The Comprehension and Production of L2 Tense-Aspect by Chinese and Arabic Learners of English: Online vs. Offline Performance

Yu Liu

PhD

University of York

Education

January 2024

Abstract

The acquisition of tense-aspect has been extensively investigated in the L2 research field. A consistent finding across these previous studies is that L2 learners often exhibit variability and experience fossilization in their usage of L2 tense-aspect. Despite this issue being widely explored, previous research has predominantly relied on learners' production and metalinguistic judgement data. While there has also been a growing number of processing studies in L2 tense-aspect research, the quantity of such studies still remains small. In addition, mixed evidence has emerged regarding the role of L1 in the processing and acquisition.

Aiming to address this potential gap, the present study investigates the online processing (eye-tracking and sentence-matching task), production (elicited imitation task), and explicit knowledge (comprehension: judgements task; production: gap-filling task) of tense-aspect structures among L2 English learners, focusing on 24 Chinese and 24 Arabic learners. It employs an experimental design with five different tasks crossing both comprehension/production and online/offline domains. The results suggest that both L2 groups could achieve native-like performance for present progressive items, which can be attributed to a positive L1 transfer effect. In contrast, the L1 Chinese group consistently underperformed compared to their L1 Arabic counterparts for past simple items, and both L2 groups exhibited non-native-like performance for present perfect items, attributable to negative L1 influences.

Overall, the results indicate a strong influence of the L1 on learners' performance in both comprehension and production; however, such influence is particularly evident in the online tasks. It is argued that L1 transfer is more likely to impact implicit knowledge than explicit knowledge. For future research that examines L2 acquisition, integrating both online

and offline tasks within a single study design is necessary to better assess learners' acquisition.

Table of Contents

List of Tables

List of Figures

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Professor Leah Roberts, for her valuable support and guidance throughout the course of my research. Thanks to her expertise and encouragement, I have acquired skills in using eye-tracker, designing experiments, and conducting quantitative data analysis. The continuous learning experience has made my PhD journey exceptionally fulfilling.

I also want to extend my sincere gratitude to my TAP member, Dr. Heather Marsden, for her invaluable feedback on the study design and my writing. In addition, I am grateful to my external examiner, Dr. Sarah Ann Liszka, and my internal examiner, Dr. Jelena O'Reilly, for their thorough review and insightful comments on the thesis.

A heartfelt thank you goes out to the participants who generously dedicated their time and effort to take part in this study. I appreciate their patience during the experimental sessions, as I am aware they took a considerable amount of time to complete.

I am thankful for my boyfriend, Cameron Malekyazdi Aitken, and all my other friends who have stood by my side throughout this PhD journey, and I deeply appreciate their support and friendships.

I extend my deepest gratitude to my parents, 刘映贵和郑秀英, whose support has been the bedrock of my PhD journey. Their financial and emotional support have paved the way for my education, and I am profoundly thankful for the opportunities they have provided.

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

Chapter 1 Introduction

1.1 The Research Background

In contrast to the generally successful process of acquiring a first language (L1), the acquisition of a second language (L2) frequently leads to differing levels of proficiency. Specifically, L2 learners encounter persistent challenges in acquiring morphosyntax and grammatical structures (DeKeyser, 2005; N. C. Ellis, 2022). The acquisition of grammatical morphology is found to be a slow, incremental, and often incomplete process (N. C. Ellis, 2022). For example, the Bottleneck Hypothesis, proposed by Slabakova (2009, 2019), suggests that functional morphology serves as a ‘bottleneck’ in L2 acquisition as it bundles a range of morphosyntactic and semantic features. The difficulty in acquiring these features acts as a constraining factor, influencing the pace of acquisition. Therefore, a considerable amount of L2 research has investigated morphosyntax phenomena (see a review in Ionin, 2012). Among these, the acquisition of L2 tense-aspect has received a lot of attention since the early 2000s (see a review in Ayoun & Rothman, 2012).

The current study focuses on the acquisition and processing of L2 tense-aspect. A consistent finding across previous studies is that L2 learners often exhibit variability, particularly in their oral production of verbal inflections when acquiring L2 tense-aspect in a variety of L2 languages, such as English, German, Spanish, French, Japanese, and Chinese (e.g., Domínguez et al., 2017; Duff & Li, 2002; Goad et al., 2003; Hawkins & Liszka, 2003; Labeau, 2005; Lardiere, 1998a, 1998b; Liszka, 2009; Prévost & White, 2000; Shirai & Kurono, 1998; Vainikka & Young-Scholten, 1996; White, 2003). This variability is not limited to developmental stages but may persist even among learners who have reached a stage where further development is unlikely. The latter phenomenon is commonly referred to as ‘fossilisation’, where L2 learners reach a stable point in their grammatical development without achieving native-like proficiency (White, 2003). For instance, in a longitudinal study by Lardiere (1998a, 1998b), Patty, an L2 English learner with Chinese as her L1 and a ten-year immersive experience in the U.S., produced past tense in only 34% of obligatory contexts at three different time points. While it may initially seem that all L2 learners, regardless of their L1, experience such variability, some studies indicate that the learner's L1 may influence the persistence of this variability. For example, in a study similar to Lardiere's (1998a, 1998b), White (2003) observed that an L1 Turkish L2 English learner produced past tense in around 80% of appropriate contexts, a much higher rate in comparison to the L1 Chinese speaker Patty. In addition, Lardiere (1998a, 1998b) found high rates of verbal inflection omissions in Patty's speech, while White (2003) observed significantly fewer omissions in her Turkish-speaking informant. These differences, potentially influenced by their respective L1 languages, suggest a role for the learner's L1 in L2 tense-aspect acquisition.

The variability observed among L2 learners has also sparked debates regarding the nature of their mental representations. Disagreements revolve around the relationship between learners' performance and the underlying representation in morphosyntax, leading to two contrasting positions: "morphology-before-syntax" and "syntax-before-morphology" (White, 2003b, pp. 183–187). Simply put, the former suggests that the acquisition of morphology precedes and drives the acquisition of syntax, while the latter argues that underlying syntactic representations emerge earlier and drive the acquisition of morphology. In empirical studies, those adopting the first position contend that the inconsistent use of inflectional morphology reflects a non-target-like mental representation. In contrast, studies adopting the second position suggest that inconsistent use is not solely caused by a non-target-like mental representation but by various other factors, such as a surface mapping failure. Both positions have prompted the formulation of various hypotheses in the field.

Although much research has been undertaken on the topic, previous empirical studies have predominantly relied on production tasks and other traditional SLA methods (e.g., judgement tasks). With the growing popularity of psycholinguistic techniques, it is argued by Clahsen (2007) that L2 processing data might be able to help resolve the theoretical controversies by exploring how grammatical structures are processed and constructed in real-time during comprehension and production. Successful processing relies on the presence of relevant underlying grammatical knowledge, and conversely, to acquire grammatical knowledge successfully, L2 learners must have access to appropriate processing mechanisms (Clahsen & Felser, 2006).

The advantages of psycholinguistic methods have led to a shift in L2 research, moving away from relying solely on learners' production and metalinguistic judgement data to incorporating real-time processing data. In fact, this shift has revealed some interesting findings. For instance, L2 learners may display metalinguistic/explicit knowledge in a judgement task, but they may still struggle to apply it during real-time processing (e.g., Roberts & Liszka, 2013). Some learners might show native-like processing, yet persistent difficulties are still reported in their production (e.g., Trenkic et al., 2014). In L2 tense-aspect research, there has also been a growing number of processing studies (e.g., Chan, 2012; Kahoul, 2014; Roberts & Liszka, 2013, 2021; Vogel, 2017; Yao & Chen, 2017); however, the quantity of such studies still remains small (Rastelli, 2020). Within this limited body of processing studies, mixed evidence has been suggested regarding the influence of the L1 in the acquisition process. For example, Chan (2012) found that Chinese learners of English, regardless of proficiency, were not sensitive to past tense violations in online processing. In contrast, their Korean and German counterparts demonstrated sensitivity comparable to the native English group. In contrast, Yao and Chen (2017) observed that advanced Chinese learners displayed sensitivity to past tense violations, and they argued that Chinese learners can develop a target-like mental representation of past tense even in the absence of grammatical tense in their L1. Thus, mixed findings have been reported.

To better examine this issue, this study utilises both processing and production data in the same participants with the aim to understand L2 representations. To further explore the role of L1 in L2 tense-aspect acquisition, the study includes both Chinese and Arabic learners of L2 English. The reason for incorporating learners from different L1 backgrounds is that Chinese and Arabic differ from each other and also differ from L2 English in how they grammatically express tense and aspect. The contrast can better tease apart L1 influences from general L2 learning processes. In terms of experimental design, the study employs visual-world eye-tracking and sentence-matching tasks to measure online comprehension and processing. The difference between these two tasks is that the former examines learners' facilitative processing of grammatical sentences, whereas the latter investigates learners' sensitivity to violations in ungrammatical sentences. For oral production, the study utilises an elicited imitation task. Both online processing and oral production tasks aim to measure learners' implicit knowledge. In addition, the study also employs an acceptability judgment task and a gap-filling task to measure learners' explicit comprehension and production knowledge. The rationale for including both implicit and explicit tasks is that learners draw on different types of knowledge when engaging in different tasks (R. Ellis, 2005). It is hoped that examining these task effects can provide us with a better understanding of whether L2 learners can gain either, both, or none of the two types of knowledge.

1.2 Outline of the Thesis

The present thesis consists of nine chapters. Chapter 2 introduces concepts related to tense and aspect and explores variations across languages in the extent to which they grammaticalise these features. The chapter also includes a discussion of perfect aspect and its connection to tense. Chapter 3 presents a discussion of the cross-linguistic variations of grammatical tense and aspect in the three languages under current investigation: English, Chinese, and Arabic. The focus is on how past tense, progressive, and perfect meanings are encoded in these languages. By highlighting the differences, the chapter aims to make predictions regarding how Arabic and Chinese learners' L1 might influence their acquisition of L2 English tense-aspect structures.

Chapter 4 provides a review of L2 research on tense-aspect acquisition. It examines two generally adopted approaches in L2 tense-aspect research and critically reviews relevant empirical studies in the field. The focus is given to the observation of varying degrees of L1 transfer effects. Building on conflicting evidence, the chapter proceeds to discuss the two main theoretical accounts for such discrepancies. With the growing popularity of L2 processing methods, the chapter also reviews processing studies on L2 tense-aspect research. It then addresses the limitations of previous studies, identifies potential research gaps, and introduces the research questions and predictions.

Chapter 5 discusses the methodology employed in the present study. Firstly, it provides information regarding the ethical considerations and the background of the participants. The comprehension and production tasks are introduced individually, with detailed discussions on methodological considerations, task rationale, stimuli design, task procedure, and data cleaning/analysis for each task. At the end of the chapter, some relevant changes made to the tasks are described based on feedback obtained during the pilot study.

Chapter 6 reports the results for each individual task. The results for each examined tense-aspect structure are provided separately. In Chapter 7, the obtained results are discussed in response to each research question, considering their relation to previous studies.

Chapter 8 presents a general discussion of the findings. It begins the discussion by focusing on cross-linguistic influences, addressing both positive and negative L1 effects.

Then, the discussion moves on to the potential influence of L1 on learners' implicit and explicit knowledge, and the developmental relationship between comprehension and production. The chapter concludes by discussing the methodological and theoretical implications of the present study. The contributions to both L2 tense-aspect research and L2 processing research are put forward.

Concluding the thesis, Chapter 9 offers a summary of the study and the overall findings. In addition, it also identifies several possible limitations of the present study and proposes directions for future research.

Chapter 2 Tense and Aspect

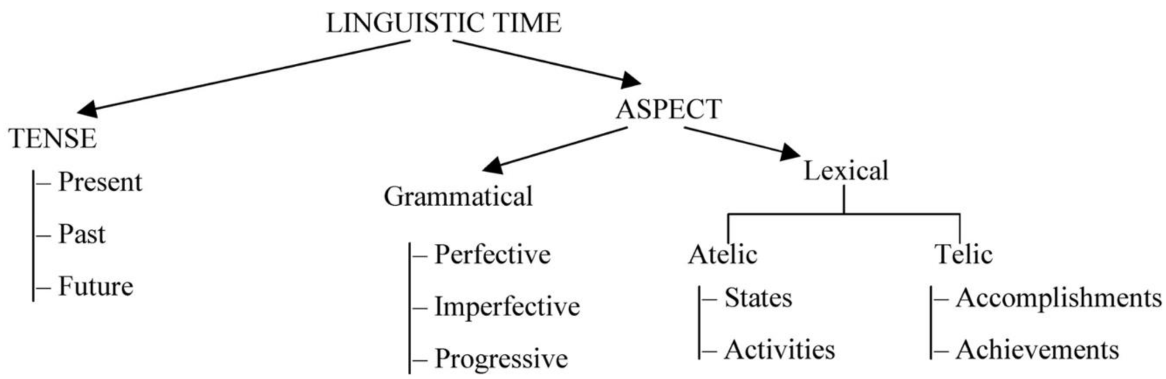
The present study focuses on the acquisition of L2 tense-aspect by adult L1 Chinese and L1 Arabic learners. This chapter aims to discuss the concepts of tense and aspect, and how they are conceptualised by different researchers in the literature.

Generally speaking, tense and aspect are two different linguistic categories related to time (Hewson, 2012). Tense is deictic, and it "relates the time of the situation referred to some other time, usually to the moment of speaking" (Comrie, 1976, pp. 1–2). In other words, it orders events sequentially on a timeline, essentially placing them in relation to other events (Ayoun & Salaberry, 2008). The most common tenses in languages are present, past, and future (Comrie, 1985). When an event is described in relation to the moment of speech, the present tense, for instance, aligns it with the very moment of speaking, and the past tense places it in a timeframe before the moment of speaking, whereas the future tense positions the event to occur after the moment of speaking (Lyons, 1968).

Aspect refers to "different ways of viewing the internal temporal constituency of a situation" (Comrie, 1976, p. 3) or "different perspectives which a speaker can take and express with regard to the temporal course of some event, action, process, etc." (Klein, 1994, p. 16). Aspect is encoded either lexically by the inherent semantic meaning of the verbs (known as lexical aspect) or morphologically through the verbal inflections and periphrastic constructions (known as grammatical or viewpoint aspect) (Salaberry & Shirai, 2002). Lexical aspect differentiates between atelic and telic verbs. For example, atelic verbs (e.g., *run, walk, want*) lack inherent endpoints within the event they describe, whereas telic verbs (e.g., *recognise a person, win the competition*) inherently signify a specific endpoint or outcome (C. S. Smith, 1997; Vendler, 1967). On the other hand, grammatical aspect distinguishes between the perfective and imperfective, and/or the progressive and non-progressive perspective. The perfective viewpoint refers to a bounded situation with endpoints, whereas the imperfective aspect presents a situation with no stress on its initial or final boundaries (Comrie, 1976). The progressive perspective, encoding part of the imperfective meaning, views an event in its progress (Comrie, 1976). Grammatical aspect varies significantly across languages, while there is less variation in terms of lexical aspect (P. Li & Shirai, 2000). The categorisation of tense and aspect is presented in Figure 2.1.

Figure 2.1

Categorisation of Tense and Aspect



Note. The figure is from Ayoun & Salaberry (2008, p. 558).

The following sections will present a discussion of the concepts related to tense and aspect before examining how tense and aspectual relations are encoded in the three languages: English, Chinese, and Arabic.

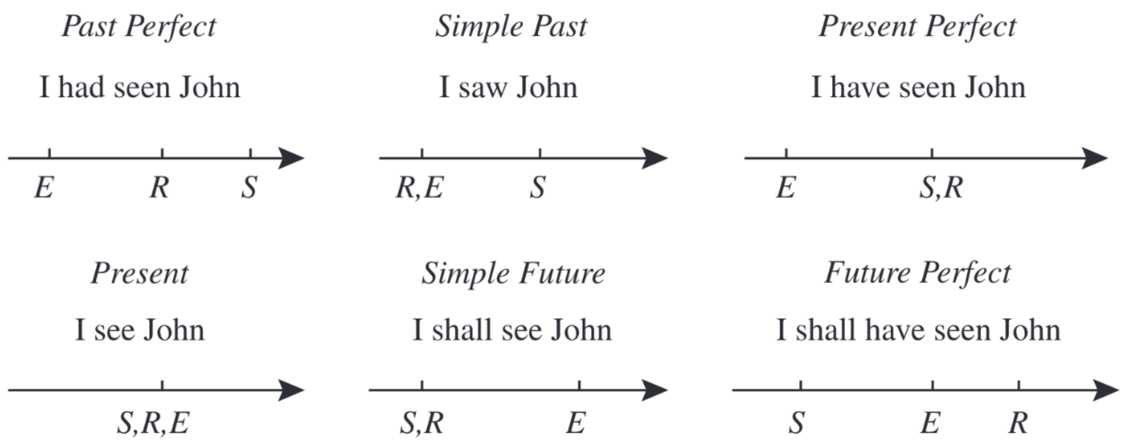
2.1 Tense

Time itself lacks inherent landmarks to locate a situation, so it is necessary to establish an arbitrary reference point which can be used to locate the situation on the timeline (Comrie, 1985). Reichenbach (1947) presented a framework in which the description of tenses involves three elements: point of speech (S), point of the event (E), and the point of reference

(R). For the three basic tenses, E aligns with R, and R's position is determined in relation to S, occurring either before, simultaneously with, or after S. Figure 2.1.1 shows how the three time points make distinctions for different tenses, and the direction of time is symbolised by the line moving from the left side to the right.

Figure 2.1.1

Reichenbach's Analysis of Tenses



Note. E = the point of event; S = the point of speech; R = the point of reference. The figure is from Reichenbach (1947, p. 72).

However, Reichenbach's analysis has been criticised for various reasons. For example, it has been found that the three-way relation can produce a greater number of possibilities than that which is typically observed in human languages (Declerck, 1986). The analysis is argued to be overly simplistic by having only one single reference point (Prior, 1967). To account for tenses more complex than the past perfect or future perfect, such as in the sentence *"we would have caught the train if we had left earlier"* (conditional perfect), at least two reference points are required. Also, Comrie (1985) argued that not every tense marker inherently includes a reference point. Instead, he proposed a clear differentiation between absolute and relative tense. For instance, absolute tense, such as the simple past, present, and future in English, directly indicates the temporal relationship between the point of event (E) and the point of speech (S), without the need for involving a reference time point (R) (Comrie, 1985), as shown in Table 2.1. In contrast, relative tense indicates how the point of event (E) relates to the point of reference (R) (Comrie, 1985). It relies on identifying a reference point that aligns with the context provided, and the present moment can also serve as a reference point for relative tenses, unless the context dictates otherwise (Comrie, 1985). The absolute-relative tenses are assumed to denote the situation when both R and S are involved, which is very common across the languages of the world, and examples are past perfect and future perfect in English, as shown in Table 2.1.1 (Comrie, 1985).

Table 2.1.1

Comrie's (1985) Analysis of Tenses

Absolute tenses		**Absolute-relative tense**	
Past	E before S	Past perfect	E before R before S
Present	E simul S	Future perfect	E before R after S
Future	E after S		

Note. E = the point of event; S = the point of speech; R = the point of reference.

Another notable difference between Comrie's and Reichenbach's framework lies in their perspectives on the present perfect. In Reichenbach's analysis, the present perfect positions an event in time relative to a reference point that aligns with the point of speech (Boland, 2006). In contrast, Comrie views it primarily as an aspectual category rather than a strictly tense-related one (Comrie, 1976, 1985). His argument is that if the reference point for the present perfect coincides with the present moment, in this case, the event in question would be situated in time prior to this reference point. However, in terms of temporal positioning, this would yield the same outcome as the past tense, which also denotes a situation as occurring before the present moment; therefore, the distinction between the perfect tense and the past tense does not lie in their temporal location (Comrie, 1985, p. 78).

While Comrie's perspective on tense is commonly adopted, it has been found that his analysis may not adequately explain certain common language phenomena (Boland, 2006). It fails to account for instances where a past tense marker is used to describe a state that remains true at the time of speaking (Klein, 1994). For example, in the sentence "*the movie was great, but it was three hours long*", both states (being great and being three hours long) are marked in the past tense, even though they are still true in the present moment. Also, a future marker can be used for a state that is already true at the point of speech. For example, in the conversation “*will you still be here at five?*” and “*yes, I will be here*”, the state of being here is already true at the moment of speech. Therefore, it cannot be argued that tense markers locate an event in time, as the events in the examples mentioned also apply to the present, yet they are not explicitly marked for the present tense (Boland, 2006).

To address this issue, Klein (1994) proposed the concept of topic time (TT), which is defined as “the time span on which the speaker’s claim on this occasion is confined” (p. 4). Tense does not locate the point of event on the timeline, but only the topic time (Klein, 1994). This suggests that the speakers do not indicate the complete temporal location of the entire event, as in Comrie's framework; rather they assert the temporal location for a specific part of the event, the one they choose to emphasise (Boland, 2006). The topic time (TT) is further differentiated from the time of situation (TSit), which refers to the actual time of the event (Klein, 1994). In the case of the absolute tense, the relationship is marked between TT, rather than TSit, and the time of utterance (TU) (i.e., the point of speech) (Boland, 2006). Specifically, the past tense represents that the speaker is asserting an event for a specific timeframe (TT) that occurred before TU. Correspondingly, the present tense can be described as TT and TU overlapping, while the future tense can be described as TT being located after

TU. In Klein's (1994) view, aspect acts to connect the time of situation (TSit) to the topic time interval (TT) (p. 99). For example, the English perfect is a combination of tense and aspect, in which tense concerns the relation between TT and TU, whereas aspect relates TT to TSit. It represents the selection of the post-state of a situation (TT) happening before TU (i.e., past perfect), overlapping with TU (i.e., present perfect), or after TU (i.e., future perfect), as summarised in Table 2.1.2.

Table 2.1.2

Klein's (1994) Analysis of Tenses

Absolute tenses		**Absolute-relative tense**	
Past	TT before TU	**Past perfect**	TT before TU, TT after TSit
Present	TT overlaps TU	**Present perfect**	TT overlaps TU, TT after TSit
Future	TT after TU	**Future perfect**	TT after TU, TT after TSit

Note. TT = topic time; TU = time of utterance; TSit = time of situation.

The current study adopts Klein's view that tense does not locate the entire duration of an event on the timeline. However, that is not to say that it does not locate the event in time; it simply focuses on the part relevant to topic time (Boland, 2006; Dik, 1997). The speaker and the addressee typically share a mutual understanding of the topic time frame that the speaker is referring to when making an assertion (Boland, 2006). The topic time frame can be large, corresponding to the entire state of an event, or it can be small, concerning only one relevant part of the event (Klein, 1994).

In specific languages, tense is either expressed grammatically through inflections (e.g., English, French) or lexically through a variety of deictic adverbials, temporal adjectives, and adverbs (e.g. Chinese) (C. S. Smith, 1997). It is necessary to address what grammaticalization and lexicalisation mean, as the boundaries between these two are sometimes not very clear-cut. As argued by Bybee et al. (1994), grammaticalization refers to a process that "grammatical morphemes develop gradually out of lexical morphemes or combinations of lexical morphemes with lexical or grammatical morphemes" (p. 4). It involves morphosyntactic transformations and becomes increasingly reliant on the phonological and semantic context in which it appears (Bybee et al., 1994).

Table 2.1.3

Examples of Difference Between Grammatical and Lexical Items

Grammatical items	Lexical items
General / abstract meaning	Specific meaning
No selection restrictions	Selection restrictions
Systematically/obligatorily used	Optionally used
Rigid syntactic position	Flexible syntactic position
Phonologically reduced	Phonologically rich

Note. The table is adapted from Boland (2006, p. 93).

From Comrie's (1985) perspective, the clear instances of grammaticalization need to satisfy two criteria: they are obligatorily expressed and morphologically bound. However, there are cases where the criteria are not sufficient to decide on the grammatical category, such as the use of auxiliaries in some languages to express tense (Comrie, 1985). Clear examples of grammatical items include tense inflections on verbs (e.g., English past *wanted* vs. non-past *want*) and other grammatical words adjacent to the verb (Comrie, 1985). The grammatical category of tense is often described as a feature associated with verbs, as in most languages with tense, it is typically integrated into the verb phrase as a grammatical feature (Comrie, 1985; Leech, 1987). In contrast, lexical items are more independent entities and have specific meanings and more flexible syntactic positions (Bybee et al., 1994). Lexicalisation does not have any necessary effects on the grammatical structure of the language (Comrie, 1985). Examples of lexical items are temporal adverbs (e.g., yesterday, tomorrow), adverbials (e.g., last year, in the morning), adjectives (e.g., recent, annual). Some differences between grammatical and lexical items are presented in Table 2.1.3.

2.2 Lexical Aspect

Different from tense which is deictic in nature, aspect is concerned with the viewpoint the speaker takes to look at the internal temporal course of an event (Comrie, 1976). It is discussed by linguists from both lexical and grammatical perspectives. Lexical aspect, also known as situation aspect (C. S. Smith, 1997), is related to verbal semantics, focusing on the inherent temporal characteristics embedded within verbs and verb phrases (Bardovi-Harlig & Reynolds, 1995). It examines how different verbs naturally convey aspects such as duration, punctuality, or a combination of both. For instance, verbs like *talk* and *sleep* inherently imply ongoing actions with duration, while verbs such as *recognise* and *fall* suggest instantaneous or punctual events (Bardovi-Harlig & Reynolds, 1995).

Table 2.2.1

Four-way Classification of Verb Predicates Based on Vendler (1957)

Verb class	Characteristics	Examples
States	Homogeneous situations without successive phases, dynamicity, or endpoints.	own, believe, love, hate
Activities	Actions with successive phases and no inherent endpoint.	walk, talk, dance, swim
Accomplishments	Actions involve successive phases but include a natural endpoint and often a change of state.	paint a picture, fix a car, build a bridge, write an essay
Achievements	Punctual, instantaneous situations with a natural endpoint and no duration.	fall, catch a ball, lose a game, recognise a face

Note. Examples are taken from Li and Shirai (2000, p. 15).

Vendler (1957) proposed one of the most influential categorisation systems for verbs based on their inherent semantic values, which distinguishes between four main classes of verbs: states, activities, accomplishments, and achievements (see Table 2.2.1). This classification has been widely used as a framework of verb analysis by many researchers (e.g., Dowty, 1979; Smith, 1997). Comrie (1976) extended Vendler's analysis and proposed classifying verbs based on contrasting pairs, such as dynamic vs. stative, telic vs. atelic, and punctual vs. durative. These pairs can be represented by binary features, including [+/-dynamic], [+/-telic], and [+/-punctual]. Specifically, dynamic verbs "consist of successive stages which occur at different moments" (Smith, 1997, p. 19) and are normally maintained by continued input of energy (e.g., play, eat) (Comrie, 1976), whereas stative verbs refer to events that are static or unchanging with an undifferentiated period (e.g., know, seem) (Smith, 1997). Telic verbs describe events that have natural endpoints, and once the endpoint is reached, a change of state occurs and the event is complete (e.g., finish, build); on the other hand, atelic verbs do not have natural endpoints and they can stop at any point without outcomes (e.g., talk) (Smith, 1997). Punctual or instantaneous verbs refer to situations that do not last in time, "one that takes place momentarily" (e.g., sneeze, hit) (Comrie, 1976, p. 42), whereas durative verbs can extend over a period of time (e.g., sleep). Table 2.2.2 is a summary resulting from the combination of Vendler's and Comrie's systems.

Table 2.2.2

Inherent Semantic Aspect Based on Vendler (1957) and Comrie (1976)

	States	**Activities**	**Accomplishments**	**Achievements**
Punctual	-	-	-	+
Telic	-	-	+	+
Dynamic	-	+	+	+

Note. The table is adapted from Andersen and Shirai (1994, p. 134).

2.3 Grammatical Aspect

Different from lexical aspect, grammatical aspect (also known as viewpoint aspect) is expressed morpho-syntactically through grammatical morphemes or periphrastic constructions in language (P. Li & Shirai, 2000; Salaberry & Shirai, 2002; C. S. Smith, 1997).

Grammatical aspect makes the basic distinction between perfective and imperfective. The perfective viewpoint refers to a bounded situation with endpoints; in other words, the situation is viewed from outside as a whole or in a completed state (Comrie, 1976). Smith (1997) used a general schema to show the initial (I) and final (F) endpoints of the event described by the aspectually perfective sentence *Lily planted a tree*, as in Figure 2.3.1 (a). It indicates that the event of planting a tree is ended and views the event entirely, and both the beginning and ending points are implied.

Figure 2.3.1

Representations of Perfective and Imperfective Viewpoints

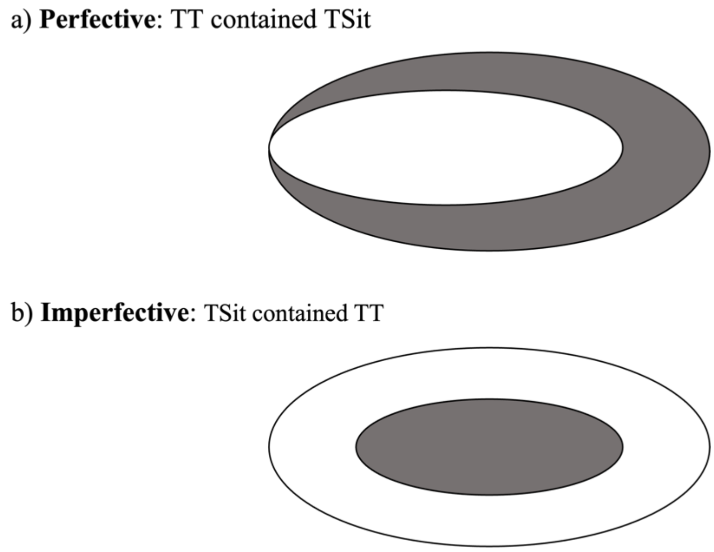
a) **Perfective**: *Lily planted a tree.*

b) **Imperfective**: *Lily was planting a tree.*

On the other hand, the imperfective viewpoint looks at part of an event from inside, with no stress on its specific initial or final endpoint (C. S. Smith, 1997). As shown in Figure 2.3.1 (b), the imperfective sentence *Lily was planting a tree* does not tell whether the event of planting a tree has been completed by Lily, unlike the perfective viewpoint. However, it does signify that the planting process was ongoing at a certain moment in time. The decision regarding which aspectual perspective to adopt largely depends on the speaker (P. Li & Shirai, 2000). When describing the event of ‘Sarah baking cookies yesterday’, the speaker has the flexibility to present it either from a perfective angle (e.g., *Sarah baked cookies*) or an imperfective angle (e.g., *Sarah was baking cookies*) based on their own perspective. This is the reason why grammatical aspect is also referred to as viewpoint aspect by Smith (1997).

Figure 2.3.2

Klein’s Analysis of Perfective-Imperfective Distinction



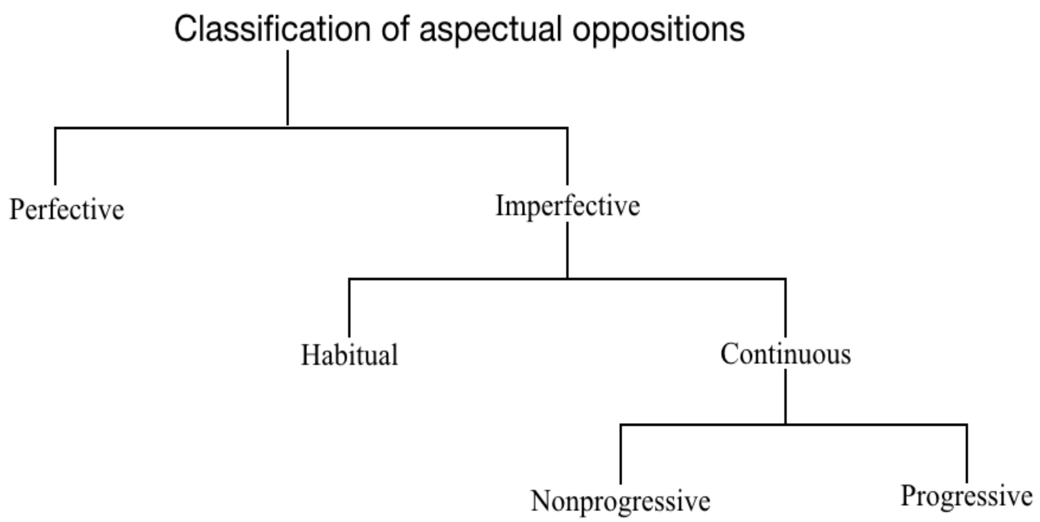
Note. TT = grey circle; TSit = white circle. The figure is adapted from Gvozdanović (2012, p. 790).

Alternatively, as discussed in Section 2.1, Klein (1994) proposed that aspect can also be represented through the temporal relationship between the topic time (TT) and the time of situation (TSit). TT refers to the time span in which the speaker makes their assertion, while TSit indicates the specific time when the situation occurs (Klein, 1994). The imperfective aspect can be defined as “the time for which an assertion is made falls entirely within the time of the situation,” meaning that TSit includes TT (Klein, 1994, p. 108). In contrast, the perfective aspect can be seen as "TT AT TSit; whenever TSit denotes two states (in the sense of a change of state), both the source and the target state are assumed to at least partly coincide with TT" (Gvozdanović, 2012, p. 786). These relationships are visually represented in Figure 2.3.2.

According to Comrie (1976), the imperfective aspect can be further divided into habitual and continuous aspect, and continuous can be subdivided into progressive and non-progressive categories (see Figure 2.3.3). Progressive aspect conveys the idea that "an event is progressing dynamically over a time frame opened up by an utterance" (Mair, 2012, p. 804). The time frame in this case can refer to the "topic time (TT)" (Klein, 1994). For example, the progressive sentences "*Sarah is/was baking cookies*" illustrate the action of baking cookies as dynamic and in progress over the topic time frame.

Figure 2.3.3

Comrie's Classification of Aspectual Categories



Note. The figure is adapted from Comrie (1976, p. 25).

On the other hand, habitual aspect refers to "a series of events or episodes, viewed as a whole and distributed over an explicitly or implicitly given interval of time" (Binnick, 2005, p. 340), or "a characteristic situation that holds at all times" (Comrie, 1985, p. 39). For example, in English, the "*used to*" construction (e.g., *Sarah used to bake cookies*), though only in the past tense, specifically denotes a habitual situation. The simple form in English may also convey habitual meaning, particularly when paired with temporal adverbs such as *usually, typically, every night*, etc. (Binnick, 2005). In the simple sentence "*John worked at the university*", it may receive the habitual interpretation, but the only excluded interpretation is the progressive meaning (Comrie, 1976).

Different languages typically have developed different grammatical aspectual categories. For instance, many Romance languages (e.g., French and Spanish) directly encode the perfective-imperfective distinction, and the imperfective aspect can denote both habitual and continuous (i.e., progressive and non-progressive) situations. This is the reason why the imperfective aspect is also referred to as general imperfective (P. Li & Shirai, 2000). For instance, the imperfective form in Spanish can express both meanings of *John was dancing* (progressive) and *John used to dance* (habitual) (Comrie, 1976), which is in contrast to languages such as English that have separate habitual and progressive aspects to encode imperfectivity. Meanwhile, there are languages that do not have grammatical aspect at all, such as Finnish and Estonian (P. Li & Shirai, 2000).

There is also variation across languages in the degree to which they grammaticalise aspectual categories. For example, progressive aspect is highly grammaticalized in English. Its use extends beyond expressing action-in-progress as it can be employed to convey habitual and futurate meanings in some cases (P. Li & Shirai, 2000; Schmiedtová & Flecken, 2008). In other words, the English progressive marker has undergone grammaticalization to become a general imperfective marker (Bybee et al., 1994). On the other hand, in languages such as Chinese and Thai, progressive markers are less grammaticalized, and their usage is restricted to only indicating action-in-progress meanings (P. Li & Shirai, 2000).

2.4 Perfect

Comrie (1976, pp. 56–61) summarised the various types of perfect as follows:

Table 2.4.1

A Summary of Different Types of Perfect

Type of perfect	Definition	Example sentence
Perfect of result	The outcome of a past situation are relevant and applicable at the moment of speaking	*John has arrived* (he is here).
Experiential perfect	An event takes place at least once within a timeframe that began in the past and continues up to the present	*John has been to London*.
Perfect of persistent	A situation that continues throughout a specific timeframe into the present	*John has lived here for ten years* (he still lives here).
Perfect of recent past	The past situation is very recent in its relevance to the present.	*John has just left the house*.

The fundamental meaning underlying the various types of perfect has often been characterized as "past situation with current relevance (CR)" (Dahl, 1985; McCoard, 1978). In fact, it can occur not only with the present tense to create the present perfect, but also with the past and future tense to create the past perfect (or pluperfect) and future perfect respectively. Therefore, the perfect can be defined as follows: “the situation occurs prior to reference time and is relevant to the situation at reference time” (Bybee et al., 1994, p. 54). For example, the present perfect is used to describe a situation that is more associated with the current moment compared to a situation expressed in the simple past tense (Ritz, 2012).

According to Bybee et al. (1994), perfects often have their origins in resultative constructions that are used to emphasise that "a state exists as a result of a past action" (p. 54). Resultative constructions differ from the perfect in their ability to co-occur with adverbs denoting an ongoing, non-limited duration (e.g., *The man is still gone/*The man has still gone*) (Lindstedt, 2000). Over time, there has been a shift in semantic meaning from current result to CR (Lindstedt, 2000). Lexically, this shift is marked by an expansion in the types of verbs used, from solely employing telic verbs within resultative constructions, to incorporating both telic and atelic verbs in perfect structures (Bybee et al., 1994).

The perfect gradually loses its CR restriction and develops to include experiential meaning (Lindstedt, 2000; Ritz, 2012). As in (2), the experiential perfect discusses a past event without referring to a specific occasion, featuring a lack of specific temporal reference to the past and less directly relevant to the present (Lindstedt, 2000). From this perspective, it is more tense-like than the canonical CR perfect. It has been observed that CR perfects are often unstable and tend to develop into past perfective tenses where the past time reference is specific and definite (Lindstedt, 2001; Ritz, 2012). For instance, in French, the perfect has evolved into a compound past tense, known as the passé composé, allowing its combination with specific past time adverbials (e.g., *Martin est parti il y a deux jours,* 'Martin has left two days ago'). Similarly, in English, both past and future perfect can be used with definite temporal adverbials (e.g., *At noon John had/will have left*). However, this usage does not apply to the English present perfect, which is considered ungrammatical, as illustrated in (1) and (2). This problem is often referred to as "the present perfect puzzle" (Klein, 1992).

(1) **John has left at noon.*

(2) **At six o'clock this morning, John has woken up.*

As argued by Lindstedt (2000, 2001), the level of incompatibility between specific time adverbials and the perfect in a given language can indicate the extent to which it has evolved into a predominantly experiential form and, consequently, a type of tense. The potential next phase in its development would be the complete disappearance of the CR meaning, and this process might be occurring in American English. For example, when a child asks their parent "Can I go now?", the response in (3a) seems more characteristic of an American parent, while (3b) is favoured by a British parent. It seems that the experiential meaning (i.e., more tense-like) is preserved much better than the CR meaning in American English (Lindstedt, 2000).

(3) a. Did you do your homework?

b. Have you done your homework?

There is more consensus that the representation of the perfect always includes some form of temporal specification, as it can be found in the past, present, and future tenses (Ritz, 2012). However, the inquiry into whether it can also be classified as an aspect has yielded different answers. Some scholars (Brinton, 1988; Dahl, 1985) have acknowledged it as a type of aspect, while some have argued that the perfect is a product of the interplay between perfective aspect and either tense or another aspect (Bauer, 1970; Comrie, 1976). In English, the perfect and progressive aspect can be combined (e.g., *John has been sleeping*). Also, the various types of perfect can be considered as an interaction with the lexical aspect (Bauer, 1970). For example, when combined with telic verbs, it takes on a resultative aspect (e.g., *John has found the key*). In contrast, when used with atelic verbs, it conveys an experiential meaning (e.g., *John has read the book*), and when paired with specific adverbials, it exhibits a continuative nature (e.g., *John has lived in London since 2010*).

In different languages, the perfect is typically expressed through inflectional forms (e.g., Swahili) or through periphrastic constructions (Dahl, 2022). The latter case is very common in European languages (e.g., English, Spanish), which involve inflected auxiliary verbs (e.g., ‘*be*’ and/or ‘*have*’) usually with non-finite forms of the main verbs (e.g., past participle form), deriving originally from resultatives (Bybee et al., 1994). In some other languages, the perfect is achieved through uninflected particles meaning ‘already’ which do not influence the form of the main verbs, such as Chinese sentence-final particle *le* and standard Indonesian *sudah* (Dahl, 2022).

Chapter 3 Cross-Linguistic Variations of Grammatical Tense and Aspect

All languages can convey fundamental concepts related to time, but they significantly differ in how they express temporal meanings, particularly in how they encode tense and aspect. Some languages require obligatory grammatical markings for tense and aspect on verbs, whereas others do not. This chapter aims to introduce language-specific facts regarding the grammaticized categories of tense and aspect in English, Chinese, and Arabic, the languages relevant to the current study. By examining the cross-linguistic differences among these three languages, predictions about the acquisition of L2 English tense and aspect can be made in the presence of L1 influence.

3.1 English

As discussed in Chapter 2, tense can be understood as a deictic category that establishes a temporal relationship between two time spans: the topic time (TT) and time of utterance (TU) (Klein, 1994). The sequence of these two time spans determines the three tenses: past, present, and future. When TT precedes TU, it conveys the past tense; if TT and TU occur at the same time, it indicates the present tenses, and when TT follows TU, it represents the future tense (Klein, 1994).

In specific languages, tense is either expressed grammatically through verbal inflections (e.g., French) or solely lexically through a variety of deictic adverbials, temporal adjectives, and adverbs (e.g. Chinese) (C. S. Smith, 1997). English falls into the former category. Tense is explicitly marked in English, and in this context, 'marked' can be interpreted as "morphologically expressing reference to a past or non-past 'temporal domain'" (Declerck, 2006, p. 23). English clearly distinguishes the contrast between past and non-past, and it is commonly believed to have two basic tenses: past and present (Lyons, 1968). The past tense, also known as preterit, is obligatorily marked on regular verbs via inflectional morphology *-ed*, such as *walked* in (1a), and on irregular verbs through unique conjugations, as exemplified by *sang* in (1b). The present tense, on the other hand, is expressed using unmarked verb forms (the stem of the verb), as illustrated by *walk* in (2a). The only exception is the third person singular, which is suffixed with *-s*, as seen in *sings* in (2b).

(1) a. *I walked to school*. (Past)

b. *She sang beautifully*. (Past)

(2) a. *I walk to school every day*. (Present)

b. *She sings beautifully*. (Present)

Many grammarians maintain that English does not have a future tense, and this is because the future tense is not expressed through an inflectional morpheme but rather by using other various means, such as modal verb *will* (3a) or simply the present tense (3b, 3c) (Declerck, 1991; Lyons, 1968). This is in contrast to the past and present distinction, which is expressed morphologically.

(3) a. *They will leave tomorrow*.

b. *They are leaving tomorrow*.

c. *They are going to leave tomorrow*.

In contrast to tense, aspect is concerned with the internal temporal structure of a situation rather than its external temporal reference points (Comrie, 1976; Klein, 1994). As discussed in Chapter 2, aspect is encoded either lexically by the inherent semantic meaning of the verbs (lexical aspect) or morphologically through verbal inflections (grammatical aspect) (Salaberry & Shirai, 2002). While there is less variation in terms of lexical aspect, there is considerable diversity in grammatical aspect among languages (P. Li & Shirai, 2000).

Based on Comrie's (1976) categorisation, grammatical aspect makes the basic distinction between perfective and imperfective. Perfective aspect views a situation externally and in a completed state, whereas imperfective aspect views an event internally with no stress on its specific initial or final endpoint (C. S. Smith, 1997). Indeed, many Romance (e.g., French, Spanish) and Slavic languages (e.g., Czech, Russian) encode this binary contrast directly via inflectional marking on the verb (Dahl, 1985). However, English does not grammaticalise this distinction in a straightforward way. It typically uses the past tense form to convey perfective aspect. It can be seen that the distinction between tense and aspect can be somewhat blurred; here, the English past tense can convey both past tense and perfective meanings at the same time (P. Li & Shirai, 2000).

English also encodes imperfectvity partially along the lines of progressivity (C. S. Smith, 1997). Prrogressive aspect typically refers to events in progress over a time frame and it is often marked using periphrastic marking (Deo, 2012); in English, the tense-marked auxiliary *be* and the participle form of the verb *V-ing* is used to convey a progressive interpretation of an event, as in (4).

(4) a. *She is writing a book.* (Porgressive)

b. *She was writing a book*. (Progressive)

This is ditinguished from the grammatical category of habitual aspect in English, which is often expressed using simple present form and the consturction *used to*, as in (5).

(5) a. *She writes*. (habitual)

b. *She used to write*. (habitual)

In terms of the perfect, as discussed in Chapter 2, while it can have various types and meanings, the concept of current relevance (CR) is considered most crucial and fundamental (Dahl, 1985; McCoard, 1978). The perfect can be found in the present, past, and future tenses; as a result, its representation always includes some form of temporal specification (Ritz, 2012). In English, perfect meaning is morphologically marked by using the auxiliary verb *have* followed by a past participle *V-ed*. The auxiliary verb can be used in the past, present, and future tenses to create past (6a), present (6b), and future (6c) perfect meanings.

(6) a. *He had finished his report*. (Past perfect)

b. *He has finished his report*. (Present perfect)

c. *He will have finished his report*. (Future perfect)

The present perfect in basically describes a past situation that is relevant to the present moment (Bardovi-Harlig, 2001; Comrie, 1976), setting it apart from the simple past, which is perceived as distinct from the present (McCoard, 1978, p. 19). It can be summarised that both the present perfect and the simple past possess the characteristic of being [+past]. However, they diverge when it comes to the feature of CR, with the present perfect possessing [+CR] and the simple past being [-CR] (Suh, 1992, cited in Bardovi-Harlig, 2001). It is well known that the English present perfect (particularly in British English) cannot co-occur with definite time adverbials in the past (7a), whereas simple past can (7b). This is often referred to as the 'the present perfect puzzle' (Klein, 1992). This can be explained by the fact that the present perfect specifies the pre-present tense, which describes a period just before the current moment but is still within the present time sphere (Declerck et al., 2006). This tense denotation explains why the temporal adverbials that include a sense of the present (e.g., *so far*, *already*, *lately*, *recently*, *since*) can be used with present perfect (8a), but not with simple past (8b) (Bardovi-Harlig, 2001). Also, the present perfect typically represents a situation occurring at an unspecified time whereas definite time adverbials (e.g., *yesterday*, *three weeks ago*) aim to identify an exact timing, leading to a semantic inconsistency between the two, as illustrated in (7a) (Bardovi-Harlig, 2001).

(7) a. *Lisa has studied French yesterday.

b. Lisa studied French yesterday.

(8) a. *Lisa has studied French since last September.*

b. **Lisa studied French since last September.*

3.2 Chinese

In contrast to English, which marks tense morphologically, Chinese has been considered a tenseless language (Lin, 2006, 2012). It is among the rare languages in the world that have received a thorough tenseless analysis within the existing literature. In Chinese, verbs are not obligatorily inflected to indicate time references (C. Li & Thompson, 1989). Unmarked or bare verbs can appear in past (9), present (10), and future (11) tense contexts. The expression of temporal locations can be achieved using temporal adverbials (e.g., yesterday, tomorrow) or simply inferred from contextual information (C. Li & Thompson, 1989; Sun, 2006).

(9) *Wo zuo-tian qu dian-ying-yuan* (Past)

I yesterday go cinema

'I went to the cinema yesterday.'

(10) *Wo tian-tian qu dian-ying-yuan* (Present)

I everyday go cinema

'I go to the cinema every day.'

(11) *Wo ming-tian qu dian-ying-yuan* (Future)

I tomorrow go cinema

'I will go to the cinema tomorrow.'

In this case, temporal adverbials can denote temporal relations, but they cannot be considered tense markers (Lin, 2012). Tense markers, by Comrie's (1985) criteria, are grammaticalized expressions that are obligatory in every sentence. In Chinese, past tense reference can be expressed both without the use of temporal adverbials denoting past time (12) and with the use of aspectual markers (13). In these examples, there is no single morphosyntactic morpheme that obligatorily appears with them. In contrast, in English, the morpheme *-ed* is seen as a tense marker, as it is always present when indicating the time sequence between the topic time and speech time, even if another expression conveys a similar temporal relation, such as the adverbial *last night* in (14).

(12) *Ta dapo bei-zi*

She break glass

'She broke a glass.'

(13) *Ta dapo-LE bei-zi*

She break-ASP glass

'She broke a glass.'

(14) *Last night, she cried.*

While Chinese does not explicitly mark tense, it has a rich array of grammatical aspect markers, as summarised in Table 3.2.1 (C. Li & Thompson, 1989; P. Li & Shirai, 2000). Unlike English, Chinese directly encodes the perfective-imperfective distinction. In general, there are two perfective aspect markers: *LE* and *GUO*. Both of them are verb-final, and their difference is that *LE* presents a situation as completed or bounded, particularly focusing on the endpoint, and this situation may not necessarily be in the past (Duff & Li, 2002). In contrast, *GUO* is used to indicate that an event has been experienced in the past and that the resulting state no longer exists at the time of speech (Klein et al., 2000; C. Li & Thompson, 1989; Smith, 1997).

Under the imperfective category, rather than grammaticalising the progressive and habitual contrast like English, Chinese only encodes the distinction between progressive and non-progressive, and there is no grammatical marker for habitual aspect (P. Li & Shirai, 2000). In Chinese, progressive aspect is overtly marked by using *ZAI*. *ZAI* initially emerged as a verb, later evolving into a locative preposition, and more recently, taking on the role of an imperfective aspect marker (Woo, 2021). While it can function as a preposition appearing both before and after the verb in its earlier uses, as an aspect marker, it is only positioned before the verb (preverbally) to indicate that an action or event is ongoing or in progress at a certain point in time (P. Li & Shirai, 2000).

Table 3.2.1

A Summary of Grammatical Aspect Markers in Chinese

	Aspect marker	**Meaning**	**Example sentence**
Perfective marker	*LE* (Postverbal)	Perfective	Wo xie-LE yi-shou shi I wrote-LE a-CI poem 'I wrote a poem.'
	GUO (Postverbal)	Experiential	Wo xie-GUO yi-shou shi I wrote-GUO a-CI poem 'I have written a poem.'
Imperfective marker	*ZAI* (Preverbal)	Progressive	Ta ZAI-du shu She ZAI-read book 'She is reading a book.'
	ZHE (Postverbal)	Durative (background event)	Ta ting-ZHE yin-yue xue-xi She listen-ZHE music study 'She studied while listening to music.'
	NE (sentence-final)	Progressive (in colloquial speech)	A: Ni ZAI gan shen-me? You ZAI do what? 'What are you doing?' B: Wo kan dian-shi NE I watch TV NE 'I am watching TV.'

Note. The table is adapted from Duff and Li (2002, p. 419).

On the other hand, the imperfective marker *ZHE* is often used to indicate an enduring and continuing situation, often serving as the background to the main event (C. S. Smith, 1997). It can co-occur with the progressive marker *ZAI* to present a durative and unbounded event (Liu, 2015).

As in (15), the progressive marker *ZAI* highlights the ongoing action of "*reading*" as an unbounded, temporary stage occurring in the middle of the process. On the other hand, the durative marker *ZHE* emphasizes the lasting, resulting state of reading. These two imperfective markers are semantically similar but emphasize different aspects of the action (Liu, 2015). Moreover, it is worth mentioning that the durative marker *ZHE* is changing into a general imperfective marker, as it can be attached to stative verbs whereas the progressive marker *ZAI* cannot (C. S. Smith, 1997). Particularly, in the northern dialects of Chinese, *ZHE* appears to be taking over the function of *ZAI* and is extending to express progressive meaning as well (Shirai, 1998).

(15) *Ta ZAI kan-ZHE shu*

She ZAI read-ZHE book

'She is (in the state of) reading a book.'

It has only been recently recognized that there is a third imperfective marker in Chinese, known as *NE*, which is positioned at the end of the sentence (P. Li & Shirai, 2000). This neglect can be attributed, in part, to the presence of pragmatic constraints on its usage in spoken language, as *NE* is only utilised to indicate progressive meaning primarily in responses to questions within informal dialogues (P. Li & Shirai, 2000).

In addition to the verb-final aspectual marker *LE*, Chinese also has a sentence-final particle *LE*. However, there has been ongoing debate about its functions or meanings in the literature (Klein et al., 2000). Some consider sentence-final *LE* as a *perfect* marker indicating the current relevance of an event at the reference time (C. Li et al., 1982; Liu, 2015). In this case, it can co-occur with the perfective verb-final *LE* in the same sentence. For example, in (16), the perfective *LE* indicates the completion of the action of *watching this movie*, while the perfect *LE* emphasizes its current relevance.

(16) *Wo kan-LE zhe-ge dianying LE*

I watch-LE this movie LE

'I have seen this movie.'

However, the use of sentence-final *LE* does not always guarantee a perfect reading. In sentence (17), a perfective meaning is the most probable interpretation, considering the presence of the definite time adverbial *last Friday* in the sentence. To better convey a perfect meaning, the adverb *yijing* ('already') is often employed to avoid misunderstanding and signal current relevance (Xiao & McEnery, 2004), as shown in (18). Also, the sentential *LE* can function as a mood or modal particle, conveying the speaker's mood or attitude (Arin, 2003; Duff & Li, 2002). An illustrative sentence is provided in (19).

(17) *Shang xing-qi-wu wo qu jian-shen-fang you-yong LE*

Up Friday I go gym swim LE

'Last Friday I went swimming at the gym.'

(18) *Wo yijing dao jia LE*

I already arrive home LE

'I have already arrived home.'

(19) *Zhe-ge ka-fei tai-tang LE*

This coffee very hot LE

'This coffee is very hot.'

As discussed in Section 2.4, the representation of the perfect always includes some form of temporal specification, as it can be found in the past, present, and future tenses (Ritz, 2012). While current relevance can be realised through the sentence-final *LE* in Chinese, encoding the perfect meaning requires the realisation of both tense and current relevance features (Bardovi-Harlig, 2001). However, since Chinese does not grammatically mark temporal relations, the present study argues that the perfect aspect is not grammatically encoded in Chinese.

3.3 Arabic

Now turning to Arabic, Arabic is an interesting case because it has a unique verb system that is believed to convey both tense and aspectual information (Ouali, 2017). This has led to ongoing debates among modern linguists regarding whether Arabic verb morphology primarily signifies aspect, with tense being conveyed through other means, or if it simultaneously marks both tense and aspect (Benmamoun, 2000). In recent literature, there appears to be a consensus that, despite the verb morphology not explicitly marking tense, Arabic is still considered a tensed language (Ouali, 2017).

In Modern Standard Arabic, the past tense is encoded by employing the perfective form of the verb (20), while the present tense is conveyed through the use of the imperfective form (21). When expressing the future tense, the imperfective form is used in combination with a future marker (22). The perfective form is formed solely through the addition of suffixes, whereas the imperfective form is created using both prefixes and suffixes (Benmamoun, 1999). In addition, the verb shows agreement with its subject in terms of person, number, and gender (Benmamoun, 2000).

(20) *katab-a ʕalj-un r-risaalat-a* (Past)

3s.m.-write.perf Ali-nom the-letter-acc.

‘Ali wrote the letter.’

(21) *ʕalj ka-j-ktəb r-risaala* (Present)

Ali 3s.m.-write.imp the-letter

‘Ali writes/is writing the letter.’

(22) *ʕalj ɣa-j-ktəb r-risaala* (Future)

Ali fut.-3s.m.-write.imp the-letter

‘Ali will write the letter.’

(Ouali, 2017, p. 90)

The two verb forms are also used to encode the perfective-imperfective aspectual opposition. As Comrie (1976) summarised, in Arabic, “perfective (verb forms) indicates both perfective meaning and relative past time reference, while the imperfective (verb forms) indicates everything else that is either imperfective meaning or relative non-past tense” (p. 80). Similar to English, which encodes perfective meaning via the simple past, Arabic also uses the simple past expressed by perfective forms to refer to a completed situation in the past (Ryding, 2005).

Unlike English, which encodes the imperfective category through the progressive/non-progressive distinction, Arabic is not morphologically marked for progressive aspect. Instead, it relies on a general or default imperfective form that can convey both progressive and habitual meanings, as exemplified in (23) (Ryding, 2005). In Arabic, the imperfective verb stem is used to convey different imperfective meanings, with the specific interpretation often relying on discourse context or the use of time adverbials (e.g., *ðaħiin* ‘right now’) (Al-Thubaiti, 2015; Mudhsh, 2021).

(23) *Ali yə-lʕəb kura*

Ali 3s.m.-play.imp football

‘Ali is playing (used to play/plays) football.’

(Al-Thubaiti, 2015, p. 189)

However, evidence from multiple modern Arabic dialects suggests an ongoing process of grammaticalization for the progressive aspect, in the forms of *gaaʕid* and *ʕammal*

(Ouali, 2022; Rubin, 2005). *gaaʕid* is derived from the posture verb ‘to sit’, which conveys the meaning of *sitting* (Al-Thubaiti, 2015). On the other hand, ʕammal is derived from the activity verb ‘to work’, carrying the literal meaning of working (Ouali, 2022). Both of the forms are fully inflected for person, number, and gender agreement. In certain dialects, both full and reduced forms of these two progressive markers are used interchangeably (Ouali, 2022). For example, in Saudi Arabic, when *gaaʕd* is employed, it conveys a progressive meaning without implying a habitual interpretation, as shown in (24).

(24) *Ali gaaʕid yə-lʕəb kura*

Ali prog 3s.m.-play.imp football

‘Ali is playing football.’

(Al-Thubaiti, 2015, p. 190)

In addition, Arabic appears to lack morphological markers for conveying perfect/non-perfect meanings. Specifically, there is no grammatical marker that indicates current relevance, which is crucial for expressing the perfect aspect (Badawi et al., 2004; Farina, 2017). In contrast to English, which distinguishes between the simple past and the present perfect, Arabic often utilises the perfective form in both contexts (Bahloul, 2008). As demonstrated in (25), both interpretations can be possible depending on the context, though the past simple meaning is inherently preferred.

(25) *Kitab Ali Darsa-h*

3s.m.-write.prf Ali lesson-his

‘Ali wrote his lesson.’ or ‘Ali has written his lesson.’

(Alruwaili, 2014, p. 61)

Instead, the intended meaning can be achieved through contextual cues and the specific verbs used (Bahloul, 2008). For example, some linguists believe that in Arabic, the particle *qad* (already/yet) can be used in conjunction with the perfective verb form to express the perfect meaning (Al-Aswad, 1983; Comrie, 1976; Fassi Fehri, 2003), as illustrated in (26). However, as argued by Bahloul (2008), if this were the case, there would not be instances of [*qad* + perfective] found even in situations where the English present perfect cannot be applied. As shown in example (27), the perfect interpretation is improbable.

(26) *qad kataba ar-risaalata*

qad 3s.m.-write.imp the-letter

'He has written the letter.'

(27) *qad žaaʔ-a al-waladu ʔamsi*

qad 3s.m.-come.prf the-boy yesterday

'The boy came yesterday.'

*'The boy has come yesterday.'

(Bahloul, 2008, pp. 66–67)

To convey a more precise perfect meaning, adverbials, such as *mundu* 'since' and *faqat* 'just', are often used by speakers (Fassi Fehri, 2003), as illustrated in example (28).

(28) *katab-a r-risaalat-a mundu xamsi saaᶜaat-in* (l-ʔaana)

3s.m.-write.prf the-letter-acc since five hours (now)

'He has written the letter for five hours (now).'

(Fassi Fehri, 2003, p. 83)

3.4 Summary

To summarise the discussion so far (see Table 3.4.1), English and Arabic both encode [+/−past] morphologically, whereas Mandarin does not grammaticalise tense, relying primarily on temporal adverbials, such as *last year*, *yesterday*, *today*, to denote temporal locations.

The three languages show significant variation in how they realise grammatical aspect. English does not straightforwardly mark the aspectual contrast between the perfective and imperfective. Instead, it distinguishes the imperfective category by differentiating between progressive and non-progressive meanings, denoted as Asp[+/−prog]. Progressive aspect in English is realised through the auxiliary *be* and the participle form of the verb *V-ing*. On the other hand, both Arabic and Chinese grammatically encode the perfective and imperfective distinction, represented as Asp[+/−perfective]. In contrast to English, Arabic does not encode Asp[+/−prog], using a default imperfective form to convey both progressive and habitual meanings. However, it is worth noting that there is an ongoing process of grammaticalization for the progressive aspect in many Arabic dialects, involving the forms of *gaaʕid* and *ʕammal*. Similar to English, Chinese uses grammatical markers, including preverbal ZAI and postverbal ZHE, to encode Asp[+/−prog].

Table 3.4.1

Grammatical Encoding of Tense and Aspect Across English, Chinese, and Arabic

	English	Chinese	Arabic
Past/non-past	+	–	+
Perfective/imperfective	–	+	+
Progressive/non-progressive	+	+	?
Perfect/non-perfect	+	–	–

Note. The question mark (?) shows that there is an ongoing process of grammaticalization.

This study maintains that the expression of present perfect meaning involves the realisation of both tense specifications and perfect aspect. In English, these features are encoded through the morphosyntactic representation of Asp[+/−perfect] and T[+/−past], which are overtly realised by the auxiliary *have* and the past participle form of the verb (*V-ed/-en*). Chinese is believed to have a perfect marker, which is the sentence-final *LE*. However, the use of this marker does not consistently indicate a perfect interpretation, and it can also carry modality meanings. More importantly, although the marker *LE* can signal current relevance, Chinese lacks T[+/−past] features to fully encode present perfect meaning. In Arabic, there is no grammatical marker that explicitly indicates current relevance. To convey the intended meanings, the particle *qad* can be used. However, similar to the Chinese marker *LE*, the use of *qad* does not guarantee a perfect reading. Adverbials are often employed instead to achieve intended perfect interpretation.

Chapter 4 L2 Acquisition of Tense-Aspect

The expression of time is a critical aspect of effective communication, and one of the earliest challenges in L2 acquisition is developing the ability to talk about time (P. Li & Shirai, 2000). This may explain the significant interest of many L2 researchers in the acquisition of L2 tense and aspect. It has been observed that acquiring L2 tense and aspect, along with their associated morphosyntactic forms, poses substantial challenges for many L2 learners. This is not surprising, given that the concept of temporality is realised differently in various languages as discussed in Chapter 2. When constructions differ between a learner's L1 and the L2, difficulty can arise.

This chapter starts by exploring relevant studies in the field, with a specific focus on how learners' L1 may influence their acquisition of tense and aspect. Then, in light of the existing conflicting evidence, various theoretical accounts are examined, including Full Access Accounts and Representational Deficit Accounts, i.e., Representational Deficit Hypothesis and Morphological Congruency Hypothesis. Following that, considering the ongoing debate regarding implicit and explicit knowledge, this chapter also reviews relevant L2 processing studies on tense and aspect. These studies have provided findings that can contribute to a more in-depth exploration of learners' mental representations and grammatical knowledge. Finally, the chapter concludes by discussing the rationale for the present study and outlining the research questions and predictions to be addressed.

4.1 L2 Studies on Tense-Aspect

The acquisition of tense and aspect has been investigated from different perspectives in the field of SLA, given its critical role as a communicative tool that enables speakers to establish temporal references and express their viewpoints on the temporal progression of a situation (Shirai & Kurono, 1998).

In the 1980s, due to the increasing interest in the temporal semantics of interlanguage, there was a shift from a focus on the acquisition of morphology as form (e.g., morpheme order studies) to a greater emphasis on morphology as the surface expression of an underlying semantic system (Bardovi-Harlig, 1999, 2000). This shift was supported by advances in the theories and concepts of the temporal system, such as tense, lexical aspect, and grammatical aspect, and cross-linguistic studies of various languages, which collectively provided a foundation for investigating the acquisition of temporal systems in SLA (Bardovi-Harlig, 2000). Within this line of research, two main approaches have emerged: the meaning-oriented approach and the form-oriented approach. The former investigates how learners express temporal concepts by using different types of linguistic devices (Bardovi-Harlig, 2000). The underlying idea of this approach is that L2 learners already have access to a comprehensive array of semantic concepts derived from their L1 experiences (Bardovi-Harlig, 2000). Therefore, unlike in L1 acquisition, L2 learners only need to acquire specific means of expressing these concepts (von Stutterheim & Klein, 1987). Most studies adopting this approach have employed longitudinal designs and elicited production data from participants through conversational interviews, film retell tasks, or written narratives (Bardovi-Harlig, 1992; Dietrich et al., 1995). These studies have revealed a potentially universal acquisitional sequence among L2 learners, progressing from reliance on pragmatic devices (e.g., discourse), to lexical devices (e.g., adverbials and connectives), and eventually to grammatical devices (e.g., verbal morphology) (Andersen, 1991; Dietrich et al., 1995; Giacalone Ramat & Banfi, 1990). Each stage is characterised by the dominant use of corresponding linguistic devices; however, this does not imply the complete absence of features from earlier or later stages (Bardovi-Harlig, 2000).

On the other hand, the form-oriented approach focuses specifically on the verbal morphology used by L2 learners and aims to understand what it means in these learners' underlying semantic system (Bardovi-Harlig, 2000). Research adopting this approach has concentrated on three main areas: the sequence of acquisition, the influence of discourse structure, and the influence of lexical aspect (Bardovi-Harlig, 2000). For instance, studies examining the sequence of acquisition have adopted both longitudinal and cross-sectional designs to investigate the development of a particular form in the target language. One such study found that the acquisition of the past tense in English progresses from past simple to past progressive, and to present perfect and past perfect (Klein, 1995). This particular order may be explained by various factors, such as morphosyntactic complexity, semantic complexity, or frequency of input, with evidence often drawn from L1 acquisition studies (C. S. Smith, 1980). In addition, studies examining the the influence of discourse structure have often used narrative discourse analysis (Bardovi-Harlig, 2000). For example, it has been observed that English learners differentiate foreground from background in oral narratives by mainly marking the foreground with the simple past tense, while background verbs are often left in their base forms (Flashner, 1989). These studies are similar to meaning-oriented studies in their reliance on small samples and descriptive approaches (i.e., oral and written narratives).

To date, a large amount of research has investigated the influence of lexical aspect on the development of tense-aspect morphology, with the Aspect Hypothesis (AH) being one of the most influential theories in this area (Bardovi-Harlig & Comajoan-Colomé, 2020). According to the AH, in both L1 and L2 acquisition of tense and aspect markers, learners tend to be initially guided by the inherent semantic aspect of the verbs (Andersen, 1991, 2002; Andersen & Shirai, 1994, 1996). In other words, verbal morphology will occur with verbs that share similar semantic values. For example, the perfective past is typically used with telic verbs that imply inherent endpoints (e.g., find, recognise); in contrast, imperfective are initially applied to statives (e.g., want, need), and progressive is used with activity verbs (e.g., run, walk) (Bardovi-Harlig & Comajoan-Colomé, 2020). The AH has been continually refined in response to a growing body of empirical evidence. Based on Vendler's (1967) categories of verbs (see Section 2.2), the latest AH predicts the developmental stages in acquiring tense and aspectual forms. Specifically, learners first employ past or perfective markers with achievement and accomplishment verbs, gradually extending this usage to activity and stative verbs (Andersen, 2002). In contrast, the use of imperfect past markers begins with stative and activity verbs before extending to accomplishment and achievement verbs (Andersen, 2002).

Given that Vendler's (1967) categories of verb predicates are present in numerous languages and reflect the cognitive and perceptual distinctions that humans use to conceptualise events and actions, the AH is considered a universal principle of acquisition, applicable across different language pairs and learning contexts (Andersen & Shirai, 1996; Bardovi-Harlig & Bergström, 1996). However, this focus on lexical aspect has led to an increasing tendency to overlook other aspectual factors, such as grammatical aspect. As Housen (2002) explained, his study focused on coding verb predicates (the unmarked verb and its main arguments) primarily for their inherent lexical aspect rather than taking into account their real-world viewpoint aspect. This research practice revolves around (i) analysing how a particular linguistic form is distributed across different classes of lexical aspect, and (ii) examining how these forms are used within specific lexical aspect categories (Bardovi-Harlig, 2002). Such an approach assumes that lexical aspect is the sole determining factor for morphological markings and disregards the potential influence of other aspectual factors (Sharma & Deo, 2009). However, in languages featuring grammatical aspect morphology, this morphology indicates the aspectual category of sentences more broadly, rather than being limited to specific lexical verbs or predicates (Sharma & Deo, 2009).

Meanwhile, while the grammatical aspect morphology is a naturally acquired feature for L1 speakers (Kihlstedt, 2002), it may not be readily unavailable to L2 learners.

More importantly, the AH is constrained by its limited explanatory power. While verb-inherent semantics by itself can account for the early stages of L2 tense and aspect development, particularly before morphological marking starts to appear (Kihlstedt, 2002), it falls short in explaining later stages of development. As learners progress to higher proficiency or end-state stages, they tend to depart from marking verbs for inherent aspect and lean towards a more native-like approach (Lardiere, 2003a; Roberts & Liszka, 2021). If the AH is universally and cross-linguistically valid, it is unable to explain the variable performance found among learners from different L1 backgrounds, which suggests an influence of L1 transfer (see studies reviewed below in Section 4.1.1).

As Gabriele et al. (2005) argued, the AH and its predictions have not specifically addressed the role of L1 transfer and the potential typological differences between learners' L1 and L2 being studied. The reason why potential L1 effects have not been studied systematically within the AH framework is that the varying acquisition patterns as a result of the L1 background, if observed, could arguably question the universality of the semantic effect and may highlight a greater role for L1 influence than previously acknowledged (Collins, 2004). Consequently, the impact of a learner's first language on their acquisition of L2 tense and aspect has been largely overlooked and remains a wide open question (P. Li & Shirai, 2000; Slabakova, 2002). Shirai (2009) thus stressed the importance of systematically examining L1 influence to separate it from natural acquisition processes.

Since then, there has been an increase in empirical studies investigating the role of L1, directly or indirectly, employing a variety of tasks and designs. The following section will present an overview of some relevant studies with evidence indicating a pronounced L1 influence even at high proficiency levels or a restricted L1 influence up to a certain proficiency level proficiency in the acquisition of L2 tense-aspect.

4.1.1 L1 Influence

The influence of L1 on the acquisition of the L2 is one of the most researched topics in the field of SLA, and it is also one of the two main issues, along with access to Universal Grammar (UG), being investigated within the generative framework (Slabakova et al., 2020). Most generative researchers agree that one clear difference between learning an L1 and L2 is prior knowledge; when we start learning an L2, we already possess knowledge of our L1 (Slabakova et al., 2020). This leads to an empirical question about transfer: How much does the L1 grammar serve as the starting point for a learner's representation of L2 grammar? (Lardiere, 2012). Considering the logical possibilities, three different scenarios can be posited. The first possibility is known as *Full Transfer* (Schwartz & Sprouse, 1996), which argues that the initial state of L2 acquisition is entirely composed of the morphosyntactic categories and features of the L1 grammar. The second scenario is *Partial Transfer*, where only some properties of the L1 transfer into L2. For example, the Minimal Trees Hypothesis (and recent Organic Grammar Approach) (Vainikka & Young-Scholten, 1994, 2011) suggests that only lexical categories are transferred from L1 at the initial state, whereas functional categories (e.g., case and tense) are not and would have to be acquired in L2. The third possibility is that there is *No Transfer* of the L1's properties to the L2 (Epstein et al., 1996). Essentially, these three proposals regarding transfer tend to focus on the initial state of L2 acquisition and concentrates on beginning or early-stage learners (Slabakova et al., 2020).

Another core issue within the generative framework is the accessibility of UG. UG posits that humans are born with a built-in capacity for language acquisition, and this innate faculty enables children to learn their first language quickly, often with minimal exposure and input (White, 2003c). This concept was proposed to explain the *poverty of the stimulus* problem, which observes that the language input children receive is insufficient to explain their sophisticated understanding of grammar (White, 2003c). Children acquire grammatical knowledge that goes far beyond the examples they hear, suggesting that certain grammatical principles must be innate rather than learned from the environment. Similarly, L2 researchers have begun to question whether UG remains accessible to L2 learners. As before, three logical positions can be considered. The first possibility is known as *Full Access* to UG (Schwartz & Sprouse, 1996), which argues that L2 learners have complete access to UG, just as they did when learning their L1. Their L2 development can be guided and restructured by universal principles, allowing them to potentially achieve native-like proficiency (Slabakova et al., 2020). In this view, UG equally facilitates the acquisition of complex grammatical knowledge in both L1 and L2 contexts (White, 2003d). The second position is *Partial Access* to UG, which suggests that L2 learners have limited access to UG in a way that only those aspects encoded in their L1 can be tapped (White, 2003d). This position predicts a significant role for the L1, with principles not activated by the L1 remaining inaccessible to L2 learners. The third position is known as *No Access* to UG, and proponents of this view (e.g., Bley-Vroman, 1989) believe that L1 and L2 acquisition are fundamentally different in their underlying learning mechanisms, with the former deriving from UG and the latter determined by non-linguistic processes. This perspective closely echoes the Critical Period Hypothesis debates at the time (Rothman et al., 2018).

The question about the role of L1 transfer and access to UG is highly connected, and the data needed to determine whether adults still access UG may depend on understanding how the L1 influences not only the early stages of acquisition but also beyond and up to ultimate attainment (Rothman et al., 2018). The rest of this section will review empirical studies specifically investigating the influence of L1 on L2 tense-aspect acquisition among learners at both developmental and late stages of learning.

Focusing on upper-beginner learners, Collins (2004) explored the acquisition of the English simple past tense by learners whose L1s were Japanese (*n* = 120) and French (*n* = 91). This study used 25-passage rational cloze tests and elicited a total of 7784 predicates from the participants. The findings revealed that both groups performed similarly in correctly using the English simple past tense in required contexts. However, differences emerged in their use of non-past responses. The French learners frequently used the present perfect form (77.2%) in past simple contexts, whereas Japanese learners did so less often (36.6%). The author ascribed the results of the French learners to L1 transfer, as French uses the *passé composé* to express both past simple and present perfect meanings.

Similarly, González and Quintana-Hernández (2018) investigated the use of past tenses by upper-beginner learners of L2 Spanish from L1 English (*n* = 22) and Dutch (*n* = 31) backgrounds using a story-retelling production task. The participants were asked to reconstruct the movie story using past tenses. The results revealed that the English learners overused past simple (Preterit) in both perfective and imperfective contexts compared to the Dutch and native groups. This overuse can be explained by the fact that English does not grammatically distinguish between perfective and general imperfective in the past. The Dutch group instead was the only group who used present perfect in perfective contexts, as present perfect in Dutch allows for both perfect and perfective interpretations.

