency could be measured with better accuracy and that measures of exposure to English could be included, it would be very interesting to consider correlation analyses between proficiency measures and dependent variables in the experiments above, as well as testing participants with the reverse pattern of language dominance, e.g., Hong Kong-raised English-Chinese bilinguals.

Previous research in the field of linguistic relativity has shown that language context can modulate bilinguals' on-line perception and conceptualisation (e.g., Athanasopoulos et al., 2015). Of the four studies reported in the previous chapters, the only one in which bilinguals were tested in both their languages was Study 3 (note that in Study 4 different

bilingual participants were tested in the two languages given that the metaphor manipulation was overt). In that Study, ERP results showed that participants manifested the influence of native spatiotemporal metaphors only when operating in their second language. Further studies will need to manipulate language context systematically and across different experimental designs in order to assess to extent to which perception and conceptualisation of time is affect by language context.

One other pervasive finding in the linguistic relativity literature is that Whorfian effects tend to weaken or even disappear when participant are subjected to verbal interference. According to explanations provided by Athanasopoulos et al. (2015), verbal interference can reduce access to a specific language. Even though it would be very interesting to test how verbal interference affects time conceptualisation in bilingual participants tested using the same paradigm as in the current thesis, this would be highly problematic given the fundamental incompatibility between a classic verbal interference task involving language production and ERP data collection. Thus, until now, no ERP experiment has used verbal interference as a means to study linguistic relativity effects. However, language production can also be replaced with other tasks, such as a silent verbalising or verbal short-term memory tasks.

The studies presented in this thesis aimed at characterising how language constrains the way in which Chinese-English bilinguals conceptualize time. The majority of the previous research has investigated whether bilinguals can fully acquire a syntactic structure that does not exist in their mother tongue (e.g. Steinhauer, White & Dury, 2009; Hawkins & Liszka, 2003; Hawkins & Chan, 1997). However, to my knowledge, no study before those presented in this thesis have investigated *semantic processing* as opposed to grammatical processing of tense in Chinese-English bilinguals. The present studies showed that Chinese-English bilinguals fail to process tense at a semantic level, to the point that they seem to have a blurred representation of time when it is conveyed by tense in English. This is true at the very least during online reading and it may be that re-evaluation of tensed stimuli allows bilinguals to recover temporal information at a later point but this is for future experiments to demonstrate. Beyond the domain of syntax, when investigating relationships between word forms that exist in the L1 (Chinese) but not in the L2 (English), we showed effect of language representations on time conceptualisation consistent with the premises of linguistic relativity and embodied cognition. More specifically, Chinese-English bilinguals conceptualize time using space in accord with spatiotemporal metaphors idiosyncratic to their native language. This would not be so surprising if these effects had been merely explicit and conscious rather than implicit and unconscious: But (a) they were observed selectively in the second language along the front-back axis; (b) they were observed along the vertical preferentially to the horizontal axis when participants were tested in the native language. The latter studies are thus the first to demonstrate that irrelevant spatial information relative to the human body is accessed when processing temporal information (embodied cognition), that these effects are predicted by constructs that only exist in one language (linguistic relativity), and that they appear across languages, i.e., when participants are tested in their second language (language non-selective access). Taken together, the results presented here provide a new picture of the highly interconnected system underpinning bilingual conceptualisation of time, in which all linguistic and perceptual representations interact in real time, whether consciously or not.

I hope that future investigations along the lines that we started drawing here will eventually provide an extended picture of time conceptualisation in Chinese individuals and show how their native language shapes said conceptualisation. I also hope that the experimental designs presented in this thesis will inspire future research in the domain of language-temporal cognition interactions.

# References

- Ainsworth-Darnell, K., Shulman, H. G., & Boland, J. E. (1998). Dissociating brain responses to syntactic and semantic anomalies: Evidence from event-related potentials. *Journal of Memory and Language*, 38(1), 112-130. doi: 10.1006/jmla.1997.2537
- Allen, M., Badecker, W., & Osterhout, L. (2003). Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes*, *18*(4), 405-430. doi: 10.1080/01690960244000054
- Alverson, H. (1994). Semantics and experience: universal metaphors of time in English, Mandarin, Hindi, and Sesotho. Baltimore: Johns Hopkins University Press.
- Ameel, E., Malt, B. C., Storms, G., & Van Assche, F. (2009). Semantic convergence in the bilingual lexicon. *Journal of Memory and Language*, 60(2), 270-290. doi: 10.1016/j.jml.2008.10.001
- Astikainen, P., Ruusuvirta, T., Wikgren, J., & Korhonen, T. (2004). The human brain processes visual changes that are not cued by attended auditory stimulation. *Neuroscience Letters*, *368*(2), 231-234. doi: 10.1016/j.neulet.2004.07.025
- Athanasopoulos, P. (2006). Effects of the grammatical representation of number on cognition in bilinguals. *Bilingualism: Language and Cognition*, 9(1), 89-96. doi: 10.1017/S1366728905002397
- Athanasopoulos, P. (2009). Cognitive representation of colour in bilinguals: The case of Greek blues. *Bilingualism:Language and Cognition*, 12(1), 83-95. doi: 10.1017/S136672890800388X
- 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. doi: 10.1111/cogs.12006
- 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. doi: 10.1111/cogs.12006
- Athanasopoulos, P., Bylund, E., Montero-Melis, G., Damjanovic, L., Schartner, A., Kibbe, A., Thierry, G. (2015). Two Languages, Two Minds: Flexible Cognitive Processing Driven by Language of Operation. *Psychological Science*, 26(4), 518-526. doi: 10.1177/0956797614567509
- Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J.-R., & Thierry, G. (2010). Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition*, *116*(3), 437-443.doi: 10.1016/j.cognition.2010.05.016
- Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in Spanish: An ERP comparison. *Journal of Cognitive Neuroscience*, *17*(1), 137-153. doi: 10.1162/0898929052880101
- Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology*, *4*, 328-328. doi: 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. doi: 10.1016/j.jml.2012.11.001
- Barsalou, L. W., Niedenthal, P. M., Barbey, A. K., & Ruppert, J. A. (2003). Social embodiment. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory*, Vol. 43, pp. 43-92. New York, NY, US: Elsevier Science.
- Bates, D., Maechler, M., & Dai, B. (2008). lme4: Linear mixed effects models using s4 classes( R package version 0.999375-28).
- Bergen, B., & Chan Lau, T. T. (2012). Writing Direction Affects How People Map Space Onto Time. *Frontiers in Psychology*, *3*(109). doi: 10.3389/fpsyg.2012.00109
- Berkum, J. J. A. v., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the N400. *Journal of Cognitive Neuroscience*, *11*(6), 657-671. doi: 10.1162/089892999563724
- Blanche, P., & Merino, B. J. (1989). Self-assessment of foreign-language skills- implications for teachers and researchers. *Language Learning*, *39*(3), 313-340. doi: 10.1111/j.1467-1770.1989.tb00595.x
- Boas, F. (1911). *Introduction to Handbook of American Indian Languages*. Washington, DC:US Government Printing Office.
- Boroditsky, L. (2000). Metaphoric structuring: understanding time through spatial metaphors. *Cognition*, 75(1), 1-28. doi: 10.1016/S0010-0277(99)00073-6
- Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, 43(1), 1-22. doi:10.1006/cogp.2001.0748
- Boroditsky, L., Fuhrman, O., & McCormick, K. (2011). Do English and Mandarin speakers think about time differently? *Cognition*, *118*(1), 123-129. doi: 10.1016/j.cognition.2010.09.010
- Boroditsky, L., & Ramscar, M. (2002). The Roles of Body and Mind in Abstract Thought. *Psychological Science*, *13*(2), 185-189. doi:10.1111/1467-9280.00434
- Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61-79). Cambridge, MA: MIT.
- Boutonnet, B., Athanasopoulos, P., & Thierry, G. (2012). Unconscious effects of grammatical gender during object categorisation. *Brain Res*, 1479, 72-79. doi: 10.1016/j.brainres.2012.08.044
- Boutonnet, B., Dering, B., Viñas-Guasch, N., & Thierry, G. (2013). Seeing objects through the language glass. *Journal of Cognitive Neuroscience*, 25(10), 1702-1710. doi: 10.1162/jocn a 00415
- Brown, C. M., Hagoort, P., & Keurs, M. T. (1999). Electrophysiological signatures of visual lexical processing: Open-and closed-class words. *Journal of Cognitive Neuroscience*, 11(3), 261-281. doi: 10.1162/089892999563382
- Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *The Journal of Abnormal and Social Psychology*, 49(3), 454. doi: 10.1037/h0057814

- Brysbaert, M., & Duyck, W. (2010). Is it time to leave behind the Revised Hierarchical Model of bilingual language processing after fifteen years of service? *Bilingualism: Language and Cognition*, 13(3), 359-371. doi: 10.1017/S1366728909990344
- Bylund, E., & Athanasopoulos, P. (2017). The Whorfian Time Warp: Representing Duration Through the Language Hourglass. *Journal of Experimental Psychology: General*, *146*(7), 911-916. doi: 10.1037/xge0000314
- Bylund, E., & Athanasopoulos, P. (2017). The Whorfian time warp: Representing duration through the language hourglass. *Journal of Experimental Psychology: General, 146*(7), 911. doi: 10.1037/xge0000314
- Caffarra, S., Molinaro, N., Davidson, D., & Carreiras, M. (2015). Second language syntactic processing revealed through event-related potentials: an empirical review. *Neuroscience & Biobehavioral Reviews*, *51*, 31-47. doi:10.1016/j.neubiorev.2015.01.010
- Carroll, J. B. (1957). Language, thought, and reality: Selected writings of Benjamin Lee Whorf. Cambridge, MA: MIT.
- Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and lefthanders. *Journal of Experimental Psychology: General*, 138(3), 351-367. doi: 10.1037/a0015854
- Casasanto, D., & Boroditsky, L. (2008). Time in the mind: using space to think about time. *Cognition*, *106*(2), 579-593. doi:10.1016/j.cognition.2007.03.004
- Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S., . . . Gil, D. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*, 186-191.
- Casasanto, D., & Jasmin, K. (2012). The Hands of Time: Temporal Gestures in English Speakers. *Cognitive Linguistics*, 23 (4), 643-674. doi: 10.1515/cog-2012-0020
- Chao, Y. R. (1968). A Grammar of Spoken Chinese. Berkeley: University of California Press.
- Chen, J.-Y. (2007). Do Chinese and English speakers think about time differently? Failure of replicating Boroditsky (2001). *Cognition*, *104*(2), 427-436. doi:https://doi.org/10.1016/j.cognition.2006.09.012
- Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, 41(1-3), 83-121. doi:10.1016/0010-0277(91)90033-z
- Chomsky, N. (1986). Knowledge of language: Its nature, origin, and use. New York: Praeger.
- Chomsky, N. (1995). The Minimalist Program. Cambridge, MA: MIT Press.
- Clark, H. H. (1973). Space, Time, Semantics and the Child. In Moore, T. E (Eds.). *Cognitive Development and Acquisition of Language* (pp. 27-63). San Diego: Academic Press.
- Clarkson, P. C. (2007). Australian Vietnamese students learning mathematics: High ability bilinguals and their use of their languages. *Educational Studies in Mathematics*, 64(2), 191. doi: 10.1007/s10649-006-4696-5

- Clarkson, P. C., & Galbraith, P. (1992). Bilingualism and mathematics learning: Another perspective. *Journal for Research in Mathematics Education*, 23(1), 34-44.
- Coch, D., & Gullick, M. M. (2012). Event-related potentials and development. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 473-511). New York, NY, US: Oxford University Press. doi: 10.1093/oxfordhb/9780195374148.013.0235
- Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea) nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive psychology*, *65*(1), 1-38. doi: 10.1016/j.cogpsych.2012.01.003
- Comrie, B. (1985). Tense. Cambridge: Cambridge University Press.
- Costa, A., Pannunzi, M., Deco, G., & Pickering, M. J. (2017). Do Bilinguals Automatically Activate Their Native Language When They Are Not Using It? *Cogn Sci*, *41*(6), 1629-1644. doi:10.1111/cogs.12434
- Coulson, S., King, J. W., & Kutas, M. (1998). Expect the unexpected: Event-related brain response to morphosyntactic violations. *Language and Cognitive Processes*, 13(1), 21-58. doi: 10.1080/016909698386582
- Cubelli, R., Paolieri, D., Lotto, L., & Job, R. (2011). The effect of grammatical gender on object categorization. *J Exp Psychol Learn Mem Cogn*, *37*(2), 449-460. doi:10.1037/a0021965
- Czigler, I., Balázs, L., & Pató, L. G. (2004). Visual change detection: event-related potentials are dependent on stimulus location in humans. *Neuroscience Letters*, *364*(3), 149-153. doi: 10.1016/j.neulet.2004.04.048
- Davidoff, J., Davies, I., & Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature, 398*(6724), 203. doi: 10.1038/18335
- de Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language, 30*(1), 90-123. doi: 10.1016/0749-596X(91)90012-9
- de la Fuente, J., Santiago, J., Román, A., Dumitrache, C., & Casasanto, D. (2014). When you think about it, your past is in front of you: how culture shapes spatial conceptions of time. *Psychol Sci, 25*(9), 1682-1690. doi:10.1177/0956797614534695
- Declerck, R., Reed, S., & Cappell, B. (2006). The Grammar of the English Verb, I: The Grammar of the English Tense System: A Comprehensive Analysis. Berlin: Mouton de Gruyter.
- Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. de Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 179-201). New York: Oxford University.
- Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5(3), 175-197. doi: 10.1017/S1366728902003012

- Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011a). Phonology by itself: Masked phonological priming effects with and without orthographic overlap. *Journal of Cognitive Psychology*, 23(2), 185-203. doi: 10.1080/20445911.2011.477811
- Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011b). Two words, one meaning: Evidence of automatic co-activation of translation equivalents. *Frontiers in Psychology*, 2, 188. doi: 0.3389/fpsyg.2011.00188
- Donchin, E. (1981). Surprise!... surprise? *Psychophysiology*, *18*(5), 493-513. doi: 10.1111/j.1469-8986.1981.tb01815.x
- Dowens, M. G., Guo, T. M., Guo, J. J., Barber, H., & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish - Evidence from Event Related Potentials. *Neuropsychologia*, 49(7), 1651-1659. doi:10.1016/j.neuropsychologia.2011.02.034
- Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., & Davies, I. R. L. (2007). Further evidence that Whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences*, 104(3), 1097-1102. doi:10.1073/pnas.0610132104
- Duff, P., & Li, D. (2002). The acquisition and use of perfective aspect in Mandarin. In Salaberry. R & Shirai. Y (Eds.), *The L2 acquisition of tense-aspect morphology* (pp. 417-453). Philadelphia, PA: John Benjamins.
- Duncan-Johnson, C. C., & Donchin, E. (1977). On quantifying surprise: The variation of event-related potentials with subjective probability. *Psychophysiology*, *14*(5), 456-467. doi: 10.1111/j.1469-8986.1977.tb01312.x
- Duyck, W., & Brysbaert, M. (2004). Forward and Backward Number Translation Requires Conceptual Mediation in Both Balanced and Unbalanced Bilinguals. *Journal of Experimental Psychology: Human Perception and Performance*, 30(5), 889-906. doi: http://dx.doi.org/10.1037/0096-1523.30.5.889
- Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 663. doi: 10.1037/0278-7393.33.4.663
- Duyck, W., & Warlop, N. (2009). Translation priming between the native language and a second language: New evidence from Dutch-French bilinguals. *Experimental Psychology*, 56(3), 173-179. doi: 10.1027/1618-3169.56.3.173
- Duñabeitia, J. A., Kinoshita, S., Carreiras, M., & Norris, D. (2011).. Language and Cognitive Processes, 26(4-6), 509-529. doi: 10.1080/01690965.2010.499215
- Duñabeitia, J. A., Perea, M., & Carreiras, M. (2009). Masked translation priming effects with highly proficient simultaneous bilinguals. *Experimental psychology*. 57(2), 98-107 doi: 10.1027/1618-3169/a000013
- Eddington, C. M., & Tokowicz, N. (2013). Examining English–German translation ambiguity using primed translation recognition. *Bilingualism: Language and Cognition, 16*(2), 442-457. doi: 10.1017/S1366728912000387

- Ellis, C., Kuipers, J. R., Thierry, G., Lovett, V., Turnbull, O., & Jones, M. W. (2015). Language and culture modulate online semantic processing. *Social Cognitive and Affective Neuroscience*, 10(10), 1392-1396. doi:10.1093/scan/nsv028
- Fausey, C. M., & Boroditsky, L. (2011). Who dunnit? Cross-linguistic differences in eyewitness memory. *Psychonomic Bulletin & Review*, 18(1), 150-157. doi:10.3758/s13423-010-0021-5
- Ferreira, F., & Karimi, H. (2015). Prosody, performance, and cognitive skill: Evidence from individual differences. In L. F. E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor* (pp. 119-132). Switzerland: Springer International Publishing.
- Ferreira, F., & Patson, N. D. (2007). The "good enough" approach to language comprehension. *Language and Linguistics Compass*, 1(1-2), 71-83. doi:10.1111/j.1749-818X.2007.00007.x
- Finkbeiner, M., Forster, K., Nicol, J., & Nakamura, K. (2004). The role of polysemy in masked semantic and translation priming. *Journal of memory and language*, *51*(1), 1-22. doi: 10.1016/j.jml.2004.01.004
- Flecken, M. (2011). What native speaker judgments tell us about the grammaticalization of a progressive aspectual marker in Dutch. *Linguistics*, 49(3), 479-524. doi:10.1515/ling.2011.015
- 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. doi:10.1016/j.cognition.2015.04.006
- Flecken, M., von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Evidence from a cross-linguistic non-verbal recognition task. *Language and Cognition*, *6*(1), 45-78. doi: 10.1017/langcog.2013.2
- Fodor, J. A. (1983). The modularity of mind: an essay on faculty psyschology. Cambridge: MIT press.
- Fox, J, & Weisberg, S. (2014). Package car.
- Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and language*, *50*(3), 259-281. doi: 10.1006/brln.1995.1048
- Friederici, A. D., & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing processes. *Journal of Psycholinguistic Research*, 25(1), 157-176. doi: 10.1007/BF01708424
- Friederici, A. D., Von Cramon, D. Y., & Kotz, S. A. (1999). Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain*, 122(6), 1033-1047. doi: 10.1093/brain/122.6.1033
- Friedrich, M., & Friederici, A. D. (2010). Maturing brain mechanisms and developing behavioral language skills. *Brain and Language*, 114(2), 66-71. doi: 10.1016/j.bandl.2009.07.004

