Author Accepted Manuscript – Applied Psycholinguistics 2026

Iconicity does not always facilitate complex sentence comprehension: Behavioural and pupillometric evidence from Mandarin-English bilingual adults

Shijie Zhang [a], Alissa Ferry [a], Silke Brandt [b], Emily Warren [a], Anna Theakston [a]

[a] Division of Psychology, Communication and Human Neuroscience, The University of Manchester, Manchester, UK. Email: shijie.zhang@manchester.ac.uk (ORCID: 0000-0002-4383-8220); alissa.ferry@manchester.ac.uk (ORCID: 0000-0003-2217-1989); emily.warren-2@manchester.ac.uk (ORCID: 0009-0007-3242-0535); anna.theakston@manchester.ac.uk (ORCID: 0000-0002-9483-7893).

[b] Department of Linguistics and English Language, Lancaster University, Lancaster, UK. Email: s.brandt@lancaster.ac.uk (ORCID: 0000-0003-3363-8740).

Corresponding author: Anna Theakston, Division of Psychology, Communication and Human Neuroscience, The University of Manchester, Manchester, M13 9PL, UK. Email: anna.theakston@manchester.ac.uk.

Conflict of interest disclosure: None.

Funding statement: This study was supported by the Economic and Social Research Council (Grant number: ES/S007113/1). For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

Ethics approval statement: This study was approved by The University of Manchester's Research Ethics Committee [Reference numbers: 13877 and 22249] and Lancaster University's Research Ethics Committee [Reference number: FASSLUMS-2025-5216-ExRev-1], and followed all ethical and legal guidelines. Prior written consent was obtained from all adult participants.

Acknowledgements: Thanks to Yiting Zhang for assisting with the piloting phase. The pilot data were included in her MRes dissertation, but neither the pilot data nor any of her writing is included in the current paper.

Iconicity does not always facilitate complex sentence comprehension: Behavioural and pupillometric evidence from Mandarin-English bilingual adults

Abstract

English adverbial sentences allow flexible clause ordering: main-adverbial and adverbial-main, and their comprehension by English monolingual speakers is influenced by iconicity. Iconic sentences, whose clause order reflects the chronological sequence of real-world events, are more easily comprehended ("We ate dinner before we went to see a movie"; "After we ate dinner, we went to see a movie") than non-iconic sentences where the order of the clauses does not match the order of events in the real world ("Before we went to see a movie, we ate dinner"; "We went to see a movie after we ate dinner"). However, Mandarin speakers strongly prefer the adverbial-main order in spontaneous Mandarin speech. This paper examines whether iconicity represents a generally preferred strategy for interpreting English adverbial sentences, or whether a speaker's first language affects processing. In two studies, we tested the comprehension of English *before*- and *after*-sentences by Mandarin-English bilingual and English monolingual adults, using online pupillometry with offline accuracy in a sentence-picture matching task. We found that bilinguals, like monolinguals, comprehended *after*-sentences in the iconic adverbial-main order (also the preferred order in Mandarin) more accurately and with less effort than in the reverse order. However, unlike monolinguals, bilinguals did not comprehend *before*-sentences better in either the iconic main-adverbial order or the preferred adverbial-main order for Mandarin. Our results suggest that iconicity is not universal; instead, language-specific usage patterns, such as preferred clause order, compete with iconicity to influence processing of adverbial sentences.

Keywords: Complex syntax, Language comprehension, Crosslinguistic influence, Iconicity, Pupillometry

1. Introduction

While in the real world, time unfolds sequentially, with one event following another, language allows us to communicate events in different orders through temporal adverbial sentences. In English, temporal adverbial sentences can occur in two orders: main-adverbial (e.g., "We ate dinner before we went to see a movie") and adverbial-main (e.g., "Before we went to see a movie, we ate dinner"). However, these two orders are not processed equally. For English monolingual children and adults, converging evidence suggests that the clause order that reflects the chronological sequence of real-world events is more easily processed (e.g., "We ate dinner before we went to see a movie"; "After we ate dinner, we went to see a movie"; De Ruiter et al., 2018; Politzer-Ahles et al., 2017). In addition to this iconic semantic mapping, the cognitive load required to process each clause order has been proposed to differ, such that main-adverbial orders require less cognitive load and are therefore easier to process than adverbial-main orders overall (e.g., Diessel, 2005). The extent to which iconicity and clause order determine ease of sentence processing is also affected by the wider discourse context and associated pragmatic considerations (e.g., De Ruiter et al., 2020, 2021), though these factors are beyond the scope of the current paper.

Despite the relatively well-understood (complex) sentence processing mechanisms employed by English monolinguals, it remains largely unexplored whether these processing mechanisms universally impact individuals from different language backgrounds, such as Mandarin-English bilinguals. In Mandarin Chinese, there is a strong preference for the temporal adverbial clause to precede the main clause (i.e., adverbial-main order), as illustrated in examples (1) and (2). Corpus studies on Mandarin show that 86.5% - 97.0% of temporal adverbial sentences, such as (*dang*/*zai*)...*de shihou* "when", *yiqian*/*zhiqian* "before", and *yihou*/*zhihou* "after", appear in adverbial-main order in spontaneous speech (Wang, 2002; Yeh, 2000). When only considering *yiqian*/*zhiqian* "before" and *yihou*/*zhihou* "after", this increases to 100.0%, although the corpus size may be small (Yeh, 2000). By contrast, in English, 41.5% of temporal adverbial sentences appear in adverbial-main order in spontaneous speech (Diessel, 2005), but when only considering *before*- and

after-sentences the mean percentage in adverbial-main order is only 32.8% - 39.8% (De Ruiter et al., 2021; Diessel, 2005). Furthermore, English shows a preference for an iconic clause order, with a mean of 71.3% - 73.8% of *before*- and *after*-sentences exhibiting iconic order (De Ruiter et al., 2021; Diessel, 2005). This difference in clause order preference between Mandarin and English may lead Mandarin-English bilinguals to use different strategies when constructing mental representations of events described in English, compared to English monolinguals.

(1) [Women chi-wan wanfan **zhihou**]$_{ADVERBIAL}$, [qu kan-le dianying]$_{MAIN}$

we eat-finish dinner **after** go see-ASP movie

"**After** we ate dinner, (we) went to see a movie."

(2) [Women qu kan dianying **zhiqian**]$_{ADVERBIAL}$, [chi-le wanfan]$_{MAIN}$

we go see movie **before** eat-ASP dinner

"**Before** we went to see a movie, (we) ate dinner."

The present studies investigate the comprehension of English temporal *before*- and *after*-sentences by Mandarin-English bilingual and English monolingual adults. We explore whether the processing strategies employed by English monolinguals uniformly affect language processing, or whether language-specific usage patterns in Mandarin influence Mandarin-English bilinguals' sentence processing, showing crosslinguistic influence. In the following sections, we review general theoretical accounts of (complex) sentence processing and the empirical evidence from English, and then discuss how these accounts apply to Mandarin and the potential crosslinguistic influence from Mandarin to English.