Izquierdo and Collins (2008) investigated the acquisition of L2 French by lower-intermediate learners whose L1s were English (*n* = 15) and Spanish (*n* = 17). Using a cloze test focused on perfective and imperfective contexts, the study found that the English learners were significantly less accurate in using imperfective morphology compared to their Spanish counterparts. English learners defaulted to the perfective form, likely due to the absence of grammatical encoding for imperfective aspect in English. In contrast, the Spanish learners could rely on L1-L2 similarities, as both Spanish and French grammatically distinguish between the perfective and imperfective aspects.

Shifting the focus to non-Indo-European target languages, Duff and Li (2002) examined the acquisition of L2 Chinese perfective aspect by L1 English learners ($N = 9$). They employed two oral production tasks (story retelling and personal narratives) and one written production task in their study. The results showed that the English participants, particularly those at lower proficiency levels, overused and oversupplied the perfective aspect marker *-le* in ungrammatical contexts across all three tasks. The researchers argued that, unlike Chinese, English inherently encodes past tense grammatically. As a result, these learners showed a strong tendency to mark past events with the perfective marker due to their sensitivity to grammatical marking of past events in English.

While there is much more research in this area (e.g., Slabakova, 2000), cumulative evidence among early-stage L2 learners in tense-aspect acquisition seems to support the *Full Transfer* hypothesis. This hypothesis, which argues that the initial state of L2 acquisition is fully guided by L1 grammar, remains highly influential in generative SLA studies (Slabakova et al., 2020). However, mixed findings have emerged regarding whether L2 learners can overcome the initial L1 transfer effect in subsequent developmental stages and ultimately acquire aspects of the L2 that are not present in their L1.

For example, McManus (2015) compared upper-intermediate English ($n = 38$) and German ($n = 37$) learners of L2 French using a picture-based oral description task. Although neither group achieved target-like performance, their production patterns differed significantly. The German learners did not exhibit clear patterns when using perfective and imperfective forms, in contrast to the English group, who faced challenges mainly with habitual contexts but not progressive ones. Also, among the German learners, there was a tendency to prefer the compound perfect tense with most lexical verbs, regardless of the specific context. These differences indicate a potential influence of L1, as aspect is realised grammatically in English and French but not in German, leading to facilitative effects for the English group.

Gujord (2013) studied the acquisition of the past tense and present perfect in L2 Norwegian among upper-intermediate learners with L1 backgrounds in Vietnamese (n = 99) and Somali (n = 97). The written texts produced by the participants were extracted from a Norwegian learner corpus (ASK) and analysed. The results of the study deviated from the general predictions of the Aspect Hypothesis since it was observed that the learners did not consistently inflect telic verb phrases before atelic verb phrases. More importantly, the study found that the Somali learners, whose L1 does not grammatically encode present perfect, encountered greater difficulties in distinguishing between the past simple and present perfect forms compared to their Vietnamese counterparts. This difficulty was evidenced by the frequent occurrence of incorrect usage of past simple in contexts where the present perfect was more appropriate.

Moving to advanced proficiency, Hawkins and Liszka (2003) focused on L2 past tense marking among advanced Chinese (n = 2), Japanese (n = 5), and German (n = 5) learners of English. They collected spontaneous oral production data through a short film story retelling task and a personal recount of a happy or exciting experience. Analysis focused on verbs in unambiguous and compulsory past tense contexts. The results showed that Japanese and German learners employed regular simple past tense morphology in 91.9% and 96.3% of obligatory contexts, respectively. In contrast, Chinese learners correctly inflected verbs in only 62.5% of these contexts, a significantly lower rate. This difference was unlikely due to the influence of L1 phonology, as both Chinese and Japanese lack word-final consonant clusters in their L1s, and therefore, if phonology were the determining factor, Chinese and Japanese learners would have exhibited similar performance. Also, it was unlikely due to performance problems in oral production, as in that case, all three groups would have displayed similar levels of optionality across different contexts. The authors, thus, concluded that the reason Chinese speakers displayed a higher degree of optional usage in their production was due to the influence of their L1, that is, Chinese lacks morphosyntactic tense marking, while Japanese and German grammatically encode tense.

Similarly, Liszka (2004) focused on the same L1 participants (2 Chinese, 5 Japanese, and 5 German speakers) but examined their production of the English present perfect. The results revealed that all three L2 groups performed similarly in obligatory contexts (47.1% for the Chinese, 55.9% for the Japanese, and 58.3% for the German), with no statistically significant differences. However, differences emerged in the use of non-present perfect forms. Among Japanese participants, there was an alternation between past simple (54.9%) and present tense usage (38.1%). Similarly, Chinese participants also displayed this alternation, with 53.7% using past simple and 46.3% using present tense. In contrast, the German group exhibited a strong preference for past simple (81%), while only 14.5% used present simple. Liszka therefore concluded that since German does not realise perfect/non-perfect distinctions and establish current reference time through grammatical means, this influence of L1 led German learners to relying on using past simple in present perfect contexts. However, any generalisable claims from the above two studies should be made with caution due to the small sample size.

Liszka (2009) also investigated the acquisition of English present simple and present progressive among advanced French learners of English ($N = 16$). The study utilised two oral production tasks (picture description and video clip description) and one written task (contextualised dialogue completion). In the picture description task, the participants demonstrated a high level of success in assigning target-like meanings to present simple forms (100%) and present progressive forms (92.7%). In the video clip description task, participants produced correct present simple forms in 94.9% of the present simple contexts, while they only provided present progressive forms in only 54.8% of the obligatory contexts. This pattern was also reflected in the written task, where participants scored 94.2% for correctly providing present simple forms in the mandatory present simple contexts but only 59.3% for supplying present progressive forms in the progressive contexts. The results suggested a tendency to overuse the present simple in progressive contexts, likely due to the influence of the French présent, which allows for both present simple and present progressive interpretations. This L1 influence may pose challenges even for advanced French learners of English.

Gabriele et al. (2005) examined whether advanced Japanese learners (*N* = 83) could accurately interpret the past progressive form in L2 English when expressing aspectual information. In Japanese, the aspectual marker *te-iru* can convey either a progressive or perfective meaning, depending on the specific verb's lexical semantics and the context of the event. In contrast, in English, the construction *be + ing* typically indicates ongoing progress regardless of the verb's lexical semantics. The results from the sentence interpretation task indicated that participants' accuracy rates were significantly lower in the past progressive contexts (e.g., around 52% with activity verbs and 56% with change of state verbs) than in the simple past contexts (e.g., around 78% with activity verbs and 83% with change of state verbs). However, since no performance difference was found between the two verb classes, the authors concluded that the learners' difficulty with the past progressive was not due to L1 transfer of lexical semantics but rather from mismatches in the associations between linguistic forms and their corresponding meanings in the L1 and L2.

Alruwaili (2014) conducted a study among L1 Arabic learners of English (*N* = 55) using an acceptability judgment task (AJT) and a gap-filling task. The findings revealed that the Arabic learners, regardless of their proficiency levels (intermediate vs. advanced) and learning contexts (immersion vs. classroom), did not accept the present perfect for the intended contexts (mean ratings = 1.3-1.9) compared to the native group (mean rating = 2.6) in the AJT task. This pattern was reflected in their written production data, where the Arabic groups scored only 32%-42% for correctly providing present perfect forms in obligatory contexts, while the native group scored 70.5%. The research showed that Arabic learners over-supplied past simple forms in present perfect contexts. This issue persisted even among immersive participants, indicating a strong influence of L1 grammar. As reviewed in Section 3.1 and 3.3, in Arabic, perfective forms can encode both past simple and present perfect interpretations, and present perfect meanings are usually achieved through adverbials and contexts. In contrast, English grammatically distinguishes between past simple and present perfect, which likely contributes to the difficulties Arabic learners face in accurately using these forms in English.

Diaubalick and Guijarro-Fuentes (2019) compared German (n = 30) and Romance (n = 30) learners (French, Italian, and Portuguese as L1s) of L2 Spanish at highly advanced levels using an AJT task. The results indicated that German learners marked ungrammatical sentences involving mismatches between adverbials and preterit/imperfect verb forms as acceptable, suggesting they struggled to differentiate between perfective and imperfective aspects. This difficulty might arise from their L1, which relies on lexical means rather than grammatical distinctions. On the other hand, Romance learners benefitted from similarities between their L1 and L2 systems, accelerating their acquisition of these aspects.

The studies reviewed above provide evidence of the persistence of L1 influence among learners, not only at developmental stages but also at very advanced stages where further development is unlikely. These findings seem to align more with the *Partial Access* to UG hypothesis, which suggests that learners only have access to UG principles encoded in their L1 and are unable to reconstruct L2 grammar not present in their L1, regardless of their proficiency. Nevertheless, other studies reveal that L2 learners can overcome L1 influence as their proficiency increases and reconstruct L2 grammar based on received input, providing evidence for the *Full Access* to UG hypothesis.

For instance, Gabriele and Maekawa (2008) extended Hawkins and Liszka's (2003) work by comparing the usage of English present and past progressive forms among L1 Chinese (n = 32), Japanese (n = 55), and Korean (n = 18) learners. The participants were categorised into intermediate and advanced proficiency levels. Unlike Hawkins and Liszka, who used production tasks, Gabriele and Maekawa employed a story interpretation task, where participants judged the compatibility of given sentences with a story they had listened to. The results revealed that both intermediate and advanced Japanese and Korean learners, as well as advanced Chinese learners, showed target-like performance. Interestingly, intermediate Chinese learners often incorrectly accepted present progressive forms for both ongoing present events (incomplete contexts) and past (complete contexts) situations. These findings suggest that the learners' L1 did indeed play a role in determining the path of L2 development; however, this influence appeared to be limited to the intermediate level of proficiency. Therefore, the authors concluded that tense acquisition is ultimately achievable, irrespective of whether one's L1 grammatically expresses tense or not. However, differences in findings between this study and Hawkins and Liszka's (2003) could be due to the nature of the tasks; oral production tasks, being more cognitively demanding and influenced by real-time pressure, may affect performance differently (Grüter et al., 2012).

Another study, conducted by Slabakova (2003), examined whether L1 Bulgarian learners of English (N = 112) within an instructed environment were able to acquire simple present and present progressive tenses, which are not grammatically instantiated in their L1. Based on the Michigan English Test (grammar) scores, the learners were categorised into low intermediate, high intermediate, and advanced proficiency levels. The elicited production task results showed that learners at all proficiency levels could accurately produce target-like forms for both tenses in obligatory contexts. Also, findings from the truth-value judgement task indicated that the Bulgarian learners improved their ability to associate forms with their appropriate semantic meanings (i.e., present progressive with continuous interpretations and present simple with habitual interpretations) as their proficiency increased. In particular, advanced learners demonstrated target-like mastery, indicating successful acquisition of the aspectual features and their semantic values.

Al-Thubaiti (2015) investigated the acquisition of the English aspectual contrast (habitual and progressive) by L1 Arabic L2 English learners (*N* = 143). The participants were divided into different proficiency level groups (beginner, low-intermediate, high-intermediate, and advanced) by their cloze test scores. A story interpretation task was used to examine whether the Arabic learners could distinguish English simple present and present progressive meanings, since their L1 does not overtly instantiate the contrast like English and their imperfective verb forms allow for both habitual and progressive interpretations. The results revealed a significant effect of L2 proficiency, with learners gradually establishing associating the simple present verb form (*V-s*) with habitual actions and the progressive form (*be + V-ing*) with ongoing events as proficiency increased. By high-intermediate and advanced levels, learners exhibited native-like performance. Unlike Liszka (2009), the Arabic learners in this study, regardless of their proficiency levels, did not show over-generalization in either aspectual context for either of the verbal patterns. Al-Thubaiti interpreted the results as indicative of positive L1 transfer, possibly due to the presence of the grammaticalized form *gaa'ad* in spoken Arabic dialects, or because the aspectual or semantic properties of the grammatical forms were not significantly influenced by L1 transfer. The differences in findings between this study and Liszka's (2009) could be once again explained by the use of different tasks. While this study employed an interpretation task, Liszka (2009) used oral production tasks. It is possible that different types of knowledge are assessed in various tasks, leading to disparate results.

Gabriele and McClure (2011) focused on the interpretation of the imperfective marker *te-i* in Japanese L2 by advanced L1 Chinese learners (*N* = 46). The results of the story interpretation task showed that the Chinese learners had difficulty in interpreting the past perfective marker correctly. The authors argued that this difficulty could not be solely attributed to L1 influence but was also due to the grammatical complexity of the marker itself. Evidence from Gabriele's (2009) study showed that even advanced native English speakers, whose L1 does encode tense, struggled with the Japanese past imperfective marker. The study suggested that positive transfer from L1 was not guaranteed and that while early learning stages might be influenced by L1 differences, these differences may have less impact on advanced learners.

In summary, the reviewed studies provide substantial evidence of L1 influence on L2 tense and aspect acquisition. However, a consensus has yet to be reached concerning the degree of L1 transfer in this process and whether it is possible to eventually acquire L2 features that do not exist in one's L1. Differences in findings across studies might be attributed to the types of tasks employed, as different tasks tend to measure different types of L2 knowledge (i.e., implicit vs. explicit). This task effect will be discussed in more detail in Section 4.2.2.

Given the conflicting evidence, different theoretical accounts within the generative framework regarding L1 transfer and access to UG have emerged. These theoretical perspectives will be discussed in the next section.

4.2 Theoretical Accounts

As reviewed above, studies have shown evidence of persistent difficulties encountered by learners with certain L1 backgrounds. For example, L1 Chinese learners of English are found to exhibit a higher degree of optional usage in their production of past tense morphology compared to L1 Japanese and German learners (Hawkins & Liszka, 2003). Various theories have been proposed to account for this phenomenon within the generative framework. These theories can be divided into two main positions: one supports the idea of a non-representational deficiency, while the other argues that there is a deficiency in the representation.

The non-representational deficit position assumes that L2 learners have full access to UG and L2 development can be guided and restructured by UG (Schwartz & Sprouse, 1996). This position is sometimes referred to as Full Access Accounts. One hypothesis under this approach is the Full transfer/Full access hypothesis proposed by Schwartz and Sprouse (1994, 1996). According to this hypothesis, learners' L1 grammar (except the phonetic structures related to lexical/morphological elements) constitutes the initial state of the L2 (i.e., Full Transfer). However, they are not bound to maintain these L1 values permanently (known as fossilization), and parameter resetting to the L2 values can take place when learners are exposed to L2 input that clash with their L1 grammar, at least in principle (i.e., Full Access). This hypothesis provides an explanation for both the initial state of the L2 and the subsequent evolution of the L2 grammar (Slabakova et al., 2020). Another hypothesis under the FAAs is the Missing Surface Inflection Hypothesis (MSIH) (Prévost & White, 2000a, 2000b), which argues that while L2 learners have full access to UG, they are also constrained by some external factors such as processing limitations. In this hypothesis, L2 learners are able to develop target-like features and linguistic mental representations, and the difficulty in their production arises under communication pressure/performance issues. Some other external or individual-level factors might include the influence of L1 prosodic representation, captured by the Prosodic Transfer Hypothesis (Goad & White, 2006), complexity involved in reassembling features into L2 bundles, captured under the Feature Reassembly Hypothesis (Lardiere, 2009), and the greater difficulty in acquiring functional morphology than acquiring syntactic structures and semantics, captured by the Bottleneck Hypothesis (Slabakova, 2008). These hypotheses share the idea that attaining native-like L2 representations is possible in principle and any difficulty encountered by the learners is the surface outcome of other various variables (Amaro et al., 2018).

On the other hand, the representational deficit position assumes that UG becomes partially inaccessible to L2 learners after the critical period, leading to partial or limited access (Rothman & Slabakova, 2018). In line with this position, a few hypotheses have emerged, including the Failed Functional Features Hypothesis (Hawkins & Chan, 1997), the Representational Deficit Hypothesis (Hawkins & Liszka, 2003), and the more recent version the Interpretability Hypothesis (Hawkins & Hattori, 2006; Tsimpli & Dimitrakopoulou, 2007). These hypotheses argue that learners' mental representations of their L2 are, in a way, hindered or handicapped due to their inability to access UG features beyond those present in their L1 (Tsimpli & Dimitrakopoulou, 2007). In other words, the non-target-like performance of learners can be attributed to the impairment of their mental representations. More recently, the Morphological Congruency Hypothesis (Jiang et al., 2011) has been proposed to explain non-target-like performance in real time processing, suggesting that achieving nativelike competence is impossible if the target morpheme is absent in the learners' L1.

The following sections will further discuss the two opposing positions by providing a detailed examination of the contrasting hypotheses. The supporting and opposing evidence for each position will also be presented.

4.2.1 Full Access Accounts

The Full Access Accounts (FAAs) maintain that adults still retain access to the full range of UG features as in L1 acquisition (Schwartz & Sprouse, 1996). According to this perspective, difficulties experienced by advanced L2 learners, particularly in the realm of functional morphology, are not attributable to an innate inability to acquire specific linguistic features imposed by maturation, i.e., after critical period. In other words, achieving native-like L2 syntactic representations is theoretically possible, and any observed difficulty among advanced learners can be attributed to limitations at the output level, such as the mapping problem between underlying representation and its surface realisation, as captured by the Missing Surface Inflection Hypothesis (MSIH).

The MSIH, proposed by Prévost and White (2000a, 2000b), partially aligns with the Full transfer/Full access (Schwartz & Sprouse, 1996) perspective, which argues that UG remains fully accessible to L2 learners. This accessibility allows learners to reconstruct their L1 grammar to align with the L2 grammar, guided by the input they receive. However, unlike the Full Transfer/Full Access, the MSIH posits that the inconsistent usage of L2 morphology in oral production does not indicate an impaired underlying syntactic representations among learners, and it is simply a mapping problem from abstract representations to the morphological forms on the surface (Prévost & White, 2000b).

Prévost and White (2000b) provided evidence for the MSIH by examining spontaneous production data collected from two adult L1 Arabic learners of French over three years, and two adult learners of German with L1 Spanish and L1 Portuguese backgrounds over two years. By analysing their speech data, the researchers found that: 1) non-finite forms were used by the learners in both finite and non-finite contexts, but finite verbs rarely occurred in non-finite contexts (below 8%); and 2) when agreement was present, it was mostly correct (above 87%). Based on these findings, Prévost and White concluded that the finite feature is present in these learners' underlying grammar rather than being impaired. This was evidenced by the fact that they rarely produced 'faulty' agreement with mismatches between person, number, and gender inflection on a verb and the associated subject. They argued that if learners did not have abstract representations for finiteness and agreement, they would show much greater variability in their production, such as finite verbs occurring in non-finite contexts and non-finite verbs in finite positions. Therefore, the learners' problem is argued to lie in the difficulty with the realisation of surface morphology.

Using the Distributed Morphology framework proposed by Halle and Marantz (1993), Prévost and White explained why non-finite verbs, as the default form, were used by the learners in the finite contexts. Within this framework, inflected forms come with features such as tense, person, number, and gender. To insert a word into a sentence, its features must align with those of the target syntax node. While syntactic nodes have fully specified features, lexical items may lack or partially specify certain features. In cases of non-exact matches, there is a competition among potential candidates for insertion, with the form closest in feature alignment to the terminal node winning. In this context, non-finite verbs are considered default forms, lacking specific details regarding finiteness, unlike other verb forms that come with such specifications. When a learner needs to produce a finite verb, it is assumed that the features linked to the terminal node in the syntactic structure are fully detailed. As a result, the lexical form possessing the highest degree of matching specifications is anticipated to be retrieved for the subsequent feature-checking process. However, Prévost and White suggest that, occasionally, due to processing constraints or communicative demands in oral production, there may be a failure in selecting the appropriate form, resulting in the insertion of a less specified form instead.

Another study supporting MSIH was conducted by Lardiere (1998a, 1998b), who collected oral production data (through naturalistic conversations) from a native Chinese speaker named *Patty* at three different time points over a time period of nine years. Patty had been living in the US for ten years when the first recording was taped and was assumed to have end-state L2 grammar due to long time of immersion in the L2 environment. However, analysis of her production data revealed that her suppliance of past tense marking in obligatory contexts was consistently low and stable across the three time points (only 34.78%, 34.85%, and 33.82%). In contrast, her performance in related morphosyntactic areas, such as nominative case marking in finite past tense contexts, was perfect (100% at the three time points). In addition, her provision of the correct forms of the copula and auxiliary *be* with agreement features appeared to be target-like (83%, 94%, and 94% at the three time points).

According to Lardiere, the omission of past-tense marking should not be interpreted as evidence of impaired syntactic representations. As outlined in Chomsky's (1995) framework, auxiliary and copula *be* verbs share an agreement feature akin to main verbs in contexts involving 3rd person singular agreement, and also nominative case assignment to the subject is facilitated by the tense head. Therefore, if Patty lacked syntactic categories for tense and agreement, she would not have demonstrated such proficient performance in nominative case assignment and the correct use of *be* forms.

In Lardiere's (2007) follow-up study, Patty's written samples were collected over six years through emails. In contrast to her spoken data (around 34%), her usage of past-tense markers in her email communication was significantly higher, with a suppliance rate of around 78% in obligatory contexts. This suggests that non-syntactic factors come into play during real time speech production. As Lardiere (2000) argued, "the problem lies in figuring out how (and whether) to spell out morphologically the categories they already represent syntactically, i.e., the 'mapping problem'" (p. 121).

Adopting a similar longitudinal design, White (2003) examined the spontaneous production data form an L1 Turkish learner of L2 English named *SD*. At the point of data collection, SD had been living in Canada for ten years. She scored 93.75% in the ELI placement test, indicating advanced proficiency. The researcher recorded five interviews with SD, with an 18-month interval between the first four recordings (Time 1) and the fifth one (Time 2). It was found that SD's suppliance of agreement and past tense on lexical verbs was very high, at 85% and 76% in both times. This performance was noticeably higher than the L1 Chinese speaker Patty's production in Lardiere's (1998a, 1998b) studies.

The comparison between Patty and SD's performance raises questions about the interpretation proposed by the MSIH. According to this hypothesis, there should be no significant differences among L2 learners. However, the results from Lardiere's and White's studies reveal a notable discrepancy. This discrepancy suggests that the MSIH may fall short in explaining why the persistent difficulty exists at advanced proficiency levels for some learners but not for others. As a result, supporters of the Full Access Approaches have put forth some supplementary hypotheses to better understand and explain the residual variation in learners' performance.

For instance, the Prosodic Transfer Hypothesis (PTH), proposed by Goad and her colleagues (2003), argues that the influence of learners' L1 prosodic structure is responsible for their failure to produce certain inflectional morphology accurately. Since Chinese and English differ in how morphology is prosodified, Goad et al. (2003) believed that this difference plays a role in explaining the inconsistent use of the past tense *-ed* morpheme among Chinese learners of English. Although the transfer of L1 phonological structures can cause difficulties in learners' oral production, this can be overcome ultimately by learners through reconstructing the prosodic representations for L2 using L1 structures (Goad & White, 2006). More recently, Lardiere (2008, 2009) built on the Full Transfer/Full Access model (Schwartz & Sprouse, 1996) and proposed the Feature Reassembly Hypothesis (FRH). The FRH argues that the difficulty encountered by L2 learners comes from the process of assembling L2 features and re-establishing connections between these features and their morphological realisations. Reassembling features can be particularly challenging when the features are present in the L1 but configured differently. In such situations, L2 learners must disassociate the feature that they have selected in their L1 and then reconstruct them to align with L2 morpholexical items. The FRH emphasises that L2 feature configurations are ultimately possible when learners have access to positive evidence in the L2 input (Lardiere, 2009).

In summary, Full Access Accounts maintain that while L2 acquisition might differ from L1 acquisition on many levels, UG remains fully accessible to L2 learners. They claim that learners' underlying syntactic representations are intact, even when L2 production deviates from the target language in certain situations. When learners are able to use a form productively, it suggests that they have acquired the underlying linguistic representation. Any variability or inconsistency in their usage can be attributed to issues related to the production or output. These problems may arise from difficulties in mapping mental representations to surface morphological forms, the complexity involved in reassembling features into L2 bundles, or negative transfer of L1 phonological structures. Despite these variables, the ultimate conclusion is that nativelike L2 syntactic representations are attainable.

4.2.2 Representational Deficit Accounts

In contrast to the Full Access Accounts, the Representation Deficit Accounts argues that persistent difficulties with certain morphosyntactic properties in end-state learners indicate that their syntactic representations may be compromised or impaired. According to this perspective, some linguistic features may undergo a critical period after which they become unattainable (Hawkins & Liszka, 2003). Hypotheses under this account believe that L2 learners have only partial access to UG, with some grammatical principles being unavailable unless they can be transferred from their L1. This section will review the Representation Deficit Hypothesis and the more recent Morphological Congruency Hypothesis.

Representation Deficit Hypothesis (RDH). The RDH was developed within the Principles and Parameters framework (Chomsky, 1986). This framework assumes that the process of L1 acquisition involves setting the appropriate value of the parameter based on linguistic input. When L2 acquisition results in different parametric values than those established in the L1, learners face challenges because a reset to L2 values is not possible (Tsimpli & Roussou, 1991). Building on this idea, Smith and Tsimpli (1995) extended the theory by emphasising the parameterisation of functional categories. They argued that after childhood, L2 learners are unlikely to acquire new functional categories, as these categories become fixed after a critical period (Snape et al., 2009).

Hawkins and Chan (1997) tested this theory by investigating whether Chinese learners of L2 English could acquire English operator movement in restrictive relative clauses (e.g., *The person who called you is waiting outside*), a feature absent in Chinese. Using an AJT task, they found that the Chinese learners were significantly different from the French learners, whose L1 has the [wh] feature, even at advanced levels. Based on these findings, Hawkins and Chan argued that L2 learners have impaired underlying representations compared to native speakers. They suggested that when L1 functional features are different from those of the L2, fossilization occurs, and reaching native-like attainment becomes impossible due to a deficit related to L1 transfer.

Tsimpli and Dimitrakopoulou (2007) further refined this theory by proposing that the persistent difficulties L2 learners experience are mainly for uninterpretable syntactic features, rather than interpretable ones. In regards to the distinction between interpretable and uninterpretable features, Chomsky (1995) explains that interpretable features convey conceptual meanings that can be understood by our Conceptual-Intentional system (i.e., it adds to the meaning of a word, essentially contributing to the semantic interpretation). For example, Number and Person are considered interpretable features, since there is a difference in semantic meaning between the pronoun *I* (referring to a first-person-singular subject) and *they* (referring to a third-person-plural subject) (Radford, 2004). In contrast, uninterpretable features do not directly contribute to semantic interpretations but instead indicate grammatical functions (Slabakova et al., 2020). For example, case is considered an uninterpretable feature, as variations in case marking (such as nominative, accusative, and genitive) do not change the underlying semantic meaning. Despite different cases in *they*, *them*, and *their*, all refer to the same third-person-plural subject without changing the core meaning (Radford, 2004).

According to Tsimpli and Dimitrakopoulou (2007), uninterpretable features, except those already present in learners' L1 grammar, cannot be modified or acquired beyond the critical period. As a result, this deficit in interlanguage grammars results in persistent challenges for L2 learners in acquiring certain features. This refined version of the proposal has been termed the Interpretability Hypothesis (IH) (Tsimpli & Dimitrakopoulou, 2007; Tsimpli, 2003) or the Representational Deficit Hypothesis (RDH).

Hawkins and Liszka's (2003) study, as reviewed Section 4.1.1, is one of the key studies in support of the RDH. Their research revealed that advanced Chinese learners had a lower accuracy rate (62.5%) in producing English past tense compared to their Japanese (91.9%) and German (96.3%) counterparts. Given that learners' proficiency levels were matched, the lower performance of the Chinese learners was likely due to L1 transfer constraints. This is because Chinese does not grammaticize past tense, whereas Japanese and German do. Hawkins and Liszka further argued that, because Chinese learners do not instantiate the uninterpretable feature [+past], the English representation for past tense is therefore absent from their interlanguage grammar. This results in an impaired L2 mental representation, leading to persistent difficulties even at advanced proficiency levels. However, one might question why Chinese learners, if they truly lacked the relevant syntactic feature, did not show a complete absence of past tense marking (i.e., a success rate of 0%). Hawkins and Liszka addressed this by suggesting that "linguistic theory appears to need to allow operations which apply to strings postsyntactically, that is in the morphological component or following vocabulary insertion" (p. 38). This means that there is a stage where the language learners' output is monitored and checked before finalisation. The process of output checking triggers the inclusion of the inflected verb form (*V-ed*) among Chinese leanrers when they can identify and monitor the ongoing conversation for indications of past events. However, this monitoring process is highly context-dependent, leading to variable use of both inflected and uninflected verb forms based on the speech context.

Another study providing support for the RDH was conducted by Tsimpli (2003), who examined the acquisition of L2 Greek by L1 Turkish and Russian learners ($N = 6$). At the time of data collection, these participants had been living in Greece for around eight to nine years. Oral production data were elicited via a 45-minute interview. Based on the similarities and differences between the participants' L1s and L2, the RDH predicts that: 1) the use of definite determiners would present challenges for L2 learners when compared to their use of indefinite determiners. This difficulty arises from the interpretability distinction, as well as the fact that the uninterpretable features required for definite determiners are not accessible in the participants' L1s; 2) L2 learners would be expected to face difficulties with pronominal 3rd person object clitics, in contrast to their handling of 1st and 2nd person clitics. This difficulty is attributed to the same reasons as above; 3) verbal morphology should pose fewer problems to L2 learners, as they are already present in the learners' L1s. The results from the oral production data confirmed all three predications. Specifically, a statistically significant difference was found in the use of indefinite and definite articles among all six participants. The accuracy rate for indefinite articles was high, averaging around 90%. In contrast, the accuracy rate for the use of definite articles was lower, around 60%, with only two participants achieving around 70%, while one participant scored below 15%. In addition, the participants' demonstrated high accuracy in using 1st and 2nd person clitics, with an average accuracy of around 90% and three participants exceeding 95%. However, their performance with 3rd person clitics was less consistent: none of the participants scored above 75%, and three participants scored below 40%. Lastly, in the domain of verbal morphology, all six participants displayed native-like performance and performed at ceiling (100%). Thus, the author argued that these results align with the RDH as difficulties only emerged in properties associated with uninterpretable features that were not present in the participants' L1s.

Hawkins and Hattori (2006) proposed two potential explanations for why uninterpretable features may be subject to a critical period while interpretable ones are not. One explanation is related to functional economy. Specifically, interpretable features are essential for daily communication and the ongoing construction or learning of new lexical items throughout a person's life. On the other hand, uninterpretable features are typically associated with a limited set of closed class items, which belong to functional categories in the language. The main function of these features is to establish stable connections or dependencies between different elements within syntactic structures. Given their specialised role, maintaining access to all uninterpretable features from the UG may not be functionally advantageous. The second explanation is rooted in brain anatomy. Hawkins and Hattori explained that retaining access to all uninterpretable features that are no longer essential for adult grammar incurs a significant energy cost. They proposed the possibility of an "energy efficiency constraint" that disconnects components not directly needed for cognitive functioning after a certain period has elapsed (Hawkins & Hattori, 2006, p. 272).

Nevertheless, some studies have provided evidence against RDH and argued that L2 learners can still exhibit a native-like pattern even if the relevant property is not present in their L1s. For instance, White et al. (2004) investigated the acquisition of L2 Spanish grammatical gender by adult learners (n = 68) of a genderless language (i.e., English), compared with French learners (n = 48) whose L1 has grammatical gender. In Spanish, nouns are categorised as either masculine or feminine, and also the determiners (articles such as "a," "an," or "the") and adjectives must agree or 'match' with the grammatical gender of the noun. This agreement is triggered by uninterpretable features. White and her colleagues categorised the participants into three proficiency levels (i.e., low, intermediate, and advanced), and administered oral production tasks and a picture identification task to the participants. The analysis focused on both interpretable gender features on nouns and uninterpretable features on determiners and adjectives. The results showed that advanced and intermediate L1 English learners performed very well in both production and picture identification task, showing no significant differences from the L1 French counterparts and the native Spanish speakers. White and her colleagues, thus, argued that the findings provide strong evidence in favour of the idea that post-puberty learners can acquire gender agreement, irrespective of the presence of gender features in their L1s. This contradicts the RDH hypothesis, indicating that there is no obstacle in accessing "new" and uninterpretable formal features.

In summary, the RDH differs from the full access approach by arguing that L2 learners only have access to UG properties that are already present in their L1s (i.e., Partial Access) (Hawkins & Chan, 1997). This hypothesis acknowledges the predominant role played by the L1; that is, the absence of certain properties in the L1 can lead to a permanent deficit in the L2 mental representation of learners. This means that these learners may not be able to acquire these L2 properties after the critical period, resulting in a non-target-like mental grammar. While the RDH assumes an impaired mental representation, it does not predict a complete absence of L2 knowledge if the learners' L1 lacks relevant features (Amaro et al., 2018). Instead, these learners are likely to exhibit a higher degree of variability or optionality in their usage compared to the native speakers or learners whose L1 has these relevant features (Hawkins & Liszka, 2003).

Morphological Congruency Hypothesis (MCH). As reviewed in Section 4.1.1, mixed findings regarding the role of L1 transfer have been accumulated over the past years, offering support for both the Full Access Accounts and Representational Deficit Accounts. One factor that could explain these conflicting results, as argued by Jiang et al. (2011), is the research method employed in the studies. Some studies have collected longitudinal production data from participants over years, while others have relied on specific elicitation tasks, often limited to a single test session. These elicitation tasks have also shown a wide variety, including picture identification, sentence or story interpretation, grammaticality judgment, story-retelling, interviewing, cloze tests, written narratives, and more. However, what is particularly important here is that these tasks may vary in their reliance on participants' explicit knowledge (Jiang et al., 2011). Some tasks emphasise grammatical accuracy and encourage the application of explicit knowledge, thereby shedding light on participants' knowledge of a particular structure rather than their ability to use it spontaneously. Other tasks are designed to minimise explicit knowledge involvement focusing instead on more naturalistic language use, thereby tapping more into implicit knowledge.

The difference between explicit and implicit knowledge has been widely discussed within the field of SLA (R. Ellis, 2005). It is believed that implicit knowledge operates without conscious awareness, whereas explicit knowledge relies on some level of conscious awareness (Rebuschat, 2013; William, 2009). In other words, implicit knowledge is utilised effortlessly, often without speakers even being conscious of its presence; in contrast, explicit knowledge is applied consciously and with awareness (Godfroid et al., 2015).

Past SLA research has empirically demonstrated that implicit and explicit knowledge may be separate constructs that can be independently measured (e.g., Bowles, 2011; Ellis, 2005, 2009; Ellis & Loewen, 2007; Godfroid et al., 2015; Gutiérrez, 2013). For example, based on a confirmatory factor analysis, Ellis (2005, 2009) argued that time-pressured tasks (e.g., timed judgement task, oral narrative task, and elicited imitation task) were more likely to draw on participants' implicit knowledge, whereas untimed tasks (e.g., untimed judgment task, metalinguistic knowledge test) were loaded onto the explicit knowledge factor. This finding was subsequently replicated in several studies involving various L2 learner populations (Bowles, 2011; Sarandi, 2015; Zhang, 2015). However, some researchers (e.g., Suzuki, 2017; Suzuki & DeKeyser, 2015; Vafaee et al., 2017) have challenged this claim and argued that time constraints alone cannot adequately restrict access to explicit knowledge to guarantee the activation of implicit knowledge. They contend that highly advanced L2 learners may still access explicit knowledge under time pressure, and in such cases, it is the automatised explicit knowledge rather than implicit knowledge that is drawn on during the task (Suzuki, 2017a). From this perspective, implicit knowledge is better measured when participants process specific grammatical structures for real-time comprehension (Suzuki & DeKeyser, 2015). In these tasks, learners are required to utilise their linguistic knowledge almost immediately upon receiving input, with minimal likelihood of consciously applying that knowledge (Suzuki, 2017a).

In this sense, online psycholinguistic methods can serve as the best measures of implicit knowledge, because they can "gather information about sentence interpretation as each word or phrase is read or heard in real time" (Keating & Jegerski, 2015, p. 2), in contrast to traditional offline methods, where data regarding sentence interpretation is gathered after participants have been presented with the entire sentence. Examples of online psycholinguistic tasks measuring sentence processing include reaction time (RT) based tests (see Jiang, 2011 for a review), such as the self-paced reading task (Roberts & Liszka, 2013) and the word-monitoring task (Suzuki, 2017a). Recently, more fine-grained techniques, such as the eye-tracking task (Grüter et al., 2012) and the event-related brain potentials (ERPs) (Armstrong et al., 2018), have been used.

Based on the differences between implicit and explicit knowledge, the Morphological Congruency Hypothesis (MCH) has been proposed by Jiang and his colleagues (2011) to specifically address the influence of L1 on L2 learners' performance in online processing tasks. According to the MCH, languages vary in their grammaticalization (i.e., morphologically marked) of a specific meaning. When two languages both use morphological markers to express a particular meaning, they are considered to be morphologically congruent; on the other hand, when one language employs a grammatical morpheme to convey a meaning while the other does not, they are morphologically incongruent (Jiang et al., 2011). Regarding nativelike attainment, the MCH predicts that:

> "*When L2 learners reach an advanced or near-native level of L2 proficiency, only congruent learners (i.e., those whose L1 has a corresponding morpheme to the target L2 morpheme) are able to reach nativelike proficiency in acquiring an L2 morpheme. Incongruent L2 learners will find it extremely difficult, if not impossible, to develop nativelike competence with respect to the same L2 morpheme*" (Jiang et al., 2011, p. 943).

The MCH is based on the concept of automatic activation: when a learner demonstrates native-like mastery of a morpheme, they will spontaneously activate its meaning when encountering it (Jiang, 2007). Therefore, the MCH can only be tested using online processing tasks to minimise learners' access to explicit knowledge (Choi & Ionin, 2021; Ionin et al., 2021; Jiang et al., 2011). From this perspective, nativelike performance in offline tasks does not serve as proof that the learner has fully acquired the morpheme. Although there is ongoing debate about tasks that tap into explicit or implicit knowledge as discussed above, the MCH considers learners' ability to quickly activate the meaning of a morpheme online as a sign of nativelike performance (Choi & Ionin, 2021). Whether their performance draws on implicit or automatised explicit knowledge is not the primary concern here.

Jiang et al. (2011) provided supporting evidence for the MCH by focusing on the English plural marking by advanced L2 ESL learners. This study included 26 L1 Japanese leaners, 24 L1 Russian learners, and a control group of 26 English native speakers. A cloze test was administered to measures L2 participants' English proficiency, and no significant difference was obtained between the two groups' scores. This suggested that the two groups' L2 proficiency levels were comparable. A self-paced reading (SPR) task was employed by the authors, due to its advantages in minimising reliance on explicit knowledge. In the SPR task, participants read sentences one word at a time, with only a single word displayed on the computer screen at any time. Each participant reads at his or her own pace and presses a button to bring up the next word. Their reaction times on each word are recorded as a measure of real-time processing. The fundamental logic of SPR task is that native speakers tend to slow down in the part of a sentence containing an error (or the next spillover region) (1b), compared to a control condition having a grammatically correct sentence (1a) (Jegerski, 2014). Therefore, if L2 learners show similar patterns and sensitivity to the errors, they are considered to be showing nativelike performance.

(1) a. *The professor noticed a few of his <u>friends</u> in the picture.*

b. **The professor noticed a few of his* <u>*friend*</u> *in the picture*.

Based on the MCH, it was predicted that the L1 Russian learners would show sensitivity to errors, slowing down when encountering the missing *-s*, as Russian requires overt plural marking (morphologically congruent with English). The L1 Japanese would not show such sensitivity since Japanese does not obligatorily mark the plural (morphologically incongruent with English). This prediction was realised based on the results from the SPR task. The Russian speakers and the native English speakers both slowed down for plural errors, whereas the Japanese speakers showed no such pattern. Therefore, Jiang and his colleagues concluded that L2 learners can become sensitive to plural errors in real-time comprehension only when a congruent plural morpheme exists in their L1, providing support for the MCH.

After this study, Jiang et al. (2017) employed a different online method, namely the sentence-picture matching task, to further test the MCH. In this task, both participants' reaction times and accuracy rates were recorded. Once again, this study focused on English plural marking and included two L2 groups: an L1 Chinese group (n = 54) and an L1 Russian group (n = 26), matched in L2 proficiency. The English singular/plural distinction is a congruent feature for the Russian speakers but an incongruent one for the Chinese speakers. The results aligned with the predictions and revealed that the Russian group displayed sensitivity to the absence of plural marking, whereas the Chinese did not exhibit such sensitivity. What is particularly interesting in this study is that a follow-up experiment was conducted with the Chinese learners to test if they were able to activate number meanings in English that are expressed lexically (e.g., several, a few, many). It was found that they did show sensitivity to the mismatch items, as indicated by their significantly slower reaction times. Taken together, the performance of the Chinese learners showed that lexicalised L2 meanings were automatically activated during online processing, whereas grammaticalized

L2 meanings that were present in L1 were not. The study provided further evidence in support of the MCH.

In contrast, a few recent studies have provided evidence against the MCH. For example, Ionin et al. (2021) investigated whether intermediate to advanced L1 Chinese (*n* = 35) and L1 Korean learners (*n* = 31) of English would show varying sensitivity to plurality in definite and indefinite contexts. Although neither Chinese nor Korean have obligatory plural marking, Korean utilises plural marking more extensively compared to Chinese. Korean's plural marker -*tul* is obligatory with definite nouns but optional with indefinite nouns. In contrast, the Chinese plural marker -*men* does not indicate definiteness and is not allowed on [-human] nouns, which is considered completely incongruent with English. This study employed both online and offline tasks: SPR and AJT task, respectively. The results suggested that in definite context (e.g., these, those), only the Korean learners showed sensitivity to the missing plural -*s*, while the Chinese learners did not, aligning with the MCH predictions of clear L1 transfer. However, in indefinite contexts (e.g., many), no differences were found between the two groups in the GJT task, as both groups overly accepted ungrammatical sentences compared to the natives. Surprisingly, both L2 groups exhibited online sensitivity to errors in the SPR task, contradicting the MCH prediction since neither Chinese nor Korean requires obligatory marking in this context. Based on the evidence, the authors concluded that it is possible for L2 learners to overcome morphological incongruency and acquire new L2 morphemes. Similar evidence against the MCH has also been found in the acquisition of L2 indefinite articles (Ionin et al., 2021) and grammatical person and number (Yao & Chen, 2017).

In summary, the predictions made by the MCH align closely with those proposed by the RDH (Ionin et al., 2021). Both hypotheses suggest that L2 learners may be unable to fully acquire certain grammatical features that are not instantiated in their L1. One significant difference between these two hypotheses is that RDH predicts L2 learning difficulty within the UG framework, whereas MCH bases its predictions on levels of morphological congruency, defined by morphemes rather than languages (Ionin, 2023). In addition, the two hypotheses differ in their methodological approaches for empirical testing. At some levels, the RDH lacks detailed specificity, allowing for various types of tasks that involve different levels of implicit and explicit knowledge. This flexibility means that RDH can be tested using a broad range of methods, from grammaticality judgments to story retellings. In contrast, MCH is specifically suited to psycholinguistic experiments (Jiang et al., 2011, 2017), as this hypothesis focuses on the automatic activation of meaning encoded in grammatical morphemes. For instance, in saying "I like the books", the speaker must activate the concept of plurality to use the plural marker correctly; similarly, when hearing this sentence, a listener must automatically recognise the plural marker *-s* and connect it to the plural meaning in mental representation to understand the reference to multiple books (Jiang et al., 2017). Therefore, to effectively test the MCH, online processing methods are necessary to minimise the involvement of explicit knowledge and focus on automatic activation of grammatical meaning. This methodological specificity may explain why the MCH has not been extensively tested beyond plural marking.

4.3 L2 Processing Studies on Tense-Aspect

As discussed in Section 4.2, psycholinguistic and neurolinguistic methods (e.g., SPR, eye-tracking, ERP) have gained increasing popularity in the field of SLA over the past decade or so, marking a shift away from exclusive reliance on traditional, often offline, methods. This shift can be attributed to the limitations of traditional SLA approaches, which struggle to disentangle learners' mental representations from their cognitive processing (Rastelli, 2020). Also, L2 researchers have become more aware of the fact that L2 acquisition requires not only accumulating L2 knowledge but also effectively applying that knowledge in real-time language processing (White, 2003b, cited in Roberts, 2012). For example, L2 learners may possess clear representations of L2 aspectual categories but encounter challenges in using these representations in real-time processing; alternatively, they might simply mimic the most commonly encountered combinations of lexemes and morphemes in the input, relying on memory-based chunks for aspectual form-function pairings even before fully establishing the corresponding mental representations (Rastelli, 2020). Spoken production data is believed to fall under real-time processing measures, as it happens in real time by its nature. However, it has limitations in its inability to measure language use in real time without placing production pressure on learners (Grüter et al., 2012). From this perspective, psycholinguistic methods (or online methods) arguably provide better tools for L2 researchers to study the mental processes involved when learners read or listen to the sentences in real time for comprehension (Marinis, 2003). In addition, as suggested by Roberts (2012), online methods can also help L2 researchers further explore and address whether persistent morphosyntactic difficulties in L2 learners, even at advanced proficiency, result from a representational deficit (RDH) or a real-time processing problem, such as knowledge retrieval in real time (MSIH).

In recent years, there has been a growing number of L2 processing studies focusing on tense-aspect acquisition. However, the quantity of these studies remains relatively small when compared to processing research on other linguistic phenomena, such as grammatical gender (Rastelli, 2020). This section will review relevant studies in detail, which have presented mixed evidence regarding the influence of L1 on L2 tense-aspect acquisition.

For instance, Chan (2012) used self-paced reading (SPR) tasks to investigate whether L1 Korean (n = 15), Chinese (n = 21), and German (n = 25) learners of L2 English are able to process temporal anomalies online in a similar way to the native English control group (n = 24). This study focused on English past simple and progressive constructions, and the stimuli included both grammatical and ungrammatical sentences with grammar- and meaning-induced violations. Examples of stimuli were shown in (2) and (3).

(2) Past simple

a. *Yesterday several large snakes escaped from their cage at the zoo.*

b. **Yesterday several large snakes escape from their cage at the zoo.*

(grammar-induced violation)

c. **Tomorrow several large snakes escaped from their cage at the zoo.*

(Meaning-induced Violation)

(3) Progressive

a. *Currently the baby is laughing while the mother tickles him.*

b. **Currently the baby laughing while the mother tickles him.*

(grammar-induced violation)

c. **Lately the baby is laughing while the mother tickles him.*

(Meaning-induced Violation)

(Chan, 2012, p. 36)

The SPR results revealed that native English speakers showed full sensitivity to tense-aspect violations in online comprehension. Specifically, they exhibited significantly longer RT when reading anomalous sentences compared to grammatical ones. However, the processing patterns among L2 learners varied. The L1 Korean group exhibited clear sensitivity to violations for both past simple and progressive, showing RT patterns similar to those of the native English group. The German learners, however, were only sensitive to grammar-induced violations but not to meaning-induced violations for progressive sentences. The L1 Chinese group showed no sensitivity to grammar-induced violations in past simple sentences, irrespective of their English proficiency levels. Yet, their RT patterns in reading progressive sentences were similar to those of the native English group, indicating that they were sensitive to both grammar- and meaning-induced violations in this context. Based on the findings, Chan argued that there was a strong influence of L1 in the L2 learners' RT patterns. German does not grammaticalise progressive aspect, thus arguably leading to their insensitivity to violations in the progressive sentences. On the other hand, Chinese does not grammaticalise past tense but allows optional usage of progressive markers, and this could explain why the Chinese learners did not exhibit sensitivity to anomalies for the past tense but that they were sensitive to violations in progressive sentences. Despite the evidence, the author acknowledged that the German participants were recruited in Germany and the other two learner groups were recruited in the US. The potential L2 immersion effect on learners' performance was not considered in this study, which might have affected the results.

Similarly, Roberts and Liszka (2013) employed a SPR task to examine if L1 French (n = 20) and German (n = 20) learners of L2 English show online sensitivity to past simple and present perfect sentences containing temporal-aspectual mismatches, as illustrated in (4) and (5). The study included a native English control group (n = 20) and the L2 were classified as advanced based on their Oxford Placement Test scores. Unlike Chan's (2012) study, Roberts and Liszka also used an offline AJT task to measure the learners' explicit knowledge of tense-aspect, in addition to the SPR task that taps into (more) implicit knowledge.

(4) Past simple

a. *Last week, James went swimming every day.*

b. **Last week, James has gone swimming every day.*

(5) Present perfect

a. *Since the summer, James has gone swimming every day.*

b. **Since the summer, James went swimming every day.*

(Roberts & Liszka, 2013, p. 421)

The results from the AJT revealed that both L1 French and L1 German groups, like the native group, displayed correct explicit knowledge of the English past simple and present perfect. This was evidenced by their significantly lower ratings for the mismatch conditions compared to the match conditions. However, in the self-paced reading task, only the L1 French showed online sensitivity to mismatch conditions in both past simple and present perfect (i.e., longer reading times). The L1 German failed to show such a processing cost in both conditions. The authors proposed that this divergence in processing behaviour may be attributed to the influence of their respective L1s. Specifically, the French participants' sensitivity to tense-aspect violations may be due to the fact that their L1 encodes aspectual distinctions (albeit differently from English). On the other hand, German does not have a grammaticalized aspect system in its language, which might affect German learners' activation of implicit knowledge related to English aspectual distinctions. Roberts and Liszka, thus suggested that even though learners show explicit knowledge of L2 aspectual distinctions, they may still not able to utilise them automatically in real-time comprehension.

Instead of adopting a violation detection paradigm as in previous processing studies, Roberts and Liszka (2021) used a SPR task and an acceptability judgement task again to investigate if L1 differences in whether aspect is grammatically realised or not influenced their online and offline comprehension of garden-path sentences. Specifically, the study aimed to investigate whether advanced L2 English learners showed sensitivity to the distinction between the [+/-progressive] Aspect. The participants included advanced L2 English learners with L1 German (*n* = 32), L1 Dutch (*n* = 24), and L1 French (*n* = 24) backgrounds, as well as a control group of native English speakers (*n* = 20). The stimuli comprised early-closure sentences, as in (6a) and (7a), and late-closure sentences, as in (6b) and (7b).

(6) Past simple

a. *As John hunted the frightened rabbit escaped through the dark trees.*

b. *As John hunted the frightened rabbit it escaped through dark trees.*

(7) Past progressive

a. *As John was hunting the frightened rabbit escaped through the dark trees.*

b. *As John was hunting the frightened rabbit it escaped through the dark trees.*

(Roberts & Liszka, 2021, p. 630)

The results of the study indicated that all the L2 English groups, regardless of their L1s, performed similarly to the native speakers in their processing and interpretation of sentences in the past simple tense. They all experienced a delay in their reading pace (suggesting they were trying to re-evaluate the sentence) as soon as they read "rabbit" in the early-closure condition (37a). This pattern was also observed in the learners' offline judgments in the AJT task. However, all three learner groups exhibited different behaviours when dealing with progressive sentences. The German learners treated both simple past and progressive similarly, showing the same preference for the direct object ('rabbit') in both online and offline performance. In contrast, the L1 French learners performed most similarly to the English native group and considered the atelic interpretation of "John was hunting" (which does not necessarily require an object). This might because the French *imparfait* can convey both ongoing and habitual actions in the past. Lastly, the L1 Dutch group, similar to the French, showed an effect of aspect in the offline AJT task. However, in the online SPR task, this effect was not observed. The authors attributed the observed online and offline differences among L2 groups to the L1 influence, i.e., whether a learner's L1 encodes the progressive aspect through syntactic means (as in French) or solely through lexical means (as in German).

Another online processing study on L2 tense-aspect was conducted by Yao and Chen (2017). They employed both self-paced reading and eye-tracking tasks (i.e., reading paradigm) to examine how the learners' L1 (i.e., Chinese) affects their online processing of English past tense and progressive. They recruited a total of 60 participants from a Chinese university and categorised these participants into either higher-proficiency level group (*n* = 30) or lower-proficiency level group (*n* = 30). The stimuli, as in (8) and (9), included both grammatical and ungrammatical sentences, and the ungrammatical sentences contained the main verb in its uninflected form.

(8) Progressive

a. *Right now, the little girl <u>is playing</u> with her doll in the house.*

b. **Right now, the little girl <u>is play</u> with her doll in the house.*

(9) Past tense

a. *Last night, my friend <u>walked</u> home after the closure of the ceremony.*

b. **Last night, my friend <u>walk</u> home after the closure of the ceremony.*

(Yao & Chen, 2017, p. 275)

The results suggested that both high- and low-proficiency learners were sensitive to the violations of progressive. This finding was in line with Chan's (2012) study, suggesting that L1 Chinese can process English progressive online, as it is grammatically marked in both Chinese and English. Since the past tense is not marked in Chinese, learners were expected to show difficulty in their processing of the English past tense, as found in Chan's (2012) study. However, it was observed that the influence of the learners' L1 was only detected among those of low proficiency. The high-proficiency learners demonstrated a sensitivity to the violations of the past tense, although this sensitivity was only detected in the eye-tracking experiment (i.e., evidenced by longer gaze time). The authors argued that this might be due to different processing load on participants. Compared to the SPR task, the eye-tracking method is less cognitively demanding and more time-sensitive. In light of these findings, Yao and Chen concluded that as the learners' L2 proficiency increases, they may indeed eventually show native-like performance in real time comprehension even if the relevant grammatical structure is absent or encoded differently in their L1s. This study provides evidence against the RDH and the MCH.

Focusing on the perfective/imperfective contrasts, Vogel (2017) employed an online SPR task and two offline cloze tasks (i.e., a story-in-context cloze task and an isolated sentence cloze task) to examine how L1 English learners (*n* = 63) comprehend aspectual contrasts in L2 Spanish. A control group of native Spanish speakers (*n* = 35) was also included. In Spanish, perfective/imperfective contrasts are encoded overtly through [+/-perfective] morphology. However, English does not grammaticalise the straight perfective/imperfective distinction but encodes imperfectivity indirectly through progressive and non-progressive meanings. It was expected that L1 English learners would encounter difficulties in learning Spanish perfective-imperfective contrasts due to the influence of L1. In this study, the stimuli in the SPR task consisted of both logical and illogical sentences, as illustrated in (10) and (11).

(10) Logical

El equipo perdía contra los rivales pero ganó en los últimos minutos del partido.

'The team was losing against their rivals but won in the last minutes of the game.'

(11) Illogical

**El equipo perdía contra los rivales pero ganó en los últimos minutos del partido*.

'*The team lost against their rivals but won in the last minutes of the game.'

(Vogel, 2017, p. 57)

The results from the two offline cloze tasks showed that the native Spanish group performed better than the L2 groups, and the advanced learner group outperformed the intermediate group. In the SPR task, both accuracy (participants were also asked to determine whether the presented sentence was logical or not) and reading times were analysed. It was revealed that L2 learners had difficulty in interpreting the meaning of the sentences, as indicated by their low accuracy rates (below 60%). Also, L2 learners experienced a decrease in processing speed, particularly in the verb region, when dealing with both logical and illogical sentences. The author interpreted these findings as indicating processing difficulty. L2 learners were both slower and less accurate in processing both logical and illogical sentences compared to native Spanish speakers. This was in line with Roberts and Liszka's (2013) findings that even though L2 learners may have acquired explicit knowledge of the Spanish perfective-imperfective contrast, possibly through classroom instruction, they were unable to apply this knowledge effectively in real-time processing. Further, Vogel interpreted the results as sopporting evidence for RDH, as even the advanced learners in the study seemed to not yet acquire the [-perfective] feature in Spanish.

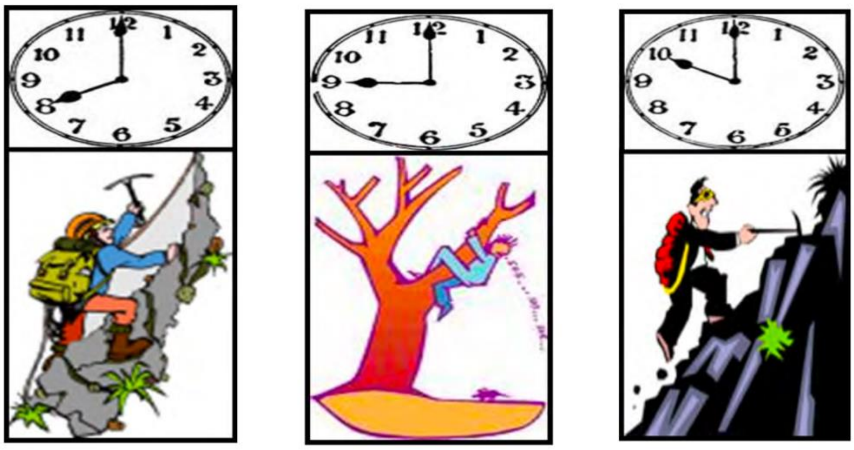
Similar to the design of the present study, Kahoul (2014) investigated the processing and production of English past tense among adult L2 learners whose L1 was either Chinese (n = 37) or Arabic (n = 34) using a picture-choice task (measuring offline responses, reaction times, and eye movements) and a sentence imitation task. The participants were matched in their proficiency at low, intermediate, and advanced levels, and a control group of English native speaker (n = 10) also participated in the study. An example of stimuli used in the picture-choice task is presented in Figure 4.3.1. In the distractor trial (a), time context (an hour before 10 o'clock) was introduced. Choosing the 9 o'clock picture would indicate that the participants had understood the tense context. In the experimental trial (b), the same pictures appeared with the critical sentence "*he climbed up the mountain.*" It was predicted that if learners were able to process past tense inflections, they would have chosen the 8 o'clock picture. On the other hand, failure to process the past tense would cause confusion between the 10 and 8 o'clock pictures. The author analysed reaction times (the speed at which a participant chose the correct picture), length of eye gaze to the target picture, and the speed of first look to the target picture as indicators of processing.

Figure 4.3.1

An Example of Stimuli Used in Kahoul's (2014) Study

a. Distractor Trial
It’s ten o’clock right now. Listen and choose:
“He was climbing something one hour ago.”

b. Experimental Trial
“He climbed up the mountain.”



Note. This is adapted from Kahoul (2014, p. 165).

The results revealed that the L1 Chinese and the L1 Arabic groups did not differ in their RTs, and both groups were significantly slower than the native speaker group. However, in terms of the length of looks to the target, only the intermediate- and advanced-level Arabic learners showed reliably more looks to the target than to the competitor pictures, whereas the Chinese learners did not show any preference for either picture, irrespective of their proficiency levels. Also, the advanced Arabic learners were as fast as the native speakers in displaying their first looks to the target picture, and the advanced Chinese group was significantly slower than their Arabic counterparts. The group differences were mirrored in the sentence imitation data, where the accuracy rates in past tense production did not differ between the Chinese and Arabic groups across the low- and intermediate-level proficiency levels. Interestingly, in the advanced proficiency groups, the Arabic learners performed significantly better than the Chinese.

The author argued that a lack of grammatical tense in Chinese has led L1 Chinese learners to be unable to process and produce English past tense in a target-like manner, compared to the Arabic learners, whose L1, similarly to English, encodes past tense grammatically. He concluded that L1 influence may not emerge until a specific developmental stage (i.e., advanced level in this case) is reached, as the L1 effect was not apparent in the performance of learners in the low and intermediate proficiency groups. This might be because the L1 effect only becomes evident when functional categories are established in a learner’s mental grammar, and at the early stage of development, learners are still in the process of acquiring the category and lack relevant syntactic representations for the L1 effect to take place (Hawkins, 2001). Nevertheless, this study suffers from its methodological issues in the analysis of the processing data. Although eye-movement data were extracted, analysing the total length of time looking at the target picture was not particularly informative and failed to capture the moment-by-moment processing information during the audio stimuli. For example, it would be more persuasive to examine a specific time period (e.g., 200 ms) after the critical word *climbed* in the spoken sentence. This approach can help determine whether knowledge of past tense marking can be immediately activated and utilised for sentence comprehension by the participants. Also, analysing the time of the first look at the target picture was inconclusive, as participants might have been in the process of deciding between pictures. Such an analysis could potentially compromise the validity of the study's results.

In Farina's (2017) study, the acquisition of the English present perfect was investigated within the same population: L1 Chinese (*n* = 44) and L1 Arabic (*n* = 11) adult learners with varying English proficiency levels, in comparison to the native English speakers (*n* = 72) and L2 English learners with other L1 backgrounds (*n* = 23) (i.e., the other L1 group). The study employed both online (a SPR task) and offline (a rating task) measures. The author focused on two relevant features associated with English present perfect: boundedness and current relevance [+CR]. It was expected that the L1 Chinese learners would experience positive transfer, as their L1 has the post-verbal perfective marker *-le* with bounded meanings and the sentence-final particle *le* with [+CR] feature, which are functionally similar to English present perfect. On the other hand, the L1 Arabic learners were predicted to experience no such transfer due to a lack of similar structure in their L1. For the stimuli, the author manipulated boundness based on whether there is a clear restriction on the direct object of the verb, as in (12). The marked current relevance was altered by introducing adverbial modifiers [+/-CR] at the beginning of the sentence, as illustrated in (13).

(12) Boundedness

a. *Deliberately, the researcher has tested <u>her theory</u> on the circus monkeys who had to identify colors.* [+BND]

b. *Deliberately, the researcher has tested <u>theories</u> on the circus monkeys who had to identify colors.* [-BND]

(13) Current relevance

a. *<u>For two minutes</u>, the dog has chased the foxes to their nearby burrow while barking to alert the hunters.* [+CR]

b. *<u>In two minutes,</u> the dog has chased the foxes to their nearby burrow while barking to alert the hunters.* [-CR]

(Farina, 2017, pp. 88, 90)

Interestingly, the predictions were not wholly realised. The results revealed that in the boundedness condition, the L1 Arabic group showed target-like performance, with shorter reading times in nonbounded contexts. In contrast, the reading times of the L1 Chinese group were not affected by the manipulations in boundedness. This difference was also reflected in the two groups' rating task results. In accounting for the results, the author suggests that although the perfective marker *-le* in Chinese indicates a bounded event, the absence of formal or structural similarity to English might inhibit the transfer of the L1. However, it was unclear why the Arabic group performed in a native-like way, possibly due to Arabic perfect and past continuous tense (the perfective and imperfective distinction) being mapped onto the English bounded and nonbounded present perfect. In the current relevance condition, the reading times of the L1 Chinese and the L1 Arabic groups did not differ significantly, and neither group showed sensitivity to the manipulation of the current relevance feature. However, in the rating task, the Arabic group showed an effect, giving higher ratings to adverbial modifiers indicating current relevance and lower ratings to those indicating a lack of current relevance. Again, unexpectedly, the Chinese group did not show such an effect. Since the L1 Arabic group and the other L1 group performed similarly, it was assumed that the Chinese group's results were likely caused by negative L1 transfer. As the author argued, one possible explanation for the negative transfer might be that English has a wider range of adverbial modifiers that signal current relevance. In other words, the absence of a single, specific form or a set of forms that clearly represents the function of sentence-final *le* (the current relevance marker in Chinese) might hinder the acquisition of adverbial markers for current relevance in English. Despite the interesting findings, the number of participants in the L2 groups of this study was highly imbalanced. The Chinese group (n=44) included more participants compared to the Arabic group (n = 11), and this might introduce potential confounds in the analysis and present challenges in separating the impact of L2 proficiency from the influence of L1 transfer.

More recently, Zeng et al. (2021) employed a SPR method to investigate the online processing of English past tense and progressive by L1 Chinese learners at lower (*n* = 30) and higher (*n* = 30) proficiency levels, compared with a control group of English native speakers (*n* = 30). This study mainly aimed to examine how lexical aspect (i.e., state, activity, and achievement verbs) interacts with grammatical aspect (i.e., past and progressive), and its influence on the processing of tense-aspect for Chinese learners whose L1 is grammatically marked differently from English. Examples of stimuli used in this study are shown in (14) and (15).

(14) Past simple

Bill loved/helped/killed the innocent child in the playground.

(state)/(activity)/(achievement)

(46) Progressive

Tom is hoping/training/beginning to win the game on Saturday.

(state)/(activity)/(achievement)

(Zeng, et al., 2021, p. 5)

The results showed that there was no significant difference between the native speakers and the L1 Chinese learners in their processing of prototypical sentences (achievement verbs with past tense and activity verbs with the progressive). However, significant differences were found between the two groups in their processing of non-prototypical items. Specifically, lower-proficiency Chinese learners were significantly slower than both the higher-proficiency learners and the native speakers when processing state verbs with past tense and progressive. This suggested that lexical aspect indeed affects the processing of L2 tense-aspect, with prototypical sentences being processed quicker than non-prototypical ones. Moreover, the processing of non-prototypical sentences appeared to be more influenced by the learner’s L1. The authors explained that compared to prototypical verbal combinations, non-prototypical ones are less firmly established or fixed, leading learners to rely more on their L1 grammar. Since the differences were found between lower-proficient learners and higher-proficient learners, the authors still argued that nativelike performance in processing is still achievable, although more studies are needed in the future to directly compare two groups with L1s that have distinct grammatical tense-aspect structures (such as Chinese and Spanish).

To summarise, L2 processing studies on tense-aspect acquisition have predominantly utilised the self-paced reading task to investigate learners’ online processing performance. Some studies have also adopted traditional offline methods to complement the online data. However, these investigations have yielded mixed evidence regarding the impact of L1 on the acquisition process. For example, Roberts and Liszka (2013, 2021) found evidence of L1 influence in advanced L2 learners’ online processing, even though all the learners demonstrated nativelike explicit knowledge of L2 tense-aspect. They argued that the L1 may influence different aspects of linguistic knowledge, with metalinguistic/explicit knowledge being less affected by the learners’ L1. In other words, the influence of L1 is more likely to be observed in implicit/online processing tasks. On the other hand, Yao and Chen's (2017) study suggested that advanced proficiency learners can eventually overcome L1 effects and show nativelike performance in online processing tasks.

One possible explanation for these varying results could be the different online methods employed. In Yao and Chen's (2017) research, learners’ sensitivity to grammatical violations was not detected in the self-paced reading task but was evident in the subsequent eye-tracking task. This difference might arise because eye-tracking, compared to the SPR task, places fewer processing demands on learners and provides a more natural reading environment (rather than a word-by-word reading approach in the SPR task) (Frenck-Mestre, 2005). Therefore, learners may exhibit their online sensitivity to the target structure more clearly in eye-tracking studies. Despite its widespread use, the SPR task requires further empirical validation, particularly with respect to its ability to truly measure participants' implicit knowledge (Marsden et al., 2018). The SPR task used in the aforementioned studies typically operated in a written modality and was not constrained by time limits imposed by the researcher (participants read the sentence by their own pace). These test designs may facilitate access to awareness and, possibly, explicit knowledge (Marsden et al., 2018). This underscores the need for further validation of the SPR task and the establishment of standardised practices for study design (Marsden et al., 2018).

4.4 The Present Study

4.4.1 Research Rationale

A substantial body of previous L2 research on tense and aspect has primarily concentrated on testing the universality of the most influential hypothesis in the field: the Aspect Hypothesis (AH) (Bardovi-Harlig & Comajoan, 2020). As reviewed earlier, the AH assumes that the inherent semantics of verb predicates play a determining role in learners' acquisition of morphological marking, and this hypothesis has received significant empirical support across various learner contexts. However, studies within this line of research have relied on elicitation tasks, typically assessing how often learners employ specific verb forms (count the numbers). In this case, even when a particular telic verb is observed to be used more frequently in the perfective aspect than in the imperfective aspect by the learner, questions about whether the learner possesses the corresponding mental representation remain unanswered (Rastelli, 2020). Shifting the focus away from this method is a fundamental step for researchers to investigate interlanguage grammar (Roberts & Liszka, 2013).

In addition, the AH tends to overlook the potential influence of grammatical aspect. While categories of verb predicates exist in many languages, grammatical aspect is less prevalent. In languages that feature grammatical aspect, native speakers naturally possess the relevant linguistic features, whereas L2 learners may lack access to these features. From this perspective, the AH might primarily account for L2 development of tense-aspect in the early stages, as learners in more advanced stages tend to move away from marking verbs solely based on inherent aspect and begin to incorporate grammatical aspect, mirroring the usage by native speakers (Roberts & Liszka, 2021). More importantly, if the AH is considered universally applicable, it struggles to explain the consistent challenges faced by learners from specific L1 backgrounds; for example, L1 Chinese L2 English learners have been found showing persistent difficulty in using past tense marking (Goad et al., 2003; Hawkins & Liszka, 2003; Lardiere, 1998a, 1998b).

Due to this limitation, the present study decides to focus on particular rather than the universal. Many prior studies in the field have also looked at the particular issues, such as the role of L1, proficiency level, task effect, etc. (Bardovi-Harlig & Comajoan, 2020). Specifically, researchers working within the generative framework have paid a great deal of attention to the role of L1 effect on the acquisition of L2 tense and aspect. However, evidence on the extent and permanence of L1 influence is mixed. Some studies (e.g., Al-Thubaiti, 2015; Gabriele & Maekawa, 2008; Gabriele & McClure, 2011) have found limited L1 influence, suggesting that L2 learners, at least those at advanced level, are able to show target-like performance even if the relevant features do not exist or are encoded differently in their L1. Other studies (e.g., Alruwaili, 2014; Gabriele et al., 2005; Hawkins & Liszka, 2003; Liszka, 2004; McManus, 2015) have found strong L1 influence among advanced L2 learners, arguing that such influence can be permanent and may prevent L2 learners from achieving nativelike performance if their L1 lacks the relevant features..

These mixed findings have been interpreted within different theoretical accounts. Specifically, the full access approaches assume that it is theoretically possible for L2 learners to attain native-like syntactic representations, and any challenges encountered by learners are typically the result of various other factors coming into play, such as mapping problems between mental representations and surface morphological production (MSIH) (Prévost & White, 2000b). On the other hand, the RDH argues for a deterministic role of the L1 influence, with the permanent deficiency in learners' L2 mental representations if relevant features are absent from their L1 (Hawkins & Liszka, 2003).

What needs to be stressed here is that no matter where these prior studies stand theoretically, the majority of them have relied on offline methods and oral production data. To some extent, the conflicting findings could be explained by the different research methods used. For example, it is commonly interpreted that when learners perform better in the comprehension than the oral production, it aligns with the Full Access Accounts. However, it is important to note that the comprehension tasks (e.g., sentence or story interpretation task) used in these studies are offline and not time constrained, tapping into a different level of knowledge compared to the oral production, which is likely to be influenced by real-time processing pressure (Grüter et al., 2012). Alternatively, when non-target-like performance is found among learners in oral production, it is often interpreted as supporting evidence for the L1 influence. Yet, as mentioned, spoken production is always affected by real-time processing constraints.

Traditional SLA methods are less able to separate learners' mental representations from real-time processing (Rastelli, 2020), which in some way restricts researchers' ability to examine the different theoretical accounts. Also, offline measures are more likely to tap into learners' explicit/metalinguistic knowledge; thus, it is not surprising that learners can exhibit target-like performance in these tasks, given the fact that many of them are instructed learners and they can utilise their learned explicit rules when performing such tasks. However, as argued by Roberts (2012), L2 acquisition requires not only using explicit L2 knowledge but also applying that knowledge in real-time comprehension. For this reason, psycholinguistic methods (online measures) have gained popularity in the field, offering tools to measure learners' implicit knowledge/online processing ability.

While there has been an increase in the number of online processing studies on L2 tense-aspect, it is still very rare compared to processing studies focusing on other linguistic phenomena (Rastelli, 2020). Among the existing processing studies, the self-paced reading task with an agreement violation detection paradigm has been predominantly employed. These studies have often relied on the use of ungrammatical sentences to test if learners showed sensitivity to the errors like native speakers. Few studies have looked at the online processing of fully grammatical sentences. More importantly, when comparing the findings of the existing processing studies, mixed evidence of L1 influence has been suggested once again.

Furthermore, it seems very common in L2 generative research to study learners' comprehension and production knowledge separately, as these two modalities are believed to develop at different paces (Marinis & Cunnings, 2018; Tasseva-Kurktchieva, 2015). Yet, measuring comprehension and production separately can make it difficult to study the asymmetric relationship between these two aspects. For L2 tense-aspect acquisition, few researchers have combined comprehension and production tasks in a single single study with the same participants. Even where both modalities are investigated, different levels of knowledge may be incorrectly compared. For example, offline comprehension results (explicit knowledge) may be directly compared to online oral production results (implicit knowledge).

To summarise the research gaps, mixed evidence for the L1 influence on the acquisition of L2 tense-aspect still exists. The conflicting results can be attributed to the limitations of the traditional SLA research methods. However, even among the small number of processing studies, contradictory findings have been observed. Therefore, to better compare learners' comprehension and production at different levels of linguistic knowledge, the present study combines both online and offline comprehension/production tasks to examine the acquisition of L2 tense-aspect by the same L1 Chinese and L1 Arabic adult learners of L2 English across all tasks The tasks are summarised in Table 4.4.1.

Table 4.4.1

Summary of Tasks Employed in the Present Study

	Online	**Offline**
Comprehension	*Task 1:* visual-world eye-tracking *Task 2:* Sentence-matching	*Task 4:* AJT
Production	*Task 3:* Elicited imitation	*Task 5:* Gap-filling

Specifically, the learners' real-time comprehension is measured using two different processing tasks: the visual-world eye-tracking examining the processing of fully grammatical sentences, and the sentence-matching tasks investigating sensitivity to tense-aspect violations. These two different paradigms are employed, as there is a possibility that processing ungrammaticalities may involve different processing procedures compared to fully grammatical items. For example, in some case (e.g., acquisition of grammatical gender), learners are able to detect errors in grammatically incorrect sentences but are unable to use grammatical information in processing grammatical sentences facilitatively (Foote, 2011; Grüter et al., 2012). It would be interesting to further explore how different experimental paradigms might lead to different processing findings. In addition, an acceptability judgement task is employed to measure learners' offline comprehension. In terms of production, an elicited imitation task is used to examine learners' real-time production accuracy of tense-aspect, and a gap-filling task is designed to test learners' offline written production knowledge. It is argued that the three online tasks are tapping into learners' implicit knowledge, whereas the two offline tasks are accessing explicit knowledge.

The L1 Chinese and L1 Arabic adult learners were chosen, as it is a common practice to include participants with different L1s. If only one L2 learner group is included, the learning problems encountered by this group may only reflect the general difficulty of acquiring the L2 structure, irrespective of learners' L1s. Also, as discussed in the cross-linguistic differences section, Chinese and Arabic manifest clear contrasts with respect to their grammaticalization of tense-aspect. This makes it possible to examine the potential influence of their respective L1s.