- Fuhrman, O., & Boroditsky, L. (2010). Cross-cultural differences in mental representations of time: Evidence from an implicit nonlinguistic task. *Cognitive science*, 34(8), 1430-1451. doi: 10.1111/j.1551-6709.2010.01105.x
- Fuhrman, O., McCormick, K., Chen, E., Jiang, H., Shu, D., Mao, S., & Boroditsky, L. (2011). How linguistic and cultural forces shape conceptions of time: English and Mandarin time in 3D. *Cogn Sci*, 35(7), 1305-1328. doi:10.1111/j.1551-6709.2011.01193.x
- Gallistel, C. R. (1989). Animal cognition: The representation of space, time and number. *Annual review of psychology*, 40(1), 155-189. doi:10.1146/annurev.ps.40.020189.001103
- Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind: Advances in the study of language and thought.* Cambridge, MA: MIT press.
- Gentner, D., & Goldin-Meadow, S. (2003). Whither Whorf. In G. D & G.-M. S (Eds.), *Language in mind:Advances in the study of language and thought* (pp. 3-14). Cambridge, MA: MIT Press.
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), 489-494. doi:10.1073/pnas.0509868103
- Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the Whorf effect beyond the realm of color discrimination. *Brain and language*, 105(2), 91-98. doi: 10.1016/j.bandl.2007.06.001
- Glenberg, A.M. and Kaschak, M.P. (2002) Grounding Language in Action. *Psychonomic Bulletin and Review*, 9, 558-565. doi: 10.3758/BF03196313
- Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and cognitive processes*, *25*(2), 149-188. doi: 10.1080/01690960902965951
- Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. In R. J. Harris (Eds.), *Advances in psychology* (Vol. 83). Cognitive process in bilinguals (pp. 207-220). Oxford, England: North-Holland. doi: 10.1016/S0166-4115(08)61496-X
- Grainger, J., & Holcomb, P. J. (2009). Watching the word go by: On the time-course of component processes in visual word recognition. *Language and linguistics compass*, 3(1), 128-156. doi: 10.1111/j.1749-818X.2008.00121.x
- Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactiveactivation model from a developmental perspective (BIA-d). In M. Kail & M. Hickman (Eds.), *Language acquisition across linguistic and cognitive systems* (pp. 267-284). Philadelphia: John Benjamins.
- Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular atficat *Electroencephalography and Clinical Neurophysiology*, 55(4), 468-484. doi:10.1016/0013-4694(83)90135-9
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3-15. doi: 10.1016/0093-934X(89)90048-5
- Gu, Y. A. N., Zheng, Y., & Swerts, M. Having a different pointing of view about the future: The effect of signs on co-speech gestures about time in Mandarin–CSL bimodal bilinguals. *Bilingualism: Language and Cognition*, 1-12. doi:10.1017/S1366728918000652
- Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, 34(6), 660-676. doi: 10.1111/j.1469-8986.1997.tb02142.x
- Hagoort, P. (2003). Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience*, *15*(6), 883-899. doi: 10.1162/089892903322370807
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439-483. doi: 10.1080/01690969308407585
- Hagoort, P., Wassenaar, M., & Brown, C. M. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research*, 16(1), 38-50. doi: 10.1016/S0926-6410(02)00208-2
- Halgren, E., Squires, N. K., Wilson, C. L., Rohrbaugh, J. W., Babb, T. L., & Crandall, P. H. (1980). Endogenous potentials generated in the human hippocampal formation and amygdala by infrequent events. *Science*, 210(4471), 803-805. doi: 10.1126/science.7434000
- Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal behaviour, 61*(1), 139-151. doi: 10.1006/anbe.2000.1518
- Hawkins, R., & Chan, C. Y. (1997). The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis'. *Second Language Research*, 13(3), 187-226. doi: 10.1191/026765897671476153
- 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*. Amsterdam: John Benjamins. Retrive from: https://doi.org/10.1075/lald.30
- Heider, E. R. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, *93*(1), 10. doi: 10.1037/h0032606
- Heider, E. R., & Olivier, D. C. (1972). The structure of the color space in naming and memory for two languages. *Cognitive Psychology*, *3*(2), 337-354. doi: 10.1016/0010-0285(72)90011-4
- Hellström, Å., & Rammsayer, T. H. (2004). Effects of time-order, interstimulus interval, and feedback in duration discrimination of noise bursts in the 50- and 1000-ms ranges. *Acta Psychologica*, *116*(1), 1-20. doi: 10.1016/j.actpsy.2003.11.003
- Hermer-Vazquez, L., Spelke, E. S., & Katsnelson, A. S. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, 39(1), 3-36. doi: 10.1006/cogp.1998.0713
- Hornstein, N. (1990). *As Time Goes By: Tense and Universal Grammar*. Cambridge, Mass: MIT Press.

- Hoshino, N., & Thierry, G. (2012). Do Spanish–English bilinguals have their fingers in two pies–or is it their toes? An electrophysiological investigation of semantic access in bilinguals. *Frontiers in psychology*, *3*, 9. doi: 10.3389/fpsyg.2012.00009
- Jaeger, F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4):434-446. doi: 10.1016/j.jml.2007.11.007
- January, D., & Kako, E. (2007). Re-evaluating evidence for linguistic relativity: Reply to Boroditsky. *Cognition*, *104*(2), 417-426. doi: 10.1016/j.cognition.2006.07.008
- Ji, L. J., Guo, T. Y., Zhang, Z. Y., & Messervey, D. (2009). Looking Into the Past: Cultural Differences in Perception and Representation of Past Information. *Journal of Personality* and Social Psychology, 96(4), 761-769. doi:10.1037/a0014498
- Jiang, N. (1999). Testing processing explanations for the asymmetry in masked crosslanguage priming. *Bilingualism: Language and Cognition*, 2(1), 59-75. doi: 10.1017/S1366728999000152
- Jiang, N., & Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of memory and language*, 44(1), 32-51. doi: 10.1006/jmla.2000.2737
- Johnson, R. (1993). On the neural generators of the P300 component of the event-related potential. *Psychophysiology*, *30*(1), 90-97. doi: 10.1111/j.1469-8986.1993.tb03208.x
- Kaiser, E. (2014). Experimental paradigms in psycholinguistics. In R. J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 135-168). Cambridge: Cambridge University Press.
- Kappenman, E. S., & Luck, S. J. (2012). ERP components: The ups and downs of brainwave recordings. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of eventrelated potential components*(pp. 3-30). New York, NY, US: Oxford University Press. doi: 10.1093/oxfordhb/9780195374148.013.0014
- Kay, P., & Kempton, W. (1984). What is the Sapir-Whorf hypothesis? American anthropologist, 86(1), 65-79. Retrieved from http://www.jstor.org/stable/679389
- Keating, G. D. (2014). Eye-tracking with text. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 69-92). London: Taylor & Francis.
- Keatley, C. W., Spinks, J. A., & De Gelder, B. (1994). Asymmetrical cross-language priming effects. *Memory & Cognition*, 22(1), 70-84. doi: 10.3758/BF03202763
- Kimura, M., Schröger, E., Czigler, I., & Ohira, H. (2010). Human visual system automatically encodes sequential regularities of discrete events. *Journal of Cognitive Neuroscience, 22*(6), 1124-1139. doi: 10.1162/jocn.2009.21299
- Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5(2), 196-214. doi: 10.1162/jocn.1993.5.2.196
- Knight, R. T. (1984). Decreased response to novel stimuli after prefrontal lesions in man. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section, 59(1), 9-20. doi: 10.1016/0168-5597(84)90016-9

- Knight, R. T. (1990). Neural mechanisms of event-related potentials: evidence from human lesion studies. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 3-18). New York, NY, US: Oxford University Press.
- Knight, R. T. (1996). Contribution of human hippocampal region to novelty detection. *Nature, 383*(6597), 256. doi: 10.1038/383256a0
- Knight, R. T., Grabowecky, M. F., & Scabini, D. (1995). Role of human prefrontal cortex in attention control. In H. H. Jasper, S. Riggio, & P. S. Goldman-Rakic (Eds.), Advances in neurology, Vol. 66. Epilepsy and the functional anatomy of the frontal lobe (pp. 21-36). New York, NY, US: Raven Press.
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557-577. doi: 10.1017/S0048577201990559
- Konishi, T. (1993). The semantics of grammatical gender: A cross-cultural study. *Journal of Psycholinguistic Research*, 22(5), 519-534. doi: 10.1007/BF01068252
- Kotz, S. A., Frisch, S., Von Cramon, D. Y., & Friederici, A. D. (2003). Syntactic language processing: ERP lesion data on the role of the basal ganglia. *Journal of the International Neuropsychological Society*, 9(7), 1053-1060. doi: 10.1017/S1355617703970093
- Kousta, S.-T., Vinson, D. P., & Vigliocco, G. (2008). Investigating linguistic relativity through bilingualism: The case of grammatical gender. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(4), 843. doi: 10.1037/0278-7393.34.4.843
- Kramer, A. F., Wickens, C. D., & Donchin, E. (1985). Processing of stimulus properties: evidence for dual-task integrality. *Journal of Experimental Psychology: Human Perception and Performance, 11*(4), 393. doi: 10.1037/0096-1523.11.4.393
- Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second language research*, 18(2), 137-171. doi: 10.1191/0267658302sr201oa
- Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33(2), 149-174. doi: 10.1006/jmla.1994.1008
- Kroll, J. F., Van Hell, J. G., Tokowicz, N., & Green, D. W. (2010). The Revised Hierarchical Model: A critical review and assessment. *Bilingualism: Language and Cognition*, 13(3), 373-381. doi: 10.1017/S136672891000009X
- Kuipers, J.-R., & Thierry, G. (2010). Event-related brain potentials reveal the time-course of language change detection in early bilinguals. *NeuroImage*, 50(4), 1633-1638. doi:https://doi.org/10.1016/j.neuroimage.2010.01.076
- Kuipers, J. R., Jones, M. W., & Thierry, G. (2018). Abstract images and words can convey the same meaning. *Scientific Reports*, 8(1), 7190. doi:10.1038/s41598-018-25441-5
- 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. doi:10.1146/annurev.psych.093008.131123

- Kutas, M., & Hillyard, S. A. (1980a). Reading between the lines event-related brain potentials during natural sentence processing *Brain and Language*, *11*(2), 354-373. doi:10.1016/0093-934x(80)90133-9
- Kutas, M., & Hillyard, S. A. (1980b). Reading senseless sentences brain potentials reflect semantic incongruity. *Science*, 207(4427), 203-205. doi:10.1126/science.7350657
- Kutas, M., & Hillyard, S. A. (1984a). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307(5947), 161-163. doi: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(Jun), 236-241. doi:10.1111/j.1749-6632.1984.tb23540.x
- Kutas, M., Lindamood, T., & Hillyard, S. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 217 237). Hillsdale, NJ: Lawrence Erlbaum.
- Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. *Advances in Psychophysiology*, *3*, 139-187.
- Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified. In M. Traxler, & M. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 83-143). Amsterdam: Elsevier.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). ImerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82.1-26. doi: 10.18637/jss.v082.i13
- Lai, V. T., & Boroditsky, L. (2013). The immediate and chronic influence of spatio-temporal metaphors on the mental representations of time in english, mandarin, and mandarinenglish speakers. *Front Psychol*, 4, 142. doi: 10.3389/fpsyg.2013.00142
- Lai, V. T., Rodriguez, G. G., & Narasimhan, B. (2014). Thinking-for-speaking in early and late bilinguals. *Bilingualism: Language and Cognition*, 17(1), 139-152. doi: 10.1017/S1366728913000151
- Lantz, D., & Stefflre, V. (1964). Language and cognition revisited. *Journal of Abnormal and Social Psychology*, 69, 472-481. doi: 10.1037/h0043769
- Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research*, *14*(4), 359-375. doi:10.1191/026765898672500216
- LeBlanc, R., & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly*, *19*, 673-687. doi: 10.2307/3586670
- Lee, P. (1996). *The Whorf theory complex: A critical reconstruction*. Amsterdam and Philadelphia: John Benjamins Publishing.
- Lehtonen, M., Hulten, A., Rodriguez-Fornells, A., Cunillera, T., Tuomainen, J., & Laine, M. (2012). Differences in word recognition between early bilinguals and monolinguals: Behavioral and ERP evidence. *Neuropsychologia*, 50(7), 1362-1371. doi:10.1016/j.neuropsychologia.2012.02.021
- Lenth R V (2016) Ismeans: Least-Squares Means, R package version 2.22.

- Lewis, P. A., & Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Current Opinion in Neurobiology*, 13(2), 250-255. doi:10.1016/s0959-4388(03)00036-9
- Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.
- Li, Y., Jones, M., & Thierry, G. (2018). Timeline blurring in fluent Chinese-English bilinguals. *Brain Research*, 1701, 93-102. doi:https://doi.org/10.1016/j.brainres.2018.07.008
- Liang, X., Xiao, F., Wu, L., Chen, Q., Lei, Y., & Li, H. (2016). The Temporal Order of Word Presentation Modulates the Amplitudes of P2 and N400 during Recognition of Causal Relations. *Frontiers in psychology*, 7, 1890. doi: 10.3389/fpsyg.2016.01890
- Luck, S. J. (2014). An introduction to the event-related potential technique. Cambridge, MA: MIT press.
- Lucy, J. (1992). *Language diversity and thought*. Cambridge, MA: Cambridge University Press.
- Lucy, J. A. (1992). Language diversity and thought: A reformulation of the linguistic relativity hypothesis. Cambridge: Cambridge University Press.
- Lucy, J. A. (1996). Grammatical categories and cognition: A case study of the linguistic relativity hypothesis. Cambridge: Cambridge University Press.
- Lupyan, G. (2012). Linguistically modulated perception and cognition: the label-feedback hypothesis. *Front Psychol, 3*, 54. doi:10.3389/fpsyg.2012.00054
- Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences of the United States of America*, 110(35), 14196-14201. doi:10.1073/pnas.1303312110
- Macnamara, J. (1967). The bilingual's linguistic performance—a psychological overview. *Journal of Social Issues, 23*(2), 58-77. doi: 10.1111/j.1540-4560.1967.tb00576.x
- Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Universality and language specificity in object naming. *Journal of Memory and Language*, 49(1), 20-42. doi: 10.1016/S0749-596X(03)00021-4
- Marinis, T. (2010). On-line sentence processing methods in typical and atypical populations.In S. Unsworth & E. Blom (Eds.), *Experimental Methods in Language Acquisition Research* (pp. 139-162). Amsterdam: John Benjamins.
- Martín-Loeches, M., Casado, P., Gonzalo, R., De Heras, L., & Fernández-Frías, C. (2006). Brain potentials to mathematical syntax problems. *Psychophysiology*, *43*(6), 579-591. doi: 10.1111/j.1469-8986.2006.00463.x
- McCarthy, G., Wood, C. C., Williamson, P. D., & Spencer, D. D. (1989). Task-dependent field potentials in human hippocampal formation. *Journal of Neuroscience*, *9*(12), 4253-4268. doi: 0.1523/JNEUROSCI.09-12-04253.1989

- McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, 88(5), 375. doi: 10.1037/0033-295X.88.5.375
- Miles, L. K., Tan, L., Noble, G. D., Lumsden, J., & Macrae, C. N. (2011). Can a mind have two time lines? Exploring space-time mapping in Mandarin and English speakers. *Psychonomic Bulletin & Review*, 18(3), 598-604. doi:10.3758/s13423-011-0068-y
- Mo, L., Xu, G., Kay, P., & Tan, L.-H. (2011). Electrophysiological evidence for the leftlateralized effect of language on preattentive categorical perception of color. *Proceedings of the National Academy of Sciences of the United States of America*, 108(34), 14026-14030. doi: doi.org/10.1073/pnas.1111860108
- Molinaro, N., Barber, H. A., & Carreiras, M. (2011). Grammatical agreement processing in reading: ERP findings and future directions. *Cortex*, 47(8), 908-930. doi: 10.1016/j.cortex.2011.02.019
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of cognitive Neuroscience*, *3*(2), 151-165. doi: 10.1162/jocn.1991.3.2.151
- 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. doi:10.1016/j.neuroimage.2006.09.007
- Ng, M. M. R., Goh, W. D., Yap, M. J., Tse, C.-S., & So, W.-C. (2017). How We Think about Temporal Words: A Gestural Priming Study in English and Chinese. *Frontiers in Psychology*, 8(974). doi:10.3389/fpsyg.2017.00974
- Ng, S., & Wicha, N. Y. Y. (2013). Meaning first: a case for language-independent access to word meaning in the bilingual brain. *Neuropsychologia*, *51*(5), 850-863. doi: 10.1016/j.neuropsychologia.2013.01.017
- 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. doi: 10.1162/jocn\_a\_00856
- Nobre, A. C., & McCarthy, G. (1995). Language-related field potentials in the anteriormedial temporal lobe: II. Effects of word type and semantic priming. *Journal of Neuroscience*, *15*(2), 1090-1098. doi: 10.1523/JNEUROSCI.15-02-01090.1995
- Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., . . . Luuk, A. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*(6615), 432. doi: 10.1038/385432a0
- Núñez, R. E., & Sweetser, E. (2006). With the future behind them: convergent evidence from aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cogn Sci*, *30*(3), 401-450. doi:10.1207/s15516709cog0000\_62
- Oppenheim, G., Wu, Y. J., & Thierry, G. (2018). Found in translation: Late bilinguals do automatically activate their native language when they are not using it. Cognitive Science, 42(5), 1700-1713. doi: 10.1111/cogs.12618

- Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*(6), 785-806. doi: 10.1016/0749-596X(92)90039-Z
- Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by gardenpath sentences: evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, memory, and cognition, 20*(4), 786. doi: 10.1037/0278-7393.20.4.786
- Ouellet, M., Santiago, J., Israeli, Z., & Gabay, S. (2010). Is the future the right time? *Experimental psychology*, *57*(4), 308-314. doi: 10.1027/1618-3169/a000036.
- Palmer, A. S., & Bachman, L. F. (1981). Basic concerns in test validation. In J. C. A. A. Hughe (Ed.), *ELTdocuments 111: Issues in language testing*. London, United Kingdom: The British Council.
- Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition*, 108(1), 155-184. doi:10.1016/j.cognition.2008.02.007
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, 6(7), 674. doi: 10.1038/nn1082
- Pavlenko, A., & Malt, B. C. (2011). Kitchen Russian: Cross-linguistic differences and firstlanguage object naming by Russian–English bilinguals. *Bilingualism: Language and Cognition*, 14(1), 19-45. doi: 10.1017/S136672891000026X
- Pazo-Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). MMN in the visual modality: a review. *Biological psychology*, *63*(3), 199-236. doi: 10.1016/S0301-0511(03)00049-8
- Perea, M., Dunabeitia, J. A., & Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *Journal of Memory and Language*, *58*(4), 916-930. doi: 10.1016/j.jml.2008.01.003
- Petten, C. V. (1993). A comparison of lexical and sentence-level context effects in eventrelated potentials. *Language and Cognitive Processes*, 8(4), 485-531. doi: 10.1080/01690969308407586
- Phillips, W., & Boroditsky, L. (2003). Can quirks of grammar affect the way you think? Grammatical gender and object concepts. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society*, 928-933. Retrive from https://escholarship.org/uc/item/31t455gf
- Phillips, W., & Santos, L. R. (2007). Evidence for kind representations in the absence of language: Experiments with rhesus monkeys (Macaca mulatta). *Cognition*, 102(3), 455-463. doi: 10.1016/j.cognition.2006.01.009
- Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, R., ... Rugg, M. D. (2000). Guidelines for using human event-related potentials to study cognition: recording standards and publication criteria. *Psychophysiology*, 37(2), 127-152. doi: 10.1111/1469-8986.3720127
- Pinker, S. (1995). The language instinct: The new science of language and mind (Vol. 7529). London: Penguin.