1.1. Theoretical accounts

1.1.1. Semantic account

The semantic principle of iconicity has been proposed to explain sentence processing (Clark, 1971; Croft, 2003; Diessel, 2008; Givón, 1985). The basic idea is that "the structure of language reflects in some way the structure of experience" (Croft, 2003, p. 102). The more closely the structure of language mirrors our real-world experience, the easier it is for us to communicate

(Givón, 1985). In the case of adverbial sentences, those in which the clause order reflects the chronological sequence of real-world events are expected to be more easily processed than those that do not. For example, in examples (4) and (5) in Table 1, the event "she hoovers the house", which is mentioned first in the sentence, also occurs first in the real world – these are iconic sentences. In contrast, in examples (3) and (6), the order of events in the sentence is reversed relative to the real-world order – these are non-iconic. Accordingly, iconic sentences are expected to be processed more easily than non-iconic ones.

[Table 1 here]

The results from child experimental studies support the effect of iconicity on English monolingual children's comprehension of adverbial sentences (e.g., Zhang et al., submitted; Blything & Cain, 2016; Blything et al., 2015; De Ruiter et al., 2018, 2020; Wagner & Holt, 2023). For example, De Ruiter and colleagues (2018) examined English monolingual children's comprehension of temporal and causal/conditional adverbial sentences (*before*, *after*, *because*, *if*) using a sentence-picture matching task, and found that five-year-olds comprehended all types of sentences more accurately when the sentences were iconic rather than non-iconic.

However, the strength of the iconicity effect may vary across different connectives. In De Ruiter et al. (2018), the iconicity effect was stronger for *after*-sentences than for *before*-sentences, with *before*-sentences being comprehended more accurately overall. A similar pattern was observed in Zhang et al. (submitted) for English monolinguals aged four to five and seven to nine. Likewise, Wagner and Holt (2023) found that English monolinguals aged seven to twelve exhibited the iconicity effect for *after*-sentences, but not for *before*-sentences[1,2]. These studies suggest that the

[1] The lack of the iconicity effect for *before*-sentences in Wagner and Holt (2023) may also reflect differences in study design and participant age, as several age groups reached near-ceiling performance, potentially masking any effect of this connective.

[2] Similar to child comprehension studies, child production studies also show a varied strength of the iconicity effect for different connectives. For example, Blything and Cain (2019) observed the iconicity effect only for *after*-sentences, but not for *before*-sentences, in English monolinguals aged three to six, whereas Zhang et al., (2025) found the iconicity effect for both sentence types in English monolinguals aged four to five. Differences in study design could partially explain these findings, but a detailed discussion of production studies falls outside the scope of the present comprehension work. Likewise, discussion of studies investigating languages

varied strength of the iconicity effect for different connectives could be related to their semantic differences. *Before*-sentences have a more consistent form-meaning mapping when used as a temporal connective, supporting more accurate comprehension overall. In contrast, "after" often appears in other constructions such as "look after" and "after all" (De Ruiter et al., 2021; Wagner & Holt, 2023), which may weaken its lexical representation in temporal adverbial sentence comprehension (see Ambridge et al. (2015) and Theakston (2020) for an overview of usage/frequency-based arguments). As a result, when processing *after*-sentences, children rely more heavily on the non-optimal semantic strategy of iconicity to map the order of events in the sentence directly onto the sequence of real-world events (Zhang et al., submitted).

Regarding English adults' comprehension of adverbial sentences, empirical evidence also supports the effect of iconicity, but differences between connectives are not consistently reported. For example, Politzer-Ahles et al. (2017) examined native English adults' online comprehension of temporal *before*- and *after*-sentences. In this study, adults read a sentence word by word while their electroencephalographic (EEG) activity was recorded to measure an anterior event-related potential (ERP) component, where greater negativity is associated with increased processing costs (e.g., Vos et al., 2001). The results showed that, in adverbial-main orders, *before*-sentences (non-iconic) elicited more negative ERPs than *after*-sentences (iconic), whereas in main-adverbial orders, *after*-sentences (non-iconic) elicited more negative ERPs than *before*-sentences (iconic), indicating greater processing costs related to non-iconic sentences (see also Münte et al. (1998) using ERPs, and Ye et al. (2012) using functional magnetic resonance imaging (fMRI)). However, unlike in child comprehension, Politzer-Ahles et al. (2017) did not find an overall difference between *before*- and *after*-sentences in adults' processing. This could be related to task design and the test stimuli used (see Zhang et al. (2025) for discussion), or it could be that the differences in the lexical

other than English and Mandarin falls outside our scope, as it is unclear whether differences exist in the frequency and usage patterns of temporal "before" and "after" in different languages, and how these might influence the results.

representations of “before” and “after” are less pronounced in adults due to their richer lexical and semantic knowledge.

1.1.2. Processing-based account

The ordering of the main and adverbial clause alone has also been proposed to affect the processing of adverbial sentences (Diessel, 2005; see also Hawkins, 1990, 1992, 1994). Main-adverbial orders have been suggested to be easier to process than adverbial-main orders. Compare examples (4) and (6) with examples (3) and (5) in Table 1. This is because, when processing main-adverbial orders, the listener can process the main clause and the adverbial clause one after the other, building the link between them only when the adverbial clause is encountered. In contrast, when processing adverbial-main orders, on hearing the adverbial the listener immediately recognises that the sentence contains at least two clauses and must keep the adverbial clause in memory until the main clause is encountered, which increases processing load.

However, previous child and adult comprehension studies did not find that one clause order was generally comprehended better than the other (e.g., De Ruiter et al., 2018; Politzer-Ahles et al., 2017). Instead, as mentioned above, the effect of iconicity (i.e., an interaction between clause order and connective type) was predominantly reported: *before*-sentences were better comprehended in the iconic main-adverbial order, and *after*-sentences were better comprehended in the iconic adverbial-main order (see Table 1 for examples). Only Zhang et al. (2023), who examined the production of *when*-sentences (e.g., “When Pig is swimming, Little Panda is drinking very fast”) by three- to five-year-old children and adults, offers some support for this processing-based hypothesis. They found that three-year-olds tended to reverse sentences in adverbial-main order into main-adverbial order in their repetitions, but this tendency decreased with age. This led the authors to suggest that younger children, due to their lower processing capacities, tend to prefer main-adverbial orders. However, the simultaneous clauses with “when” in Zhang et al. (2023) differ from *before*- and *after*-sentences in that they do not encode a clear order of events, thereby limiting the study’s ability to investigate the effect of iconicity, and the test sentences were preceded by a

discourse context which has also been found to affect sentence processing (e.g., De Ruiter et al., 2020).

1.1.3. Clause order in Mandarin adverbial sentences and potential crosslinguistic influence

As in sentence comprehension, English monolingual adults tend to be influenced by iconicity when ordering main and temporal adverbial clauses in spontaneous English speech (71.3% - 73.8% of adult *before*- and *after*-sentences are produced in iconic order; De Ruiter et al., 2021; Diessel, 2005). In contrast, Mandarin Chinese speakers show a strong preference for the temporal adverbial clause to precede the main clause (i.e., adverbial-main order) in spontaneous Mandarin speech (86.5% - 97.0% of temporal adverbial sentences, increasing to 100% for *before-/after*-sentences only; Wang, 2002; Yeh, 2000).

The preference for the adverbial-main order in Mandarin can be linked to its information structure. Mandarin is widely recognised as a topic-comment language where the topic typically occupies the sentence-initial position (e.g., Chao, 1968; Li & Thompson, 1981). Despite some differences in the precise definition of "topic", it is commonly assumed to encode given or shared information (e.g., Halliday & McDonald, 2004; Li & Thompson, 1981). This aligns with the functional role of initial temporal adverbial clauses, which serve to provide pragmatically presupposed information and establish the thematic ground for subsequent clauses (Diessel, 2001, 2005).