4.4.2 Research Questions and Predictions

Based on the research aims and design, the present study attempts to address the following four main research questions:

RQ1: How do Chinese and Arabic learners of English process English tense-aspect markings in comparison to native English speakers in the visual-world eye-tracking task?

Predictions:

a) *The L1 English group will be able to process morphological markings in past simple (-ed), present perfect (have+V-ed), and present progressive (be+V-ing) and use them to encode event information immediately.*
b) *The L1 Chinese group is expected to show target-like processing of be+V-ing in present progressive sentences but will display slower processing of past simple (-ed) and present perfect (have+V-ed) constructions.*

c) *The L1 Arabic group will demonstrate target-like processing of past simple (-ed) and present progressive (be+V-ing) sentences but will exhibit slower processing of present perfect (have+V-ed) sentences.*

There is ample evidence that native speakers are able to rapidly use morphosyntactic information during online processing (Clahsen & Felser, 2006). Particularly for tense-aspect processing, Zhou et al. (2014) found that both children and adult native speakers of Mandarin could use the perfective and durative morphemes (*-le* and *-zhe*) encoded in the verb immediately (before the verb's arguments were introduced) to compute the event information during real-time comprehension. Based on their findings, it is thus predicted that native English speakers will also process morphological markings (*-ed*, *have+V-ed*, and *be+V-ing*) incrementally and use them to encode temporal/aspectual information and facilitate event recognition.

Although it has been argued that inflectional morphology tends to pose a persistent challenge for L2 learners in both comprehension and production (DeKeyser, 2005), we predict that both Chinese and Arabic groups will show target-like processing of present progressive sentences (*be+V-ing)*, in light of some previous findings suggesting that English present progressive does not pose challenges for either Chinese (H. L. Chan, 2012; Yao & Chen, 2017) or Arabic speakers (Alruwaili, 2014; Al-Thubaiti, 2015). Based on the MCH, it is predicted that the Arabic group will demonstrate an L1-based advantage over the Chinese group in processing past tense morpheme *-ed*, as past tense is morphologically marked in Arabic but not in Chinese. However, both L2 groups are predicted to exhibit non-target-like patterns (slower than the native group) in processing *have+V-ed* in present perfect sentences, as present perfect is not grammatically marked in either Chinese or Arabic.

RQ2: Do Chinese and Arabic learners of English exhibit grammaticality effects for tense-aspect structures similar to native English speakers in the sentence-matching task?

Predictions:

a) *The L1 English group is anticipated to demonstrate online sensitivity to violations in present perfect and present progressive sentences, but not in past simple sentences.*
b) *The L1 Chinese group is expected to exhibit online sensitivity only to violations in present progressive sentences, with no sensitivity to violations in past simple and present perfect sentences.*
c) *The L1 Arabic group will also demonstrate online sensitivity to violations in present progressive and past simple sentences, without exhibiting sensitivity to violations in present perfect sentences.*

The reason why the L1 English group is predicted not to be sensitive to violations in past simple sentences is based on the findings of Roberts and Liszka (2013). Their study suggested that the changes in how the present perfect is used in American English are currently occurring in British English, making it gradually acceptable to use sentences such as "*I already did that*" instead of "*I have already done that.*" Given that Chinese lacks a tense feature, it is predicted that Chinese speakers will not be sensitive to violations in both past simple and present perfect sentences. However, because Chinese grammaticalises progressive meaning through its aspectual marker, Chinese learners are expected to show sensitivity to errors in present progressive sentences.

Furthermore, since Arabic encodes tense grammatically, it is predicted that Arabic speakers will exhibit sensitivity to violations in past simple sentences. These participants were instructed learners, and it was assumed that they had not been exposed to the English environment sufficiently to fully adopt a native-like processing pattern for mismatched past simple sentences. Although Arabic does not grammatically realise progressiveness, we still predict that they will be sensitive to violations in present progressive sentences based on the findings of previous studies (Alruwaili, 2014; Al-Thubaiti, 2015) and the fact that the use of *gaa'ad* in Arabic dialects is becoming grammaticalised to convey progressive meaning. Lastly, as Arabic does not grammatically encode the present perfect, the Arabic learners are expected to show no processing cost for violations in present perfect sentences.

RQ3: Do Chinese and Arabic learners of English perform differently from the native English speakers in the sentence imitation task?

<u>Predictions:</u>

a) *The L1 English group will consistently display target-like patterns, correcting more ungrammatical sentences rather than introducing errors into grammatical sentences across all three tense contexts.*
b) *The L1 Chinese group will demonstrate target-like patterns in the present progressive context, but not in the past simple and present perfect contexts.*
c) *The L1 Arabic group is expected to perform in a target-like manner in both past simple and present progressive contexts, but not in the present perfect context.*

It is assumed that participants' automatic correction of ungrammatical sentences during repetition can serve as evidence of their internalised knowledge (Erlam, 2006). The native speakers are expected to perform at a high level in correcting ungrammatical items, as demonstrated in previous studies using an imitation task (Cristante, 2020; Erlam, 2006; Spada et al., 2015). For the L2 groups, predictions are based on cross-linguistic differences discussed in RQ1 and RQ2.

RQ4: Do Chinese and Arabic learners of English exhibit explicit knowledge of English tense-aspect compared to native English speakers, as measured by the offline acceptability judgement task (comprehension) and gap-filling task (production)?

Predictions:

a) *All three groups are expected to demonstrate correct explicit knowledge of English past simple, present prefect, and present progressive.*

b) *All three groups will rate the ungrammatical sentences significantly lower than the grammatical sentences in the AJT task.*

c) *All three group will display a high level of accuracy in supplying English tense-aspect forms in the gap-filling Task.*

These predictions are grounded in the fact that these learners had received formal instruction in English and had spent varying number of years in a classroom setting before pursuing their studies at a UK university. Given their level of English proficiency, it is anticipated that they will exhibit target-like performance in offline tasks designed to tap into their explicit knowledge.

Chapter 5 Methodology

The present study aims to investigate the acquisition of tense-aspect by adult L1 Chinese and L1 Arabic learners of L2 English. The study examines their varying levels of knowledge in English tense-aspect through a combination of online and offline comprehension and production tasks. This chapter presents information about the methodology of the research, starting with a discussion of ethical considerations, followed by the details of both L1 English and L2 (L1 Chinese and L1 Arabic) participants.

The three comprehension tasks, comprising the visual-world eye-tracking task, sentence-matching task, and acceptability judgment task, are methodically explained, including stimuli design, task administration, data cleaning, and analysis. Following this, the two production tasks, elicited imitation and gap-filling, are introduced in a similar manner with a focus on methodological aspects and procedures for data cleaning and analysis. The chapter also outlines the data collection procedures and concludes with the pilot study section.

5.1 Ethical Considerations

Since human participants were involved in this study, ethical issues were considered and addressed throughout the entire research process. The research project obtained prior approval from the ethical committee at the University of York before data collection. The recruited participants were adults (18 years or older) and did not belong to any vulnerable groups. They were invited to participate in the study via email invitations from their departmental administrators, recruitment posters displayed on campus noticeboards, or through social media platforms.

Participants who voluntarily signed up for the study were invited to the eye-tracking lab at the university. Before starting the experiment, each participant was presented with a consent form and an accompanying information sheet. Participants were required to read all the information and were encouraged to ask any additional questions they might have had. The information sheet provided details about the purpose of the study, the tasks involved in the experiment, procedures about data anonymisation, the usage of data in public, and contact information for both the researcher and the Ethics Committee. After signing the consent form, each participant was assigned a numerical code to ensure anonymity. The code was the only form of identification in the database. The code linking their name with the data was stored in a separate file protected with a strong password, and only the researcher has access to it. Participants were given the right to withdraw at any point during the experiment and up to two weeks afterward, after which their data would be fully anonymized. In addition, participants were provided with a copy of their signed consent form in case they wish to contact the researcher for data withdrawal or further inquiries.

All experimental data were securely stored and treated with confidentiality. Only the researcher has access to the raw data, and the researcher's supervisor had access to the anonymised data. Any utilization of the data in public, such as during conferences, talks, or poster presentations, was conducted in anonymised formats.

5.2 Participants

5.2.1 Native English Participants

A total of 24 native English participants were recruited from the university of York, consisting of 19 females and 5 males. Their age ranged from 19 to 37 years old, with a mean age of 23.7 years (SD = 5.39). To ensure that variations in English dialects did not influence participants' responses to specific sentences, only native British speakers were included in the study. None of the participants were early bilinguals or spoke another language other than

English at an advanced level. They all had normal or corrected-to-normal vision, and no hearing disorders were reported.

5.2.2 L2 Participants

Another 48 L2 participants were also recruited from the university of York. They were from L1 Chinese (n = 24) and L1 Arabic (n = 24) backgrounds. The L1 Chinese group consisted of 19 females and 5 males, with their age ranging from 19 to 36 years old (M = 25.9). The L1 Arabic group had 16 females and 8 males, and their age ranged from 19 to 38 years old (M = 27.5).

As noted by Hulstijn (2011), measuring language proficiency has proven challenging, and there is currently a lack of consensus on a standardised measure of proficiency in SLA research. For this reason, three different yet commonly utilised tools for assessing overall English proficiency were adopted in this study. Proficiency measure 1 was the Oxford Quick Placement Test (QPT) (2001), which consists of 60 items. Initially, the L1 Chinese group achieved a mean score of 43 (SD = 6), while the L1 Arabic group scored a mean of 49 (SD = 7.2). These scores categorised the L2 participants as upper-intermediate/near-advanced learners. However, to ensure comparable proficiency levels between the two groups, three participants' data from each group were excluded. In the Chinese group, three participants scoring below 35 were removed, and in the Arabic group, three participants scoring above 57 were excluded. This adjustment resulted in a final sample of 21 participants in each group, with a mean QPT score of 44.8 for the Chinese group and 48.5 for the Arabic group. To statistically compare the QPT scores between the groups, a Shapiro-Wilk test was first performed to assess the normality of the data distribution within each group. The results indicated that the QPT scores were normally distributed (L1 Chinese: W = 0.97, p = .82; L1 Arabic: W = 0.92, p = .10). Then, a parametric independent samples t test was conducted and revealed no significant differences between the L2 groups' QPT scores: t (40) = 1.20, p = .053.

Table 5.2.1

Summary of L2 Groups' English Proficiency Levels and Learning Backgrounds

	L1 Chinese group (*n* = 21)		L1 Arabic group (*n* = 21)	
	***M* (*SD*)**	**Range**	***M* (*SD*)**	**Range**
QPT scores	44.8 (4.9)	37-56	48.5 (7)	32-57
IELTS	7.0 (0.5)	5.5-7.5	7.0 (0.6)	5.5-8.0
Self_rating				
Listening	4.8 (0.7)	3-6	4.3 (0.7)	3-6
Reading	4.9 (0.7)	3-6	4.6 (0.9)	3-6
Speaking	3.4 (0.9)	1-5	4.4 (0.8)	3-6
Writing	3.3 (0.6)	2-4	4.1 (0.8)	3-6
Average	4.1 (0.6)	2.5-5	4.3 (0.7)	3-5.8
Age of learning	8 (2.4)	3-13	9 (4.4)	4-20
Years of formal English instruction	13 (2.7)	6-16	10 (4.0)	4-20
Number of months residing in the UK	30.5 (27)	6-98	38.8(30.5)	2-132

Proficiency measure 2 involved standardised English proficiency test scores (e.g., IELTS, TOEFL) reported by the L2 participants in their language background questionnaires. These tests were undertaken as part of the entry requirements for their university programmes. Both groups had a mean score of 7 in IELTS or equivalent tests.[1] The Shapiro-Wilk test suggested that the Arabic group's scores were normally distributed ($W = 0.92$, $p = .09$), while the Chinese group's scores were not ($W = 0.84$, $p = .003$). Due to this violation of normality, a non-parametric Mann-Whitney U test was performed, which indicated no significant difference in IELTS scores between the two groups: $U = 237$, $p = .67$.

Proficiency measure 3 involved the L2 participants' self-reported English proficiency levels in listening, reading, speaking, and writing, rated on a scale from 1 (very low) to 6 (native). Previous research has shown that self-report language measures have demonstrated high reliability and validity across various contexts (Marian et al., 2007). The two groups had similar mean self-rating scores (L1 Chinese: $M = 4.1$; L1 Arabic: $M = 4.3$). The Shapiro-Wilk test suggested that the scores for the Chinese group were not normally distributed ($W = 0.90$, $p = .03$), although the scores for the Arabic group were ($W = 0.90$, $p = .03$). Therefore, a non-parametric Mann-Whitney U test was conducted and found no significant difference in the self-reported scores between the two groups: $U = 260.5$, $p = .31$.

A composite proficiency score was also calculated based on the three measures. Once again, there was no significant differences in the composite proficiency score between the Chinese and Arabic groups:[2] $t(40) = 1.96$, $p = .06$.

In addition to proficiency measures, the L2 participants were asked to report their English learning experiences in the language background questionnaire. Both L2 groups started learning English at a similar mean age: 8 years old for the Chinese group and 9 years old for the Arabic group. No significant difference was found in the onset age of learning: t

[1] IELTS global scales were used as the majority of the participants could not remember their scores for the four sub-skills and did not report them in the language background questionnaire.

[2] Normality of composite proficiency scores: the assumption of normality was met in both groups (Chinese group: $W = 0.98$, $p = .91$; Arabic group: $W = 0.93$, $p = .13$).

(40) = 1.39, $p = .17$.[3] The Chinese group received slightly longer years of formal English instruction ($M = 13$) compared to the Arabic group ($M = 10$), but this difference was not significant: $U = 145.5$, $p = .06$.[4] Although the Arabic group had longer immersion experiences in the UK at the time of testing compared to the Chinese group, no significant difference was found: $U = 264$, $p = .28$.[5] Based on these results, the L2 groups were well-matched in terms of their English learning experiences.

The next section will discuss the different tasks used in the present study, and the three comprehension tasks will be introduced before the two production tasks.

5.3 Visual-World Eye-tracking Task

5.3.1 Methodological Considerations

Eye-tracking is an online method that involves using an eye-tracking device to track and record the precise movements and positions of people's eyes as they observe content displayed on a computer screen (Godfroid, 2019). There are two common paradigms: text-based and visual-world. The former is often used to investigate participants' reading behaviour and comprehension of text-based materials such as words, sentences, and texts, and the latter is commonly used to study participants' processing of spoken language while viewing a visual scene at the same time (Roberts & Siyanova-Chanturia, 2013). Compared to traditional SLA methods, eye-tracking provides a direct measure of cognitive processing during a language task (Rayner, 2009). It does not rely on participants making explicit

[3] Normality of onset age of learning: the assumption of normality was met in both groups (Chinese group: $W = 0.96$, $p = .62$; Arabic group: $W = 0.91$, $p = .06$).

[4] Normality of formal instruction time: the assumption of normality was met for the Chinese group ($W = 0.93$, $p = .14$), but it was violated for the Arabic group ($W = 0.91$, $p = .048$).

[5] Normality of immersion time: the assumption of normality was violated in both groups (Chinese group: $W = 0.83$, $p = .002$; Arabic group: $W = 0.87$, $p = .01$).

judgements, allowing researchers to directly assess how participants comprehend and process information in real time; in addition, it collects participants' eye movements data with precise temporal resolution, measured in milliseconds, allowing for accurate identification of subtle changes in participants' comprehension (Conklin et al., 2018). In particular, the visual-world paradigm (VMP) can be utilised with participants of various ages, including preschool children, as it does not require reading skills or completing complex tasks (Berends et al., 2016). The present study adopted a visual-world design; therefore, the following discussion will focus on this paradigm.

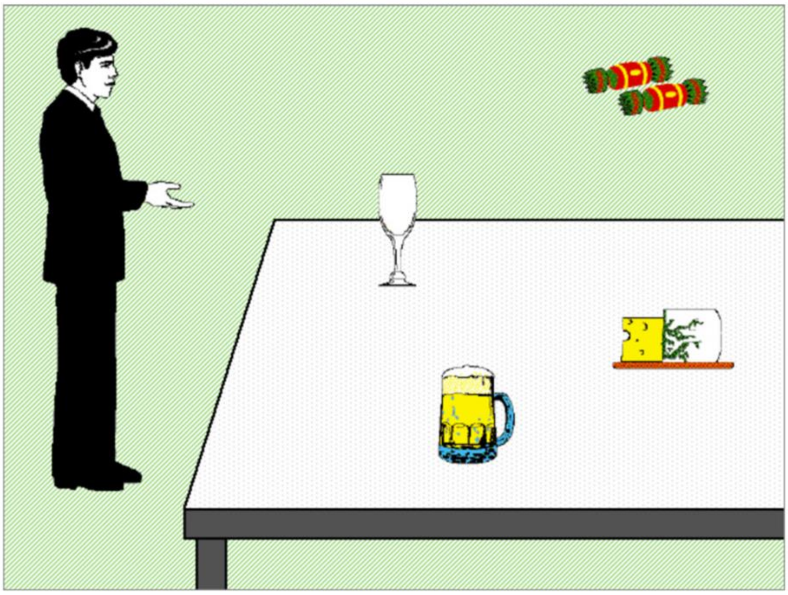
The visual-world paradigm was developed based on an observation made by Cooper (1974). He found that his participants spontaneously directed their gaze to the relevant visual objects described in the spoken narratives. For example, their eyes were drawn to the picture of a lion when they heard a mention of the word 'lion' in the narratives. These eye movements were closely synchronised to the incremental processing of the linguistic input, and eye fixations were sometimes initiated before the spoken word was completed. His study had a pioneering impact and provided evidence in support of the use of VWP task for investigating real-time language comprehension. However, this method had failed to get noticed until Tanenhaus et al. (1995) employed it 20 years later. In their study, participants were instructed to touch one of four blocks that varied in marking, colour, or shape (e.g., "Touch the starred yellow square"). They found that individuals typically made eye movements to the target block approximately 250 ms after the word that uniquely specified the target in terms of visual characteristics (e.g., "starred" or "square"). This study further validated visual-world paradigm as a method investigating the cognitive processes involved in language comprehension.

Visual-world research is grounded in the linking hypothesis (i.e., eye-mind link), which proposes a connection between auditory-linguistic processing and visual processing

(Godfroid, 2019). This hypothesis suggests that individuals tend to look at objects mentioned in the auditory stimuli they hear, indicating that eye movements can provide insights into the linguistic representations formed during auditory processing (Godfroid, 2019). Essentially, visual-world researchers use eye movements as a way to measure representations (Tanenhaus & Trueswell, 2006). This means that by tracking where people look in a visual scene, researchers can infer which linguistic representations are being activated in the listener's mind at specific moments. For example, the pattern of eye fixations can indicate the timing and manner in which the phonology and meaning of a word are accessed from the mental lexicon during auditory comprehension (e.g., Allopenna et al., 1998; Mercier et al., 2016). In other words, it can uncover what is represented and activated in the mind, as processing and representations are inherently connected (Godfroid, 2019).

Figure 5.3.1

Example of a Visual Display used in Altmann and Kamide's (2007, p.505) Study



The past research has revealed two main types of eye movements in the visual-world paradigm: referential eye movements, which happen in align with named objects in the auditory input, and anticipatory eye movements, which occur before the objects are named (Godfroid, 2019). A third type is described in Altmann and Kamide's study (2007), where the objects were neither named nor going to be named in the auditory input. In this study, the researchers asked a group of English native speakers to listen to sentences such as *The man will drink all of the beer* and *The man has drunk all of the wine,* while presenting them with a visual display containing a full glass of beer, an empty glass of wine, and two other distractors (as illustrated in Figure 5.3.1).

Their results suggest that native speakers were able to use the morphosyntactic cues encoded in the verbs, as they directed more looks to the full glass of beer in the future tense condition ("*will drink")* and to the empty glass of wine in the past tense condition ("*has drunk")*. In this study, the eye movements were observed towards target visual objects that were not mentioned or intended to be mentioned in the auditory input. When listening to sentence *The man has drunk all of the wine*, the participants looked more to the empty glass which was not named in the auditory sentence. The authors suggested that language comprehension occurs with consideration of object affordances, which refer to "the actions which an object permits its viewer" (Altmann & Kamide, 2007, p. 510). For instance, an empty glass here does not afford drinking, and participants were able to use their real-world knowledge to infer how the glass may have become empty, thus suggesting that an empty glass represents the outcome of a drinking event (Godfroid, 2019). The study has contributed to the revised linking hypothesis, which emphasises the significance of object affordances, event representations, and shared featural representations (encompassing visual, phonological, semantic, and morphosyntactic aspects) as crucial elements influencing eye movements during language processing (Altmann & Kamide, 2007).

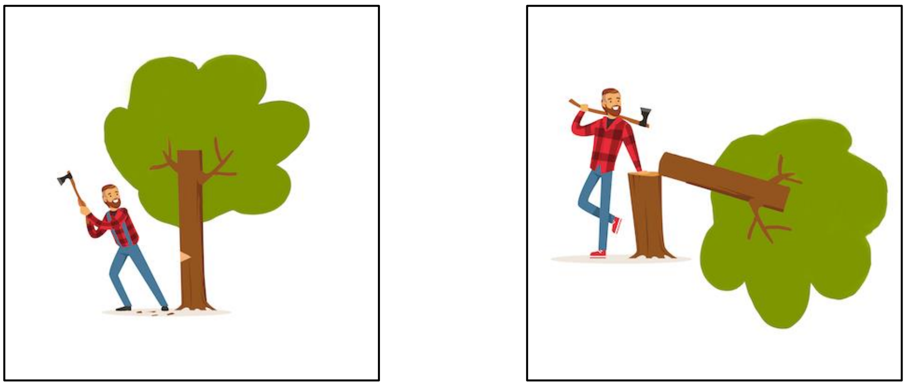
Since then, researchers have begun using this paradigm to investigate sentence processing among both L1 speakers and L2 learners (e.g., Ahn & Song, 2023; Chan et al., 2018; Cunnings et al., 2017; Ellert, 2013; Grüter et al., 2012; E. Kim et al., 2015; H. Kim & Grüter, 2021; Koch et al., 2021; Lago et al., 2022; Minor et al., 2022; Mornati et al., 2022; Perdomo & Kaan, 2021; Trenkic et al., 2014; Zhou et al., 2014). In the current study, the visual-world paradigm was employed to look at participants' use of morphological cues to encode event information in online sentence processing. It focuses on English past simple, present perfect, and present progressive sentences, and attempts to explore whether participants are able to make use of the temporal and aspectual information (imperfective vs. perfective) conveyed by the morphology of the verbs (i.e., *V-ing, V-ed and have+V-ed*) during real-time comprehension. It is expected that if the L2 participants have truly acquired the grammatical structures, they will be able to do so.

5.3.2 Stimuli Design

A total number of 54 critical items were constructed for the eye-tracking experiment. Each item consisted of three auditory sentences and one visual image comprising two pictures, both with dimensions of 400 × 400 pixels. One of the pictures depicted an ongoing event (e.g., *The young man is chopping the tree down*), while the other picture depicted a completed event (e.g., *The young man chopped the tree down* or *The young man has chopped the tree down*). An example of a critical item is presented in Figure 5.3.2. The visual representation does not capture the distinction between past simple and present perfect tenses, as it is challenging to portray these differences pictorially. Both tenses can be associated with the result of the action shown in the picture. However, this does not compromise the quality of the data. The rationale behind this design is that participants' eye movements and gaze patterns reflect their interpretation of tense as guided by the auditory cues, rather than the visual stimuli alone, as described by the linking hypothesis (see Section 5.3.1). Participants were presented with only one sentence per image item, and the specific tense was conveyed through the accompanying auditory sentence, which allows us to examine whether participants can use the verb information to identify the target picture. One limitation is that this design only examines resultative perfect, among the four perfect aspect types. This limitation will be discussed in the study limitations section.

Figure 5.3.2

An Example of a Critical Item Used in the Visual-World Eye-Tracking Task



Auditory sentences:
a. The young man is chopping the tree down. (Present progressive)
b. The young man chopped the tree down. (Past simple)
c. The young man has chopped the tree down. (Present perfect)

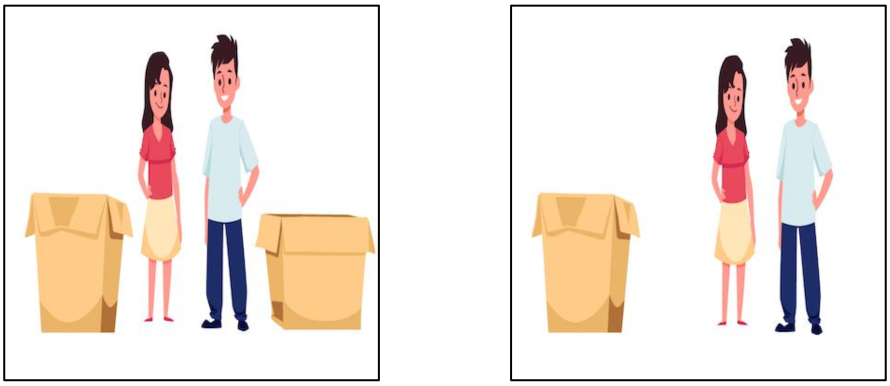
To prevent any potential bias towards specific displayed objects, the positioning of the completed event and the ongoing event picture was counterbalanced throughout the trials. In half of the trials, the completed event appeared on the left side of the screen while the ongoing event was shown on the right side. In the remaining half of the trials, this positioning was reversed.

The three sentences of each item were put into three different experimental lists, and each participant was randomly assigned to one of the lists and listened to only one version of each item. This helps to avoid repetition effects where participants respond to the stimulus in an abnormal manner (e.g., skimming through it) due to their prior exposure to it (Keating &

Jegerski, 2015). Each experimental list also included 54 filler items to conceal the critical items from the participants. Similar to the critical item, every filler item consisted of one auditory sentence and one visual image comprising two different pictures. In half of the filler trials, the auditory sentences corresponded to the pictures displayed on the left side, while in the remaining trials, the auditory sentences described the pictures shown on the right side. In each experimental list, the order of the critical and filler items was randomised for each participant.

Figure 5.3.3

An Example of a Filler Item Used in the Visual-World Eye-Tracking Task



Auditory sentences:
The man and women stand between the boxes.

5.3.3 Administration of the Task

This task was conducted using an Eyelink 1000 plus eye-tracker developed by SR Research company. Participants were seated approximately 80cm away from the desktop computer, which had a resolution of 1024×768 pixels. An adjustable padded chinrest and forehead rest was used to securely fix participants' head position and maintain their eye positions within the eye tracker's range throughout the experiment. Eye movements were recorded at a sampling rate of 500 Hz (i.e., 500 data points or frames per second). The recording was conducted monocularly, which is a very common practice in visual-world eye-tracking studies.

Before the experiment, each participant underwent a calibration and validation process using a 9-point calibration grid to ensure the accuracy of the eye-tracking data. Throughout the entire experiment, participants were instructed to maintain a consistent head position to minimise any potential disruptions to the recording. After the calibration, participants were shown the task instruction page on the monitor to ensure a clear understanding of the task requirements. Following this, they completed 5 practice trials to familiarise themselves with the task. Each trial began with a fixation dot at the centre of the screen, which served the purpose of drift correction. To proceed to the next trial, participants were required to maintain their gaze fixed on the dot. The visual images appeared on the screen 500ms before the corresponding spoken sentence was played through headphones. Participants then had to select the picture that best matched the spoken sentence by pressing either the Left arrow key for “the left picture” or the Right arrow key for “the right picture”, after which the next trial began. The entire experiment lasted 15 to 20 minutes.

5.3.4 Data Cleaning

After the data collection, the eye-tracking data was cleaned and prepared for subsequent analyses. Following Godfroid's (2019) guidelines, trials with abnormalities should be removed first. Therefore, each individual trial recording was inspected by the researcher using the software DataViewer developed by the SR Research company. A total of 17 trials with abnormalities were deleted at this stage. These abnormalities could have resulted from participant body movement or distractions from their mobile phones, as indicated by the recordings or noted down by the researcher during the experiment.

After the initial screening, the data were exported from DataViewer. The exported data were then pre-processed and cleaned using the *VWPre* package (Porretta et al., 2020) in *RStudio* (version 3.6.2, R Core Team, 2019). The offline responses for both critical and filler items were used to examine the engagement of the participants with the task. If the participants selected the correct pictures, they were given a score of 1, otherwise 0. All participants scored at least 60% across all the trials, and no participant data were excluded as they appropriately performed the task and paid attention to the auditory sentences. Following this, all the filler trials were removed from the remaining data pool. For the critical trials, time series were created from 200ms before the onset of the verb and extended over a period of 3000 ms. Trials with over 50% track loss (i.e., blink and offscreen data) during this period were excluded from further analysis, and this resulted in the exclusion of 6 events (0.16% of the remaining data points). Finally, the remaining data were computed in each 50 ms time window. Each time window consisted of 25 observations or samples, and this calculation was performed for the entire 3000 ms period.

5.3.5 Data Analysis

Analysing time-course data from the visual-world task can be challenging, since they often violate the assumptions of many parametric statistical tests (Baayen et al., 2018; Porretta et al., 2017). For example, in this case, participants cannot simultaneously look at both pictures. If they fixate on one image area, they cannot look at the other area at the same time. This violates the assumption of independent observations for parametric tests when comparing looks to these two interest areas (Ito & Knoeferle, 2022). Also, it is important to note that consecutive eye movements are often correlated with each other (Barr et al., 2011; Ito & Knoeferle, 2022). Although the eye-tracker can record eye fixations at the millisecond level, participants do not constantly move their eyes at such high frequencies (Stone et al., 2021). It has been found that an adult typically takes around 200 to 250 milliseconds to plan and execute an eye movement (Saslow, 1967; Q. Yang et al., 2002). Consequently, neighbouring eye fixations often reflect the same stage of processing and exhibit a robust correlation (Stone et al., 2021). In this case, conducting parametric tests such as ANOVA at multiple time points can lead to an overestimation of data variance and an inflated Type I error rate (Huang & Snedeker, 2020). Grouping data points into larger time bins (e.g., 150ms bins) helps decrease autocorrelation, but it cannot entirely eliminate the issue (Stone et al., 2021). Also, when the analysis is carried out in numerous time bins, it can elevate the Type I error rate because of the multiple analyses performed, known as family-wise error rate. Although the multiple comparisons problem can be addressed (e.g., through Bonferroni correction), this correction can be overly cautious and might consequently raise the Type II error rate (Ito & Knoeferle, 2022).

Alternatively, some non-parametric tests provide an effective way to solve the multiple comparisons problem and control for autocorrelation of eye movement data (Maris & Oostenveld, 2007). For example, the non-parametric cluster-based permutation analysis (CPA) has been increasingly used to analyse visual-world data in recent years (e.g., Amos et al., 2022; Andringa, 2020; Baker & Love, 2022; Barr et al., 2014; Chan et al., 2018; Huang & Snedeker, 2020; Koch et al., 2021; Marimon et al., 2022; Minor et al., 2022, 2023; Mornati et al., 2022; W. Yang et al., 2020). In this test, condition labels (e.g., target/competitor) are randomly reassigned multiple times to disrupt the experimental manipulation, creating a null distribution, and the significance of the original dataset's test statistic is determined by where it stands in this distribution (Stone et al., 2021). This analytical method controls for autocorrelation since the temporal structure of the data remains intact during permutation, and it can also identify temporal clusters or time windows in which participants’ looks to the target and competitor differed significantly (Barr et al., 2014; Maris & Oostenveld, 2007).

However, one limitation of CPA is that the identified significant time windows cannot indicate when the effect emerged and its duration (Sassenhagen & Draschkow, 2019). This is because the *p* value is calculated based on cluster-level statistics, which means there is not a statistical confidence regarding a particular time bin within the cluster (Ito & Knoeferle, 2022). For example, if an identified time cluster spans an entire time window of 500-1000ms, which only indicates that there was a significant effect within the time window. It cannot be interpreted that the significant effect started at 500ms and ended at 1000ms (i.e., lasted for a duration of 500ms).

If the research question also concerns comparing the onset of the effect among groups, the bootstrapping divergence point analysis (DPA) can be used as a complement to CPA. Similarly to CPA, DPA controls for the multiple comparisons problem and autocorrelation of visual-world data (Ito & Knoeferle, 2022). In this analysis, a non-parametric bootstrap procedure is used to iteratively resample the original dataset multiple times to generate new datasets, and after each resampling, a statistical test is performed (Stone et al., 2021). The process aims to derive an estimation of the distribution of test results. Some recent studies (e.g., Corps et al., 2022; Ito et al., 2023; Lago et al., 2022; Minor et al., 2022) have employed this analysis for visual-world data, since it can also predict when an effect begins (i.e., the time when the looks to the target start to diverge from the looks to the competitor) and provide a confidence interval, enabling researchers to statistically compare the onset of effect between various groups (Ito et al., 2023; Stone et al., 2021). However, this analysis does not estimate the duration of an effect and can only identify the first divergence point (Ito & Knoeferle, 2022; Stone et al., 2021). More importantly, it is not appropriate to test whether an effect is present, since it assumes the effect exists already and only detects its onset (Seedorff et al., 2018). Therefore, the DPA is normally used in combination with the CPA that tests whether the effect is significant in the first place (Minor et al., 2022).

In the current study, the cluster-based permutation analyses were performed first in each group using *permutes* package (Voeten, 2022) in *RStudio* to assess whether there was a significant effect (i.e., the probability of participants' looks to the target picture was significantly more than the non-target picture). The test was conducted in the pre-selected interest period of 3000ms (starting 200ms before the onset of the verb), and the analysis procedures followed Ito and Knoeferle's (2022) steps. Since fixation proportions are inherently bound between 0 and 1, they might not suitable to be the dependent variable for many statistical analyses (Godfroid, 2019). Thus, an empirical logit transformation was conducted base on Barr (2008) using the following formula:

$$"elog = ln\ (\frac{\text{looks} + 0.5}{\text{non} - \text{looks} + 0.5})"$$

(Godfroid, 2019, p. 298)

Following this, the linear mixed-effects model was applied to each consecutive 50ms time bin (i.e., a total of 60 bins). The model predicted the logit transformed fixations by the interest area (i.e., target vs. non-target), and included by-participant and by-item random intercepts. When consecutive time bins exhibited a significant effect ($t > 2$; $p < .05$), they were grouped into a cluster. The size of each cluster was determined by summing the Likelihood Ratio Tests statistics (i.e., cluster mass statistics). To ensure that these clusters were not the result of random chance, 1000 permutations were performed. In each permutation, the picture labels (target and non-target) were randomly reassigned, and the same mixed-effects model analysis was repeated where the cluster results were compared with the ones identified in the first step. Clusters in the original data were considered significant if the probability of observing a cluster of equal or larger size in the permuted data was less than 5% ($p < .05$) (Maris & Oostenveld, 2007).

As discussed earlier, although the CBP analyses helped identify if each group showed significantly more looks to the target area in the interested time period, they could not compare the onset of the effect among the three groups statistically. Therefore, the bootstrapping divergence point analyses were performed complementarily. The analysis was conducted using the *boot* package (Canty & Ripley, 2022), and the procedures were identical to that which is described in Stone et al.'s (2021) study. Firstly, a one sample t-test testing the effect of interest area on logit transformed fixations was conducted for each time bin. The t-test was chosen in this case due to its efficiency in terms of convergence and computational time (Stone et al., 2021). To establish a divergence point where the target attracted significantly more fixations than the non-target area, the first of four consecutive bins exhibiting a significant effect in the same direction was selected. With a bin size of 50ms in the present study, the analysis presumed that a fixation should last for a minimum of 200ms (i.e., equivalent to 4 bins). Then, a bootstrap approach was employed to create "new" datasets through resampling the original dataset with replacement. This resampling was stratified, taking participant, time bin, and interest area (target/non-target) into account, ensuring that data were resampled within these specific categories. This step was iterated 1000 times with new divergence points being obtained for each resampled dataset. The means of these generated timepoints were taken as the overall divergence point for the group. To estimate the variability, the percentile method was used to calculate a 95% confidence interval (CI).

When presenting the results of CBP analyses, the report included the start and end times of the identified temporal cluster, the cluster mass statistic (summed LRT statistics), the summed t-values associated with the fixed effect (indicating the direction of the effect), and the p-values corresponding to the cluster-mass test. For DPA results, the estimated onset of divergence point in each group and the mean differences among their onset times were reported. The 95% CI of these mean differences were used to assess the significance of group differences. Specifically, if 95% CI includes zero, it suggests that the true difference between the groups could be zero, indicating a lack of statistical significance; on the other hand, if the CI does not contain zero, it implies a statistically significant difference between the groups (see Cumming, 2012; Larson-Hall & Plonsky, 2015).

5.4 Sentence-Matching Task

5.4.1 Methodological Considerations

The sentence-matching (SM) experiment has a long history in psycholinguistics research. The procedure of the task is rather simple: participants are sequentially presented with sentence pairs on a computer screen, and they are asked to decide whether the sentences in each pair are identical in form or not. Previous research indicates that response times (RTs) are significantly faster for matched grammatical sentence pairs (1a) compared to matched ungrammatical sentence pairs (1b) among native speakers, suggesting that response latencies can reveal implicit knowledge of grammaticality (Bley-Vroman & Masterson, 1989; Freedman & Forster, 1985).

(1) *a. The living cost is much higher in London.*

The living cost is much higher in London.

b. **The living cost is much high in London.*

**The living cost is much high in London.*

Two possible explanations have been put forward to account for this phenomenon. One theory, advocated by Freedman and Forster (1985), suggests that participants can rapidly create a coherent, high-level mental representation when processing grammatical sentences; however, they struggle to do the same for ungrammatical sentences, leading to a less efficient word-by-word comparison strategy, resulting in longer response times. Another explanation, proposed by Crain and Fodor (1987), is that participants tend to correct the ungrammatical sentences in their minds, and it is the mental correction, rather than the ungrammaticality itself, that delays their response times. For our purposes, the specific debate between these two theories is not crucial. What matters is that the SM task does not ask learners for explicit judgements of grammaticality; instead, it relies on participants' implicit linguistic knowledge, as evidenced by their reaction times (Gass, 2001).

Bley-Vroman and Masterson (1989) examined the applicability of the SM experiment in SLA research. They conducted a study with Korean learners of English and native English speakers as controls. The results showed that both groups took significantly longer to match ungrammatical English sentences (with incorrect determiner placement) than grammatical ones. They argued that the SM task could supplement overt judgment data, which often fails to assess learners' developing interlanguage grammar accurately. Following this study, the SM task has been widely used in L2 research and has proven to be sensitive to various L2 structures. For example, Clahsen et al. (1995) found that some Korean learners of German exhibited longer RTs in the SM task when matching sentence pairs that violated subject-verb agreement, similar to native German speakers. Similarly, Duffield et al. (2002) discovered that both English and Spanish learners of French clearly distinguished between grammatical and ungrammatical stimuli, taking more time to respond to ungrammatical pairs involving illicit clitic placement than to grammatical pairs. As a reaction time task, SM provides insights into whether learners have integrated grammatical structures into their interlanguage representation; in contrast, offline judgment data are more influenced by learners' explicit knowledge (Jiang, 2012). In this aspect, the SM task appears to be superior to the traditional grammatical judgement task in assessing learners' implicit knowledge of grammar.

On the other hand, some researchers, such as Gass (2001), have raised caution about the application of SM methodology and its validity. A few studies have reported that the SM task failed to produce grammaticality effects in native speakers for certain grammatical structures, such as subjacency violations in wh-questions (Crain & Fodor, 1987) and sentences violating the verb-second constraint of German (Eubank, 1993). However, other researchers have argued that this could be because these seemingly ungrammatical structures are, in fact, underlyingly grammatical, supported by relevant syntactic theory. Regardless of the explanation, the crucial point is that before using the SM task to study L2 learners, it is essential to establish that the structure under investigation indeed elicits grammaticality effects in native speakers (Clahsen et al., 1995). If this is confirmed, the SM procedure can be utilised to L2 learners.

In addition, Gass (2001) conducted an SM experiment with beginning and intermediate L2 learners of French and surprisingly found no significant difference in their RTs for grammatical and ungrammatical pairs across three structures (i.e., adverb placement, subject-verb agreement, and clitic-pronoun placement). These results did not align with the same participants' offline acceptability judgment task outcomes. Consequently, she concluded that the SM task is only valid among L2 learners with a certain level of proficiency. Below that level, learners might not be able to construct high-level representations and may resort to word-by-word matching instead of actual reading and processing. In contrast, Duffield et al. (2007) argued that the failure of the SM experiment in Gass's article does not necessarily invalidate this methodology. In fact, using the SM task becomes redundant if one assumes that results from traditional judgment tasks are always valid (Duffield et al., 2007). Instead, what Gass's study (2001) may indicate is that the suitability of the SM task may depend on the learners' proficiency level and the specific grammatical structures being investigated.

Setting aside the mixed results, previous studies have emphasised the need for increased application of the SM methodology, particularly in SLA research. By doing so, we can gather more empirical evidence to gain a better understanding of this design, given its advantages over offline judgment tasks in certain aspects. In this study, an SM task was employed to investigate participants' underlying grammatical knowledge of English past simple, present perfect, and present progressive structures. Specifically, if participants are sensitive to ungrammaticalities, they will exhibit longer RTs when judging ungrammatical sentence pairs compared to grammatical ones.

5.4.2 Stimuli Design

For the SM task, a total of 48 critical items were constructed. Each item contains the three structures in two versions: the grammatical version and the ungrammatical version. The ungrammatical version involved a tense-aspect violation and was designed following Roberts and Liszka's (2013) study, wherein a mismatch between the temporal adverbial in the topic position and the verb was introduced. Note that all critical sentences were presented in pairs, which means each sentence was presented twice to participants.

Six experimental lists were created in total, each containing a single version of the critical item. Alongside the critical item, each list comprised 48 filler item pairs. An example of created experimental sentences is presented in Table 5.4.1. Since all the experimental items were presented in matching pairs, all the filler items appeared in non-matching pairs. The sentence in the filler pair (2) differed lexically from each other by only one word. All the filler items were grammatical. Participants were randomly assigned to one of the six lists, and the presentation order of both experimental and filler items was randomized for each participant.

Table 5.4.1

Examples of Critical Stimuli in the Sentence-Matching Experiment

Structure	Condition	Example sentence
Simple Past	Grammatical	a. *Last year, my grandmother taught me things about growing plants.*
	Ungrammatical	b. **Since last year, my grandmother taught me things about growing plants.*
Present Perfect	Grammatical	c. *Since last year, my grandmother has taught me things about growing plants.*
	Ungrammatical	d. **Last year, my grandmother has taught me things about growing plants*
Present Progressive	Grammatical	e. *Right now, my grandmother is teaching me things about growing plants*
	Ungrammatical	f. **Lately, my grandmother is teaching me things about growing plants*

(2) a. *I am used to drinking tea every morning before going to work.*

b. *I am used to drinking coffee every morning before going to work.*

5.4.3 Administration of the Task

The SM task was built using Experiment Builder SR Research (version 2.3.1) and administered using the desktop computer in the eye-tracking lab. The experiment started with an instruction page and five practise trials to familiarise the participants with the task. Participants were also given opportunities to ask questions if they had any at the end of the practise sessions. Each trial began with a fixation cross displayed at the centre of the screen, lasting for 500 ms. Following this, the first sentence of the pair appeared at the top left of the screen and remained visible for 3500 ms before being replaced by the second sentence of the pair, which appeared at the bottom right of the screen for 3500 ms as well. Participants were given clear instructions to promptly indicate whether the second sentence was identical to the first by pressing either the Left arrow key for "same" or the Right arrow key for "different". The response time was measured starting from the presentation of the second sentence and stopped when the participants pressed either of the keys. If a participant failed to respond within 3500 ms after the disappearance of the second sentence, the next trial was initiated automatically after an interval of 750 ms. The entire task took approximately 15 minutes for each participant to complete.

5.4.4 Data Cleaning

The data cleaning of the response times involved the following steps. Firstly, the percentages of correctly matching the sentences were calculated for each participant in each group. All the participants in the three groups achieved at least a 60% accuracy rate; thus no participant’s data was excluded at this stage. Following this, all the filler items were removed from the subsequent analysis, and trials with no responses provided by the participants within the given time limit were removed, which affected 1.14% of the data. After this, the means and standard deviations of the response latencies were calculated for each participant in each condition, and response latencies falling outside a cut-off of 2 standard deviations of each participant’s individual mean per condition were excluded from the analysis, resulting in the removal of 4.82% of the remaining data. It is important to mention that this criterion, where data points were excluded if they fell outside of mean ± 2 standard deviations, was chosen based on the practices of previous researchers (Duffield et al., 2002, 2007; Duffield & White, 1999; Gass, 2001) who used the same task.

5.4.5 Data Analysis

In this study, linear mixed-effects models were employed to analyse the sentence-matching data. Over the last 10 years, this statistical method has been widely used in L2 research, psycholinguistic studies, and language testing research (Cunnings, 2012). It is believed that it may gradually replace ANOVA in the next ten years, following the similar trend in the field of psychology (Godfroid, 2019).

When analysing reaction time data, it is important to consider that there is often a correlation in reaction times both within a single participant and across different items. This correlation arises from the fact that some participants generally respond more quickly, and certain items also tend to elicit faster reactions (V. A. Brown, 2021). Therefore, when working with data where observations are not dependent due to same participants responding to multiple trials, employing a statistical test that considers these interdependencies becomes crucial for analysis. For this reason, repeated ANOVAs have been used to model participant-level (F_1 analysis) and item-level variance (F_2 analysis) (H. H. Clark, 1973). However, repeated ANOVAs do not take both sources of variability into account simultaneously, and F_1 and F_2 analyses are conducted separately (Linck & Cunnings, 2015). On the contrary, mixed-effects model provides an alternative solution to the problem, as the model can consider the variability within and across participants and items at the same time and includes crossed random effects for both participants and items in a single model (Cunnings, 2012).

Another limitation with traditional ANOVA is that it requires a balanced dataset (Linck & Cunnings, 2015). If one observation is missing from the individual or item, the entire case is deleted from the analysis. This could potentially reduce the sample size, reducing the statistical power (V. Brown, 2021). While averaging over participants and items helps avoid the problem related to missing data, it could introduce the possibility that the averages for each participant or item may not be derived from the same number of observations (Linck & Cunnings, 2015). On the other hand, a mixed-effects model does not require a balanced design and handles missing data well, and analyses can be conducted on raw data without prior averaging (Cunnings & Finlayson, 2015). While a missing value leads to the removal of an observation, each observation in the mixed-modelling framework represents just one of many responses within an individual. Thus, the impact of single observation removal is minor compared to the ANOVA framework where all participant responses are treated as one observation (V. Brown, 2021).

ANOVAs have stricter assumptions, assuming a continuous dependent variable and categorical independent variables. Within ANOVA, continuous independent variables require categorisation, unlike in mixed-effects models (Baayen, 2010). Also, generalised linear mixed-effects models can be applied to analyse binary and categorical dependent variables (Jaeger, 2008; Linck & Cunnings, 2015). Furthermore, ANOVA results provide information about the significance of an effect, but they do not offer insights into the magnitude or direction of the effect, which can be indicated in mixed-effects models using coefficient estimates (V. Brown, 2021). Therefore, compared to ANOVAs, mixed effects models offer greater flexibility, which makes them more suitable to analyse experimental data that contain missing values in most cases.

In mixed-effects models, both fixed and random effects are included in a single analysis (Baayen, 2007). Fixed effects refer to predictors or independent variables that are held constant across the experiment (e.g., gender differences), and they are repeatable and can be tested with new samples (Winter, 2019). Random effects represent clusters of related data points within higher-level groups (e.g., a participant or an item) to account for individual variations within these groups (V. Brown, 2021). As random effects are sampling discrete units from a group, they are categorical in nature (Winter, 2019). Random effects can be modelled with only random intercepts or with random intercepts plus random slopes. Specifically, random intercepts account for variations in mean values among participants and items, while varying slopes account for variations in sensitivity to the manipulation, for example, participants may differ in how they respond to the manipulation – some showing strong effects, others weaker, and some even no difference (Cunnings, 2012). In L2 research, many scholars have chosen the default random intercepts only models and have failed to include random slope components when required (e.g., repeated measures variables), which can lead to an inflated Type 1 error rate (Cunnings & Finlayson, 2015; Schielzeth & Forstmeier, 2009). In fact, the problem lies in the fact that there is not a universally established guideline for selecting random effects in mixed-effects models. For example, Barr et al. (2013) suggested that random effects structures should be kept in a 'maximal' way, which has led linguistic researchers to fit models with overly complex random effects structures by adding varying slope terms for all critical variables in the study (Winter, 2019). As Matuschek et al. (2017) argued, these maximal models often result in estimation problems or fail to converge. The current study followed Winter's (2019) suggestion, that there is no single recipe for all circumstances, and all components of the model, including random effect structures, should be justified by the experimental design and theory.

For the analyses, three overall models were constructed to analyse participants' response time for past simple, present perfect, and present progressive structures respectively. The overall models aimed to compare group differences. Therefore, the fixed effects in the overall model included the two independent variables: *group* (L1 English vs. L1 Chinese/L1 Arabic) and *condition* (grammatical vs. ungrammatical), and their interactions. The random effects included both *participant* and *item*. The *group* variable was manipulated between participants, and each participant belonged to only one group, so it remained constant across participants. In contrast, the variable *group* was manipulated within items because the same items were presented to all groups; as such, introducing a random slope was necessary. As for the variable *condition*, it was manipulated within participants as all participants encountered both grammatical and ungrammatical conditions. It was also manipulated within items since items appeared in both conditions. Based on the design, the maximal random effect structures should include by-participant and by-item random intercepts, and by-participant random slopes for condition and by-item random slopes for group and condition.

Also, for each structure, three within-group analyses were conducted to examine whether each group showed sensitivity towards the ungrammatical condition. In within-group models, the fixed effect included *condition*, and the maximal random effect structures included by-participant and by-item random intercepts, and by-participant random slopes for *condition* and by-item random slopes for *condition*. However, in reality, the maximal models often fail to converge, which usually happens when the model is overly complex and includes multiple random slope parameters (Linck & Cunnings, 2015). In the current study, when the maximal models failed to converge, the optimising function (different algorithms for estimation) was used as the first step to find the model that converges (Winter, 2019). If this did not address the problem, then the model was simplified by gradually removing the random effect that was accounting for the least amount of variance until convergence was attained (Linck & Cunnings, 2015).

The mixed-effects model analyses were all conducted using the *lme4* package (Bates et al., 2015) in *RStudio*. Also, the *lmerTest* package (Kuznetsova et al., 2017) was utilised to compute p values, and the *MuMIn* package (Bartoń, 2023) was employed to compute values of variance (R^2). In cases where a significant interaction between the fixed effects was identified, the *emmeans* package (Lenth et al., 2022) was employed to execute Tukey-adjusted pairwise comparisons. When performing the analyses, a dummy coding scheme was adopted. For the *group* variable, the L1 English group was set as the reference level, and for the *condition* variable, grammatical condition was assigned as the reference level.

5.4.6 Data Reporting

In the present study, when reporting the results, descriptive statistics, such as means, standard deviations and 95% confidence intervals (CI), and some visual representations were presented first, followed by the mixed-effects model results. Following Cunnings' (2012) guidelines, the structure of converged mixed-effects models and the *R* formula employed will be reported clearly. Regarding the model results, the reported information included coefficient estimates (β) with 95% CI, *t* values (or *z* values), standard errors (*SE*), and the associated *p*-values (V. Brown, 2021). The significance level was set at 0.05 ($p < .05$). This cut-off point for *p* value, according to Plonsky (2015), is in fact somewhat arbitrary. Also, *p* values can be influenced by both sample size and the degree of variability in the data pool. A small *p* value might not necessarily reflect a significant result, while a larger *p* value, indicating non-significance, could potentially conceal an important result (Larson-Hall & Plonsky, 2015). To address the flaws of *p* values and move away from the dichotomous thinking that categorises results as either significant or non-significant (Cohen, 1994), effect sizes were reported as well when necessary, in this study. Unlike *p* values, effect sizes offer a tool to estimate the actual intensity of the relationship or the size of the effect being examined (Cumming, 2012; Plonsky, 2015). In other words, it tells us how important the result is in the relevant context (Larson-Hall, 2016). More importantly, it is a standardised measure and is comparable, which allows future researchers to compare studies across different contexts and to aggregate them through meta-analysis (Plonsky, 2012, 2015).

In the SLA field, common types of effect sizes include mean differences (e.g., Cohen’s d), correlations (coefficient r), and metrics related to variance accounted for (r^2, R^2, and η^2) (Plonsky, 2015). Any of these indices can be used, as they can be easily converted into one another; however, certain effect sizes are often more closely associated with specific statistical tests (Larson-Hall, 2016). For instance, relationship indices are typically utilised in correlation or regression analyses (Larson-Hall, 2016). In the present study, Cohen's d was calculated as the focus of the analyses was more on between-group or within-group differences. In terms of interpretations, the study did not use Cohen's benchmarks. As suggested by Plonsky and Oswald (2014), these benchmarks tend to underestimate the effects commonly observed in L2 research. Instead, they proposed a new scale specific to L2 research, as presented in Table 5.4.2. These reporting practices were consistently applied in other tasks using mixed-effects models.

Table 5.4.2

Effect Size Cohen's d Value Interpretations in the L2 Research Field

	Interpretations		
Effect size	**Small**	**Medium**	**Large**
Between-groups	0.4	0.7	1.0
Within-groups	0.6	1.0	1.4

Note. The table is adapted from Plonsky (2015, p. 38).

5.5 Acceptability Judgement Task

5.5.1 Methodological Considerations

Acceptability judgement task (AJT), sometimes termed as grammaticality judgement task (GJT), has been extensively used from the early days of L2 research since it can provide information that cannot be easily obtained from other types of data sources (Schütze & Sprouse, 2013). Specifically, it helps identify which utterances are possible and which are impossible, even if those utterances have never been naturally produced (Schütze & Sprouse, 2013). Unlike brain measure, which cannot reliably show sensitivity to all ungrammatical sentences, or large corpora, which cannot confirm grammaticality based solely on sentence occurrence or absence, AJT data in this sense provide a clearer picture (Schütze & Sprouse, 2013)..

In an AJT task, participants are presented with one sentence at a time, and they are required to judge whether the sentence is grammatically acceptable or not by using either a dichotomous response (yes/no) or a Likert-scale response. The task stimulus usually consists of both grammatical and ungrammatical sentences (R. Ellis, 1991). Judgement data is often preferred to traditional production data as the latter can include production errors that are caused by slips of the tongue, avoidance, phonological complexity, difficulties in memory retrieval, amongst other factors (Ionin & Zyzik, 2014). In other words, judgement data allows the researcher to investigate linguistic structures that occur rarely or are difficult to elicit from learners' production data (R. Ellis, 1991; Loewen, 2009). Also, the task itself is relatively simple to design and administer to a large number of participants at once, and the scoring process is much less complicated than analysing production data (Plonsky et al., 2020). This practicality side has contributed to the widespread use of AJT in L2 research.

Researchers who employ AJTs often need to make various decisions related to the design. One of the major decisions is whether to make the task time-constrained or not, as this decision may affect the types of L2 knowledge that can be elicited (Plonsky et al., 2020). Many studies, employing both timed and untimed versions of AJTs, have found that timed judgement tasks may be more likely to tap into participants' implicit knowledge due to the limited time available, making it challenging for participants to access explicitly learned rules (Bowles, 2011; R. Ellis, 2005; R. Ellis & Loewen, 2007; Loewen, 2009). Godfroid et al. (2015) studied learners' performance in timed and untimed judgment tasks by tracking their eye movements. They found that learners regressed more often on untimed items, leading them to suggest that untimed judgment tasks may elicit more use of explicit knowledge compared to the timed ones. However, Gutiérrez (2013) employed both timed and untimed AJTs to test a group of L2 Spanish learners, and argued that time condition does not affect knowledge types whereas stimulus type does. By using both exploratory and confirmatory factor analyses, the researcher found that learners' scores for ungrammatical sentences in both timed and untimed AJTs loaded on the same construct interpreted as explicit knowledge and their scores for grammatical sentences in both tasks loaded on the same construct interpreted as implicit knowledge (Suzuki, 2017b). More recently, Vafaee et al. (2017) further investigated if manipulating the construct of AJTs leads to measures of different knowledge types by comparing AJT scores with two online processing tests (self-paced reading and word-monitoring task) and one metalinguistic knowledge test. They found that AJTs draw participants' attention to form, therefore regardless of being timed or not, they are more likely to tap into explicit knowledge. The authors concluded that online processing tasks are better measures of implicit knowledge.

Some other design decisions may also influence learners' utilisation of implicit or explicit knowledge and, in a broader sense, impact their performance in AJTs. For example, Murphy (1997) adopted both written and aural AJT tasks in his study, and found that L2 learners were not only slower but also less accurate when the sentences were presented aurally than when presented visually. Kim and Nam (2017) went further and argued that timed aural AJT task can measure stronger implicit knowledge. Apart from task modality (written/aural), if the AJT task requires participants to identify errors or provide corrections during the judgment process, it may lead to an increased reliance on explicit and/or metalinguistic knowledge (R. Ellis, 1991). In addition, other design features, such as response types (Ionin & Zyzik, 2014) and the difficulty of target features (Shiu et al., 2018), can all play a role in impacting learner performance in AJTs.

Despite the ongoing debate about the types of knowledge AJTs tap into, judgement tasks without time pressure seem to be more effective in eliciting explicit knowledge compared to tasks with time constraints. For this reason, this study employed an untimed AJT in written modality to measure participants' explicit knowledge. The purpose was to complement the data from the two online processing tasks (i.e., eye-tracking and sentence-matching) that arguably measure (more) implicit knowledge. This type of combination of implicit and explicit tasks has been widely adopted in previous studies. By comparing participants' performance in the implicit and explicit tasks, we can better examine their knowledge of English tense-aspect structures.

5.5.2 Stimuli Design

The critical stimuli in the AJT were identical to those in the two online tasks. Participants were presented with the same critical items they had seen in the eye-tracking (54 items) and SM task (48 items). For example, if one participant was assigned to List 1 in the eye-tracking and SM task, he/she would have been presented with the critical items from the same lists in the AJT. Since critical sentences from the eye-tracking task were all grammatical, another 54 ungrammatical filler items were constructed to ensure a balance between the number of grammatical and ungrammatical sentences. An example of filler item is presented in (3). Also, since three different target structures were investigated, distractors were self-incorporated (Mackey & Gass, 2005); for example, the past simple items served as distractors for present progressive items, and so on.

(3) **The only piece of furniture George has in his bedroom are a bed.*

The meta-analysis on judgement tasks in L2 research, conducted by Plonsky et al. (2020), suggests that the median number of total items in previous AJT studies was 50, but the range of minimum and maximum item numbers varied from 4 to 300, which can potentially lead to participant fatigue in many instances. However, it was not realistic to decrease the number of items in the current AJT because the same items as in the two online tasks had to be used for methodological purposes. Therefore, to minimise the potential effect of participant fatigue or boredom, all participants were informed that they could stop and take breaks during the task if they needed to. Some researchers (e.g., Ionin et al., 2021) also suggest that the AJT can better measure learners' explicit knowledge if the target region is underlined or bolded. This approach is to make sure that participants are aware of the structures they are judging, particularly when the stimulus materials involve multiple sentences. For example, in Ionin et al.'s (2021) study, each item included two sentences: the first established the context, and the second contained a grammatical error (e.g., "*Mary felt lonely last week. So she finally got cat from a shelter*") (p. 128). In contrast, the present study used only one isolated sentence per stimulus. Therefore, the target structure was not underlined, as participants could clearly identify the sentence they were judging.

5.5.3 Administration of the Task

The AJT was constructed and administered using the same Experiment Builder on the desktop computer in the lab. The task began with an instruction page, followed by five practice trials aimed at familiarising the participants with the task. Each stimulus sentence was presented in isolation, and participants were asked to use the mouse to indicate their judgements on a 7-point scale, from completely unacceptable to completely acceptable. An example is presented below (4).

(4) *Two weeks ago, Mike went to the birthday party of his best friend.*

1	2	3	4	5	6	7
○	○	○	○	○	○	○
Completely unacceptable						Completely acceptable

At the beginning of each new trial, the mouse position reset to the centre. This helps minimise the influence of previous responses or cursor placement and prevent potential biases in participants' responses. The entire task lasted approximately 25-30 minutes, and participants were offered breaks during the task.

5.5.4 Data Analysis

Prior to analysis, participants' rating responses were z-score transformed. The z-score transformation is a linear transformation that can maintain all of the relationships within the data, and it does not introduce any distortion or change in the relative relationships between data points (Winter, 2019). To compute the z-score of a participant's responses, the mean and standard deviation of all their judgments were first calculated; then, each individual judgment was subtracted from this mean, and the result was divided by the participant's standard deviation (Schütze & Sprouse, 2013). These resulting z-score values express each response in units of standard deviation from that participant's mean; for example, a positive z-score indicates a rating higher than the mean, whereas a negative z-score suggests a rating lower than the mean. This z-score transformation has been recommended as a routine for analysing Likert scale data by many researchers (e.g., Casasanto et al., 2010; Schütze & Sprouse, 2013; Spinner & Gass, 2019), since points on the scale may have different meanings to different participants even if clear instructions are provided. Therefore, converting their responses to z-scores can help address potential biases in Likert scale data, such as scale compression (e.g., participants predominantly use only 3-5 on a 1-7 scale) and scale skew (e.g., participants mostly use the ends of the scale: 1 or 7) (Schütze & Sprouse, 2013). The primary critique of using z-score transformation is that it treats Likert scale data, which is ordinal in nature, as continuous (Busch, 1993). However, as Schütze and Sprouse (2013) argued, when using parametric statistical tests such as t-tests, ANOVAs, or linear mixed effects models with Likert scale data, there is an implicit assumption that treating it as continuous is suitable for practical analysis.

In this study, the mixed-effects models were employed to analyse the z-score transformed AJT data. Firstly, three separate overall models were constructed for the three structures: past simple, present perfect, and present progressive. The fixed effects included *group* (L1 English vs. L1 Chinese/L1 Arabic) and *condition* (grammatical vs. ungrammatical), and their interaction. A dummy coding scheme was adopted. For the *group* variable, the L1 English group was set as the reference level, and for the *condition* variable, the grammatical condition was assigned as the reference level. Similar to the sentence-matching data analysis, the maximal random effect structures included by-participant and by-item random intercepts, and by-participant random slopes for condition and by-item random slopes for group and condition. When the maximal model failed to converge, the random effect structures were gradually trimmed until convergence was achieved.

In addition, within-group analyses were performed for each structure to determine whether the grammatical and ungrammatical conditions were rated differently. For these within-group models, the fixed effect only included *condition*, and the maximal random effect structures contained *by-participant* and *by-item* random intercepts, and *by-participant* random slopes for condition and *by-item* random slopes for condition.

5.6 Elicited Imitation Task

5.6.1 Methodological Considerations

Elicited imitation (EI) is a task that involves having participants listen to a number of spoken sentences which they are asked to repeat as precisely as they can (Erlam, 2006). The underlying assumption is that if learners possess a relevant mental representation of the linguistic information embedded in the sentence, they will be capable of decoding the acoustic information, associating sounds with corresponding structures and meanings, and ultimately reconstructing the intended meaning by converting selected structures back into sounds (Yan et al., 2016). Conversely, if they do not have an internal representation of relevant features, repeating the stimuli that are too long to be stored in short-term memory will be very challenging (Erlam & Wei, 2021; Rebuschat & Mackey, 2013).

There have been two main research purposes identified in the utilisation of the EI task. The first trend is the use of EI as a global measure of L2 proficiency (see Kostromitina & Plonsky, 2022; Yan et al., 2016 for reviews). The EI task has been found to be a valid and reliable tool to discriminate learners of different proficiency levels across various L2 languages, such as English (S.-L. Wu et al., 2022), French (Gaillard & Tremblay, 2016), Korean (Y. Kim et al., 2016), Chinese (S.-L. Wu & Ortega, 2013), and Spanish (Bowden, 2016).

The second trend, particularly relevant to this study, involves the application of EI as a potential measure of implicit knowledge. The supporting evidence for the validity comes from a series of studies. For example, Erlam (2006) administered an EI task to 20 native English speakers and 95 L2 English learners. The stimuli included 34 sentences/statements focusing on 17 grammar features, and each grammatical feature had two sentences: one grammatical and one ungrammatical. The sentences ranged in length from 8 to 18 syllables. The participants were asked to listen to a sentence/statement (e.g., *To speak English well you must study for many months*), then express their opinion about the statement using a beliefs questionnaire (i.e., true, not true, not sure), and lastly repeat the sentence they heard using correct English. The results indicated that the native speakers corrected 91% of the ungrammatical sentences while L2 learners corrected 35% of these sentences. The correlation analysis found no significant relationship between the syllable length and repetition success, indicating that the elicited utterances were not merely memorized verbatim. Also, significant correlations were found between participants' performance in the EI task and the other two measures tapping into implicit knowledge (i.e., the oral narrative test, IELTS speaking test).

Erlam concluded that the correlations observed provided evidence that the EI was indeed measuring learners' implicit knowledge. Similarly, some other studies using either exploratory or confirmatory factor analysis also indicated that the EI task, along with other implicit tests (i.e., oral native tests, timed AJT), loaded on the same construct labelled as implicit knowledge (Bowles, 2011; R. Ellis, 2005; R. Ellis & Loewen, 2007; Spada et al., 2015; Zhang, 2015). Ellis (2009) even argued that among the three tests, the EI task was their best measure of implicit knowledge.

However, Suzuki and DeKeyser (2015) cast doubts on the validity of EI as a measure of implicit knowledge. Their study focused on advanced-level Japanese L2 learners, and compared their performance in three tasks, including an EI task with the built-in word monitoring task, a metalinguistic knowledge task (MKT), and a probabilistic serial reaction time (SRT) task (i.e., a measure of aptitude for implicit learning). The results revealed that participants' EI performance correlated significantly with their metalinguistic knowledge, but not with SRT performance. Conversely, the word monitoring task showed a significant correlation with SRT, while no significant correlation with the metalinguistic test was observed. The authors argued that psycholinguistic measures (e.g., word monitoring tasks) are better for assessing implicit knowledge whereas EI tasks seem to only measure automatised explicit knowledge (i.e., involves awareness). However, a point to note with this study is that the participants were explicitly instructed to 'correct' the errors in the sentences, which could have potentially made them aware of the existence of ungrammatical sentences and led them to utilise explicit knowledge (Erlam & Wei, 2021; Sarandi, 2020). Moreover, given half of the participants had backgrounds in Japanese linguistics and all were advanced learners, the possibility of accessing well-automatised explicit knowledge was likely enhanced (Suzuki & DeKeyser, 2015).

For the EI task design, manipulating certain aspects can increase the likelihood of accessing participants' implicit knowledge (Erlam, 2006, 2009). Firstly, the task must be conducted under time pressure to prevent participants from planning or monitoring their responses (R. Ellis, 2005). Secondly, to avoid rote repetition, the stimulus sentences should be carefully designed. Their length should exceed participants' short-term memory capacity because if the stimuli are too short, they can be memorised easily as chunks (Bley-Vroman & Chaudron, 1994). However, determining the suitable length of stimuli is tricky due to differences in participants' age, working memory capacity, and language proficiency (Spada et al., 2015). For example, Mackey and Gass (2005) recommended a sentence length between 12 and 17 syllables, while Graham et al. (2008) proposed a range of 6 to 19 syllables for adults. Erlam (2006) suggested that if there is no significant correlation found between stimuli length and participants' success in imitating stimuli, it can be interpreted as evidence that participants are not simply engaging in rote memorization but using their linguistic knowledge. Also, inserting a delay between the presentation of the stimuli and the repetition can help prevent rote repetition (Vinther, 2002). For example, McDade et al. (1982) found that introducing a three to five-second time interval helped prevent rote repetition. To further prevent participants from planning their responses during this time interval, an interruptive or secondary task can be introduced, such as an opinion questionnaire (Erlam, 2006), a word clicking task (Yan, 2015), or a picture selection task (Erlam & Ellis, 2018).

In addition, the inclusion of ungrammatical sentences in the stimuli is also more likely to elicit implicit knowledge. Many researchers (e.g., Hamayan et al., 1977; Munnich et al., 1994; Schimke, 2011; Schimke & Dimroth, 2018; Smith, 1973; Verhagen, 2009) have argued that the automatic correction of ungrammatical stimuli by participants serves as convincing evidence that they have internalised the target grammatical knowledge. In other words, if participants correct the errors in ungrammatical sentences and leave grammatical sentences unchanged, it can be assumed that they have implicit knowledge of relevant grammar. Nevertheless, a few studies (Gallimore & Tharp, 1981; Markman et al., 1975) have found that native speakers often reproduce ungrammatical sentence without making corrections, particularly when task instructions do not explicitly require them to do so. This poses a challenge because if there are explicit instructions focusing on error correction, it might make participants aware of the presence of ungrammatical sentences and trigger the utilisation of explicit knowledge (Suzuki & DeKeyser, 2015). However, a recent study conducted by Granena (2016) revealed that there was no significant difference in L2 English participants' EI performance based on whether they were instructed to 'repeat the sentence in correct English' or 'repeat the sentence'.

Despite the challenges in its design, this study employed an EI task aimed to measure participants' online production of English tense-aspect based on their implicit knowledge. The reason for selecting EI over other production tasks is that, as a more controlled task, EI tends to be more reliable in eliciting specific linguistic structures (Erlam, 2006; Jessop et al., 2007). While participants can avoid using certain structures in spontaneous oral production, they cannot avoid them in the EI task. In terms of design features, the task was performed by the participants under time pressure, and ungrammatical sentences were included in the stimuli, and, more importantly, an interruptive task (i.e., a drawing task) was introduced to create a delay and prevent rote memorization. All these elements aimed to maximise the possibility of tapping into participants' implicit knowledge.

5.6.2 Stimuli Design

A total of 48 critical items were created, with each item consisting of the three tense-aspect structures in both grammatical and ungrammatical versions. The ungrammatical sentences were constructed similarly to those in the sentence-matching task by introducing a mismatch between the temporal adverbial at the beginning of the sentence and the verb that follows. The length of the critical sentences ranged from 12 to 21 syllables, with an average of 16 (*SD* = 1.8), which aligns with similar studies involving second language learners of comparable proficiency (e.g., Eisenstein et al., 1982; Jensen & Vinther, 2003; Munnich et al., 1994). Despite the lack of consensus in the literature regarding optimal stimulus length, the chosen length was considered suitable due to the participants' upper-intermediate English proficiency and higher educational background. In addition, focusing on three different tense-aspect structures allows sentences from each structure to serve as distractors for one another.

Table 5.6.1

An Example of a Critical Item Used in Elicited Imitation Task

Structure	**Condition**	**Example auditory sentence**
Simple Past	Grammatical	a. *Last Christmas, James thought about buying a new car for my dad.*
	Ungrammatical	b. **Since last Christmas, James thought about buying a new car for my dad.*
Present Perfect	Grammatical	c. *Since last Christmas, James has thought about buying a new car for my dad.*
	Ungrammatical	d. **Last Christmas, James has thought about buying a new car for my dad.*
Present Progressive	Grammatical	e. *Right now, James is thinking about buying a new car for my dad.*
	Ungrammatical	f. **Lately, James is thinking about buying a new car for my dad.*

Six experimental lists were generated, each containing an additional 16 filler sentences (8 grammatical and 8 ungrammatical sentences). An example of filler sentences is presented in (5). The filler sentences were of similar length to the critical sentences.

Participants were presented with one of the six lists, ensuring they were exposed to only one version of each critical item. The stimuli in each list were randomised for every participant.

(5) a. *This time tomorrow, my friend and I will be flying to Paris.* (Grammatical)

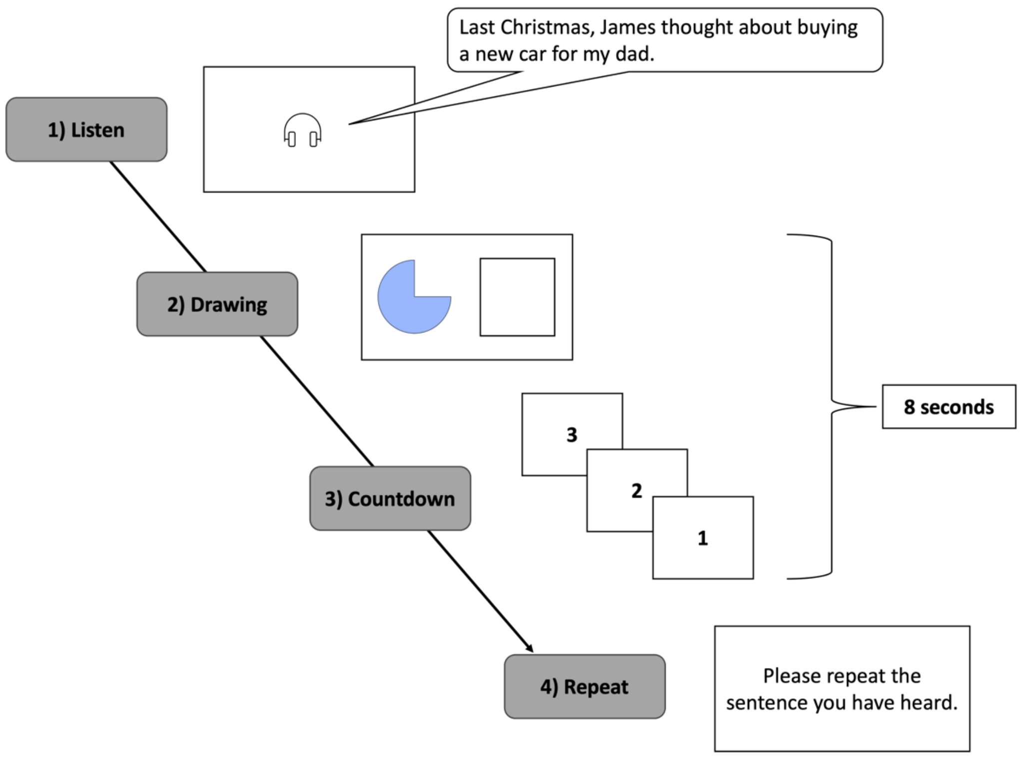
b. **There is five man and four women standing in the room.* (Ungrammatical)

5.6.3 Administration of the Task

The EI task was built on Gorilla Experiment Builder and administered through a laptop in the lab. An illustration of the task procedure is presented in Figure 5.6.1.