- Pinker, S. (2007). Toward a consilient study of literature. *Philosophy and Literature, 31*(1), 162-178. doi: 10.1353/phl.2007.0016
- Polich, J. (2012). Neuropsychology of P300. In S. J. Luck & E. S. Kappenman (Eds.), The Oxford handbook of event-related potential components (pp. 159-188). New York, NY, US: Oxford University Press. doi: 10.1093/oxfordhb/9780195374148.013.0089
- Qiu, Y. C., & Zhou, X. L. (2012). Processing temporal agreement in a tenseless language: An ERP study of Mandarin Chinese. *Brain Research*, 1446, 91-108. doi:10.1016/j.brainres.2012.01.051
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English Language*. London: Longman.
- R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rea, P. M. (1981). Formative assessment of student performance: The role of selfappraisal. *Indian Journal of Applied Linguistics*, 7, 66-88.
- Reichenbach, H. (1947). Elements of Symbolic Logic. New York: Macmillan & Co.
- Reichle, E. D., Pollatsek, A., & Rayner, K. (2007). Modeling the effects of lexical ambiguity on eye movements during reading. In R. van Gompel (Ed.), *Eye Movements: A Window on Mind and Brain* (pp. 271-292). Oxford: Elsevier Science.
- Reid, V. M., Hoehl, S., Grigutsch, M., Groendahl, A., Parise, E., & Striano, T. (2009). The neural correlates of infant and adult goal prediction: evidence for semantic processing systems. *Developmental Psychology*, 45(3), 620. doi: 10.1037/a0015209
- Ritter, P., & Villringer, A. (2006). Simultaneous EEG–fMRI. *Neuroscience & Biobehavioral Reviews*, *30*(6), 823-838. doi: 10.1016/j.neubiorev.2006.06.008
- Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: The effect of verbal interference. *Memory & Cognition, 28*(6), 977-986. doi:10.3758/bf03209345
- Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, 50(4), 378-411. doi: 10.1016/j.cogpsych.2004.10.001
- Rösler, F., Pütz, P., Friederici, A., & Hahne, A. (1993). Event-related brain potentials while encountering semantic and syntactic constraint violations. *Journal of Cognitive Neuroscience*, 5(3), 345-362. doi: 10.1162/jocn.1993.5.3.345
- Salillas, E., & Wicha, N. Y. Y. (2012). Early Learning Shapes the Memory Networks for Arithmetic: Evidence From Brain Potentials in Bilinguals. *Psychological Science*, 23(7), 745-755. doi:10.1177/0956797612446347
- Sams, M., Paavilainen, P., Alho, K., & Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 62(6), 437-448. doi: 10.1016/0168-5597(85)90054-1

- Sanoudaki, E., & Thierry, G. (2015). Language non-selective syntactic activation in early bilinguals: the role of verbal fluency. *International Journal of Bilingual Education and Bilingualism*, *18*(5), 548-560. doi: 10.1080/13670050.2015.1027143
- Sapir, E. (1949). *Selected writings in language, culture and personality*. California: Univ of California Press.
- Sassenhagen, J., & Fiebach, C. (2018). Finding the P3 in the P600: Decoding shared neural mechanisms of responses to syntactic violations and oddball target detection. *BioRxiv*, 449322. doi: 10.1101/449322
- Schmiedtová, B., Von Stutterheim, C., & Carroll, M. (2011). Language-specific patterns in event construal of advanced second language speakers. In A. Pavlenko (Ed.), *Thinking and Speaking in Two Languages* (pp. 112-141). Clevedon: Multilingual Matters.
- Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition*, 37(5), 569-586. doi: 10.3758/MC.37.5.569
- Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, 55(2), 197-212. doi: 10.1016/j.jml.2006.03.004
- Sera, M. D., Elieff, C., Forbes, J., Burch, M. C., Rodríguez, W., & Dubois, D. P. (2002). When language affects cognition and when it does not: An analysis of grammatical gender and classification. *Journal of Experimental Psychology: General*, 131(3), 377. doi: 10.1037/0096-3445.131.3.377
- Singmann, H., Bolker, B., Westfall, J., & Aust, F. (2017). Afex: analysis of factorial experiments. R package version 0.17-8. http://cran.r-project.org/package=afex.
- Slobin, D. (1996). From 'thought and language' to 'thinking for speaking'. In J. Gumperz & S. Levinson (Eds.), *Rethinking Linguistic Relativity* (pp. 70-96). Cambridge: Cambridge University Press.
- Smith, C. (1991). *The Parameter of Aspect* (1st ed.). Dordrecht: Kluwer Academic Publishers.
- Smith, C. (2008). Time with and without tense. In J. Guéron & J. Lacarme (Eds.), *Time and Modality* (pp. 227–249). Berlin: Springer.
- Spalek, K., Hoshino, N., Wu, Y. J., Damian, M., & Thierry, G. (2014). Speaking two languages at once: Unconscious native word form access in second language production. *Cognition*, 133(1), 226-231. doi: 10.1016/j.cognition.2014.06.016
- Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology*, 38(4), 387-401. doi: 10.1016/0013-4694(75)90263-1
- Steinhauer, K., & Drury, J. E. (2012). On the early left-anterior negativity (ELAN) in syntax studies. *Brain and language*, *120*(2), 135-162. doi: 10.1016/j.bandl.2011.07.001
- Steinhauer, K., & Ullman, M. T. (2002). Consecutive ERP effects of morpho-phonology and morpho-syntax. *Brain and Language*, 83(1), 62-65.

- 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. doi: https://doi.org/10.1177/0267658308098995
- Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, 150(3700), 1187-1188. doi: 10.1126/science.150.3700.1187
- Swaab, T. Y., Baynes, K., & Knight, R. T. (2002). Separable effects of priming and imageability on word processing: an ERP study. *Cognitive Brain Research*, 15(1), 99-103. doi: 10.1016/S0926-6410(02)00219-7
- Thierry, G. (2016). Neurolinguistic Relativity: How Language Flexes Human Perception and Cognition. *Lang Learn*, *66*(3), 690-713. doi:10.1111/lang.12186
- Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences of the United States of America*, 106(11), 4567-4570. doi:10.1073/pnas.0811155106
- Thierry, G., & Wu, Y. J. (2004). Electrophysiological evidence for language interference in late bilinguals. *NeuroReport*, 15(10), 1555-1558. doi: 10.1097/01.wnr.0000134214.57469.c2
- Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, 104(30), 12530-12535. doi:10.1073/pnas.0609927104
- Thomas, M. S. C., & Allport, A. (2000). Language switching costs in bilingual visual word recognition. *Journal of Memory and Language*, 43(1), 44-66. doi: 10.1006/jmla.1999.2700
- Titone, R. (1990). La lingua straniera nella scuola elementare. Roma: Armando Editore.
- Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes*, 22(5), 727-779. doi: 10.1080/01690960601057068
- Traugott, E. C. (1978). On the expression of spatio-temporal relations in language. In J. H. Greeberg, C. A. Ferguson, and E. Moravcsit (Eds.), *Universals of human language*, 3(pp. 369-400). Stanford: Stanford University Press.
- Tse, C.-S., & Altarriba, J. (2008). Evidence against linguistic relativity in Chinese and English: A case study of spatial and temporal metaphors. *Journal of Cognition and Culture, 8*(3-4), 335-357. doi: 10.1163/156853708X358218
- Tversky, B., Kugelmass, S., & Winter, A. (1991). Cross-cultural and developmental trends in graphic productions. *Cognitive Psychology*, 23(4), 515-557. doi: 10.1016/0010-0285(91)90005-9
- van Hell, J. G., & De Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, *128*(3), 431-451. doi: 10.1016/j.actpsy.2008.03.010

- van Heuven, W. J. B., & Dijkstra, T. (2010). Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models. *Brain Research Reviews, 64*(1), 104-122. doi: 10.1016/j.brainresrev.2010.03.002
- van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*(3), 458-483. doi: 10.1006/jmla.1998.2584
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and wordfrequency in event-related brain potentials. *Memory & Cognition*, 18(4), 380-393. doi: 10.3758/BF03197127
- Van Petten, C., & Luka, B. J. (2006). Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language*, 97(3), 279-293. doi: 10.1016/j.bandl.2005.11.003
- VanPatten, B. (2014). The psycholinguistics of SLA. In (Eds.), *Theories in second language* acquisition: An introduction (pp. 1-19). New York: Routledge.
- Vaughan-Evans, A., Kuipers, J. R., Thierry, G., & Jones, M. W. (2014). Anomalous transfer of syntax between languages. *Journal of Neuroscience*, 34(24), 8333-8335. doi: 10.1523/JNEUROSCI.0665-14.2014
- Vigliocco, G., Vinson, D. P., Paganelli, F., & Dworzynski, K. (2005). Grammatical gender effects on cognition: implications for language learning and language use. *Journal of Experimental Psychology: General, 134*(4), 501. doi: 10.1037/0096-3445.134.4.501
- von Elek, T. (1981). Self-assessment of Swedish as a second language(Sja1vbedomning av fardigheter i svenska som andra sprak). Gijteborg, Sweden: University of Goteborg, Language Teaching Research Center.
- von Elek, T. (1982). *Test of Swedish as a second language: An experiment in self-assessment.* Goteborg, Sweden: University of Goteborg, Language Teaching Research Center.
- von Humboldt, W. F. (1963). Humanist without portfolio: An anthology of the writings of Wilhelm von Humboldt: Wayne State University Press.
- von Stutterheim, C., & Carroll, M. (2006). The impact of grammatical temporal categories on ultimate attainment in L2 learning. In H. Byrnes., H. Weger-Guntharp., & K. Sprang. (Eds.), *Educating for Advanced Foreign Language Capacities* (pp. 40–53). Georgetown: GUP.
- Wang, P. P., & Bellugi, U. (1994). Evidence from two genetic syndromes for a dissociation between verbal and visual-spatial short-term memory. *Journal of Clinical and Experimental Neuropsychology*, 16(2), 317-322. doi: 10.1080/01688639408402641
- Wang, W. S. Y., & Sun, C. (2015). The Oxford handbook of Chinese linguistics. New York, NY, US: Oxford University Press. doi: 10.1093/oxfordhb/9780199856336.001.0001
- Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, 8(3), 231-256. doi: 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). doi:10.1371/journal.pone.0052318
- Whorf, B. (1956). Language, thought, and reality. Cambridge, MA: MIT Press.
- Whorf, B. (1940). Science and linguistics. Cambridge, MA: MIT Press.
- Wijers, A. A., Mulder, G., Okita, T., Mulder, L. J. M., & Scheffers, M. K. (1989). Attention to color: An analysis of selection, controlled search, and motor activation, using eventrelated potentials. *Psychophysiology*, 26(1), 89-109. doi: 10.1111/j.1469-8986.1989.tb03137.x
- Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of United States of America*, 104(19), 7780-7785. doi: 10.1073/pnas.0701644104
- Witzel, C., & Gegenfurtner, K. R. (2011). Is there a lateralized category effect for color? *Journal of Vision*, 11(12), 16. doi: 10.1167/11.12.16.
- Wu, Y. J., & Thierry, G. (2010). Chinese-English Bilinguals Reading English Hear Chinese. *Journal of Neuroscience*, 30(22), 7646-7651. doi:10.1523/jneurosci.1602-10.2010
- Wu, Y. J., & Thierry, G. (2012). Unconscious translation during incidental foreign language processing. *Neuroimage*, *59*(4), 3468-3473. doi:10.1016/j.neuroimage.2011.11.049
- Yang, W., Sun, & Y. J. P. B., (2016). A monolingual mind can have two time lines: Exploring space-time mappings in Mandarin monolinguals. *Psychon Bull Rev.*, 23(3), 857-864. doi:10.3758/s13423-015-0964-7

 $\infty$ 

## Appendix A

Bilingual Language inventory questionnaire

P	articipant #							Date
			Languag	ge Histo	ory Ques	stionnair	e	
This 70u	questionnaire is desig be as accurate as poss	gned to give ible when a	us a bet nswering	ter unde g the fol	erstandii llowing	ng of yo questior	ur lang 1s.	uage experience. We ask that
l.C	ountry of Birth		Wh	at is yo	ur first l	anguage	e	
2.	When did you start lea	arning Engl	ish? A	.ge:				
	How long have you le	arnt Englis	h?					
	How long have you li	ved-in Engl	ish-speal	king co	untries i	n total?	_	yearsmon
	Have you taken any E	nglish Prof	iciency 7	fest, suc	ch as IE	LTS or T	FOFEL	
	If Yes. How muc	h have vou	got and V	When?				
	Please rate your Engl 1 not literate	<b>ish</b> <u>reading</u> 2 3	proficier 4	ncy on a 5	ten-poi 6	nt scale 7	8	9 10 very literate
	Please rate your Eng	lish writing	g proficie	ency on	a ten-po	oint scale	e.	·
	1 not literate	2 3	4	5	6	7	8	9 10 very literate
	Please rate your Engl	ish <u>convers</u>	ational fl	uency o	on a ten-	point sc	ale.	
	1 not fluent	2 3	4	5	6	7	8	9 10 very fluent
	Please rate your Engl	ish speech	comprehe	ension a	bility o	n a ten-p	point sc	cale.
	l unable to understand conversat	2 3 ion	4	5	6	7	8	9 10 perfectly able to understand conversation
0.	Please rate your Chi	nese readin	g profici	ency on	a ten-p	oint scal	le.	
	1 not literate	2 3	4	5	6	7	8	9 10 very literate
1.	Please rate your Chi	nese <u>writin</u>	g proficie	ency on	a ten-p	oint scal	e.	
	1 not literate	2 3	4	5	6	7	8	9 10 very literate
2.	Please rate your Chi	nese conve	rsational	fluency	on a te	n-point s	scale.	
	1 not fluent	2 3	4	5	6	7	8	9 10 very fluent
3.	Please rate your Chi	nese speech	o compre	hensior	ability	on a ten	-point	scale.
	1 unable to understand conversat	2 3	4	5	6	7	8	9 10 perfectly able to understand conversation

14.	What p addition	ercenta of Q14	age of t 4 & Q1	he time 5 is 100	do you ) %.	spend u	ising Ei	nglish?	-	%.	* Please 1	nake sure	that the
15.	What p	ercenta	age of t	he time	do you	spend u	ising Cl	hinese?	_	%.			
	16. Apart langu	t from l ages?	English	, have y	ou lear	nt any c	other for	reign la	nguage?			If yes	, which
	17. Wher 18. How 19. How	n you si many y often d	tart to l years ha lo you ι	earn tho ave you ise thos	ose lang learnt f e langu	uages? hose lar ages rig	nguages	?			 		
	20. Have	you w	ritten ir	n vertica	ıl colun	nns arra	nged fro	om left	to right	?			
		1	2	3	4	5	6	7	8	9	10		
	21. Have	you re	ad in v	ertical c	olumns	arrange	ed from	left to 1	right?				
		1	2	3	4	5	6	7	8	9	10		
	22. Have	you w 1	ritten in 2	n horizo 3	ntal col 4	umns an 5	rranged 6	from u 7	p to dov 8	vn? 9	10		
	23. Have	you re 1	ad in h 2	orizonta 3	ıl colun 4	nns arrai 5	nged fro 6	om up to 7	o down? 8	9	10		

## Appendix B

Experimental Stimuli used in Study 1

### Presentation List A <sup>14</sup>

Sentence	Condition	Block
After my aunt had worked as a nurse for forty years, she retired from the hospital.	Past perfect control	1
After my aunt has worked as a nurse for forty years, she retired from the hospital.	Present past violation	2
After my aunt will have worked as a nurse for forty years, she retired from the hospital.	Future past violation	3
After my aunt had worked as a nurse for forty years, she retired from the fish.	Semantic violation	4
After the director of the school had resigned from the university, he will work for a multinational.	Fillers	1
After the director of the school has resigned from the university, he will work for a multinational.	Fillers	2
After the director of the school will have resigned from the university, he will work for a multinational.	Fillers	3
After the director of the school has resigned from the university, he will work for a meter.	Fillers	4
After my uncle had taken part in the local racing event, he repaired his car.	Past perfect control	4
After my uncle has taken part in the local racing event, he repaired his car.	Present past violation	1
After my uncle will have taken part in the local racing event, he repaired his car.	Future past violation	2
After my uncle had taken part in the local racing event, he repaired his cake.	Semantic violation	3
After the ship had sailed for three days, it will arrive at its destination.	Fillers	4
After the ship has sailed for three days, it will arrive at its destination.	Fillers	1
After the ship will have sailed for three days, it will arrive at its destination.	Fillers	2
After the ship has sailed for three days, it will arrive at its pizza.	Fillers	3
After a scientist had decoded monkey genes, other scholars also wanted to study them.	Past perfect control	3
After a scientist has decoded monkey genes, other scholars also wanted to study them.	Present past violation	4
After a scientist will have decoded monkey genes, other scholars also wanted to study them.	Future past violation	1
After a scientist had decoded monkey genes, other scholars also wanted to study cheese.	Semantic violation	2
After George had received a formal invitation, he will visit the White House.	Fillers	3
After George has received a formal invitation, he will visit the White House.	Fillers	4
After George will have received a formal invitation, he will visit the White House.	Fillers	1
After George has received a formal invitation, he will visit the egg.	Fillers	2
After the two companies had reached an agreement, the lawyer drafted a contract.	Past perfect control	2
After the two companies have reached an agreement, the lawyer drafted a contract.	Present past violation	3
After the two companies will have reached an agreement, the lawyer drafted a contract.	Future past violation	4
After the two companies had reached an agreement, the lawyer drafted a butter.	Semantic violation	1
After he had confirmed the number of guests, he will book the rooms.	Fillers	2
After he has confirmed the number of guests, he will book the rooms.	Fillers	3
After he will have confirmed the number of guests, he will book the rooms.	Fillers	4
After he has confirmed the number of guests, he will book the mile.	Fillers	1
After the German football team had defeated Brazil, they faced Argentina in the finals.	Past perfect control	1
After the German football team has defeated Brazil, they faced Argentina in the finals.	Present past violation	2
After the German football team will have defeated Brazil, they faced Argentina in the finals.	Future past violation	3
After the German football team had defeated Brazil, they faced Argentina in the sister.	Semantic violation	4

<sup>&</sup>lt;sup>14</sup> Presentation list B was the same as presentation list A with future forms replaced by simple past forms and simple past forms replaced by future forms.