Furthermore, this preference for the adverbial-main order in Mandarin does not necessarily contradict processing-based explanations (e.g., Diessel, 2005). While processing-based accounts suggest that adverbial-main orders in English impose higher processing load (relative to main-adverbial orders), Mandarin avoids this through its clause-final connective structure (see Diessel (2001) for more details on typological differences in adverbial sentences across languages). Specifically, English connectives occur in the initial position of adverbial clauses, immediately signalling the sentence type and requiring the listener to anticipate a subsequent main clause. However, Mandarin connectives occur at the end of adverbial clauses. As shown in examples (1) and (2), the listener does not necessarily identify the adverbial clause until the connective is encountered

at the end; thus, the link between the two clauses is established only at that point. Due to this structural difference, the processing load for adverbial-main orders in Mandarin does not mirror that of English, also providing a processing-based justification for its prevalence in spontaneous speech.

However, Mandarin's strong preference for the adverbial-main order limits the ability to fully test the iconicity effect. Chen et al. (2022) provided ERP evidence for iconicity in Mandarin, showing that Mandarin adults exhibited greater processing costs for *before*-sentences than for *after*-sentences. This is because, although both types of test sentences followed the adverbial-main order, in *before*-sentences this order is non-iconic, whereas in *after*-sentences it is iconic. Yet, these findings offer only a partial view. If the iconicity effect is truly universal, one would expect an advantage of iconic *after*-sentences over non-iconic *after*-sentences, and an advantage of iconic *before*-sentences over non-iconic *before*-sentences. Such an investigation, however, is not feasible in Mandarin due to its relatively rigid word order in temporal adverbial sentences.

Examining how Mandarin-English bilinguals comprehend English adverbial sentences, which allow for both adverbial-main and main-adverbial orders, provides an ideal test of whether the iconicity effect observed in English monolinguals is universal, or modulated by Mandarin-specific usage patterns, leading to crosslinguistic influence (e.g., Hulk & Müller, 2000; Müller & Hulk, 2001). If Mandarin usage patterns influence Mandarin-English bilinguals' processing of English, we would expect the adverbial-main order that dominates in Mandarin to facilitate their processing of the same order in English, whereas the dispreferred main-adverbial order may pose a greater difficulty. Put another way, based on the crosslinguistic influence and usage patterns, Mandarin-English bilingual adults should comprehend adverbial-main orders better than main-adverbial orders in English.

Note that for *after*-sentences, the predictions of the crosslinguistic influence and usage patterns and of iconicity converge: the dominant adverbial-main order in Mandarin is also the iconic order for *after*-sentences. Therefore, according to both hypotheses, Mandarin-English bilinguals should comprehend *after*-sentences in the iconic adverbial-main order better than in the non-iconic

main-adverbial order, similar to what has been found in English monolinguals (Münte et al., 1998; Politzer-Ahles et al., 2017; Ye et al., 2012). However, the crosslinguistic influence and usage patterns conflict with iconicity for *before*-sentences: according to the former, the adverbial-main order should be better comprehended, whereas according to the latter, the iconic main-adverbial order should be better comprehended (see Table 1 for examples).

To our knowledge, no studies have examined Mandarin-English bilinguals' processing of English temporal adverbial sentences. However, several corpus-based studies have investigated the ordering distribution of English temporal adverbial sentences in the writing of Chinese learners of English. For example, Li and Long (2022) examined English temporal adverbial sentences (e.g., *when*, *before*, *after*, *until*) in the writing of Chinese learners of English, comparing it to native English and native Chinese written texts. They found that Chinese learners of English tended to use adverbial-main orders. However, as English language proficiency increased, the proportion of adverbial-main orders decreased—from 78.0% in beginners to 56.5% in advanced learners. The proportion of adverbial-main orders in beginners was similar to the proportion found in native Chinese texts (96.5%), while the proportion in advanced learners was closer to the proportion in native English essays (36.0%). Therefore, Li and Long's (2022) results indicate a clear crosslinguistic influence from Chinese to English, with this influence decreasing as English language proficiency increases. However, they did not analyse different types of temporal adverbial sentences separately, so it remains unclear whether there were differences between *after*- and *before*-sentences (and whether Chinese learners of English would also show an effect of iconicity).

Fang (2009) also found a general preference for adverbial-main orders in the writing of Chinese learners of English. However, when examining *after*- and *before*-sentences separately, they found that only *after*-sentences occurred more frequently in the adverbial-main order (iconic, 65.98%). *Before*-sentences more often appeared in the main-adverbial order (iconic, 64.37%). This suggests an effect of iconicity in Chinese learners' English writing, despite crosslinguistic influence. However, they did not account for the Chinese learners' English proficiency.

To summarise, previous studies have reported crosslinguistic influence on Chinese learners' use of English temporal adverbial sentences, but the results remain inconclusive. Moreover, these studies focused exclusively on written data, where individuals have more time to plan their sentences than in spoken language. In two studies, we examined Mandarin-English bilinguals' comprehension of spoken English temporal adverbial sentences both offline and online, and compared this to that of English monolinguals. Our goal was to disentangle whether Mandarin-English bilinguals employ the same processing strategies (e.g., iconicity) as English monolinguals, or whether their processing is affected by language-specific usage patterns in Mandarin.

1.2. The present studies

The present studies investigated Mandarin-English bilingual and English monolingual adults' comprehension of sentences containing two types of connectives ("after" and "before") in two clause orders (main-adverbial and adverbial-main), using an offline sentence-picture matching task (Experiment 1) and an online version of the task employing pupillometry (Experiment 2). Additionally, we investigated whether Mandarin-English bilingual adults' comprehension of temporal adverbial sentences was affected by their English language proficiency. Based on theoretical predictions and previous empirical evidence, the following hypotheses were formulated in relation to behaviour accuracy, reaction times (RTs), and changes in pupil size:

(i) Semantic account of iconicity: Iconic clause orders are comprehended better (i.e., with higher accuracy, faster RTs, and smaller changes in pupil size) than non-iconic clause orders (e.g., Clark, 1971; Croft, 2003; Diessel, 2008; Givón, 1985).

(ii) Semantic consistency: *Before*-sentences are comprehended better than *after*-sentences (e.g., Ambridge et al, 2015; Theakston, 2020).

(iii) Processing-based account: Main-adverbial clause orders are comprehended better than adverbial-main clause orders (e.g., Diessel, 2005).

(iv) Crosslinguistic influence and usage patterns: Adverbial-main clause orders are comprehended better than main-adverbial clause orders by Mandarin-English bilingual adults (e.g., Hulk & Müller, 2000).

(v) Individual differences: The higher the English proficiency of Mandarin-English bilingual adults, the less crosslinguistic influence from Mandarin on their English sentence comprehension (Li & Long, 2022).

2. Experiment 1

2.1. Participants

A total of 46 Mandarin-English bilingual and 40 English monolingual adults were included in the final sample[3]. The bilinguals (age range = 18-38 years, mean age = 22.74 years, SD = 3.73 years, 39 females) were recruited through social media and the University of Manchester's credit system. They all spoke both Mandarin and English, with Mandarin as their native language, and were living in the UK at the time of testing. They were either pursuing or had obtained university degrees: undergraduate (N=20), master's (N=21), and PhD (N=5). Most bilinguals (N=38) had taken the International English language Testing System (IELTS), and we used their IELTS scores as a proxy for their English language proficiency in this experiment. The remaining bilinguals (N=8) either did not have any standardised English test scores or had scores from other English tests (e.g., the Chinese College English test), which could not be directly compared.