Figure 5.6.1

An Illustration of the Elicited Imitation Task Procedure



The task began with the instruction page, where the participants were not informed of the presence of ungrammatical sentences and were simply instructed to repeat what they had heard. If they could not recall the exact wording, they were given the flexibility to use paraphrasing as long as the intended meaning remained consistent. This helped to direct participants' attention to the meaning of the sentences and enhance the likelihood of assessing their implicit knowledge. To ensure familiarity with the task procedure, the participants were given five practice trials.

During the task, the participants heard a stimulus sentence through headphones. Then, the screen proceeded automatically, and a figure (e.g., a heart, a triangle) was shown on the left-hand side. The participants were required to draw the same figure on the right side of the screen using either the touchpad or the mouse. They had five seconds to complete their drawing. This interruptive task was aimed at preventing immediate repetition, which could potentially lead learners to rely on rote memorisation (McDade et al., 1982). Following the drawing task, a three-second countdown appeared on the screen. After the countdown, the participants were prompted with the message, "Please repeat the sentence you have heard." They were given 15 seconds to reproduce the sentence before the next trial began. Their responses were recorded automatically. If a participant found themselves unable to repeat the sentence, they were provided the option to simply skip the trial. The entire task lasted approximately 20 minutes.

5.6.4 Data Cleaning and Scoring

After the task, the recorded responses of all critical stimuli were transcribed and scored by the researcher. During the transcribing process, certain answers were marked as missing values and were excluded from the analysis. These exclusions were made for various reasons, including cases where participants omitted the imitation of verbs, skipped trials, or provided responses that were difficult to transcribe due to background noise or unclear pronunciation. Approximately 3.06% of the entire dataset was impacted by these exclusions.

Table 5.6.2

Example of the Scoring Method for the Elicited Imitation Task, Exemplified by Participants' Imitation Responses

	Change type	Scoring	Stimulus sentence	Imitation answer
Grammatical	G → G	0	Last week, Linda planted new flowers in her garden.	Last week, Linda planted new flowers in her garden.
	G → UG	1	For the past year, my father has taught English to foreign students at the school.	*For the past year, my father taught English to foreign students at the school.
Ungrammatical	UG → UG	0	*Since last Christmas, James thought about buying a car for my dad.	* Since last Christmas, James thought about buying a car for my dad.
	UG → G	1	*Since 2015, Richard studied music at the University of York.	Since 2015, Richard has studied music at the University of York.

Regarding the scoring process, only the first response from the participants for each trial was scored, regardless of whether the participants corrected themselves in later attempts. The scoring method focused on changes made by the participants to the forms of structures under investigation and not on the overall accuracy of the imitation, following the practices of some previous researchers (e.g., Cristante, 2020; Schimke, 2011; Schimke & Dimroth, 2018). Therefore, an unchanged response (i.e., verbatim repetition) was given a score of “0” and a changed response was given a score of “1”. More specifically, for the grammatical sentences, if the participants correctly repeated the grammatical sentence without changing it, they received a score of “0”. On the other hand, if they inserted an error into the grammatical sentence, they received a score of “1”. Similarly, for the ungrammatical sentences, if participants corrected them, they were given a score of “1”. If they simply repeated them without correction, a score of “0” was given to them. An example of the scoring procedure is presented in Table 5.6.2.The rationale for adopting this scoring method is that when learners repeat sentences and make changes to them, these changes can reflect their internalisation of grammatical knowledge. If learners insert errors into the grammatical sentences while repeating, it may indicate a lack of relevant grammatical knowledge. Conversely, if they correct the ungrammatical sentences, it shows that they have the knowledge of relevant grammatical structures. In this case, correctly repeating grammatical sentences and imitating ungrammatical sentence verbatim are not very informative because it could be due to learners memorising the sentences as a chunk without constructing the meaning themselves (Cristante, 2020). Even though several task design features were incorporated (e.g., time pressure, an interruptive task, a delay between the presentation of the sentences and repetition) to direct participants’ attention to meaning and maximally tap into their implicit knowledge, there could still be chances where participants focused on form and accessed their explicit knowledge, particular when repeating the sentence verbatim and not making any changes.

It is important to emphasise that the scoring process solely concentrated on the specific linguistic elements being targeted (i.e., the time adverb and the following verbs). For example, in the obligatory past simple context, the scoring focused on whether the verb was inflected with *-ed*, and in the present perfect context, whether the auxiliary verb *have/has* and the past participle of the verb were present. The analysis did not take into account the use of substitute words and any grammatical mistakes (e.g., articles, pronouns, plurals) that occurred outside the scope of that focus.

5.6.5 Data Analysis

Since the dependent variable was binary (i.e., 0/1) in this case, the logistic regression models, a type of a generalised linear mixed-effects model (Linck & Cunnings, 2015), were employed for the analysis using the *lme4* package (Bates et al., 2015). Firstly, three overall models were built separately for each grammatical structure. The fixed effects included *change type* (grammatical → ungrammatical vs. ungrammatical → grammatical), *group* (L1 English vs. L1 Chinese/L1 Arabic) and their interaction. A dummy coding scheme was adopted. For the *change type* variable, grammatical → ungrammatical (G → UG) was set as the reference level, and for the *group* variable, the L1 English group was assigned as the reference level. The model started with maximal random effect structures, which included by-participant and by-item random intercepts, and by-participant random slopes for *change type* and by-item random slopes for *group*. When the maximal model failed to converge, the optimising function was applied first before trimming down the random effect structures until convergence was achieved.

Apart from the overall models, three within-group models were also constructed for each structure to examine whether each group showed a significant preference for one of the change types (i.e., if they showed a tendency to correct the ungrammatical sentences or introducing errors into the grammatical sentences). In within-group models, the fixed effect only included *change type*, and the maximal random effect structures included by-participant and by-item random intercepts, and by-participant random slopes for *change type*.

In addition, the errors introduced by the L2 participants when repeating grammatical sentences were coded and analysed. The frequency of each error type was calculated and compared between groups when necessary.

5.7 Gap-Filling Task

5.7.1 Methodological Considerations

Gap-filling test is a type of cloze test which involves exercises where certain words are removed or deleted from a text, and the participant taking the test is then required to fill in the omitted words (Davies et al., 1999). In some cases, specific words are rationally removed for a particular purpose; for example, prepositions are deleted to test participants' knowledge of preposition usage (McCray & Brunfaut, 2018). Some scholars (e.g., Alderson, 2000) describe this type as gap-filling test, while others (e.g., Oller & Jonz, 1994) use the term rational-deletion cloze test.

Gap-filling test has been employed by numerous studies as a measure of overall L2 proficiency (e.g., J. Brown, 2002; Bylund et al., 2012; Jonz, 1990; Slabakova, 2000; Tremblay & Garrison, 2008), since this test requires learners to employ multiple language skills simultaneously, engaging in complex language processing while focusing on the content (Slabakova, 2000). It has been found that the gap-filling task results showed a significant correlation with standardised ESL tests in predicting learner's proficiency levels (Hanania & Shikhani, 1986; Oller & Conrad, 1971). In addition, different formats of gap-filling tests have also been used to investigate L2 vocabulary and collocation development (Nizonkiza, 2012; Sonbul et al., 2023). As argued by Sonbul et al. (2023), they provide a more direct measure of learners' vocabulary knowledge compared to corpus-based evidence, and researchers can manipulate variables such as frequency and congruency.

More relevant to the present study, the gap-filling task has also been used to study the developing (productive) grammar of L2 learners and specifically, their ability to produce certain elements such as verbs, articles, prepositions, and particles explicitly (Hyltenstam, 1983; Macrory & Stone, 2000; Oller & Inal, 1971). Among these studies, one commonly used task format is to provide a sentential context (not a whole paragraph) and ask learners to complete the missing element. It is believed that by focusing on one sentence at a time, L2 learners, particularly from formal learning settings, are more likely to employ their conscious knowledge of rules they have formally studied, even though these rules may not yet have become integrated into their competence (Dulay et al. 1982, cited in Slabakova, 2000). From this perspective, learners' explicit knowledge is more likely to be tapped into using the gap-filling task (Asiyaban et al., 2020; R. Ellis, 2005).

In the present study, a gap-filling test was used to assess participants' use of morphosyntactic markings in written production, with a specific focus on their use of temporal/aspectual morphology. The purpose of this test was to measure their explicit knowledge in written production and to make a comparison between their explicit knowledge and implicit productive knowledge, which was measured in the EI task.

5.7.2 Stimuli Design

Given that the gap-filling task was intended to complement the online production data collected from the elicited imitation task, it was crucial that the stimuli used in both tasks were identical. Therefore, if the participant was assigned to list 1 in the EI task, the person would be given the same critical sentences from that list. An example of stimuli presentation is provided in (6). All filler items were excluded in this task, leaving 48 items in each list.

(6) *Last Christmas, James ______ (think) about buying a new car for my dad.*

5.7.3 Administration of the Task

The gap-filling task was administered using pencil-and-paper. All the participants sat by a desk and completed the test in the lab. They received instructions to fill in the blanks with the suitable forms of the provided verbs. In addition, participants were directed to inflect verbs when needed and to incorporate the necessary auxiliaries as required. Three example sentences were filled in as an illustration. In case the participants forgot the test requirements, the instructions were printed on a separate page and were provided for them during the test. Although the participants performed the task without time pressure, the majority of them could complete it in approximately 5 to 10 minutes.

5.7.4 Data Scoring and Analysis

The scoring process for the gap-filling task was straightforward. Participants were given a score of "1" if they provided the correct verb form for the given context (i.e., v-ed for past simple, have/has+v-ed for present perfect, and be +v-ing for present progressive); otherwise, they received a score of "0". The suppliance of incorrectly conjugated verbs such as *feeled* instead of *felt* was considered correct, since the emphasis was on eliciting past morphology in past simple contexts. If the participants demonstrated knowledge about the appropriate usage of past tense, it was marked as correct.

Similar to the EI analysis, logistic regression models were employed due to the binary nature of the dependent variable (i.e., 1/0). Three separate models were constructed for each linguistic structure to compare group differences. The fixed effect included *group* (L1 English vs. L1 Chinese/L1 Arabic) with the L1 English group assigned as the reference level, and the maximal random effect structures consisted of by-participant and by-item random intercepts, and by-item random slopes for *group*. All three maximal models successfully converged through the optimization process.

5.8 Overall Data Collection Procedures

All participants were tested individually in a single session in the eye-tracking laboratory at the university. Upon arrival, the participants received a consent form accompanied by an information sheet. They were required to read all the information on the sheet thoroughly before signing the consent form. They were also given a chance to ask any questions they had about the experiment. Following this, all the participants completed a language background questionnaire aimed at collecting data regarding about their age, gender, academic level, first language(s), other languages they speak and proficiency levels. The L2 participants were also asked about their English learning experiences, IELTS test or other standardised English test scores, and they were also asked to self-report their English proficiency level on a 1 to 6 scale.

Following the language background questionnaire, participants completed three tasks designed to tap into implicit knowledge. These tasks were conducted sequentially: the visual-world eye-tracking task, the sentence-matching task, and the elicited imitation task. These implicit tasks were followed by a more explicit measure of comprehension (i.e., acceptability judgement task). Finally, participants undertook the explicit production task (i.e., gap-filling test). The order of task administration was arranged in this format to prevent, as far as possible, participants from becoming conscious of the tested structures during the implicit tasks and to maintain their sensitivity to the stimuli. Participants could take breaks between tasks if they needed to. Also, since the AJT lasted longer and was conducted without time pressure, the participants could stop during the task and have a break to prevent fatigue.

After the experiment, the L2 participants took a short break and undertook the Oxford Quick Placement test in a paper-and-pencil format. The entire experimental session ranged from approximately 1 to 1.5 hours for native speakers and from 1.5 up to 2 hours for L2 participants. All participants were offered Amazon vouchers (£5 for native speakers and £10 for L2 participants) as a token of appreciation for their participation and time.

5.9 The Pilot Study

Piloting refers to administering the instrument to a small group of people from the target population (Mackey & Gass, 2011). In particular, when the study involves a control group, it is essential to ensure that the control group can perform at ceiling on the test materials (Mackey & Gass, 2011). For example, when both grammatical and ungrammatical sentences are included in the stimuli, as in the present study, if native speakers do not show any grammaticality effect, this may indicate some potential problems with the stimulus sentences. Also, the pilot results can provide feedback on the task procedures, the clarity of item phrasing and instructions, and completion time of the whole experiment, etc. (Mackey & Gass, 2011).

The pilot study involved a total of six participants, including two native English speakers, two Chinese learners of English, and two Arabic learners of English. For the visual-world eye-tracking task, it was found that participants' offline responses for a critical item (e.g., The strong boy is lifting/lifted the weights up his head) did not align with expectations. One possible reason could be that the pictures used were not clear enough to convey the differences. To address this issue, an up-arrow sign '↑' was added to the image to indicate the action of *lifting*.

For the sentence-matching task, two L2 participants expressed that the 3000 ms time limit for viewing the sentences was not enough. In response to their feedback, the time limit was adjusted to 3500 ms to make sure that participants had sufficient time to read the sentence while preventing excessive time that may allow them to examine the form of the sentences.

In terms of the acceptability judgment task, since it included all stimulus sentences from the eye-tracking and sentence-matching tasks, participants were observed to experience fatigue easily. However, reducing the number of stimulus items was not feasible due to the experimental design. Therefore, during the official data collection stage, participants were allowed to pause the task and take breaks, as the task itself was not timed.

For the elicited imitation task, the original instruction explicitly stated 'Please repeat the sentence in correct English', in line with practices adopted by other researchers (e.g., Erlam, 2006; Suzuki & DeKeyser, 2015) who used the same task. However, participants expressed that this instruction made them pay extra attention to the grammar of their repetitions. Therefore, the task instruction was revised to 'Please repeat the sentence you have heard' to ensure that participants remained unaware of the presence of ungrammatical sentences in the stimuli and to increase the likelihood of assessing their implicit grammatical knowledge in production.

Chapter 6 Results

This chapter reports the results of the five tasks, starting with comprehension tasks, which include visual-world eye-tracking (VWP), sentence-matching (SM), and acceptability judgment tasks (AJT). It then progresses to the production tasks, specifically the elicited imitation (EI) and gap-filling tasks. Each task section begins with a descriptive presentation of the results, including visualisation plots, followed by inferential statistical outcomes. A brief summary of findings is also provided at the end of each task section.

6.1 Visual-World Task Results

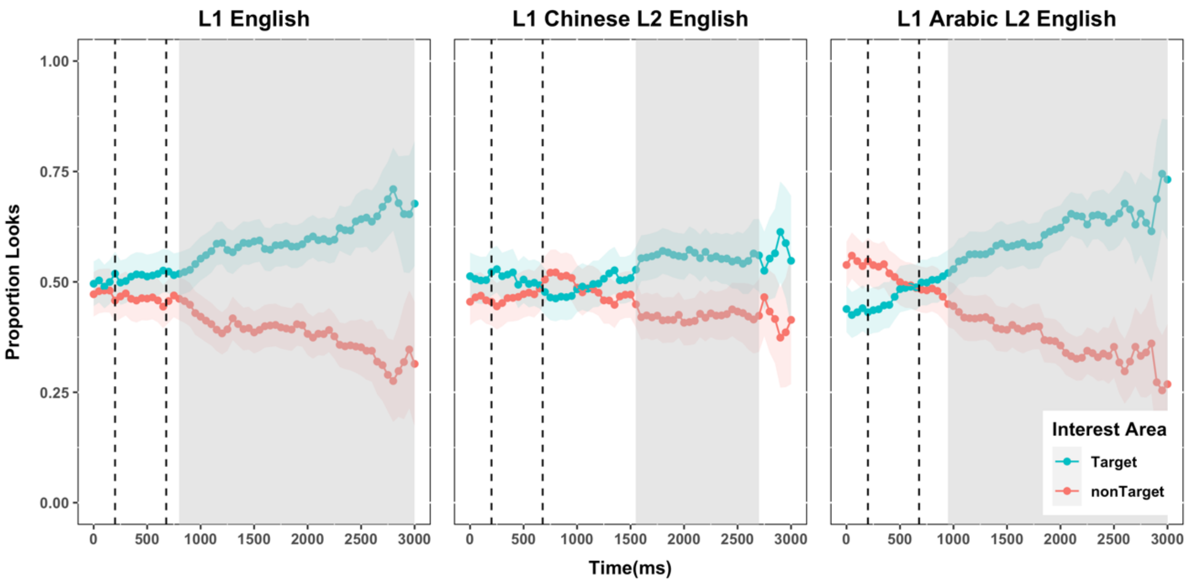
This section presents the results from the visual-world eye-tracking task. This task was used to investigate whether participants were able to use morphological cues to encode event information in online sentence comprehension. It was assumed that if L2 participants could process the verbal morphology (i.e., *V-ing, V-ed and have+V-ed*), native-like fixation patterns would be observed. The results for the past simple items are reported first, followed by present perfect and present progressive items.

6.1.1 Past Simple

Figure 6.1.1 illustrates the proportion of fixations in the target area (i.e., completed event) and non-target area (i.e., ongoing event) within a 3000 ms window as participants listened to simple past sentences. This window extended from 200 ms before the verb onset to 2800 ms after the verb onset. The figure shows that the L1 English group exhibited a steady increase in their looks to the target area after the onset of the verb (the first vertical dashed line). In contrast, the L1 Arabic group initially fixated on the non-target area, but their focus on this area decreased after the verb onset, and they looked much more at the target after the verb offset. Interestingly, the L1 Chinese group continued to increase their looks to the non-target area even after the onset of the verb, and the point at which they started looking more at the target occurred much later than in the other two groups.

Figure 6.1.1

Average Fixation Proportions of Looking at the Interest Areas by the Groups When Hearing Past Simple Sentences in the Visual-World Eye-Tracking Task

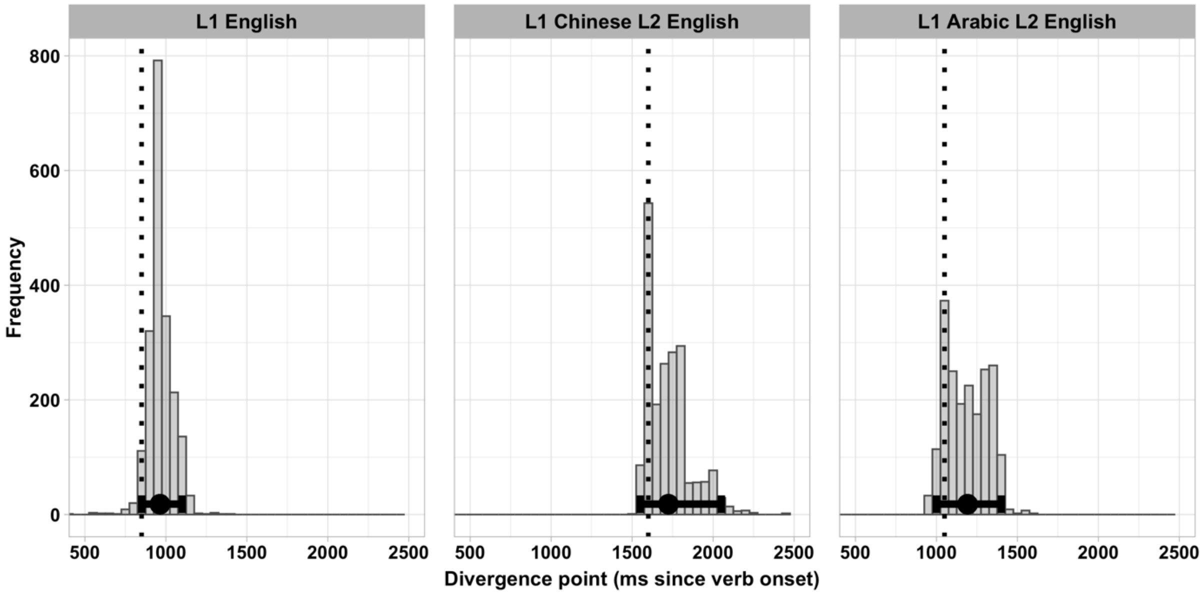


Note. The two vertical dashed lines represent the average verb duration (200-678 ms). Shaded areas represent the significant time clusters identified through the cluster-based permutation analyses.

The cluster-based permutation (CBP) analyses revealed clusters of time windows in which looks to the target and non-target areas significantly differed within the three groups. In the L1 English group, the significant time clusters extended from 800 to 3000 ms (cluster mass statistic = 2554, sum t = 240, p < .001). Since the t values were above 0, the direction was positive, indicating that their proportions of looking at the target were significantly higher than those at the non-target. Similarly, for the L2 learner groups, significant positive time clusters were also identified: 1550-2700 ms in the L1 Chinese group (cluster mass statistic = 554, sum t = 80, $p < .001$) and 950-3000 ms in the L1 Arabic group (cluster mass statistic = 2618, sum t = 232, $p < .001$). This implies that both L2 groups exhibited a preference for the target area at some point during the identified time windows.

Figure 6.1.2

Bootstrap Distributions of Divergence Points of Target Over Non-Target for the Past Simple Items in the Groups

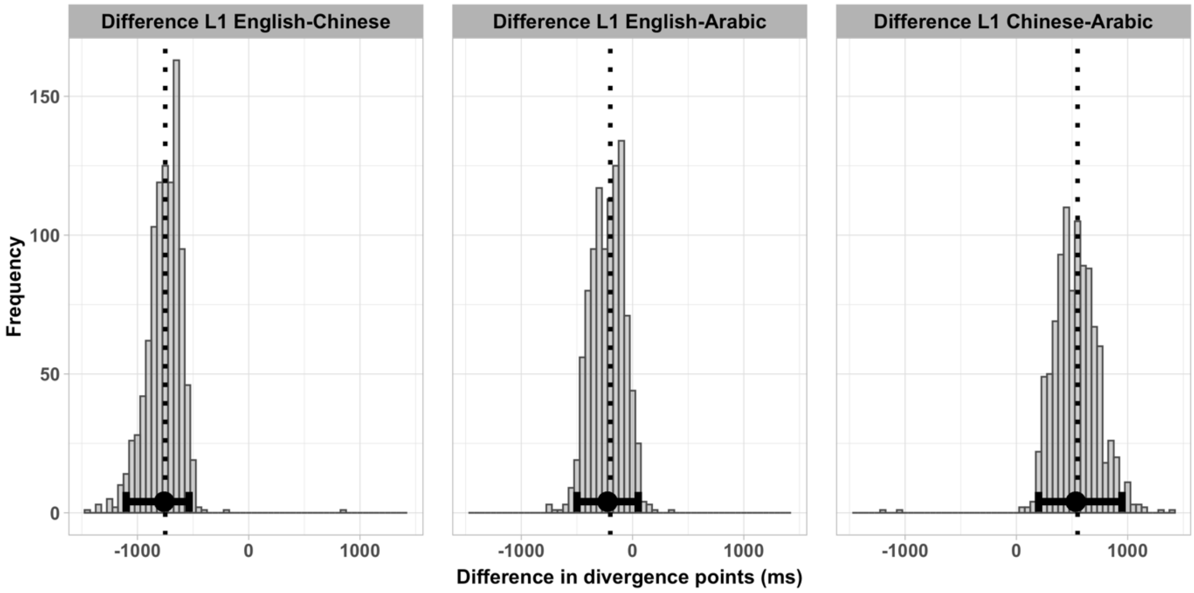


Note. The x-axis shows the distribution of divergence points based on 1000 bootstraps. The y-axis shows the number of resamples where a particular divergence point was observed. The error bars indicate the bootstrap mean and its 95% CI. Dotted vertical lines correspond to the divergence points in the original dataset.

As discussed in Section 5.3.5, CBP analyses only indicate the presence of an effect, without providing information on when the effect started. To estimate when the looks to the target began to diverge from the looks to the non-target in each group, complementary divergence point analyses were also conducted. The results of the bootstrap distribution are plotted in Figure 6.1.2. In the L1 English group, it was found that the mean divergence point was 963 ms, 95% CI = [850, 1100] ms. For the L1 Chinese group, the mean divergence point was 1724 ms, CI = [1550, 2050] ms. In the L1 Arabic group, the mean divergence point was 1192 ms, CI = [1000, 1400] ms.

Figure 6.1.3

Bootstrap Distributions of the Difference in Divergence Points Between Groups for the Past Simple Items



Note. Points and error bars indicate bootstrap means and 95% CI. Dotted vertical lines indicate mean divergence point differences in the original data.

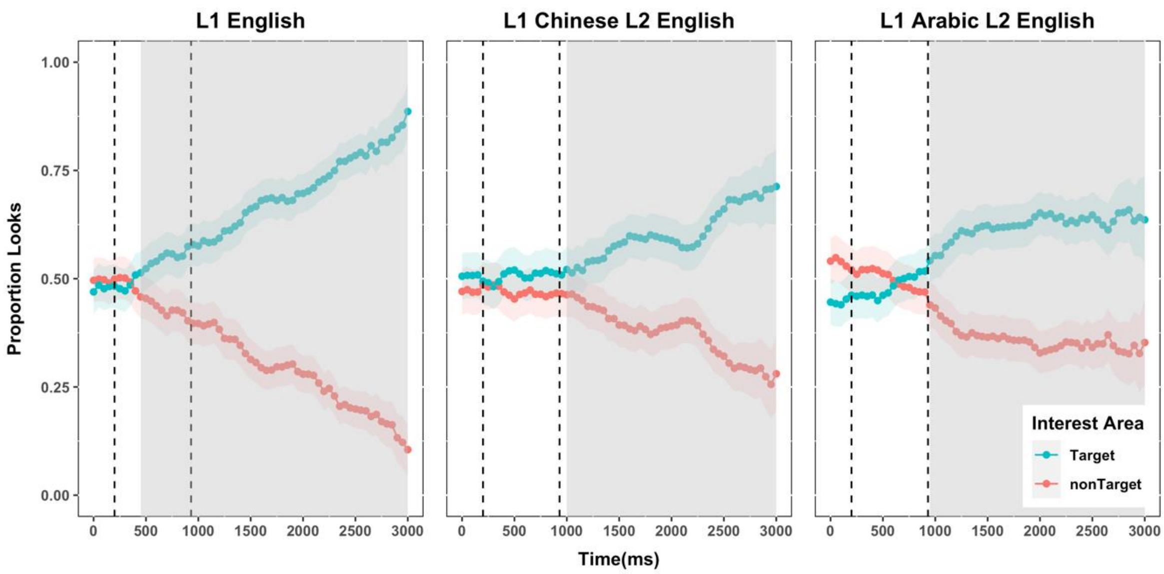
The distributions of bootstrapped differences between groups are presented in Figure 6.1.3. The mean difference in divergence points between the L1 English and L1 Chinese groups was -810 ms, CI = [-1150, -585] ms. This finding aligns with the much earlier divergence point observed in the English group. Importantly, the confidence interval does not contain zero, supporting a reliable difference between the two groups. Similarly, between the L1 English and Arabic groups, the divergence point was 274 ms earlier in the English group, with a CI of [-550, 0] ms. However, the CI of the between-group difference contains zero, thus failing to support a significant difference. The comparison of L2 groups revealed a 483 ms difference in their divergence points, with a CI of [150, 900] ms. Since the CI does not contain zero, the difference is therefore significant. The L1 Chinese group was significantly slower than the Arabic group in terms of how quickly they looked at the target area.

6.1.2 Present Perfect

Figure 6.1.4 illustrates the proportion of fixations among three groups in both target and non-target areas while processing the present perfect sentences.

Figure 6.1.4

Average Fixation Proportions of Looking at the Target and Non-Target Areas by the Groups When Hearing Present Perfect Sentences in the Visual-World Eye-Tracking Task



Note. The two vertical dashed lines represent the average verb duration (200-930 ms). Shaded areas represent significant time clusters identified through the cluster-based permutation analyses.

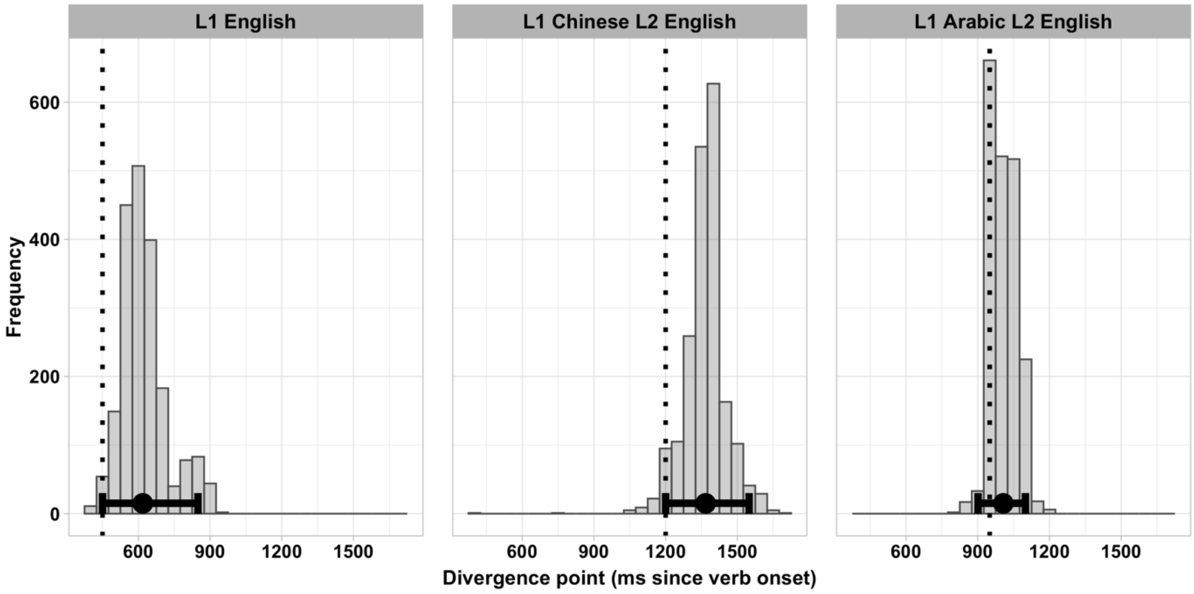
In the case of the L1 English group, they began directing their looks more towards the target area (i.e., completed event area) approximately 150 ms after the onset of the verb, exhibiting significantly more fixations on the target area compared to the other two L2 groups in general. The L1 Chinese group demonstrated similar patterns, fixating on the target area shortly after the verb onset. Although the L1 Arabic group took a slightly longer time compared to the other two groups to look towards the target area, they still displayed more fixations on the target before the verb offset.

The cluster-based permutation analyses once again revealed significant positive time clusters in all three groups: a cluster from 450 to 3000 ms (cluster mass statistic = 11982, sum $t = 586$, $p < .001$) in the L1 English group; a cluster from 1000 to 3000 ms (cluster mass statistic = 2971, sum $t = 241$, $p < .001$) in the L1 Chinese group; and a cluster from 950 to 3000 ms (cluster mass statistic = 3713, sum $t = 283$, $p < .001$) in the L1 Arabic group. These findings indicate that, at some point during the identified time windows, all three groups exhibited a significant preference for looking more towards the target over the non-target areas.

The distribution of bootstrapped divergence points in each group is presented in Figure 6.1.5. For the L1 English group, the mean divergence point was 618 ms, with a CI of [450, 850] ms. This suggests that the English group looked significantly more towards the target area even before the verb offset. In contrast, for the L1 Chinese group, the mean divergence point was much later, at 1367 ms, with a CI of [1200, 1550] ms. The L1 Arabic group exhibited a mean divergence point of 1007 ms, CI of [901, 1100] ms. The divergence points appeared slower in the L2 groups compared to the English group.

Figure 6.1.5

Bootstrap Distributions of Divergence Points of Target Over Non-Target for the Present Perfect Items in the Groups

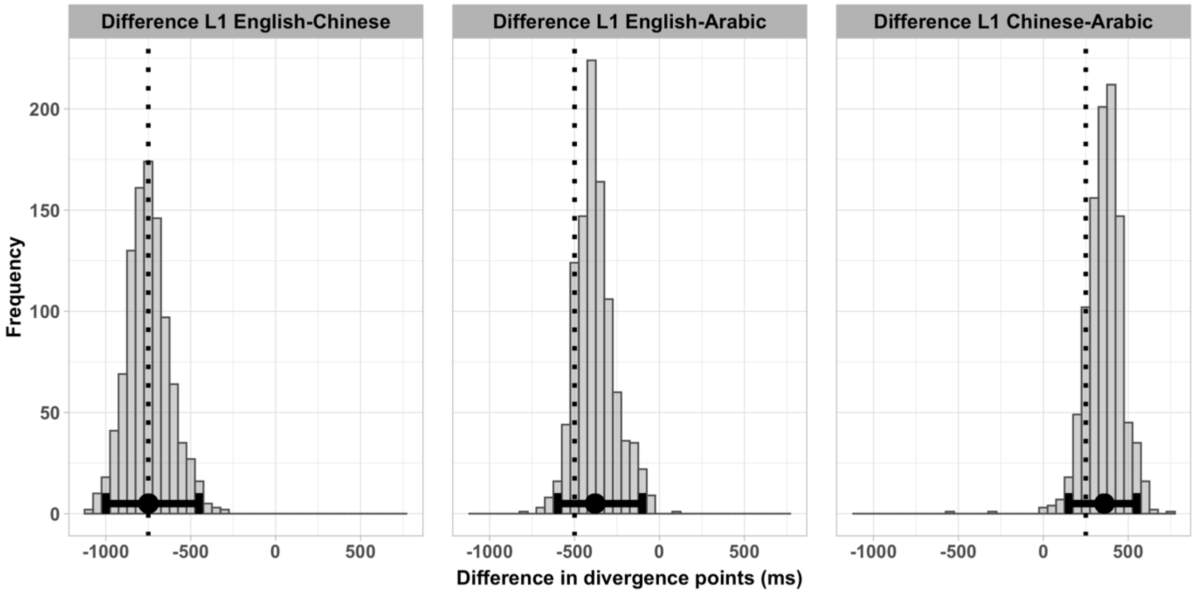


Note. The x-axis shows the distribution of divergence points based on 1000 bootstraps. The y-axis shows the number of resamples where a particular divergence point was observed. The error bars indicate the bootstrap mean and its 95% CI. Dotted vertical lines correspond to the divergence points in the original dataset.

The bootstrap distributions of divergence point differences between groups are illustrated in Figure 6.1.6. The mean difference between the L1 English and Chinese groups was -798 ms, with a CI of [-1050, -500] ms. Similarly, the mean difference between the L1 English and Arabic groups was -428 ms, with a CI of [-650, -150] ms. This suggests that the onset of the divergence point in the L1 English group occurred significantly earlier than in the L2 groups, as the confidence interval does not contain zero. The comparison between the L2 groups also revealed a significant difference. The estimated divergence point in the Chinese group occurred 309 ms significantly later than in the Arabic group, with a CI of [100, 500] ms.

Figure 6.1.6

Bootstrap Distributions of the Difference in Divergence Points Between Groups for the Present Perfect Items



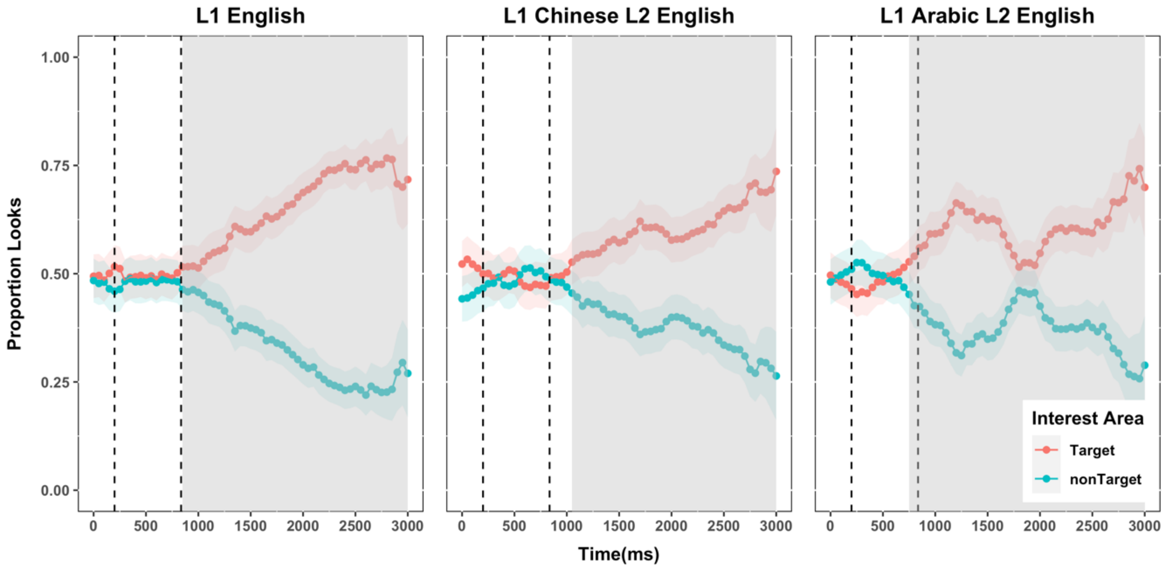
Note. Points and error bars indicate bootstrap means and 95% CI. Dotted vertical lines indicate mean divergence point differences in the original data.

6.1.3 Present Progressive

When processing the present progressive sentences, as illustrated in Figure 6.1.7, the L1 English group looked slightly more to the target area (i.e., ongoing event area) following verb onset. There was a sudden increase in their fixations to the target around the verb offset. In contrast, the L1 Chinese group initially directed slightly more fixations to the non-target area after the verb onset, gradually increasing their looks to the target after the verb offset. On the other hand, the L1 Arabic group showed a steady increase in fixations on the target area after the onset of the verb, with the diverging point occurring before the verb offset, earlier than the other two groups.

Figure 6.1.7

Average Fixation Proportions of Looking at the Interest Areas by Group When Hearing Present Progressive Sentences in the Visual-World Eye-Tracking Task



Note. The two vertical dashed lines represent the average verb duration (200-835ms). Shaded areas represent significant time clusters identified through the cluster-based permutation analyses.

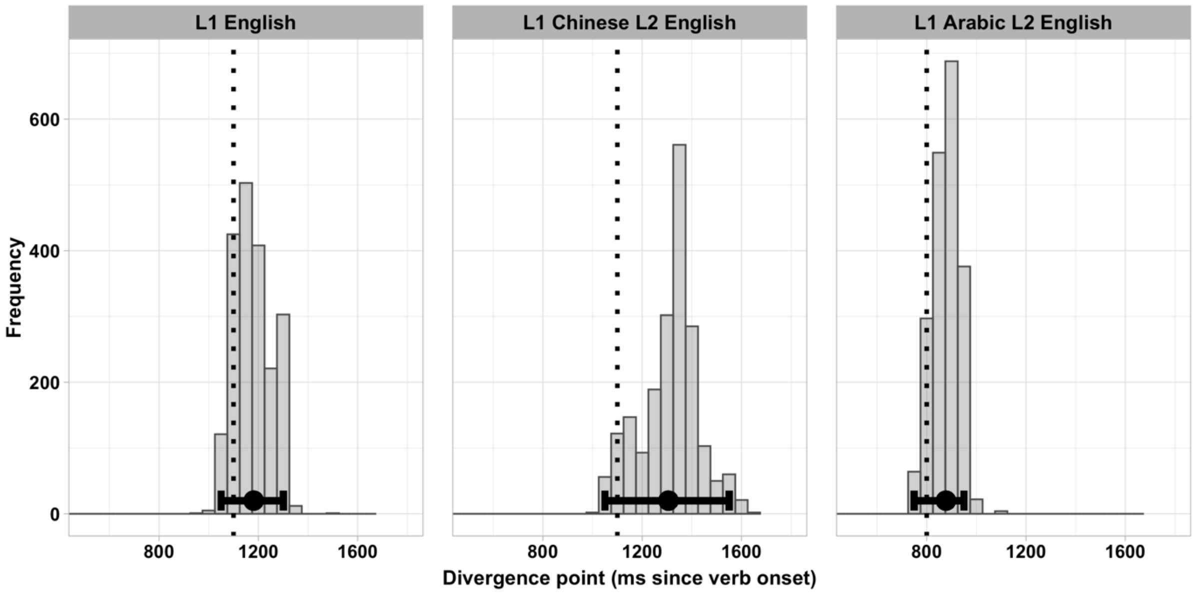
The cluster-based permutation analysis revealed a significant positive time cluster that spanned from 850 ms to 3000 ms (cluster mass statistic = 7839, sum $t = 412$, $p < .001$) in the L1 English group. Similarly, positive time clusters were also identified in the L2 group: from 1050 ms to 3000 ms (cluster mass statistic = 2782, sum $t = 236$, $p < .001$) in the L1 Chinese group, and from 750 ms to 3000 ms (cluster mass statistic = 3211, sum $t = 267$, $p < .001$) in the L1 Arabic group.

The results of the divergence point analyses are presented in Figure 6.1.8. It was found that for the L1 English group, the estimated divergence point was 1181 ms, CI= [1050, 1300]. The L1 Chinese group exhibited a slightly later divergence point at 1305 ms, with a

CI= [1050, 1550]. In contrast, the estimated divergence point in the L1 Arabic group occurred earlier than in the other two groups, at 877 ms, CI= [750, 950].

Figure 6.1.8

Bootstrap Distributions of Divergence Points of Target Over Non-Target for the Present Progressive Items in the Groups

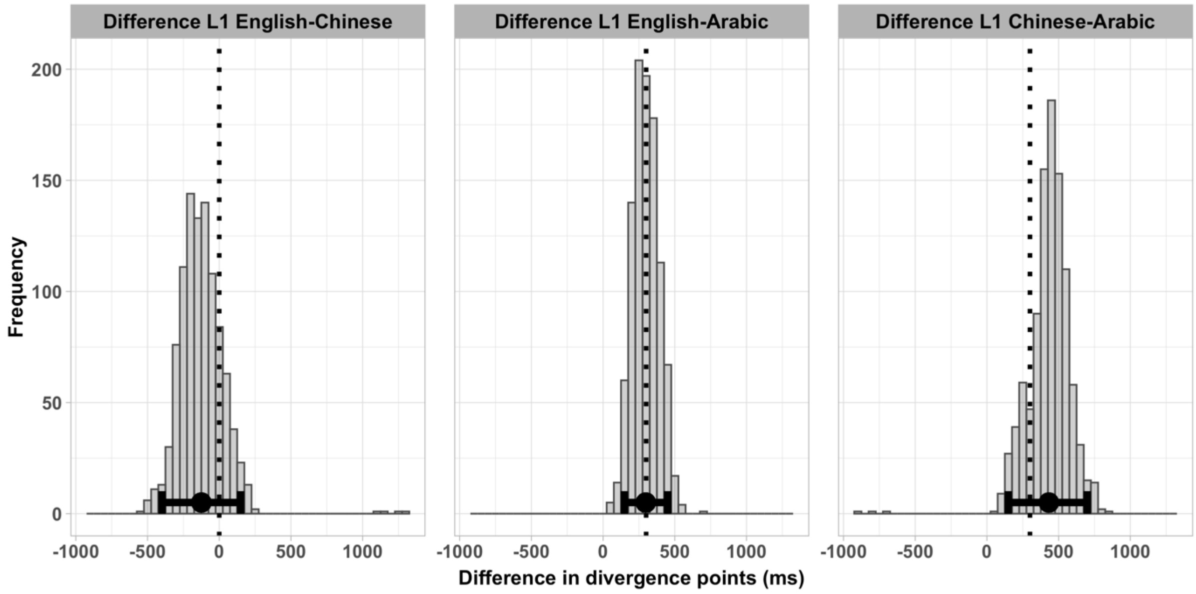


Note. The x-axis shows the distribution of divergence points based on 1000 bootstraps. The y-axis shows the number of resamples where a particular divergence point was observed. The error bars indicate the bootstrap mean and its 95% CI. Dotted vertical lines correspond to the divergence points in the original dataset.

As illustrated in Figure 6.1.9, the mean difference between the L1 English and L1 Chinese groups in their divergence points was -174 ms, CI = [-450, 100]. This suggests that the L1 English group fixated more on the target 174 ms earlier than the L1 Chinese group; however, the difference was not statistically significant, as the CI contains zero. Between the L1 English and L1 Arabic groups, the mean difference was 247 ms, CI = [100, 400]. The divergence point in the L1 English groups occurred significantly later than in the Arabic group, as the CI does not contain zero. Similarly, compared to the L1 Arabic group, compared to the L1 Arabic group, the divergence point in the L1 Chinese group was also found to occur significantly later by 381 ms, with a CI of [100, 650].

Figure 6.1.9

Bootstrap Distributions of the Difference in Divergence Points Between Groups for the Present Progressive Items



Note. Points and error bars indicate bootstrap means and 95% CI. Dotted vertical lines indicate mean divergence point differences in the original data.

To summarise the results of the VWP, all three participant groups fixated significantly more on the target area representing a completed event when they heard verbs marked for the past simple (*V-ed*) and present perfect (*have+V-ed*). Similarly, they exhibited significantly more looks to the target area depicting ongoing event when they heard verbs marked for the present progressive (V-ing). However, differences emerged in their divergence points (i.e., how quickly they started looking more to the target area). For past simple items, the L1 Chinese group demonstrated a significantly delayed divergence point compared to the L1

English and L1 Arabic groups. Conversely, the Arabic group displayed gaze patterns similar to those of the native speakers, with no significant difference between their divergence point and that of the English group. Regarding the present perfect items, both L2 groups were significantly slower in diverting their looks more to the target over the non-target compared to the L1 English group. Surprisingly, for present progressive items, the L1 Arabic group exhibited a significantly earlier divergence point than the native English group. The L1 Chinese group displayed a pattern similar to that of the native speakers, with no significant difference found in their divergence point compared to the English group.

6.2 Sentence-Matching Task Results

This section presents the results obtained from the sentence-matching (SM) task, which aimed to investigate participants' online sensitivity to tense and aspect mismatches. The assumption was that if participants were indeed sensitive, longer response times (RT) would be observed when judging ungrammatical sentence pairs compared to grammatical ones. The discussion begins with the results for the past simple items, followed by the findings for the present perfect and present progressive items.

6.2.1 Past Simple

The descriptive statistics are summarised in Table 6.2.1. It can be seen that, for all three groups, the grammatical sentence pairs were responded to slightly faster than the ungrammatical pairs. In both conditions, the L2 groups were generally slower in responding compared to the native English group. The distribution of the RTs in each group is presented in Figure 6.2.1.

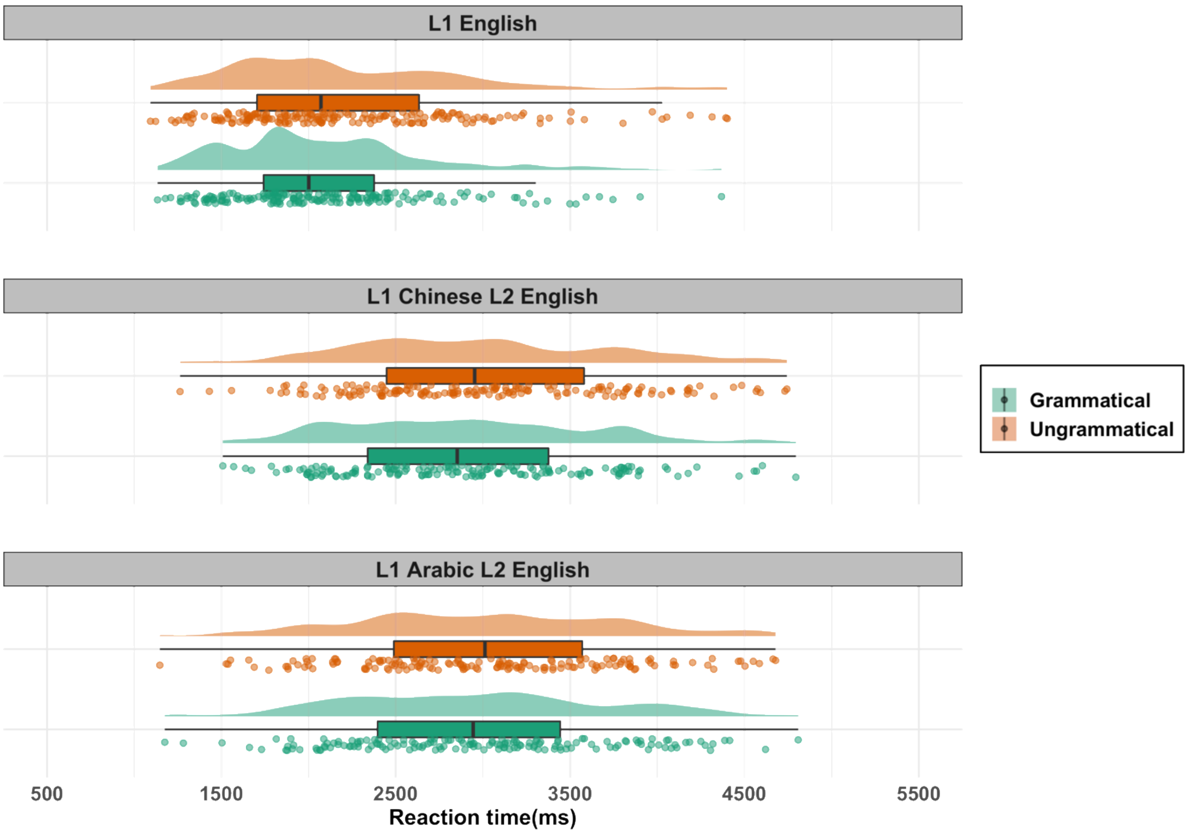
Table 6.2.1

Descriptive Statistics of RT (ms) for Past Simple Items by Group in the Sentence-Matching Task

	Condition			
	Grammatical		Ungrammatical	
Group	***M* (*SD*)**	**95% *CI***	***M* (*SD*)**	**95% *CI***
L1 English	2096 (579)	[2013, 2180]	2223 (685)	[2125, 2322]
L1 Chinese	2906 (709)	[2795, 3018]	3007 (720)	[2891, 3124]
L1 Arabic	2969 (741)	[2852, 3086]	3022 (753)	[2903, 3141]

Figure 6.2.1

The RT Distribution for Past Simple Items by Group in the Sentence-Matching Task



Note. Data are presented as half-violin plot, box plots, and raw data as scatters for each group and condition.

Table 6.2.2

Tukey-Adjusted Pairwise Comparisons of RTs Between Groups for Past Simple Items in the Sentence-Matching Task

Condition	Group	Estimate	*SE*	*t*	*p*	*d*
Grammatical	English vs. Chinese	-826.22	124.96	-6.61	< **.001**	1.4
	English vs. Arabic	-869.36	125.02	-6.95	< **.001**	**1.47**
	Chinese vs. Arabic	-43.14	129.51	-0.33	.94	.07
Ungrammatical	English vs. Chinese	-798.99	124.9	-6.4	< **.001**	**1.35**
	English vs. Arabic	-807.77	124.98	-6.46	< **.001**	**1.37**
	Chinese vs. Arabic	-8.78	129.48	-0.07	.99	.01

Note. Fitted model formula:$RT \sim group * condition + (1|participant\_id) + (1|item\_id)$). Marginal R^2 = 0.25; Conditional R^2 = 0.47. The interpretations of Cohen's *d* values for between-group comparisons using the following benchmarks: small (d = .40), medium (d = .70), and large (d = 1.00) (Plonsky & Oswald, 2014).

In terms of inferential statistics, since the maximal model failed to converge due to overfitting, the random structures in the overall model were slowly trimmed down to only include random intercepts for both participants and items. The fitted model revealed a significant main effect of group (χ^2 (2) = 48.49, p < .001) and condition (χ^2 (1) = 6.4, p = .011), and no interaction between group and condition was found (χ^2 (2) = 0.46, p = .794). Since the between-group differences were significant, the Tukey-adjusted pairwise analysis was conducted to examine more closely how the three groups differ from each other in both grammatical and ungrammatical conditions. The results are reported in Table 6.2.2. There were significant differences between the L1 English group and both of the L2 groups (*p*s <.001, *d*s > 1.0). The results were aligned with the descriptive statistics that native speakers were significantly faster than the L2 groups in responding to the sentence pairs regardless of whether they were in grammatical or ungrammatical conditions. None of the interactions reached significance, suggesting there was no difference among groups in how they responded to both grammatical and ungrammatical conditions.

Table 6.2.3

Fixed Effects Results From Within-Group Mixed-Effects Models for Past Simple Items in the Sentence-Matching Task

	Estimate (95% CI)	***SE***	***t***	***p***	***d***
(a) L1 English group					
(Intercept)	2106.73 (1947.61 – 2265.84)	81.18	25.95	< **.001**	-
Ungrammatical	127.11 (22.01 – 232.22)	53.63	2.37	**.02**	**.26**
Model formula: *RT~ condition + (1+condition|participant_id) + (1+condition|item_id);* Marginal R^2 = 0.01; Conditional R^2 = 0.44.					
(b) L1 Chinese group					
(Intercept)	2926.26 (2740.5 – 3112.02)	94.78	30.88	< **.001**	-
Ungrammatical	110.46 (-31.18 – 252.09)	72.26	1.53	.131	.19
Model formula*: RT~ condition + (1+condition|participant_id) + (1+condition|item_id);* Marginal R^2 = 0.01; Conditional R^2 = 0.35.					
(c) L1 Arabic group					
(Intercept)	2794.81 (2801.42 – 3148.21)	88.47	33.63	< **.001**	-
Ungrammatical	58.63 (-93.61 – 210.87)	77.68	0.76	.461	.09
Model formula: *RT~ condition + (1+condition|participant_id) + (1+condition|item_id);* Marginal R^2 = 0.002; Conditional R^2 = 0.24.					

Note. The interpretations of Cohen's *d* values for within-group comparisons using the following benchmarks: small (d = .60), medium (d = 1.00), and large (d = 1.40) (Plonsky & Oswald, 2014).

Since the main effect of condition was significant, three separate within-group models were constructed to determine whether the effect existed in each individual group. The results are presented in Table 6.2.3. There was a significant effect of condition only in the native English group ($\beta = 127.11$, $p = .02$, $d = .26$). Although the L2 groups both showed longer RTs in ungrammatical conditions, the differences were not statistically significant in either group (both *p*s > .05).

6.2.2 Present Perfect

For the present perfect items, as set out in Table 6.2.4, only the L1 English group responded slightly faster to grammatical pairs than ungrammatical ones. In contrast, both L2 groups displayed divergent patterns, showing slower response times for grammatical pairs as opposed to ungrammatical ones. In addition, similar to the past simple items, the L2 groups demonstrated slower response times across both conditions when compared to the L1 English group.

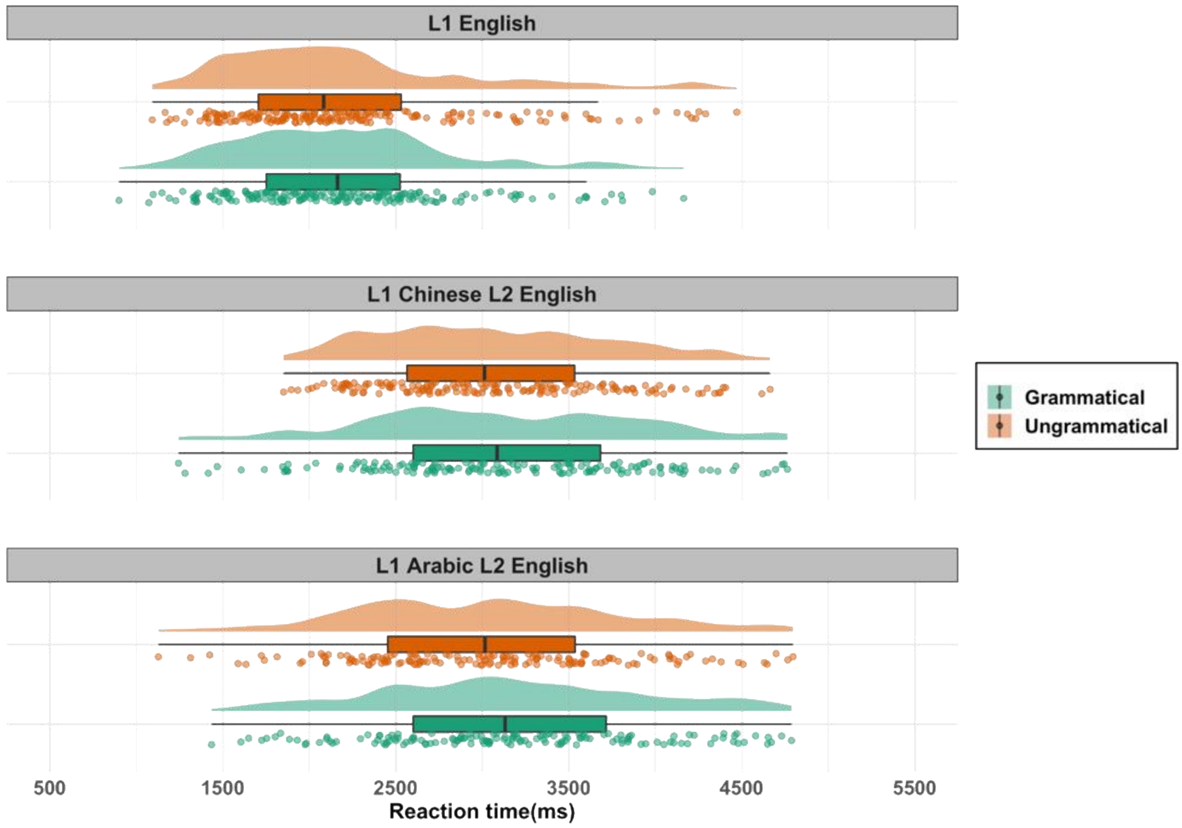
Table 6.2.4

Descriptive Statistics of RT (ms) for Present Perfect Items by Group in the Sentence-Matching Task

	Condition			
	Grammatical		**Ungrammatical**	
Group	***M* (*SD*)**	***95% CI***	***M* (*SD*)**	***95% CI***
L1 English	2198 (625)	[2107, 2288]	2235 (728)	[2130, 2341]
L1 Chinese	3119 (783)	[2994, 3244]	3072 (664)	[2968, 3176]
L1 Arabic	3166 (769)	[3041, 3291]	3032 (764)	[2909, 3155]

Figure 6.2.2

The RT Distribution for Present Perfect Items by Group in the Sentence-Matching Task



Note. Data are presented as half-violin plot, box plots, and raw data as scatters for each group and condition.

The overall mixed-effects model revealed a significant main effect of group (χ^2 (2) = 46.03, p < .001), and no main effect of condition was observed (χ^2 (1) = 1.59, p = .207), and there was no significant interaction between group and condition (χ^2 (2) = 3.91, p = .141) was found. The between-group comparisons are presented in Table 6.2.5. Similar to the past simple items, significant differences were found between the L1 English group and both L2 groups in both conditions (ps <.001, ds > 1.0). The L1 English group responded significantly quicker than the L2 groups to the items. However, since the main effect of condition was not significant, it suggests that there were no statistical differences in RTs between the grammatical and ungrammatical conditions for any of the groups. In other words, the participants' response times did not vary significantly between the two conditions, regardless of the group they belonged to.

Table 6.2.5

Tukey-Adjusted Pairwise Comparisons of RTs Between Groups for Present Perfect Items in the Sentence-Matching Task

Condition	Group	Estimate	*SE*	*t*	*p*	*d*
Grammatical	English vs. Chinese	-947.61	136.42	-6.94	< **.001**	**1.58**
	English vs. Arabic	-978.36	136.76	-7.16	< **.001**	**1.63**
	Chinese vs. Arabic	-30.74	141.92	-0.23	.97	.05
Ungrammatical	English vs. Chinese	-843	140.27	-6.02	< **.001**	**1.41**
	English vs. Arabic	-794.55	140.73	-5.65	< **.001**	**1.33**
	Chinese vs. Arabic	48.45	145.63	-0.33	.95	.08

Note. Fitted model formula: *RT ~ group * condition + (1+condition|participant_id) + (1+condition|item_id)*. Marginal R^2 = 0.26; Conditional R^2 = 0.51.

6.2.3 Present progressive

Table 6.2.6 summarises the descriptive statistics of RT (ms) for present progressive items.

Table 6.2.6

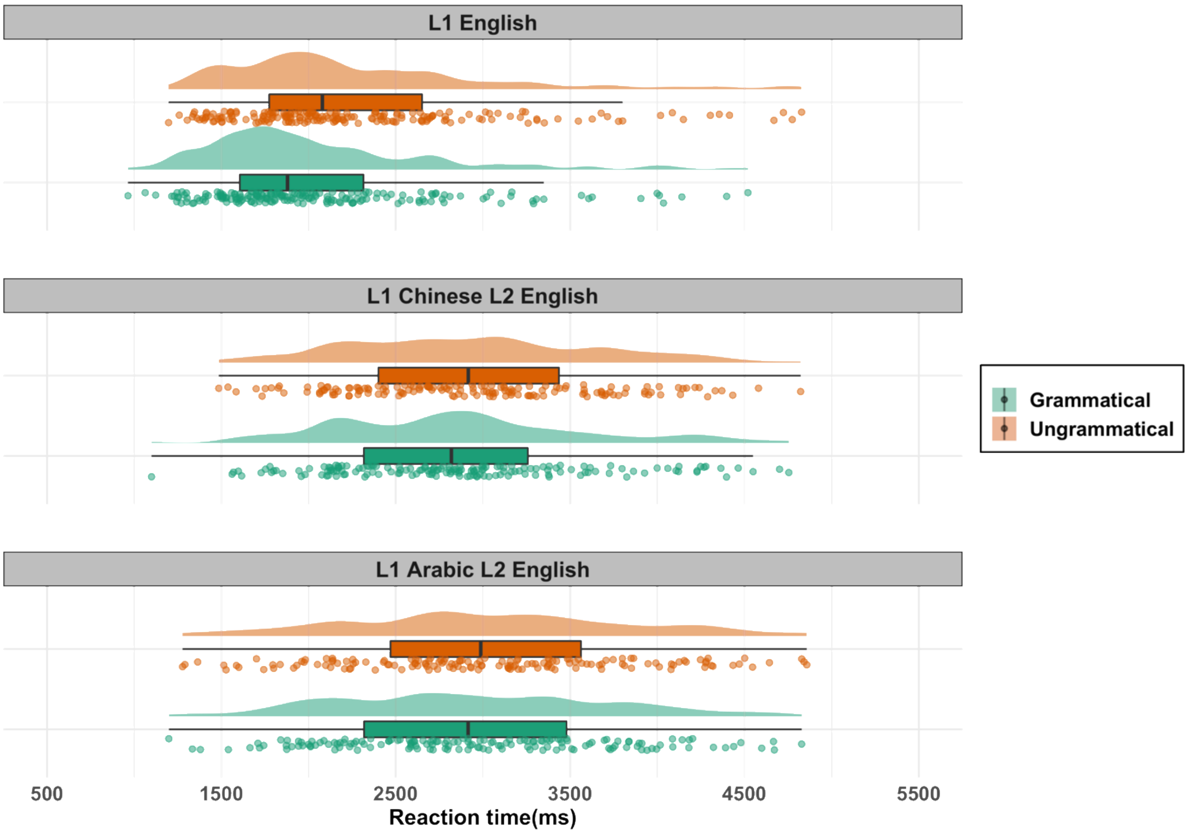
Descriptive Statistics of RT (ms) for Present Progressive Items by Group in the Sentence-Matching Task

	Condition			
	Grammatical		Ungrammatical	
Group	***M* (*SD*)**	***95% CI***	***M* (*SD*)**	***95% CI***
L1 English	2060 (674)	[1963, 2156]	2287 (764)	[2178, 2396]
L1 Chinese	2862 (728)	[2748, 2976]	2948 (719)	[2833, 3062]
L1 Arabic	2953 (786)	[2827, 3078]	3036 (819)	[2904, 3168]

All three groups responded slightly faster to the grammatical pairs compared to the ungrammatical pairs, and the RT differences were larger in the L1 English group (227 ms) versus the L2 groups (less than 100 ms). Once again, the L1 English group was generally faster compared to the two L2 groups.

Figure 6.2.3

***The** RT Distribution for Present Progressive Items by Group in the Sentence-Matching Task*



Note. Data are presented as half-violin plot, box plots, and raw data as scatters for each group and condition.

The overall mixed-effects model revealed a main effect of group (χ^2 (2) = 40.84, p < .001) and condition (χ^2 (1) = 11.35, p < .001). No significant interaction between group and condition was found (χ^2 (2) = 2.61, p = .271). Further comparisons between groups were conducted, and the results are outlined in Table 6.2.7. Once again, the L1 English group exhibited significantly quicker response times than both L2 groups in both conditions (*p*s < .001, *d*s > 1.0).

Table 6.2.7

Tukey-Adjusted Pairwise Comparisons of RTs Between Groups for Present Progressive Items in the Sentence-Matching Task

Condition	**Group**	**Estimate**	***SE***	***t***	***p***	***d***
Grammatical	English vs. Chinese	-829.84	135.88	-6.12	**< .001**	**1.30**
	English vs. Arabic	-899.31	136.21	-6.6	**< .001**	**1.41**
	Chinese vs. Arabic	-69.48	141.16	-0.49	.88	.11
Ungrammatical	English vs. Chinese	-693.13	136.25	-5.09	**< .001**	**1.09**
	English vs. Arabic	-767.57	136.41	-5.63	**< .001**	**1.20**
	Chinese vs. Arabic	-74.44	141.77	-0.53	.86	.12

Note. Fitted model formula: *RT ~ group * condition + (1 | participant_id) + (1 | item_id)* Marginal R^2 = 0.22; Conditional R^2 = 0.44.

To explore the significance of the condition within each group, three within-group models were constructed. The results (see Table 6.2.8) suggest that the effect was only significant in the English group (β = 227.43, *p* = .002, *d* = .40). Although both L2 groups displayed longer response times in the ungrammatical condition, these differences were not statistically significant in either group (both *p*s > .05).

Table 6.2.8

Fixed Effects Results from Within-Group Mixed-Effects Models for Present Progressive Items in the Sentence-Matching Task

	Estimate (95% CI)	***SE***	***t***	***p***	***d***
(a) L1 English group					
(Intercept)	2065.11 (1878.54 – 2251.68)	95.19	21.7	< **.001**	**-**
Ungrammatical	227.43 (103.10 – 351.76)	63.44	3.59	**.002**	**.40**
Model formula: RT~ condition + (1+condition|participant_id) + (1|item_id); Marginal R^2 = 0.024; Conditional R^2 = 0.39.					
(b) L1 Chinese group					
(Intercept)	2897.64 (2693.86 – 3101.42)	103.97	27.87	< **.001**	**-**
Ungrammatical	89.9 (-46.01 – 225.80)	69.34	1.3	.202	.15
Model formula: RT~ condition + (1+condition|participant_id) + (1+condition|item_id); Marginal R^2 = 0.004; Conditional R^2 = 0.34.					
(c) L1 Arabic group					
(Intercept)	2958.08 (2779.92 – 3136.23)	90.90	32.54	< **.001**	**-**
Ungrammatical	100.76 (-65.82 – 267.34)	84.99	1.19	.237	.14
Model formula: RT~ condition + (1+condition|participant_id) + (1+condition|item_id); Marginal R^2 = 0.004; Conditional R^2 = 0.17.					

In summary, the English group consistently showed longer RTs in ungrammatical conditions compared to grammatical conditions. Significant RT differences were observed for past simple and present progressive items, while no significant differences were found for present perfect items. The L2 groups exhibited RT patterns similar to the English group for past simple and present progressive items, but the RT differences were not significant in either of the L2 groups. For present perfect items, the L2 groups displayed contrasting RT patterns compared to the native English group. Surprisingly, they took less time to judge ungrammatical pairs than grammatical pairs.

6.3 Acceptability Judgement Task Results

This section presents the results obtained from the acceptability judgement task (AJT). The purpose of this task was to measure the participants' explicit knowledge of English tense-aspect, complementing the two online implicit tasks (the sentence-matching and visual-world eye-tracking tasks). The stimuli used in the task were identical to those used in the two online tasks. The participants were asked to judge the grammaticality of each sentence using a scale ranging from 1 (completely unacceptable) to 7 (completely acceptable). The assumption underlying the task was that participants with explicit knowledge of English tense-aspect would provide lower ratings to ungrammatical sentences. The discussion of the AJT results starts with the past simple items, followed by the findings for the present perfect and present progressive items.

6.3.1 Past Simple

Descriptive statistics from Table 6.3.1 reveal that all three participant groups consistently rated sentences with grammatical violations (containing mismatches between tense and aspect) as less acceptable than grammatically correct sentences. This implies that all the groups were able to distinguish between the correct and incorrect usage of the simple past construction. Moreover, in the grammatical condition, the ratings among the three groups were rather comparable. However, in the context of ungrammatical sentences, the L1 Arabic group provided slightly higher ratings (*M* = 5.28) compared to the L1 Chinese (*M* = 4.12) and English (*M* = 4.88) groups. This suggests that the L1 Arabic participants tended to find sentences with grammatical violations more acceptable compared to the other two groups.

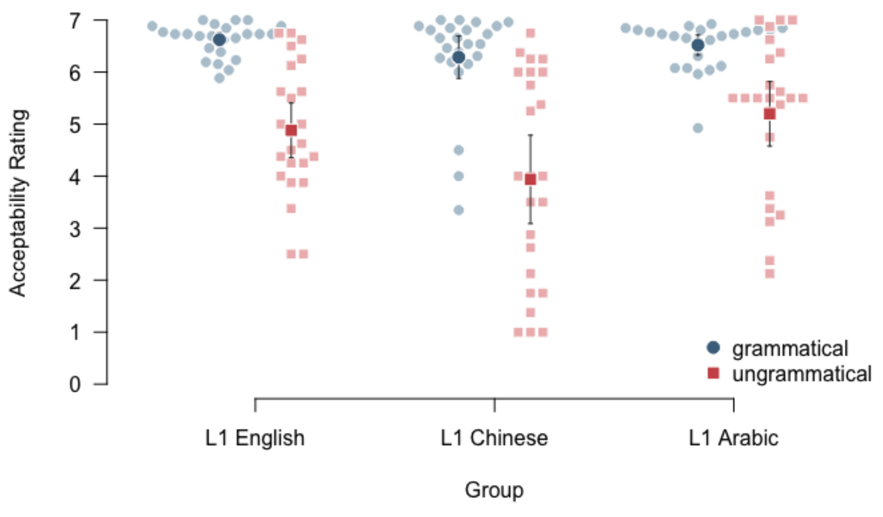
Table 6.3.1

Descriptive Statistics of Rating Scores for Past Simple Items by Group in the Acceptability Judgment Task

	Condition			
	Grammatical		Ungrammatical	
Group	***M* (*SD*)**	***95% CI***	***M* (*SD*)**	***95% CI***
L1 English	6.62 (1.0)	[6.54, 6.70]	4.88 (1.94)	[4.60, 5.16]
L1 Chinese	6.34 (1.49)	[6.21, 6.46]	4.12 (2.4)	[3.75, 4.49]
L1 Arabic	6.6 (1.15)	[6.51, 6.69]	5.28 (2.22)	[4.94, 5.62]

Figure 6.3.1

Mean Rating of Each Condition for Simple Past Items by Group (Error Bars = 95% Confidence Interval)



The overall mixed-effects model yielded significant main effects for both group (χ^2 (2) = 6.93, p = .031) and condition (χ^2 (1) = 54.08, p < .001) with no significant difference between the two (χ^2 (2) = 3.29, p = .193). The Tukey-adjusted pairwise analysis (see Table 6.3.2) revealed that the observed main effect of group might be driven by the disparities between the two L2 groups, as no significant differences were found between the English group and either of the L2 groups in both conditions. Instead, a marginally significant difference was observed between the L1 Chinese group and the L1 Arabic group in their ratings for ungrammatical sentences (β = -0.58, p = .058). The Arabic group exhibited a slightly higher acceptance of ungrammatical sentences compared to the Chinese group.

Table 6.3.2

Tukey-Adjusted Pairwise Comparisons of RTs Between Groups for Past Simple Items in the Sentence-Matching Task

Condition	Group	Estimate	*SE*	*t*	*p*	*d*
Grammatical	English vs. Chinese	0.15	0.09	1.56	.27	.24
	English vs. Arabic	0.01	0.09	0.07	.1	.01
	Chinese vs. Arabic	-0.14	0.09	-1.47	.31	.23
Ungrammatical	English vs. Chinese	0.4	0.24	1.64	.24	.67
	English vs. Arabic	-0.18	0.24	-0.76	.73	.31
	Chinese vs. Arabic	-0.58	0.25	-2.33	**.058**	**.98**

Note. Fitted model formula: *zrating ~ group * condition + (1+condition | participant_id) + (1+group+condition| item_id).* Marginal R^2 = 0.21; Conditional R^2 = 0.54.

Given the observed main effect of condition, within-group analyses were performed for each of the three groups. Results (see Table 6.3.3) indicated a significant effect of condition within each group. This suggests that all three groups rated the ungrammatical sentences as significantly more unacceptable than the grammatical sentences.

Table 6.3.3

Fixed Effects Results from Within-Group Mixed-Effects Models for Past Simple Items in the Acceptability Judgment Task

	Estimate (95% CI)	***SE***	***t***	***p***	***d***
(a) L1 English group					
(Intercept)	0.38 [0.3, 0.47]	0.05	8.53	< **.001**	-
Ungrammatical	-0.88 [-1.14, -0.62]	0.13	-6.55	< **.001**	**1.86**
Model formula: zrating ~ condition + (1 + condition| participant_id) + (1+condition| item_id). Marginal R^2 = 0.24; Conditional R^2 = 0.62.					
(b) L1 Chinese group					
(Intercept)	0.24 [0.04, 0.45]	0.11	2.3	**.03**	-
Ungrammatical	-1.14 [-1.57, -0.72]	0.22	-5.27	< **.001**	**1.71**
Model formula*: Model formula: zrating ~ condition + (1+condition|participant_id) + (1+condition| item_id).* Marginal R^2 = 0.22; Conditional R^2 = 0.58.					
(c) L1 Arabic group					
(Intercept)	0.38 [0.31, 0.45]	0.04	10.85	< **.001**	-
Ungrammatical	-0.7 [-1.04, -0.35]	0.18	-3.98	< **.001**	**1.09**
Model formula: zrating ~ condition + (1+condition|participant_id) + (1 |item_id). Marginal R^2 = 0.13; Conditional R^2 = 0.38.					

6.3.2 Present Perfect

For present perfect items (see Table 6.3.4), all three groups rated grammatical sentences higher than ungrammatical ones. This indicates that, overall, participants across all groups displayed correct knowledge of the present perfect usage. Also, the mean ratings for grammatical sentences were relatively consistent across the three groups, ranging from 6.18 to 6.38. In contrast, the mean ratings for ungrammatical sentences exhibited a wider range, with the L1 Arabic group assigning higher ratings (M = 5.27) compared to the L1 Chinese (M = 3.97) and English (M = 3.42) groups. This suggests that L1 Arabic participants were more likely to accept ungrammatical present perfect sentences compared to the other two groups.