After the oil company had found the technical fault, it will dispatch engineers to the platforms. Fillers After the oil company has found the technical fault, it will dispatch engineers to the platforms. Fillers After the oil company will have found the technical fault, it will dispatch engineers to the platforms. Fillers After the oil company has found the technical fault, it will dispatch engineers to the soup. Fillers After the doctor had thoroughly examined the patient, she prescribed expensive drugs. Past perfect control After the doctor has thoroughly examined the patient, she prescribed expensive drugs. Present past violation After the doctor will have thoroughly examined the patient, she prescribed expensive drugs. Future past violation After the doctor had thoroughly examined the patient, she prescribed expensive fire. Semantic violation After he had enjoyed a night at the opera, he will walk back home. Fillers After he has enjoyed a night at the opera, he will walk back home. Fillers After he will have enjoyed a night at the opera, he will walk back home. Fillers After he has enjoyed a night at the opera, he will walk back rice. Fillers After he had bought a painting, he donated it to the Royal Museum. Past perfect control After he has bought a painting, he donated it to the Royal Museum. Present past violation After he will have bought a painting, he donated it to the Royal Museum. Future past violation After he had bought a painting, he donated it to pig. Semantic violation After the new hospital had received new beds, it will open to the public. Fillers After the new hospital has received new beds, it will open to the public. Fillers After the new hospital will have received new beds, it will open to the public. Fillers After the new hospital has received new beds, it will open to the pencil. Fillers After the government had announced tax increase, citizens protested outside parliament. Past perfect control After the government has announced tax increase, citizens protested outside parliament. Present past violation After the government will have announced tax increase, citizens protested outside parliament. Future past violation After the government had announced tax increase, citizens protested outside noodle. Semantic violation Fillers After we had finished the meal, my mother will wash the dishes. After we have finished the meal, my mother will wash the dishes. Fillers After we will have finished the meal, my mother will wash the dishes. Fillers Fillers After we have finished the meal, my mother will wash the sky. After George had served as branch manager for many years, the company promoted him to general manager. Past perfect control After George has served as branch manager for many years, the company promoted him to general manager. Present past violation After George will have served as branch manager for ten years, the company promoted him to general manager. Future past violation After George had served as branch manager for many years, the company promoted him to general snake. Semantic violation After the last customer had finished his meal, the restaurant will close its door. Fillers Fillers After the last customer has finished his meal, the restaurant will close its door. After the last customer will have finished his meal, the restaurant will close its door. Fillers After the last customer has finished his meal, the restaurant will close its boyfriend. Fillers After my supervisor had finished data collection, she analysed it for two weeks. Past perfect control After my supervisor has finished data collection, she analysed it for two weeks. Present past violation After my supervisor will have finished data collection, she analysed it for two weeks. Future past violation After my supervisor had finished data collection, she analysed it for two mirror. Semantic violation After his leg had recovered, he will play football again. Fillers After his leg has recovered, he will play football again. Fillers After his leg will have recovered, he will play football again. Fillers After his leg has recovered, he will play football chopstick. Fillers

1

2

3

4

Λ

1

2

3

4

1

2

3

3

4

1

2

3

4

1

2

2

3

4

1

2

3

2 3

4

1

2

3

4

4

1

2

3

4

1

2

After my sister had bought the house, she painted the wall in dark green. F Ρ After my sister has bought the house, she painted the wall in dark green. After my sister will have bought the house, she painted the wall in dark green. F After my sister had bought the house, she painted the wall in dark psychology. S After the company had launched a new project, it will recruit new employees. F After the company has launched a new project, it will recruit new employees. F After the company will have launched a new project, it will recruit new employees. F After the company has launched a new project, it will recruit new March. F After the advertisement had finished, they watched the football match. F F After the advertisement has finished, they watched the football match. F After the advertisement will have finished, they watched the football match. After the advertisement had finished, they watched the football meet. S F After he had arrived in Beijing, he will taste the local specialties. After he has arrived in Beijing, he will taste the local specialties. F After he will have arrived in Beijing, he will taste the local specialties. F After he has arrived in Beijing, he will taste the local cousin. F After my mother had tasted steak, she cooked some every week F After my mother has tasted steak, she cooked some every week. P After my mother will have tasted steak, she cooked some every week. F After my mother had tasted steak, she cooked some every lake S After my sister had obtained her driving license, she will plane to drive across Canada. F After my sister has obtained her driving license, she will plane to drive across Canada. F After my sister will have obtained her driving license, she will plane to drive across Canada. F After my sister has obtained her driving license, she will plane to drive across cherry. F Ρ After my brother had obtained this orchard, he planted many apple trees. After my brother has obtained this orchard, he planted many apple trees. F After my brother will have obtained this orchard, he planted many apple trees. F After my brother had obtained this orchard, he planted many apple film. S After my son had graduated from University, he will apply for a high-paying job. F After my son has graduated from University, he will apply for a high-paying job. F After my son will have graduated from University, he will apply for a high-paying job. F After my son has graduated from University, he will apply for a high-paying rabbit. F After the match had ended, a politician awarded prizes to the winner. F Р After the match has ended, a politician awarded prizes to the winner. After the match will have ended, a politician awarded prizes to the winner. F After the match had ended, a politician awarded prizes to the jacket. S F After the actress had played in a Hollywood movie, she will increase her fees. After the actress has played in a Hollywood movie, she will increase her fees. F After the actress will have played in a Hollywood movie, she will increase her fees. F F After the actress has played in a Hollywood movie, she will increase her banana. After my uncle had found a new job, he stopped collecting unemployment benefit. Р After my uncle has found a new job, he stopped collecting unemployment benefit. F After my uncle will have found a new job, he stopped collecting unemployment benefit. F After my uncle had found a new job, he stopped collecting unemployment zoo. S

ast perfect control	3
resent past violation	4
uture past violation	1
emantic violation	2
illers	3
illers	4
illers	1
illers	2
ast perfect control	2
resent past violation	3
uture past violation	4
emantic violation	1
illers	2
illers	3
illers	4
illers	1
ast perfect control	1
resent past violation	2
uture past violation	3
emantic violation	4
illers	1
illers	2
illers	3
illers	4
ast perfect control	4
resent past violation	1
uture past violation	2
emantic violation	3
illers	4
illers	1
illers	2
illers	3
ast perfect control	3
resent past violation	4
uture past violation	1
emantic violation	2
illers	3
illers	4
illers	1
illers	2
ast perfect control	2
resent past violation	3
uture past violation	4
emantic violation	1

After my cousin had married, she will move to China.	Fillers	2
After my cousin has married, she will move to China.	Fillers	3
After my cousin will have married, she will move to China.	Fillers	4
After my cousin has married, she will move to pork.	Fillers	1
After the factory had applied the new environmental policy, the authorities approved production.	Past perfect control	1
After the factory has applied the new environmental policy, the authorities approved production.	Present past violation	2
After the factory will have applied the new environmental policy, the authorities approved production.	Future past violation	3
After the factory had applied the new environmental policy, the authorities approved sun.	Semantic violation	4
After the teenager had bought chocolate, she will bake chocolate cookies.	Fillers	1
After my teenager has bought chocolate, she will bake chocolate cookies.	Fillers	2
After my teenager will have bought chocolate, she will bake chocolate cookies.	Fillers	3
After the teenager has bought chocolate, she will bake chocolate disease.	Fillers	4
After the war had ceased, people returned to their home country.	Past perfect control	4
After the war has ceased, people returned to their home country.	Present past violation	1
After the war will have ceased, people returned to their home country.	Future past violation	2
After the war had ceased, people returned to their home forefinger.	Semantic violation	3
After my father had replaced the broken parts, the washing machine will run once again.	Fillers	4
After my father has replaced the broken parts, the washing machine will run once again.	Fillers	1
After my father will have replaced the broken parts, the washing machine will run once again.	Fillers	2
After my father has replaced the broken parts, the washing machine will run once King.	Fillers	3
After the war had begun, many refugees fled their homeland.	Past perfect control	3
After the war has begun, many refugees fled their homeland.	Present past violation	4
After the war will have begun, many refugees fled their homeland.	Future past violation	1
After the war had begun, many refugees fled their pink.	Semantic violation	2
After my aunt had read the recipe, she will make pasta the right way.	Fillers	3
After my aunt has read the recipe, she will make pasta the right way.	Fillers	4
After my aunt will have read the recipe, she will make pasta the right way.	Fillers	1
After my aunt has read the recipe, she will make pasta the right poem.	Fillers	2
After the police had started to hunt him, he flew to France.	Past perfect control	2
After the police have started to hunt him, he flew to France.	Present past violation	3
After he police will have started to hunt him, he flew to France.	Future past violation	4
After the police had started to hunt him, he flew to desk.	Semantic violation	1
After the general had fought many fierce battles, he will leave the army.	Fillers	2
After the general has fought many fierce battles, he will leave the army.	Fillers	3
After the general will have fought many fierce battles, he will leave the army.	Fillers	4
After the general has fought many fierce battles, he will leave the orange.	Fillers	1
After the government had installed a new traffic system, it became more efficient.	Past perfect control	1
After the government has installed a new traffic system, it became more efficient.	Present past violation	2
After the government will have installed a new traffic system, it became more efficient.	Future past violation	3
After the government had installed a new traffic system, it became more coffee.	Semantic violation	4
After my uncle had agreed that I borrow his car, I will drive it to London.	Fillers	1
After my uncle has agreed that I borrow his car, I will drive it to London.	Fillers	2
After my uncle will have agreed that I borrow his car, I will drive it to London.	Fillers	3
After my uncle has agreed that I borrow his car, I will drive it to cash.	Fillers	4

After the concert had begun, the famous artist sung a masterpiece.	Past perfect control	4
After the concert has begun, the famous artist sung a masterpiece.	Present past violation	1
After the concert will have begun, the famous artist sung a masterpiece.	Future past violation	2
After the concert had begun, the famous artist sung a strawberry.	Semantic violation	3
After I had topped up, I will ring my parents.	Fillers	4
After I have topped up, I will ring my parents.	Fillers	1
After I will have topped up, I will ring my parents.	Fillers	2
After I have topped up, I will ring my jacket.	Fillers	3
After the waitress had confirmed the booking, she led us to our table.	Past perfect control	3
After the waitress has confirmed the booking, she led us to our table.	Present past violation	4
After the waitress will have confirmed the booking, she led us to our table.	Future past violation	1
After the waitress had confirmed the booking, she led us to our snow.	Semantic violation	2
After he had suffered a headache, he will take two pills.	Fillers	3
After he has suffered a headache, he will take two pills.	Fillers	4
After he will have suffered a headache, he will take two pills.	Fillers	1
After he has suffered he had a headache, he will take two railways.	Fillers	2
After the weather had turned cold, she wore a thick blue jacket.	Past perfect control	2
After the weather has turned cold, she wore a thick blue jacket.	Present past violation	3
After the weather will have turned cold, she wore a thick blue jacket.	Future past violation	4
After the weather had turned cold, she wore a thick blue week.	Semantic violation	1
After the enemy had strengthened massively, the general will withdraw his soldiers.	Fillers	2
After the enemy has strengthened massively, the general will withdraw his soldiers.	Fillers	3
After the enemy will have strengthened massively, the general will withdraw his soldiers.	Fillers	4
After the enemy has strengthened massively, the general will withdraw his oven.	Fillers	1
After I had learnt how to hold my breath under water, the coach taught me freestyle.	Past perfect control	1
After I have learnt how to hold my breath under water, the coach taught me freestyle.	Present past violation	2
After I will have learnt how to hold my breath under water, the coach taught me freestyle.	Future past violation	3
After I had learnt how to hold my breath under water, the coach taught me refrigerator.	Semantic violation	4
After my mother had mended her bicycle, she will ride home with my father.	Fillers	1
After my mother has mended her bicycle, she will ride home with my father.	Fillers	2
After my mother will have mended her bicycle, she will ride home with my father.	Fillers	3
After my mother has mended her bicycle, she will ride home with my room.	Fillers	4
After security had inspected the stadium, it held its first match.	Past perfect control	4
After security has inspected the stadium, it held its first match.	Present past violation	1
After security will have inspected the stadium, it held its first match.	Future past violation	2
After security had inspected the stadium, it held its first boyfriend.	Semantic violation	3
After I had said sorry for breaking her vase, my mother will forgive me.	Fillers	4
After I have said sorry for breaking her vase, my mother will forgive me.	Fillers	1
After I will have said sorry for breaking her vase, my mother will forgive me.	Fillers	2
After I have said sorry for breaking her vase, my mother will forgive computer.	Fillers	3
After the war between the two oil countries had ended, oil prices fell by a big margin.	Past perfect control	3
After the war between the two oil countries has ended, oil prices fell by a big margin.	Present past violation	4
After the war between the two oil countries will have ended, oil prices fell by a big margin.	Future past violation	1
After the war between the two oil countries had ended, oil prices fell by a big movie.	Semantic violation	2

After the company had recruited new administrative employees, it will build a bigger reception. Fillers After the company has recruited new administrative employees, it will build a bigger reception. Fillers After the company will have recruited new administrative employees, it will build a bigger reception. Fillers After the company has recruited new administrative employees, it will build a bigger daughter. Fillers After my daughter had finished her ice cream, I gave her one more scoop. Past perfect control After my daughter has finished her ice cream, I gave her one more scoop. Present past violation After my daughter will have finished her ice cream, I gave her one more scoop. Future past violation After my daughter had finished her ice cream, I gave her one more autumn. Semantic violation After my brother had sought advice from a senior, he will choose his modules. Fillers After my brother has sought advice from a senior, he will choose his modules. Fillers After my brother will have sought advice from a senior, he will choose his modules. Fillers After my brother has sought advice from a senior, he will choose his cloudy. Fillers Past perfect control After the summer sale had started, she bought many expensive clothes. After the summer sale has started, she bought many expensive clothes. Present past violation After the summer sale will have started, she bought many expensive clothes. Future past violation After the summer sale had started, she bought many expensive wars. Semantic violation After my uncle had joined the army, he probably will fight every year. Fillers After my uncle has joined the army, he probably will fight every year. Fillers After my uncle will have joined the army, he probably will fight every year. Fillers After my uncle has joined the army, he probably will fight every juice. Fillers After we had finished the meal, my grandpa paid the bill. Past perfect control After we have finished the meal, my grandpa paid the bill. Present past violation After we will have finished the meal, my grandpa paid the bill. Future past violation After we had finished the meal, my grandpa paid the road. Semantic violation After my brother had revised, he will sit his final exams. Fillers After my brother has revised, he will sit his final exams. Fillers After my brother will have revised, he will sit his final exams. Fillers After my brother has revised, he will sit his final clock. Fillers After the club had appointed a famous coach, it won many matches. Past perfect control After the club has appointed a famous coach, it won many matches. Present past violation Future past violation After the club will have appointed a famous coach, it won many matches. After the club had appointed a famous coach, it won many winds. Semantic violation After the apples had turned red, the farmers will sell them. Fillers Fillers After the apples have turned red, the farmers will sell them. After the apples will have turned red, the farmers will sell them. Fillers After the apples have turned red, the farmers will sell neck. Fillers After the student had read the novel, she wrote an e-mail to the author. Past perfect control After the student has read the novel, she wrote an e-mail to the author. Present past violation After the student will have read the novel, she wrote an e-mail to the author. Future past violation After the student had read the novel, she wrote an e-mail to the sugar. Semantic violation After my cousin had looked at a map, he will know the location of Paris. Fillers After my cousin has looked map, he will know the location of Paris. Fillers After my cousin will have looked map, he will know the location of Paris. Fillers After my cousin has looked map, he will know the location of smile. Fillers

3

Δ

1

2

2

3

4

1

2

3

Λ

1

2

3

4

2

3

4

4

1

2

3

4

2

з

3

4

1

2

3

4

1

2

2

3

4

1

2

3

4

After the police had searched the crime scene, they found the weapon.	Past perfect control
After the police have searched the crime scene, they found the weapon.	Present past violation
After the police will have searched the crime scene, they found the weapon.	Future past violation
After the police had searched the crime scene, they found the laugh.	Semantic violation
After the little boy had agreed to go to bed, his mother will tell him a bedtime story.	Fillers
After the little boy has agreed to go to bed, his mother will tell him a bedtime story.	Fillers
After the little boy will have agreed to go to bed, his mother will tell him a bedtime story.	Fillers
After the little boy has agreed to go to bed, his mother will tell him a bedtime furniture.	Fillers
After the world cup had ended, the winners drank champagne.	Past perfect control
After the world cup has ended the winners drank champagne.	Present past violation
After the world cup will have ended, the winners drank champagne.	Future past violation
After the world cup had ended, the winners drank house.	Semantic violation
After the holiday had begun, my brother will sleep late into the morning.	Fillers
After the holiday has begun, my brother will sleep late into the morning.	Fillers
After the holiday will have begun, my brother will sleep late into the morning.	Fillers
After the holiday has begun, my brother will sleep late into the medicine.	Fillers
After my aunt had announced her marriage, she got lots of presents.	Past perfect control
After my sister has announced her marriage, she got lots of presents.	Present past violation
After my sister will have announced her marriage, she got lots of presents.	Future past violation
After my aunt had announced her marriage, she got lots of science.	Semantic violation
After she had given birth, she will eat much less than before.	Fillers
After she has given birth, she will eat much less than before.	Fillers
After she will have given birth, she will eat much less than before.	Fillers
After she has given birth, she will eat much less than rubbish.	Fillers

# Appendix C

Supplementary materials of Study 1





ERPs elicited by the critical verb and the post-critical word in the three experimental conditions in Chinese-English bilingual participants at 12 electrode sites.



Appendix C2 - Supplementary Figure 2

ERPs elicited by the critical verb and the post-critical word in the three experimental conditions in Native speakers of English at 12 electrode sites.