The English monolingual adults (age range = 18-30 years, mean age = 19.35 years, SD = 2.01 years, 35 females) were all recruited through the University of Manchester's credit system, and all were undergraduates. Neither the monolingual nor the bilingual participants had any known speech and language difficulties, nor any vision or hearing impairments. In addition to the 46 bilinguals and

[3] Given that no prior studies have conducted a similar investigation, we could not perform a power analysis based on existing data to determine the sample size. Therefore, the sample size for Experiment 1 was chosen as a reasonable number of participants for an initial exploratory study. Using the data from Experiment 1, we then conducted a power analysis for Experiment 2, which indicated that 40 participants per language group were sufficient to achieve at least 80% power. Additionally, Experiment 1 was not pre-registered, whereas Experiment 2 was pre-registered.

40 monolinguals in the final sample, two additional bilinguals were tested, but their data had to be excluded due to experimenter error.

2.2. Offline sentence-picture matching task

The participants' comprehension of adverbial sentences was tested using an offline sentence-picture matching task. In the task, as shown in Table 2, the participant first listened to an audio-recorded adverbial sentence that ended with a beep sound while viewing a blank screen. Immediately after the beep, two picture sequences were displayed at the top and bottom of the screen respectively, and the participant was asked to select the picture sequence that matched the sentence they had heard. Both picture sequences depicted the same pair of events described in the sentence but in different orders: one matched the order of events in the sentence, and the other presented them in reverse. Within each picture sequence, the pictures followed a left-to-right reading order: the left picture represented the event that happened first, and the right picture represented the event that happened second, a convention shared by native English and Mandarin speakers. RTs were measured from the offset of the beep sound to the participant's click on a picture sequence.

[Table 2 here]

The task was presented using the experiment builder Gorilla.sc. The experimenter provided participants with the Gorilla link to access the task and observed them as they completed it, either in person in a lab at the University of Manchester or online via Zoom. The task took between 15 and 20 minutes to complete.

2.2.1. Design

The offline sentence-picture matching task had three factors: one between-subjects factor (LanguageGroup), and two within-subject factors (ConnectiveType, ClauseOrder), each with the following levels:

- LanguageGroup: bilingual, monolingual
- ConnectiveType: *before*, *after*

- ClauseOrder: main-adv (main-adverbial order), adv-main (adverbial-main order)

2.2.2. Materials

A total of 24 test sentences were constructed, with six test sentences per condition (four conditions: two ConnectiveType × two ClauseOrder). Each sentence consisted of a main and an adverbial clause, describing two events performed by a single actor (either a boy or a girl). To create these sentences, we used 12 event pairs (e.g., the "paint the old fence" and "hoover the house" pair), with each pair being connected once with "after" and once with "before". Among the *after*-sentences (N=12), half followed a main-adv order, while the other half followed an adv-main order. The same distribution applied to the *before*-sentences (N=12). A full list of test sentences can be found in Table S1 in the Supplemental Material.

The pair of events in each test sentence was arbitrarily related; the events could happen in either order, with only the connective "before/after" indicating the correct order. The subject in each test sentence was pronominal ("he" in half of the sentences and "she" in the other half), and the objects of the transitive verbs were always inanimate. Each event was described using a different verb in the present tense, and each event description was five syllables long.

Four different experimental lists were constructed. Each list consisted of two sessions, with the test sentence containing the same event pair appearing separately in each session (see examples (7) and (8)). The order of the test sentences within each session was pseudo-randomised, allowing a maximum of two consecutive test sentences in the same condition. The position of the correct picture sequence in each session was counterbalanced: in half of the trials, the correct picture sequence was at the top and in the other half of the trials at the bottom. List 2 was the same as List 1, with the difference that all main-adv sentences were turned into adv-main sentences and vice versa (see examples (9) and (10)). Lists 3 and 4 were created by swapping session 1 and session 2 of Lists 1 and 2. Participants were randomly assigned to one of the four experimental lists.

(7) List 1 Session 1: *After she paints the old fence, she hoovers the house.*

(8) List 1 Session 2: *Before she paints the old fence, she hoovers the house.*

(9) List 2 Session 1: *She hoovers the house after she paints the old fence.*

(10) List 2 Session 2: *She hoovers the house before she paints the old fence.*

In addition to the 24 test sentences, four warm-up trials were provided before the test to familiarise participants with the task. The first two warm-up trials were simple sentences (e.g., "Seal is catching a ball") with only two pictures (e.g., a seal catching a ball vs. a seal opening a door). The next two warm-up trials resembled the test trials (i.e., two picture sequences, e.g., water his plant + switch the light on vs. switch the light + water his plant), except the sentences followed a "First..., then..." structure (e.g., "First he switches the light on, then he waters his plant"). In addition, seven filler trials were interspersed, one after every three test trials. They were also simple sentences, like the first two warm-up trials. All sentences were audio-recorded by a young female native speaker of British English, and all audio and visual stimuli were borrowed from De Ruiter et al. (2018).

2.3. Data analysis

We analysed participants' behavioural accuracy and RTs statistically, one after another, followed by simple correlation analyses between bilingual participants' accuracy and RT performance in sentence comprehension and their IELTS scores.

2.3.1. Behavioural accuracy

In total, we collected 1104 responses from bilinguals, of which 955 responses were correct and 149 were incorrect, achieving an accuracy of 86.50%. We collected 960 responses from monolinguals, of which 898 responses were correct and 62 were incorrect, achieving an accuracy of 93.54%. Figure 1 shows the proportion of correct and incorrect responses for both language groups across the four conditions (two types of adverbial sentences and two clause orders). It reveals an apparent three-way interaction between ClauseOrder and Type across the two language groups, which we analysed statistically to examine its significance.

[Figure 1 here]

Statistical analysis was carried out using Generalised Linear Mixed-effects Models (GLMMs; Baayen et al., 2008) with the lme4 package (Bates et al., 2015) in R, version 4.5.2. GLMMs are well-

suited for analysing repeated measures data in psycholinguistics, because they account for both the fixed effects of experimental conditions and the random variation across participants and test items (Baayen et al., 2008). The fixed effects included ClauseOrder (sum-coded: main-adv = 0.5, adv-main = -0.5), Type (sum-coded: *before* = 0.5, *after* = -0.5), and LanguageGroup (sum-coded: bilingual = 0.5, monolingual = -0.5). Random intercepts were included for participants and items to account for individual and item-level variability. Model building followed a stepwise forward selection procedure. First, the contribution of each fixed effect was tested by comparing it individually to the null model containing only random intercepts using the ANOVA function. The fixed effects that significantly improved model fit were retained for the next stage. Next, two-way interactions (ClauseOrder × Type, ClauseOrder × LanguageGroup, Type × LanguageGroup) were added one at a time to the model containing the significant fixed effects. The two-way interactions that significantly improved model fit (along with the included fixed effects) were retained. Finally, the model containing the significant two-way interactions was compared to the model containing the three-way interaction (ClauseOrder × Type × LanguageGroup). Model comparison indicated that the three-way interaction significantly improved model fit, χ^2 (3) = 10.44, p = 0.015, and was therefore retained in the final model (see Table 3). By-participant random slopes were not included, as they did not improve the fit of the final model. We interpreted the three-way interaction using post hoc comparisons using the emmeans package (Lenth et al., 2023) with Bonferroni correction.