Table 6.3.4

Descriptive Statistics of Rating Scores for Present Perfect Items by Group in the Acceptability Judgment Task

	Condition			
	Grammatical		**Ungrammatical**	
Group	***M* (*SD*)**	***95% CI***	***M* (*SD*)**	***95% CI***
L1 English	6.3 (1.43)	[6.19, 6.41]	3.42 (1.9)	[3.15, 3.69]
L1 Chinese	6.18 (1.56)	[6.05, 6.31]	3.97 (2.49)	[3.59, 4.35]
L1 Arabic	6.38 (1.4)	[6.27, 6.50]	5.27 (2.32)	[4.92, 5.63]

The overall mixed-effects model revealed significant main effects of group (χ^2 (2) = 11.21, p = .004), condition (χ^2 (1) = 61.41, p < .001), and an interaction between the two (χ^2 (2) = 13.15, p = .001). Between-group analyses indicated that the group differences arise from distinctions in the ungrammatical condition. As presented in Table 6.3.5, there were no group differences in the grammatical condition. All three groups rated the grammatical sentences similarly. However, in the ungrammatical condition, the Arabic group exhibited a significant difference from the other two groups, with their ratings for ungrammatical sentences significantly higher than those of the other two groups (ps< .05, ds > 1.0).

Figure 6.3.2

Mean AJT Rating of Each Condition for Present Perfect Items by Group (Error Bars = 95% Confidence Interval)

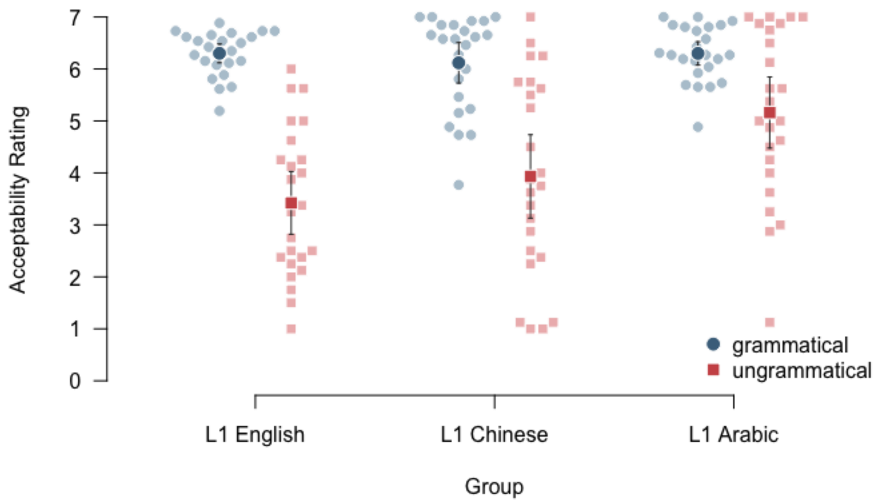


Table 6.3.5

Tukey-Adjusted Pairwise Comparisons of RTs Between Groups for Present Perfect Items in the Sentence-Matching Task

Condition	Group	Estimate	*SE*	*t*	*p*	*d*
Grammatical	English vs. Chinese	0.06	0.09	0.61	.81	.09
	English vs. Arabic	-0.04	0.09	-0.44	.90	.06
	Chinese vs. Arabic	-0.10	0.1	-1.01	.57	.15
Ungrammatical	English vs. Chinese	-0.28	0.26	-1.09	.52	.42
	English vs. Arabic	-0.96	0.26	-3.69	**.001**	**1.42**
	Chinese vs. Arabic	-0.67	0.27	-2.51	**.038**	**1.00**

Note. Fitted model formula: *zrating ~ group * condition + (1+condition | participant_id) + (1+condition| item_id).* Marginal R^2 = 0.24; Conditional R^2 = 0.55.

The within-group analyses (Table 6.3.6) revealed a significant effect of condition in all three groups (all *p*s < .05, *d*s > 2.0). This shows that in each group, participants consistently rated the ungrammatical sentences significantly lower than the grammatical sentences.

Table 6.3.6

Fixed Effects Results from Within-Group Mixed-Effects Models for Present Perfect Items in the Acceptability Judgment Task

	Estimate (95% CI)	*SE*	*t*	*p*	*d*
(a) L1 English group					
(Intercept)	0.2 [0.08, 0.31]	0.06	3.26	**.002**	-
Ungrammatical	-1.44 [-1.72, -1.16]	0.14	-10.17	< **.001**	**2.38**
Model formula: zrating ~ condition + (1 + condition| participant_id) + (1+condition| item_id). Marginal R^2 = 0.36; Conditional R^2 = 0.64.					
(b) L1 Chinese group					
(Intercept)	0.15 [-0.06, 0.36]	0.11	1.4	**0.173**	-
Ungrammatical	-1.12 [-1.58, -0.65]	0.24	-4.72	< **.001**	**2.17**
Model formula: zrating ~ condition + (1+condition|participant_id)+(1+condition| item_id). Marginal R^2 = 0.2; Conditional R^2 = 0.61.					
(c) L1 Arabic group					
(Intercept)	0.25 [0.14, 0.36]	0.06	4.54	< **.001**	-
Ungrammatical	-0.56 [-0.89, -0.23]	0.17	-3.33	**.003**	**2.02**
Model formula: zrating~condition + (1+condition|participant_id) + (1+condition|item_id). Marginal R^2 = 0.07; Conditional R^2 = 0.37.					

6.3.3 Present Progressive

Table 6.3.7 presents a summary of the descriptive statistics for the present progressive items.

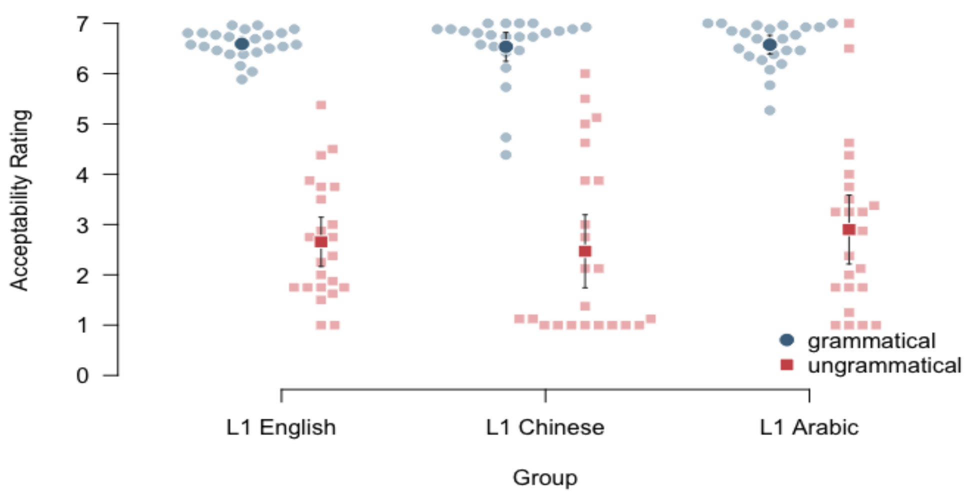
Table 6.3.7

Descriptive Statistics of Rating Scores for Present Progressive Items by Group in the Acceptability Judgment Task

	Condition			
	Grammatical		Ungrammatical	
Group	***M* (*SD*)**	***95% CI***	***M* (*SD*)**	***95% CI***
L1 English	6.59 (1.03)	[6.51, 6.67]	2.66 (1.61)	[2.43, 2.89]
L1 Chinese	6.61 (1.03)	[6.53, 6.70]	2.46 (2.13)	[2.13, 2.78]
L1 Arabic	6.64 (1.07)	[6.55, 6.73]	3.02 (2.41)	[2.65, 3.38]

Figure 6.3.3

Mean AJT Rating of Each Condition for Present Progressive Items by Group (Error Bars = 95% Confidence Interval)



Similar to the findings for past simple and present perfect items, all three groups consistently rated ungrammatical present progressive sentences lower than their grammatical counterparts. The mean ratings for grammatical sentences were relatively consistent across the three groups, ranging from 6.59 to 6.64. While the mean ratings for ungrammatical sentences exhibited a slightly broader range (2.46-3.02), this range was considerably smaller compared to the variability observed for past simple and present perfect items.

The overall mixed-effects model revealed no main effect of group (χ^2 (2) = 1.48, p = .478), and there was no significant interaction between group and condition (χ^2 (2) = 1.02, p = .6). This suggests that the three groups performed similarly in both grammatical and ungrammatical conditions. The model only indicated a significant main effect of condition (χ^2 (1) = 126.18, $p < .001$). To further explore this effect in each group and determine the size of the effect, within-group analyses were conducted. As summarised in Table 6.3.8, all three groups rated the ungrammatical sentences as significantly more unacceptable than the grammatical sentences (all ps < .001, ds > 3.0), with large effect sizes obtained.

In summary, the group differences in the AJT task were relatively smaller compared to those in the online comprehension tasks. All three groups consistently rated ungrammatical sentences as significantly less acceptable than grammatical ones. This indicates that they all displayed explicit knowledge of English past simple, present perfect, and present progressive usages. One interesting result emerged in the Arabic participants' ratings for past simple and present perfect sentences. It was observed that their ratings for ungrammatical sentences were higher than those of the other two groups, suggesting a greater likelihood of accepting ungrammatical past simple (e.g., **For the last two weeks, Mike went to the birthday party of his best friend*) and present perfect constructions (e.g., **Six months ago, my two brothers have studied Italian at a school in Rome*).

Table 6.3.8

Fixed Effects Results from Within-Group Mixed-Effects Models for Present Progressive Items in the Acceptability Judgment Task

	Estimate (95% CI)	***SE***	***t***	***p***	***d***
(a) L1 English group					
(Intercept)	0.37 [0.28, 0.45]	0.04	8.29	< **.001**	-
Ungrammatical	-2.02 [-2.29, -1.76]	0.14	-14.88	< **.001**	**4.56**
Model formula: zrating ~ condition + (1 + condition| participant_id) + (1+condition| item_id). Marginal R^2 = 0.65; Conditional R^2 = 0.82.					
(b) L1 Chinese group					
(Intercept)	0.39 [0.25, 0.52]	0.07	5.7	< **.001**	-
Ungrammatical	-2.14 [-2.58, -1.71]	0.22	-9.65	< **.001**	**4.29**
Model formula: zrating ~ condition + (1+condition|participant_id)+(1 | item_id). Marginal R^2 = 0.62; Conditional R^2 = 0.81.					
(c) L1 Arabic group					
(Intercept)	0.4 [0.32, 0.48]	0.04	10.01	< **.001**	-
Ungrammatical	-1.86 [-2.27, -1.46]	0.21	-9.06	< **.001**	**3.06**
Model formula: zrating~condition + (1+condition|participant_id) + (1+condition|item_id). Marginal R^2 = 0.51; Conditional R^2 = 0.7.					

6.4 Elicited Imitation Task Results

This section presents the results of the elicited imitation (EI) task, which aimed to measure participants' online oral production of English tense-aspect morphology. Since it was performed under time pressure, it was assumed to tap into participants' implicit knowledge. The analysis focused on the changes made by the participants during imitation, as it is believed that when learners repeat sentences and make changes to them, these modifications can reflect their internalisation of grammatical knowledge. If they automatically correct the ungrammatical sentences, it shows that they have the knowledge of relevant grammatical structures. It was found that participants demonstrated a high frequency of modifications in their replies, implying an inability to maintain the presented sentences as a single chunk in working memory. This suggests that they need to reconstruct the meaning themselves based on their grammatical knowledge. The results will still be presented separately for past simple items, followed by present perfect and present progressive items.

6.4.1 Past Simple

Table 6.4.1 demonstrates the percentage of changes made by the participants when repeating the past simple sentences in grammatical and ungrammatical conditions. The results indicate that the L1 English group corrected the majority of the ungrammatical sentences (73.2%). In contrast, the L2 groups performed more poorly, with the Arabic group correcting only 33.3%, and the Chinese group correcting just 25%. Regarding the preference for the change type, the L1 English and Arabic groups exhibited similar patterns, more frequently normalizing ungrammatical sentences than introducing errors to grammatical ones. However, the L1 Chinese group demonstrated an opposite trend, changing more grammatical sentences to their ungrammatical counterparts than correcting the ungrammatical sentences. The differences are visualised in Figure 6.4.1.

Table 6.4.1.

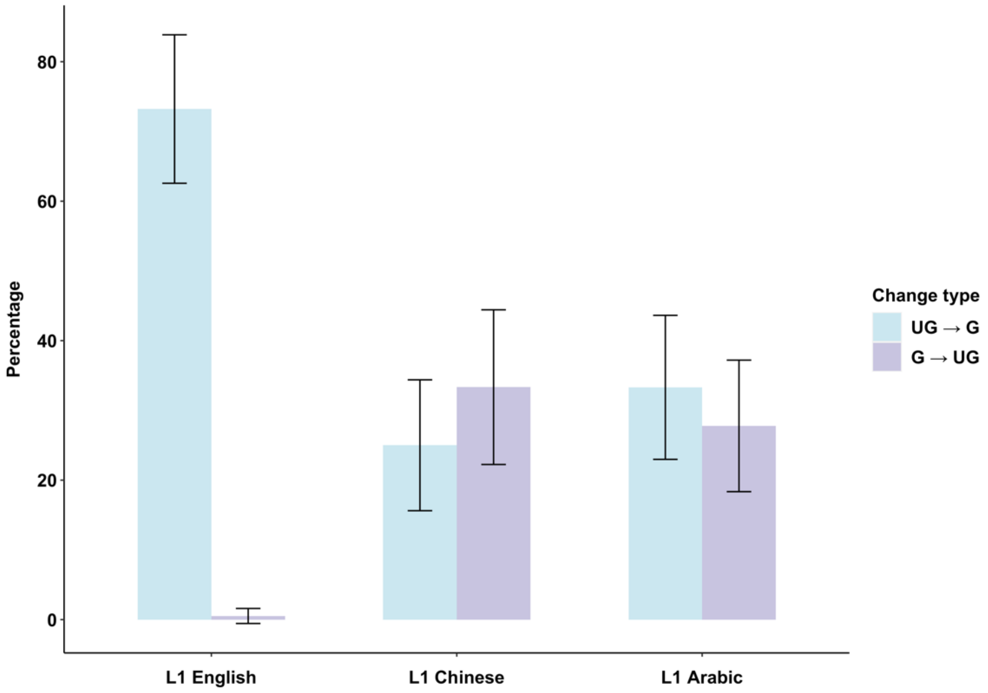
Mean Percentage of Changes Made to the Grammatical and Ungrammatical Past Simple Sentences by Group in the Elicited Imitation Task

		Change type	
Group		**UG → G**	**G → UG**
L1 English	***M* (*SD*)**	73.2% (25.2)	0.5% (2.6)
	95% *CI*	[62.6, 83.9]	[-0.6, 1.6]
L1 Chinese	***M* (*SD*)**	25% (20.6)	33.3% (24.4)
	95% *CI*	[15.6, 34.4]	[22.2, 44.4]
L1 Arabic	***M* (*SD*)**	33.3% (22.7)	27.8% (20.7)
	95% *CI*	[23.0, 43.6]	[18.4, 37.2]

Note. Change type: UG → G (ungrammatical to grammatical); G → UG (grammatical to ungrammatical).

Figure 6.4.1

Percentage of Changed Imitations When Repeating Past Simple Items by Group in the Elicited Imitation Task (Error Bars = 95% CI)



Note. Change type: UG → G (ungrammatical to grammatical); G → UG (grammatical to ungrammatical).

The mixed-effects logistic regression model revealed significant main effects for both group (χ^2 (2) = 7.03, p =.03) and change type (χ^2 (1) = 49.23, p < .001). The interaction between group and change type was also statistically significant (χ^2 (2) = 83.03, p < .001), suggesting significant differences among the groups in terms of how much they changed items in both the grammatical and ungrammatical conditions. As presented in Table 6.4.2, the effect of group was due to differences between the English group and both L2 groups. Specifically, when repeating ungrammatical sentences, the English group exhibited a significantly higher likelihood of correcting these sentences during repetitions compared to both L2 groups (*p*s < .001). Likewise, when repeating grammatical sentences, the English group was less likely to introduce errors during repetitions compared to both L2 groups (*p*s < .001).

Table 6.4.2

Between-Group Comparisons for Past Simple Items in the Elicited Imitation Task

Change type	**Group**	**Estimate**	***SE***	***z***	***p***
UG → G	English vs. Chinese	2.47	0.39	6.33	**< .001**
	English vs. Arabic	1.98	0.38	5.20	**< .001**
	Chinese vs. Arabic	-0.49	0.39	-1.27	.41
G → UG	English vs. Chinese	-4.72	1.06	-4.47	**< .001**
	English vs. Arabic	-4.40	1.06	-4.16	**< .001**
	Chinese vs. Arabic	0.32	0.35	0.90	.64

Note. Fitted model formula: *score ~ group * change_type + (1 + change_type| participant_id) + (1 | item_id).* Marginal R^2 = 0.53; Conditional R^2 = 0.62.

Table 6.4.3

Fixed Effects Results from Within-Group Mixed-Effects Models for Past Simple Items in the Elicited Imitation Task

	Estimate (95% CI)	*SE*	*z*	*p*
(a) L1 English group				
(Intercept)	-9.36 [-17.24, -1.47]	4.02	-2.33	.02
UGtoG	10.61 [2.74, 18.48]	4.02	2.64	**.008**
Model formula: score ~ change_type + (1 + change_type| participant_id) + (1 | item_id). Marginal R^2 = 0.64; Conditional R^2 = 0.93.				
(b) L1 Chinese group				
(Intercept)	-0.83 [-1.33, -0.32]	0.26	-3.18	.001
Ungrammatical	-0.41 [-1.12, 0.31]	0.37	-1.11	.267
Model formula: score ~ change_type + (1 + change_type| participant_id) + (1 | item_id). Marginal R^2 = 0.01; Conditional R^2 = 0.18.				
(c) L1 Arabic group				
(Intercept)	-1.12 [-1.58, -0.67]	0.23	-4.83	< .001
Ungrammatical	0.38 [-0.13, 0.89]	0.26	1.46	.145
Model formula: score ~ change_type + (1 + change_type| participant_id) + (1 | item_id). Marginal R^2 = 0.01; Conditional R^2 = 0.15.				

The within-group analyses were conducted to investigate the preferences for the change type in all three groups. The results (Table 6.4.3) revealed that only the L1 English group exhibited a significant preference for correcting ungrammatical sentences rather than adding errors to grammatical sentences (β = 10.61, p = .008, d = 8.26). The Arabic group also showed a similar preference to the English group (β = 0.38), but this tendency did not reach statistical significance (p = .145). In contrast, the Chinese group demonstrated an opposite preference, although no significant difference was observed.

As discussed earlier, both L2 groups exhibited a tendency to introduce errors when repeating grammatical sentences (see Table 6.4.4). Among the L1 Chinese participants, 33.3% of the grammatical sentences were changed into ungrammatical ones during their repetitions, while the L1 Arabic group changed 27.8%. A qualitative analysis of the errors introduced by each group was conducted, and the findings are summarised in Table 6.4.4. The Chinese group, in particular, frequently used bare verbs in contexts where past morphology was necessary (41%). Examples of such errors included imitations such as "**Three hours ago, my friend prepare food and drinks for our movie night*". On the other hand, the Arabic group tended to introduce errors by using present perfect forms in past simple obligatory contexts (46%). For instance, they produced imitations such as "**An hour ago, the student has talked about some issues with her tutor*".

Table 6.4.4

Errors Produced in Past Simple Contexts by L2 Groups in the Elicited Imitation Task

Error type	**L1 Chinese L2 English**	**L1 Arabic L2 English**
	***N* (%)**	***N* (%)**
Omit inflections	**29 (41%)**	11 (24%)
Omit temporal adverbials	5 (7%)	9 (20%)
Use present perfect forms	19 (27%)	**21 (46%)**
Other errors	17 (25%)	5 (11%)
Total	70 (100%)	46 (100%)

Notes. Other errors include other forms such as present progressive and present simple tenses.

6.4.2 Present Perfect

Table 6.4.5 summarises the percentage of changes introduced to the present perfect sentences by the three participant groups. Consistent with the results for the past simple items, the L1 English group corrected the majority of the ungrammatical items (around 80%) to their grammatical counterparts, whereas the L2 groups only corrected approximately 31%. In terms of change type preference, the L1 English group showed a clear tendency to correct ungrammatical sentences over adding errors to grammatical sentences. In contrast, both L2 groups displayed opposite patterns, with more errors added to grammatical sentences than normalising ungrammatical sentences, as illustrated in Figure 6.4.2.

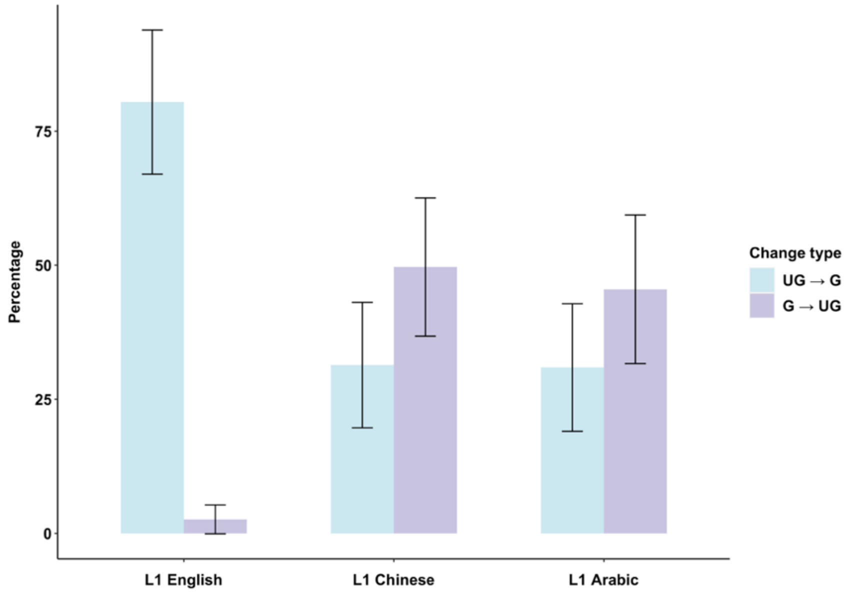
Table 6.4.5

Mean Percentage of Changes Made to the Grammatical and Ungrammatical Present Perfect Sentences by Group in the Elicited Imitation Task

		Change type	
Group		**UG → G**	**G → UG**
L1 English	***M* (*SD*)**	80.4% (31.8)	2.6% (6.4)
	95% *CI*	[67.0, 93.9]	[-0.1, 5.29]
L1 Chinese	***M* (*SD*)**	31.4% (25.7)	49.7% (28.3)
	95% *CI*	[19.7, 43.1]	[36.8, 62.5]
L1 Arabic	***M* (*SD*)**	30.9% (26.1)	45.5% (30.5)
	95% *CI*	[19.0, 42.8]	[31.7, 59.4]

Figure 6.4.2

Percentage of Changed Imitations When Repeating Present Perfect Items by Group in the Elicited Imitation Task (Error Bars = 95% CI)



Note. Change type: UG → G (ungrammatical to grammatical); G → UG (grammatical to ungrammatical).

The mixed-effects logistic regression results once again revealed a significant effect of change type (χ^2 (2) = 24.07, $p < .001$), and there was a significant interaction between group and change type (χ^2 (2) = 90.18, $p < .001$), suggesting that the three groups differed from each other in terms of the changes made in the grammatical and ungrammatical conditions. As outlined in Table 6.4.6, significant differences were observed between the L1 English group and L2 groups for both change types (*p*s < .001). This implies that the English group performed significantly better compared to the L2 groups, and they normalised more ungrammatical sentences and left more grammatical sentences unchanged.

Table 6.4.6

Between-Group Comparisons for Present Perfect Items in the Elicited Imitation Task

Change type	Group	Estimate	*SE*	*z*	*p*
UG → G	English vs. Chinese	3.47	0.62	5.56	**< .001**
	English vs. Arabic	3.57	0.63	5.67	**< .001**
	Chinese vs. Arabic	0.10	0.61	0.16	.99
G → UG	English vs. Chinese	-4.38	0.73	-5.60	**< .001**
	English vs. Arabic	-4.15	0.73	-5.66	**< .001**
	Chinese vs. Arabic	0.23	0.48	0.49	.88

Note. Fitted model formula: *score ~ group * change_type + (1 + change_type| participant_id) + (1+group| item_id).* Marginal R^2 = 0.44; Conditional R^2 = 0.68.

The within-group analyses (see Table 6.4.7) revealed that in the L1 English group, changes were significantly more frequent from ungrammatical to grammatical direction than from grammatical to ungrammatical direction (β = 8.58, p < .001). In both L2 groups, the effect of change type also reached statistical significance (*p*s < .05) but in the opposite direction, as indicated by the negative β *values*. This suggests that both L2 groups exhibited a clear tendency to add errors to grammatical sentences rather than correcting the ungrammatical sentences.

When repeating grammatical sentences, both L2 groups demonstrated a high tendency to change them into ungrammatical versions: 49.7% in the L1 Chinese group, and 45.5% in the L1 Arabic group. The errors introduced by the two groups are summarised in Table 6.4.8. For the L1 Chinese group, a prevailing issue was using present forms, accounting for 39% of their errors. In addition, there were instances of using alternative past simple forms in present perfect contexts, constituting 25% of the errors. In the case of the L1 Arabic group, over half of their produced errors (55%) involved past simple forms. Furthermore, there were occurrences where they opted not to produce temporal adverbials, reflecting a tendency to avoid using them (21%).

Table 6.4.7

Fixed Effects Results from Within-Group Mixed-Effects Models for Present Perfect Items in the Elicited Imitation Task

	Estimate (95% CI)	***SE***	***z***	***p***
(a) L1 English group				
(Intercept)	-4.72 [-6.23, -3.20]	0.77	-6.11	< .001
UGtoG	8.58 [6.36, 10.79]	1.13	7.58	**< .001**
Model formula: score ~ change_type + (1 + change_type| participant_id) + (1 | item_id). Marginal R^2 = 0.61; Conditional R^2 = 0.89.				
(b) L1 Chinese group				
(Intercept)	-0.03 [-0.66, 0.60]	0.32	-0.09	.928
UGtoG	-1.01 [-1.78, -0.23]	0.4	-2.54	**.011**
Model formula: score ~ change_type + (1 + change_type| participant_id) + (1 | item_id). Marginal R^2 = 0.05; Conditional R^2 = 0.36.				
(c) L1 Arabic group				
(Intercept)	-0.27 [-0.93, 0.40]	0.34	-0.79	< .001
UGtoG	-0.74 [-1.30, -0.19]	0.28	-2.62	**.009**
Model formula: score ~ change_type + (1 + change_type| participant_id) + (1 | item_id). Marginal R^2 = 0.03; Conditional R^2 = 0.34.				

Table 6.4.8

Errors Produced in Present Perfect Contexts by L2 Groups in the Elicited Imitation Task

Error type	**L1 Chinese**	**L1 Arabic**
	***N* (%)**	***N* (%)**
Omit temporal adverbials	12 (11%)	**16 (21%)**
Use past simple forms	**27 (25%)**	**42 (55%)**
Use present forms	**42 (39%)**	11 (14%)
Omit verbal inflections	14 (13%)	5 (6%)
Other errors	13 (12%)	3 (4%)
Total	108 (100%)	77 (100%)

Note. Other errors include using other forms such as past progressive and past perfect tenses.

6.4.3 Present Progressive

Table 6.4.9 presents the frequency of changes made to the present progressive items by the three groups.

Table 6.4.9

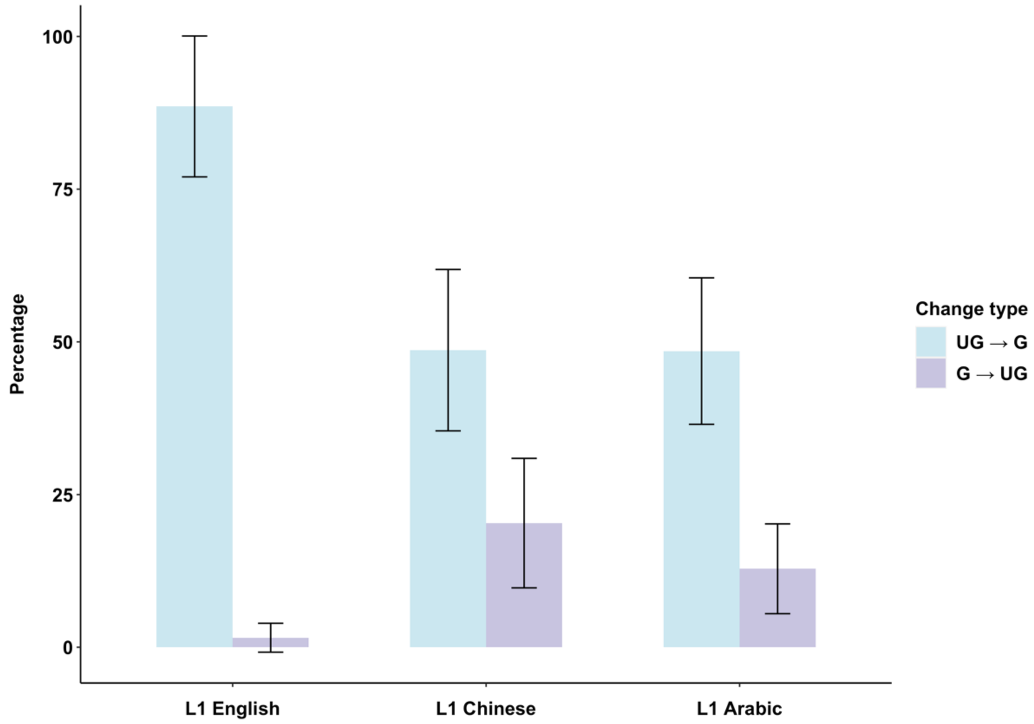
Mean Percentage of Changes Made to the Grammatical and Ungrammatical Present Progressive Sentences by Group in the Elicited Imitation Task

Group		**Change type**	
		UG → G	**G → UG**
L1 English	***M* (*SD*)**	88.5% (27.3)	1.6% (5.6)
	95% *CI*	[77.0, 100]	[-0.8, 3.9]
L1 Chinese	***M* (*SD*)**	48.6% (29)	20.3% (23.3)
	95% *CI*	[35.4, 61.9]	[9.7, 30.9]
L1 Arabic	***M* (*SD*)**	48.5% (26.4)	12.8% (16.1)
	95% *CI*	[36.5, 60.5]	[5.5, 20.2]

Once again, the L1 English group corrected a majority of the ungrammatical sentences (88.5%). Although the L2 groups only corrected around 48% of the ungrammatical sentences, their performance was far better compared to their results for the past simple and present perfect items. Regarding the change type tendency (see Figure 6.4.3), the L2 groups, similar to the L1 English group, displayed a higher frequency of correcting the ungrammatical sentences than introducing errors into the grammatical sentences.

Figure 6.4.3

Percentage of Changed Imitations When Repeating Present Progressive Items by Group in the Elicited Imitation Task (Error Bars = 95% CI)



The mixed-effects logistic regression model found a significant main effect of change type (χ^2 (1) = 75.69, p < .001), as well as an interaction between group and change type (χ^2 (2) = 43.57, p < .001). Consistent with the observed patterns in other structures, the English group performed significantly better than the L2 groups when repeating both ungrammatical and grammatical sentences (*p*s < .005) (see Table 6.4.10). The L2 groups displayed a tendency to correct fewer ungrammatical sentences and introduce more errors into grammatical sentences compared to the English group.

Table 6.4.10

Between-Group Comparisons for Present Progressive Items in the Elicited Imitation Task

Change type	**Group**	**Estimate**	***SE***	***z***	***p***
UG → G	English vs. Chinese	3.59	0.70	5.17	< **.001**
	English vs. Arabic	3.55	0.69	5.20	< **.001**
	Chinese vs. Arabic	-0.04	0.59	-0.07	.10
G → UG	English vs. Chinese	-3.11	0.95	-3.28	**.003**
	English vs. Arabic	-2.51	0.96	-2.62	**.02**
	Chinese vs. Arabic	0.60	0.58	1.04	.55

Note. Fitted model formula: *score ~ group * change_type + (1 + change_type| participant_id) + (1+group| item_id).* Marginal R^2 = 0.56; Conditional R^2 = 0.76.

While the L2 groups performed poorly compared to the English group, the within-group analyses (see Table 6.4.11) revealed that they still demonstrated a significant tendency to correct ungrammatical sentences instead of introducing errors into grammatical sentences (*p*s < .001). A similar pattern was also observed in the English group.

As presented in Table 6.4.12, despite both L2 groups demonstrating an ability to correct more ungrammatical sentences, they still introduced more errors—20.3% in the Chinese group and 12.8% in the Arabic group—when repeating grammatical sentences, compared to the native speaker group. Within the L1 Chinese group, a common error was the production of incorrect forms (53%) in present progressive contexts, such as present simple and past tenses. Also, similar to their errors in other constructions, they frequently omitted verb inflections (28%). In contrast, the L1 Arabic group exhibited a lower error rate, with 36% of their errors involving the omission of temporal adverbials.

Table 6.4.11

Fixed Effects Results from Within-Group Mixed-Effects Models for Present Progressive Items in the Elicited Imitation Task

	Estimate (95% CI)	***SE***	***z***	***p***
(a) L1 English group				
(Intercept)	-8.47 [-14.82, -2.12]	3.24	-2.61	.009
UGtoG	17.86 [7.72, 28]	5.17	3.45	**< .001**
Model formula: score ~ change_type + (1 + change_type| participant_id) + (1 | item_id). Marginal R^2 = 0.54; Conditional R^2 = 0.98.				
(b) L1 Chinese group				
(Intercept)	-1.94 [-2.73, -1.14]	0.4	-4.78	< .001
UGtoG	1.85 [0.82, 2.88]	0.52	3.53	**< .001**
Model formula: score ~ change_type + (1 + change_type| participant_id) + (1 | item_id). Marginal R^2 = 0.14; Conditional R^2 = 0.46.				
(c) L1 Arabic group				
(Intercept)	-2.28 [-2.94, -1.62]	0.34	-6.78	< .001
UGtoG	2.18 [1.28, 3.08]	0.46	4.74	**< .001**
Model formula: score ~ change_type + (1 + change_type| participant_id) + (1 | item_id). Marginal R^2 = 0.23; Conditional R^2 = 0.38.				

Table 6.4.12

Errors Produced in Present Progressive Contexts by L2 Groups in the Elicited Imitation Task

Error type	L1 Chinese	L1 Arabic
	***N* (%)**	***N* (%)**
Omit inflections	**13 (28%)**	3 (14%)
Omit temporal adverbials	4 (9%)	**8 (36%)**
Omit auxiliaries	5 (11%)	4 (18%)
Other errors	**25 (53%)**	7 (32%)
Total	47 (100%)	22 (100%)

Note. Other errors include using other forms such as present simple and past tenses.

In summary, the L2 groups generally performed significantly worse than the native English group in the imitation task. They corrected fewer ungrammatical sentences and introduced more errors into grammatical sentences compared to the English group. Preferences for change types also differed among the groups. For past simple items, both L1 English and L1 Arabic groups corrected more ungrammatical sentences (UG→G) than adding errors to grammatical sentences (G→UG), though this preference was only statistically significant in the English group. Conversely, the L1 Chinese group displayed a contrasting pattern, favouring more G→UG than UG→G. In the case of present perfect items, both L2 groups exhibited patterns opposite to the native English group, introducing more errors into grammatical sentences than normalizing ungrammatical sentences. These non-target-like preferences reached statistical significance in both L2 groups. However, in the imitation of present progressive sentences, both L2 groups showed a significant tendency to correct ungrammatical sentences, aligning with the preferences observed in the English group. Lastly, the qualitative analyses of L2 participants' errors in repeating grammatical sentences revealed specific patterns. Chinese participants frequently omitted verb inflections, while Arabic participants tended to use present perfect forms in past simple contexts and past simple forms in present perfect contexts.

6.5 Gap-Filling Task Results

This section discusses the findings of the gap-filling task, which was designed to assess the participants’ explicit knowledge of English temporal/aspectual morphology or forms in written production. The objective of this task was to complement the online production data obtained from the elicited imitation task, and as such, the stimuli presented to the participants in both tasks were identical. Given the proficiency level of the L2 participants and their years of formal English instruction, it was predicted that target-like performance would be observed in both L2 groups. The presentation of the results begins with the past simple items, followed by the findings for the present perfect and present progressive items.

6.5.1 Past Simple

As demonstrated in Table 6.5.1.1, each of the three groups achieved comparable scores, all surpassing 95%, in the context of the past simple. This implies that all groups were able to use accurate forms of the English past simple within the respective contexts.

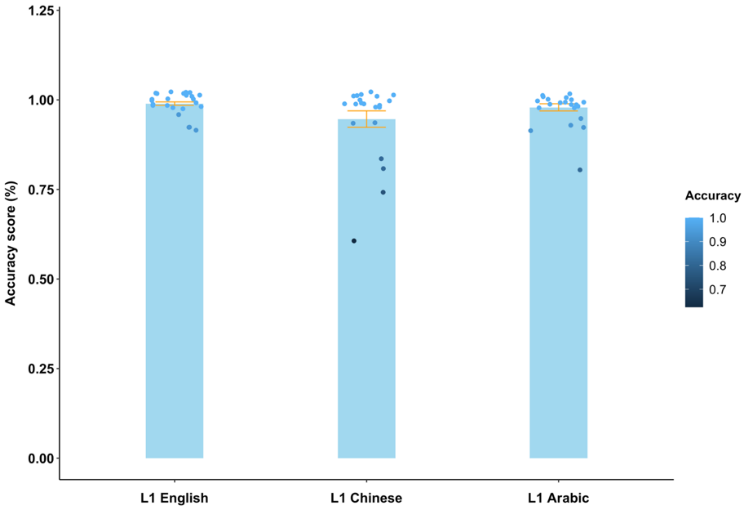
Table 6.5.1

Descriptive Statistics of Accuracy Scores for Past Simple Items by Group in the Gap-Filling Task

	Accuracy scores	
Group	***M* (*SD*)**	***95% CI***
L1 English	0.99 (0.02)	[0.98, 1.00]
L1 Chinese	0.95 (0.11)	[0.90, 0.99]
L1 Arabic	0.98 (0.05)	[0.96, 1.00]

Figure 6.5.1

Accuracy Scores (Percentage of Correctness) for Past Simple Items by Group in the Gap-Filling Task



Note. Each bar represents the mean, and the error bars correspond to the standard error. The jittered points represent individual data points.

Table 6.5.2

Fixed Effects Results from a Mixed-Effects Logistic Regression Model Fitted to the Accuracy Scores for Past Simple Items in the Gap-Filling Task

Fixed effects	**Estimate *[95% CI]***	***SE***	***z***	***p***
(Intercept)	6.06 [4.21, 7.91]	0.94	6.42	**< .001**
L1 Chinese	-1.49 [-3.27, 0.29]	0.91	-1.64	.101
L1 Arabic	-0.64 [-2.52, 1.24]	0.96	-0.67	.504

Note. Fitted model formula: *score ~ group + (1|participant_id) + (1 | item_id).* Marginal R^2 = 0.05; Conditional R^2 = 0.58.

The results from the mixed-effects logistic regression model revealed a non-significant effect of group (χ^2 (2) = 2.55, p = .279). The fixed effects results (see Table 6.5.2) revealed no significant differences between the L1 English group and either of the L2 groups (both *p*s >.05). Both L2 groups demonstrated target-like performance in their usage of past simple morphology.

6.5.2 Present Perfect

Compared to the results obtained in the past simple context, there was greater variability among the three groups in the present perfect context (see Table 6.5.3). The L1 English group attained the highest mean score of 85%, with the L1 Chinese group following closely at 75%. Conversely, the L1 Arabic group achieved the lowest score at only 45%. The relatively larger 95% CI within the groups may also suggest variability within each group.

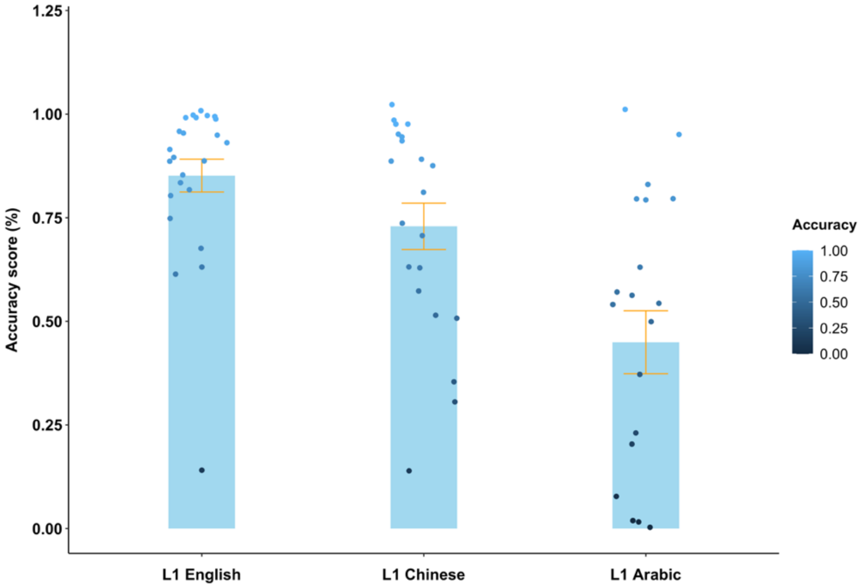
Table 6.5.3

Descriptive Statistics of Accuracy Scores for Present Perfect Items by Group in the Gap-Filling Task

	Accuracy scores	
Group	**Mean (*SD*)**	**95% *CI***
L1 English	0.85 (0.19)	[0.77, 0.93]
L1 Chinese	0.73 (0.26)	[0.61, 0.85]
L1 Arabic	0.45 (0.35)	[0.29, 0.61]

Figure 6.5.2

Accuracy Scores (Percentage of Correctness) for Present Perfect Items by Group in the Gap-Filling Task



Note. Each bar represents the mean, and the error bars correspond to the standard error. The jittered points represent individual data points.

Table 6.5.4

Between-Group Comparisons for Present Perfect Items in the Gap-Filling Task

Group	Estimate	*SE*	*z*	*p*
English vs. Chinese	0.75	0.69	1.10	**.52**
English vs. Arabic	3.15	0.67	4.67	**< .001**
Chinese vs. Arabic	2.40	0.67	3.60	**.001**

Note. Fitted model formula: *score ~ group + (1|participant_id) + (1 | item_id).* Marginal R^2 = 0.19; Conditional R^2 = 0.65.

The mixed-effects logistic regression analysis revealed a significant effect of group (χ^2 (2) = 22.98, $p < .001$). The between-group comparisons (see Table 6.5.4) indicated a significant difference between the English and Arabic groups (β = 3.15, $p < .001$). Also, there was a significant difference between the Chinese and Arabic groups (β = 2.40, p = .001). The Arabic group performed significantly worse than the other two groups in the present perfect context.

To better examine the observed differences, it was crucial to analyse the incorrect forms provided by the participants. Table 6.5.5 presents the number and percentage of incorrect form usage in the context of the present perfect by the three groups. The table indicates a tendency among all groups to use past simple forms in present perfect contexts. Specifically, the Arabic participants demonstrated a much higher frequency of past simple forms in present perfect contexts (n = 138) compared to the L1 English (n = 47) and Chinese (n = 53) groups. This suggests that Arabic learners encountered greater difficulty in accurately using present perfect forms compared to the other two groups.

Table 6.5.5

Incorrect Forms Used in the Present Perfect Contexts by Group in the Gap-Filling Task

	Past simple	**Past progressive**	**Other forms**	**Total**
Group	***N* (%)**	***N* (%)**	***N* (%)**	***N* (%)**
L1 English	47 (82.5%)	4 (7%)	6 (10.5%)	57 (100%)
L1 Chinese	53 (58.2%)	22 (24.2%)	16 (17.6%)	91 (100%)
L1 Arabic	138 (74.6%)	24 (13%)	23 (12.4%)	185 (100%)

Note. Other forms include using past perfect, present simple, present progressive tenses.

6.5.3 Present Progressive

In the present progressive contexts, all three groups achieved high scores (above 94%). In addition, there was less variability within each group when compared to the present perfect contexts. This implies that all the participants could use the present progressive forms correctly.

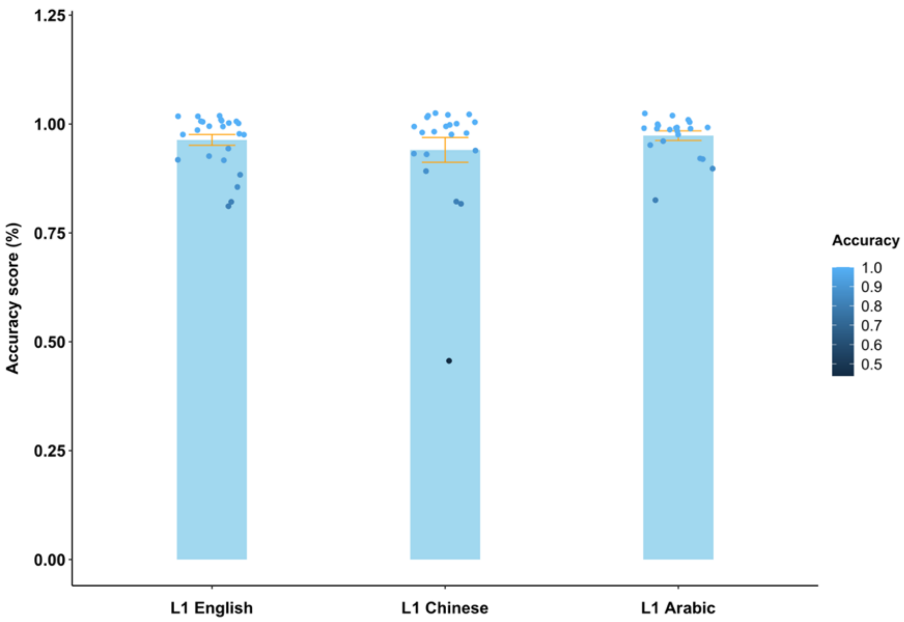
Table 6.5.6

Descriptive Statistics of Accuracy Scores for Present Progressive Items by Group in the Gap-Filling Task

	Accuracy scores	
Group	***M* (*SD*)**	**95% CI**
L1 English	0.96 (0.06)	[0.94, 0.99]
L1 Chinese	0.94 (0.13)	[0.88, 1.00]
L1 Arabic	0.97 (0.05)	[0.95, 1.00]

Figure 6.5.3

Accuracy Scores (Percentage of Correctness) for Present Progressive Items by Group in the Gap-Filling Task



Note. Each bar represents the mean, and the error bars correspond to the standard error. The jittered points represent individual data points.

The results from the mixed-effects logistic regression model aligned with the descriptive findings, and the effect of the group did not reach statistical significance (χ^2 (2) = 4.3, p = .116). As indicated in Table 6.5.7, there were no significant differences between the L1 English group and either of the L2 groups (both ps > .05). These findings suggest that both L2 groups exhibited native-like performance in the use of present progressive morphology in production.

To summarise the main findings of the gap-filling task, all three groups performed almost at ceiling in supplying correct verb forms for the past simple and present progressive contexts, showing no group differences. However, in the present perfect context, differences emerged, as the L1 Arabic group performed significantly worse than the other two groups. An examination of their errors revealed a tendency to supply past simple forms in situations that required obligatory present perfect usage.

Table 6.5.7

Fixed Effects Results from a Mixed-Effects Logistic Regression Model Fitted to the Accuracy Scores for Present Progressive Items in the Gap-Filling Task

Fixed effects	**Estimate *[95% CI]***	***SE***	***z***	***p***
(Intercept)	6.2 [2.80, 9.60]	1.74	3.57	< **.001**
L1 Chinese	-1.57 [-4.68, 1.56]	1.59	-0.98	**.326**
L1 Arabic	2.99 [-3.17, 9.16]	3.15	0.95	**.342**

Note. Fitted model formula: *score ~ group + (1|participant_id) + (1+group | item_id).* Marginal R^2 = 0.18; Conditional R^2 = 0.83.

6.6 Summary of Main Results

Table 6.6.1 presents a summary of the main results obtained from the different online and offline tasks.

Table 6.6.1

A Summary of Main Results of the Study

	Main Results
Online Measures:	
Visual-world task	1) All participant groups fixated significantly more on the target event areas during the selected time periods. 2) The Chinese group showed significantly delayed divergence compared to the English and Arabic groups for past simple items. 3) Both L2 groups were significantly slower to diverge to the target compared to the English group for present perfect items. 4) Both L2 groups showed native-like divergence patterns for present progressive items.
Sentence-matching task	1) The English group consistently showed longer RTs in ungrammatical conditions compared to grammatical ones. 2) The L2 groups had similar RT patterns to the English group for past simple and present progressive items, but the RT differences were not significant in the L2 groups. 3) Both L2 groups displayed contrasting RT patterns compared to the native English group for present perfect items.
Elicited imitation task	1) The English group consistently showed a preference for correcting ungrammatical sentences (UG→G) over adding errors to grammatical sentences (G→UG). 2) The Chinese group preferred more G→UG changes over UG→G changes compared to the English and Arabic groups for past simple items. 3) Both L2 groups showed a preference for more G→UG changes over UG→G changes for present perfect items.

	Main Results
	4) Both L2 groups significantly corrected more ungrammatical sentences for present progressive items, in line with the patterns observed in the English group.
Offline Measures:	
AJT task	1) All groups rated ungrammatical sentences significantly less acceptable than grammatical ones. 2) The Arabic group rated ungrammatical past simple and present perfect sentences higher than the other two groups.
Gap-filling task	1) All three groups performed nearly perfectly in supplying correct verb forms for past simple and present progressive contexts, with no significant group differences observed. 2) In the present perfect context, the Arabic group performed significantly worse compared to the other two groups.

Chapter 7 Discussion

This study explores the acquisition of L2 tense-aspect by adult learners of English whose L1 backgrounds are Chinese and Arabic. The research aims to contribute to the ongoing debate concerning the extent of L1 influence and learnability issues in adult L2 acquisition within the generative framework. For data collection, each L1 group (English, Chinese, and Arabic) initially consisted of 24 participants. However, in order to make meaningful comparisons between the performance of the Chinese and Arabic groups, three participants from each group were excluded from the analysis based on their English proficiency levels. Comprehension and production knowledge of learners were assessed through a combination of online and offline tasks. This chapter discusses the obtained results (see Chapter 6) in response to each of the research questions and the relevant predictions.

7.1. RQ1: How do Chinese and Arabic Learners of English Process English Tense-Aspect Markings in Comparison to Native English Speakers in the Visual-World Eye-Tracking Task?

<u>Predictions:</u>

a) *The L1 English group will be able to process morphological markings in past simple (-ed), present perfect (have+V-ed), and present progressive (be+V-ing) and use them to encode event information immediately*.

b) *The L1 Chinese group is expected to show target-like processing of be+V-ing in present progressive sentences but will display slower processing of past simple (-ed) and present perfect (have+V-ed) constructions*.

c) *The L1 Arabic group will demonstrate target-like processing of past simple (-ed) and present progressive (be+V-ing) sentences but will exhibit slower processing of present perfect (have+V-ed) sentences*.

The predictions were entirely realised. The results indicate that all three groups fixated significantly more on the target area during the selected time window, suggesting that they could use morphological markings to encode event information and achieve target interpretations. Specifically, similar to native speakers, the L2 participants could construct mental representations of completed events when exposed to past simple and present perfect morphemes. Likewise, they could construct mental representations of ongoing or progressive events when hearing the present progressive morpheme. Subsequently, they directed their gaze towards the visual scene with a temporal structure compatible with their relevant mental representations. In general, this adds to the evidence that event information associated with verbs plays a significant role in both L1 and L2 sentence processing (Madden & Zwaan, 2003; Minor et al., 2022; Zhou et al., 2014). Previous research has suggested that, like L1 speakers, L2 learners can use grammatical features, such as grammatical gender, to facilitate their comprehension of sentences (Grüter et al., 2012; Hopp, 2013). The present study results show that L2 learners can also use temporal/aspectual markings to enhance their sentence comprehension.

On the other hand, some differences were observed in the onset times of divergence points among the groups. Firstly, for present progressive sentences, no between-group differences were found between the onset times of divergence points for the English and Chinese groups. This finding aligns with the results of studies conducted by Chan (2012) and Yao and Chen (2017), wherein the Chinese learners performed in a native-like way, demonstrating sensitivity to errors in present progressive sentences during the SPR and eye-tracking reading tasks. The results of the present studies, coupled with their findings, suggest that Chinese learners not only can detect errors in ungrammatical present progressive sentences but also can utilise morphological cues in a manner consistent with native speakers during online processing. More surprisingly, the Arabic group outperformed both the English and Chinese groups, displaying a significantly quicker divergence point. This result contributes to one of the few rare examples (e.g., Lee & Phillips, 2023; Wu & Juffs, 2022) where L2 learners surpass native speakers in processing by extending the evidence to the morphological level. While it is difficult to provide explanations, one possible hypothesis could be that Arabic learners have a more complex morphological system in their L1, potentially providing them with a processing advantage (Z. Wu & Juffs, 2022).

Nevertheless, taken together, both L2 groups performed in a native-like way in their fixation patterns and processing speeds. This shows that they could compute the meaning of a spoken sentence and encode an event using morphological cues as rapidly and efficiently as native speakers. The results add to the growing body of evidence showing native-like morphosyntactic processing among adult L2 learners (e.g., Alemán Bañón et al., 2014; Dowens et al., 2010; Hopp, 2006; Jegerski, 2015; Keating, 2009; Tokowicz & MacWhinney, 2005; Tokowicz & Warren, 2010; Trenkic et al., 2014; Yuan, 2017). The findings challenge the notion that L2 learners tend to rely more on lexical and pragmatic factors, underusing grammatical information during sentence processing, as captured in the Shallow Structure Hypothesis (Clahsen & Felser, 2006, 2018). In particular, the results contradict previous findings, which suggested that L2 speakers are generally slower in processing and less sensitive to morphological information compared to native speakers (e.g., Clahsen, Felser, et al., 2010; Lew-Williams & Fernald, 2007).

As predicted, non-native-like patterns were observed in the processing of past simple items. Specifically, the L1 Chinese group exhibited a significantly slower onset time of divergence points compared to their L1 Arabic counterparts and the native group. This suggests that the Chinese participants took much longer to process the morphological information *-ed* encoded in the verb, even though they ultimately displayed a correct preference for the target visual area. In other words, their ability to use the morphological cues to encode tense information facilitatively was compromised compared to the other two groups. In contrast, the Arabic group displayed a native-like processing speed, with no difference from the English group in onset times. The findings are consistent with Kahoul's (2014) study, where advanced Arabic learners outperformed advanced Chinese learners in the speed of showing their first looks to the target picture. However, a difference is also observed. While in Kahoul's study, the Chinese group did not show more looks to the target over competitor pictures, in the present study, the Chinese group exhibited a target preference, as suggested by the cluster analysis. Nevertheless, they were significantly slower than the other two groups in the onset times of divergence points. The present study validates these findings using more robust statistical analyses instead of simply counting the total length of time looking at the target and calculating means of the time of the first look to the target, as was performed in Kahoul's study.

At the same time, the findings contradict those of Yao and Chen (2017), who found that highly proficient Chinese learners exhibited online sensitivity to past tense violations in both SPR and eye-tracking reading tasks. The contrasting results could be attributed to differences in the experimental paradigm. In their research, Yao and Chen employed a violation paradigm, using an uninflected form within a past simple context (e.g., **last night, my friend walk home after the closure of the ceremony*). As Trenkic et al. (2014) argued, detecting errors in grammatically incorrect sentences can be very different from utilising grammatical information facilitatively in processing fully grammatical sentences. For example, Foote (2011) observed that near-native English learners of Spanish exhibited online sensitivity to gender agreement violations in the SPR task, while Grüter et al. (2012) found that the same learner population failed to use grammatical gender cues in a native-like manner in the VWP task. In the current study, the Chinese learners indeed demonstrated the ability to employ morphological information in real-time processing. However, the issue lay in their inability to achieve a native-like pattern, in contrast to their Arabic counterparts.

In alignment with the prediction, both L2 groups exhibited significantly slower processing of present perfect sentences, displaying non-native-like patterns. Their divergence points emerged much later compared to the English group, who looked significantly more to the target area approximately 310ms *before* the verb offset. This suggests that the Chinese and Arabic participants were unable to utilise the morphological information associated with the verb *have+V-ed* as rapidly and efficiently as the native speakers. Another interesting result relates to the significant differences observed between the L2 groups. The mean divergence points detected were 70ms and 430ms *after* the verb offset in the Arabic and Chinese groups, respectively, with the Chinese group proving to be significantly slower than their Arabic counterparts. It is suspected that these observed differences may be attributed to the nature of the materials used and different aspects of L1 influences in the L2 groups (see detailed discussion below).

Considering the combined findings, it appears that native-like processing patterns and speeds are attainable when the forms are constructed similarly between learners' L1 and L2. This aligns with some previous studies focusing on the influence of cross-linguistic similarities (e.g., Foucart & Frenck-Mestre, 2011; Sabourin & Stowe, 2008; Tokowicz & MacWhinney, 2005; Tolentino & Tokowicz, 2011). As discussed in Section 3.2, similar to English, Chinese conveys progressive meaning grammatically through the aspect marker *ZAI* (Lin, 2006; Liu, 2015). This marker is considered a pre-verbal morpheme and is used to present an event an event in the midst of its occurrence. On the other hand, Arabic does not grammaticize progressive aspect; however, there is evidence suggesting that the language is currently undergoing a process of grammaticalization through the inflected forms of *gaaʕid* (Ouali, 2022). This phenomenon is observed in various Arabic dialects, including both Saudi and Egyptian Arabic. Whenever this form is employed, it enables the expression of progressive meaning. Therefore, the similarities between L1 and L2 helped both L2 groups in processing and utilising English progressive forms automatically during online comprehension in a native-like manner.

In contrast, non-native-like processing is more likely to be observed when the L2 grammatical forms are not present or differ in learners' L1, consistent with some previous processing research findings (e.g., Chan, 2012; L. Chen et al., 2007; Jiang et al., 2011; Lew-Williams & Fernald, 2010; Roberts & Liszka, 2013; Wang et al., 2022). For instance, with past simple items, it was observed that the Chinese group exhibited significantly slower processing speed compared to the other two groups. Since the L2 groups were matched in terms of their English proficiency, the processing pattern observed in the Chinese group is likely to be caused by the influence of their L1. Chinese does not grammatically express tense, whereas Arabic, similar to English, marks tense explicitly. These cross-linguistic differences contribute to the different online behaviours observed in the learner groups. The absence of grammatical tense in their L1 has resulted in Chinese learners being less efficient than their Arabic counterparts in utilising past tense marking *-ed* during online comprehension.

In the case of present perfect items, the results can also be explained by considering learners' L1 grammar. Both Chinese and Arabic lack a similar marking system for present perfect as seen in English. Recall that the English present perfect form (*have+v-ed/en*) conveys both [+perfect] and [+past] meanings. While Chinese employs a perfect marker (i.e., sentence-final "LE"), it lacks a grammatical tense to fully instantiate the present perfect, often relying on adverbs (e.g., "already") to convey the intended meaning. Similarly, Arabic does not grammatically express the perfect aspect and lacks a specific marker to explicitly indicate current relevance. However, it does instantiate [+past] meaning grammatically. In Arabic, adverbs are frequently used to establish current relevance and convey the present perfect meaning. Thus, the absence of grammatical forms for the present perfect in L1 may have led to slower processing patterns for both Chinese and Arabic learners when utilising the present perfect form in their L2 English. In addition, the comparison between the L2 groups reveals that the Arabic group still demonstrated a quicker divergence point compared to the Chinese group.

Figure 7.1.1

Examples of Target Pictures for Present Perfect Items Used in the Eye-Tracking Task

(1) The young man has chopped the tree down. (Left)
(2) The boy and girl have planted a tree together. (Right)

This difference could be attributed to the nature of the tested materials. The visual-world paradigm mostly focused on the resultative perfect constructions (e.g., "I have broken my leg") since they could be more easily depicted through pictures, unlike experiential perfect (e.g., "I have read that book") and perfect of persistent situation (e.g., "I have lived here for 10 years"). The resultative perfect describes a past event with a lasting outcome in the present (Comrie, 1976). In images, only the end-result stage was portrayed (see Figure

7.1.1). The actions of chopping and planting were completed in the past. Due to L1 influences, the Arabic group may have been able effectively to encode the [+past] meaning associated with present perfect forms, while the Chinese group might not be as sensitive to such meaning. This likely contributed to the observed differences between the two groups.

The combined findings provide some evidence for the Competition Model (MacWhinney, 2005), which argues that learners rely on cues from their L1 in comprehending and processing an L2. Since adult learners' L1 is deeply ingrained, the model predicts that they will initially process L2 grammatical features resembling their L1 more easily but encounter challenges with features differ from their L1, leading to errors when relying on L1 cues in such cases (Tokowicz & Warren, 2010). That is to say that positive transfer is likely to occur in situations where two languages are similar, whereas negative transfer might take place when two languages differ, arising from competition between L1 and L2 cues (Van Hell & Tokowicz, 2010). For features absent in learners' L1, the Competition Model predicts that learners will experience no competition and are likely to process the features easily if there is sufficient input exposure (e.g., Foucart & Frenck-Mestre, 2011; Tokowicz & MacWhinney, 2005; Trenkic et al., 2014). However, the present results contradict this prediction, as learners in the current study still struggled to process features that are absent in their L1s and unique to their L2 (i.e., past tense marking for the Chinese group). This is consistent with other studies suggesting that L2 learners face difficulties in acquiring features that are not present in their L1 (e.g., L. Chen et al., 2007; Jiang, 2004, 2007).

Essentially, the findings from the eye-tracking task offer supporting evidence for the Representational Deficit proposals, i.e., the RDH and MCH, both of which argue that L1 plays a predominant role in L2 ultimate attainment. The absence of certain grammatical properties in the L1 can pose significant challenges for L2 learners, and this learning process may be influenced by maturational constraints or critical periods. The eye-tracking results suggest that when grammatical morphemes are congruent (i.e., those present in the L1; in this case, present progressive), L2 learners can automatically activate their meanings during processing and apply them in event recognition similarly to native speakers. When morphemes are incongruent (i.e., not present in L1; here, past simple and present perfect), L2 learners struggle to process them. However, to acquire grammatical knowledge successfully, L2 learners must be able to effectively process and apply this knowledge in real-time. From this perspective, incongruent L2 morphemes may not be fully integrated into L2 grammar, making it impossible to achieve native-like competence, as learners might always struggle to activate their meanings automatically during real-time processing, as suggested by the MCH.

7.2 RQ2: Do Chinese and Arabic Learners of English Exhibit Grammaticality Effects for Tense-Aspect Structures Similar to Native English Speakers in the Sentence-Matching Task?

<u>Predictions:</u>

a) *The L1 English group is anticipated to demonstrate online sensitivity to violations in present perfect and present progressive sentences, but not in past simple sentences.*
b) *The L1 Chinese group is expected to exhibit online sensitivity only to violations in present progressive sentences, with no sensitivity to violations in past simple and present perfect sentences.*
c) *The L1 Arabic group will also demonstrate online sensitivity to violations in present progressive and past simple sentences, without exhibiting sensitivity to violations in present perfect sentences.*

The predictions were not realised, and surprisingly, opposite results were suggested in some cases. In the native English group, reliable grammaticality effects were found for the past simple and present progressive sentences, as indicated by significantly longer RTs in the ungrammatical conditions compared to the grammatical ones. This suggests that participants were sensitive to the adverbial-tense mismatches. These results partially contradict the findings from Roberts and Liszka's (2013) study, which adopted the same experimental manipulation (i.e., a mismatch introduced between the temporal adverbial and the following verb). Specifically, in Roberts and Liszka's (2013) study, native English participants showed online sensitivity to mismatches in present perfect sentences but not in past simple sentences in the SPR task. The researchers explained that this reflects an ongoing change in British English, influenced by American English, which allows the usage of past simple forms to convey present perfect meaning (Roberts & Liszka, 2013). This is why we made the prediction; however, the results of the current study suggest that the past simple mismatches/ungrammaticalities did indeed cause a processing cost for the English speakers (e.g., *Last month, Susan ate only*… vs. **For a month now, Susan ate only*…). This implies that our English participants (British) still found it somewhat ungrammatical to associate a past simple form with a present perfect context, at least in online comprehension. However, it is crucial to approach these findings with some caution. Although the p value reached statistical significance (p = .02), the effect size was considered very small (d = .21) based on Plonsky and Oswald (2014) benchmarks. Similarly, the English group also exhibited a significant grammaticality effect for present progressive sentences (p = .002), but once again, the effect size was extremely small (d = .07).

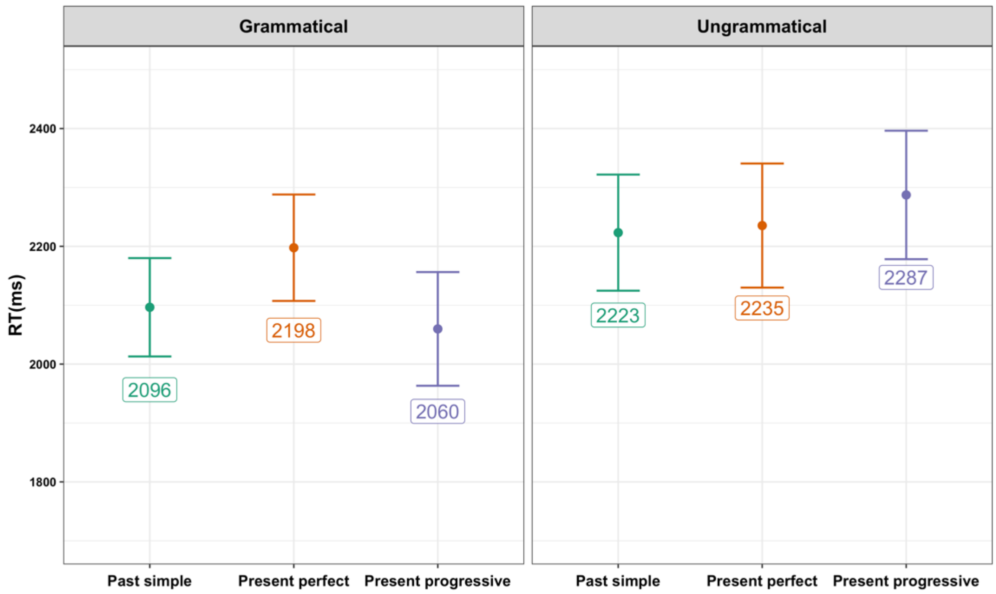
Contrary to the findings of Roberts and Liszka (2013), the English participants in the current study did not exhibit sensitivity to ungrammaticalities in present perfect sentences (e.g., *For the past year, my two brothers have studied*... vs. **Last year, my two brothers have* *studied…*). This result is somewhat surprising given that it is considered fully ungrammatical to use present perfect forms in the past simple context. In this case, when *have studied* is used, it implies that the time being discussed includes the moment of utterance. Consequently, employing adverbs specifying a specific past time *last year* becomes incompatible, as it would exclude the time of utterance and result in complete ungrammaticality (Klein, 1992). The lack of sensitivity among English participants could be attributed to tense-aspect violations posing a lesser processing burden compared to semantic and morphosyntactic violations, as suggested by the ERP data in Flecken et al.'s (2015) study. However, this explanation does not clarify why the same English participants in this study showed sensitivity to manipulations in past simple and present progressive contexts, despite the observed grammaticality effects being very small.

Another likely explanation could be the different reaction-time tasks employed in the present study (SM task) and the Roberts and Liszka's (2013) study (SPR tasks). The SPR task measures sentence processing on a word-by-word basis, whereas the SM task measures processing time only after participants fully read the sentence (Jiang, 2012). Thus, processing information about the critical words cannot be examined within the SM task. As discussed in Section 5.4, the SM task has yielded mixed results in the existing literature and has sometimes failed to detect grammaticality effects in both non-native and native speakers for certain grammatical structures (Duffield, White, Garavito, et al., 2002; Duffield & White, 1999; Eubank, 1993; Freedman & Forster, 1985; Gass, 2001; Verhagen, 2011). This, in fact, motivated the adoption of the SM task in the current study, with the aim of gathering more empirical evidence to better assess the validity of this method.

Upon closer examination of the RT differences in the present perfect items, it can be seen that the English group actually exhibited the expected pattern: taking longer time to match ungrammatical sentence pairs (M = 2235, CI = [2130, 2341]) compared to grammatical pairs (M = 2198, CI = [2107, 2288]). However, the RT differences between these two conditions simply failed to reach statistical significance. Two possible explanations can be proposed. Firstly, the lack of significance might arise from processing difficulty with grammatical pairs causing inhibition, or alternatively, from easier processing of ungrammatical pairs causing facilitation (Duffield et al., 2007). If inhibition is the cause, response times for grammatical pairs should be longer than those for comparable grammatical pairs; on the other hand, if facilitation is the cause, response times for ungrammatical pairs should be shorter than those for comparable ungrammatical pairs (Duffield et al., 2007). Secondly, the lack of a significant difference might simply be attributed to the size of the grammaticality effect being too small to reach statistical significance (Duffield, White, Garavito, et al., 2002).

Figure 7.2.1

Mean Response Times (RTs) of the English Group in Grammatical and Ungrammatical Conditions in the Sentence-Matching Task (Error Bars = 95% CI)



Note. The rounded mean RTs are displayed in the text boxes under each error bar.

In the present study, for the native speakers (see Figure 7.2.1), the average RT in grammatical conditions (across all three structures) was 2118 ms, while for ungrammatical conditions it was 2248 ms. A comparison of the average RTs with their RTs for present perfect structures reveals that the native speakers exhibited a stronger inhibition effect for grammatical pairs (+80 ms) than a facilitation effect for ungrammatical pairs (-13 ms). This indicates that the lack of an effect here can be attributed to the fact that the English group took much longer to match the present perfect grammatical pairs compared to other grammatical pairs.

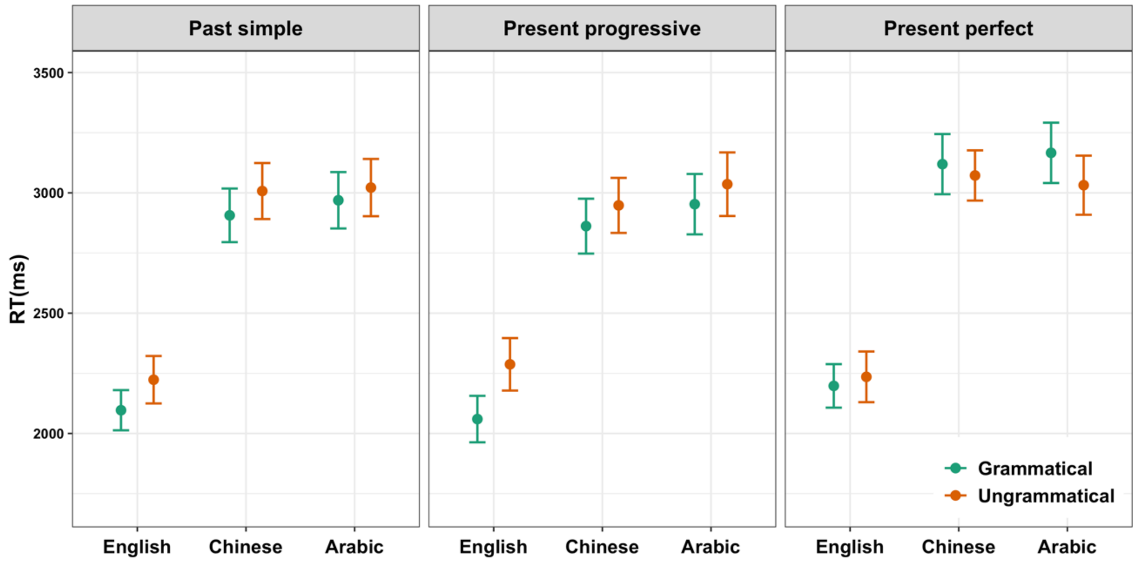
This raises the question: why did native English speakers respond more slowly to present perfect grammatical pairs? Drawing on some corpus evidence (Hundt & Smith, 2009; Schlüter, 2002; Werner, 2014), a likely explanation is that in both British and American varieties of English, roughly two thirds of the utterances employing the present perfect form are not accompanied by explicit temporal adverbials. This suggests that the use of present perfect in real life relies on verbal structures without temporal specifications. Hundt and Smith (2009) have even argued that the observation that most present perfect forms are used without specific triggering adverbials helps explain ongoing changes in their usage, particularly given that the concept of current relevance in present perfect often involves subjective interpretations by the speaker. In the current study, the grammatical pairs were constructed by associating indefinite temporal adverbials that extend to the present and the present perfect forms (e.g., *For the past year, my two brothers have studied Italian at a school in Rome*), which are less commonly used in real-life contexts. In addition, corpus evidence suggests that in American English, when an indefinite temporal adverb is used, the past simple is used more frequently than the present perfect, particularly in less formal contexts (Werner, 2013). Therefore, due to their less common usage and the ongoing influence from American English, the present perfect grammatical pairs might have caused a certain level of inhibition among the native English speakers in this study, leading to longer RTs for these participants.

Now turning to the L2 participants, largely contrary to the predictions, neither of the L2 groups exhibited a grammaticality effect in any of the three structures. This finding seems to align with the results of the SM experiment conducted by Gass (2001), where L2 participants similarly failed to demonstrate a grammaticality effect in any of the tested conditions. Gass (2001) suggested that this lack of effect could be attributed to the validity of the SM task being limited to L2 learners with a certain proficiency level. Below this threshold, learners might struggle to construct high-level representations and may resort to a word-by-word matching strategy instead of engaging in genuine reading and processing. However, this proposed explanation seems unlikely in the present study. The participants in this research were classified as upper-intermediate in proficiency. Also, the task design implemented in this study involved the disappearance of the first sentence before the onset of the second sentence, rather than displaying both sentences simultaneously on the screen. This design choice could make word-by-word matching extremely difficult for the L2 participants (Jiang, 2012). Nevertheless, it cannot be ruled out that this choice might have introduced new factors, such as working memory (Jiang, 2012), that could have influenced the results.