#### Appendix C3 - Supplementary Analyses

**Split-group analyses.** Yao and Chen (2016) investigated how late Chinese-English bilinguals are affected by cross-linguistic differences, using tense and progressive aspect violation tasks. They divided their participants into two groups based on proficiency in L2 English: (a) a high proficiency group including participants having obtained a score of 7 or above in the IELTS, having passed the TEM-8 test, or having achieved 100 or above in the TOEFL; and (b) a low proficiency group having passed the CET-4 test (equivalent to a score of 6 in the IELTS). They found that Chinese-English bilinguals with a score of 7 or above in the IELTS were able to detect tense violations.

Accordingly, we divided our group of 27 Chinese-English bilingual participants into two proficiency subgroups: (1) A high proficiency group including bilingual participants with an IELTS score of 7 or above (n = 14; Mean = 7.2),; (2) A lower proficiency group including bilingual participants with an IELTS score of 6.5 (n = 8) and we dismissed five participants who received instruction exclusively through the medium of English since high school and were thus high functioning bilinguals although we did not have a quantitative measurement of proficiency for them.

N400 mean amplitudes were analysed by means of repeated measures analyses of variance (ANOVA) with group (high / low proficiency) as between-subject factor and conditions (correct control, PPM, FPM) as within subject factor. There was no significant effect of group (F(1,20) = 1.36, p = 0.26,  $\eta_p^2 = 0.06$ ) or condition (F(2, 40) = 0.27, p = 0.76,  $\eta_p^2 = 0.01$ ) on N4-1 mean amplitude and the interaction between group and condition was not significant either (F(2,40) = 2.37, p = 0.11,  $\eta_p^2 = 0.11$ ). There was no significant effect of group (F(1,20) = 2.07, p = 0.17,  $\eta_p^2 = 0.09$ ) or condition (F(2,40) = 0.2, p = 0.82,  $\eta_p^2 = 0.01$ ) on the mean amplitude of the N4-2 either and the interaction between group and condition was not significant (F(2,40) = 1.53, p = 0.23,  $\eta_p^2 = 0.07$ ).

**Correlation analyses.** We found no significant correlations between participants' performance in an offline judgement task and N400 modulations in the N4-1 (PPM: r = -0.18, p = 0.43, n = 22; *FPM:* r = -0.06, p = 0.78, n = 22) or the N4-2 (PPM: r = -0.31, p = 0.17, n = 22; *FPM:* r = -0.05, p = 0.83, n = 22) analysis windows.

Additional proficiency test. In addition, we collected overt time alignment judgement accuracy using the most difficult misalignment condition PPM and the correct control condition in a new group of Chinese-English bilingual participants with proficiency closely matched to the group tested in the main experiment (n=21,  $M_{IELTS \ score}$ = 6.86, SD = 0.5, range 6.5–8).

These participants with a very similar proficiency to those tested in the ERP experiment did not significantly differ in accuracy from a control group of English native speakers tested on the same task, F (1,37) = 0.95, p=0.34 (Mean Accuracy PPM, English: 77%, Bilingual: 64%; Mean Accuracy control, English: 93%, Bilingual: 96%), making it very unlikely for the lack of differences between participants in the ERP experiment to hide an overall difference in proficiency.

### Appendix D

Experimental stimuli used in Study 2

### Presentation List A <sup>15</sup>

#### Sentence Condition

Sentence	Condition	Block
After my aunt had worked as a nurse, she will retire to Spain.	Past Perfect Gap	1
After my aunt has worked as a nurse, she will retire to Spain.	Present Perfect Correct	2
After my aunt worked as a nurse, she will retire to Spain.	Past Simple Gap	3
After my aunt works as a nurse, she will retire to Spain.	Present Simple Correct	4
After the head of school had resigned, he worked for a bank.	Past Perfect Correct	1
After the head of school has resigned, he worked for a bank.	Present Perfect Clash	2
After the head of school resigned, he worked for a bank.	Past Simple Correct	3
After the head of school resigns, he worked for a bank.	Present Simple Clash	4
After my uncle had raced his car, he will repair his gearbox.	Past Perfect Gap	4
After my uncle has raced his car, he will repair his gearbox.	Present Perfect Correct	1
After my uncle raced his car, he will repair his gearbox.	Past Simple Gap	2
After my uncle races his car, he will repair his gearbox.	Present Simple Correct	3
After the ship had sailed at full speed, it arrived at its destination.	Past Perfect Correct	4
After the ship has sailed at full speed, it arrived at its destination.	Present Perfect Clash	1
After the ship sailed at full speed, it arrived at its destination.	Past Simple Correct	2
After the ship sails at full speed, it arrived at its destination.	Present Simple Clash	3
After the scientist had decoded the genes, everyone will want to study them.	Past Perfect Gap	3
After the scientist has decoded the genes, everyone will want to study them.	Present Perfect Correct	4
After the scientist decoded the genes, everyone will want to study them.	Past Simple Gap	1
After the scientist decodes the genes, everyone will want to study them.	Present Simple Correct	2
After George had received a formal invitation, he visited the White House.	Past Perfect Correct	3
After George has received a formal invitation, he visited the White House.	Present Perfect Clash	4
After George received a formal invitation, he visited the White House.	Past Simple Correct	1
After George receives a formal invitation, he visited the White House.	Present Simple Clash	2
After the company had recruited new employees, the lawyer will draft the employment contracts.	Past Perfect Gap	2
After the company has recruited new employees, the lawyer will draft the employment contracts.	Present Perfect Correct	3
After the company recruited new employees, the lawyer will draft the employment contracts.	Past Simple Gap	4
After the company recruits new employees, the lawyer will draft the employment contracts.	Present Simple Correct	1
After he had confirmed the number of guests, he booked the rooms.	Past Perfect Correct	2
After he has confirmed the number of guests, he booked the rooms.	Present Perfect Clash	3
After he confirmed the number of guests, he booked the rooms.	Past Simple Correct	4
After he confirms the number of guests, he booked the rooms.	Present Simple Clash	1
After the wrestler had defeated his opponent, he will face an Argentinean in the finals.	Past Perfect Gap	1
After the wrestler has defeated his opponent, he will face an Argentinean in the finals.	Present Perfect Correct	2

<sup>&</sup>lt;sup>15</sup> Presentation list B was the same as presentation list A with future forms replaced by simple past forms and simple past forms replaced by future forms.

After the wrestler defeated his opponent, he will face an Argentinean in the finals After the wrestler defeats his opponent, he will face an Argentinean in the finals. After the engineer had found the fault, he dispatched technicians to the platforms. After the engineer has found the fault, he dispatched technicians to the platforms. After the engineer found the fault, he dispatched technicians to the platforms. After the engineer finds the fault, he dispatched technicians to the platforms. After the doctor had examined the patient, she will prescribe the most expensive drugs After the doctor has examined the patient, she will prescribe the most expensive drugs. After the doctor examined the patient, she will prescribe the most expensive drugs. After the doctor examines the patient, she will prescribe the most expensive drugs After he had enjoyed a night at the opera, he walked back home After he has enjoyed a night at the opera, he walked back home After he enjoyed a night at the opera, he walked back home After he enjoys a night at the opera, he walked back home. After he had bought a painting, he will donate it to the Museum. After he has bought a painting, he will donate it to the Museum. After he bought a painting, he will donate it to the Museum After he buys a painting, he will donate it to the Museum After the new hospital had received beds, it opened to the public. After the new hospital has received beds, it opened to the public After the new hospital received beds, it opened to the public. After the new hospital receives beds, it opened to the public. After John had reached the airport, he will contact me immediately. After John has reached the airport, he will contact me immediately After John reached the airport, he will contact me immediately. After John reaches the airport, he will contact me immediately. After John had finished his meal, his brother washed the dishes After John has finished his meal, his brother washed the dishes After John finished his meal, his brother washed the dishes. After John finishes his meal, his brother washed the dishes After Ian had developed the software by himself, the company will promote him to director. After Ian has developed the software by himself, the company will promote him to director. After lan developed the software by himself, the company will promote him to director. After Ian develops the software by himself, the company will promote him to director. After the last customer had finished her dessert, the restaurant closed for the night. After the last customer has finished her dessert, the restaurant closed for the night. After the last customer finished her dessert, the restaurant closed for the night After the last customer finishes her dessert, the restaurant closed for the night. After my supervisor had finished data collection, she will analyse it for two weeks. After my supervisor has finished data collection, she will analyse it for two weeks After my supervisor finished data collection, she will analyse it for two weeks After my supervisor finishes data collection, she will analyse it for two weeks

Past Simple Gap	3
Present Simple Correct	4
Past Perfect Correct	1
Present Perfect Clash	2
Past Simple Correct	3
Present Simple Clash	4
Past Perfect Gap	4
Present Perfect Correct	1
Past Simple Gap	2
Present Simple Correct	3
Past Perfect Correct	4
Present Perfect Clash	1
Past Simple Correct	2
Present Simple Clash	3
Past Perfect Gap	3
Present Perfect Correct	4
Past Simple Gap	1
Present Simple Correct	2
Past Perfect Correct	3
Present Perfect Clash	4
Past Simple Correct	1
Present Simple Clash	2
Past Perfect Gap	2
Present Perfect Correct	3
Past Simple Gap	4
Present Simple Correct	1
Past Perfect Correct	2
Present Perfect Clash	3
Past Simple Correct	4
Present Simple Clash	1
Past Perfect Gap	1
Present Perfect Correct	2
Past Simple Gap	3
Present Simple Correct	4
Past Perfect Correct	1
Present Perfect Clash	2
Past Simple Correct	3
Present Simple Clash	4
Past Perfect Gap	4
Present Perfect Correct	1
Past Simple Gap	2
Present Simple Correct	3

After he had recovered from his injury, he played football again. After he has recovered from his injury, he played football again. After he recovered from his injury, he played football again. After he recovers from his injury, he played football again. After my sister had bought the house, she will paint the kitchen light blue. After my sister has bought the house, she will paint the kitchen light blue. After my sister bought the house, she will paint the kitchen light blue. After my sister buys the house, she will paint the kitchen light blue. After the company had launched a new project, it recruited twenty new employees. After the company has launched a new project, it recruited twenty new employees After the company launched a new project, it recruited twenty new employees After the company launches a new project, it recruited twenty new employees After the advertisement had finished, he will watch the rest of the match. After the advertisement has finished, he will watch the rest of the match. After the advertisement finished, he will watch the rest of the match. After the advertisement finishes, he will watch the rest of the match. After he had arrived in Beijing, he tasted the local delicacies After he has arrived in Beijing, he tasted the local delicacies After he arrived in Beijing, he tasted the local delicacies After he arrives in Beijing, he tasted the local delicacies After my mother had tasted steak, she will cook meat every week. After my mother has tasted steak, she will cook meat every week. After my mother tasted steak, she will cook meat every week. After my mother tastes steak, she will cook meat every week After Jane had obtained her visa, she travelled across China. After Jane has obtained her visa, she travelled across China After Jane obtained her visa, she travelled across China After Jane obtains her visa, she travelled across China After my brother had purchased an orchard, he will plant many apple trees. After my brother has purchased an orchard, he will plant many apple trees After my brother purchased an orchard, he will plant many apple trees After my brother purchases an orchard, he will plant many apple trees. After my son had graduated from University, he established his own company. After my son has graduated from University, he established his own company. After my son graduated from University, he established his own company. After my son graduates from University, he established his own company. After the match had ended, a celebrity will award prizes to the winner. After the match has ended, a celebrity will award prizes to the winner. After the match ended, a celebrity will award prizes to the winner. After the match ends, a celebrity will award prizes to the winner After the actress had played in a Hollywood movie, she increased her fees After the actress has played in a Hollywood movie, she increased her fees

Past Perfect Correct	4
Present Perfect Clash	1
Past Simple Correct	2
Present Simple Clash	3
Past Perfect Gap	3
Present Perfect Correct	4
Past Simple Gap	1
Present Simple Correct	2
Past Perfect Correct	3
Present Perfect Clash	4
Past Simple Correct	1
Present Simple Clash	2
Past Perfect Gap	2
Present Perfect Correct	3
Past Simple Gap	4
Present Simple Correct	1
Past Perfect Correct	2
Present Perfect Clash	3
Past Simple Correct	4
Present Simple Clash	1
Past Perfect Gap	1
Present Perfect Correct	2
Past Simple Gap	3
Present Simple Correct	4
Past Perfect Correct	1
Present Perfect Clash	2
Past Simple Correct	3
Present Simple Clash	4
Past Perfect Gap	4
Present Perfect Correct	1
Past Simple Gap	2
Present Simple Correct	3
Past Perfect Correct	4
Present Perfect Clash	1
Past Simple Correct	2
Present Simple Clash	3
Past Perfect Gap	3
Present Perfect Correct	4
Past Simple Gap	1
Present Simple Correct	2
Past Perfect Correct	3
Present Perfect Clash	4

After the actress played in a Hollywood movie, she increased her fees.	Past Simple Correct
After the actress plays in a Hollywood movie, she increased her fees.	Present Simple Clash
After my uncle had lost his job, he will collect unemployment benefits.	Past Perfect Gap
After my uncle has lost his job, he will collect unemployment benefits.	Present Perfect Correct
After my uncle lost his job, he will collect unemployment benefits.	Past Simple Gap
After my uncle loses his job, he will collect unemployment benefits.	Present Simple Correct
After my cousin had married, she moved back to China.	Past Perfect Correct
After my cousin has married, she moved back to China.	Present Perfect Clash
After my cousin married, she moved back to China.	Past Simple Correct
After my cousin marries, she moved back to China.	Present Simple Clash
After the factory had applied a new environmental policy, the minister will approve production.	Past Perfect Gap
After the factory has applied a new environmental policy, the minister will approve production.	Present Perfect Correct
After the factory applied a new environmental policy, the minister will approve production.	Past Simple Gap
After the factory applies a new environmental policy, the minister will approve production.	Present Simple Correct
After the teenager had bought chocolate, she baked delicious cookies.	Past Perfect Correct
After the teenager has bought chocolate, she baked delicious cookies.	Present Perfect Clash
After the teenager bought chocolate, she baked delicious cookies.	Past Simple Correct
After the teenager buys chocolate, she baked delicious cookies.	Present Simple Clash
After the war had ended, the refugee will return to his homeland.	Past Perfect Gap
After the war has ended, the refugee will return to his homeland.	Present Perfect Correct
After the war ended, the refugee will return to his homeland.	Past Simple Gap
After the war ends, the refugee will return to his homeland.	Present Simple Correct
After my father had replaced the broken parts, the engine ran again.	Past Perfect Correct
After my father has replaced the broken parts, the engine ran again.	Present Perfect Clash
After my father replaced the broken parts, the engine ran again.	Past Simple Correct
After my father replaces the broken parts, the engine ran again.	Present Simple Clash
After the war had begun, Rebecca will flee her homeland.	Past Perfect Gap
After the war has begun, Rebecca will flee her homeland.	Present Perfect Correct
After the war began, Rebecca will flee her homeland.	Past Simple Gap
After the war begins, Rebecca will flee her homeland.	Present Simple Correct
After my aunt had got the recipe, she made authentic Italian pasta.	Past Perfect Correct
After my aunt has got the recipe, she made authentic Italian pasta.	Present Perfect Clash
After my aunt got the recipe, she made authentic Italian pasta.	Past Simple Correct
After my aunt gets the recipe, she made authentic Italian pasta.	Present Simple Clash
After the inquiry had started to become public, he will fly out of the country.	Past Perfect Gap
After the inquiry has started to become public, he will fly out of the country.	Present Perfect Correct
After the inquiry started to become public, he will fly out of the country.	Past Simple Gap
After the inquiry starts to become public, he will fly out of the country.	Present Simple Correct
After the general had broken his leg in combat, he left the army.	Past Perfect Correct
After the general has broken his leg in combat, he left the army.	Present Perfect Clash
After the general broke his leg in combat, he left the army.	Past Simple Correct

After the government had installed a new traffic system, it will become easier to drive. After the government has installed a new traffic system, it will become easier to drive. After the government installed a new traffic system, it will become easier to drive. After the government installs a new traffic system, it will become easier to drive. After my uncle had agreed to let Jenny borrow his car, she drove to London. After my uncle has agreed to let Jenny borrow his car, she drove to London. After my uncle agreed to let Jenny borrow his car, she drove to London. After my uncle agrees to let Jenny borrow his car, she drove to London. After the concert had begun, the artist will sing his biggest hit. After the concert has begun, the artist will sing his biggest hit. After the concert began, the artist will sing his biggest hit. After the concert begins, the artist will sing his biggest hit. After John had topped up his phone, he rang his parents. After John has topped up his phone, he rang his parents. After John topped up his phone, he rang his parents. After John tops up his phone, he rang his parents. After the waitress had confirmed the booking, she will lead us to our table. After the waitress has confirmed the booking, she will lead us to our table. After the waitress confirmed the booking, she will lead us to our table. After the waitress confirms the booking, she will lead us to our table. After he had undergone keyhole surgery, he took some time off. After he has undergone keyhole surgery, he took some time off. After he underwent keyhole surgery, he took some time off. After he undergoes keyhole surgery, he took some time off. After the president had arrived, he will meet the Queen at Buckingham Palace. After the president has arrived, he will meet the Queen at Buckingham Palace. After the president arrived, he will meet the Queen at Buckingham Palace. After the president arrives, he will meet the Queen at Buckingham Palace. After the enemy had strengthened massively, the general withdrew his soldiers. After the enemy has strengthened massively, the general withdrew his soldiers. After the enemy strengthened massively, the general withdrew his soldiers. After the enemy strengthens massively, the general withdrew his soldiers. After Peter had bought a fishing pole, his cousin will teach him how to fish. After Peter has bought a fishing pole, his cousin will teach him how to fish. After Peter bought a fishing pole, his cousin will teach him how to fish. After Peter buys a fishing pole, his cousin will teach him how to fish. After my mother had adjusted her saddle, she rode across the field. After my mother has adjusted her saddle, she rode across the field. After my mother adjusted her saddle, she rode across the field. After my mother adjusts her saddle, she rode across the field. After the security guard had inspected the stadium, he will hold a safety workshop. After the security guard has inspected the stadium, he will hold a safety workshop.