[Table 3 here]

As shown in Figure 1 and supported by post-hoc analyses in Table S2 (Supplemental Material), monolinguals comprehended *before*-sentences more accurately when they were in the iconic main-adv order than in the non-iconic adv-main order, and they comprehended *after*-sentences more accurately when they were in the iconic adv-main order than in the non-iconic main-adv order. Further support for this pattern is seen in main-adv orders where *before*-sentences were comprehended more accurately than *after*-sentences, and vice versa in adv-main orders. These results suggest that monolinguals were significantly more accurate at comprehending temporal

adverbial sentences in iconic clause order. However, they did not show a consistent advantage for main-adv orders, or for *before*- over *after*-sentences.

Bilinguals also comprehended *after*-sentences more accurately when they were in the iconic adv-main order (which is also the preferred order in Mandarin). However, they comprehended *before*-sentences equally, regardless of whether they were in main-adv or adv-main orders thus showing no consistent preference for either order. Furthermore, in main-adv orders *before*-sentences were comprehended similarly to *after*-sentences, while in adv-main orders *after*-sentences were comprehended more accurately than *before*-sentences. This indicates that, unlike monolinguals, bilinguals showed the effect of iconicity and/or clause order only in *after*-sentences, but not in *before*-sentences.

2.3.2. Behavioural reaction times

For the analyses of RTs, only the correct responses were included. After inspecting the correct responses, we removed the responses with RTs that exceeded three standard deviations above or below the mean RT. In total, 941 out of 955 correct responses from bilinguals (excluded 14 outliers), and 883 out of 898 responses from monolinguals (excluded 15 outliers) were included in the statistical analysis. The mean RT was 2746.75ms (SD = 1742.42ms) for bilinguals and 1505.70ms (SD = 761.12ms) for monolinguals. Figure 2 shows that monolinguals comprehended adverbial sentences more quickly than bilinguals in general. Furthermore, there seems to be an interaction between ClauseOrder and Type for bilinguals, but not for monolinguals, which we analysed statistically to examine its significance.

We analysed the RTs using the same strategy as for accuracy. Model comparison indicated that the interaction between LanguageGroup and Type significantly improved model fit, χ^2 (1) = 5.17, p = 0.023, and was therefore retained in the final model (see Table 4). As shown in Figure 2 and supported by post-hoc analyses in Table S3, monolinguals comprehended both types of adverbial sentences significantly more quickly than bilinguals. Moreover, monolinguals comprehended *before*-sentences as quickly as *after*-sentences, whereas bilinguals comprehended *before*-sentences

significantly more slowly than *after*-sentences. Figure 2 further shows that bilinguals' slower comprehension of *before*-sentences was likely driven by the fact that *before*-sentences in the iconic main-adv order were comprehended the slowest compared to other conditions. However, the three-way interaction between ClauseOrder, Type and LanguageGroup did not improve model fit, χ^2 (4) = 7.23, p = 0.125, so was not retained in the final model.

[Figure 2 here]

[Table 4 here]

To conclude, in Experiment 1 using the offline sentence-picture matching task, monolinguals comprehended iconic *before*- and *after*-sentences more accurately than non-iconic ones but showed no differences in their comprehension of these sentences in terms of RTs. Bilinguals also showed the effect of iconicity and/or clause order in *after*-sentences in terms of accuracy, but not in *before*-sentences. Instead, *before*-sentences were comprehended more slowly than *after*-sentences, largely driven by sentences in the iconic main-adv order (which is the dispreferred order in Mandarin). Neither group showed a consistent advantage for either clause order, nor an advantage for *before*- over *after*-sentences.

2.3.3. Bilinguals' individual differences

We then ran simple correlations to examine whether the bilingual participants' performance in the offline sentence-picture matching task (i.e., mean accuracy and mean RTs) correlated with their IELTS scores across the board and separately for specific sentence types (as we observed differences in accuracy and RTs according to sentence type). The eight bilinguals who did not have IELTS scores were excluded from the individual differences analyses, leaving 38 bilinguals.

The results showed that bilinguals' IELTS scores marginally significantly correlated with their overall mean accuracy, r(36) = 0.31, p = 0.054, and significantly correlated with their overall mean RTs, r(36) = -0.59, p < 0.001. Furthermore, as shown in Table 5, bilinguals' IELTS scores significantly correlated with the mean RTs of all sentence types, and with the accuracy of *after*-sentences in the main-adv order. These results suggest that bilinguals with better English proficiency were more

accurate and quicker in comprehending both *before*- and *after*-sentences overall compared to those with lower proficiency, and this correlation was especially pronounced in the accuracy of *after*-sentences in the main-adv order (which is the non-iconic and dispreferred order in Mandarin).

[Table 5 here]

However, these individual difference results should be interpreted with caution, as we relied on bilinguals' self-reported IELTS scores as a proxy for their English language proficiency. The timing of when bilinguals took the IELTS test likely varied and was not formally recorded, so their scores may not accurately reflect their current English proficiency. Moreover, the offline behavioural measure used in Experiment 1 may have lacked sensitivity to detect fine-grained differences in sentence comprehension, where both language groups performed with high accuracy. To address these concerns, we conducted Experiment 2—an online version of the sentence-picture matching task—to examine potential differences in the cognitive load required during the comprehension of different types of adverbial sentences. Additionally, we measured bilinguals' English proficiency at the time of testing to ensure a more accurate reflection of their current English proficiency.

3. Experiment 2

To measure cognitive load during the comprehension of adverbial sentences, we tracked participants' pupil size changes. Pupil size measurement (pupillometry) has been shown to be a reliable psychophysiological indicator of processing effort during syntactic comprehension (e.g., Ayasse et al., 2021; Chapman & Hallowell, 2021; Fernandez et al., 2018; Scherger et al., 2021). The more difficult a sentence is to comprehend, the greater the cognitive resources required, leading to larger increases in pupil size. Moreover, pupillometry has the advantage of allowing for more natural auditory stimulus presentation, resembling real-life listening scenarios, compared to methods like self-paced reading. It also provides a non-invasive and easy-to-set-up alternative to techniques such as EEG and fMRI.

Based on the results from Experiment 1, we pre-registered Experiment 2 in the Open Science Framework (OSF) repository, which specified research hypotheses, study design, sampling and analysis plan (https://osf.io/m7u6v).

3.1. Participants

A new group of 54 Mandarin-English bilingual adults and 47 English monolingual adults was included in the final sample[4]. They were also recruited through social media and the University of Manchester's credit system. The bilinguals (age range = 18-34 years, mean age = 21.80 years, SD = 3.58 years; 40 Females) spoke both Mandarin and English, with Mandarin as their native language, and were living in the UK at the time of testing. Their current English language ability was assessed using a standardised English language test (the Oxford Quick Placement Test, Oxford University Press, 2001), as part of this study. They were either pursuing or had obtained university degrees: undergraduates (N=39), master's (N=6), or PhD (N=9). The monolingual adults (age range = 18-22 years, mean age = 19.53 years, SD = 1.03 years; 28 Females) spoke only English and were all undergraduates. Neither the monolinguals nor the bilinguals had any known speech and language difficulties, nor vision or hearing impairments. In addition to the 54 bilingual and 47 monolingual adults in the final sample, one additional monolingual adult came to the lab but did not proceed with the study due to poor calibration.