Alternatively, the absence of effects in the L2 groups could be interpreted as the evidence that the L2 groups were unable to distinguish between ungrammatical and grammatical representations, especially considering that the English group showed reliable grammaticality effects for both past simple and present progressive items. However, as discussed above, the size of the grammaticality effects observed in the English group was extremely small. This raises the possibility that the L2 groups may have indeed displayed RT tendencies in the correct direction, but these differences were not significant due to their limited magnitude. This speculation seems to be the case when looking at the RT patterns exhibited by the L2 groups (see Figure 7.2.2).

Figure 7.2.2

Mean Response Times (RTs) of the Three Groups in Grammatical and Ungrammatical Conditions in the Sentence-Matching Task (Error Bars = 95% CI)



For the past simple and present progressive items, both the Chinese and Arabic groups exhibited RT patterns that aligned with target-like behaviour. Specifically, longer RTs were consistently found in the ungrammatical conditions compared to the grammatical conditions. This suggests that both L2 groups might have demonstrated a certain level of sensitivity to ungrammaticalities for these two structures. It is possible that the design feature of the current SM task, which required participants to retain the first sentence of the pair in their minds while reading the second sentence, imposed a greater cognitive load. This increased processing burden could have contributed to the grammaticality effect failing to reach significance, unlike the native speaker group.

The speculation gains more certainty when examining the RT patterns of L2 groups for the present perfect items, where a reversed direction from the English group was observed. Specifically, the L2 participants took longer to match the grammatical pairs than the ungrammatical pairs, as shown in Figure 7.2.2. This reversed direction in RTs for the present perfect items could be interpreted as stronger evidence that the L2 groups did not show any sensitivity to the ungrammaticality in these sentences. It indicates that their grammatical representation might indeed differ from that of the native English group.

To sum up, the findings of the L2 groups in the SM task can be interpreted as follows: both L2 groups demonstrated a certain level of grammaticality effect for the past simple and present progressive items, as evidenced from their target-like RT directions (although these RT differences were not statistically significant). In contrast, neither group displayed any grammaticality effects for present perfect structures, as suggested by their reversed RT patterns.

Firstly, for past simple items, it is surprising to see that the Chinese group responded to them in the right direction, showing some sensitivities to ungrammaticalities (e.g., **Since last year, my grandmother taught me things about plants.*), which is contrary to the initial prediction. One possible explanation is related to their strategy use. Since Chinese does not encode tense morphologically, Chinese learners of English tend to rely on lexical or pragmatical devices to encode temporality, regardless of their proficiency levels (Ma et al., 2022). Moreover, the Chinese participants in the present study all received years of formal English instruction. In the classroom context, a "signal word" or context approach is often employed by teachers and grammar reference books, where temporal adverbials are emphasised to introduce and reinforce learners’ knowledge of tense-aspect marking (Fuchs et al., 2016; S. Yang & Huang, 2004). This may lead the Chinese group to use temporal adverbials as cues in their online interpretation. For example, they might apply the explicit rule they learned: when "for" or "since" is used, present perfect should be employed rather than past simple. However, if this were the case, it is puzzling why they did not apply the same strategy to ungrammatical present perfect pairs (e.g., **Last year, my grandmother has taught me things about plants*). The combined evidence suggests that, due to time pressure during the task, participants were unlikely to rely on explicit rules. This implies that their implicit knowledge was likely the primary measure. Therefore, a more plausible conclusion is that the Chinese group indeed demonstrated some underlying grammatical knowledge of English past simple. Similar to the English group, they found it somewhat ungrammatical to associate past simple forms (e.g., *taught*) and past time adverbials denoting current relevance (e.g., *since last year*). Nevertheless, since the RT difference did not reach significance, it can only be said that they were moving towards a target-like pattern. Given that they were able to use the grammatical information for successful online comprehension, as shown in the VW study, the results of the SM task confirm the earlier argument that detecting errors in grammatically incorrect sentences might differ from encoding grammatical information facilitatively in fully grammatical sentences.

The Arabic group demonstrated a certain grammaticality effect for the past simple items. It was predicted that Arabic learners would perform better than their Chinese counterparts due to an L1-based advantage. However, this advantage was not observed, as no difference was found between the L2 groups. The RT differences within the Arabic group also failed to reach significance. It is suspected that this lack of difference could also be caused by the influence of their L1 grammar representations. While Arabic does grammaticalise past tense, it does not explicitly mark perfect meaning. Both past simple and present perfect are conveyed using the perfective form. Although adverbials are commonly utilised to signify present perfect interpretations, the same perfective forms are employed. It is possible that learners were guided by their L1 grammar, and thus did not perceive a semantic clash when encountering such sentences. This finding is reflected in the Arabic group's AJT ratings, where they exhibited a slightly higher acceptance of such sentences.

For the present perfect items, the L2 groups did not show any grammaticality effect for present perfect, as indicated by their reversed RT directions from the English group, which aligns with the predictions. This suggests that the use of definite past time adverbs (e.g., *last year*) with the present perfect (e.g., *has taught*) did not result in a clash for the learners, implying that they might have non-target-like underlying knowledge. These results can be attributed to a transfer effect from their L1s. As discussed in Section 3.2, while Chinese uses perfect marker sentence-final *LE* to indicate current relevance of an event, it can also be used in the presence of the definite past time adverbial to simply convey a perfective meaning (e.g., *Shang xing-qi-wu wo qu jian-shen-fang you-yong LE* / 'Last Friday I went swimming at the gym'). This might explain why a clash did not occur for them when encountering such sentences in their L2 English.

Similar evidence is presented in two recent ERP studies conducted by Li et al. (2018, 2023). In these studies, unlike native speakers, upper-intermediate Chinese learners of English did not display a N400 modulation elicited by time clashes between clauses (e.g., **After the director of the school has resigned from the university, he worked for a multinational*). More interestingly, these learners showed reversed patterns compared to native speakers in processing time gaps between clauses, exhibiting more negative N400 amplitudes for compatible sentences with present perfect (e.g., *After the director… has resigned ..., he will work ...*) than in incorrect gap conditions with past perfect. In addition, these participants displayed more positive N400 amplitudes for sentences with present simple (e.g., *After the director ... resigns ..., he will work ...*) compared to other acceptable combinations involving present perfect. The authors concluded that the Chinese learners might overly rely on the adverbial *after* rather than tense morphology to compute temporal order and experience a form of tense blindness in real-time comprehension. Particularly, they seem to conceptualise English present perfect as part of the past time sphere, whereas native speakers conceptualise it in a pre-present zone that precedes the current moment (Y. Li et al., 2023). This evidence can also account for the non-target-like performance of the Chinese group in the present study.

The results of the Arabic group aligns with Farina's (2017) findings from the SPR task, where Arabic learners of English also demonstrated a lack of sensitivity to similar violations in present perfect sentences (e.g., **In two minutes, the dog has chased the foxes to their nearby burrow while barking to alert the hunters*). Following a similar line of explanation as presented by Farina (2017), their performance is possibly due to transfer from their L1 grammar. As discussed in Section 3.3, while English overtly grammaticalises perfect meaning through morphological forms, Arabic encodes it covertly and extralinguistically through adverbials and context. Specifically, Arabic does not explicitly mark current relevance, which is crucial for encoding present perfect meaning in English. It is plausible that the Arabic learners encountered difficulty integrating the concept of current relevance associated with the present perfect into their interlanguage grammar. Consequently, they may have encoded only the past meaning, making the sentences semantically compatible for them.

In the case of present progressive items, both L2 groups showed a certain grammaticality effect. They found ungrammatical sentences such as *"*Lately, Ella is dancing to her favourite song in her bedroom*" to be somewhat ungrammatical. This aligns with Chan's (2012) findings, where Chinese learners were also sensitive to clashes between temporal adverbials (e.g., recently) and present progressive forms in the SPR tasks. One explanation could be that the English progressive aspect holds a special position in the tense-aspect developmental sequence for all learners, as its morpheme is often acquired earlier than other morphological forms (e.g., *-ed*) due to its salient nature (Housen, 2002). However, this is unlikely because much pervious research has shown that adult learners whose L1 lacks the progressive aspect encounter persistent problem in both processing and production of progressive forms (e.g., Chan, 2012; Hawkins et al., 2008; Liszka, 2009). The absence of a grammaticalized progressive aspect can limit the learners' ability to represent the idea of ongoing actions grammatically. Thus, in the present case, positive L1 transfer might better explain the observed effect in the L2 groups. Chinese marks progressive aspect through preverbal marker ZAI. The marker ZAI can be used to describe an ongoing situation both in the present (the default reading) and the past (Lin, 2006); however, when explicit temporal adverbials are used to denote the present (e.g., right now, currently), only present progressive meaning is possible. This can explain why the Chinese learners were able to detect the clash between progressive forms and adverbials (e.g., *lately, recently*) that do not allow for an ongoing interpretation. Similarly, the Arabic learners experience positive transfer due to the ongoing grammaticalization of the progressive aspect through the inflected forms of *gaaʕid* in many Arabic spoken dialects. Whenever this form is used, a progressive meaning can be expressed.

In summary, the findings of the sentence-matching task were unexpected. The L2 groups did not exhibit any grammaticality effects for the three structures under investigation. The English group showed a reliable effect for past simple and present progressive items but not for present perfect items. However, the detected effects were relatively small, making direct comparison with those of the L2 groups challenging. Despite this, insights can be gathered by examining RT directions. The English group displayed longer RTs in ungrammatical conditions across all three structures, whereas such consistent patterns were not observed in the L2 groups. Specifically, both L2 groups demonstrated target-like RT directions for present progressive items, which mirrors their performance in the visual-world task where they also exhibited native-like processing patterns of the progressive morpheme.

Both L2 groups showed target-like RT directions for past simple items; however, the Arabic group did not demonstrate the predicted advantage over the Chinese group, contrary to what was observed in the visual-world task. This discrepancy may be attributed to potential negative L1 transfer experienced by Arabic learners in the sentence-matching task. Lastly, for the present perfect items, both L2 groups displayed reversed RT directions compared to the English group, indicating non-native-like underlying knowledge. This finding also mirrors their performance in the visual-world task.

The combine findings from the two online processing tasks, each employing different experimental paradigms, point to a persistent influence of learners' L1 grammar on L2 processing. This generally supports the RDH and MCH by providing evidence from tense-aspect acquisition research, consistent with prior processing studies in this domain (e.g., Chan, 2012; Roberts & Liszka, 2013; Vogel, 2017).

7.3 RQ3: Do Chinese and Arabic Learners of English Perform Differently from the Native English Speakers in the Sentence Imitation Task?

Predictions:

a) *The L2 groups will perform worse than the L1 English group.*
b) *The L1 English group will consistently display target-like patterns, correcting more ungrammatical sentences rather than introducing errors into grammatical sentences across all three tense contexts.*
c) *The L1 Chinese group will demonstrate target-like patterns in the present progressive context, but not in the past simple and present perfect contexts.*
d) *The L1 Arabic group is expected to perform in a target-like manner in both past simple and present progressive contexts, but not in the present perfect context.*

These predictions were mostly confirmed. Firstly, the English group exhibited consistent patterns, showing a strong inclination to normalise ungrammatical sentences rather than introducing errors into grammatical sentences across the three tense contexts. Specifically, they corrected 73.2%, 80.4%, and 88.5% of the ungrammatical sentences in the past simple, present perfect, and present progressive contexts, respectively. In addition, it was rare for them to transform grammatical sentences into ungrammatical ones when repeating grammatical items (all below 2.6%). However, one might wonder why the native group displayed greater variability in their percentage scores when repeating ungrammatical sentences compared to grammatical sentences. Also, it raises the question of why they did not consistently perform at a ceiling level, correcting all ungrammatical sentences during repetition. These questions can be addressed by examining the standard deviations.

Table 7.3.1

Standard Deviations (SD) Within the English Group in Their Percentage Scores in the Elicited Imitation Task

	SD	
Type	**UG → G**	**G → UG**
Past simple	25.2%	2.6%
Present perfect	31.8%	6.4%
Present progressive	27.3%	5.6%

Note. Change type: UG → G (ungrammatical to grammatical); G → UG (grammatical to ungrammatical).

As illustrated in Table 7.3.1, higher standard deviations are observed within the English group when repeating ungrammatical sentences, which indicates greater variability among individual participants in their imitation of these sentences. This means that some English participants may have consciously or unconsciously corrected the ungrammatical sentences based on their own grammatical knowledge, while others may have intentionally repeated the sentences verbatim, possibly believing that this was the purpose of this task. It is important to note that participants were explicitly instructed to repeat the sentences as given in the task instruction. This was intended to prevent them from becoming aware of the presence of ungrammatical sentences, thereby minimising the use of their explicit knowledge. Although a drawing task was inserted before the repetition, the observed variability indicates that some native-speaking participants may still have memorised the sentences without constructing the meaning themselves. Consequently, their imitative responses might reflect memory span rather than grammatical knowledge. On the other hand, other participants may have processed and reconstructed the items internally, aligning with the task's purpose and rationale.

Turning to the L2 groups, both Chinese and Arabic groups performed much worse than the native English group, aligning with the predictions. As illustrated in Figure 7.3.1, in comparison to the English group, both L2 groups corrected significantly fewer ungrammatical sentences. Also, they tended to introduce more errors into grammatical sentences, as shown in Figure 7.3.2. These results are consistent with a substantial body of previous research on L2 tense-aspect, which has observed non-native-like performance in the oral production of learner groups, even at advanced proficiency levels (e.g., Domínguez et al., 2017; Hawkins & Liszka, 2003; Liszka, 2004) and in contexts where predictions suggested no difficulty due to L1 facilitative effects (e.g., Amaro et al., 2018).

Figure 7.3.1

Percentage of Corrected Ungrammatical Sentences by Three Groups in the Elicited Imitation Task (Error Bars = 95% CI)

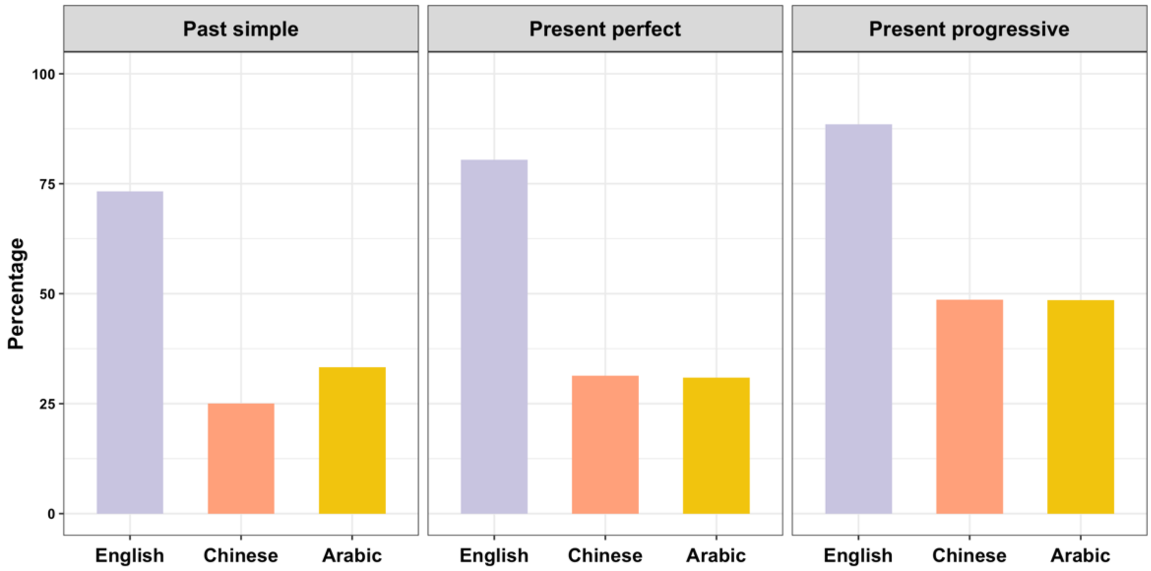
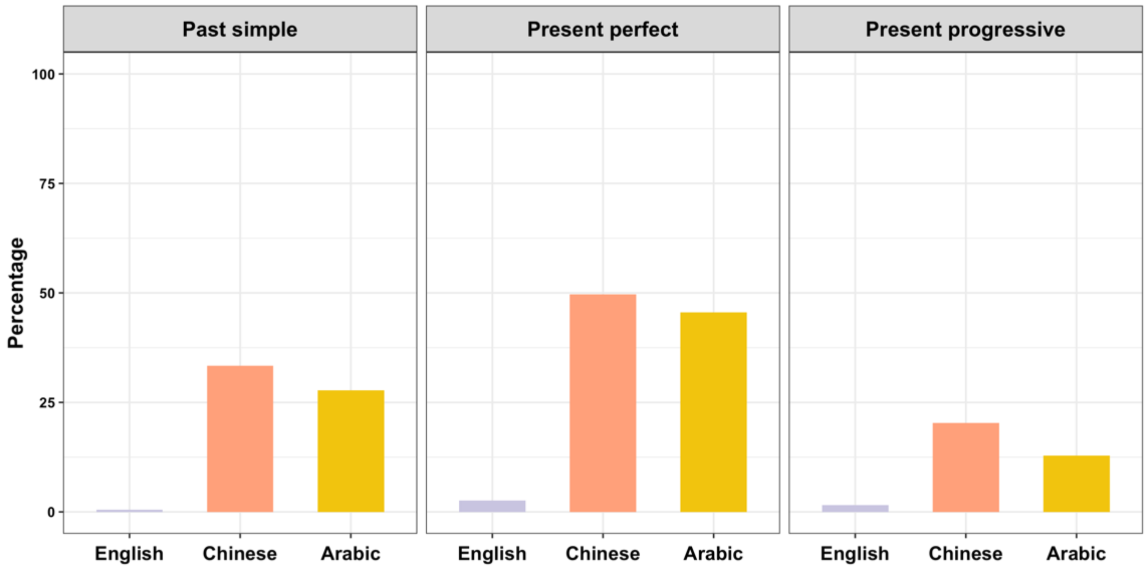


Figure 7.3.2

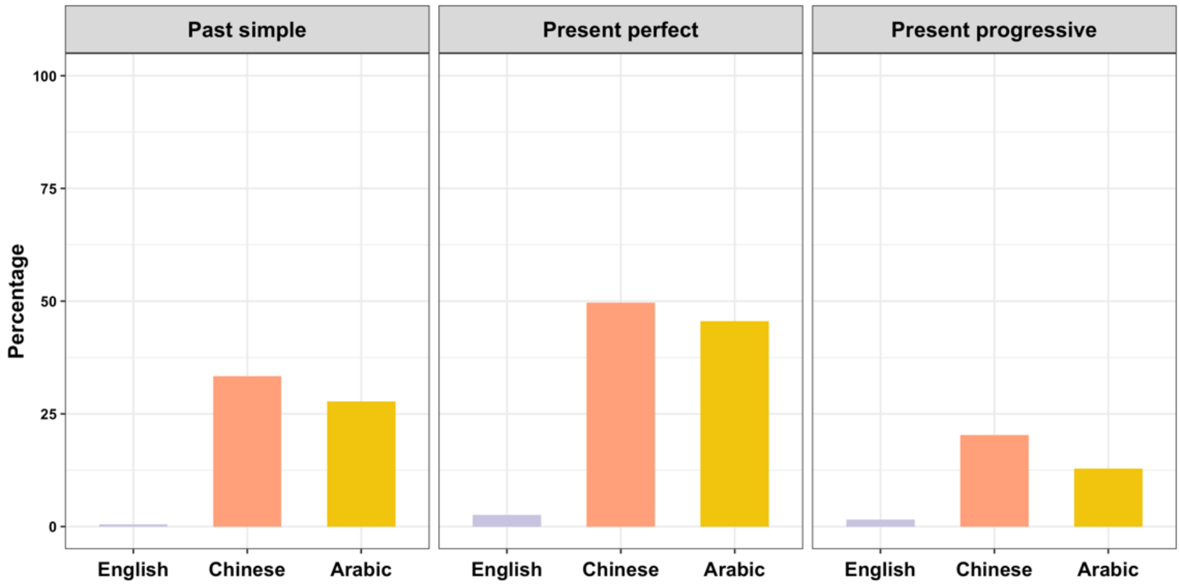
Percentage of Errors Introduced to Grammatical Sentences by Three Groups in the Elicited Imitation Task (Error Bars = 95% CI)



One plausible explanation is related to the nature of the oral production task itself. As noted by Grüter et al. (2012), spoken language production takes place in real-time, which is inherently influenced by the demands of real-time processing. Similar to other online tasks, communicative pressure can impose a higher processing burden, requiring L2 learners to retrieve and articulate words quickly (Alemán Bañón et al., 2017). Another explanation attributes the non-native-like performance to production-specific issues. This aligns with the accounts under the MSIH (Hazendar & Schwartz, 1997; Prévost & White, 2000b; White, 2011). Specifically, this hypothesis argues that persisting errors during oral communication are likely due to performance issues rather than a lack of underlying grammatical features. In the current case, errors in learners' spoken production may result from challenges in selecting and inserting correct inflectional morphology while speaking.

Figure 7.3.3

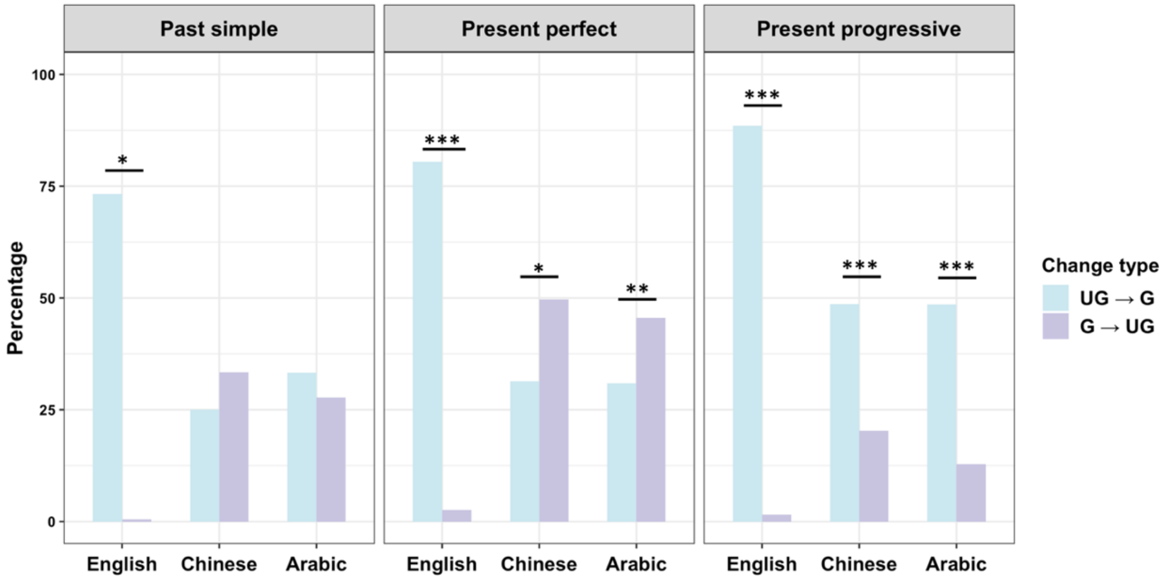
Percentage of Errors Introduced to Grammatical Sentences by Three Groups in the Elicited Imitation Task (Error Bars = 95% CI)



The first explanation appears unlikely because if processing pressure in real-time more were a general factor affecting both expressive and receptive domains, we would expect to see similar non-target-like performance in the L2 groups across the two online comprehension tasks. However, this was not observed. For instance, both L2 group exhibited native-like patterns in processing present progressive morphemes in the eye-tracking task, and they also demonstrated a target-like grammaticality effect for this structure in the sentence-matching task. In contrast, in the imitation task, they performed much worse than the native group. This combined evidence suggests that the difficulties encountered by the L2 groups in imitation were not solely due to processing demands during speech production. Similarly, the second explanation of production-specific problems (i.e., MSIH account) also fails to fully account for the results in the L2 groups. For example, both the Chinese and Arabic groups performed much worse and displayed a reversed pattern in direction of change compared to the English group (see Figure 7.3.4) when producing present perfect forms orally. These non-target-like patterns extended to their comprehension. Both groups were significantly slower in processing present perfect morphemes in the eye-tracking task and showed no grammaticality effect for the similar structure in the sentence-matching task. This evidence indicates that the difficulties experienced by L2 learners are not confined to oral production but also extend to online comprehension (Clahsen, Martzoukou, et al., 2010; Grüter et al., 2012).

Figure 7.3.4

A Comparison of the Three Groups in Their Change Types in the Elicited Imitation Task



Note. Change type: UG → G (ungrammatical to grammatical); G → UG (grammatical to ungrammatical). Asterisk sign indicates significant difference.

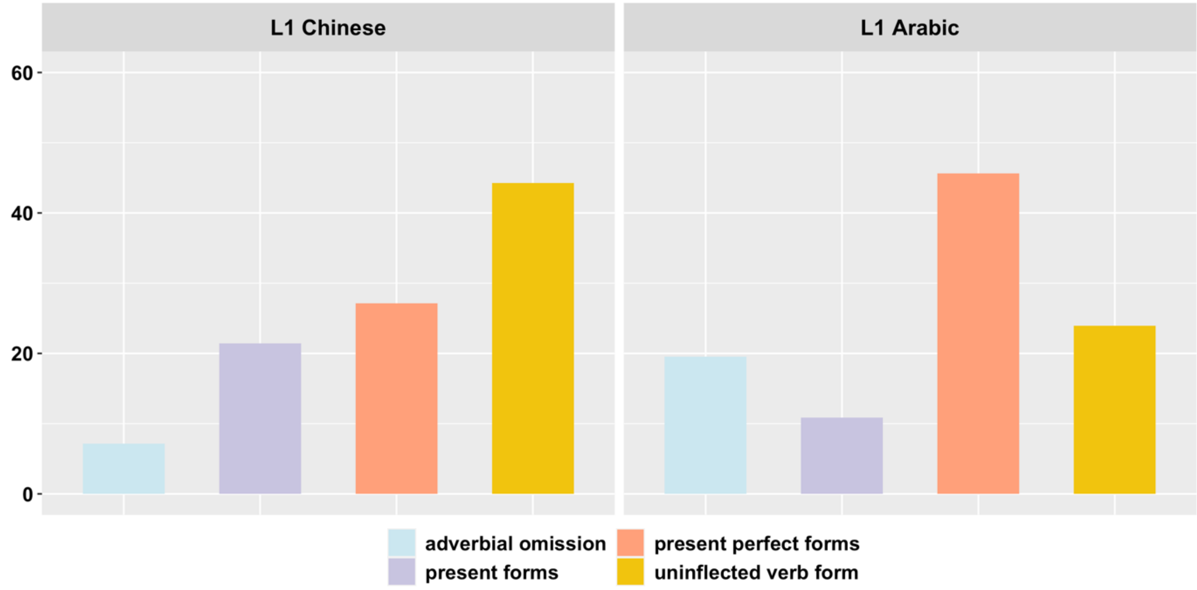
So, what has caused the non-target-like performance in the sentence production of L2 groups? This question requires a closer examination of the change patterns within the groups (see Figure 7.3.4). Across all three structures, the English group displayed a greater tendency to correct ungrammatical sentences rather than adding errors to grammatical sentences. These tendencies were all statistically significant. Similar to the English group, both L2 groups also demonstrated a significant tendency to correct errors in present progressive sentences while leaving grammatically correct ones unchanged. This indicates that the learner groups have underlying knowledge of the structure. Although their overall performance lagged behind that of the native group, their change tendencies suggest a progression towards a native-like level. They had not reached that level yet, possibly because these learners were at upper-intermediate/near advanced levels. These findings align with those reported by Zeng et al.'s (2021), who analysed spoken data from the learner corpus (LINSEI) of Chinese learners.

Their study revealed that learners at both intermediate and advanced proficiency levels exhibited no difficulty in producing progressive forms, especially with activity verbs. In addition, the results of the Arabic group also corroborate Kleinmann's (1977) finding, which showed that Arabic learners of English, whose L1 does not grammaticise progressive aspect, performed similarly to the Spanish and Portuguese learners, whose L1s explicitly mark progressive meaning in an oral task.

For the past simple items, the L2 groups demonstrated different change preferences. The Arabic group, similar to the English group, displayed a tendency to normalise ungrammatical sentences. On the other hand, the Chinese group exhibited an opposite trend; they tended to introduce errors into grammatically correct sentences. Although the observed tendency in the Arabic group was not statistically significant, it did align with the predicted direction, suggesting that Arabic learners were moving towards a more native-like level. In contrast, the Chinese group performed worse and exhibited a reversed direction of change, indicating a potential lack of underlying grammatical knowledge. This finding aligns with numerous studies documenting the consistent difficulty Chinese learners face with past tense forms in oral production (e.g., Amaro et al., 2018; C. Chen, 2009; Goad et al., 2003; Goad & White, 2006; Hawkins & Liszka, 2003; Kahoul, 2014; Lardiere, 1998). For example, the end-state learner Patty from Lardiere's (1998) study produced past simple morphology in less than 40% of obligatory contexts. Similarly, in Goad et al.'s (2003) study, high-intermediate Chinese learners correctly produced past simple about 57% of the time in an oral picture-description task.

Figure 7.3.5

Error Introduced into Grammatical Past Simple Sentences by the L2 Groups in the Elicited Imitation Task



In this task, the changes made to the orginal sentences are indicators of meaning reconstruction, and thus are more likely to reflect participants' underlying grammatical knowledge. In the Chinese group (see Figure 7.3.5), 44% of the errors (31/70) introduced into grammatical past simple sentences were related to the omission of verbal inflections (e.g., **Last year, many teachers apply new technologies in their teaching*). This result aligns with Hawkins & Liszka's (2003) study, which suggested that advanced Chinese learners produced uninflected verb forms in 38% of unambiguous past simple contexts during a spontaneous oral production task, a rate much higher than their Japanese and German counterparts. In the present study, the Chinese group also exhibited a significantly higher frequency of producing uninflected verbs compared to the Arabic group ($\chi^2 = 5.95$, $df = 1$, $p = .01$). This higher optionality with past tense forms can be attributed to the absence of grammatical past tense marking in Chinese, where uninflected verb forms are used in both past and present contexts.

It has resulted in persistent difficulties for learners in correctly producing English past tense forms in oral production.

In contrast, Arabic marks tense in a way similar to English, which may explain why the Arabic group performed better than the Chinese group and exhibited target-like change patterns when imitating past simple sentences. The lack of statistical significance in their performance could possibly be understood by examining the types of changes or errors introduced into grammatical sentences. As illustrated in Figure 7.7, approximately 46% (21/46) of the Arabic group's errors involved using present perfect forms. Although the Chinese group also used present perfect forms in their changes (27%, 19/70), the frequency of such error was significantly higher in the Arabic group ($\chi^2 = 4.71$, $df = 1$, $p = .03$). This suggests that the Arabic group was more likely to associate present perfect forms with the past simple context. Recall that in Arabic, the perfective form can express both preterit and present perfect meanings, whereas in English, the two meanings are realised by two different forms. Such differences in form-meaning mapping (DeKeyser, 2005; Slabakova, 2008) may pose persistent challenges for Arabic learners. This issue will be further explored after examining their performance in the present perfect context to provide a more comprehensive picture.

For present perfect structures, both L2 groups exhibited a different trend compared to the English group. They demonstrated more instances of introducing errors when repeating grammatically correct sentences than correcting ungrammatical sentences. This reversal in change direction suggests a potential lack of underlying grammatical knowledge about the present perfect structure, leading to learners' difficulty in production. These findings align with some previous research using oral production tasks (Erica & Wong, 2016; Liszka, 2004). For example, Liszka's (2004) found that advanced Chinese learners of English produced correct present perfect forms in approximately 47.4% of obligatory contexts.

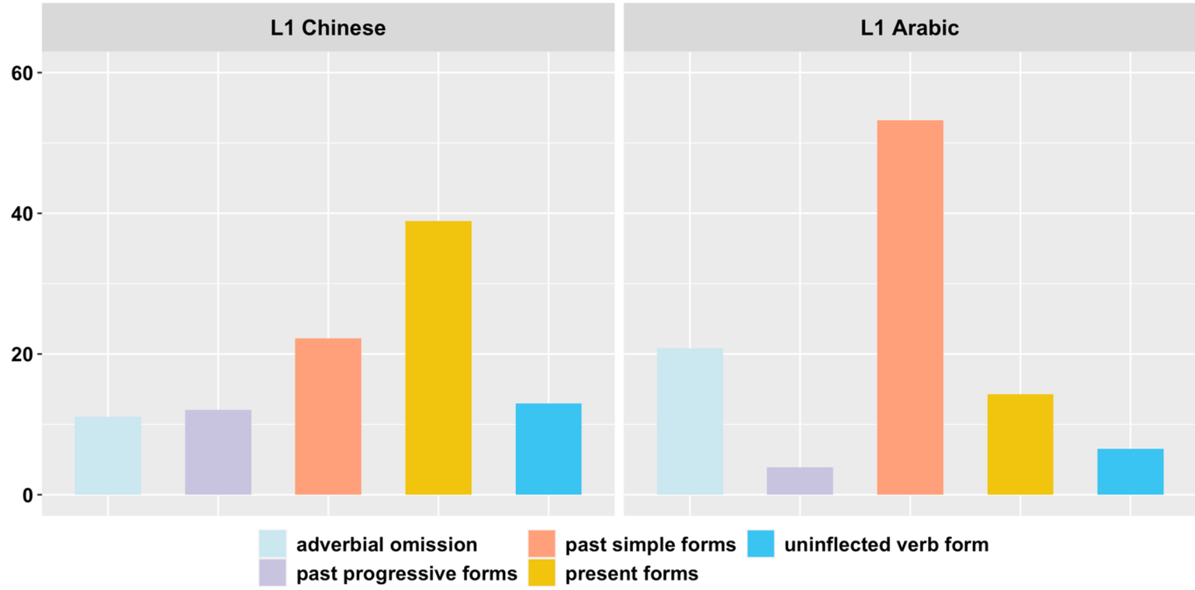
Similarly, Erica and Wong (2016) reported that both advanced and intermediate Chinese learners of English produced correct present perfect forms in less than 40% of obligatory contexts. More broadly, the results are also in line with Bardovi-Harlig's (2001) study, which observed that learners, including those from an Arabic background, often underused present perfect forms, particularly in contexts involving with temporal adverbials (e.g., *for*- and *since*- phrases).

One explanation for the target-deviant performance of the Arabic and Chinese learners in producing present perfect forms could be associated with the difficulty of the structure itself. As argued by Housen (2002), the persent perfect is the second-last constrcution to be acquired by L2 learners of English and its production only emerges after the complete acquisition of the past simple by learners. This late acquisiton is not affected by explicit instruction (Housen, 2002). However, this explanation alone cannot fully account for the observed results when examining the changes/errors made by the participants when repeating grammatical present perfect sentences.

The qualitative results (see Figure 7.3.6) regarding the errors introduced by the two groups into grammatical present perfect sentences revealed interesting patterns. The Arabic group produced past simple forms (55%, 42/77) or simply omitted temporal adverbials (21%, 16/77). However, the frequency of using past simple forms was sigificantly higher ($\chi^2 = 15.2$, $df = 1$, $p < .001$), indicating a stronger perference for using past simple forms in contexts requiring present perfect. In contrast, the Chinese group showed a more balanced pattern, with 39% of their errors (42/108) involving the use of present forms and 25% (27/108) involving past simple forms. The alternating use of these two forms did not show a difference in frequency ($\chi^2 = 3.06$, $df = 1$, $p = .08$). These different error patterns seem to reflect the influences of the L1 grammar of each group.

Figure 7.3.6

Error Introduced into Grammatical Present Perfect Sentences by the L2 Groups in the Elicited Imitation Task



According to Bardovi-Harlig (2001), when learners mistakenly use past or present tense in the present perfect context, it suggests that they encode only part of the two meanings associated with the present perfect: [+past] and [+current relevance]. Using preterit forms indicates association with events before the present, encoding [+past] but not [+current relevance]. On the other hand, using present forms suggests an association with the present moment, encoding [+current relevance] but not [+past]. These errors are viewed as undergeneralisations, reflecting an emerging form-meaning association rather than being entirely incorrect; instead, they are considered incomplete associations (Bardovi-Harlig, 2001). In the case of the Arabic learners, Arabic grammatically encodes past meaning but does not mark current relevance explicitly. The meaning of current relevance is established through context and adverbials. As a result, the Arabic learners faced difficulties integrating the meaning of current relevance into their interlanguage grammar. This challenge is reflected in their tendency to prefer past simple forms, as these forms align more closely with the [+past] feature they are accustomed to from their L1. However, Bardovi-Harlig (2001) also argued that the use of past simple in present perfect contexts is not necessarliy a direct result of L1 influence. Instead, L2 learners, irespective of their L1 backgrounds, are likely to use past simple forms due to the semantic proximity between past simple and present perfect, both of which share the [+past] feature. Neverthless, this explanation falls short in addressing the Arabic group's production of present perfect forms in past simple contexts, which involved incorporating [+current relevance] where it is not appropriate. More importantly, such patterns were not observed in the Chinese group. Thus, the plausible conclusion is that the Arabic group struggled with establishing accurate form-meaning relationships. In English, the past simple and present perfect meanings are overly realised in two different forms: *v-ed* and *have+v-ed/en*; in contrast, these two meanings are mapped onto one form in Arabic. This difference results in Arabic learners equating the two English forms and producing both forms interchangeably.

Now, turning to the Chinese group, we observed their alternation between past simple and present forms in the context of the present perfect—a pattern also found in Liszka's (2004) study. The use of present forms suggests that Chinese learners can convey [+current relevance] but struggle with [+past] meaning. This challenge appears rooted in the grammaticalization patterns of their L1. Chinese employs the current relevance marker LE but lacks grammaticalised tense, leading to their difficulties in expressing [+past] meaning grammatically. However, how can their use of past forms be accounted for? In line with Liszka's (2004) speculation, it is possible that Chinese learners might employ past forms not necessarily to indicate past meaning but rather as a marker of the perfective aspect. In Chinese, the perfective marker, verbal-final *LE,* signifies completion without strictly placing the event in the past. This perfective marker can co-occur with the current relevance marker *LE* to convey present perfect meanings.

In summary, the results from the elicited imitation task cannot be simply attributed to real-time performance issues or explained by the MSIH account. Instead, the observed differences in performance between the L2 groups provide some evidence for the Representational Deficit Accounts, which emphasise the influential role of the L1 and propose that absences in L1 properties can result in persistent deficits in L2 mental representations. For the past simple items, the Chinese group, compared to the other two groups, showed a reversed direction of change. They constantly substituted verb forms with uninflected ones when imitating grammatical sentences. For the present perfect items, both L2 groups demonstrated a reversed direction of change compared to the English group, indicating a lack of underlying grammatical knowledge. The Arabic group had difficulty distinguishing past simple and present perfect forms and their associated meanings. In Arabic, current relevance meaning is established through lexical means, so learners could not differentiate between present perfect and past simple in English, often producing them interchangeably. The difficulty of the Chinese group seemed to arise from their inability to encode [+past] meaning associated with present perfect, leading to their production of present forms in the present perfect context, unlike their Arabic counterparts.

7.4 RQ4: Do Chinese and Arabic Learners of English Exhibit Explicit Knowledge of English Tense-Aspect Compared to Native English Speakers, as Measured by the Offline Acceptability Judgement Task and Gap-Filling Task?

<u>Predictions:</u>

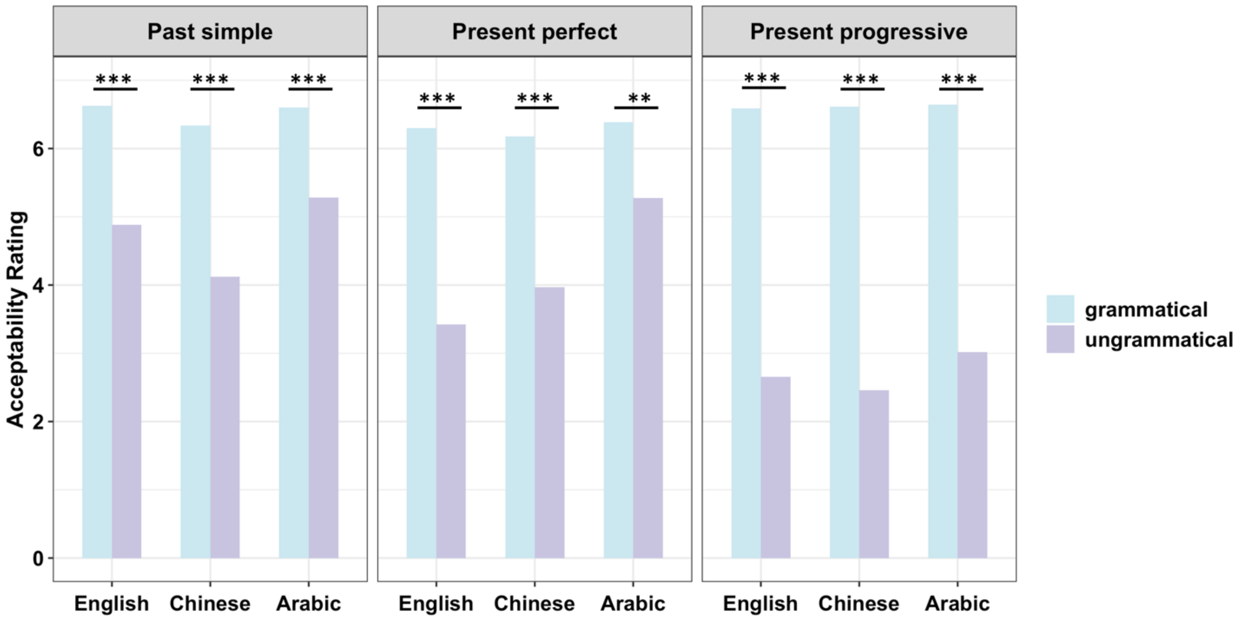
d) *All three groups are expected to demonstrate correct explicit knowledge of English past simple, present prefect, and present progressive.*
e) *All three groups will rate the ungrammatical sentences significantly lower than the grammatical sentences in the AJT task.*

f) *All three group will display a high level of accuracy in supplying English tense-aspect forms in the gap-filling Task.*

In the AJT task, the predictions were all confirmed. All groups successfully differentiated between correct and incorrect usages of the three structures, showing target-like explicit knowledge of English tense-aspect (see Figure 7.4.1).

Figure 7.4.1

Mean Ratings for the Three Structures by the Three Groups in the AJT Task



Specifically, for the past simple items, both L2 groups rated ungrammatical sentences significantly lower than grammatical ones. No differences were observed between the English group and either of the L2 groups in their ratings for both conditions. One interesting observation emerged between the L2 groups: the Arabic learners rated ungrammatical past simple sentences slightly higher compared to the Chinese learners. This difference was marginally significant ($\beta = 0.58$, $SE = 0.25$, $t = 2.33$, $p = .058$), which suggests that the Arabic group possibly displayed a greater acceptance of using past simple forms in a present perfect context (e.g., *For a month now, Susan ate only…*). In fact, this trend mirrors their performance in the sentence-matching task, where no L1-based advantage was observed compared to the Chinese group. To better understand this performance, it is crucial to examine the results for present perfect items as well.

For the present perfect items, both L2 groups performed similarly to the English group, exhibiting target-like patterns with significantly lower ratings for ungrammatical sentences compared to grammatical ones. However, a between-group comparisons revealed that the Arabic group once again assigned significantly higher ratings to ungrammatical sentences than the other two groups. This indicates a greater acceptance of adverbial-verb tense mismatch among the Arabic learners (e.g., **Last month, Susan has eaten only…*). The performance of the Arabic group, in fact, aligns with findings from Alruwaili's (2014) study, where Arabic learners struggled to reject English present perfect forms in a past simple context during the AJT task. This reflects a similarity to what is conveyed in their L1 grammar. As discussed earlier, Arabic lacks a grammatical marker to indicate current relevance, a crucial feature for expressing the perfect aspect (Badawi et al., 2004; Farina, 2017). The perfective form in Arabic can convey both past simple and present perfect meanings, which in English are distinctly marked by the use of *V-ed* and *have + V-ed* forms. Adverbials and particles with perfective forms are often employed in Arabic to indicate current relevance and differentiate present perfect from past simple meanings. Since Arabic maps these two meanings onto one form, Arabic learners might struggle to differentiate English present perfect from past simple forms. They may equate the form-meaning association of the present perfect and past simple, treating them similarly to their L1 grammar. This could explain their acceptance for both past simple and present perfect ungrammatical sentences. Despite the potential influence of their L1 grammar, the within-group results indicated that Arabic learners were still able to distinguish between correct and incorrect associations for both past simple and present perfect, showing target-like judgement patterns.

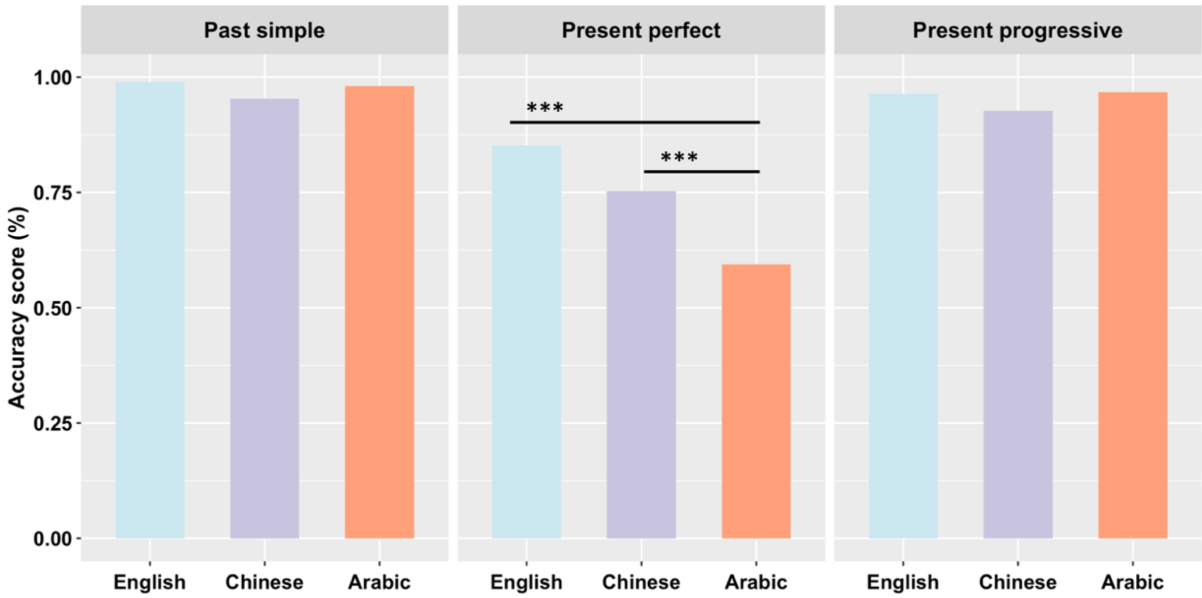
Another observation emerged within the native English group, where they demonstrated a higher level of acceptance for ungrammatical past simple sentences (mean rating = 4.88) compared to ungrammatical present perfect sentences (mean rating = 3.42). A subsequent analysis revealed that this difference was statistically significant (β = 0.75, *SE* = 0.08, *t* = 9.08, *p* < .001). This finding suggests that the English group was more tolerant of using past simple forms to convey present perfect meanings, a tendency not observed in their sentence-matching results. As noted in Section 7.2, this could reflect an ongoing change in British English influenced by the American variety, as Roberts and Liszka (2013) suggested. It is becoming increasingly acceptable to employ past simple to indicate current relevance, as seen in examples such as "*I already finished the work"*. Semantically, this acceptability might arise from the temporal adverbial's (e.g., *for a month now*) reference to a time span that includes both past and present. As explained by Roberts and Liszka (2013), this inclusion of past time in the adverbial makes the use of past simple forms not entirely ungrammatical.

Lastly, for present progressive items, no group differences were observed, and both L2 groups rated the sentences in a native-like manner. The findings are consistent with some previous research investigating the acquisition of English present progressive using offline comprehension/interpretation tasks (L1 Aabic: Alruwaili, 2014; Al-Thubaiti, 2015; L1 Chinese: Dudley & Slabakova, 2020; Gabriele & Maekawa, 2008).

Note that the materials utilised in the AJT task were identical to those employed in the VWP and SM tasks. However, a comparison of the results reveals that target-like performance was observed in the offline AJT task across the three structures, whereas such performance was not consistently demonstrated in the two online tasks. This discrepancy can be attributed to the different types of knowledge involved in these tasks. In the AJT task, learners were not under time pressure and had the flexibility to take their time in making judgments about language forms (Vafaee et al., 2017). As a result, their explicit knowledge was more likely to be accessed. Given that all learners had received English instruction in a formal classroom setting for years, it is not surprising that they all exhibited target-like knowledge of English tense-aspect usage in the AJT task. Further discussion regarding the contrasting performance between online and offline tasks will be presented in the next chapter.

Figure 7.4.2

A Comparison of the Three Groups in Their Accuracy Rates in the Gap-Filling Task



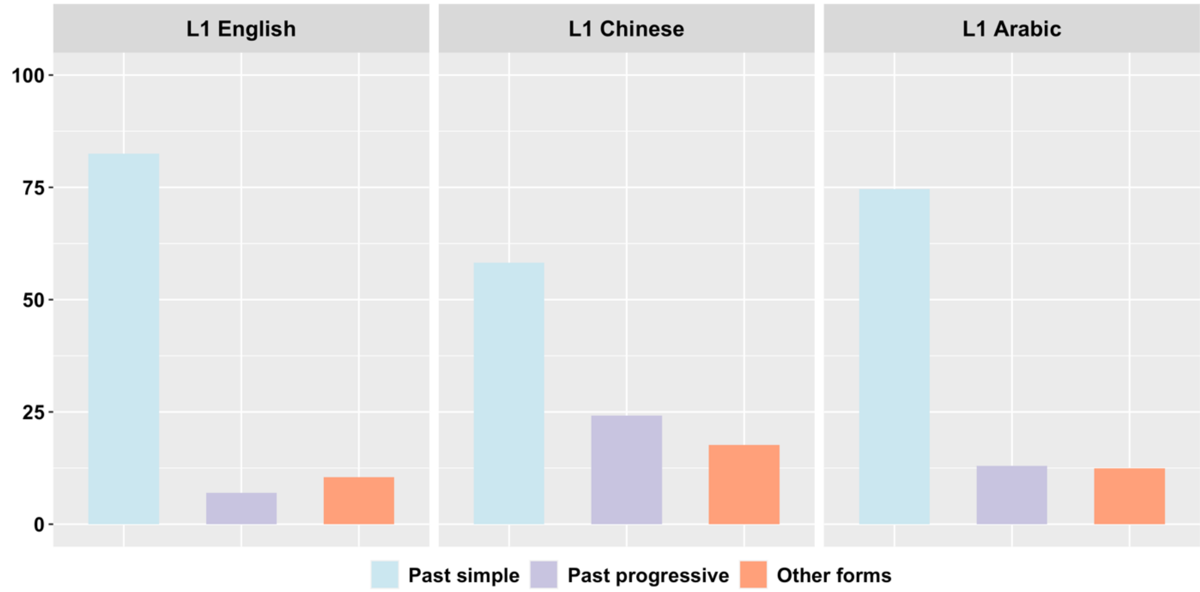
Now, turning to the gap-filling task, the predictions were mostly realised, except for the Arabic group. As illustrated in Figure 7.4.2, all three groups performed at ceiling in providing correct forms in obligatory past simple context, achieving accuracy rates above 95%. No significant group differences were observed. These results align with previous research on learners' written production of past tense forms (L1 Arabic: Alruwaili, 2014; L1 Chinese: Hawkins & Liszka, 2003; Lardiere, 2003). In those studies, both Arabic and

Chinese learners could produce correct past simple morphology in a native-like manner. The present results found similar results, suggesting that learners had correct metalinguistic/explicit knowledge.

In addition, all groups exhibited no difficulty in supplying present progressive morphology in the obligatory context, with accuracy rates above 94%. Both L2 groups demonstrated native-like performance not only in interpreting the appropriate context but also in providing the correct morphological forms.

Figure 7.4.3

Incorrect Forms Supplied by the Three Groups in the Present Perfect Context in the Gap-Filling Task



In the present perfect context, the results revealed that all three groups performed numerically worse than they did in the other two contexts. The Chinese group performed similarly to the English group and no difference was found between them. However, the Arabic group performed significantly worse compared to the other two groups, with a mean accuracy rate of only 45%, considerably lower than that of the Chinese (73%) and English groups (85%). To gain a better understanding, it is necessary to examine the incorrect forms provided by the three groups (see Figure 7.4.3).

Among the cases of incorrect supply, the English group exhibited a preference for utilising past simple forms (82.5%; 47/57). This tendency is somewhat consistent with their performance in the AJT task, where they demonstrated a greater acceptance of past simple forms in a present perfect context compared to present perfect forms in a past simple context. Although the experimental materials in these two tasks were not entirely identical, they were constructed in a similar manner.

Similarly, the Arabic group also displayed a consistent tendency to provide past simple forms (74.6%; 138/185). While past simple forms were also provided by the other two groups, the Arabic group's number of past simple forms was considerably higher than that of the English group (n = 47) and the Chinese group (n = 53). This higher number contributed to their statistically worse performance than the other two groups in the task overall. Their over-supply of past simple forms reflects their performance in the elicited imitation task, where they also incorrectly produced past simple forms in the present perfect context. This tendency can be attributed to the influence of their L1, where both past and present perfect meanings are often mapped onto a single form. They can associate the *-ed* marking with past events, which explains their native-like performance for past simple context. However, they appeared to extend the use of the past simple form to present perfect contexts. Therefore, they encounter prolonged challenges in employing the perfect form (*have+v-ed/en*) in their L2 English even in the written production task.

To sum up, the L2 groups performed better in the gap-filling task (written production) compared to the elicited imitation task (oral production). Note that the experimental materials were identical for the two tasks. In the imitation task, the errors in learners' production indicated a clear influence of their L1 grammar. In contrast, in the gap-filling task, this influence was not consistently observed, especially in the case of the Chinese group. On the other hand, the Arabic group's performance in the present perfect context lagged, as they demonstrated a tendency to provide past simple forms—a reflection of the pattern represented in their L1 grammar.

Chapter 8 General Discussion

This chapter aims to provide an overall discussion based on the combined results of the individual experiments. It begins the discussion by focusing on cross-linguistic influences, addressing both positive and negative L1 effects. Then, the discussion moves on to the comparison of participants' online and offline performance, exploring the potential influence of L1 on learners' implicit and explicit knowledge. Given that the present study measured both comprehension and production knowledge, the chapter also presents a discussion of the relationship between comprehension and production based on the obtained results. Finally, the chapter concludes by discussing the methodological and theoretical implications of the findings.

8.1 Cross-Linguistic Influences

The present study investigated the acquisition of L2 tense-aspect by adult Chinese and Arabic learners of English. To ensure a better comparison between the L2 groups, the Chinese and Arabic learners were matched in terms of their English proficiency levels, age of acquisition, years of formal instruction, and immersion experience in the UK. Their English proficiency was assessed by using the QPT test, IELTS (or other equivalent) tests results, and self-reported proficiency scores. The learners were categorised as being at an upper-intermediate/near-advanced level. The study employed five different tasks that crossed both online/offline and comprehension/production domains. Based on the discussion in Chapter 7, the findings across all the experiments suggest a persistent influence of L1 on learners' performance in both comprehension and production. This section will briefly recap the cross-linguistic differences before discussing the L1 transfer effects observed in the learner groups.

According to Barto-Sisamout et al. (2009), when comparing morphosyntactic systems between an L1 and an L2, four types of relationships may exist: 1) same/similar, where the grammatical feature is marked very similarly between L1 and L2; 2) L1-L2+, where the feature is not explicitly marked in L1 but in L2; 3) similar but different, where the featured is marked in both L1 and L2 but in different ways; and 4) L1+L2-, where a feature is marked in L1 but not in L2. With this background information in mind, the relationship between English and Chinese/Arabic languages in their grammatical making of tense and aspect can be better understood.

Firstly, English encodes the [+/-past] feature by distinguishing past and non-past events overtly through the morphological form *-ed*. Similarly, Arabic also grammaticalises tense morphologically: past tense is realised by using the perfective form of the verb, and present tense is encoded by using the imperfective form of the verb. On the other hand, Chinese differs from the other two languages and is considered a tenseless language, relying largely on lexical devices (e.g., temporal adverbials) to refer to time locations.

Secondly, the Asp[+/-prog] feature is encoded grammatically in English, and the progressive aspect is overtly marked by the verbal pattern of *be+v-ing*. As a language with rich aspect markers, Chinese also grammaticalises progressive aspect through the usage of preverbal marker ZAI. This marker, similar to the English progressive form, specifically denotes an ongoing situation that is occurring in the middle of the process. Although its usage can be optional when enough context is provided, it is required in most cases. In contrast to English and Chinese, Arabic lacks progressive aspect, and its imperfective verb forms express both progressive and non-progressive meaning. Nevertheless, currently, in spoken Arabic, there is arguably an ongoing process of grammaticalization of progressive aspect, involving the inflected form of *gaaʕd*. Whenever it is used, the progressive meaning can be achieved.

Thirdly, English distinguishes the [+/-perfect] meanings. For example, the present perfect meaning is overtly marked by the form *have+v-ed/en*, which essentially encodes two semantic features: [+past] and [+current relevance] (i.e., a past event that extends to the present). In contrast, Arabic does not realise the perfect/non-perfect distinction morphologically. The perfective verb forms that indicate past tense meaning are also used to express perfect meaning. Adverbials (e.g., already, just) are often employed to achieve the perfect interpretation. As for Chinese, it has a perfect marker, the sentence-final *LE*, to express perfect meaning, which specifically encodes feature [+current relevance]. However, it lacks grammatical tense to encode feature [+past].

The results of the present study indicate a positive L1 transfer effect for structures that are marked similarly in both L1 and L2. It was revealed that both Chinese and Arabic learners performed in a target-like manner in their comprehension and production of the English present progressive. Firstly, in the eye-tacking task, they exhibited native-like patterns and speed in processing progressive morpheme, and they were able to utilise the morphological marking to encode event information and achieve the target ongoing interpretation as rapidly as the native speakers. Furthermore, in the sentence-matching task, they showed certain sensitivities to ungrammaticalities, with longer RT observed in judging ungrammatical sentence pairs compared to grammatical ones. Although the observed differences were not statistically significant, it is suspected that this lack of significance may be attributed to the small size of the effect, especially when considering the fact that the effect size in the native English group was very small. This has made the results challenging to interpret, and there is a possibility that sentence-matching tasks might not be ideal for testing grammaticality effects among L2 learners. The validity of this method will be discussed in the following section (see Section 8.4). In addition, in the offline comprehension task (AJT), both L2 groups displayed no differences from the English group, demonstrating explicit knowledge of the correct and incorrect usage of the present progressive. Similarly, in the elicited imitation task, both L2 groups demonstrated a significant tendency to correct errors in ungrammatical present progressive sentences while leaving grammatically correct ones unchanged, like the English group. They also performed almost at ceiling in the gap-filling written production tasks, suggesting that they were able to produce the target form in the obligatory context.

These results confirm the findings from some previous research that has investigated the acquisition of English present progressive employing various online processing tasks (L1 Chinese: Chan, 2012; Yao & Chen, 2017), offline comprehension/interpretation tasks (L1 Chinese: Dudley & Slabakova, 2020; Gabriele & Maekawa, 2008; L1 Arabic: Alruwaili, 2014; Al-Thubaiti, 2015), oral production tasks (L1 Chinese: Dudley & Slabakova, 2020; L1 Arabic: Kleinmann, 1977), and written production tasks (L1 Arabic: Alruwaili, 2014). The present results, along with those previous studies, collectively suggest that the acquisition of English progressive does not pose a significant or persistent challenge to either Chinese or Arabic learners, due to a positive L1 transfer.

The present results suggest a negative L1 transfer effect for structures that are absent in the learners' L1 (i.e., L1-L2+). Notably, the Arabic group outperformed their Chinese counterparts in past simple items, particularly in the online processing and production tasks. Specifically, in the eye-tracking task, the Arabic group performed similarly to the native speakers and displayed a quick divergence point to the target area. Although the Chinese group could also utilise the past marking *-ed* and achieve the target interpretation of a completed event, the divergence point of their eye movements appeared much later than those of the Arabic and English groups. The results of the sentence-matching task did not reveal any difference between the L2 groups; however, as discussed earlier in Section 7.2, this was possibly due to negative transfer effects experienced by the Arabic group from their L1 when processing the tested materials. Moreover, in the imitation task, the Arabic group displayed a target-like direction of change, showing a tendency to correct ungrammatical sentences. In contrast, the Chinese group exhibited a reversed direction, suggesting a lack of underlying knowledge. They showed a tendency to omit verbal inflections when repeating grammatically correct sentences, due to a lack of grammaticalized tense in their L1.

These findings are consistent with Kahoul's (2014) study, where Arabic learners also performed better than their Chinese counterparts in the processing and production tasks. Furthermore, the results of the imitation task align with previous research that has documented persistent problems experienced by L1 Chinese learners in their oral production of L2 English past tense (e.g., Amaro et al., 2018; C. Chen, 2009; Goad et al., 2003; Goad & White, 2006; Hawkins & Liszka, 2003; Lardiere, 1998). The current results further suggest that such difficulty is not production-specific and can extend to online comprehension as well (C. Chen, 2009; Clahsen, Martzoukou, et al., 2010; Grüter et al., 2012). Essentially, the results provide evidence against the MSIH account and indicate that Chinese learners' poor production of the past tense is not simply caused by difficulty in realising surface forms; their non-native-like performance in processing implies that the cause might likely be an underlying representational issue.

Negative L1 influences were also observed for structures marked differently in the L1 and the L2. In the eye-tracking task, both L2 groups exhibited significantly slower processing of the present perfect morpheme compared to the native control group. They also displayed non-target-like RT directions in the sentence-matching task, indicating a lack of grammaticality effect. In addition, in the imitation task, both L2 groups demonstrated a reversed direction of change compared to the English group. They displayed a preference for introducing errors into grammatical sentences, and their error patterns reflected influences from their respective L1s. The Chinese group, for instance, incorrectly produced both present and past simple forms in present perfect contexts. This pattern is consistent with production findings from previous studies (Erica & Wong, 2016; Liszka, 2004). The use of present forms by the Chinese group suggests that they were able to establish current relevance but not the past meaning associated with present perfect. In fact, current relevance meaning is encoded in their L1 through the perfect marker sentence-final *LE*. Their use of past simple forms does not signify the establishment of past meaning, and there is suspicion that they utilised the past simple form as a perfective marker. This may be attributed, at least in part, to an L1 influence related to the encoding of perfective aspect verbal-final LE, which is in clear contrast with the past tense meaning encoded in the present perfect.

On the other hand, the Arabic group showed a tendency to use past simple forms, as the current relevance associated with present perfect is inferred through context and adverbials in their L1. This has led to difficulty in encoding current relevance meaning grammatically. The findings were reflected in their gap-filling results, where the Arabic group performed significantly worse than the other two groups, consistently supplying past simple forms in present perfect context. More generally, the Arabic group had persistent difficulty in establishing form-meaning associations regarding English preterit and present perfect forms, as these two meanings are mapped onto one form in their L1.

In summary, the findings of the study are more aligned with the Representational Deficit Accounts. The online processing findings provide supportive evidence for the Morphological Congruency Hypothesis (MCH) (Jiang et al., 2011). Only congruent grammatical morphemes (those present in L1) are acquirable for adult L2 learners, while incongruent ones (not present in L1) pose extreme difficulty. Native-like competence in learning incongruent morphemes is unattainable, as learners will struggle to activate their meanings spontaneously. Much of the previous support for the MCH has come from the acquisition of L2 plural marking (Jiang et al., 2011, 2017), and the present study builds upon the existing evidence by extending its scope to include the L2 tense-aspect phenomenon. Essentially, the combined findings are in alignment with the Representation Deficit

Hypothesis (RDH) (Hawkins & Chan, 1997; Hawkins & Liszka, 2003). The L1 plays a significant role in L2 acquisition, and the absence of certain morphosyntactic properties in the L1 can lead to a potentially permanent deficit in the L2 mental representation of learners. This means that these learners may not be able to acquire these L2 properties after their critical period, resulting in a non-target-like mental grammar.

8.2 Online vs. Offline Performance

The results of the present study suggest that group differences were more noticeable in online tasks compared to offline tasks. Specifically, in comprehension, both L2 groups, like the native control group, demonstrated target-like judgment patterns in the offline AJT task. However, group disparities surfaced in the two online tasks due to L1 effects, wherein the Chinese group exhibited non-native-like processing patterns for past simple, and both L2 groups failed to display target-like performance in processing and detecting violations of present perfect sentences. In production, both L2 groups, like the English group, successfully produced correct tense-aspect forms in the relevant obligatory contexts, as indicated by the gap-filling task. The exception was the Arabic group, which performed significantly worse in supplying present perfect forms, arguably due to negative L1 influence. Similar group differences were once again observed in the online elicited imitation task. The Chinese group displayed a target-deviant direction of changes compared to the other two groups for past simple items, and both L2 groups exhibited a similar target-deviant direction for present perfect items. This discrepancy suggests that while L1 effects are not always evident in offline tasks, they are consistently observed in online tasks.

As reviewed in Section 4.2.2, online and offline tasks typically involve different types of knowledge (implicit vs. explicit knowledge). Implicit knowledge operates without conscious awareness, whereas explicit knowledge relies on some level of conscious awareness (R. Ellis, 2005; Rebuschat, 2013). When performing linguistic tasks, implicit knowledge can be utilised effortlessly, often without learners even being conscious of its presence; in contrast, explicit knowledge is often applied consciously and learners are aware of what they know (Godfroid et al., 2015). However, there is currently a lack of consensus regarding the type of knowledge employed by participants in a specific task (Cho, 2022). Nevertheless, some researchers argue that online processing tasks (e.g., eye-tracking, SPR) and specific task-related factors (e.g., time constraints) are more likely to measure participants' implicit knowledge (Jiang et al., 2011; Suzuki & DeKeyser, 2015). On the other hand, non-time-constrained tasks are believed to tap into explicit knowledge (R. Ellis, 2005). Under this assumption, in the present study, the three online tasks (i.e., the eye-tracking, sentence-matching, and imitation tasks) have measured participants' implicit knowledge, and the two offline tasks (i.e., the AJT and gap-filling tasks) have assessed their explicit knowledge.

Based on this distinction, the present results indicate that L1 transfer effect is more likely to be observed in implicit tasks than explicit tasks. This observation aligns with findings from previous studies that combined both online and offline tasks (e.g., Hopp, 2010; Roberts & Liszka, 2013, 2021; Vogel, 2017). For example, Roberts and Liszka's (2013) found that L1 German learners demonstrated target-like explicit knowledge of L2 English past simple and present perfect in the offline AJT task; however, they failed to show any sensitivity to aspectual violations in the online SPR tasks, presumably due to the absence of a grammaticalized aspect system in their L1. Similarly, Vogel (2017) observed that advanced L1 English L2 Spanish learners, whose L1 does not grammatically distinguish the perfective-imperfective contrast, performed similarly to the native Spanish group in offline cloze tasks and displayed explicit knowledge of the Spanish perfective-imperfective contrast.

Nevertheless, they did not exhibit sensitivity to violations in online SPR tasks, indicating that their implicit knowledge remained non-target-like due to L1 influence.

As noted by Hopp (2010), the failure to show target-like performance in the online tasks among L2 learners can be attributed to their reduced L2 processing resources and efficiency in general. There has been evidence of this in findings from neuroimaging research, which reveals a substantial overlap in the cortical areas involved in processing both L1 and L2 (Stowe & Sabourin, 2005). Also, L2 learners tend to exhibit greater and more extensive activation in these areas compared to native speakers, suggesting that L2 processing might be less efficient than L1 processing (Stowe & Sabourin, 2005). This speculation is supported by a study conducted by Roberts et al. (2008). In their study, L1 German L2 Dutch learners demonstrated native-like performance in resolving ambiguous subject pronouns in the offline AJT task, unlike their L1 Turkish counterparts, who exhibited non-target-like performance. The observed difference is attributed to L1 transfer, as German shares similarities with Dutch as a non-null subject language, while Turkish is a null subject language. Interestingly, the online eye-tracking task revealed that both learner groups performed poorly compared to the native Dutch group, suggesting that a general L2 processing limitations in online tasks may override transfer effects, as concluded by the authors. However, such a general processing disadvantage cannot account for the results in the present study, as native-like processing was observed in L2 groups for L1-L2 similar structures, and non-native-like processing was found for structures with predicted influence from L1.

The observed L1 influences in the implicit tasks can be better understood by considering fundamental differences in L1 and L2 acquisition. In L1 acquisition, children typically demonstrate better performance in online tasks involving implicit knowledge compared to offline tasks requiring explicit knowledge (e.g., pronoun acquisition in Sekerina et al., 2004). This is because L1 grammar is acquired implicitly through natural exposure to language input rather than reliance on explicit rules (N. C. Ellis, 2008). In contrast, L2 acquisition can be different, as implicit learning from natural linguistic exposure is often limited compared to L1 acquisition, and explicit learning resources are essential in the learning process (N. C. Ellis, 2008). More importantly, the challenge lies in applying implicit learning processes, as L2 learners' implicit representations are already adapted to the structures and patterns of their L1, resulting in an entrenched configuration (N. C. Ellis, 2008). Such L1-tuned mechanism can potentially interfere with the optimal processing of the L2 particularly in tasks requiring the automatic activation of grammatical meanings, as described under the MCH (Jiang et al., 2011). This view is well-established in psycholinguistics, where it is recognised that when individuals learn their L1, they also internalise which meanings are encoded grammatically and should therefore be automatically activated during language processing (Jiang et al., 2017). As Levelt (1989) explicitly states, "languages vary in the semantic features that are grammatically acknowledged. Consequently, the encoding of messages differs among speakers of different languages" (p. 106). These cross-linguistic differences can significantly influence the meaning activation in L2 processing. For L2 learners, processing congruent morphemes involves linking morphological markers with related meanings already activated due to their L1. However, processing incongruent morphemes presents a challenge. Here, learners must automatically activate a meaning that is not typically activated in their L1. This process is where difficulty arises, as it requires establishing new connections and adjusting to unfamiliar linguistic patterns (Jiang et al., 2017). This difficulty in activate grammatical meaning might explain why L1 transfer effects manifest more prominently in implicit tasks, as explicit tasks that lack time pressure and information about moment-by-moment processing are less likely to assess learners' automatised or implicit knowledge.

However, this is not to suggest that transfer effects cannot be observed in offline explicit tasks at all. In fact, previous studies have consistently shown their presence in explicit tasks (e.g., Alruwaili, 2014; Gabriele et al., 2005; Izquierdo & Collins, 2008), and in the present study, the Arabic group's poor performance in the present perfect context during the offline gap-filling task can also be attributed to L1 influence. This simply suggests that the persistence or strength of L1 influence can be strong enough to affect L2 performance in explicit tasks. While overcoming this influence is possible, especially for learners engaged in explicit tasks and those who have undergone years of formal classroom learning, it may not always be easily achieved.

Nevertheless, the observation of L1 effects in implicit tasks is not always consistent. Some previous studies have noticed the L1 transfer effect more prominently in offline explicit tasks rather than in online implicit tasks (Cho, 2022; Ionin et al., 2021; M. H. Kim & Ionin, 2022; Tokowicz & MacWhinney, 2005; Zufferey et al., 2015). For instance, Zufferey et al. (2015) conducted a study with L1 Dutch and French learners' comprehension of discourse connectives (i.e., *if* and *when*) in L2 English. They used both an offline AJT task and an online eye-tracking reading task. The results revealed that both L2 groups performed similarly to the native control group and showed sensitivity to inappropriate usages in the online task. However, in the offline AJT task, an L1 transfer effect was observed. The French group performed poorly on the uses of *if* and the Dutch group struggled with the uses of *when* due to ambiguity of their respective L1 connectives. Kim and Ionin (2022) employed an online elicited imitation task and an offline force-choice task to investigate the knowledge of L2 English definite article *the* among L1 Korean learners. The results showed that, while in the imitation task, Korean learners exhibited similar patterns of performance as native speakers, differences emerged in the force-choice task where they underused *the* in the bridging context (the uniqueness is inferred other than previous mention) due to a lack of articles in their L1.

One explanation for the different patterns observed in these studies is that L1 transfer might affect learners' performance in explicit tasks more. However, this is unlikely, as argued earlier, because L1 transfer is commonly viewed in L2 acquisition models as an implicit process influencing learners' underlying knowledge (Schwartz & Sprouse, 1996). Thus, implicit tasks that assess learners' automatic activation of grammatical meanings should be more susceptible to L1 effects. Another explanation proposed by Cho (2022) is related to task effects: in offline explicit tasks, the cognitive load associated with metalinguistic reasoning may be more taxing when making conscious judgments of sentences, which potentially has a negative impact on L2 performance (Hopp, 2009; Orfitelli & Polinsky, 2017). This is also doubtful, as many studies have found that online implicit tasks can indeed impose a higher processing load compared to explicit tasks (e.g., Hopp, 2010). If learners' non-target-like performance in explicit tasks is due to processing load, similar performance should be observed in online processing tasks. A third explanation is based on R. Ellis's (2006) argument that the development of implicit and explicit knowledge among L2 learners does not follow a synchronous pattern. Implicit knowledge tends to develop earlier than explicit knowledge for some syntactic structures, but the reverse is true for others (R. Ellis, 2006). For the acquisition of L2 articles and discourse connectives, implicit knowledge might emerge before explicit knowledge (Cho, 2022; Zufferey et al., 2015). If this were the case, for L2 tense-aspect acquisition, explicit knowledge seems to emerge earlier than implicit knowledge. Yet another question remains: can explicit knowledge be transformed into implicit knowledge? In fact, this relationship is widely debated as the interface issue in L2 acquisition (R. Ellis, 2005; Hulstijn, 2015). The current study suggests that explicit knowledge can become implicit knowledge, particularly when positive L1 transfer is present.

In such cases, learners can activate grammatical meanings automatically in real-time processing, as these meanings are already activated in their L1.

Overall, the observed discrepancy in performance between these two types of tasks underscores the need for a comprehensive approach in assessing L2 knowledge (Cho, 2022; Roberts & Liszka, 2013). To address theoretical questions about whether L2 learners can form target-like underlying representations and the role of L1 in this process, future studies should employ both online and offline tasks within a single study .

8.3 Comprehension vs. Production

The present findings also suggest an asymmetric relationship between learners' comprehension and production. On both implicit and explicit levels, it was found that learners generally perform better in comprehension tasks than in production tasks. It is important to note that a direct statistical comparison of their performance in the two modalities is not possible due to different outcome variables and materials used in comprehension and production tasks. Therefore, only an indirect comparison can be made by assessing how the L2 groups perform relative to the native control group.