Past Perfect Gap 1 Present Perfect Correct 2 Past Simple Gap 3 Present Simple Correct 4 Past Perfect Correct 1 Present Perfect Clash 2 Past Simple Correct 3 Present Simple Clash 4 Past Perfect Gap 4 Present Perfect Correct 1 Past Simple Gap 2 Present Simple Correct з Past Perfect Correct 4 Present Perfect Clash 1 Past Simple Correct 2 Present Simple Clash 3 Past Perfect Gap 3 Present Perfect Correct 4 Past Simple Gap 1 Present Simple Correct 2 Past Perfect Correct 3 Present Perfect Clash 4 Past Simple Correct 1 Present Simple Clash 2 Past Perfect Gap 2 Present Perfect Correct 3 Past Simple Gap 4 Present Simple Correct 1 Past Perfect Correct 2 Present Perfect Clash 3 Past Simple Correct 4 Present Simple Clash 1 Past Perfect Gap Present Perfect Correct 2 Past Simple Gap 3 Present Simple Correct 4 Past Perfect Correct 1 Present Perfect Clash 2 Past Simple Correct 3 Present Simple Clash 4 Past Perfect Gap 4 Present Perfect Correct 1

After the security guard inspected the stadium, he will hold a safety workshop.	Past
After the security guard inspects the stadium, he will hold a safety workshop.	Prese
After Peter had said sorry for breaking a vase, his mother forgave him reluctantly.	Past I
After Peter has said sorry for breaking a vase, his mother forgave him reluctantly.	Prese
After Peter said sorry for breaking a vase, his mother forgave him reluctantly.	Past
After Peter says sorry for breaking a vase, his mother forgave him reluctantly.	Prese
After the dispute between the two countries had ended, the euro will fall against the dollar.	Past I
After the dispute between the two countries has ended, the euro will fall against the dollar.	Prese
After the dispute between the two countries ended, the euro will fall against the dollar.	Past
After the dispute between the two countries ends, the euro will fall against the dollar.	Prese
After he had secured more materials, Richard built a bigger garage.	Past I
After he has secured more materials, Richard built a bigger garage.	Prese
After he secured more materials, Richard built a bigger garage.	Past
After he secures more materials, Richard built a bigger garage.	Prese
After my daughter had bought flowers, she will give them to me.	Past I
After my daughter has bought flowers, she will give them to me.	Prese
After my daughter bought flowers, she will give them to me.	Past
After my daughter buys flowers, she will give them to me.	Prese
After my brother had sought advice from a teacher, he chose his modules.	Past I
After my brother has sought advice from a teacher, he chose his modules.	Prese
After my brother sought advice from a teacher, he chose his modules.	Past
After my brother seeks advice from a teacher, he chose his modules.	Prese
After the summer sale had started, she will buy many clothes.	Past I
After the summer sale has started, she will buy many clothes.	Prese
After the summer sale started, she will buy many clothes.	Past
After the summer sale starts, she will buy many clothes.	Prese
After my uncle had joined the army, he fought in a war.	Past I
After my uncle has joined the army, he fought in a war.	Prese
After my uncle joined the army, he fought in a war.	Past
After my uncle joins the army, he fought in a war.	Prese
After Anna had finished her meal, her grandpa will pay the bill.	Past I
After Anna has finished her meal, her grandpa will pay the bill.	Prese
After Anna finished her meal, her grandpa will pay the bill.	Past
After Anna finishes her meal, her grandpa will pay the bill.	Prese
After my brother had revised, he sat seven final exams.	Past I
After my brother has revised, he sat seven final exams.	Prese
After my brother revised, he sat seven final exams.	Past
After my brother revises, he sat seven final exams.	Prese
After the club had appointed a famous coach, it will win the league.	Past
After the club has appointed a famous coach, it will win the league.	Prese
After the club appointed a famous coach, it will win the league.	Past
After the club appoints a famous coach, it will win the league.	Prese

Past Simple Gap	2
Present Simple Correct	3
Past Perfect Correct	4
Present Perfect Clash	1
Past Simple Correct	2
Present Simple Clash	3
Past Perfect Gap	3
Present Perfect Correct	4
Past Simple Gap	1
Present Simple Correct	2
Past Perfect Correct	3
Present Perfect Clash	4
Past Simple Correct	1
Present Simple Clash	2
Past Perfect Gap	2
Present Perfect Correct	3
Past Simple Gap	4
Present Simple Correct	1
Past Perfect Correct	2
Present Perfect Clash	3
Past Simple Correct	4
Present Simple Clash	1
Past Perfect Gap	1
Present Perfect Correct	2
Past Simple Gap	3
Present Simple Correct	4
Past Perfect Correct	1
Present Perfect Clash	2
Past Simple Correct	3
Present Simple Clash	4
Past Perfect Gap	4
Present Perfect Correct	1
Past Simple Gap	2
Present Simple Correct	3
Past Perfect Correct	4
Present Perfect Clash	1
Past Simple Correct	2
Present Simple Clash	3
Past Perfect Gap	3
Present Perfect Correct	4
Past Simple Gap	1
Present Simple Correct	2

After the fruit had ripened, the farmer sold them
After the fruit has ripened, the farmer sold them
After the fruit ripened, the farmer sold them
After the fruit ripens, the farmer sold them

After the student had finished reading the novel, she will write a book review. After the student has finished reading the novel, she will write a book review. After the student finished reading the novel, she will write a book review. After the student finishes reading the novel, she will write a book review. After my cousin had looked at a map, he knew the location of Limerick. After my cousin has looked at a map, he knew the location of Limerick. After my cousin looked at a map, he knew the location of Limerick. After my cousin looks at a map, he knew the location of Limerick. After the inspector had searched the crime scene, she will find the weapon. After the inspector has searched the crime scene, she will find the weapon. After the inspector searched the crime scene, she will find the weapon. After the inspector searches the crime scene, she will find the weapon. After the little boy had agreed to go to bed, his mother told him a story. After the little boy has agreed to go to bed, his mother told him a story. After the little boy agreed to go to bed, his mother told him a story. After the little boy agrees to go to bed, his mother told him a story. After the tournament had ended, the winner will drink too much wine. After the tournament has ended, the winner will drink too much wine. After the tournament ended, the winner will drink too much wine. After the tournament ends, the winner will drink too much wine. After the holiday had begun, my brother slept late every morning. After the holiday has begun, my brother slept late every morning. After the holiday began, my brother slept late every morning After the holiday begins, my brother slept late every morning. After my aunt had announced her marriage, she will get lots of presents. After my aunt has announced her marriage, she will get lots of presents. After my aunt announced her marriage, she will get lots of presents. After my aunt announces her marriage, she will get lots of presents. After she had given birth, she ate much less. After she has given birth, she ate much less. After she gave birth, she ate much less.

Past Perfect Correct	3
Present Perfect Clash	4
Past Simple Correct	1
Present Simple Clash	2
Past Perfect Gap	2
Present Perfect Correct	3
Past Simple Gap	4
Present Simple Correct	1
Past Perfect Correct	2
Present Perfect Clash	3
Past Simple Correct	4
Present Simple Clash	1
Past Perfect Gap	1
Present Perfect Correct	2
Past Simple Gap	3
Present Simple Correct	4
Past Perfect Correct	1
Present Perfect Clash	2
Past Simple Correct	3
Present Simple Clash	4
Past Perfect Gap	4
Present Perfect Correct	1
Past Simple Gap	2
Present Simple Correct	3
Past Perfect Correct	4
Present Perfect Clash	1
Past Simple Correct	2
Present Simple Clash	3
Past Perfect Gap	3
Present Perfect Correct	4
Past Simple Gap	1
Present Simple Correct	2
Past Perfect Correct	3
Present Perfect Clash	4
Past Simple Correct	1
Present Simple Clash	2

After she gives birth, she ate much less.

# Appendix E

Experimental stimuli used in Study 4
# **Vertical Experiment**

Critical Stimulus	Translation	Translation Arrow Direction Condition	
上	up	Up	Spatial Congruent
下	down	Up	Spatial Incongruent
上	up	Down	Spatial Incongruent
下	down	Down	Spatial Congruent
上个月	last month	Up	Spatiotemporal Congruent
下个月	next month	Up	Spatiotemporal Incongruent
去年	last year	Up	Temporal Congruent
明年	next year	Up	Temporal Incongruent
上个月	last month	Down	Spatiotemporal Incongruent
下个月	next month	Down	Spatiotemporal Congruent
去年	last year	Down	Temporal Incongruent
明年	next year	Down	Temporal Congruent

# Horizontal Experiment

Critical Stimulus	Translation	Arrow Direction	Condition
左	left	Left	Spatial Congruent
右	right	Left	Spatial Incongruent
左	left	Right	Spatial Incongruent
右	right	Right	Spatial Congruent
上个月	last month	Left	Spatiotemporal Congruent
下个月	next month	Left	Spatiotemporal Incongruent
去年	last year	Left	Temporal Congruent
明年	next year	Left	Temporal Incongruent
上个月	last month	Right	Spatiotemporal Incongruent
下个月	next month	Right	Spatiotemporal Congruent
去年	last year	Right	Temporal Incongruent
明年	next year	Right	Temporal Congruent

Filler stimulus	s Translation	Filler stimulus	Translation	Filler stimulus	Translation
苹果汁	apple juice	橄榄油	olive oil	糖果	candy
鱼馅饼	fish pie	洋葱圈	onion ring	櫻桃	cherry
冰咖啡	iced coffee	排骨	pork ribs	蜂蜜	honey
红茶	black tea	豌豆汤	pea soup	鸭肉	duck
蓝芝士	blue cheese	红酒	red wine	葡萄	grapes
辣椒酱	chill sauce	烤牛肉	roast beef	羊肉	lamb
果干	dried fruit	豆奶	soya milk	橙	orange
水果酒	fruit cider	燕麦饼	oat cookies	梨	pear
豆沙拉	bean salad	芒果酱	mango jam	花生	peanut
热狗	hot dog	苏打水	soda water	披萨	pizza
冰淇淋	ice cream	煎鸡蛋	fried egg	*	rice
柠檬塔	lemon tart	烤火腿	roast ham	酸奶	yogurt
肉丸	meat ball	绿豌豆	green pea	金枪鱼	tuna

# Appendix F

# Publication

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

#### Brain Research 1701 (2018) 93-102



# Contents lists available at ScienceDirect

### Brain Research



journal homepage: www.elsevier.com/locate/bres

#### Research report

## Timeline blurring in fluent Chinese-English bilinguals



# Yang Li, Manon Jones, Guillaume Thierry\*

School of Psychology, Bangor University, Wales LL57 2AS, UK

#### ARTICLE INFO

Article history: Received 8 August 2017 Received in revised form 2 May 2018 Accepted 9 July 2018 Available online 19 July 2018

Keywords: Reference time Tenselessness Event-related potentials Syntax-semantics interference Linguistic relativity Sentence processing

#### ABSTRACT

Linguistic relativity effects arising from differences in terminology and syntax between languages have now been established in various domains of human cognition. Although metaphors have been shown to affect time conceptualisation, there is little evidence to date that the presence or absence of tense within a given language can affect how one processes temporal sequences of events. Here, we set out to characterise how native speakers of Mandarin Chinese – a tenseless language- deal with reference time misalignment using event-related brain potentials. Fluent Chinese-English participants and native speakers of English made acceptability judgements on sentences in which the adjunct clause started with the connective 'after' and was either temporally aligned or not with the main clause in terms of reference time conveyed by the verb. Native speakers of English failed to overtly report such reference time misalignments between clauses, but significant N400 modulations showed that they nevertheless required additional semantic processing effort. Chinese speakers, however, showed no such N400 modulation suggesting that they did not covertly detect reference time misalignments between clauses in real time. Critically, all participants manifested normal sentence comprehension as shown by a standard N400 semantic violation elicited by incongruent endings. We conclude that Chinese speakers of English experience difficulties locating events on a timeline in relation to one another when temporal information is conveyed by tense.

Crown Copyright © 2018 Published by Elsevier B.V. All rights reserved.

#### 1. Introduction

Recent research has provided evidence that language influences cognitive functioning (Athanasopoulos, 2009; Boroditsky, 2001; Boutonnet et al., 2012; Choi and Bowerman, 1991; Centner and Goldin-Meadow, 2003; Lantz and Stefflre, 1964; Lucy, 1992; Lupyan and Ward, 2013; Whorf, 1956). Such effects have been demonstrated at the level of elementary visual perception (Thierry et al., 2009) and object categorisation (Boutonnet et al., 2012; Cubelli et al., 2011; Phillips and Boroditsky, 2003), through to high-level, abstract meaning processing such as event conceptualization (Flecken et al., 2015) and cultural semantics (Ellis et al., 2015).

Grammatical variations between languages can influence event conceptualization as shown by studies of motion event categorization both in language tasks (Flecken, 2011; von Stutterheim and Carroll, 2006) and non-verbal tasks (Athanasopoulos, 2009; Athanasopoulos and Bylund, 2013; Flecken et al., 2015; Flecken et al., 2014). For example, both English and Arabic speakers – whose native languages have aspect markers– spontaneously mention the temporal properties of motion events (e.g., "Two women are walking down a path"). In contrast, native speakers of German -whose mother tongue lacks aspect categorization altogetherdescribe the same events in more holistic terms, including the mention of a possible endpoint (von Stutterheim and Carroll, 2006).

Here, we set out to examine whether linguistic differences in tense marking can affect the representation of temporal relationships between events. Tense is a linguistic device that locates a given situation in time (Declerck et al., 2006). It is accepted that English is a tensed language although there is a controversy over the existence of the future tense (Comrie, 1985; Declerck et al., 2006; Quirk et al., 1985). Other languages lack absolute tense altogether. That is, they do not mark either present-future or past-non-past distinctions in their grammar. In Mandarin Chinese, for example, specifying the temporal location of an event is not compulsory (Comrie, 1985; Li and Thompson, 1989). Instead, temporal information is optionally expressed through time adverbials (e.g. zuó tiān - 'yesterday'; míng tiān - 'tomorrow'), modal auxiliaries (e.g. yào - 'will'; jiāng - 'will'), or through context (Duff and Li, 2002; Smith, 1991), the default position being that the event unfolds in the present (Smith, 2008), Oiu and Zhou (2012), for instance, found that native speakers of Mandarin Chinese are sensitive to the disagreement between a modal auxiliary (e.g., jiā ng yào - 'will') or an aspectual particle (e.g., the marker of perfect

<sup>\*</sup> Corresponding author.

E-mail address: g.thierry@bangor.ac.uk (G. Thierry).

https://doi.org/10.1016/j.brainres.2018.07.008 0006-8993/Crown Copyright © 2018 Published by Elsevier B.V. All rights reserved.

#### Y. Li et al./Brain Research 1701 (2018) 93-102



Fig. 1. Timeline representation of the critical stimuli. E: Event time; R: Reference time; S: Speech time. Index number identifies the clause (1 for the adjunct clause and 2 for the main clause).

#### Table 1

Examples of sentence stimuli in each of the experimental conditions.

Condition	Stimulus example	Chinese translation (pin-yin)
Control Correct	After the director of the school had resigned from the university, he worked for a multinational.	院长从大学辞职后, 他去了一家跨国公司工作。
Present-Past Misalignment	*After the director of the school has resigned from the university, he worked for a multinational.	yuàn zhăng cóng dà xué cí zhí hòu, tā qù-le yī jiā kuà guó gōng sī gōng zuò
Future-Past Misalignment	*After the director of the school will have resigned from the university, he worked for a multinational.	
Semantic violation	*After the director of the school had resigned from the university, he worked for a meter.	院长从大学辞职后, 他去了一家米工作。 yuàn zhăng cóng dà xué cí zhí hòu, tā qù-le yī jiā mī göng zuò
Filler incorrect	*After the director of the school had resigned from the university, he will work for a multinational.	院长从大学辞职后, 他将要去一家跨国公司工作。
Filler correct	After the director of the school has resigned from the university, he will work for a multinational.	yuàn zhăng cóng dà xué cí zhí hòu, tã jiãng yào qù yī jiã kuà guó gōng sī gōng zuò
Filler correct	After the director of the school will have resigned from the university, he will work for a multinational.	
Filler semantic violation	*After the director of the school had resigned from the university, he will work for a meter.	院长从大学辞职后, 他将要去一家米工作。 yuàn zhāng cóng dà xué cí zhí hòu, tā jiāng yào qù yī jiā mī gông zuò

aspect -gu) and a temporal noun phrase (e.g., shàng gè yuè – 'last month'), as in the following sentence:

\* 上个月 联合国 将要 派出 特别 调查组。<sup>1</sup> Last month UN jiangyao Dispatch Special investigation unit "Last month, the United Nations will dispatch a special investigation unit".

[(From Qiu and Zhou, 2012, p. 94)]

To investigate whether cross-linguistic differences in tense marking can influence readers' perception of time, we tested fluent Mandarin-English bilinguals reading English sentences. According to the approach proposed by Reichenbach (1947), the timeline corresponding to a situation described by an utterance involves three time points: (a) Speech time (the time at which the utterance is produced), (b) reference time (the perspective from which a situation is perceived), and (c) event time (the time at which the event happens). In order to understand the temporal order of events in a given sentence, and therefore its overall meaning, it is necessary to encode on the one hand the relationship between Speech Time and Reference Time (theoretically encoded by tense), and on the

<sup>1</sup> Pin yin version of sentence: Shàng-gè yuè lián hé guó jiāng yào pài chũ tèbié diào chá zǔ. Time (theoretically encoded by aspect). In the case of a tensed language, the three time points and their relationships are coded directly by inflection (Smith, 2008). However, in the case of tenseless Mandarin Chinese, the relationship between Speech Time and Reference Time can remain unspecified because it is not encoded by an inflectional morpheme within the verb (Smith, 2008) and specifying temporal information is not compulsory (Smith, 1991).

We thus created complex English sentences featuring a reference time misalignment (RTM) between their adjunct and main clauses. In all cases, adjunct clauses began with the connective *'after*' and systematically described a first event with perfect aspect - a grammatical category that exists in both English and Chinese. In the RTM conditions (see Fig. 1B and C) the adjunct clause was in the present or the future tense, whereas the main clause was in the absolute past tense (simple past). Note that such RTM is different from tense violation, since the latter entails grammatically incorrect tense forms within a given clause, as in "Yesterday, I sail Diane's Boat to Boston" (from Steinhauer and Ullman, 2002). We also created a semantic violation condition in which the statement was made meaningless by the presence of an incongruent word ending designed to serve as a semantic control, to test participants' understanding of the materials presented (see Table 1).