3.2. Online sentence-picture matching task

In the online sentence-picture matching task (see Figure 3), each trial began with a 3000ms silent baseline period to establish the participant's baseline pupil size. The participant then listened to an audio-recorded test sentence (3185ms), followed by a 5000ms silent delay period. During this 8185ms time window, the participant fixated on a cross at the centre of the screen, and changes in their pupil size while processing the sentence were recorded. The test sentences were matched in duration to ensure that differences in pupil size were not due to variability of sentence duration. The

[4] The sample size was determined based on a priori power analysis using data from Experiment 1. To achieve at least 80% power, 40 participants were required in each language group. However, to account for potential dropouts and technical issues, we pre-registered to recruit up to 55 participants per language group.

delay period was included to fully capture changes in the participant's pupil size. Next, two pictures depicting the same events as described in the test sentence appeared on the screen. The participant was asked to determine whether the sequence of events described in the test sentence matched the sequence shown in the pictures using a gamepad (i.e., a Logitech controller) to respond. This step ensured that participants were attending to the test sentences and also allowed us to record their behavioural responses. RTs were measured from the onset of the picture presentation to the participant's click on a picture sequence. Pupil size changes were not recorded during this step.

[Figure 3 here]

3.2.1. Design

As in Experiment 1, the online sentence-picture matching task had three factors: one between-subjects factor (LanguageGroup: bilingual, monolingual), and two within-subject factors (ConnectiveType: *before*, *after*; ClauseOrder: main-adv, adv-main).

3.2.2. Materials

48 test sentences were constructed, with 12 test sentences per condition (four conditions: two ConnectiveType × two ClauseOrder). The first 24 test sentences were the same as those in Experiment 1 (see examples (11) and (12)). Then, by reversing the events in each of the first 24 sentences (while maintaining the original clause order), we created another 24 test sentences (see examples (13) and (14)). We doubled the number of test sentences in Experiment 2 to enhance statistical power, anticipating potential data loss (e.g., due to blinks) during pupillometry recordings.

(11) Block 1: *After she paints the old fence, she hoovers the house.*

(12) Block 2: *Before she paints the old fence, she hoovers the house.*

(13) Block 3: *After she hoovers the house, she paints the old fence.*

(14) Block 4: *Before she hoovers the house, she paints the old fence.*

Four different experimental lists were constructed. Each list consisted of four blocks, with one of the four versions of each test sentence containing the same event pair occurring in each block (see examples (11)-(14)). The order of the four blocks within each list was randomised. The order of

the test sentences within each block (N=12) was pseudo-randomised, allowing a maximum of two consecutive test sentences in the same condition. The corresponding picture sequence in each block was counterbalanced: in half of the test sentences, the sequence of events described in the test sentence matched the sequence shown in the pictures, and in the other half, it did not. List 2 was the same as List 1, with the difference that all main-adv sentences were turned into adv-main sentences and vice versa. Lists 3 and 4 were the same as Lists 1 and 2, respectively, but with the picture sequence corresponding to each test sentence reversed (see Table 6 below for examples). Participants were randomly assigned to one of the four experimental lists. All test sentences were generated as audio using the voice from a young female native speaker of British English. The corresponding picture sequences were borrowed from De Ruiter et al. (2018).

[Table 6 here]

3.2.3. Procedure

Participants completed the online sentence-picture matching task individually in an eye-tracking lab at the University of Manchester. The lab was kept quiet, with all curtains closed and the room lights left on to ensure consistent brightness across participants. Each participant sat in an adjustable chair, resting their chin on a chinrest positioned in front of an EyeLink 1000 Plus eye tracker (sampling rate: 500Hz; equipped with a 35mm lens; pupil size resolution: 0.2% of diameter) and a 13-inch Dell display screen. The eye tracker recorded the participant's pupil size from one eye only. The task was programmed using the SR Research Experiment Builder software.

The task began with adjusting the eye position and calibrating the eye tracker. Once calibrated, the participant put on headphones (for listening to test sentences) and held the gamepad, with two fingers on each hand positioned over the left and right buttons. The participant was then provided with the task instructions on the display screen, followed by four warm-up trials to familiarise themselves with the task, particularly with pressing the corresponding buttons on the gamepad. Note that we provided two versions of the gamepad setup to control for potential effects of handedness: In one version, the left button indicated that the sequence of events described in the

sentence matched the sequence shown in the pictures, and the right button indicated a mismatch. In the other version, the left button indicated a mismatch and the right button indicated a match. Participants were randomly assigned to one of the two versions. Participants had to complete all four warm-up trials correctly to proceed to the test phase. If they did not, the warm-up trials were repeated until completed correctly. The warm-up trials were similar to the test trials, except that the sentences followed the structure "First …, then…" (e.g., "First he switches the light on, then he waters his plant").

Once the participant passed the warm-up phase, they moved on to the test trials. After completing two blocks (i.e., half of the test sentences, N=24), the participant was given a break of up to 5 minutes. Overall, the online sentence-picture matching task took around 30 to 40 minutes to complete, including calibration and the break. After completing the task, monolingual participants left the lab, while bilingual participants proceeded to complete the standardised English language test, which took around 20 minutes.

3.3. Data analysis

Data processing and analysis were conducted as outlined in the pre-registration. We analysed participants' changes in pupil size, and their behavioural accuracy and RTs statistically one by one. Finally, we performed simple correlation analyses between bilingual participants' performance in sentence comprehension and their English test scores.

3.3.1 Changes in pupil size

The raw pupil data required pre-processing to identify and interpolate missing data (e.g., due to blinks), and to convert the data into a format suitable for statistical analysis (i.e., averaging). During blinks, the eye tracker either failed to record any pupil data, or recorded inaccurate pupil sizes (e.g., due to partial closures of the eyelids). Partial closure of the eyelids was identified by steep changes in pupil size adjacent to missing data and was marked as missing, along with an additional 25ms before and after the excluded data to ensure accurate pupil sizes were used in interpolation Mathôt & Vilotijević, 2023). Similarly, when participants were not fixating on the central cross or

when artifacts occurred, the recorded pupil size did not accurately reflect the true pupil size and was also coded as missing. While we could have simply removed the missing data, doing so would have significantly reduced statistical power. Therefore, we linearly interpolated sequences of missing data that were up to 750ms in duration. A threshold of 750ms was used to account for the standard duration of a blink (100-400ms; e.g., Schiffman, 2000) and an estimated 175ms of missing data on each side due to the eyelid opening and closing (de Gee et al., 2014; Nyström et al., 2016; Satterthwaite et al., 2007; van Rijn et al., 2012).

We then performed data exclusion on the pre-processed data. We excluded: 1) any trial that contained missing pupil data that could not be interpolated (i.e., gaps longer than 750ms), or any trial in which more than 50% of the data had been interpolated (to ensure data reliability at the trial level); 2) any trial with extreme baseline pupil sizes (i.e., those more than three standard deviations from the participant's mean baseline pupil size); and 3) any participant who contributed fewer than 20% of trials (N=10; to ensure data reliability at the participant level).