For present progressive items, native-like processing behaviour and certain grammaticality effects were observed in both L2 groups in the online comprehension tasks. In contrast, their oral production performance was significantly worse than the native group, despite exhibiting a target-like pattern of changes. Explicitly, the learners displayed native-like patterns in the AJT task (comprehension) and gap-filling task (written production). Thus, no differences were found in offline comprehension and production performance (explicit level), but oral production appears to lag behind online comprehension (implicit level). However, target-like comprehension is not always evident when the L1 influences come into play for the other two structures. For example, the Chinese group performed poorly in both online comprehension and oral production tasks compared to the English group. Still, they exhibited native-like performance in the offline AJT and written production tasks. The Arabic group showed similar performance as the Chinese group for present perfect structures, but they demonstrated target-deviant patterns in the gap-filling task, in contrast to their target-like performance in the AJT task.

These findings suggest that, for L2 learners, the development of comprehension tends to precede that of production particularly at the implicit level. This general tendency aligns with observations in language acquisition studies (Hendriks & Koster, 2010; Swain, 1985). It is also consistent with some previous L2 empirical research that has explicitly discussed the relationship between comprehension and production (e.g., García-Tejada et al., 2023; Trenkic et al., 2014). This relationship appears logical. As argued by Clark (1993), for effective meaning production, learners must initially comprehend and map meaning to specific linguistic forms, storing these associations in memory for future use. Nevertheless, some studies have reported a reversed asymmetry in the L2 acquisition, particularly in cases involving gender and number agreement (Montrul et al., 2008; Tasseva-Kurktchieva, 2015). These instances suggest that, in certain linguistic features, the development of production may outpace comprehension in L2 learners.

Furthermore, it is worth noting that the learners' oral production performance was far less native-like compared to their performance in online comprehension. Even when native-like processing patterns were observed in online comprehension, their oral production was considerably less accurate in comparison to the native group. While they were able to exhibit a target-like direction of changes, they corrected a significantly lower number of ungrammatical sentences and were more prone to producing errors when repeating grammatically correct sentences, as opposed to the native group. Apart from the explanation that comprehension precedes production, another factor contributing to these differences is the fundamental differences between comprehension and production. Comprehending language involves decoding, where the input is matched against its internal language rules. Oral production, on the other hand, is more complex. It goes beyond decoding, requiring tasks such as making lexical choices, assigning syntactic representations, converting mental representation to the phonological form for articulation, and monitoring speech etc. (Levelt et al., 1999). While L1 speakers generally experience automatic speech formulation, L2 speakers might need to engage in serial processing when it comes to lexical retrieval and encoding.

8.4 Implications of Findings

8.4.1 Methodological Implications

As mentioned in Section 5.4, the adoption of sentence-matching (SM) task in the present study aimed to gain our understandings of this method. The results were unexpected. Specifically, no significant effect was observed in the L2 groups for all three structures, and in the English group, no significant effect was found for the present perfect structure. Although significant effects were identified in the English group for past simple and present progressive structures, the effect sizes were extremely small based on Plonsky and Oswald's (2014) L2 field-specific criteria. This raises the question of whether the SM task is effective in examining underlying grammaticality among L2 learners, particularly when the detected effect in the native group was small.

Indeed, previous studies have reported null results for certain grammatical structures in both native speakers (e.g., Freedman & Forster, 1985; Hubers et al., 2020) and L2 learners (e.g., Duffield & White, 1999; Gass, 2001; Verhagen, 2011). In the present study, the English group did take longer to respond to ungrammatical pairs than grammatical pairs and the RT differences were in the predicted directions. It is suspected that the lack of significance for present perfect structures occurred because the English group took much longer to match the present perfect grammatical pairs compared to the other grammatical pairs. This delay led to RT differences that were too small to reach significance. Their RTs might be inhibited by the less common usage of present perfect forms accompanied by temporal adverbial in real-life contexts (e.g., *since last year*), as indicated by corpus evidence (e.g., Werner, 2014).

The L2 groups demonstrated target-like RT directions for past simple and present progressive items, but the effects were not statistically significant. Following Duffield et al. (2002), the lack of significance might have two interpretations. On one hand, it could suggest that the L2 groups failed to distinguish between grammatical and ungrammatical sentences, showing no clear grammaticality effect. On the other hand, it might indicate that learners exhibited tendencies in the right direction, but the effect was too subtle to reach statistical significance. The difficulty in differentiating between these two interpretations could be considered a limitation of the SM methodology's effectiveness in L2 studies.

In the present study, we opted for the second interpretation for two reasons. Firstly, both L2 groups displayed target-like RT directions, which differed from the reversed RT directions observed for present perfect items. These differences might imply that the L2 groups demonstrated a certain level of grammaticality effect for past simple and present progressive structures compared to their performance on present perfect items. Secondly, the task design feature might have imposed additional cognitive demands on L2 learners compared to native speakers. In this task, to prevent L2 learners from relying on word-to-word matching, the first sentence of each pair disappeared before the onset of the second sentence, rather than displaying both sentences simultaneously on the screen. This design choice could potentially introduce a new variable, i.e., working memory (Jiang, 2012), as learners were required to store information about the first sentence in short-time memory while processing the second sentence. Particularly, reading-span scores in L2 learners are found to be lower compared to those in L1 speakers (Reichle et al., 2016), and learners' working memory capacity is positively associated with the outcomes of L2 processing tasks (Linck et al., 2014). However, the present study did not include a working memory test to further investigate this possibility.

In addition, the SM task is unable to provide more detailed processing information, as it measures processing time only after participants have fully read the sentences. This differs from other reaction-time tasks, such as the self-paced reading task, where processing is measured incrementally right from the beginning of the reading process and on a word-by-word basis. The difference in methodology may account for the contradictory results between our study and that of Roberts and Liszka (2013), who tested similar tense-aspect mismatch violations. In fact, a recent study conducted by Hubers et al. (2020) provided evidence for this speculation. Their research focused on grammatical norm violations that defy prescriptive grammar rules but are commonly used by native speakers. Using a sentence-matching task, they examined the sensitivity of native German and Dutch speakers to such violations in comparative constructions. The results suggested no significant differences in RTs between sentences containing grammatical norm violations and fully ungrammatical sentences. However, a follow-up eye-tracking reading experiment showed that fully ungrammatical construction led to longer RTs compared to the grammatical norm violation, suggesting processing differences between the two structures. Therefore, the authors concluded that the eye-tracking task, being more time-sensitive, was able to detect differences in processing that the SM task could not.

In summary, the sentence-matching task shows potential as a practical method for assessing learners' underlying grammatical knowledge due to its simplicity in design and ease of data analysis. However, its application in L2 research requires further validation. For example, the null results can be difficult to interpret in some cases. This is particularly true for ungrammatical structures that might not induce a big processing cost. In such cases, employing more sensitive processing measures, such as SPR and eye-tracking tasks, might be preferable, as they can offer detailed word-to-word processing information.

8.4.2 Theoretical Implications

Contributions to the L2 Tense-Aspect Research. The acquisition of tense-aspect has been widely explored in the field of SLA. For example, numerous previous studies have investigated the influence of universal lexical aspect on the development of L2 tense-aspect morphology (see Bardovi-Harlig & Comajoan-Colomé, 2020 for a review). However, the role of a learner's L1 in shaping their acquisition of tense and aspect has been largely overlooked and remains an open question (Li & Shirai, 2000; Slabakova, 2002). As argued by Shirai (2009), there is a pressing need to systematically investigate cross-linguistic influences in L2 tense-aspect research and to disentangle these from the natural acquisition process guided by the lexical aspect hypothesis. The present study contributes to this line of research by examining the potential influence of L1 on L2 acquisition of tense-aspect among learners from two different L1 backgrounds, which can better isolate L1 effects from the general difficulties associated with L2 learning.

Methodologically speaking, previous research on L2 tense-aspect, investigating the role of L1, has primarily utilised oral production and offline judgment/interpretation tasks (e.g., Alruwaili, 2014; Al-Thubaiti, 2015; Amaro et al., 2018; Ayoun & Salaberry, 2008; Collins, 2004; Diaubalick & Guijarro Fuentes, 2017; Domínguez et al., 2017; Gabriele et al., 2005; Hawkins & Liszka, 2003; Izquierdo & Collins, 2008; Liszka, 2009, 2015; Muroya, 2019; Yang & Huang, 2004). While these studies have yielded mixed evidence, they have provided many insights into learners' production and offline comprehension performance at both the early and late stages of acquisition. However, here remains a gap in our understanding of learners' real-time comprehension performance. It is acknowledged that L2 acquisition requires not only accumulating L2 knowledge but also effectively applying that knowledge in real-time language processing (Roberts, 2012). Although the quantity of existing L2 literature examining tense-aspect processing is growing (e.g., Chan, 2012; Kahoul, 2014; Roberts & Liszka, 2013, 2021; Vogel, 2017; Yao & Chen, 2017), it still remains small compared to processing research on other linguistic phenomena (Rastelli, 2020). Thus, the investigation of learners’ processing performance in the present study can contribute to addressing this research gap.

The findings of the present study suggest a persistent L1 influence that not only restricts learners’ production but also extends to their online comprehension. This finding aligns with a few previous studies on L2 tense-aspect processing (H. L. Chan, 2012; Kahoul, 2014; Roberts & Liszka, 2013). The evidence indicates that the acquisition of English tense-aspect by Chinese and Arabic learners is heavily influenced by whether the relevant form is grammaticalised in their L1. It is important to note that such L1 influences were consistently observed in the online tasks, and both L2 groups performed in a native-like manner in the offline tasks (except for the Arabic group's performance in the present perfect context in the gap-filling task). It is not surprising that these learners could exhibit target-like performance in offline tasks, given their near-advanced proficiency and the fact that they were instructed learners who could utilise their acquired explicit knowledge during such tasks. However, this raises doubts about the results of previous studies that examined L1 influences in a similar population using only offline tasks. Native-like performance in offline tasks does not necessarily indicate that learners have fully acquired the structure, as they might still struggle to utilise and activate its meaning in real-time comprehension (Roberts & Liszka, 2013). Future research examining L1 transfer effects should incorporate both implicit and explicit tasks to better measure learners’ knowledge.

Theoretically, the findings are more aligned with the Representation Deficit Accounts. In the acquisition of L2 tense-aspect, both comprehension and production knowledge of learners can be influenced by their L1 grammar. This influence might be permanent, and L2 learners might not be able to attain native-like representations if the relevant features are absent from their L1s. However, since the learners in this study were at an upper-intermediate/near-advanced proficiency level, the question of whether native-like representations are eventually acquirable remains unanswered. Nevertheless, it is suspected that the influence of L1 could persist for a long period.

Overall, the results do not support the Full Access accounts, which argue that learners can acquire L2 grammatical features that are not instantiated in their L1, and any observed difficulty in their usage can be attributed to issues related to the production or realisation of surface forms. If this were the case, no differences would be observed between the L2 groups, and target-like performance should always be evident in the processing task since they are assumed to have underlying representations.

Contributions to the L2 Processing Research. One major hypothesis in L2 processing research is the Shallow Structure Hypothesis (Clahsen & Felser, 2006), which argues that "the syntactic representations adult L2 learners compute for comprehension are shallower and less detailed than those of native speakers" (p. 32). In other words, L2 learners rely less on syntactic information compared to native speakers. Instead, they construct sentence representations mostly based on lexical, semantic, and pragmatic information. This hypothesis attributes the reduced utilisation of grammatical information to age-related constraints (Clahsen & Felser, 2006). Some previous studies have supported this hypothesis and found that L2 learners are much slower in processing and less sensitive to morphological information compared to native speakers (e.g, Clahsen, Felser, et al., 2010; Lew-Williams & Fernald, 2007). However, the present study provides contradictory evidence and indicates that L2 learners were able to utilise morphological cues to encode event information in online sentence comprehension. Native-like processing patterns and speed were observed in both L2 groups for the progressive morpheme and also in the Arabic group for the past tense morpheme. These results contribute to the growing body of evidence suggesting that native-like morphosyntactic processing is achievable among adult L2 learners (e.g., Hopp, 2006; Tokowicz & MacWhinney, 2005; Tokowicz & Warren, 2010; Trenkic et al., 2014; Yuan, 2017). These instances collectively show that the phenomenon of shallow processing in L2 learners is not absolute.

The existing body of research on the role of L1 in L2 morphosyntactic processing has yielded mixed evidence. Some studies have observed the influence of learners' L1 on processing (H. L. Chan, 2012; L. Chen et al., 2007; Jiang, 2004; Roberts & Liszka, 2013, 2021; Tokowicz & MacWhinney, 2005; Wang et al., 2022). On the other hand, other studies have found no such influence (Foote, 2011; Roberts et al., 2008; Silva & Clahsen, 2008; Yao & Chen, 2017). The current study aligns with the perspective of persistent L1 influence. Specifically, the Chinese group was significantly slower than the other two groups in processing past tense, attributable to the absence of grammaticalised tense in their L1. Also, both groups exhibited non-native-like processing patterns of present perfect and demonstrated no sensitivity to its violations, due to the fact that the perfect meaning is encoded differently in their respective L1s. In contrast, the native-like processing of the progressive aspect in both groups can be linked to the grammatical marking of progressive meaning in their L1 (Chinese) or its ongoing grammaticalization (Arabic). These processing findings lend support to the Morphological Congruency Hypothesis (MCH) (Jiang et al., 2011). Native-like competence is attainable for congruent grammatical morphemes, while it is extremely difficult for incongruent morphemes. The difficulty arises from the need to first treat the related meaning as grammaticalised, which entails the automatic activation and representation of this meaning during language processing (Jiang et al., 2011). For L2 learners, this task can be extremely challenging.

Lastly, the present results suggest a potential influence of experimental paradigms on processing outcomes. Many prior studies in language processing have primarily employed a violation detection paradigm, relying on the use of ungrammatical sentences to examine learners' sensitivity to errors compared to the native speakers (e.g., Chan, 2012; Roberts & Liszka, 2013; Vogel, 2017; Yao & Chen, 2017). For instance, Yao and Chen (2017) constructed sentences with past tense violations (i.e., omission) to test if L1 Chinese L2 English learners with high proficiency were sensitive to errors in both SPR and eye-tracking tasks. Their results revealed that these learners demonstrated target-like sensitivity, leading the authors to conclude that Chinese learners can eventually attain native-like competence, even if past tense is absent in their L1. However, the present study found that Chinese learners could not utilise past tense information as efficiently as their Arabic counterparts in the eye-tracking task, despite demonstrating a certain level of sensitivity to ungrammaticalities in the sentence-matching task. These conflicting results suggest that detecting errors in grammatically incorrect sentences can be very different from utilising grammatical information facilitatively in processing fully grammatical sentences (Trenkic et al., 2014). From this perspective, the non-violation paradigm can be advantageous, as it allows researchers to assess whether a grammatical feature genuinely facilitates language processing, rather than merely indicating a response to a grammatical violation (Vainio et al., 2016).

Chapter 9 Conclusion

9.1 Summary of the Findings

The present study investigated L2 English learners' online processing, production, and explicit knowledge of tense-aspect structures. A total of 72 participants were recruited for the study, including 24 Chinese learners of English, 24 Arabic learners of English, 24 native British English speakers as the control group. The English proficiency of L2 participants were measured through the Oxford Quick Placement test scores, IELTS (or other equivalent) tests results, and self-reported proficiency scores. The learners were categorised as being at an upper-intermediate/near-advanced level. The regression results revealed no significant differences between the L2 groups in their QPT scores, IELTS scores, and self-reported proficiency scores (all *p*s > .05). Similarly, no significant differences were found in the age of L2 English acquisition, number of years of formal instruction, and time of residence in the UK (all *p*s > .05).

The experimental design included five tasks that crossed both comprehension/production and online/offline domains. For online processing, a visual-world eye-tracking paradigm was employed to investigate whether learners could effectively use morphological information to achieve target event interpretations. A sentence-matching task tested learners' sensitivity to tense-aspect violations. The study used an elicited imitation task to examine learners' oral production. Two additional tasks were also employed to assess learners' explicit knowledge, with an acceptability judgement task measuring learners' offline comprehension and a gap-filling task evaluating learners' written production.

The results of the eye-tracking task show that both L2 groups could use morphological markers to encode event information and achieve target interpretations. However, differences were observed in their processing speed. The Chinese group was significantly slower in processing past tense marking (*v-ed)* compared to their Arabic counterparts, whereas the Arabic groups exhibited native-like processing pattern. Both L2 groups demonstrated a significantly delayed divergence point in processing present perfect morpheme (*have+v-ed*) compared to the English group. However, they exhibited native-like processing patterns for progressive morpheme (*v-ing*), and the Arabic group even outperformed the English group.

The results of the sentence-matching task produced unexpected findings. The English group displayed predicted RT directions, with longer RTs for ungrammatical pairs compared to grammatical pairs, across all three structures. However, the RT differences were only significant for past simple and present progressive items. The lack of significance in present perfect items may be attributed to the inhibition of the grammatical items (e.g., *For the past year, my two brothers have studied…*). Based on corpus evidence, this inhibition could be linked to the less common association of present forms with temporal adverbials and the ongoing influence of American English, where past simple is used to convey present perfect meaning. This influence was mirrored in the English groups' AJT ratings, indicating a higher level of acceptance for such usage. The RT difference in L2 groups were not significant across all three structures. In contrast, the RT differences in L2 groups were not statistically significant for all three structures. Nevertheless, both L2 groups demonstrated longer RTs when evaluating ungrammatical pairs for past simple and present progressive items, suggesting a target-like RT direction. The lack of significance may be attributed to a weak effect induced by certain task design features (see Section 8.3). It is thus interpreted that both L2 groups exhibited a certain level of grammaticality effect for past simple and present progressive structures, in contrast to the absence of a grammaticality effect observed for present perfect structures, as evidenced by the reversed RT directions. Contrary to the prediction, the Arabic group did not perform better than the Chinese group with the past simple items. It is suggested that the Arabic group experienced a negative L1 influence, as their L1 maps both preterit and present perfect meanings onto a single form. The differentiation between perfect and preterit meanings relies on contexts and adverbials. Consequently, the Arabic group may have tended to over-accept past simple forms in present perfect contexts, a trend also reflected in their AJT results, where they exhibited a higher acceptance of such usage compared to the Chinese group.

The results of the elicited imitation task indicate a consistent preference among the English group for correcting ungrammatical sentences rather than introducing errors into grammatical sentences across all three structures in their imitations. This preference serves as evidence supporting the existence of relevant grammatical knowledge. Both L2 groups displayed similar target-like preferences when imitating present progressive items. However, a contrasting pattern emerged when imitating present perfect items, suggesting a deficiency in their underlying knowledge. In addition, different patterns were observed for past simple items, where the Arabic group aligned with the English group, while the Chinese group exhibited a reversed tendency.

The results of the two offline tasks largely align together. In the AJT task, both L2 groups, like the English group, rated the ungrammatical sentences significantly lower than grammatical ones. This implies that they had explicit knowledge of correct and incorrect tense-aspect usage. Similarly, in the gap-filling task, both L2 groups performed similarly to the English group in supplying past simple and present perfect forms in the obligatory context. However, the Arabic group performed worse than the other two groups in the present perfect context, and they tended to produce past simple forms instead.

The combined findings indicate a strong influence of the L1 on learners' performance in both comprehension and production, particularly in the implicit tasks. The Arabic group exhibited an L1-based advantage over their Chinese counterparts in processing and producing the past tense. This advantage arises because Arabic, like English, grammatically encodes

[+past], whereas the Chinese learners' L1 lacks such grammaticalization. Neither L2 group demonstrated target-like performance for the present perfect. In Arabic, the perfect and preterit meanings are mapped onto one single form, with the two meanings distinguished through lexical means (e.g., temporal adverbials). This has arguably led Arabic learners to struggle with acquiring both semantic meanings of the present perfect—past and current relevance. They tended to encode only the past meaning, evident in their overproduction of the past simple in a present perfect context. The Chinese group, despite their L1 encoding [+perfect] through a current relevance marker, faced challenges in associating past meaning with the English present perfect due to the absence of grammaticalized tense in their language. Finally, both L2 groups demonstrated target-like performance in processing and producing the present progressive. This success can be attributed to the fact that both of their L1s grammaticize [+progressive] meaning or it is in the process of becoming more grammaticalized.

9.2 Limitations and Future Research

The present study has several limitations. Firstly, the proficiency levels and sample sizes within the L2 groups might pose potential problems. At the time of data collection, all L2 participants were enrolled in a UK university, and they had all fulfilled the language requirements for admission into their study programmes. Despite no significant differences being observed between their IELTS test results and self-reported proficiency scores, the Arabic group achieved a higher average score (M = 48.8) in the QPT test compared to the Chinese group (M = 44.8). While no statistical differences emerged in their QPT test results, it should be noted that the obtained p-value (p = .053) approached significance. This should be considered when interpreting the results. In addition, the final analysis included only 21 participants in each L2 group. While such sample sizes are very common in processing research involving adult L2 learners, it is essential to acknowledge that the sample size, in this case, was still relatively small. As suggested by Keating and Jegerski (2015), a minimum of 24 participants per group is recommended for sentence processing studies when testing experimental items involving two conditions (e.g., grammatical and ungrammatical sentences).

The second limitation relates to the experimental materials used in the eye-tracking task. When constructing sentence stimuli, telic verbs (i.e., accomplishments and achievements) were mostly used, as the different stages of the action (in the process and the end-result stage) could be more easily portrayed through images compared to atelic verbs (i.e., states and activities). The inherent verb semantics might have an influence on learners' performance. In addition, for each visual image, two pictures were created. One depicted the ongoing process of an event, while the other one illustrated the end-result stage of the event. Due to this limitation, the present perfect sentences focused mainly on resultative constructions, among the total four types of perfect.

The third limitation is the experimental manipulation of present progressive items in the sentence-matching and elicited imitation tasks. When selecting adverbials that clash with the present progressive, temporal adverbials, such as lately, recently, and every week, were used. This selection follows Chan's (2012) study, as these adverbs can disrupt the ongoing interpretation in sentences like "**Lately, my grandmother is teaching me things about growing plants*". This creates a clear mismatch or violation, which can be used to test participants' sensitivity to the incorrect usage of the present progressive. However, these adverbials signal past actions that often extend into the present, inadvertently testing tense violations rather than purely aspectual violations, because they can also collocate with past tense verbs. In this context, adverbials (e.g., every week) that indicate present simple seem to be more appropriate, as in English, the progressive directly contrasts with verbs unmarked for aspect, which are interpreted as describing habitual or generic information (Comrie, 1976). Even though violations are created, it should be borne in mind that this limitation in the stimuli design.

Another potential limitation is related to the coding of elicited imitation data. The imitative responses of participants were transcribed by the researcher, who is an L2 English speaker. Next, a native British English speaker re-transcribed the responses involving the past simple and present perfect tenses. This step was taken because the presence or absence of words-final *-t/-d* might not be very discernible to non-native ears. Any discrepancies in the transcriptions were thoroughly discussed and resolved. Nevertheless, the coding process for the qualitative analysis was conducted solely by the researcher. As suggested by Révész (2011), when coding oral production data, both intra-coder and inter-coder reliability should be considered. In the present study, the researcher coded the data twice to ensure consistency in assigning error types. However, it is important to note that the inter-coder reliability was not checked.

A further limitation concerns the overall duration of the experimental session. The L2 participants were required to complete five individual tasks and an English test, which could extend the session to 2 hours. Although participants were allowed to take breaks between tasks, some of them still showed signs of fatigue when performing the explicit tasks towards the end of the session. It is possible that the fatigue might have had a potential impact on the participants' performance.

Future research on L1 influences can consider employing online processing methods to examine the acquisition of structures in less-studied L2 languages, in addition to English. This approach would help to widely attest the phenomenon of L1 transfer across various grammatical structures in different L2 languages. Moreover, it is crucial to utilise both online and offline tasks in future research when investigating cross-linguistic influences. Analysing learners' performance in both implicit and explicit tasks can offer a more comprehensive understanding of the acquisition process. Future research should also explore further about the differences between violation-detection and non-violation paradigms. For example, exploring why variations arise among learners when exposed to different paradigms, such as their sensitivity to grammatical violations but their inability to employ grammatical information facilitatively when processing correct sentences, or vice versa, can be a direction for future investigation. In addition, it should be noted that the L2 learners in the present study were at near-advanced proficiency levels, which poses challenges in determining whether the observed negative L1 influences can eventually be overcome. Thus, future research might consider including end-state/near-native learners with extensive immersion experiences, which can provide valuable insights into the persistence of L1 influences over time.

Appendices

Appendix 1: The Consent Form

Information

The research is part of a PhD study and conducted in the Department of Education at the University of York. The purpose of this project is to investigate the comprehension and production of English by second language learners. Firstly, you will need to fill in a language background questionnaire. Then, you will be asked to complete three comprehension tasks and two production tasks. A detailed description about each task will be given, and a practice session will need to be completed before performing the task. The experiment takes around 1 to 2 hours to complete, and you can take short breaks between each task if you want to.

Your participation in this study is voluntary, and you are free to withdraw at any point during the experiment without giving reasons. All the information you provide will remain confidential. A random code will be assigned to you once you decide to participate, and it will be the only form of identification in the database. The code associating your name and your data will be kept in a separate file with strong password protection, and only the researcher will have access to it. After you have completed the study, you can still contact the researcher to withdraw your data within two weeks. After two weeks, your data will be fully anonymised and it will be impossible to delete it.

No identifiable information will be included when using the data publicly (e.g., in presentations and online). All the data may be kept indefinitely and could be used for future analysis or in other research projects in anonymous format.

If you have any questions or concerns about the study, please contact the researcher, Yu Liu, at yl4674@york.ac.uk, or the supervisor of this project, Professor Leah Roberts, at leah.roberts@york.ac.uk, or the Education Ethics Committee, Professor Carole Torgerson, at education-research-admin@york.ac.uk.

By signing your name below, you are giving your consent to participate in this study, and you have read all the information provided for you.

Name of participant

Signature of participant

Date

Appendix 2: The Language Background Questionnaires

Language Background Questionnaire (L2 groups)

This section is to be filled out by the researcher: Date: _______________ Participant code: _______________

Part 1: Biographic information

1. Age: ____________
2. Sex: □male □female
3. Academic level: □Undergraduate □Postgraduate
4. Your native language(s): ____________________
5. Other foreign languages you have known other than English:

   Language: _____________ Proficiency level: _____________

   Language: _____________ Proficiency level: _____________
6. Do you have any hearing or visual problems? □YES □NO

   If your answer is YES, please specify: ____

Part 2: English learning and use

7. At what age did you start learning English?

   ____
8. How many years of formal English instruction have you received?

   Number of years: _____________
9. How long have you been living in the UK?

   Number of months: _____________
10. Have you ever lived or studied in another English-speaking country other than the UK?

    □YES □NO

If your answer is YES, please list the country you have lived and for how long you have lived there.

Country: ____________ Number of months: ____________

Country: ____________ Number of months: ____________

Country: ____________ Number of months: ____________

11. What was your most recent IELTS test score?

____

If you do not have an IELTS score but have taken another standardised English test (e.g., TOEFL, Duolingo), please specify the name of the test and the score you have received.

Name: ____________ Score: ____________

Name: ____________ Score: ____________

12. How would you rate each of your English skills on a scale from 1 to 6 (please tick the box)?

	1 Very low	2 Low	3 Intermediate	4 Advanced	5 Near-native	6 Native
Listening	☐	☐	☐	☐	☐	☐
Reading	☐	☐	☐	☐	☐	☐
Speaking	☐	☐	☐	☐	☐	☐
Writing	☐	☐	☐	☐	☐	☐

Language Background Questionnaire (English group)

This section is to be filled out by the researcher:
Date: _______________ Participant code: _______________

Biographic information

13. Age: ____________

14. Sex: □male □female

15. Academic level: □Undergraduate □Postgraduate

16. Your native language(s): ____________________

17. Other foreign languages you have known:

    Language: _____________ Proficiency level: _____________

    Language: _____________ Proficiency level: _____________

    Language: _____________ Proficiency level: _____________

18. Do you have any hearing or visual problems? □YES □NO

    If your answer is YES, please specify: ____

Appendix 3: The Eye-tracking Task Materials

Example images:

Spoken sentence: The little girl is washing/washed/has washed her hands with the soap after playing.

All the other images have been unloaded on OSF: DOI 10.17605/OSF.IO/VQ32D

Spoken sentences:

1. The boy and girl are planting/planted/have planted a small tree together.

2. The old man is planting/planted/have planted a small apple tree in his garden.

3. The cute dog is eating/ate/has eaten his food on the floor.

4. The cute girl is eating/ate/has eaten an apple before her breakfast.

5. The young man is repairing/repaired/has repaired the broken air conditioner.

6. The male worker is repairing/repaired/has repaired the green car on his own.

7. The boy student is writing/wrote/has written his name on the blackboard.

8. The physics teacher is writing/wrote/has written a formula on the board.

9. The boss is writing/wrote/has written his new plan for the company on the board.

10. The two farmers are picking/picked/have picked the ripe apples off the tree.

11. The young woman is picking/picked/have picked the heavy box up on her own.

12. The husband is painting/painted/has painted the wall of the living room all blue.

13. The artist is painting/painted/has painted a vase for himself on the board.

14. The little girl is painting/painted/has painted a goldfish for herself on the board.

15. The big elephant is drinking/drank/has drunk the water in the little pond.

16. The young boy is drinking/drank/has drunk a glass of water after exercise.

17. The doctor is pushing/pushed/has pushed the patient’s wheelchair up the ramp.

18. The young man is pushing/pushed/has pushed the closet against the wall.

19. The two brothers are cleaning/cleaned/have cleaned the living room together.

20. The little boy is cleaning/cleaned/has cleaned his desk after using it.

21. The father and the son are washing/washed/have washed the blue car outside.

22. The little girl is washing/washed/has washed her hands with the soap after playing.

23. The two kids are washing/washed/have washed the dishes in the kitchen.

24. The father and the son are building/built/have built a big snowman in the park.

25. The boy and the girl are building/built/have built a house with the blocks.

26. The little girl is making/made/has made the bed by herself after getting up.

27. The young man is making/made/has made a wooden chair in the workshop.

28. The mother and the daughter are making/made/has made a strawberry cake together.

29. The little girl is hanging/hung/has hung her pink jacket on the hook.

30. The young artist is hanging/hung/has hung a new painting on the wall.

31. The young man is pouring/poured/has poured milk into his coffee.

32. The bartender is pouring/poured/has poured himself a glass of wine after work.

33. The little boy is watering/watered/has watered the flowers in the garden.

34. The old lady is sweeping/swept/has swept the yellow leaves in her garden.

35. The little girl is brushing/brushed/has brushed her teeth before sleep.

36. The two kids are crossing/crossed/have crossed the street by themselves.

37. The photographer is taking/took/has taken a picture of the young couple.

38. The male worker is fixing/fixed/has fixed the bicycle in the workshop.

39. The little kid is filling/filled/has filled the vase with the tap water.

40. The two men are moving/moved/ have moved the heavy sofa downstairs.

41. The businessman is blowing/blew/has blown the candles out on the cake.

42. The girl student is drawing/drew/has drawn a picture of a giraffe on the paper.

43. The little boy is wiping/wiped/has wiped the dirty window with the cloth.

44. The two boys are digging/dug/ have dug a hole together in the backyard.

45. The two young men are assembling/assembled/ have assembled the bookshelf together.

46. The middle-aged man is chopping/chopped/has chopped the tree down with an axe.

47. The young woman is cutting/cut/has cut the two oranges into slices.

48. The small rabbit is pulling/pulled/has pulled a big carrot from the ground.

49. The strong boy is lifting/lifted/has lifted the weights up his head.

50. The two sisters are folding/folded/ have folded all the clothes on their own.

51. The two angry businessmen are tearing/tore/ have torn the contract in half.

52. The young couple are decorating/decorated/ have decorated the Christmas tree together.

53. The old lady is knitting/knitted/has knitted a Christmas sock for her grandson.

54. The male driver is changing/changed/has changed the tyre of his red car.

Appendix 4: The Sentence-matching Task Materials

1. Last month, Susan ate only fruits and vegetables to lose weight.
   *For a month now, Susan ate only fruits and vegetables to lose weight.
   For the past month, Susan has eaten only fruits and vegetables to lose weight.
   *Last month, Susan has eaten only fruits and vegetables to lose weight.
   Right now, Susan is eating only fruits and vegetables to lose weight.
   *Lately, Susan is eating only fruits and vegetables to lose weight.

2. Last year, my grandmother taught me things about growing plants.
   *Since last year, my grandmother taught me things about growing plants.
   Since last year, my grandmother has taught me things about growing plants.
   *Last year, my grandmother has taught me things about growing plants.
   Right now, my grandmother is teaching me things about growing plants.
   *Lately, my grandmother is teaching me things about growing plants.

3. Last summer, my two bothers studied Italian at a school in Rome.
   *For the past year, my two brothers studied Italian at a school in Rome.
   For the past year, my two brothers have studied Italian at a school in Rome.
   *Last summer, my two brothers have studied Italian at a school in Rome.
   Currently, my two brothers are studying Italian at a school in Rome.
   *Lately, my two brothers are studying Italian at a school in Rome.

4. Two years ago, Chris lived with three other people in a shared house.
   *For the last two years, Chris lived with three other people in a shared house.
   For the last two years, Chris has lived with three other people in a shared house.
   *Two years ago, Chris has lived with three other people in a shared house.
   Now, Chris is living with three other people in a shared house.
   *Recently, Chris is living with three other people in a shared house.

5. A week ago, Monica found it difficult to talk about her problems.
   *For the past week, Monica found it difficult to talk about her problems.
   For the past week, Monica has found it difficult to talk about her problems.
   *A week ago, Monica has found it difficult to talk about her problems.

Now, Monica is finding it difficult to talk about her problems.
*Lately, Monica is finding it difficult to talk about her problems.

6. Last Friday, Lucy met two new friends from her book club.
   *Since last Friday, Lucy met two new friends from her book club.
   Since last Friday, Lucy has met two new friends from her book club.
   *Last Friday, Lucy has met two new friends from her book club.
   Right now, Lucy is meeting two new friends from her book club.
   *Every day, Lucy is meeting two new friends from her book club.

7. Three years ago, Matt worked at the biggest software company in India.
   *For the last three years, Matt worked at the biggest software company in India.
   For the last three years, Matt has worked at the biggest software company in India.
   *Three years ago, Matt has worked at the biggest software company in India.
   Currently, Matt is working at the biggest software company in India.
   *Lately, Matt is working at the biggest software company in India.

8. Last night, Mike considered taking a job in the local café for extra cash.
   *Since last night, Mike considered taking a job in the local café for extra cash.
   Since last night, Mike has considered taking a job in the local café for extra cash.
   *Last night, Mike has considered taking a job in the local café for extra cash.
   Right now, Mike is considering taking a job in the local café for extra cash.
   *Recently, Mike is considering taking a job in the local café for extra cash.

9. Last June, the two kids stayed with their grandparents in Montreal.
   *Since last June, the two kids stayed with their grandparents in Montreal.
   Since last June, the two kids have stayed with their grandparents in Montreal.
   *Last June, the two kids have stayed with their grandparents in Montreal.
   Now, the two kids are staying with their grandparents in Montreal.
   *Recently, the two kids are staying with their grandparents in Montreal.

10. A few hours ago, Emma felt unwell after the cold shower in the morning.
    *For a few hours now, Emma felt unwell after the cold shower in the morning.
    For a few hours now, Emma has felt unwell after the cold shower in the morning.

*A few hours ago, Emma has felt unwell after the cold shower in the morning.
At the moment, Emma is feeling unwell after the cold shower in the morning.
*Every morning, Emma is feeling unwell after the cold shower in the morning.

11. A few days ago, I helped my neighbour with his garden work.
    *For the last few days, I helped my neighbour with his garden work.
    For the last few days, I have helped my neighbour with his garden work.
    *A few days ago, I have helped my neighbour with his garden work.
    Right now, I am helping my neighbour with his garden work.
    *Every day, I am helping my neighbour with his garden work.

12. Yesterday, Linda thought about buying an expensive dress for the party.
    *Since yesterday, Linda thought about buying an expensive dress for the party.
    Since yesterday, Linda has thought about buying an expensive dress for the party.
    *Yesterday, Linda has thought about buying an expensive dress for the party.
    At the moment, Linda is thinking about buying an expensive dress for the party.
    *Every day, Linda is thinking about buying an expensive dress for the party.

13. Last Sunday, my neighbour planted different types of flowers in her garden.
    *Since last Sunday, my neighbour planted different types of flowers in her garden.
    Since last Sunday, my neighbour has planted different types of flowers in her garden.
    *Last Sunday, my neighbour has planted different types of flowers in her garden.
    Right now, my neighbour is planting different types of flowers in her garden.
    *Every day, my neighbour is planting different types of flowers in her garden.

14. Last night, I ate vanilla ice cream and a piece of cheesecake.
    *Since last night, I ate vanilla ice cream and a piece of cheesecake.
    Since last night, I have eaten vanilla ice cream and a piece of cheesecake.
    *Last night, I have eaten vanilla ice cream and a piece of cheesecake.
    Right now, I am eating vanilla ice cream and a piece of cheesecake.
    *Every day, I am eating vanilla ice cream and a piece of cheesecake.

15. A year ago, David wrote a story about his trip on the website.
    *For the last year, David wrote a story about his trip on the website.

For the last year, David has written a story about his trip on the website.
*A year ago, David has written a story about his trip on the website.
Currently, David is writing a story about his trip on the website.
*Recently, David is writing a story about his trip on the website.

16. Yesterday, my colleagues talked about going to the cinema together.
*Since yesterday, my colleagues talked about going to the cinema together.
Since yesterday, my colleagues have talked about going to the cinema together.
*Yesterday, my colleagues have talked about going to the cinema together.
At the moment, my colleagues are talking about going to the cinema together.
*Every day, my colleagues are talking about going to the cinema together.

17. In 2008, Benjamin and his family planned to move back to Germany.
*Since 2008, Benjamin and his family planned to move back to Germany.
Since 2008, Benjamin and his family have planned to move back to Germany.
*In 2008, Benjamin and his family have planned to move back to Germany.
Currently, Benjamin and his family are planning to move back to Germany.
*Recently, Benjamin and his family are planning to move back to Germany.

18. A few days ago, Anna made cookies for her friends from the yoga class.
*For the last few days, Anna made cookies for her friends from the yoga class.
For the last few days, Anna has made cookies for her friends from the yoga class.
*A few days ago, Anna has made cookies for her friends from the yoga class.
Right now, Anna is making cookies for her friends from the yoga class.
*Lately, Anna is making cookies for her friends from the yoga class.

19. Many years ago, many doctors used this medicine for the treatment of cancer.
*For years now, many doctors used this medicine for the treatment of cancer.
For years now, many doctors have used this medicine for the treatment of cancer.
*Many years ago, many doctors have used this medicine for the treatment of cancer.
Now, many doctors are using this medicine for the treatment of cancer.
*Lately, many doctors are using this medicine for the treatment of cancer.

20. A few hours ago, Jason slept in the guest room after a long driving here.

*For the last few hours, Jason slept in the guest room after a long driving here.
For the last few hours, Jason has slept in the guest room after a long driving here.
*A few hours ago, Jason has slept in the guest room after a long driving here.
At the moment, Jason is sleeping in the guest room after a long driving here.
*Every morning, Jason is sleeping in the guest room after a long driving here.

21. Two years ago, this local artist painted pictures of his childhood memories.
    *For the last two years, this local artist painted pictures of his childhood memories.
    For the last two years, this local artist has painted pictures of his childhood memories.
    *Two years ago, this local artist has painted pictures of his childhood memories.
    Right now, this local artist is painting pictures of his childhood memories.
    *Every day, this local artist is painting pictures of his childhood memories.

22. Last year, Sam learnt to play the guitar by watching YouTube tutorials.
    *Since last year, Sam learnt to play the guitar by watching YouTube tutorials.
    Since last year, Sam has learnt to play the guitar by watching YouTube tutorials.
    *Last year, Sam has learnt to play the guitar by watching YouTube tutorials.
    At the moment, Sam is learning to play the guitar by watching YouTube tutorials.
    *Every day, Sam is learning to play the guitar by watching YouTube tutorials.

23. Last Saturday, the students sold drinks and snacks in the local park.
    *For the past week, the students sold drinks and snacks in the local park.
    For the past week, the students have sold drinks and snacks in the local park.
    *Last Saturday, the students have sold drinks and snacks in the local park.
    Right now, the students are selling drinks and snacks in the local park.
    *Every day, the students are selling drinks and snacks in the local park.

24. Yesterday, Betty’s mother baked different pies for her birthday party.
    *Since yesterday, Betty’s mother baked different pies for her birthday party.
    Since yesterday, Betty’s mother has baked different pies for her birthday party.
    *Yesterday morning, Betty’s mother has baked different pies for her birthday party.
    At the moment, Betty’s mother is baking different pies for her birthday party.
    *Recently, Betty’s mother is baking different pies for her birthday party.

25. Two months ago, John trained himself for his second marathon in London.
    *For the last two months, John trained himself for his second marathon in London.
    For the last two months, John has trained himself for his second marathon in London.
    *Two months ago, John has trained himself for his second marathon in London.
    Currently, John is training himself for his second marathon in London.
    *Recently, John is training himself for his second marathon in London.

26. A few hours ago, the plumber tried to fix the broken washing machine.
    *For the last few hours, the plumber tried to fix the broken washing machine.
    For the last few hours, the plumber has tried to fix the broken washing machine.
    *A few hours ago, the plumber has tried to fix the broken washing machine.
    Right now, the plumber is trying to fix the broken washing machine.
    *Every day, the plumber is trying to fix the broken washing machine.

27. Two days ago, Julie created colourful graphs by using this new software.
    *For the last two days, Julie created colourful graphs by using this new software.
    For the last two days, Julie has created colourful graphs by using this new software.
    *Two days ago, Julie has created colourful graphs by using this new software.
    At the moment, Julie is creating colourful graphs by using this new software.
    *Every morning, Julie is creating colourful graphs by using this new software.

28. Ten minutes ago, the children screamed at their parents for more toys.
    *For the last ten minutes, the children screamed at their parents for more toys.
    For the last ten minutes, the children have screamed at their parents for more toys.
    *Ten minutes ago, the children have screamed at their parents for more toys.
    Right now, the children are screaming at their parents for more toys.
    *Every day, the children are screaming at their parents for more toys.

29. An hour ago, the kids in the class fought over the largest piece of cake.
    *For the past hour, the kids in the class fought over the largest piece of cake.
    For the past hour, the kids in the class have fought over the largest piece of cake.
    *An hour ago, the kids in the class have fought over the largest piece of cake.
    At the moment, the kids in the class are fighting over the largest piece of cake.
    *Every Monday, the kids in the class are fighting over the largest piece of cake.

30. Last night, the couple argued about where to go for the coming holiday.
*Since last night, the couple argued about where to go for the coming holiday.
Since last night, the couple have argued about where to go for the coming holiday.
*Last night, the couple have argued about where to go for the coming holiday.
Right now, the couple are arguing about where to go for the coming holiday.
*Every day, the couple are arguing about where to go for the coming holiday.

31. Three hours ago, Daniel’s assistant waited for him at the airport.
*For the last three hours, Daniel’s assistant waited for him at the airport.
For the last three hours, Daniel’s assistant has waited for him at the airport.
*Three hours ago, Daniel’s assistant has waited for him at the airport.
At the moment, Daniel’s assistant is waiting for him at the airport.
*Every day, Daniel’s assistant is waiting for him at the airport.

32. Last year, the company developed new products for the existing market.
*Since last year, the company developed new products for the existing market.
Since last year, the company has developed new products for the existing market.
*Last year, the company has developed new products for the existing market.
Now, the company is developing new products for the existing market.
*Lately, the company is developing new products for the existing market.

33. Last summer, Adam designed short courses for beginning learners of French.
*For the last year, Adam designed short courses for beginning learners of French.
For the last year, Adam has designed short courses for beginning learners of French.
*Last summer, Adam has designed short courses for beginning learners of French.
Currently, Adam is designing short courses for beginning learners of French.
*Lately, Adam is designing short courses for beginning learners of French.

34. A few months ago, the professor drafted proposals for his new projects.
*For the last few months, the professor drafted proposals for his new projects.
For the last few months, the professor has drafted proposals for his new projects.
*A few months ago, the professor has drafted proposals for his new projects.
Now, the professor is drafting proposals for his new projects.

*Lately, the professor is drafting proposals for his new projects.

35. Last Monday, the politicians discussed the matter from different perspectives.
    *Since last Monday, the politicians discussed the matter from different perspectives.
    Since last Monday, the politicians have discussed the matter from different perspectives.
    *Last Monday, the politicians have discussed the matter from different perspectives.
    Right now, the politicians are discussing the matter from different perspectives.
    *Every day, the politicians are discussing the matter from different perspectives.

36. Several hours ago, my father wrapped all my textbooks in old newspapers.
    *For the last few hours, my father wrapped all my textbooks in old newspapers.
    For the last few hours, my father has wrapped all my textbooks in old newspapers.
    *Several hours ago, my father has wrapped all my textbooks in old newspapers.
    At the moment, my father is wrapping all my textbooks in old newspapers.
    *Every day, my father is wrapping all my textbooks in old newspapers.

37. Yesterday, I cleaned the dust under the furniture in the living room.
    *Since yesterday, I cleaned the dust under the furniture in the living room.
    Since yesterday, I have cleaned the dust under the furniture in the living room.
    *Yesterday, I have cleaned the dust under the furniture in the living room.
    Right now, I am cleaning the dust under the furniture in the living room.
    *Every day, I am cleaning the dust under the furniture in the living room.

38. Last night, Tracy and her little brother were very hungry and tired.
    *Since last night, Tracy and her little brother were very hungry and tired.
    Since last night, Tracy and her little brother have been very hungry and tired.
    *Last night, Tracy and her little brother have been very hungry and tired.
    Right now, Tracy and her little brother are very hungry and tired.
    *Recently, Tracy and her little brother are very hungry and tired.

39. Several weeks ago, Sarah’s tutor encouraged her to sign up for the chess class.
    *For the last few weeks, Sarah’s tutor encouraged her to sign up for the chess class.
    For the last few weeks, Sarah’s tutor has encouraged her to sign up for the chess class.
    *Several weeks ago, Sarah’s tutor has encouraged her to sign up for the chess class.

Currently, Sarah’s tutor is encouraging her to sign up for the chess class.
*Recently Sarah’s tutor is encouraging her to sign up for the chess class.

40. Yesterday, my neighbour told me everything about his holiday in Paris.
*For the last few hours, my neighbour told me everything about his holiday in Paris.
For the last few hours, my neighbour has told me everything about his holiday in Paris.
*Yesterday, my neighbour has told me everything about his holiday in Paris.
At the moment, my neighbour is telling me everything about his holiday in Paris.
*Every day, my neighbour is telling me everything about his holiday in Paris.

41. Ten years ago, Jack’s grandfather made wooden furniture for a living.
*For the last ten years, Jack’s grandfather made wooden furniture for a living.
For the last ten years, Jack’s grandfather has made wooden furniture for a living.
*Ten years ago, Jack’s grandfather has made wooden furniture for a living.
Now, Jack’s grandfather is making wooden furniture for a living.
*Lately, Jack’s grandfather is making wooden furniture for a living.

42. Last year, the Student Union organised different talks for the students.
*Since last year, the Student Union organised different talks for the students.
Since last year, the Student Union has organised different talks for the students.
*Last year, the Student Union has organised different talks for the students.
Currently, the Student Union is organising different talks for the students.
*Lately, the Student Union is organising different talks for the students.

43. Three months ago, Charlie prepared different speeches for the conference.
*For the last three months, Charlie prepared different speeches for the conference.
For the last three months, Charlie has prepared different speeches for the conference.
*Three months ago, Charlie has prepared different speeches for the conference.
Now, Charlie is preparing different speeches for the conference.
*Lately, Charlie is preparing different speeches for the conference.

44. An hour ago, the photographer took pictures of the beautiful sunset.
*For the past hour, the photographer took pictures of the beautiful sunset.
For the past hour, the photographer has taken pictures of the beautiful sunset.

*An hour ago, the photographer has taken pictures of the beautiful sunset.
Right now, the photographer is taking pictures of the beautiful sunset.
*Recently, the photographer is taking pictures of the beautiful sunset.

45. In 2020, Laura refused to eat any meat and fish in her diet.
*Since 2020, Laura refused to eat any meat and fish in her diet.
Since 2020, Laura has refused to eat any meat and fish in her diet.
*In 2020, Laura has refused to eat any meat and fish in her diet.
Now, Laura is refusing to eat any meat and fish in her diet.
*Lately, Laura is refusing to eat any meat and fish in her diet.

46. Last week, Alice and her team were at the biggest hotel in New York.
*Since last week, Alice and her team were at the biggest hotel in New York.
Since last week, Alice and her team have been at the biggest hotel in New York.
*Last week, Alice and her team have been at the biggest hotel in New York.
Right now, Alice and her team are at the biggest hotel in New York.
*Every day, Alice and her team are at the biggest hotel in New York.

47. Last March, the London office assisted us with this marketing campaign.
*Since last March, the London office assisted us with this marketing campaign.
Since last March, the London office has assisted us with this marketing campaign.
*Last March, the London office has assisted us with this marketing campaign.
Now, the London office is assisting us with this marketing campaign.
*Lately, the London office is assisting us with this marketing campaign.

48. In 2015, Mary struggled to balance work and family commitments.
*Since 2015, Mary struggled to balance work and family commitments.
Since 2015, Mary has struggled to balance work and family commitments.
*In 2015, Mary has struggled to balance work and family commitments.
Currently, Mary is struggling to balance work and family commitments.
*Lately, Mary is struggling to balance work and family commitments.

Appendix 5: The Elicited Imitation Task Materials

1. Last week, Linda planted new flowers in her garden.
   *For the past week, Linda planted new flowers in her garden.
   For the past week, Linda has planted new flowers in her garden.
   *Last week, Linda has planted new flowers in her garden.
   Right now, Linda is planting new flowers in her garden.
   *Lately, Linda is planting new flowers in her garden.

2. Last year, my father taught English to foreign students at the school.
   *For the past year, my father taught English to foreign students at the school.
   For the past year, my father has taught English to foreign students at the school.
   *Last year, my father has taught English to foreign students at the school.
   Currently, my father is teaching English to foreign students at the school.
   * Lately, my father is teaching English to foreign students at the school.

3. Last night, she was so shocked by the news about her friend.
   *Since last night, she was so shocked by the news about her friend.
   Since last night, she has been so shocked by the news about her friend.
   *Last night, she has been so shocked by the news about her friend.
   At the moment, she is so shocked by the news about her friend.
   *Every night, she is so shocked by the news about her friend.

4. Last Friday, the lawyer advised him to leave the country soon.
   *For the past week, the lawyer advised him to leave the country soon.
   For the past week, the lawyer has advised him to leave the country soon.
   *Last Friday, the lawyer has advised him to leave the country soon.
   Right now, the lawyer is advising him to leave the country soon.
   *Recently, the lawyer is advising him to leave the country soon.

5. In 2015, Richard studied Music at the University of York.
   *Since 2015, Richard studied Music at the University of York.
   Since 2015, Richard has studied Music at the University of York.
   *In 2015, Richard has studied Music at the University of York.
   Currently, Richard is studying Music at the University of York.
   *Recently, Richard is studying Music at the University of York.

6. A few years ago, my parents lived in a beautiful Greek island.
   *For the last few years, my parents lived in a beautiful Greek island.
   For the last few years, my parents have lived in a beautiful Greek island.
   *A few years ago, my parents have lived in a beautiful Greek island.
   Currently, my parents are living in a beautiful Greek island.
   *Lately, my parents are living in a beautiful Greek island.

7. Last Christmas, James thought about buying a new car for my dad.
   *Since last Christmas, James thought about buying a new car for my dad.
   Since last Christmas, James has thought about buying a new car for my dad.
   *Last Christmas, James has thought about buying a new car for my dad.
   Right now, James is thinking about buying a new car for my dad.
   *Every day, James is thinking about buying a new car for my dad.

8. Yesterday, the man felt very weak after the operation.
   *Since yesterday, the man felt very weak after the operation.
   Since yesterday, the man has felt very weak after the operation.
   *Yesterday, the man has felt very weak after the operation.
   At the moment, the man is feeling very weak after the operation.
   *Every day, the man is feeling very weak after the operation.

9. A few days ago, this married couple were in Italy for their honeymoon.
   *For the last few days, this married couple were in Italy for their honeymoon.
   For the last few days, this married couple have been in Italy for their honeymoon.
   *A few days ago, this married couple have been in Italy for their honeymoon.
   Now, this married couple are in Italy for their honeymoon.
   *Lately, this married couple are in Italy for their honeymoon.

10. Last month, the little girl learned how to play the piano.
    *Since last month, the little girl learned how to play the piano.
    Since last month, the little girl has learnt how to play the piano.
    *Last month, the little girl has learnt how to play the piano.
    Currently, the little girl is learning how to play the piano.

*Recently, the little girl is learning how to play the piano.

11. Two hours ago, my friends cleaned the room for the surprise party.
    *For the last two hours, my friends cleaned the room for the surprise party.
    For the last two hours, my friends have cleaned the room for the surprise party.
    *Two hours ago, my friends have cleaned the room for the surprise party.
    Right now, my friends are cleaning the room for the surprise party.
    *Every day, my friends are cleaning the room for the surprise party.

12. An hour ago, the student talked about some issues with her tutor.
    *For the past hour, the student talked about some issues with her tutor.
    For the past hour, the student has talked about some issues with her tutor.
    *An hour ago, the student has talked about some issues with her tutor.
    At the moment, the student is talking about some issues with her tutor.
    * Every morning, the student is talking about some issues with her tutor.

13. Last weekend, the princess made her own clothes for the party.
    *Since last weekend, the princess made her own clothes for the party.
    Since last weekend, the princess has made her own clothes for the party.
    *Last weekend, the princess has made her own clothes for the party.
    Right now, the princess is making her own clothes for the party.
    *Every night, the princess is making her own clothes for the party.

14. In 2010, my auntie worked as a researcher at the university.
    *Since 2010, my auntie worked as a researcher at the university.
    Since 2010, my auntie has worked as a researcher at the university.
    *In 2010, my auntie has worked as a researcher at the university.
    Currently, my auntie is working as a researcher at the university.
    *Lately, my auntie is working as a researcher at the university.

15. Last night, the little baby slept soundly in the next room.
    *Since last night, the little baby slept soundly in the next room.
    Since last night, the little baby has slept soundly in the next room.
    *Last night, the little baby has slept soundly in the next room.

At the moment, the little baby is sleeping soundly in the next room.
*Every night, the little baby is sleeping soundly in the next room.

16. A few days ago, Sarah and her husband considered selling the house.
*For the last few days, Sarah and her husband considered selling the house.
For the last few days, Sarah and her husband have considered selling the house.
*A few days ago, Sarah and her husband have considered selling the house.
Right now, Sarah and her husband are considering selling the house.
*Recently, Sarah and her husband are considering selling the house.

17. Many years ago, my uncle gave my advice about running business.
*For the past years, my uncle gave my advice about running business.
For the past years, my uncle has given my advice about running business.
*Many years ago, my uncle has given my advice about running business.
Currently, my uncle is giving my advice about running business.
*Recently, my uncle is giving my advice about running business.

18. Last summer, Amy stayed at her sister's big house in town.
*Since last summer, Amy stayed at her sister's big house in town.
Since last summer, Amy has stayed at her sister's big house in town.
*Last summer, Amy has stayed at her sister's big house in town.
Now, Amy is staying at her sister's big house in town.
*Lately, Amy is staying at her sister's big house in town.

19. Last year, many teachers applied technology in their teaching.
*For the past year, many teachers applied technology in their teaching.
For the past year, many teachers have applied technology in their teaching.
*Last year, many teachers have applied technology in their teaching.
Now, many teachers are applying technology in their teaching.
*Lately, many teachers are applying technology in their teaching.

20. A few hours ago, Ella's mother baked cookies for the kids.
*For the last few hours, Ella's mother baked cookies for the kids.
For the last few hours, Ella's mother has baked cookies for the kids.

*A few hours ago, Ella’s mother has baked cookies for the kids.
Right now, Ella’s mother is baking cookies for the kids.
*Every day, Ella’s mother is baking cookies for the kids.

21. Yesterday, my friends asked me questions about English history.
*Since yesterday, my friends asked me questions about English history.
Since yesterday, my friends have asked me questions about English history.
*Yesterday, my friends have asked me questions about English history.
Right now, my friends are asking me questions about English history.
*Recently, my friends are asking me questions about English history.

22. Last Monday, the professor encouraged us to read more widely.
*For the last week, the professor encouraged us to read more widely.
For the last week, the professor has encouraged us to read more widely.
*Last Monday, the professor has encouraged us to read more widely.
Currently, the professor is encouraging us to read more widely.
*Every day, the professor is encouraging us to read more widely.

23. Last year, they organised meetings between the teachers and parents.
*Since last year, they organised meetings between the teachers and parents.
Since last year, they have organised meetings between the teachers and parents.
*Last year, they have organised meetings between the teachers and parents.
Now, they are organising meetings between the teachers and parents.
*Recently, they are organising meetings between the teachers and parents.

24. Three weeks ago, Emma tried to lose weight before her wedding.
*For the last three weeks, Emma tried to lose weight before her wedding.
For the last three weeks, Emma has tried to lose weight before her wedding.
*Three weeks ago, Emma has tried to lose weight before her wedding.
Now, Emma is trying to lose weight before her wedding.
*Lately, Emma is trying to lose weight before her wedding.

25. Three hours ago, we discussed different solutions to the problem.
*For the last three hours, we discussed different solutions to the problem.

For the last three hours, we have discussed different solutions to the problem.
*Three hours ago, we have discussed different solutions to the problem.
At the moment, we are discussing different solutions to the problem.
*Every morning, we are discussing different solutions to the problem.

26. Last winter, the students collected money and clothes for the homeless.
*Since last winter, the students collected money and clothes for the homeless.
Since last winter, the students have collected money and clothes for the homeless.
*Last winter, the students have collected money and clothes for the homeless.
Currently, the students are collecting money and clothes for the homeless.
* Lately, the students are collecting money and clothes for the homeless.

27. A few years ago, she used the local newspaper to express her views.
*For the past years, she used the local newspaper to express her views.
For the past years, she has used the local newspaper to express her views.
*A few years ago, she has used the local newspaper to express her views.
Now, she is using the local newspaper to express her views.
*Lately, she is using the local newspaper to express her views.

28. Last month, Ben was busy with school because of the coming exams.
*Since last month, Ben was busy with school because of the coming exams.
Since last month, Ben has been busy with school because of the coming exams.
*Last month, Ben has been busy with school because of the coming exams.
Now, Ben is busy with school because of the coming exams.
*Lately, Ben is busy with school because of the coming exams.

29. Ten years ago, Sam dreamed to be a successful businessman.
*Since ten years ago, Sam dreamed to be a successful businessman.
Since ten years ago, Sam has dreamed to be a successful businessman.
*Ten years ago, Sam has dreamed to be a successful businessman.
Currently, Sam is dreaming to be a successful businessman.
*Recently, Sam is dreaming to be a successful businessman.

30. An hour ago, I had coffee and cakes in my favourite coffee shop.
*For the last hour, I had coffee and cakes in my favourite coffee shop.

For the last hour, I have had coffee and cakes in my favourite coffee shop.
*An hour ago, I have had coffee and cakes in my favourite coffee shop.
Right now, I am having coffee and cakes in my favourite coffee shop.
*Every day, I am having coffee and cakes in my favourite coffee shop.

31. Three days ago, Jenny planned to go to Rome for holiday.
*For the last three days, Jenny planned to go to Rome for holiday.
For the last three days, Jenny has planned to go to Rome for holiday.
*Three days ago, Jenny has planned to go to Rome for holiday.
Now, Jenny is planning to go to Rome for holiday.
*Lately, Jenny is planning to go to Rome for holiday.

32. A few days ago, the teacher made a reading list for the students.
*For the last few days, the teacher made a reading list for the students.
For the last few days, the teacher has made a reading list for the students.
*A few days ago, the teacher has made a reading list for the students.
At the moment, the teacher is making a reading list for the students.
*Every morning, the teacher is making a reading list for the students.

33. Two hours ago, the children drew pictures of their family members.
*For the last two hours, the children drew pictures of their family members.
For the last two hours, the children have drawn pictures of their family members.
*Two hours ago, the children have drawn pictures of their family members.
At the moment, the children are drawing pictures of their family members.
*Every morning, the children are drawing pictures of their family members.

34. Five minutes ago, the two birds fought over a piece of bread.
*For the last five minutes, the two birds fought over a piece of bread.
For the last five minutes, the two birds have fought over a piece of bread.
*Five minutes ago, the two birds have fought over a piece of bread.
Right now, the two birds are fighting over a piece of bread.
*Every day, the two birds are fighting over a piece of bread.

35. Three years ago, Vicky wrote books and poems for children.

*For the last three years, Vicky wrote books and poems for children.
For the last three years, Vicky has written books and poems for children.
*Three years ago, Vicky has written books and poems for children.
Currently, Vicky is writing books and poems for children.
*Lately, Vicky is writing books and poems for children.

36. Twenty minutes ago, my two brothers argued about what game to play.
    *For the last twenty minutes, my two brothers argued about what game to play.
    For the last twenty minutes, my two brothers have argued about what game to play.
    *Twenty minutes ago, my two brothers have argued about what game to play.
    Right now, my two brothers are arguing about what game to play.
    *Recently, my two brothers are arguing about what game to play.

37. Last month, the coach picked students to play for the football team.
    *Since last month, the coach picked students to play for the football team.
    Since last month, the coach has picked students to play for the football team.
    *Last month, the coach has picked students to play for the football team.
    Currently, the coach is picking students to play for the football team.
    *Lately, the coach is picking students to play for the football team.

38. A few years ago, the company provided free meals to all employees.
    *For the last few years, the company provided free meals to all employees.
    For the last few years, the company has provided free meals to all employees.
    *A few years ago, the company has provided free meals to all employees.
    Now, the company is providing free meals to all employees.
    *Lately, the company is providing free meals to all employees.

39. In 2018, Kate designed clothes and shoes especially for shorter women.
    *Since 2018, Kate designed clothes and shoes especially for shorter women.
    Since 2018, Kate has designed clothes and shoes especially for shorter women.
    *In 2018, Kate has designed clothes and shoes especially for shorter women.
    Currently, Kate is designing clothes and shoes especially for shorter women.
    *Lately, Kate is designing clothes and shoes especially for shorter women.

40. Last Monday, we explored new ways to market our products.
    *Since last Monday, we explored new ways to market our products.
    Since last Monday, we have explored new ways to market our products.
    *Last Monday, we have explored new ways to market our products.
    Now, we are exploring new ways to market our products.
    *Lately, we are exploring new ways to market our products.

41. Yesterday, the team developed a new marketing strategy.
    *For the past day, the team developed a new marketing strategy.
    For the past day, the team has developed a new marketing strategy.
    *Yesterday, the team has developed a new marketing strategy.
    Now, the team is developing a new marketing strategy.
    *Recently, the team is developing a new marketing strategy.

42. Last summer, Jane created characters and plots for her new novel.
    *Since last summer, Jane created characters and plots for her new novel.
    Since last summer, Jane has created characters and plots for her new novel.
    *Last summer, Jane has created characters and plots for her new novel.
    Right now, Jane is creating characters and plots for her new novel.
    *Every day, Jane is creating characters and plots for her new novel.

43. Two years ago, the company built hotels and restaurants on the island.
    *For the last two years, the company built hotels and restaurants on the island.
    For the last two years, the company has built hotels and restaurants on the island.
    *Two years ago, the company has built hotels and restaurants on the island.
    Currently, the company is building hotels and restaurants on the island.
    *Lately, the company is building hotels and restaurants on the island.

44. Three hours ago, my friend prepared food and drinks for our movie night.
    *For the last three hours, my friend prepared food and drinks for our movie night.
    For the last three hours, my friend has prepared food and drinks for our movie night.
    *Three hours ago, my friend has prepared food and drinks for our movie night.
    At the moment, my friend is preparing food and drinks for our movie night.
    *Every morning, my friend is preparing food and drinks for our movie night.

45. Many months ago, Chloe struggled to find her true passion in life.
    *For months now, Chloe struggled to find her true passion in life.
    For months now, Chloe has struggled to find her true passion in life.
    *Many months ago, Chloe has struggled to find her true passion in life.
    Now, Chloe is struggling to find her true passion in life.
    *Recently, Chloe is struggling to find her true passion in life.

46. A few days ago, the manager persuaded me to invest more money.
    *For the last few days, the manager persuaded me to invest more money.
    For the last few days, the manager has persuaded me to invest more money.
    *A few days ago, the manager has persuaded me to invest more money.
    Currently, the manager is persuading me to invest more money.
    *Lately, the manager is persuading me to invest more money.

47. Last month, she found it difficult to concentrate on her work.
    *For the last month, she found it difficult to concentrate on her work.
    For the last month, she has found it difficult to concentrate on her work.
    *Last month, she has found it difficult to concentrate on her work.
    Right now, she is finding it difficult to concentrate on her work.
    *Every day, she is finding it difficult to concentrate on her work.

48. Two hours ago, Lisa cooked steak and potatoes for her boyfriend.
    *For the last two hours, Lisa cooked steak and potatoes for her boyfriend.
    For the last two hours, Lisa has cooked steak and potatoes for her boyfriend.
    *Two hours ago, Lisa has cooked steak and potatoes for her boyfriend.
    Right now, Lisa is cooking steak and potatoes for her boyfriend.
    * Every day, Lisa is cooking steak and potatoes for her boyfriend.

Appendix 6: The Gap-filling Task Materials

This task is to complement data from the elicited imitation task, which looks at online production. So the items in this task are the same with the imitation task.

Instructions:

This is a gap-filling task. Please *read* the sentences and *fill in the blanks* by using the verbs in the brackets. Try to provide the appropriate form of the verbs. The first five have been filled in as examples.

Examples:

1. Sam wants (want) to learn a foreign language in his spare time.

2. This time tomorrow, the plane will be arriving (arrive) in London.

3. Brad used to play (play) with his cousin on the beach nearby.

4. Anna usually cycles (cycle) to work in the morning.

5. Next year, Sarah and her family will move (move) to Japan.

Please fill in the blanks of the following sentences.

1. Last week, Linda ____________ (plant) new flowers in her garden.

   For the past week, Linda ____________ (plant) new flowers in her garden.

   Right now, Linda ____________ (plant) new flowers in her garden.

2. Last year, my father ____________ (teach) English to foreign students at the school.

   For the past year, my father ____________ (teach) English to foreign students at the school.

   Currently, my father ____________ (teach) English to foreign students at the school.

3. Last night, she ____________ (shock) about the news about her friend.

   Since last night, she ____________ (shock) about the news about her friend.

   At the moment, she ____________ (shock) about the news about her friend.

4. Last Friday, the lawyer ____________ (advise) him to leave the country soon.

For the past week, the lawyer ___________ (advise) him to leave the country soon.

Right now, the lawyer ___________ (advise) him to leave the country soon.

5. In 2015, Richard ___________ (study) Music at the University of York.

   Since 2015, Richard ___________ (study) Music at the University of York.

   Currently, Richard ___________ (study) Music at the University of York.

6. A few years ago, my parents ___________ (live) on a beautiful Greek island.

   For the last few years, my parents ___________ (live) on a beautiful Greek island.

   Currently, my parents ___________ (live) on a beautiful Greek island.

7. Last Christmas, James ___________ (think) about buying a new car for my dad.

   Since last Christmas, James ___________ (think) about buying a new car for my dad.

   Right now, James ___________ (think) about buying a new car for my dad.

8. Yesterday, the man ___________ (feel) very weak after the operation.

   Since yesterday, the man ___________ (feel) very weak after the operation.

   At the moment, the man ___________ (feel) very weak after the operation.

9. A few days ago, this married couple ___________ (are) in Italy for their honeymoon.

   For the last few days, this married couple ___________ (are) in Italy for their honeymoon.

   Now, this married couple ___________ (are) in Italy for their honeymoon.

10. Last month, the little girl ___________ (learn) how to play the piano.

    Since last month, the little girl ___________ (learn) how to play the piano.

    Currently, the little girl ___________ (learn) how to play the piano.

11. Two hours ago, my friends ___________ (clean) the room for the surprise party.

    For the last two hours, my friends ___________ (clean) the room for the surprise party.

    Right now, my friends ___________ (clean) the room for the surprise party.

12. An hour ago, the student ___________ (talk) about some issues with her tutor.

    For the past hour, the student ___________ (talk) about some issues with her tutor.

At the moment, the student ___________ (talk) about some issues with her tutor.

13. Last weekend, the princess ___________ (make) her own clothes for the party.

    Since last weekend, the princess ___________ (make) her own clothes for the party.

    Right now, the princess ___________ (make) her own clothes for the party.

14. In 2010, my auntie ___________ (work) as a researcher at the university.

    Since 2010, my auntie ___________ (work) as a researcher at the university.

    Currently, my auntie ___________ (work) as a researcher at the university.

15. Last night, the little baby ___________ (sleep) soundly in the next room.

    Since last night, the little baby ___________ (sleep) soundly in the next room.

    At the moment, the little baby ___________ (sleep) soundly in the next room.

16. A few days ago, Sarah and her husband ___________ (consider) selling the house.

    For the last few days, Sarah and her husband ___________ (consider) selling the house.

    Right now, Sarah and her husband ___________ (consider) selling the house.

17. Many years ago, my uncle ___________ (give) me advice about running a business.

    For the past years, my uncle ___________ (give) me advice about running a business.

    Currently, my uncle ___________ (give) me advice about running a business.

18. Last summer, Amy ___________ (stay) at her sister's big house in town.

    Since last summer, Amy ___________ (stay) at her sister's big house in town.

    Now, Amy ___________ (stay) at her sister's big house in town.

19. Last year, many teachers ___________ (apply) technology in their teaching.

    For the past year, many teachers ___________ (apply) technology in their teaching.

    Now, many teachers ___________ (apply) technology in their teaching.

20. A few hours ago, Ella's mother ___________ (bake) cookies for the kids.

    For the last few hours, Ella's mother ___________ (bake) cookies for the kids.

    Right now, Ella's mother ___________ (bake) cookies for the kids.

21. Yesterday, my friends ___________ (ask) me questions about English history.

    Since yesterday, my friends ___________ (ask) me questions about English history.

    Right now, my friends ___________ (ask) me questions about English history.

22. Last Monday, the professor ___________ (encourage) us to read more widely.

    For the last week, the professor ___________ (encourage) us to read more widely.

    Currently, the professor ___________ (encourage) us to read more widely.

23. Last year, they ___________ (organise) meetings between the teachers and parents.

    Since last year, they ___________ (organise) meetings between the teachers and parents.

    Now, they ___________ (organise) meetings between the teachers and parents.

24. Three weeks ago, Emma ___________ (try) to lose weight before her wedding.

    For the last three weeks, Emma ___________ (try) to lose weight before her wedding.

    Now, Emma ___________ (try) to lose weight before her wedding.

25. Three hours ago, we ___________ (discuss) different solutions to the problem.

    For the last three hours, we ___________ (discuss) different solutions to the problem.

    At the moment, we ___________ (discuss) different solutions to the problem.

26. Last winter, the students ___________ (collect) money and clothes for the homeless.

    Since last winter, the students ___________ (collect) money and clothes for the homeless.

    Currently, the students ___________ (collect) money and clothes for the homeless.

27. A few years ago, she ___________ (use) the local newspaper to express her views.

    For the past years, she ___________ (use) the local newspaper to express her views.

    Now, she ___________ (use) the local newspaper to express her views.

28. Last month, Ben ___________ (busy) with school because of the coming exams.

    Since last month, Ben ___________ (busy) with school because of the coming exams.

    Now, Ben ___________ (busy) with school because of the coming exams.

29. Ten years ago, Sam ___________ (dream) to be a successful businessman.

    Since ten years ago, Sam ___________ (dream) to be a successful businessman.

    Currently, Sam ___________ (dream) to be a successful businessman.

30. An hour ago, I ___________ (have) coffee and cakes in my favourite coffee shop.

    For the last hour, I ___________ (have) coffee and cakes in my favourite coffee shop.

    Right now, I ___________ (have) coffee and cakes in my favourite coffee shop.

31. Three days ago, Jenny ___________ (plan) to go to Rome for a holiday.

    For the last three days, Jenny ___________ (plan) to go to Rome for a holiday.

    Now, Jenny ___________ (plan) to go to Rome for a holiday.

32. A few days ago, the teacher ___________ (make) a reading list for the students.

    For the last few days, the teacher ___________ (make) a reading list for the students.

    At the moment, the teacher ___________ (make) a reading list for the students.

33. Two hours ago, the children ___________ (draw) pictures of their family members.

    For the last two hours, the children ___________ (draw) pictures of their family members.

    At the moment, the children ___________ (draw) pictures of their family members.

34. Five minutes ago, the two birds ___________ (fight) over a piece of bread.

    For the last five minutes, the two birds ___________ (fight) over a piece of bread.

    Right now, the two birds ___________ (fight) over a piece of bread.

35. Three years ago, Vicky ___________ (write) books and poems for children.

    For the last three years, Vicky ___________ (write) books and poems for children.

    Currently, Vicky ___________ (write) books and poems for children.

36. Twenty minutes ago, my two brothers ___________ (argue) about what game to play.

    For the last twenty minutes, my two brothers ___________ (argue) about what game to play.

Right now, my two brothers ___________ (argue) about what game to play.

37. Last month, the coach ___________ (pick) students to play for the football team.

Since last month, the coach ___________ (pick) students to play for the football team.

Currently, the coach ___________ (pick) students to play for the football team.

38. A few years ago, the company ___________ (provide) free meals to all employees.

For the last few years, the company ___________ (provide) free meals to all employees.

Now, the company ___________ (provide) free meals to all employees.

39. In 2018, Kate ___________ (design) clothes and shoes especially for shorter women.

Since 2018, Kate ___________ (design) clothes and shoes especially for shorter women.

Currently, Kate ___________ (design) clothes and shoes especially for shorter women.

40. Last Monday, we ___________ (explore) new ways to market our products.

Since last Monday, we ___________ (explore) new ways to market our products.

Now, we ___________ (explore) new ways to market our products.

41. Yesterday, the team ___________ (develop) a new marketing strategy.

For the past day, the team ___________ (develop) a new marketing strategy.

Now, the team ___________ (develop) a new marketing strategy.

42. Last summer, Jane ___________ (create) characters and plots for her new novel.

Since last summer, Jane ___________ (create) characters and plots for her new novel.

Right now, Jane ___________ (create) characters and plots for her new novel.