In control (correct) sentences, the adjunct clause was in the past perfect tense (see Fig. 1A). It shared its reference time and speech time with the main clause in the simple past tense, and thus was

94

#### 2. Results

#### 2.1. Behavioural data

#### 2.1.1. Reference time misalignment

speech time and reference time with the main clause. We tested monolingual speakers of English as controls and Chinese-English speakers, who were fluent in both Mandarin Chinese and English, that is, able to hold a normal, fluent conversation in either of the two languages (Titone, 1972; Macnamara, 1967; Grosjean, 1989). We did not expect marked differences between fluent Chinese-English bilinguals and native speakers of English in a sentence acceptability task. However, we expected eventrelated brain potentials (ERPs) recorded simultaneously to index differences in online processing of temporal information. For native English speakers, we predicted that RTMs would elicit greater negativity in the N400 range compared with correct sentences (see Nieuwland, 2015; Liang et al., 2016), owing to semantic difficulties in reconciling the incongruous timelines presented in the adjunct and main clauses (Kutas and Federmeier, 2011: Kutas and Hillyard, 1984; Van Petten and Kutas, 1990). However, because tense encodes the relationship between reference time and speech time, and since Mandarin Chinese does not encode tense directly, we predicted that native Chinese speakers would be less sensitive to RTMs than their native English peers. This should translate into a relatively weaker N400 modulation in bilinguals and we thus

predicted a group x condition interaction in the N400 range.

correct according to the rule of temporal connectives (Hornstein,

1990). The RTM conditions were of two kinds: (1) a Present-Past Misalignment (PPM; Fig. 1B) and a Future-Past Misalignment

(FPM; Fig. 1C), in which the tense of adjunct clauses does not share

Collinearity was not an issue in the models used for analysis: Fixed-effects correlations (|r|) were <0.7 for all predictors. For both accuracy and reaction time (RT) models, the intercept represents the average likelihood that English participants were accurate in the control condition. Each coefficient compares the average for a different combination of fixed factor levels against this intercept, and *p* values are derived from the normal approximation method (Barr et al., 2013).

Both PPM and FPM yielded significantly lower accuracy compared with the past perfect control condition (Fig. 2 and Table 2, FE1 and FE2). There was no fixed effect of Native language (FE3), indicating that English and Chinese participants had similar performance for baseline control sentences. No interaction was found (11 and 12), suggesting that bilingual Chinese-English participants were similarly *inaccurate* in detecting either type of RTM.

As for RTs, English monolingual participants showed no processing time costs for RTM conditions compared with the intercept condition, and Chinese-English bilinguals did not differ from English monolinguals at baseline (Table 2, FE1, FE2, FE3). However, a significant interaction in the analysis showed that Chinese bilinguals had longer RTs in the PPM condition relative to the additive



Fig. 2. Mean accuracy and Reaction time of semantic judgement for control condition, Present-Past time clash and Future-past time clash conditions for both English native participants and Chinese-English bilingual participants. Error bars represent s.e.m.

Table 2

LMM analyses for Reference Time Misalignment behavioral data.

Model 1 (English at int.)	Accuracy		Reaction Time			Model 2 (Chinese at int.)	Accuracy		Reaction Time				
	b	SE	z	b	SE	t		b	SE	z	b	SE	t
Int. English/Control	4.59	0.51	8.99***	6.40	0.09	71.17	Int. Chinese/Control	3.91	0.38	10.41***	6.37	0.08	80.54***
FE1. English/PPM	-8.64	0.74	-11.65***	-0.02	0.04	-0.52	FE1. Chinese/PPM	-8.07	0.62	-12.92***	0.09	0.04	2.36*
FE2. English/FPM	-8.27	0.74	-11.12***	0.03	0.04	-0.59	FE2. Chinese/FPM	-7.33	0.59	-12.52***	0.03	0.04	0.9
FE3. Chinese/Control	-0.60	0.57	-1.05	-0.03	0.11	-0.27	FE3. English/Control	0.78	0.58	1.34	0.02	0.11	0.27
<ol><li>Chinese/PPM</li></ol>	0.46	0.86	0.54	0.11	0.06	2.00*	<ol> <li>English/PPM</li> </ol>	-0.71	0.86	-0.83	-0.11	0.06	-2*
I2. Chinese/FPM	0.84	0.88	0.96	0.06	0.06	1.04	<ol><li>English/FPM</li></ol>	-1.06	0.87	-1.22	-0.06	0.06	-1.04

p < 0.05; p < 0.01; m p < 0.001; Int. is the intercept (baseline) condition against which all other contrasts are compared. Fixed Effects FE1 & FE2 consider only the data from the group used for intercept, and examine whether performance differs between the specified conditions and the intercept. FE3 examines differences between groups in the control condition. Interaction effects I1 & I2 examine whether comparison group participants' performance on the PPM condition (I1) or FPM condition (I2) differs from their own performance in the control condition, and how the magnitude of this difference compares to that shown by the English group. Thus, whilst a significant interaction shows that both groups behaved similarly across conditions.

contribution to the model of their RTs in the control condition, and the RTs of the English natives in the control and PPM conditions (*I1*). No such interaction emerged for the FPM, however (*I2*).

#### 2.1.2. Semantic violation

Accuracy and reaction times (RT) were modeled as a function of native language (English, Mandarin Chinese) as between-group factor, and semantic violation (control, semantic violation) as within-participant factor. In all other respects, our models were implemented similarly to those described in the previous section: Intercept values comprised the average likelihood that English monolingual participants were accurate in the baseline control condition. For accuracy data (see Fig. 3), all participants performed at ceiling (>97% accuracy on average). Thus, no significant differences emerged in the model (Table 3,

Thus, no significant differences emerged in the model (Table 3, *FE1, FE2 and 11*). For RT data, English monolinguals took the same amount of time to respond to the control as the semantic violation sentences (*FE1*), and Chinese-English bilinguals did not differ from English controls at baseline (*FE2*). However, a significant interaction emerged, such that Chinese-English bilinguals were significantly slower responding to semantic violations than the additive contribution to the model of their own performance on correct trials and English monolinguals' performance on correct and semantic violation trials (*I1*).

#### 2.2. Electrophysiological data

#### 2.2.1. Reference time misalignment

ERP mean amplitudes were analysed with repeated-measures analyses of variance (ANOVA). Mean amplitudes were modelled as a function of native language (English, Mandarin Chinese) and condition (past perfect control, PPM and FPM). The N400 effect was examined after the onset of the critical verb (henceforth N4-1) and after the onset of the next word in line (henceforth N4-2).

2.2.1.1. N4-1. There was a significant main effect of native language (F(1, 44) = 7.35, p = 0.01,  $\eta_p^2 = 0.14$ ) on N4-1 mean amplitude, and a significant interaction between native language and condition (F(2, 88) = 4.84, p = 0.01,  $\eta_p^2 = 0.1$ ). The condition main effect was not significant (F(2, 88) = 0.97, p = 0.38,  $\eta_p^2 = 0.02$ ). In English controls, N400 negativity was significantly greater in the PPM than in the baseline condition (t(18) = 1.387, p = 0.09; one-tailed *t*-test). In Chinese-English bilinguals, however, the difference between PPM and baseline condition did not attain statistical significance (t(26) = -0.55; p = 0.29; one-tailed test; Fig. 4a; Supplementary Fig. 1). As for the FPM versus baseline comparison, native speakers of English had significantly more negative N400 amplitudes in response to FPM (t(18) = 2.637, p = 0.01; one-tailed test; Supplementary Fig. 2) but we found no such difference in



Fig. 3. Mean accuracy and Reaction time of semantic judgement for control condition and Semantic Violation condition for both English native participants and Chinese English bilingual participants. Error bars represent SE of the mean.

#### Table 3

LMM analyses for Semantic Violation behavioral data.

Model 1 (English at int.)	Accuracy		Reaction Time			Model 2 (Chinese at int.)	Accuracy		Reaction Time				
	b	SE	z	b	SE	t		b	SE	z	b	SE	t
Int. English/Control	4.45	0.52	8.52***	6.40	0.09	70.87***	Int. Chinese/Control	3.84	0.37	10.36***	6.37	0.08	80.35***
FE1. English/PSV	0.67	0.97	0.69	-0.03	0.06	-0.48	FE1. Chinese/PSV	1.11	0.73	1.53	0.20	0.06	3.18**
FE2. Chinese/Control	-0.63	0.59	-1.07	-0.03	0.11	-0.26	FE2. English/Control	0.63	0.59	1.06	0.03	0.11	0.27
I1. Chinese/PSV	0.65	1.14	0.57	0.23	0.09	2.63**	I1. English/PSV	-0.04	1.33	-0.03	-0.23	0.08	-2.71**

p < 0.01; p < 0.001; Int. is the intercept (baseline) condition against which all other contrasts are compared. FF1 consider only English participants' data, and examine whether performance differs on the PSV vs. Control condition (FF1). FF2 examines differences between Chinese and English participants on the Past Perfect control condition. Interaction effects: 11 examine whether Chinese participants' performance on the PSV condition (11) differs from their own performance on the control condition, and how the magnitude of this difference compares to that shown by the English group. Thus, whilst a significant interaction signals group differences as a function of condition, no significant interaction infers that both groups behave similarly across conditions.



Fig. 4. ERPs elicited by reference time alignment manipulations. (a) ERPs elicited by the critical word and post-critical word in the past perfect control condition (black lines) and the PPM condition (blue line); (b) ERPs elicited by the critical word and post-critical word in the past perfect control condition (black lines) and the FPM condition (red line). Left, English natives; Right, Chinese-English bilinguals. ERP graphs depict variations of a linear derivation of channels C1, C2, C2, CP1, CP2, and CP2.

184

Chinese-English bilinguals (t(26) = -1.62, p = 0.06; one-tailed test). In fact, the difference tended to go in the opposite direction in Chinese participants (Fig. 4b).

2.2.1.2. N4-2. There was a significant main effect of native language on N4-2 mean amplitudes (F(1,44) = 42.50, p < 0.001,  $\eta_p^2 = 0.25$ ), and a significant interaction between native language and

Y. Li et al./Brain Research 1701 (2018) 93-102



Fig. 5. ERPs elicited by the last word in the past perfect control condition (black line) and the semantic violation condition (yellow line). Left, English native speakers; Right, Chinese English bilinguals. ERP graphs depict variations of a linear derivation of channels C1, C2, C2, CP1, CP2, and CP2.

condition (*F*(2,88) = 3.48; *p* = 0.035,  $\eta_p^2$  = 0.073). However, the main effect of condition was not significant (*F*(2,88) = 2.66, *p* = 0.08,  $\eta_p^2$  = 0.06). In English native controls, N400 mean amplitudes elicited by the post-critical word (N4-2) differed statistically between PPM and baseline conditions (*t*(18) = 2.54, p = 0.01; one-tailed test) but this difference was not significant in the Chinese-English bilingual group (*t*(26) = -0.3, *p* = 0.38; one-tailed test; Fig. 4a; Supplementary Fig. 1). FPM and baseline conditions also differed significantly in native speakers of English (*t*(18) = 2.74, *p* < 0.01; one-tailed test; Supplementary Fig. 2) and again, this difference was not significant in the Chinese participants (*t*(26) = -0.29, *p* = 0.39; one-tailed test; Fig. 4b).

#### 2.2.2. Semantic violation

98

The N400 elicited by the sentence-final word was analysed using a repeated-measures ANOVA with native language (Mandarin Chinese, English) as between-group factor and semantic violation (control condition, semantic condition) as within-subject factor. Only the main effect of semantic violation proved statistically different (F(1,44) = 20.58, p < 0.001,  $\eta_p^2 = 0.32$ ; Fig. 5); there was no significant effect of native language (F(1,44) = 0.75, p = 0.75,  $\eta_p^2 = 0.002$ ), nor was there an interaction between native language and semantic violation (F(1,44) = 3.43, p = 0.07,  $\eta_p^2 = 0.07$ ). Both native speakers of English and Chinese-English bilinguals showed greater negativity in the N400 range for semantic violation tions as compared to control sentences (English: t(18) = 3.39, p = 0.002; one-tailed t-test; Mandarin Chinese: t(26) = 2.6, p = 0.008;

#### 3. Discussion

Here we examined whether linguistic differences between Mandarin Chinese and English in the domain of temporal encoding influences speakers' sensitivity to temporal sequence violations. Despite showing a lack of metacognitive awareness regarding the semantic acceptability of sentences featuring a reference time misalignment, native speakers of English produced a significant N400 modulation in response to the verb in critical position as well as the post-critical word. As expected, N400 mean amplitude was increased for verbs inducing an RTM as compared to verbs that did not conflict with the reference time of the adjunct clause. Critically, no such effect was found in fluent Chinese-English bilinguals with a good command of English grammar: Not only were they indistinguishable from their monolingual peers in terms of acceptability judgement, but in addition, there was no measurable shift in N400 amplitude in any of the conditions. Nevertheless, lexicalsemantic violations elicited a classical N400 modulation in both groups, suggesting good levels of overall sentence meaning integration.

Most previous research investigating tense processing in L2 learners, healthy L1 speakers, or language impaired L1 speakers have tended to use simple constructs in which a time adverbial and the tense form are incompatible (i.e., the tense violation paradigm; see Steinhauer and Ullman, 2002; White et al., 2012; Newman et al., 2007). These studies focused on ERP amplitude modulations of either the P600, the left anterior negativity (LAN), or a biphasic LAN-P600 complex. However, in the current study, we did not investigate tense violations occurring within a given clause, but rather reference time misalignment between two clauses, each of them being grammatically correct when considered independently. Tense, thus, had global rather than local temporal relevance in our study. It was therefore only when participants encountered the verb of the main clause that they were in a position to perceive a reference time misalignment, bearing in mind that the tense used in the main clause did not constitute a tense violation until they were able to recall the temporal

information encoded in the first clause. And indeed, an explorative analysis revealed no significant modulation of P600 mean amplitudes in either group of participants and for any experimental contrast.

We expected that native speakers of English would identify RTMs or -at the very least- that they would detect them more often than their Chinese-English bilingual peers. This is because an RTM produces a semantic incongruence at the level of the entire statement and results in a content that effectively does not 'make sense'. The absence of RTM detection in the behavioural data suggests that the information conveyed by tense can be subtle, especially when the misalignment depends on long-range integration of information across two clauses. This may be explained by automatic repair mechanisms in reading, especially in the context of this experiment in which we used word-by-word presentation and given that RTM differences are rather difficult to identify in general. Indeed, word-by-word presentation (Kaiser, Marinis, 2010; VanPatten, 2014) is very unnatural (even though it is often imperative in ERP studies of reading) and it is likely to tap into working memory more than natural reading, which may have contributed to blurring the events' timeline. Also, the task used in the experiment likely biased the participants to make basic semantic adequacy judgements because of the presence of a clearly aberrant word in the semantic violation condition. In a recent study by Nieuwland (2015), participants were required to either explicitly assess stimulus plausibility or simply read the same statements for comprehension. In both cases, participants displayed larger N400 amplitudes for stimuli which were inconsistent with real-world knowledge. In addition, our data is consistent with recent findings from the language comprehension literature, in which language processing is construed as "good enough" (characterized by underspecified grammatical representations). For the purposes of our offline task, participants may have been using a simple heuristic to interpret these sentences according to existing schemata; avoiding full linguistic computation since this was taskirrelevant (e.g. Ferreira and Karimi, 2015; Ferreira and Patson 2007). Thus, the subtle between-clause violations in FPM and PPM conditions may have escaped participants' initial scrutiny in terms of the degree to which these sentences "made sense".

Critically, however, English speakers did process the tense configuration of the matrix clauses as indicated by a modulation of the N400 elicited by the post-critical word following the locus of a reference time misalignment in the case of PPMs, and both the critical verb and the following word in the case of FPMs. We interpret this result as showing that the temporal representation of events was successfully extracted on the basis of tense information by native speakers of English, even though this did not translate into behavioural effects. Note that the RTM resulted in an N400 modulation as early as the critical verb for FPM but only at the post-critical word in the case of PPM. Even though we did not predict such a difference, we could have anticipated this on the basis of the magnitude of the misalignment. Indeed, an FPM is arguably more salient than a PPM, due to the time gap being wider. In addition, recall that it is a matter of debate whether or not the future form in English gualifies as tense, due to the mandatory use of the auxiliary 'will'. In other words, it could be that the auxiliary produced a strong expectation for a shift of the reference time into the future, leading to more salient incongruence than in the case of the PPM

It must be noted that although the reference time of the present perfect is the present, it is mostly used to describe events that have happened in the recent past, that is, in the pre-present zone (Declerck et al., 2006). We propose that this contributes to making the PPM condition relatively less contrastive than the FPM condition. In this case, one could reasonably expect N400 modulations to appear later for PPM than FPM conditions, an effect akin to a spill-over, which is commonly observed in eye-tracking studies (Kaiser, 2014; Keating, 2014; Reichle et al., 2007). Beyond the fact that spill-over effects have previously been identified in studies of tense violation in L1 (e.g. Qiu and Zhou, 2012), it is unsurprising that the N400 modulation elicited by RTM should carry over to the post-critical word because of the requirement for crossclause integration in order to retrieve the temporal relationship between the two events described.

One may wonder if the language proficiency of our Chinese-English bilingual participants was high enough to detect RTMs. First, the native speakers of Mandarin Chinese involved in this study performed with a very high level of accuracy in the semantic violation conditions, on a par with their English native peers, indicating that their level of comprehension was indeed excellent. Second, most of the bilingual participants obtained a high score at the IELTS, a standard test of English proficiency. Although 5 participants did not provide a score, their English proficiency level was expected to be high since they received instruction exclusively through the medium of English from high-school onwards. We also conducted a split-group analysis excluding the participants without a numerical score to test for potential differences in RTM sensitivity in relation to IELTS score and found that the latter failed to relate to the former (see Supplementary Analysis).

Importantly, all bilingual participants involved in this study reported having high English proficiency (See Methods). Based on an extensive review of the literature (e.g. LeBlanc and Painchaud, 1985; Palmer and Bachman, 1981; Rea, 1981; von Elek, 1981, 1982), Blanche and Merino (1989) concluded that self-reports provide "good or very good" measures of proficiency, and such measures are often used in ERP experiments involving bilingual participants since they are very quick to obtain (e.g. Dowens et al., 2011; Lehtonen, et al., 2012).

To further assess the role of proficiency in the results, we tested an additional group of 21 Chinese-English bilinguals closely matched in IELTS score with the participants tested here on an overt time alignment judgment task along with a new group of native English controls. This new group of Chinese-English bilingual performed similarly to their English native peers (see Supplementary Analyses). Therefore, we assume that low proficiency in English is not the reason why Chinese participants failed to detect PPMs and FPMs.

Note that Chinese-English bilinguals needed a longer time to judge whether PPM sentences were acceptable as compared to control ones. It may be that re-evaluation mechanisms taking place over the processing of the entire sentence were longer in this condition because the sentences were in fact perceived as 'strange', but this effect could simply stem from the ambiguity of the present perfect form itself: is it a past form or a present form? Independently of whether there is an RTM between clauses, the delay in the PPM condition would then be due to internal processing issues rather than RTM resolution.