After data pre-processing and exclusion, 1810 out of 2592 responses (69.83%) from bilinguals, and 1910 out of 2256 responses (84.66%) from monolinguals were included in the statistical analysis for pupil responses. Figure 4 shows the pupil size changes of bilinguals and monolinguals during their comprehension of two types of adverbial sentences in two clause orders. Specifically, pupil size changes refer to the average change in pupil size during the comprehension of each trial relative to the baseline pupil size (pupil size was recorded in arbitrary units), with positive values indicating an increase in pupil size. It reveals that bilinguals showed larger changes in pupil size overall, compared to monolinguals. Additionally, there appears to be a three-way interaction between ClauseOrder, Type and LanguageGroup, such that although both groups show a similar general pattern of results, the differences between conditions appear more pronounced in monolinguals than in bilinguals.

[Figure 4 here]

We analysed the pupil data using the same strategy as for behavioural accuracy and RTs in Experiment 1, but with participants' pupil size changes used as the dependent variable. Model comparison indicates that only the fixed effect of LanguageGroup significantly improved model fit, χ^2 (1) = 18.71, $p < 0.001$, and was retained in the final model, confirming that bilinguals showed significantly larger changes in pupil size than monolinguals, $\beta = 78.31$, $SE(\beta) = 17.22$, $t = 4.55$, $p < 0.001$.

The exploratory model including the three-way interaction between ClauseOrder, Type and LanguageGroup only marginally improved the fit of the final model, χ^2 (6) = 11.53, $p = 0.073$ (see Table 7 for this exploratory model). As shown in Figure 4 and supported by post-hoc analyses in Table S4, monolinguals exhibited smaller changes in pupil size when comprehending *before*-sentences in the iconic main-adv order compared to the non-iconic adv-main order, and when comprehending *after*-sentences in the iconic adv-main order compared to the non-iconic main-adv order. Further support for this pattern is seen in main-adv orders, where monolinguals exhibited smaller changes in pupil size for iconic *before*-sentences than for non-iconic *after*-sentences, and vice versa for their comprehension in adv-main orders. However, no significant differences between orders were observed for either *before*- or *after-sentences* among bilinguals, although the largest differences were observed between *after*-sentences in the iconic (and preferred in Mandarin) adv-main clause order compared to the non-iconic (and dispreferred in Mandarin) main-adv order.

For monolinguals, these pupillometry results are consistent with the offline accuracy results from Experiment 1, showing a significant effect of iconicity for both *before*- and *after*-sentences, but no advantage for main-adv over adv-main orders or for *before*- over *after*-sentences alone. However, no significant effect of iconicity or clause order was observed in bilinguals' pupillometry data, although the largest differences occurred for *after*-sentences in iconic/preferred clause order compared to non-iconic/dispreferred clause order. The lack of clear findings in the pupillometry data from bilinguals may be due to a potential ceiling effect, given that our bilinguals exhibited relatively large changes in pupil size across the board, much larger than seen in monolinguals.

[Table 7 here]

3.3.2. Behavioural accuracy

In total, we collected 2592 responses from bilinguals, with 2352 correct and 240 incorrect, resulting in an accuracy of 90.74%, and 2256 responses from monolinguals, with 2141 responses correct and 115 incorrect, resulting in an accuracy of 94.90%. Both groups of participants in Experiment 2 were slightly more accurate than those in Experiment 1. This improvement may be attributed to the pupillometry design used in Experiment 2, which required participants to be more attentive, as well as the 5000ms delay period that allowed for more processing time before providing behavioural responses. In contrast, participants in Experiment 1 had to respond immediately. Despite the slight differences in study design, the accuracy patterns observed in this experiment mirrored those in Experiment 1 (see Figure 5). The final model revealed the same significant effects as those outlined for Experiment 1, demonstrating the robustness of the results (see Tables S5 and S6, Supplemental Material).

[Figure 5 here]

3.3.3. Behavioural reaction times

As in Experiment 1, we included only correct responses in the RT analyses and excluded responses with RTs that exceeded three standard deviations above or below the mean. In total, 2316 out of 2352 correct responses from bilinguals (excluded 36 outliers), and 2103 out of 2141 correct responses from monolinguals (excluded 38 outliers) were included. The mean RT for bilinguals was 2337.00ms (SD = 1614.40ms) and for monolinguals was 1824.25ms (SD = 1081.42ms).

Model comparison indicates that only the fixed effect of LanguageGroup significantly improved model fit, χ^2 (1) = 12.88, $p < 0.001$, and was retained in the final model. That is, monolinguals comprehended adverbial sentences significantly more quickly than bilinguals, β = 547.83, *SE(β)* = 147.82, $t = 3.71$, $p < 0.001$, consistent with the findings from Experiment 1. However, unlike in Experiment 1, we did not observe a significant interaction between LanguageGroup and Type. In Experiment 1, monolinguals comprehended *after*-sentences as quickly as *before*-sentences,

while bilinguals comprehended *before*-sentences significantly more slowly than *after*-sentences, apparently showing a particular disadvantage for *before*-sentences in the iconic main-adv order (the dispreferred order in Mandarin). In Experiment 2, as shown in Figure 6, there appears to be an overall advantage for iconic *before*-sentences for both language groups, but the model that included this interaction between ClauseOrder and Type (see Table 8) only marginally improved the fit of the final model, χ^2 (3) = 7.35, $p < 0.062$. We will return to these inconsistent RT results across the experiments in the Discussion section.

[Figure 6 here]

[Table 8 here]

3.3.4. Bilinguals' individual differences

Lastly, as outlined in our pre-registration, we ran simple correlations to test whether bilingual participants' performance in the online sentence-picture matching task (i.e., mean pupil size changes, mean accuracy and mean RTs) correlated with their English test scores across the board and separately for specific sentence types. As in Experiment 1, the results show that bilinguals' English test scores significantly correlated with their overall mean accuracy, $r(52) = 0.52$, $p < 0.001$, and with their overall mean RTs, $r(52) = -0.29$, $p = 0.033$, suggesting that bilinguals with better English were more accurate and quicker in comprehending adverbial sentences overall compared to those with poorer English ability. Furthermore, as shown in Table 9, bilinguals' English test scores significantly correlated with their accuracy for each sentence type, with the strongest correlation observed for *before*-sentences in the main-adv order (iconic, but the dispreferred order in Mandarin). Scores also significantly correlated with their RTs for both *before*- and *after*-sentences in the main-adv order.

On the other hand, we did not observe a significant correlation between bilinguals' English language ability and their overall mean pupil size changes, nor between English language ability and the mean pupil size change for each type of sentence (see also Table 9). These results suggest that bilinguals' English language ability did not affect differences in pupil size changes during sentence

comprehension. However, the large changes in pupil size across the board among bilinguals, along with the possibility that some may have reached a ceiling effect, could have obscured potential differences.

[Table 9 here]

Following an anonymous reviewer's suggestion, and given that bilinguals' English test scores significantly correlated with their overall mean accuracy and with accuracy for each sentence type, we further conducted an exploratory analysis by adding English test scores as a predictor into the final accuracy model for bilinguals to examine its independent contribution. The results showed that English test scores significantly improved model fit, χ^2 (1) = 17.63, $p < 0.001$, with higher English test scores associated with more accurate comprehension of adverbial sentences overall, $\beta = 0.09$, $SE(\beta) = 0.02$, $z = 4.56$, $p < 0.001$. However, no significant interactions were found between English test scores and ClauseOrder, Type, or the interaction of ClauseOrder and Type. This lack of interaction is likely due to limited statistical power or the relatively homogeneous proficiency levels within our UK university student sample.