43. Two years ago, the company ___________ (build) hotels and restaurants on the island.

For the last two years, the company ___________ (build) hotels and restaurants on the island.

Currently, the company ___________ (build) hotels and restaurants on the island.

44. Three hours ago, my friend ___________ (prepare) food and drinks for our movie night.

For the last three hours, my friend ____________ (prepare) food and drinks for our movie night.

At the moment, my friend ____________ (prepare) food and drinks for our movie night.

45. Many months ago, Chloe ____________ (struggle) to find her true passion in life.

For months now, Chloe ____________ (struggle) to find her true passion in life.

Now, Chloe ____________ (struggle) to find her true passion in life.

46. A few days ago, the manager ____________ (persuade) me to invest more money.

For the last few days, the manager ____________ (persuade) me to invest more money.

Currently, the manager ____________ (persuade) me to invest more money.

47. Last month, she ____________ (find) it difficult to concentrate on her work.

For the last month, she ____________ (find) it difficult to concentrate on her work.

Right now, she ____________ (find) it difficult to concentrate on her work.

48. Two hours ago, Lisa ____________ (cook) steak for her boyfriend.

For the last two hours, Lisa ____________ (cook) steak for her boyfriend.

Right now, Lisa ____________ (cook) steak for her boyfriend.

References

Ahn, H., & Song, M.-J. (2023). L1-Korean speakers' definiteness processing in L2 English: A visual world paradigm eye tracking study. *Bilingualism (Cambridge, England)*, *26*(3), 556–569. https://doi.org/10.1017/S1366728922000840

Al-Aswad, M. K. (1983). *Contrastive analysis of Arabic and English verbs in tense, aspect, and structure* [Thesis, University of Michigan]. http://deepblue.lib.umich.edu/handle/2027.42/159694

Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press. https://doi.org/10.1017/CBO9780511732935

Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1–L2 similarity and structural distance. *Second Language Research*, *30*(3), 275–306. https://doi.org/10.1177/0267658313515671

Alemán Bañón, J., Miller, D., & Rothman, J. (2017). Morphological variability in second language learners: An examination of electrophysiological and production data. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(10), 1509–1536. https://doi.org/10.1037/xlm0000394

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Alruwaili, R. (2014). *Second language acquisition of aspect and tense by Saudi-Arabic learners of English* [Doctoral Dissertation]. University of York.

Al-Thubaiti, K. A. (2015). L2 acquisition of English aspect by L1 Arabic speakers: The role of interpretable features at the syntax-semantics interface. In D. Ayoun (Ed.), *The* *acquisition of the present* (pp. 185–214). John Benjamins Publishing Company. https://doi.org/10.1075/z.196.07alt

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Amaro, J. Cabrelli., Campos-Dintrans, G., & Rothman, J. (2018). The role of L1 phonology in L2 morphological production: L2 English past tense production by L1 Spanish, Mandarin, and Japanese Speakers. *Studies in Second Language Acquisition*, *40*(3), 503–527. https://doi.org/10.1017/S0272263117000122

Amos, R. M., Seeber, K. G., & Pickering, M. J. (2022). Prediction during simultaneous interpreting: Evidence from the visual-world paradigm. *Cognition*, *220*, 104987. https://doi.org/10.1016/j.cognition.2021.104987

Andersen, R. W. (1991). Developmental sequences: The emergence of aspect marking in second language acquisition. In T. Huebner & C. A. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theories* (pp. 305–324). Benjamins.

Andersen, R. W. (2002). Dimensions of "pastness". In M. R. Salaberry & Y. Shirai (Eds.), *L2 Acquisition of Tense–Aspect Morphology* (pp. 79–105). John Benjamins Publishing Company.

Andersen, R. W., & Shirai, Y. (1994). Discourse motivations for some cognitive acquisition principles. *Studies in Second Language Acquisition*, *16*(2), 133–156. https://doi.org/10.1017/S0272263100012845

Andersen, R. W., & Shirai, Y. (1996). The primacy of aspect in first and second language acquisition: The pidgin-creole connection. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 527–570). Academic Press.

Andringa, S. (2020). The emergence of awareness in uninstructed L2 learning: A visual world eye tracking study. *Second Language Research, 36*(3), 335–357. https://doi.org/10.1177/0267658320915502

Arin, M. L. (2003). *Aspect, tense and mood: Context dependency and the marker 'le' in Mandarin Chinese* [Doctoral Dissertation]. Lund University.

Armstrong, A., Bulkes, N., & Tanner, D. (2018). Quantificational cues modulate the processing of English subject-verb agreement by native Chinese speakers: An ERP study. *Studies in Second Language Acquisition*, *40*(4), 731–754. https://doi.org/10.1017/S0272263118000013

Asiyaban, A. R., Yamini, M., Bagheri, M.-S., & Yarmohammadi, L. (2020). Implicit/explicit knowledge and its contribution towards tense consistency employment across EFL learners' proficiency levels. *Cogent Education*, *7*(1), 1727129. https://doi.org/10.1080/2331186X.2020.1727129

Ayoun, D., & Rothman, J. (2012). Generative approaches to the acquisition of tense, aspect and mood in adult L2. In M. R. Salaberry & L. Comajoan-Colomé (Eds.), *Research design and methodology in studies on L2 tense and aspect* (pp. 119–156). Mounton de Gryuter.

Ayoun, D., & Salaberry, M. R. (2008). Acquisition of English tense-aspect morphology by advanced French instructed learners. *Language Learning*, *58*(3), 555–595. https://doi.org/10.1111/j.1467-9922.2008.00450.x

Baayen, R. H. (2007). Mixed models. In *Analyzing linguistic data: A practical introduction to statistics using R* (pp. 241–302). Cambridge University Press.

http://lib.myilibrary.com/browse/open.asp?id=239009&entityid=https://shib.york.ac.uk/shibboleth

Baayen, R. H. (2010). A real experiment is a factorial experiment? *The Mental Lexicon*, *5*(1), 149–157. https://doi.org/10.1075/ml.5.1.06baa

Baayen, R. H., Van Rij, J., De Cat, C., & Wood, S. (2018). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by generalized additive mixed models. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed-Effects Regression Models in Linguistics* (pp. 49–69). Springer International Publishing. https://doi.org/10.1007/978-3-319-69830-4_4

Badawi, E.-S. M., Carter, M. G., & Gully, A. (2004). *Modern written Arabic: A comprehensive grammar*. Routledge.

Bahloul, M. (2008). *Structure and function of the Arabic verb* (1st edition). Routledge.

Baker, C., & Love, T. (2022). It's about time! Time as a parameter for lexical and syntactic processing: An eye-tracking-while-listening investigation. *Language, Cognition and Neuroscience*, *37*(1), 42–62. https://doi.org/10.1080/23273798.2021.1941147

Bardovi-Harlig, K. (1992). The Use of Adverbials and Natural Order in the Development of Temporal Expression. *International Review of Applied Linguistics in Language Teaching*, *30*(4), 299–320.

Bardovi-Harlig, K. (1999). From morpheme studies to temporal semantics: Tense-aspect research in SLA. *Studies in Second Language Acquisition*, *21*(3), 341–382. https://doi.org/10.1017/S0272263199003010

Bardovi-Harlig, K. (2000). *Tense and aspect in second language acquisition: Form, meaning, and use*. Blackwell.

Bardovi-Harlig, K. (2001). Another piece of the puzzle: The emergence of the present perfect. *Language Learning*, *51*(1), 215–264.

Bardovi-Harlig, K. (2002). Analyzing aspect. In M. R. Salaberry & Y. Shirai (Eds.), *The L2 acquisition of tense–aspect morphology* (Vol. 27, pp. 129–154). John Benjamins Publishing Company. https://doi.org/10.1075/lald.27

Bardovi-Harlig, K., & Bergström, A. (1996). Acquisition of tense and aspect in second language and foreign language learning: Learner narratives in ESL and FFL. *The Canadian Modern Language Review*, *52*(2), 308–330. https://doi.org/10.3138/cmlr.52.2.308

Bardovi-Harlig, K., & Comajoan-Colomé, L. (2020). The Aspect Hypothesis and the acquisition of L2 past morphology in the last 20 years: A state-of-the-scholarship review. *Studies in Second Language Acquisition*, *42*(5), 1137–1167. https://doi.org/10.1017/S0272263120000194

Bardovi-Harlig, K., & Reynolds, D. W. (1995). The role of lexical aspect in the acquisition of tense and aspect. *TESOL Quarterly*, *29*(1), 107–131. https://doi.org/10.2307/3587807

Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. https://doi.org/10.1016/j.jml.2007.09.002

Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, *137*(2), 201–207. https://doi.org/10.1016/j.actpsy.2010.09.011

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General*, *143*(1), 404–413. https://doi.org/10.1037/a0031813

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). https://doi.org/10.1016/j.jml.2012.11.001

Bartoń, K. (2023). *MuMIn: Multi-model inference* (1.47.5) [Computer software]. https://cran.r-project.org/web/packages/MuMIn/index.html

Barto-Sisamout, K., Nicol, J., Witzel, J., & Witzel, N. (2009). Transfer effects in bilingual sentence processing. *Arizona Working Papers in SLA & Teaching, 16,* 1–26.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bauer, G. (1970). The English ‘Perfect’ reconsidered. *Journal of Linguistics*, *6*(2), 189–198. https://doi.org/10.1017/S0022226700002590

Benmamoun, E. (1999). Arabic morphology: The central role of the imperfective. *Lingua*, *108*(2–3), 175–201. https://doi.org/10.1016/S0024-3841(98)00045-X

Benmamoun, E. (2000). *The feature structure of functional categories: A comparative study of Arabic dialects*. Oxford University Press.

Berends, S. M., Brouwer, S. M., & Sprenger, S. A. (2016). Eye-Tracking and the Visual World Paradigm. In M. S. Schmid, S. M. Berends, C. Bergmann, S. M. Brouwer, N. Meulman, B. J. Seton, S. A. Sprenger, & L. A. Stowe, *Designing Research on Bilingual Development* (pp. 55–80). Springer International Publishing. https://doi.org/10.1007/978-3-319-11529-0_5

Binnick, R. I. (2005). The markers of habitual aspect in English. *Journal of English Linguistics*, *33*(4), 339–369. https://doi.org/10.1177/0075424205286006

Bley-Vroman, R. (1989). What is the logical problem of foreign language learning? In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (1st ed., pp. 41–68). Cambridge University Press. https://doi.org/10.1017/CBO9781139524544.005

Bley-Vroman, R., & Chaudron, C. (1994). Elicited imitation as a measure of second-language competence. In E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 245–261). L. Erlbaum Associates.

Bley-Vroman, R., & Masterson, D. (1989). Reaction-time as a supplement to grammaticality judgments in the investigation of second language learners' competence. *University of Hawai'i Working Papers in ESL, 8*(2), 207–237.

Boland, A. (2006). *Aspect, tense and modality: Theory, typology, acquisition*. LOT.

Bowden, H. W. (2016). Assessing second-language oral proficiency for research: The Spanish elicited imitation task. *Studies in Second Language Acquisition*, *38*(4), 647–675. https://doi.org/10.1017/S0272263115000443

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition*, *33*(2), 247–271. https://doi.org/10.1017/S0272263110000756

Brinton, L. J. (1988). *The development of English aspectual systems: Aspectualizers and post-verbal particles*. Cambridge University Press.

Brown, J. D. (2002). Do cloze tests work? Or, is it just an illusion? *Second Language Studies*, *21*(1), 79–125.

Brown, V. A. (2021). An introduction to linear mixed-effects modeling in R. *Advances in Methods and Practices in Psychological Science*, *4*(1). https://doi.org/10.1177/2515245920960351

Busch, M. (1993). Using likert scales in L2 research. *TESOL Quarterly*, *27*(4), 733–736. https://doi.org/10.2307/3587408

Bybee, J. L., Perkins, R. D., & Pagliuca, W. (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. University of Chicago Press.

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2012). Does L1 maintenance hamper L2 nativelikeness? A study of L2 ultimate attainment in early bilinguals. *Studies in Second Language Acquisition*, *34*. https://doi.org/10.1017/S0272263112000034

Canty, A., & Ripley, B. (2022). *Boot: Bootstrap R (S-plus) functions* (1.3-28.1) [Computer software]. https://cran.r-project.org/web/packages/boot/boot.pdf

Casasanto, L. S., Hofmeister, P., & Sag, I. A. (2010). Understanding acceptability judgments: Additivity and working memory effects. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 224–229). Cognitive Science Society. https://doi.org/10/papers/0037/ind

Chan, A., Yang, W., Chang, F., & Kidd, E. (2018). Four-year-old Cantonese-speaking children's online processing of relative clauses: A permutation analysis. *Journal of Child Language*, *45*(1), 174–203. https://doi.org/10.1017/S0305000917000198

Chan, H. L. (2012). *Tense-aspect processing in second language learners* [Doctoral Dissertation, University of Pittsburgh]. http://d-scholarship.pitt.edu/12499/

Chen, C. (2009). *The representation and processing of past tense in Chinese English-language learners* [Doctoral Dissertation]. The University of Arizona.

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

Cho, J. (2022). Online processing and offline judgments of L2-English articles. *Linguistic Approaches to Bilingualism*, *12*(3), 280–309. https://doi.org/10.1075/lab.18053.cho

Choi, S. H., & Ionin, T. (2021). Plural marking in the second language: Atomicity, definiteness, and transfer. *Applied Psycholinguistics*, *42*(3), 549–578. https://doi.org/10.1017/S0142716420000569

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Praeger.

Chomsky, N. (1995). *The minimalist program*. The MIT Press.

Clahsen, H. (2007). Psycholinguistic perspectives on grammatical representations. In S. Featherston & W. Sternefeld (Eds.), *Roots: Linguistics in search of its evidential base* (pp. 97–132). Mouton de Gruyter.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*(1), 3–42. https://doi.org/10.1017/S0142716406060024

Clahsen, H., & Felser, C. (2018). Some notes on the shallow structure hypothesis. *Studies in Second Language Acquisition*, *40*(3), 693–706. https://doi.org/10.1017/S0272263117000250

Clahsen, H., Felser, C., Neubauer, K., Sato, M., & Silva, R. (2010). Morphological structure in native and nonnative language processing. *Language Learning*, *60*(1), 21–43. https://doi.org/10.1111/j.1467-9922.2009.00550.x

Clahsen, H., Hong, U., & Sonnenstuhl-Henning, I. (1995). Grammatical constraints in syntactic processing: Sentence-matching experiments on German. *The Linguistic Review*, *12*(1), 5–34. https://doi.org/10.1515/tlir.1995.12.1.5

Clahsen, H., Martzoukou, M., & Stavrakaki, S. (2010). The perfective past tense in Greek as a second language. *Second Language Research*, *26*(4), 501–525. https://doi.org/10.1177/0267658310373880

Clark, E. V. (1993). *The lexicon in acquisition* (pp. xii, 306). Cambridge University Press. https://doi.org/10.1017/CBO9780511554377

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*(12), 997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Collins, L. (2004). The particulars on universals: A comparison of the acquisition of tense-aspect morphology among Japanese- and French-speaking learners of English. *The Canadian Modern Language Review*, *61*(2), 251–274. https://doi.org/10.3138/cmlr.61.2.251

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge University Press.

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

Conklin, K. author, Pellicer-Sanchez, A. author, & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research*. Cambridge University Press. https://doi.org/10.1017/9781108233279

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Corps, R. E., Brooke, C., & Pickering, M. J. (2022). Prediction involves two stages: Evidence from visual-world eye-tracking. *Journal of Memory and Language*, *122*, 104298. https://doi.org/10.1016/j.jml.2021.104298

Crain, S., & Fodor, J. D. (1987). Sentence matching and overgeneration. *Cognition*, *26*(2), 123–169.

Cristante, V. (2020). Sentence imitations in monolingual and early second language learners and the implications for grammar teaching. *Pedagogical Linguistics*, *1*(2), 149–183. https://doi.org/10.1075/pl.20001.cri

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge. http://site.ebrary.com/lib/york/Doc?id=10545628

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382. https://doi.org/10.1177/0267658312443651

Cunnings, I., & Finlayson, I. (2015). Mixed effects modeling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159–181). Routledge.

Cunnings, I., Fotiadou, G., & Tsimpli, I. (2017). Anaphora resolution and reanalysis during L2 sentence processing: Evidence from the visual world paradigm. *Studies in Second Language Acquisition*, *39*(4), 621–652. https://doi.org/10.1017/S0272263116000292

Dahl, Ö. (1985). *Tense and aspect systems*. Blackwell.

Dahl, Ö. (2022). Perfects across languages. *Annual Review of Linguistics*, *8*(1), 279–297. https://doi.org/10.1146/annurev-linguistics-031120-123428

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge University Press.

Declerck, R. (1986). From Reichenbach (1947) to Comrie (1985) and beyond: Towards a theory of tense. *Lingua*, *70*(4), 305–364. https://doi.org/10.1016/0024-3841(86)90048-3

Declerck, R. (1991). *Tense in English: Its structure and use in discourse*. Routledge. https://www.routledge.com/Tense-in-English-Its-Structure-and-Use-in-Discourse/Declerck/p/book/9781138919372

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

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*(S1), 1–25. https://doi.org/10.1111/j.0023-8333.2005.00294.x

Deo, A. (2012). Morphology. In R. I. Binnick (Ed.), *The Oxford handbook of tense and aspect* (pp. 155–183). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195381979.013.0005

Diaubalick, T., & Guijarro Fuentes, P. (2017). L1 effects as manifestations of individual differences in the L2 acquisition of the Spanish tense-aspect-system. In K. Bellamy, M. W. Child, P. González, A. Muntendam, & M. C. Parafita Couto (Eds.), *Multidisciplinary approaches to bilingualism in the Hispanic and Lusophone world* (pp. 9–40). John Benjamins Publishing Company. https://doi.org/10.1075/ihll.13.02dia

Diaubalick, T., & Guijarro-Fuentes, P. (2019). The strength of L1 effects on tense and aspect: How German learners of L2 Spanish deal with acquisitional problems. *Language Acquisition*, *26*(3), 282–301. https://doi.org/10.1080/10489223.2018.1554663

Dietrich, R., Klein, W., & Noyau, C. (1995). *The acquisition of temporality in a second language*. Benjamins.

Dik, S. C. (1997). *The Theory of Functional Grammar: The structure of the clause*. Walter de Gruyter.

Domínguez, L., Arche, M. J., & Myles, F. (2017). Spanish imperfect revisited: Exploring L1 influence in the reassembly of imperfective features onto new L2 forms. *Second Language Research*, *33*(4), 431–457. https://doi.org/10.1177/0267658317701991

Dowens, M. G., Vergara, M., Barber, H. A., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, *22*(8), 1870–1887. https://doi.org/10.1162/jocn.2009.21304

Dowty, D. R. (1979). *Word meaning and Montague grammar: The semantics of verbs and times in generative semantics and in Montague's PTQ*. D. Reidel.

Dudley, A., & Slabakova, R. (2020). Aspectual contrasts in the English present tense revisited: Exploring the role of input and L1 influence. *Pedagogical Linguistics*, *1*(1), 66–93. https://doi.org/10.1075/pl.19010.dud

Duff, P. A., & Li, D. (2002). The acquisition and use of perfective aspect in Mandarin. In M. R. Salaberry & Y. Shirai (Eds.), *The L2 acquisition of tense–aspect morphology* (Vol. 27, pp. 417–453). John Benjamins Publishing Company. https://doi.org/10.1075/lald.27

Duffield, N., Matsuo, A., & Roberts, L. (2007). Acceptable ungrammaticality in sentence matching. *Second Language Research*, *23*(2), 155–177. https://doi.org/10.1177/0267658307076544

Duffield, N., & White, L. (1999). Assessing L2 knowledge of Spanish clitic placement: Converging methodologies. *Second Language Research*, *15*(2), 133–160. https://doi.org/10.1191/026765899668237583

Duffield, N., White, L., Garavito, J. B. D., Montrul, S., & Prévost, P. (2002). Clitic placement in L2 French: Evidence from sentence matching. *Journal of Linguistics*, *38*(3), 487–525. https://doi.org/10.1017/S0022226702001688

Eisenstein, M., Bailey, N., & Madden, C. (1982). It take two: Contrasting tasks and contrasting structures. *TESOL Quarterly*, *16*(3), 381–393. https://doi.org/10.2307/3586637

Ellert, M. (2013). Resolving ambiguous pronouns in a second language: A visual-world eye-tracking study with Dutch learners of German. *International Review of Applied Linguistics in Language Teaching*, *51*(2). https://doi.org/10.1515/iral-2013-0008

Ellis, N. C. (2008). Implicit and explicit Knowledge about language. In J. Cenoz & N. H. Hornberger (Eds.), *Encyclopedia of language and education: Vol. 6. Knowledge about language* (pp. 1–13). Springer Academic.

Ellis, N. C. (2022). Second language learning of morphology. *Journal of the European Second Language Association*, *6*(1), Article 1. https://doi.org/10.22599/jesla.85

Ellis, R. (1991). Grammatically judgments and second language acquisition. *Studies in Second Language Acquisition*, *13*(2), 161–186. https://doi.org/10.1017/S0272263100009931

Ellis, R. (2005). Measuring implicit and explicit knowlegde of a secong language: A psychometric study. *Studies in Second Language Acquisition*, *27*(2), 141–172. https://doi.org/10.1017/S0272263105050096

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics*, *27*(3), 431–463. https://doi.org/10.1093/applin/aml022

Ellis, R. (2009). Measuring implicit and explicit knolwedge of a second language. In S. Loewen, C. Elder, H. Reinders, R. Erlam, J. Philp, & R. Ellis (Eds.), *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching* (pp. 31–64). Multilingual Matters. https://doi.org/10.21832/9781847691767

Ellis, R., & Loewen, S. (2007). Confirming the operational definitions of explicit and implicit knowledge in Ellis (2005): Responding to Isemonger. *Studies in Second Language Acquisition*, *29*(1), 119–126. https://doi.org/10.1017/S0272263107070052

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, *19*(4), 677–714. https://doi.org/10.1017/S0140525X00043521

Erica, L. Y., & Wong, B. E. (2016). Acquisition of the present perfect and the simple past by Malaysian Chinese ESL learners. *Pertanika Journal of Social Science and Humanities*, *24*(2), 885–906.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics*, *27*(3), 464–491. https://doi.org/10.1093/applin/aml001

Erlam, R. (2009). The elicited oral imitation test as a measure of implicit knowledge. In D. Singleton (Ed.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 65–93). Multilingual Matters.

Erlam, R., & Ellis, R. (2018). Task-based language teaching for beginner-level learners of L2 French: An exploratory study. *The Canadian Modern Language Review*, *74*(1), 1–26. https://doi.org/10.3138/cmlr.3831

Erlam, R., & Wei, L. (2021). The importance of increased processing demands in the design of elicited imitation tests. *Language Teaching Research*, *0*(0), 1–29. https://doi.org/10.1177/13621688211026032

Eubank, L. (1993). Sentence matching and processing in L2 development. *Second Language Research*, *9*(3), 253–280. https://doi.org/10.1177/026765839300900303

Farina, C. J. (2017). *Adult L2 processing and acquisition of the English present perfect* [Doctoral Dissertation]. University of South Carolina.

Fassi Fehri, A. (2003). Arabic perfect and temporal adverbs. In A. Alexiadou, M. Rathert, & A. von Stechow (Eds.), *Perfect explorations* (pp. 69–100). De Gruyter Mouton.

Flashner, V. E. (1989). Transfer of aspect in the English oral narratives of native Russian speakers. In H. W. Dechert & M. Raupach (Eds.), *Transfer in language production / edited by Hans W. Dechert and Manfred Raupach.* (pp. 71–97). Ablex Pub. Corp.

Flecken, M., Walbert, K., & Dijkstra, T. (2015). ‘Right now, Sophie ∗swims in the pool?!’: Brain potentials of grammatical aspect processing. *Frontiers in Psychology*, *6*, 1764.

Foote, R. (2011). Integrated knowledge of agreement in early and late English–Spanish bilinguals. *Applied Psycholinguistics*, *32*(1), 187–220. https://doi.org/10.1017/S0142716410000342

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*(3), 379–399. https://doi.org/10.1017/S136672891000012X

Freedman, S. E., & Forster, K. I. (1985). The psychological status of overgenerated sentences. *Cognition*, *19*(2), 101–131. https://doi.org/10.1016/0010-0277(85)90015-0

Frenck-Mestre, C. (2005). Eye-movement recording as a tool for studying syntactic processing in a second language: A review of methodologies and experimental findings. *Second Language Research*, *21*(2), 175–198. https://doi.org/10.1191/0267658305sr257oa

Fuchs, R., Götz, S., & Werner, V. (2016). The present perfect in learner Englishes: A corpus-based case study on L1 German intermediate and advanced speech and writing. In V. Werner, E. Seoane, & C. Suárez-Gómez (Eds.), *Re-assessing the present perfect* (pp. 297–338). De Gruyter Mouton. https://doi.org/10.1515/9783110443530-013

Gabriele, A., & Maekawa, J. (2008). Interpreting tense in a second language. *EUROSLA Yearbook*, *8*, 79–106. https://doi.org/10.1075/eurosla.8.07gab

Gabriele, A., Martohardjono, G., & McClure, W. (2005). Evaluating the role of the L1 in the L2 acquisition of aspect: A study of Japanese learners of English. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *Proceedings of the International Symposium on Biligualism* (pp. 808–826). Cascadilla Press.

Gabriele, A., & McClure, W. (2011). Why some imperfectives are interpreted imperfectly: A study of Chinese learners of Japanese. *Language Acquisition*, *18*(1), 39–83. https://doi.org/10.1080/10489223.2011.530515

Gaillard, S., & Tremblay, A. (2016). Linguistic proficiency assessment in second language acquisition research: The elicited imitation task. *Language Learning*, *66*(2), 419–447. https://doi.org/10.1111/lang.12157

Gallimore, R., & Tharp, R. G. (1981). The interpretation of elicited sentence imitation in a standardized context. *Language Learning*, *31*(2), 369–392. https://doi.org/10.1111/j.1467-1770.1981.tb01390.x

García-Tejada, A., Cuza, A., & Lustres Alonso, E. G. (2023). The production and comprehension of Spanish se use in L2 and heritage Spanish. *Second Language Research*, *39*(2), 301–331. https://doi.org/10.1177/02676583211016537

Gass, S. M. (2001). Sentence matching: A re-examination. *Second Language Research*, *17*(4), 421–441. https://doi.org/10.1177/026765830101700407

Giacalone Ramat, A., & Banfi, E. (1990). The acquisition of temporality: A second language perspective. *Folia Linguistica*, *24*(3–4), 405–428. https://doi.org/10.1515/flin.1990.24.3-4.405

Goad, H., & White, L. (2006). Ultimate attainment in interlanguage grammars: A prosodic approach. *Second Language Research*, *22*(3), 243–268. https://doi.org/10.1191/0267658306sr268oa

Goad, H., White, L., & Steele, J. (2003). Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations? *Canadian Journal of Linguistics*, *48*(2), 243–263. https://doi.org/10.1353/cjl.2004.0027

Godfroid, A. (2019). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide* (1st ed.). Routledge. https://doi.org/10.4324/9781315775616

Godfroid, A., Loewen, S., Jung, S., Park, J.-H., Gass, S., & Ellis, R. (2015). Timed and untimed grammaticality measure distinct types of knowledge: Evidence from eye-movement patterns. *Studies in Second Language Acquisition*, *37*(2), 269–297. https://doi.org/10.1017/S0272263114000850

González, P., & Quintana-Hernández, L. (2018). Inherent aspect and L1 transfer in the L2 acquisition of Spanish grammatical aspect. *The Modern Language Journal*, *102*(3), 611–625. https://doi.org/10.1111/modl.12502

Graham, C. R., McGhee, J., & Millard, B. (2008). The role of lexical choice in elicited imitation item difficulty. In M. T. Prior, Y. Watanabe, & S. Lee (Eds.), *Proceedings of the 2008 Second Language Research Forum*. Casadila Press.

Granena, G. (2016). Elicited imitation as a measure of implicit L2 knowledge: The role of working memory and short-term memory capacity. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 112–124). John Benjamins Publishing Company.

Grüter, T., Lew-Williams, C., & Fernald, A. (2012). Grammatical gender in L2: A production or a real-time processing problem? *Second Language Research*, *28*(2), 191–215. https://doi.org/10.1177/0267658312437990

Gujord, A.-K. H. (2013). *Grammatical encoding of past time in L2 Norwegian* [Doctoral Dissertation]. The University of Bergen.

Gutiérrez, X. (2013). The construct validity of grammaticality judgment tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition*, *35*(3), 423–449. https://doi.org/10.1017/S0272263113000041

Gvozdanović, J. (2012). Perfective and imperfective aspect. In R. I. Binnick (Ed.), *The Oxford handbook of tense and aspect* (pp. 781–802). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195381979.013.0027

Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (Vol. 70, Issue 4, pp. 111–176). https://doi.org/10.2307/416329

Hamayan, E., Saegert, J., & Larudee, P. (1977). Elicited imitation in second language learners. *Language and Speech*, *20*(1), 86–97. https://doi.org/10.1177/002383097702000109

Hanania, E., & Shikhani, M. (1986). Interrelationships among three tests of language proficiency: Standardized ESL, cloze, and writing. *TESOL Quarterly*, *20*(1), 97–109. https://doi.org/10.2307/3586391

Hawkins, R. (2001). *Second language syntax: A generative introduction*. Blackwell.

Hawkins, R., Casillas, G., Hattori, H., Hawthorne, J., Husted, R., Lozano, C., Okamoto, A., Thomas, E., & Yamada, K. (2008). The semantic effects of verb raising and its consequences in second language grammars. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (pp. 328–351). Lawrence Erlbaum.

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. https://doi.org/10.1191/026765897671476153

Hawkins, R., & Hattori, H. (2006). Interpretation of English multiple wh-questions by Japanese speakers: A missing uninterpretable feature account. *Second Language Research*, *22*(3), 269–301. https://doi.org/10.1191/0267658306sr269oa

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. J. Towell (Eds.), *The lexicon–syntax interface in second language acquisition* (pp. 21–44). John Benjamins Publishing Company.

Hazendar, B., & Schwartz, B. D. (1997). Are there optional infinitives in child L2 acquisition? In E. Hughes, M. Hughes, & A. Greenhill (Eds.), *Proceedings of the 21st Annual Boston University Conference on Language Development* (pp. 257–268). Cascadilla Proceedings Project.

Hendriks, P., & Koster, C. (2010). Production/comprehension asymmetries in language acquisition. *Lingua*, *120*(8), 1887–1897. https://doi.org/10.1016/j.lingua.2010.02.002

Hewson, J. (2012). Tense. In R. I. Binnick (Ed.), *The Oxford handbook of tense and aspect* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195381979.013.0017

Hopp, H. (2006). Syntactic features and reanalysis in near-native processing. *Second Language Research*, *22*(3), 369–397. https://doi.org/10.1191/0267658306sr272oa

Hopp, H. (2009). The syntax–discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, *12*(4), 463–483. https://doi.org/10.1017/S1366728909990253

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, *120*(4), 901–931. https://doi.org/10.1016/j.lingua.2009.06.004

Hopp, H. (2013). Grammatical gender in adult L2 acquisition: Relations between lexical and syntactic variability. *Second Language Research*, *29*(1), 33–56. https://doi.org/10.1177/0267658312461803

Housen, A. (2002). The development of tense-aspect in English as a second language and the variable influence of inherent aspect. In M. R. Salaberry & Y. Shirai (Eds.), *The L2 acquisition of tense–aspect morphology* (Vol. 27, pp. 155–197). John Benjamins Publishing Company. https://doi.org/10.1075/lald.27

Huang, Y., & Snedeker, J. (2020). Evidence from the visual world paradigm raises questions about unaccusativity and growth curve analyses. *Cognition*, *200*, 104251. https://doi.org/10.1016/j.cognition.2020.104251

Hubers, F., Redl, T., De Vos, H., Reinarz, L., & De Hoop, H. (2020). Processing prescriptively incorrect comparative particles: Evidence from sentence-matching and eye-tracking. *Frontiers in Psychology*, *11*, 186. https://doi.org/10.3389/fpsyg.2020.00186

Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, *8*(3), 229–249. https://doi.org/10.1080/15434303.2011.565844

Hulstijn, J. H. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning: The virtues and pitfalls of a two-system view. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 25–46). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.48.02hul

Hundt, M., & Smith, N. (2009). The present perfect in British and American English: Has there been any change, recently? *International Computer Archive of Modern and Medieval English*, *33*, 45–63.

Hyltenstam, K. (1983). Data types and second language variability. In H. Ringbom (Ed.), *Psycholinguistics and Foreign Language Learning. Papers from a Conference (Stockholm, Sweden and Abo, Finland, October 25-26, 1982).* (pp. 57–74). Research Institute of the Abo Akademi Foundation. https://eric.ed.gov/?id=ED276313

Ionin, T. (2012). Morphosyntax. In J. Herschensohn & M. Young-Scholten (Eds.), *The Cambridge handbook of second language acquisition* (pp. 505–528). Cambridge University Press. https://doi.org/10.1017/CBO9781139051729.030

Ionin, T. (2023). Online vs. Offline sensitivity to missing plural marking in L2 English. In P. Gappmayr & J. Kellogg (Eds.), *Proceedings of the 47th annual Boston University Conference on Language Development* (pp. 366–379). Cascadilla Proceedings Project.

Ionin, T., Choi, S. H., & Liu, Q. (2021). Knowledge of indefinite articles in L2-English: Online vs. offline performance. *Second Language Research*, *37*(1), 121–160. https://doi.org/10.1177/0267658319857466

Ionin, T., & Zyzik, E. (2014). Judgment and interpretation tasks in second language research. *Annual Review of Applied Linguistics*, *34*, 37–64. https://doi.org/10.1017/S0267190514000026

Ito, A., & Knoeferle, P. (2022). Analysing data from the psycholinguistic visual-world paradigm. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01969-3

Ito, A., Nguyen, H. T. T., & Knoeferle, P. (2023). German-dominant Vietnamese heritage speakers use semantic constraints of German for anticipation during comprehension in Vietnamese. *Bilingualism: Language and Cognition*, 1–18. https://doi.org/10.1017/S136672892300041X

Izquierdo, J., & Collins, L. (2008). The facilitative role of L1 Influence in tense-aspect marking: A comparison of Hispanophone and Anglophone learners of French. *The Modern Language Journal*, *92*(3), 350–368. https://doi.org/10.1111/j.1540-4781.2008.00751.x

Jaeger, T. 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. https://doi.org/10.1016/j.jml.2007.11.007

Jegerski, J. (2014). Self-paced reading. In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 20–49). Routledge.

Jegerski, J. (2015). The processing of case in near-native Spanish. *Second Language Research*, *31*(3), 281–307. https://doi.org/10.1177/0267658314563880

Jensen, E. D., & Vinther, T. (2003). Exact repetition as input enhancement in second language acquisition. *Language Learning*, *53*(3), 373–428. https://doi.org/10.1111/1467-9922.00230

Jessop, L., Suzuki, W., & Tomita, Y. (2007). Elicited Imitation in second language acquisition research. *The Canadian Modern Language Review*, *64*(1), 215–238. https://doi.org/10.3138/cmlr.64.1.215

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*(4), 603–634. https://doi.org/10.1017/S0142716404001298

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning*, *57*(1), 1–33. https://doi.org/10.1111/j.1467-9922.2007.00397.x

Jiang, N. (2011). *Conducting Reaction Time Research in Second Language Studies* (0 ed.). Routledge. https://doi.org/10.4324/9780203146255

Jiang, N. (2012). *Conducting reaction time research in second language studies*. Routledge.

Jiang, N., Hu, G., Chrabaszcz, A., & Ye, L. (2017). The activation of grammaticalized meaning in L2 processing: Toward an explanation of the morphological congruency effect. *International Journal of Bilingualism*, *21*(1), 81–98. https://doi.org/10.1177/1367006915603823

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, *61*(3), 940–967. https://doi.org/10.1111/j.1467-9922.2010.00627.x

Jonz, J. (1990). Another turn in the conversation: What does cloze measure? *TESOL Quarterly*, *24*(1), 61–83. https://doi.org/10.2307/3586852

Kahoul, W. (2014). *Arabic and Chinese learners' production, perception and processing of past tense and verbal agreement morphology in L2 English* [Doctoral Dissertation]. Newcastle University.

Keating, G. D. (2009). Sensitivity to Violations of Gender Agreement in Native and Nonnative Spanish: An Eye-Movement Investigation. *Language Learning*, *59*(3), 503–535. https://doi.org/10.1111/j.1467-9922.2009.00516.x

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition*, *37*(1), 1–32. https://doi.org/10.1017/S0272263114000187

Kihlstedt, M. (2002). Reference to past events in dialogue. In M. R. Salaberry & Y. Shirai (Eds.), *The L2 acquisition of tense–aspect morphology* (Vol. 27, pp. 323–362). John Benjamins Publishing Company. https://doi.org/10.1075/lald.27

Kim, E., Montrul, S., & Yoon, J. (2015). The on-line processing of binding principles in second language acquisition: Evidence from eye tracking. *Applied Psycholinguistics*, *36*(6), 1317–1374. https://doi.org/10.1017/S0142716414000307

Kim, H., & Grüter, T. (2021). Predictive processing of implicit causality in a second language: A visual-world eye-tracking study. *Studies in Second Language Acquisition*, *43*(1), 133–154. https://doi.org/10.1017/S0272263120000443

Kim, J., & Nam, H. (2017). Measures of implicit knowledge revisited: Processing modes, time pressure, and modality. *Studies in Second Language Acquisition*, *39*(3), 431–457. https://doi.org/10.1017/S0272263115000510

Kim, M. H., & Ionin, T. (2022). Potential L1 transfer effects in explicit and implicit knowledge of articles in L2 English. *Australian Review of Applied Linguistics*. https://doi.org/10.1075/aral.21010.kim

Kim, Y., Tracy-Ventura, N., & Jung, Y. (2016). A measure of proficiency or short-term memory? Validation of an elicited imitation test for SLA research. *The Modern Language Journal*, *100*(3), 655–673. https://doi.org/10.1111/modl.12346

Klein, W. (1992). The present perfect puzzle. *Language*, *68*(3), 525–552. https://doi.org/10.2307/415793

Klein, W. (1994). *Time in language*. Routledge. https://doi.org/10.4324/9781315003801

Klein, W. (1995). The acquisition of English. In R. Dietrich, W. Klein, & C. Noyau (Eds.), *The acquisition of temporality in a second language* (pp. 31–70). Benjamins.

Klein, W., Li, P., & Hendriks, H. (2000). Aspect and assertion in Mandarin Chinese. *Natural Language & Linguistic Theory*, *18*, 723–770.

Kleinmann, H. H. (1977). Avoidance behavior in adult second language acquisition. *Language Learning*, *27*(1), 93–107.

Koch, E. M., Bulté, B., Housen, A., & Godfroid, A. (2021). Using verb morphology to predict subject number in L1 and L2 sentence processing: A visual-world eye-tracking experiment. *Journal of the European Second Language Association*, *5*(1), Article 1. https://doi.org/10.22599/jesla.79

Kostromitina, M., & Plonsky, L. (2022). Elicited imitation tasks as a measure of L2 proficiency: A meta-analysis. *Studies in Second Language Acquisition*, *44*(3), 886–911. https://doi.org/10.1017/S0272263121000395

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26. https://doi.org/10.18637/jss.v082.i13

Labeau, E. (2005). Beyond the Aspect Hypothesis: Tense–aspect development in advanced L2 French. *EUROSLA Yearbook*, *5*(1), 77–101. https://doi.org/10.1075/eurosla.5.06lab

Lago, S., Stone, K., Oltrogge, E., & Veríssimo, J. (2022). Possessive processing in bilingual comprehension. *Language Learning*, *73*(3), 904–941. https://doi.org/10.1111/lang.12556

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

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

Lardiere, D. (2000). Mapping features to form in second language acquisition. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 102–129). Blackwell.

Lardiere, D. (2003a). Revisiting the comparative fallacy: A reply to Lakshmanan and Selinker, 2001. *Second Language Research*, *19*(2), 129–143. https://doi.org/10.1191/0267658303sr216oa

Lardiere, D. (2003b). Second language knowledge of [+/-Past] vs. [+/-Finite]. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)* (pp. 176–189). Cascadilla Proceedings Project.

Lardiere, D. (2007). *Ultimate attainment in second language acquisition: A case study*. Routledge.

Lardiere, D. (2008). Feature assembly in second language acquisition. In J. M. Liceras, H. Zobl, & H. Goodluck (Eds.), *The role of formal features in second language acquisition* (1st ed., pp. 106–140). Lawrence Erlbaum Associates. https://doi.org/10.4324/9781315085340-5

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Second Language Research, 25*(2), 173–227. https://doi.org/10.1177/0267658308100283

Lardiere, D. (2012). Linguistic approaches to second language morphosyntax. In A. Mackey & S. M. Gass (Eds.), *The Routledge handbook of second language acquisition* (1st ed., pp. 106–126). Routledge. https://doi.org/10.4324/9780203808184-10

Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R* (Second Edition). Routledge.

Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning, 65*(S1), 127–159. https://doi.org/10.1111/lang.12115

Lee, E.-K. R., & Phillips, C. (2023). Why non-native speakers sometimes outperform native speakers in agreement processing. *Bilingualism: Language and Cognition, 26*(1), 152–164. https://doi.org/10.1017/S1366728922000414

Leech, G. N. (1987). *Meaning and the English verb* (2nd ed.). Longman.

Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2022). *Emmeans: Estimated marginal means, aka least-squares means* (1.7.2) [Computer software]. https://cran.r-project.org/web/packages/emmeans/index.html

Levelt, W. J. M. (1989). *Speaking: From intention to articulation* (pp. xiv, 566). The MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *The Behavioral and Brain Sciences*, *22*(1), 1–75. https://doi.org/10.1017/s0140525x99001776

Lew-Williams, C., & Fernald, A. (2007). How first and second language learners use predictive cues in online sentence interpretation in Spanish and English. In H. Caunt-Nulton, S. Kulatilake, & I. Woo (Eds.), *Proceedings of the 31st Annual Boston University Conference on Language Development* (Vol. 2, pp. 382–393). Cascadilla Proceedings Project. https://babylab.princeton.edu/sites/default/files/papers/Lew-WilliamsFernald2007-BU.pdf

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464. https://doi.org/10.1016/j.jml.2010.07.003

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. University of California Press.

Li, C. N., Thompson, S. A., & Thompson, M. R. (1982). The discourse motivation for the perfect aspect: The Mandarin particle LE. In P. J. Hopper (Ed.), *Typological studies in language* (Vol. 1, pp. 19–44). John Benjamins Publishing Company. https://doi.org/10.1075/tsl.1.05li

Li, P., & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. Mouton de Gruyter. https://doi.org/10.1515/9783110800715

Li, Y., Casaponsa, A., Jones, M., & Thierry, G. (2023). Chinese learners of English are conceptually blind to temporal differences conveyed by tense. *Language Learning*, *74*(1), 184–217. https://doi.org/10.1111/lang.12584

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

Lin, J.-W. (2006). Time in a language without tense: The case of Chinese. *Journal of Semantics*, *23*(1), 1–53. https://doi.org/10.1093/jos/ffh033

Lin, J.-W. (2012). Tenselessness. In R. I. Binnick (Ed.), *The Oxford handbook of tense and aspect* (pp. 669–695). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195381979.013.0023

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research: Mixed-effects models. *Language Learning*, *65*(S1), 185–207. https://doi.org/10.1111/lang.12117

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Lindstedt, J. (2000). The perfect – aspectual, temporal and evidential. In Ö. Dahl (Ed.), *Tense and aspect in the languages of Europe* (pp. 365–383). Mouton de Gruyter.

Lindstedt, J. (2001). Tense and aspect. In *Language typology and language universals: An international handbook* (Vol. 1, pp. 768–783). Walter de Gruyter.

Liszka, S. A. (2004). Exploring the effects of first language influence on second language pragmatic processes from a syntactic deficit perspective. *Second Language Research*, *20*(3), 212–231. https://doi.org/10.1191/0267658304sr238oa

Liszka, S. A. (2009). Associating meaning to form in advanced L2 speakers: An investigation into the acquisition of the English present simple and present progressive. In N. Snape, Y. I. Leung, & M. Sharwood Smith (Eds.), *Representational deficits in SLA:*

Studies in honor of Roger Hawkins (pp. 229–246). John Benjamins Publishing Company.

Liszka, S. A. (2015). The L2 acquisition of the English present simple – present progressive distinction: Verb-raising revisited. In D. Ayoun (Ed.), *The acquisition of the present* (pp. 57–86). John Benjamins Publishing Company. https://doi.org/10.1075/z.196.03lis

Liu, M. (2015). Tense and aspect in Mandarin Chinese. In W. S.-Y. Wang & C. Sun (Eds.), *The Oxford handbook of Chinese linguistics* (pp. 274–289). Oxford University Press.

Loewen, S. (2009). Grammaticality judgment tests and the measurement of implicit and explicit L2 Knowledge. In D. Singleton (Ed.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94–112). Multilingual Matters. https://doi.org/10.21832/9781847691767-006

Lyons, J. (1968). *Introduction to theoretical linguistics*. Cambridge University Press.

Ma, L., Martohardjono, G., & McClure, W. (2022). The functional roles of lexical devices in second language learners' encoding of temporality: A study of Mandarin Chinese-speaking ESL learners. *International Review of Applied Linguistics in Language Teaching*, *60*(3), 563–597. https://doi.org/10.1515/iral-2018-0298

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Erlbaum. https://doi.org/10.4324/9781410612564

Mackey, A., & Gass, S. M. (2011). *Research methods in second language acquisition: A practical guide* (Vol. 13). Wiley-Blackwell. https://doi.org/10.1002/9781444347340

Macrory, G., & Stone, V. (2000). Pupil progress in the acquisition of the perfect tense in French: The relationship between knowledge and use. *Language Teaching Research*, *4*(1), 55–82. https://doi.org/10.1177/136216880000400104

MacWhinney, B. (2005). Extending the competition model. *International Journal of Bilingualism*, *9*(1), 69–84. https://doi.org/10.1177/13670069050090010501

Madden, C. J., & Zwaan, R. A. (2003). How does verb aspect constrain event representations? *Memory & Cognition*, *31*(5), 663–672. https://doi.org/10.3758/BF03196106

Mair, C. (2012). Progressive and continuous aspect. In R. I. Binnick (Ed.), *The Oxford handbook of tense and aspect* (pp. 803–827). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195381979.013.0028

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marimon, M., Höhle, B., & Langus, A. (2022). Pupillary entrainment reveals individual differences in cue weighting in 9-month-old German-learning infants. *Cognition*, *224*, 105054. https://doi.org/10.1016/j.cognition.2022.105054

Marinis, T. (2003). Psycholinguistic techniques in second language acquisition research. *Second Language Research*, *19*(2), 144–161. https://doi.org/10.1191/0267658303sr217ra

Marinis, T., & Cunnings, I. (2018). Using psycholinguistic techniques in a second language teaching setting. In C. Wright, T. Piske, & M. Young-Scholten (Eds.), *Mind matters in SLA* (pp. 185–202). Multilingual Matters. https://doi.org/10.21832/9781788921626-012

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Markman, B. R., Spilka, I. V., & Tucker, G. R. (1975). The use of elicited imitation in search of an interim French grammar. *Language Learning*, *25*(1), 31–41. https://doi.org/10.1111/j.1467-1770.1975.tb00107.x

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics*, *39*(5), 861–904. https://doi.org/10.1017/S0142716418000036

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McCoard, R. W. (1978). *The English perfect: Tense-choice and pragmatic inferences*. North-Holland Publishing Company.

McCray, G., & Brunfaut, T. (2018). Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. *Language Testing*, *35*(1), 51–73. https://doi.org/10.1177/0265532216677105

McDade, H. L., Simpson, M. A., & Lamb, D. E. (1982). The use of elicited imitation as a measure of expressive grammar: A question of validity. *The Journal of Speech and Hearing Disorders*, *47*(1), 19–24. https://doi.org/10.1044/jshd.4701.19

McManus, K. (2015). L1-L2 differences in the acquisition of form-meaning pairings in a second language. *The Canadian Modern Language Review*, *71*(2), 155–181. https://doi.org/10.3138/cmlr.2070.51

Mercier, J., Pivneva, I., & Titone, D. (2016). The role of prior language context on bilingual spoken word processing: Evidence from the visual world task. *Bilingualism: Language and Cognition*, *19*(2), 376–399. https://doi.org/10.1017/S1366728914000340

Minor, S., Mitrofanova, N., Guajardo, G., Vos, M., & Ramchand, G. (2023). Aspect processing across languages: A visual world eye-tracking study. *Frontiers in Language Sciences*, *1*. https://www.frontiersin.org/articles/10.3389/flang.2022.1052205

Minor, S., Mitrofanova, N., & Ramchand, G. (2022). Fine-grained time course of verb aspect processing. *PLOS ONE*, *17*(2). https://doi.org/10.1371/journal.pone.0264132

Montrul, S., Foote, R., & Perpiñán, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, *58*(3), 503–553. https://doi.org/10.1111/j.1467-9922.2008.00449.x

Mornati, G., Riva, V., Vismara, E., Molteni, M., & Cantiani, C. (2022). Infants aged 12 months use the gender feature in determiners to anticipate upcoming words: An eye-tracking study. *Journal of Child Language*, 1–19. https://doi.org/10.1017/S030500092200006X

Mudhsh, B. A. D. M. (2021). A comparative study of tense and aspect categories in Arabic and English. *Cogent Arts & Humanities*, *8*(1), 1899568. https://doi.org/10.1080/23311983.2021.1899568

Munnich, E., Flynn, S., & Martohardjono, G. (1994). Elicited imitation and grammaticality judgment tasks: What they measure and how they relate to each other. In E. Tarone, S. M. Gass, & A. D. Cohen (Eds.), *Research methodology in second-language acquisition* (pp. 227–245). L. Erlbaum Associates.

Muroya, A. (2019). L1 transfer in L2 acquisition of English verbal morphology by Japanese young instructed learners. *Languages*, *4*(1), Article 1. https://doi.org/10.3390/languages4010001

Murphy, V. A. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research*, *13*(1), 34–65. https://doi.org/10.1191/026765897671676818

Nizonkiza, D. (2012). Quantifying controlled productive knowledge of collocations across proficiency and word frequency levels. *Studies in Second Language Learning and Teaching*, *2*(1), Article 1. https://doi.org/10.14746/ssllt.2012.2.1.4

Oller, J. W., & Conrad, C. A. (1971). The Cloze Technique and ESL Proficiency. *Language Learning*, *21*(2), 183–194. https://doi.org/10.1111/j.1467-1770.1971.tb00057.x

Oller, J. W., & Inal, N. (1971). A cloze test of English prepositions. *TESOL Quarterly*, *5*(4), 315–326. https://doi.org/10.2307/3585498

Oller, J. W., & Jonz, J. (1994). Why cloze procedure? In J. W. Oller & J. Jonz (Eds.), *Cloze and coherence* (pp. 1–20). Bucknell University Press. https://doi.org/10.13140/RG.2.1.1835.6969

Orfitelli, R., & Polinsky, M. (2017). When performance masquerades as comprehension: Grammaticality judgments in non-native speakers. In M. Kopotev, O. Lyashevskaya, & A. Mustajoki (Eds.), *Quantitative approaches to the Russian language* (1st ed., pp. 197–214). Routledge. https://doi.org/10.4324/9781315105048

Ouali, H. (2017). The syntax of tense in Arabic. In E. Benmamoun & R. Bassiouney (Eds.), *The routledge handbook of Arabic linguistics* (1st ed., pp. 89–103). Routledge. https://doi.org/10.4324/9781315147062

Ouali, H. (2022). The imperfective verb and the progressive aspect in Arabic. In A.-K. Ali & A. Hachimi (Eds.), *Perspectives on Arabic linguistics XXXIII* (Vol. 11, pp. 121–146). John Benjamins Publishing Company. https://doi.org/10.1075/sal.11

Perdomo, M., & Kaan, E. (2021). Prosodic cues in second-language speech processing: A visual world eye-tracking study. *Second Language Research*, *37*(2), 349–375. https://doi.org/10.1177/0267658319879196

Plonsky, L. (2012). Effect size. In P. J. Robinson (Ed.), *The Routledge encyclopedia of second language acquisition* (pp. 200–202). Routledge.

Plonsky, L. (2015). Statistical power, p values, descriptive statistics, and effect sizes: A ‘back-to-basics’ approach to advancing quantitative methods in L2 research. In L. Plonsky (Ed.), *Advancing Quantitative Methods in Second Language Research* (1st ed., pp. 23–45). Routledge. https://doi.org/10.4324/9781315870908

Plonsky, L., Marsden, E., Crowther, D., Gass, S. M., & Spinner, P. (2020). A methodological synthesis and meta-analysis of judgment tasks in second language research. *Second Language Research*, *36*(4), 583–621. https://doi.org/10.1177/0267658319828413

Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 Research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Porretta, V., Kyröläinen, A.-J., Rij, J., & Järvikivi, J. (2017). *Visual world paradigm data: From preprocessing to nonlinear time-course analysis*. 268–277. https://doi.org/10.1007/978-3-319-59424-8_25

Porretta, V., Kyröläinen, A.-J., Rij, J. van, & Järvikivi, J. (2020). *VWPre: Tools for preprocessing visual world data* (1.2.4) [Computer software]. https://cran.r-project.org/web/packages/VWPre/index.html

Prévost, P., & White, L. (2000a). Accounting for morphological variation in second language acquisition: Truncation or missing inflection? In M.-A. Friedemann & L. Rizzi (Eds.), *The acquisition of syntax: Studies in comparative developmental linguistics* (1st ed., pp. 202–235). Routledge. https://doi.org/10.4324/9781315839899

Prévost, P., & White, L. (2000b). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, *16*(2), 103–133. https://doi.org/10.1191/026765800677556046

Prior, A. N. (1967). *Past, present and future*. Clarendon Press.

Quick Placement Test. (2001). Oxford University Press.

R Core Team. (2019). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Radford, A. (2004). *Minimalist syntax: Exploring the structure of English*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511811319

Rastelli, S. (2020). *The acquisition of aspect in a second language*. Cambridge University Press. https://doi.org/10.1017/9781108903455

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology (2006)*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595–626. https://doi.org/10.1111/lang.12010

Rebuschat, P., & Mackey, A. (2013). Prompted production. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Wiley-Blackwell. https://doi.org/10.1002/9781405198431.wbeal0964

Reichenbach, H. (1947). *Elements of symbolic logic*. The Free Press.

Reichle, R. V., Tremblay, A., & Coughlin, C. (2016). Working memory capacity in L2 processing. *Probus*, *28*(1), 29–55. https://doi.org/10.1515/probus-2016-0003

Révész, A. (2011). Coding second language data validly and reliably. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition* (1st ed., pp. 201–221). Wiley. https://doi.org/10.1002/9781444347340.ch11

Ritz, M.-E. (2012). Perfect tense and aspect. In R. I. Binnick (Ed.), *The Oxford handbook of tense and aspect* (pp. 881–907). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195381979.013.0031

Roberts, L. (2012). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research*, *28*(1), 113–127. https://doi.org/10.1177/0267658311418416

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse—L1 influence and general learner effects. *Studies in Second Language Acquisition*, *30*(3), 333–357. https://doi.org/10.1017/S0272263108080480

Roberts, L., & Liszka, S. A. (2013). Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English. *Second Language Research*, *29*(4), 413–439. https://doi.org/10.1177/0267658313503171

Roberts, L., & Liszka, S. A. (2021). Grammatical aspect and L2 learners' online processing of temporarily ambiguous sentences in English: A self-paced reading study with German, Dutch and French L2 learners. *Second Language Research*, *37*(4), 619–647. https://doi.org/10.1177/0267658319895551

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, *35*(2), 213–235. https://doi.org/10.1017/S0272263112000861

Rothman, J., Bayram, F., Kupisch, T., Lohndal, T., & Westergaard, M. (2018). Generative approaches to second language (L2) acquisition and advanced L2 proficiency. In P. A. Malovrh & A. G. Benati (Eds.), *The handbook of advanced proficiency in second language acquisition* (1st ed., pp. 72–93). Wiley. https://doi.org/10.1002/9781119261650.ch5

Rothman, J., & Slabakova, R. (2018). The generative approach to SLA and its place in modern second language studies. *Studies in Second Language Acquisition*, *40*(2), 417–442. https://doi.org/10.1017/S0272263117000134

Rubin, A. D. (2005). *Studies in semitic grammaticalization*. Brill.

Ryding, K. C. (2005). *A reference grammar of modern standard Arabic*. Cambridge University Press.

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

Salaberry, M. R., & Shirai, Y. (2002). L2 acquisition of tense-aspect morphology. In M. R. Salaberry & Y. Shirai (Eds.), *The L2 Acquisition of Tense-Aspect Morphology* (pp. 1–20). John Benjamins Publishing Company.

Sarandi, H. (2015). Reexamining elicited imitation as a measure of implicit grammatical knowledge and beyond…? *Language Testing*, *32*(4), 485–501. https://doi.org/10.1177/0265532214564504

Sarandi, H. (2020). The structural sensitivity of elicited imitation as a measure of implicit grammatical knowledge. *Eurasian Journal of Applied Linguistics*, *6*(2), 265–284. https://doi.org/10.32601/ejal.775806

Saslow, M. G. (1967). Latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*(8), 1030–1033. https://doi.org/10.1364/JOSA.57.001030

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

Schimke, S. (2011). Variable verb placement in second-language German and French: Evidence from production and elicited imitation of finite and nonfinite negated sentences. *Applied Psycholinguistics*, *32*(4), 635–685. https://doi.org/10.1017/S0142716411000014

Schimke, S., & Dimroth, C. (2018). The influence of finiteness and lightness on verb placement in L2 German: Comparing child and adult learners. *Second Language Research*, *34*(2), 229–256. https://doi.org/10.1177/0267658317723071

Schlüter, N. (2002). Temporal specification of the present perfect: A corpus-based study. In P. Peters, P. Collins, & A. S. Cohen (Eds.), *New frontiers of corpus research* (pp. 307–315). Brill.

Schmiedtová, B., & Flecken, M. (2008). Aspectual concepts across languages: Some considerations for second language learning. In S. De Knop & T. De Rycker (Eds.), *Cognitive approaches to pedagogical grammar* (pp. 357–384). Mouton de Gruyter. https://doi.org/10.1515/9783110205381.3.357

Schütze, C. T., & Sprouse, J. (2013). Judgment data. In R. J. Podesva & D. Sharma (Eds.), *Research methods in linguistics* (pp. 27–50). Cambridge University Press. https://doi.org/10.1017/cbo9781139013734.004

Schwartz, B. D., & Sprouse, R. A. (1994). Word order and nominative case in non-native language acquisition: A longitudinal study of (L1 Turkish) German Interlanguage. In T. Hoekstra & B. D. Schwartz (Eds.), *Language Acquisition and Language Disorders* (Vol. 8, pp. 317–368). John Benjamins Publishing Company. https://doi.org/10.1075/lald.8.14sch

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*(1), 40–72. https://doi.org/10.1177/026765839601200103

Seedorff, M., Oleson, J., & McMurray, B. (2018). Detecting when timeseries differ: Using the Bootstrapped Differences of Timeseries (BDOTS) to analyze visual world paradigm data (and more). *Journal of Memory and Language*, *102*, 55–67. https://doi.org/10.1016/j.jml.2018.05.004

Sekerina, I. A., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of Child Language*, *31*(1), 123–152.

Sharma, D., & Deo, A. (2009). Contact-based aspectual restructuring: A critique of the Aspect Hypothesis. *Queen Mary Occasional Papers Advancing Linguistics (OPALS) Series*, *12*.

Shirai, Y. (1998). Where the progressive and the resultative meet: Imperfective aspect in Japanese, Chinese, Korean and English. *Studies in Language*, *22*(3), 661–692. https://doi.org/10.1075/sl.22.3.06shi

Shirai, Y. (2009). Temporality in first and second language acquisition. In W. Klein & P. Li (Eds.), *The expression of time* (pp. 167–193). Mouton de Gruyter.

Shirai, Y., & Kurono, A. (1998). The acquisition of tense-aspect marking in Japanese as a second language. *Language Learning*, *48*(2), 279–244. https://doi.org/10.1111/1467-9922.00041

Shiu, L.-J., Yalçın, Ş., & Spada, N. (2018). Exploring second language learners' grammaticality judgment performance in relation to task design features. *System*, *72*, 215–225. https://doi.org/10.1016/j.system.2017.12.004

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, *11*(2), 245–260. https://doi.org/10.1017/S1366728908003404

Slabakova, R. (2000). L1 transfer revisited: The L2 acquisition of telicity marking in English by Spanish and Bulgarian native speakers. *Linguistics*, *38*(4), 739–770. https://doi.org/10.1515/ling.2000.004

Slabakova, R. (2002). Recent research on the acquisition of aspect: An embarrassment of riches? *Second Language Research*, *18*(2), 172–188. https://doi.org/10.1191/0267658302sr202ra

Slabakova, R. (2003). Semantic evidence for functional categories in interlanguage grammars. *Second Language Research*, *19*(1), 42–75. https://doi.org/10.1191/0267658303sr214oa

Slabakova, R. (2008). *Meaning in the Second Language*. Mouton de Gruyter. https://doi.org/10.1515/9783110211511

Slabakova, R. (2009). What is easy and what is hard to acquire in a second language? In M. Bowles, T. Ionin, S. Montrul, & A. Tremblay (Eds.), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)* (pp. 280–294). Cascadilla Proceedings Project.

Slabakova, R. (2019). The Bottleneck Hypothesis updated. In T. Ionin & M. Rispoli (Eds.), *Three streams of generative language acquisition research* (pp. 319–345). John Benjamins Publishing Company. https://doi.org/10.1075/lald.63.16sla

Slabakova, R., Leal, T., Dudley, A., & Stack, M. (2020). *Generative second language acquisition* (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781108762380

Smith, C. (1973). An experimental approach to children's linguistic competence. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp. 497–521). Holt, Rinehart and Winston.

Smith, C. S. (1980). The acquisition of time talk: Relations between child and adult grammars. *Journal of Child Language*, *7*(2), 263–278. https://doi.org/10.1017/S0305000900002622

Smith, C. S. (1997). *The parameter of aspect* (2nd ed.). Kluwer Academic.

Smith, N. V., & Tsimpli, I.-M. (1995). *The mind of a savant: Language learning and modularity* (1st edition). Blackwell.

Snape, N., Leung, Y. I., & Smith, M. Sharwood. (Eds.). (2009). *Representational deficits in SLA: Studies in honor of Roger Hawkins*. John Benjamins Publishing Company.

Sonbul, S., El-Dakhs, D. A. S., & Masrai, A. (2023). Second language productive knowledge of collocations: Does knowledge of individual words matter? *Studies in Second Language Acquisition*, *45*(2), 480–502. https://doi.org/10.1017/S0272263122000341

Spada, N., Shiu, J. L.-J., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparisons with other validation studies. *Language Learning*, *65*(3), 723–751. https://doi.org/10.1111/lang.12129

Spinner, P., & Gass, S. M. (2019). *Using judgments in second language acquisition research* (1st ed.). Routledge.

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, *24*(5), 833–841. https://doi.org/10.1017/S1366728920000607

Stowe, L. A., & Sabourin, L. (2005). Imaging the processing of a second language: Effects of maturation and proficiency on the neural processes involved. *International Review of Applied Linguistics in Language Teaching*, *43*(4), 329–353. https://doi.org/10.1515/iral.2005.43.4.329

Sun, C. (2006). *Chinese: A linguistic introduction*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511755019

Suzuki, Y. (2017a). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, *38*(5), 1229–1261. https://doi.org/10.1017/S014271641700011X

Suzuki, Y. (2017b). Validity of new measures of implicit knowledge: Distinguishing implicit knowledge from automatized explicit knowledge. *Applied Psycholinguistics*, *38*(5), 1229–1261. https://doi.org/10.1017/S014271641700011X

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge: Elicited imitation and word monitoring. *Language Learning*, *65*(4), 860–895. https://doi.org/10.1111/lang.12138

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. M. Gass & C. G. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House Publishers.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634. https://doi.org/10.1126/science.7777863

Tanenhaus, M. K., & Trueswell, J. C. (2006). Eye Movements and Spoken Language Comprehension. In *Handbook of Psycholinguistics* (pp. 863–900). https://doi.org/10.1016/B978-012369374-7/50023-7

Tasseva-Kurktchieva, M. (2015). Can production precede comprehension in L2 acquisition? *Second Language Research*, *31*(4), 493–522. https://doi.org/10.1177/0267658315581725

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, *27*(2), 173–204. https://doi.org/10.1017/S0272263105050102

Tokowicz, N., & Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of*

Cognitive Psychology, *22*(7), 1092–1106. https://doi.org/10.1080/09541440903325178

Tolentino, L. C., & Tokowicz, N. (2011). Across languages, space, and time: A review of the role of cross-language similarity in L2 (morpho)syntactic processing as revealed by fMRI and ERP methods. *Studies in Second Language Acquisition*, *33*(1), 91–125. https://doi.org/10.1017/S0272263110000549

Tremblay, A., & Garrison, M. D. (2008). Cloze Tests: A Tool for Proficiency Assessment in Research on L2 French. In M. T. Prior, Y. Watanabe, & S. Lee (Eds.), *Proceedings of the 2008 Second Language Research Forum* (pp. 73–88). Casadila Press.

Trenkic, D., Mirkovic, J., & Altmann, G. T. M. (2014). Real-time grammar processing by native and non-native speakers: Constructions unique to the second language. *Bilingualism: Language and Cognition*, *17*(2), 237–257. https://doi.org/10.1017/S1366728913000321

Tsimpli, I. (2003). Clitics and Determiners in L2 Greek. *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference*, 331–339.

Tsimpli, I.-M. (2003). Interrogatives in the Greek/English interlanguage: A minimalist account. *Selected Papers on Theoretical and Applied Linguistics, Vol 15*, 214–225. https://doi.org/10.26262/ISTAL.V15I0.5682

Tsimpli, I.-M., & Dimitrakopoulou, M. (2007). The Interpretability Hypothesis: Evidence from wh-interrogatives in second language acquisition. *Second Language Research*, *23*(2), 215–242. https://doi.org/10.1177/0267658307076546

Tsimpli, I.-M., & Roussou, A. (1991). Parameter-resetting in L2? *UCL Working Papers in Linguistics*, *3*, 149–169.

Vafaee, P., Suzuki, Y., & Kachisnke, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition*, *39*(1), 59–95. https://doi.org/10.1017/S0272263115000455

Vainikka, A., & Young-Scholten, M. (1994). Direct access to X'-theory: Evidence from Turkish and Korean adults learning German. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 265–316). John Benjamins Publishing Company. https://doi.org/10.1075/lald.8.13vai

Vainikka, A., & Young-Scholten, M. (1996). Gradual development of L2 phrase structure. *Second Language Research*, *12*(1), 7–39.

Vainikka, A., & Young-Scholten, M. (2011). *The acquisition of German: Introducing Organic Grammar*. De Gruyter Mouton. https://doi.org/10.1515/9783110263848

Vainio, S., Pajunen, A., & Hyönä, J. (2016). Processing modifier–head agreement in L1 and L2 Finnish: An eye-tracking study. *Second Language Research*, *32*(1), 3–24. https://doi.org/10.1177/0267658315592201

Van Hell, J. G., & Tokowicz, N. (2010). Event-related brain potentials and second language learning: Syntactic processing in late L2 learners at different L2 proficiency levels. *Second Language Research*, *26*(1), 43–74. https://doi.org/10.1177/0267658309337637

Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, *66*(2), 143–160. https://doi.org/10.2307/2182371

Vendler, Z. (1967). Verbs and times. In Z. Vendler (Ed.), *Linguistics in philosophy* (pp. 97–121). Cornell University Press. https://www.jstor.org/stable/10.7591/j.ctvr7f2mf

Verhagen, J. (2009). *Finiteness in Dutch as a second language* [Doctoral Dissertation]. Vrije Universiteit Amsterdam.

Verhagen, J. (2011). Verb placement in second language acquisition: Experimental evidence for the different behavior of auxiliary and lexical verbs. *Applied Psycholinguistics*, *32*(4), 821–858. https://doi.org/10.1017/S0142716411000087

Vinther, T. (2002). Elicited imitation: A brief overview. *International Journal of Applied Linguistics*, *12*(1), 54–73. https://doi.org/10.1111/1473-4192.00024

Voeten, C. C. (2022). *Permutes: Permutation tests for time series data* (2.6) [Computer software]. https://cran.r-project.org/web/packages/permutes/index.html

Vogel, E. K. (2017). *Native vs. Non-Native processing of Spanish: The role of lexical and grammatical aspect* [Doctoral Dissertation]. Florida State University.

von Stutterheim, C., & Klein, W. (1987). A concept-oriented approach to second language studies. In C. Wollman. Pfaff (Ed.), *First and second language acquisition processes* (pp. 191–205). Newbury House Publishers.

Wang, D., Wang, S., Zinszer, B., Sheng, L., & Jasińska, K. (2022). Cross-linguistic influences of L1 on L2 morphosyntactic processing: An fNIRS study. *Journal of Neurolinguistics*, *63*, 101063. https://doi.org/10.1016/j.jneuroling.2022.101063

Werner, V. (2013). Temporal adverbials and the present perfect/past tense alternation. *English World-Wide*, *34*(2), 202–240. https://doi.org/10.1075/eww.34.2.04wer

Werner, V. (2014). *The present perfect in World Englishes: Charting unity and diversity*. University of Bamberg Press. https://www.semanticscholar.org/paper/The-Present-Perfect-in-World-Englishes%3A-Charting-Werner/9543a29d86163564db3a342bfa3e1966dcc88099

White, L. (2003a). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and Cognition*, *6*(2), 129–141. https://doi.org/10.1017/S1366728903001081

White, L. (2003b). Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology. *Bilingualism: Language and Cognition*, *6*(2), 129–141. https://doi.org/10.1017/S1366728903001081

White, L. (2003c). On the nature of interlanguage representation: Universal grammar in the second language. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (1st ed., pp. 18–42). Wiley. https://doi.org/10.1002/9780470756492.ch2

White, L. (2003d). *Second language acquisition and universal grammar*. Cambridge University Press. https://doi.org/10.1017/CBO9780511815065

White, L. (2011). Second language acquisition at the interfaces. *Lingua*, *121*(4), 577–590. https://doi.org/10.1016/j.lingua.2010.05.005

White, L., Valenzuela, E., Kozlowska–Macgregor, M., & Leung, Y.-K. I. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, *25*(01), 105–133. https://doi.org/10.1017/S0142716404001067

William, J. N. (2009). Implicit learning in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *The new handbook of second language acquisition* (2nd ed., pp. 319–353). Emerald.

Winter, B. (2019). *Statistics for linguists: An introduction using R* (1st ed.). Routledge. https://doi.org/10.4324/9781315165547

Woo, I. (2021). On preverbal zai in Mandarin Chinese: Its progressive and prepositional functions. *Linguistics*, *59*(3), 513–539. https://doi.org/10.1515/ling-2021-0044

Wu, S.-L., & Ortega, L. (2013). Measuring global oral proficiency in SLA research: A new elicited imitation test of L2 Chinese. *Foreign Language Annals*, *46*(4), 680–704. https://doi.org/10.1111/flan.12063

Wu, S.-L., Tio, Y. P., & Ortega, L. (2022). Elicited imitation as a measure of L2 proficiency: New insights from a comparison of two L2 English parallel forms. *Studies in Second Language Acquisition*, *44*(1), 271–300. https://doi.org/10.1017/S0272263121000103

Wu, Z., & Juffs, A. (2022). Effects of L1 morphological type on L2 morphological awareness. *Second Language Research*, *38*(4), 787–812. https://doi.org/10.1177/0267658321996417

Xiao, R., & McEnery, T. (2004). *Aspect in Mandarin Chinese: A corpus-based study*. John Benjamins Publishing Company.

Yan, X. (2015). *The processing of formulaic language on elicited imitation tasks by second language speakers* [Doctoral Dissertation]. Purdue University.

Yan, X., Maeda, Y., Lv, J., & Ginther, A. (2016). Elicited imitation as a measure of second language proficiency: A narrative review and meta-analysis. *Language Testing*, *33*(4), 497–528. https://doi.org/10.1177/0265532215594643

Yang, Q., Bucci, M. P., & Kapoula, Z. (2002). The latency of saccades, vergence, and combined eye movements in children and in adults. *Investigative Ophthalmology & Visual Science*, *43*(9), 2939–2949.

Yang, S., & Huang, Y. Y. (2004). The impact of the absence of grammatical tense in L1 on the acquisition of the tense-aspect system in L2. *International Review of Applied Linguistics in Language Teaching*, *42*(1), 49–70. https://doi.org/10.1515/iral.2004.002

Yang, W., Chan, A., Chang, F., & Kidd, E. (2020). Four-year-old Mandarin-speaking children's online comprehension of relative clauses. *Cognition*, *196*, 104103. https://doi.org/10.1016/j.cognition.2019.104103

Yao, P., & Chen, B. (2017). Cross-linguistic differences affect late Chinese-English learners on-line processing of English tense and aspect. *International Journal of Bilingualism*, *21*(3), 268–290. https://doi.org/10.1177/1367006915624248

Yuan, B. (2017). Can L2 sentence processing strategies be native-like? Evidence from English speakers' L2 processing of Chinese base-generated-topic sentences. *Lingua, 191–192*, 42–64. https://doi.org/10.1016/j.lingua.2017.04.001

Zeng, X., Chen, X., & Shirai, Y. (2021). Lexical and grammatical aspect in on-line processing of English past tense and progressive aspect by Mandarin speakers. *Frontiers in Psychology*, *12*, 661923.

Zeng, X., Shirai, Y., & Chen, X. (2021). Universals and transfer in the acquisition of the progressive aspect: Evidence from L1 Chinese, German, and Spanish learners' use of the progressive -ing in spoken English. *International Review of Applied Linguistics in Language Teaching*, *59*(2), 267–292. https://doi.org/10.1515/iral-2017-0078

Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *37*(3), 457–486. https://doi.org/10.1017/S0272263114000370

Zhou, P., Crain, S., & Zhan, L. (2014). Grammatical aspect and event recognition in children's online sentence comprehension. *Cognition*, *133*(1), 262–276. https://doi.org/10.1016/j.cognition.2014.06.018

Zufferey, S., Mak, W., Degand, L., & Sanders, T. (2015). Advanced learners' comprehension of discourse connectives: The role of L1 transfer across on-line and off-line tasks. *Second Language Research*, *31*(3), 389–411. https://doi.org/10.1177/0267658315573349