It is thus likely that the lack of detection of RTMs in the bilingual participants relates in some way to cross-linguistic differences between Mandarin Chinese and English, and more particularly, to the way temporal information about events is conveyed by language. Although Mandarin Chinese, just like English, features the perfective aspect, it has no direct equivalent for tense. This means that Chinese-English bilinguals reading a perfect form in English will know that the particular event described in the adjunct clause is completed but will have difficulty figuring out when completion occurs: past, present, or post-present.

The relative inability of Chinese-English bilinguals to perceive RTMs in English may have implications well beyond the domain of second language sentence comprehension and indeed concern time conceptualisation more generally. Given that native speakers of Mandarin Chinese tend to culturally care about the past more than their Canadian counterparts (Ji et al., 2009), difficulties in identifying temporal relations in English may lead to significant misunderstandings in everyday language use. In other words, Chinese-English bilinguals may be expected to experience a blurred relationship between past and present when interacting in English, which would stand in stark contrast with their experience of the same relationship in their native language.

In conclusion, Chinese-English bilinguals do not register reference time misalignments between clauses when event time is encoded by tense in English, and probably over-rely on the adverbial form "after" to figure out temporal order, since the same is used in their native language to specify temporal sequencing. Despite such strategy, they fail to accurately position two events in relation to one another on the timeline, which becomes blurred as a result. Future studies will determine whether such effects remain when participants are directly instructed to process temporal sequences.

#### 4. Materials and methods

#### 4.1. Participants

Thirty Chinese-English bilinguals and 25 native speakers of English took part in this study. Data from three bilingual participants and 6 native speakers of English were discarded due to poor ERP data quality, such as heavy blinking and excessive alpha elicitation. Of the remaining 27 bilingual participants, 10 were males and 17 females, with a mean age of 22.3 (SD = 2.7) and were all righthanded. In the English native group, 19 participants dataset were kept (8 males and 11 females; *Mean* age = 22.4, SD = 9.3; one lefthander and 18 right-handers). All participants were students at Bangor University, UK, and received either payment or course credits for their participation.

Table 4

Chinese-English bilingual participants' language background.

Measure	Mean	SD
Age of L2 acquisition	6.9	3.2
Length of L2 learning (years)	15.2	3.5
Length of staying in an English speaking country (Months)	44.7	61.4
Daily Chinese usage (%)	50.7	20.4
Daily English usage (%)	49.3	20.4



Fig. 6. Chinese-English bilingual participants' self-estimation of their Chinese and English level (10 point- scale). Error bar represents stand error. The average age at which Chinese-English bilinguals started to learn English was 6.9 years (SD = 3.2), and all were living in the UK at the time of testing. The average IELTS score for the bilingual group was 6.9 (SD = 0.5, from 6.5 to 8). Five further bilinguals did not provide IELTS scores, since they received English medium instruction since high school. Table 4 summarizes the Chinese-English bilinguals language background.

Bilingual participants also self-reported their proficiency in both English and Mandarin Chinese (see Fig. 6). All participants had normal or corrected to normal vision. The study was approved by the School of Psychology, Bangor University ethics committee.

#### 4.2. Stimuli

The materials consisted of 70 sentence sets, each containing 8 sentences. Four were experimental sentences featuring either a (i) correctly tensed verb, (ii) PPM, (iii) FPM, or (iv) semantic violation (see Table 1) and 4 sentences served as fillers. The locus of the reference time misalignment coincided always with the second word of the main clause. For the main analyses, we compared the control condition (i) with the two RTM conditions (ii) and (iii). An additional analysis comprised (i) and (iv), in order to ascertain that the Chinese-English bilinguals comprehended the overall meaning of the sentences.

In order to dilute the critical experimental manipulations, filler sentences were included, in which the matrix sentences used the simple future tense. There were two presentation lists, which alternated so as to present experimental items and fillers in a fully counterbalanced fashion. Each presentation list featured 4 blocks and a given sentence from a given condition was only presented together in the same block. In addition, verb regularity was systematically manipulated such that half were regular and the other half irregular. There was no significant difference in lexical frequency between regular and irregular lists even though there was a trend for irregular vers to be more frequent (U = 451.5; p = 0.06).

## 4.3. Procedure

Bilingual participants first filled out a language background questionnaire. All participants were seated 100 cm away from a 19-inch computer monitor and responded by pressing buttons on a reaction time box. The first clause of each sentence was presented at once and participants were instructed to press any button when they had finished reading. The rest of the sentence then comprised individually presented words, in the centre of the screen, for a duration of 300 ms (ISI 400 ms), in order to minimise eye movements. Once the whole sentence had been read, participants were required to judge whether or not it made sense (see Fig. 7).

#### 4.4. Design and behavioural data analysis

In this experiment, we compared two groups (English native speakers, Chinese-English bilinguals) and, within-subject, three reference time alignment conditions (correct, PPM, FPM). In addition, participants' understanding of the sentences was assessed by analysing effects of semantic violations in sentence completions (final word). Accuracy and reaction times (RT) were modeled as a function of one between-groups factor: Native language (English, Mandarin Chinese), and one within-subject factor: RTM (correct, PPM, FPM). Accuracy was analyzed using a binomial logistic regression. Reaction time data were log transformed and analyzed based on linear mixed effects modeling using R (R Development Core Team, 2008) and the lime4 library (Bates et al., 2008).  $\beta$ -values are reported and tested at p < 0.05. As recommended by



Fig. 7. Structure of an experimental trial.

Barr et al. (2013), we modeled the maximal random effect values of participant and items intercepts and slopes across groups and condition in both models (when models successfully converged).

#### 4.5. ERP recording and analysis

Electrophysiological data were recorded at a rate of 1 kHz from 64 Ag/AgCl electrodes according to the extended 10-20 convention using an online (0.05-200 Hz) bandpass filter. Two additional electrodes were used to monitor eye movements, one below and one above the right eye. Electrode Cz was the reference electrode and impedances were kept below 5 k $\Omega$ . EEG data was filtered bandpass using zero-phrase shift digital filtering (0.1 Hz, 24 dB/oct- 20 Hz, 48 dB/oct). Periods of EEG instability corresponding to experiment pauses were removed manually as well as major artefacts through visual inspection of the data and then we adopted the procedure proposed by Gratton et al. (1983) to mathematically correct eyeblink artefacts. ERPs were computed from epochs ranging from -200 ms to 1500 ms after the onset of critical word, always in second position within the main clause. For the semantic violation condition, epochs ranged from -200 ms to 1500 ms, so as to coincide with onset of the sentence-final word. Epochs with any activity exceeding  $\pm\,100~\mu V$  at any electrode site except electroocculogram channels were eliminated. More than 30 trials in each participant and condition were included in the averaging procedure. Baseline correction was performed in reference to pre-stimulus activity and individual averages were digitally re-referenced to the global average reference. All analyses were conducted again using the average of the mastoid electrodes as reference and all effects reported based on the global average reference were qualitatively replicated in this analysis.

101

For RTM analyses, we measured ERP amplitudes over 6 centroparietal electrodes, CP1, CPz, CP2, Cz, C1, C2 at which the N400 is usually maximal (Kutas and Hillyard, 1980a,b; Kutas and Hillvard, 1984). We identified two time-windows for analysis, the usual N400 time-window between 350 and 500 ms after the onset of the critical word (the verb in the main clause: e.g., worked) and a window between 1200 and 1350 ms corresponding to the N400 window of the post-critical word. For semantic violation analyses, N400 modulations were analysed between 350 and 500 ms after the onset of the final word.

#### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.brainres.2018.07. 008.

#### References

- Athanasopoulos, P., 2009. Cognitive representation of colour in bilinguals: the case
- Athanasopoulos, P., 2009. Cognitive representation of colour in bilinguals: the case of Greek blues. Bilingualism-Lang. Cognit. 12, 83–95.
  Athanasopoulos, P., Bylund, E., 2013. Does grammatical aspect affect motion event cognition? a cross-linguistic comparison of english and swedish speakers. Cognit. Sci. 37, 286–309.
  Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68, 255–278. confirmatory hypothesis testing: keep it maximal. J. Mem. Lang. 68, 255–278. Bates, D., Maechler, M., & Dai, B. (2008). Ime4: Linear mixed effects models using s4
- classes

- Blanche, P., Merino, B.J., 1989. Self-assessment of foreign-language skills implications for teachers and researchers. Lang. Learn. 39, 313–340.
  Boroditsky, L., 2001. Does language shape thought?: mandarin and English speakers' conceptions of time. Cogn. Psychol. 43, 1–22.
  Boutonnet, B., Athanasopoulos, P., Thierry, G., 2012. Unconscious effects of grammatical gender during object categorisation. Brain Res. 1479, 72–79.
  Chei S. Brauerster, M. 2001. Internet and conception and the processing in the process
- Choi, S., Bowerman, M., 1991, Learning to express motion events in English and Korean: the influence of language-specific lexicalization patterns. Cognition 41, 83-121

- 83-121. Comrie, B., 1985. Tense. Cambridge University Press, Cambridge. Cubelli, R., Paolieri, D., Lotto, L., Job, R., 2011. The effect of grammatical gender on object categorization. J. Exp. Psychol. Learn. Mem. Cognit. 37, 449-460. 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 Counter Bealin. Gruyter, Berlin
- Gruyter, Berni, Dowens, M.G., Guo, T.M., Guo, J.J., Barber, H., Carreiras, M., 2011. Gender and number processing in Chinese learners of Spanish evidence from event related potentials. Neuropsychologia 49, 1651–1659.
  Duff, P., Li, D., 2002. The acquisition and use of perfective aspect in Mandarin. In: S.
- Duri, F., Li, D., 2002. The acquisition and use of perfective aspect in Mandarin. In: S. R. SY, (Eds.), The L2 acquisition of tense-aspect morphology. John Benjamins, Philadelphia, PA, pp. 417–453.
   Ellis, C., Kuipers, J.R., Thierry, G., Lovett, V., Turnbull, O., Jones, M.W., 2015. Language and culture modulate online semantic processing. Social Cognit. Affect. Neurosci. 10, 1392–1396.
   Ferreira, F., Karimi, H., 2015. Prosody, performance, and cognitive skill: Evidence for micidated difference in C.G. (2010). Device morphology.

- Perietia, F., Natimi, F., 2015. Frusody, periormatice, and cognitive skin: Evidence from individual differences. In: Gibson, I.F.E. (Ed.), Explicit and implicit prosody in sentence processing: Studies in honor of Janet Dean Fodor. Springer International Publishing, Switzerland, pp. 119–132.
  Ferreira, F., Patson, N.D., 2007. The "good enough" approach to language comprehension. Lang. Linguist. Compass 1, 71–83.
  Flecken, M., 2011. What native speaker judgments tell us about the grammaticalization of a progressive aspectual marker in Dutch. Linguistics for final statistics. 49, 479-524.

- 49, 479–524.
   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.
   Flecken, M., von Stutterheim, C., Carroll, M., 2014. Grammatical aspect influences motion event perception: evidence from a cross-linguistic non-verbal recognition task. Lang. Cognit. 6, 45–78.
   Gentner, D., Goldin-Meadow, S., 2003. Whither Whorf. In: G.D, G.-M. S (Eds.), Language in mid: Advances in the study of language and thought. MIT Press, Cambridge, MA, pp. 3–14.
   Gratton, G., Coles, M.G.H., Donchin, E., 1983. A new method for off-line removal of ocular atficiat. Electroencephalogr. Clin. Neurophysiol. 55, 468–484.
- ocular atficat. Electroencephalogr. Clin. Neurophysiol. 55, 468-484. Grosjean, F., 1989. Neurolinguists, beware! The bilingual is not two monolinguals in
- one person. Brain Lang. 36, 3-15.
- one person. Brain Lang. 36, 3–15. Hornstein, N., 1990. As Time Goes By: Tense and. Universal Grammar: MIT Press. Ji, L.J., Guo, T.Y., Zhang, Z.Y., Messervey, D., 2009. Looking into the past: cultural differences in perception and representation of past information. J. Pers. Soc. Psychol. 96, 761–769. Kaiser, E., 2014. Experimental paradigms in psycholinguistics. In: Podesva, R.J., Sharma, D. (Eds.), Research methods in linguistics. Cambridge University Press, Nature 1, 2014. Statemethol Science 2014.
- pp. 135-168
- pp. 135–168. Keating, G.D., 2014. Eye-tracking with text. In: Jegerski, J., VanPatten, B. (Eds.), Research methods in second language psycholinguistics. Taylor & Francis, London, pp. 69–92. Kutas, M., Federmeier, K.D., 2011. Thirty years and counting: finding meaning in the
- n400 component of the event-related brain potential (ERP), Annu, Rev. Psychol. 62, 621-647.

- 62, 621–647.
  Kutas, M., Hillyard, S.A., 1980a. Reading between the lines event-related brain potentials during natural sentence processing. Brain Lang. 11, 354–373.
  Kutas, M., Hillyard, S.A., 1980b. Reading senseless sentences brain potentials reflect semantic incompruity. Science 207, 203–205.
  Kutas, M., Hillyard, S.A., 1984. Event-related brain potentials (ERPs) elicited by novel stimuli during sentence processing. Ann. N. Y. Acad. Sci. 425, 236–241.
  Lantz, D., Steffler, V., 1964. Language and cognition revisited. J. Abnorm. Soc. Psychol. 69, 472–481.
  LeBlanc, R., Painchaud, G., 1985. Self-assessment as a second language placement instrument. TESOL Quarterly 19, 673–687.

- Lehtonen, M., Hulten, A., Rodriguez-Fornells, A., Cunillera, T., Tuomainen, J., Laine,
- Lehtonen, M., Hulten, A., Rodriguez-Fornelis, A., Cunillera, T., Tuomainen, J., Laine, M., 2012. Differences in word recognition between early bilinguals and monolinguals: Behavioral and ERP evidence. Neuropsychologia 50, 1362–1371. Li, C.N., Thompson, S.A., 1989. Mandarin Chinese: A functional reference grammar. University of California Press, Berkeley.
  Liang, X., Xiao, F., Wu, L., Chen, Q., Lei, Y., Li, H., 2016. The temporal order of word presentation modulates the amplitudes of P2 and N400 during recognition of presentatione. Durabel 3, 1800.
- causal relations. Front. Psychol. 7, 1890.
   y, J., 1992. Language Diversity and Thought. Cambridge University Press,
- Lucy, Ca
- Lucy, J., 1992. Language Diversity and Thought. Cambridge University Press, Cambridge.
   Lupyan, G., Ward, E.J., 2013. Language can boost otherwise unseen objects into visual awareness. PNAS 110, 14196–14201.
   Macnamara, J., 1967. The bilingual's linguistic performance. J. Soc. Issues 23, 58–77.
   Marinis, T., 2010. On-line sentence processing methods in typical and atypical populations. In: Unsworth, S., Blom, E. (Eds.), Experimental Methods in Language Acquisition Research. John Benjamins, pp. 139–162.
   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, 435–445.
- 435-445
- 435–445.
   Nieuwland, M. 2015. The truth before and after: brain potentials reveal automatic activation of event-knowledge during sentence comprehension. J. Cog. Neurosci. 27, 2215–2228.
   Palmer, A.S., Bachman, L.F., 1981. Basic concerns in test validation. In: Hughe, J.C.A. A. (Ed.), ELTdocuments 111: Issues in language testing. London, United Viscotar The Device.
- Kingdom, The British. Phillips, W., & Boroditsky, L. (2003), Can guirks of grammar affect the way you
- https://www.ebiodowsy.ic/2003/can/quinso/grammaa anct: the way for think? Grammatical gender and object concepts. Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society, Pts 1 and 2, 928-933. , /.C., Zhou, X.L., 2012. Processing temporal agreement in a tenseless language: an ERP study of Mandarin Chinese. Brain Res. 1446, 91–108. Oi
- Quirk, R., Greenbaum, S., Leech, G., Svartvik, J., 1985. A Comprehensive Grammar of
- the English Language. Longman, London. R Development Core Team, 2008, R: A language and environment for statistical

- R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. URL http://www.R-project.org.
   Rea, P.M., 1981. Formative assessment of student performance: the role of self-appraisal. Indian J. Appl. Linguist. 7, 66-88.
   Reichenbach, H., 1947. Elements of Symbolic Logic. Macmillan & Co, New York.
   Reichle, E.D., Pollatsek, A., Rayner, K., 2007. Modeling the effects of lexical ambiguity on eye movements during reading. In: van Gompel, R. (Ed.), Eye Movements: A Window on Mind and Brain. Elsevier Science, Oxford, pp. 271-202 292 th, C., 1991. The Parameter of Aspect. Kluwer Academic Publishers, Dordrecht.
- Simuti, C., 1991. Ine rarameter of Aspect. Kluwer Academic Publishers, Dordrecht, Smith, C., 2008. Time with and without tense. In: Scieron, J., Lacarme, J. (Eds.), Time and Modality. Springer, Berlin, pp. 227–249.
  Steinhauer, K., Ullman, M.T., 2002. Consecutive ERP effects of morpho-phonology and morpho-syntax. Brain Lang, 83, 62–65.
  Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., Kujpers, J.-R., 2009. Unconscious effects of language-specific terminology on preattentive color perception. PNAS 106, 4567–4570.

- perception. PNAS 106, 4567-4570. Titone, R., (1972). Bilinguismo pecoce ed educazione bilingue. Amando (Ed.) Roma. VanPatten, B. (2014). The psycholinguistics of SLA. In B. VanPatten & J. Williams (Eds.). Theories in second language acquisition: An introduction: Routledge. Van Petten, C., Kutas, M., 1990. Interactions between sentence context and word-frequency in event-related brain potentials. Memory Cognition 18, 380-393. von Elek, T., 1981. Self-assessment ofSwedish as a second language (Sja1Vbedominig av fardigheter i svenska som andra sprak). University of Goteborg, Language Teaching Research Center, Gijteborg, Sweden. wan Elek, T. 1982. Tea for Swedish as a second in relf.
- von Elek, T., 1982. Test of Swedish as a second language: An experiment in self-assessment. University of Goteborg, Language Teaching Research Center,
- assessment. University of Goteborg, Language Teaching Research Center, Goteborg, Sweden. von Stutterheim, C., Carroll, M., 2006. The impact of grammatical temporal categories on ultimate attainment in L2 learning. In: Byrnes, H., Weger-Guntharp, H., Sprang, K. (Eds.), Educating for Advanced Foreign Language Capacities. GUP, Georgetown, pp. 40–53. White, E.J., Genesee, F., Steinhauer, K., 2012. Brain Responses before and after
- White, E.J., Genesee, F., Stehmauer, K., 2012. Dath Responses before and after Intensive Second Language Learning: proficiency based changes and first language background effects in adult learners. PLoS One 7.
  Whorf, B., 1956. Language, Thought, and Reality. MIT Press, Cambridge, MA.

 $\infty$ 

102