4. Discussion

The present studies are novel in examining Mandarin-English bilingual adults' comprehension of English temporal *before*- and *after*-sentences compared to English monolingual adults, and in incorporating pupillometry to measure cognitive load associated with the comprehension of adverbial sentences, in addition to purely behavioural measures. We tested whether the general theoretical accounts of (complex) sentence processing apply to Mandarin-English bilingual adults as they do to monolinguals: (i) the semantic account relating to iconicity predicts that *after*-sentences in the iconic adv-main order and *before*-sentences in the iconic main-adv order should be better comprehended than non-iconic orders; (ii) the semantic account relating to semantic consistency predicts that *before*-sentences should be better comprehended than *after*-sentences; and (iii) the processing-based account predicts that main-adv orders should be better comprehended than adv-main orders overall. Alternatively, language-specific usage patterns in

Mandarin might lead bilinguals to exhibit crosslinguistic influence from Mandarin to English, such that, unlike monolinguals, bilinguals might comprehend adv-main orders better than main-adv orders, as Mandarin strongly prefers adv-main orders (Hypothesis (iv)). Additionally, we explored the link between bilinguals' sentence comprehension and their English language proficiency (Hypothesis (v)).

4.1. The competition between semantic, processing-based, and crosslinguistic influence and usage accounts

Using both offline and online pupillometry sentence-picture matching tasks, we found that English monolingual adults comprehended both *before*- and *after*-sentences more accurately when the sentences were in iconic order than in non-iconic order. Iconic order also elicited smaller changes in pupil size during monolinguals' processing compared to non-iconic order, indicating that less cognitive effort was required. Therefore, our monolingual results are consistent with findings from previous studies using EEG and fMRI with the same population (Münte et al., 1998; Politzer-Ahles et al., 2017; Ye et al., 2012), supporting the semantic account of iconicity (Hypothesis (i)), but not the processing-based account, which predicts a general processing advantage for main-adv orders (Hypothesis (iii)).

Moreover, consistent with the previous study by Politzer-Ahles et al. (2017), we did not observe that monolinguals showed any overall difference in comprehension between *before*- and *after*-sentences, despite the semantic differences associated with them (contradicting Hypothesis (ii)). As mentioned in the Introduction, young children have been found to comprehend *before*-sentences more accurately than *after*-sentences overall, likely because "before" exhibits a more consistent form-meaning mapping as a temporal connective than "after" (e.g., De Ruiter et al., 2018). However, the difference between "before" and "after" may be less pronounced in adults' comprehension due to their richer lexical and semantic knowledge.

Similar to English monolinguals, Mandarin-English bilinguals also showed no overall difference in comprehension between *before*- and *after*-sentences. However, unlike English

monolinguals, bilinguals' comprehension patterns for each type of sentence were different. They showed an effect of iconicity and/or clause order only for *after*-sentences, as reflected in the accuracy data and with the largest differences going in that direction for pupil size changes. We suggest that this asymmetry arises from an interaction between crosslinguistic influence from the bilinguals' dominant language, Mandarin, on English and iconicity. Specifically, for *after*-sentences, the preferred order in Mandarin, the adv-main order, is also the iconic order. Thus, both cues facilitate bilinguals' comprehension of *after*-sentences in the adv-main order, compared to non-iconic *after*-sentences in the main-adv order, for which neither cue is preferred. In contrast, for *before*-sentences, the two cues compete. The preferred adv-main order in Mandarin is non-iconic, whereas the iconic order corresponds to the dispreferred main-adv order. This competition may have reduced the facilitative effect of either cue, which explains why we observed no clear preference for either order in *before*-sentences[5]. This explanation is consistent with the Competition Model of second language (L2) learning (e.g., MacWhinney, 1987, 2018; Tokowicz & MacWhinney, 2005), which proposes that adult L2 learning involves massive transfer from the first language (L1). When processing cues in L1 and L2 align, positive transfer occurs, facilitating L2 processing. When cues in L1 and L2 compete, this leads to processing difficulty and negative transfer from the L1.

Further support for this interaction between crosslinguistic influence and iconicity was found in the bilinguals' RTs in Experiment 1. They processed *before*-sentences more slowly than *after*-sentences, most likely driven by the fact that *before*-sentences in the main-adv order (iconic, but the

[5] As suggested by an anonymous reviewer, we acknowledge an alternative explanation for the absence of the iconicity effect in *before*-sentences: the effect of iconicity is mediated by the position of the connective (e.g., Makrodimitris & Schulz, 2025a, b). This explanation suggests that non-iconic *after*-sentences (in main-adv order) should be the most difficult to comprehend because the connective "after" appears mid-sentence. This requires the listener to revise their initial representation of the first-mentioned event, potentially leading to more errors or a higher processing load. In contrast, in non-iconic *before*-sentences (in adv-main order), the sentence-initial connective "before" immediately signals that reordering is required. This allows the listener to construct the correct event representation from the outset, thereby facilitating better comprehension. Under this view, the absence of the iconicity effect in *before*-sentences results from high accuracy in both iconic and non-iconic conditions, whereas the iconicity effect for *after*-sentences is driven by considerably lower accuracy in the non-iconic condition. Nevertheless, this explanation does not align with our bilingual data; across both experiments, non-iconic *after*-sentences were not statistically the most difficult to comprehend regarding accuracy, pupil size changes, or RTs.

dispreferred order in Mandarin) were comprehended the slowest compared to other conditions (whereas monolinguals did not show any differences in RTs). However, we must interpret the RT results with caution, as in Experiment 2, bilinguals, like monolinguals, showed a trend toward processing *before*-sentences in the main-adv order more quickly than in the reverse order. We suspect that differences in study design may explain these differences in RTs. In Experiment 1, participants viewed two picture sequences immediately after hearing the test sentence and had to select the sequence that matched the sentence, whereas in Experiment 2, after hearing the sentence there was a 5000ms delay period before participants viewed a single picture sequence and indicated whether the picture sequence they saw was a match or mismatch for the sentence. The immediate response demands in Experiment 1 may have increased processing difficulty for bilinguals when hearing *before*-sentences in the main-adv order, where iconicity conflicted with their preferred word order, resulting in slower responses. Monolinguals, who do not face the same difficulty (since the iconic order of *before*-sentences is also the preferred order in English), were not affected in the same way. In Experiment 2, however, the long delay period may have given bilinguals enough time to resolve this difficulty (i.e., to determine event order). Alternatively, the long delay may have prompted all participants to reconstruct the sentence from memory rather than verbatim, and sentences that are both iconic and in the main-adv order (referencing the processing-based account) may be easier to reconstruct. However, these interpretations remain speculative, and further investigation will be needed to determine how task demands and delay duration affect RTs.

Overall, our findings indicate that although iconicity affects English monolinguals' processing of temporal *before*- and *after*-sentences, it does not uniformly apply to individuals with different language backgrounds—iconic sentences were not always easier for our bilinguals to comprehend. However, nor did they rely exclusively on clause order or connective type. There appeared to be an interaction between crosslinguistic influence and iconicity. Sentences where both cues pointed to the same meaning, hypothesised to be easier to process (*after*-sentences in adv-main order, iconic and preferred order in Mandarin), had an advantage over sentences where both cues pointed to the

meaning hypothesised to be more difficult to process (*after*-sentences in main-adv order, non-iconic and dispreferred order in Mandarin). In contrast, sentences where the cues pointed to conflicting meanings in terms of ease of processing (*before*-sentences that are either iconic and in the dispreferred order for Mandarin, or non-iconic and in the preferred order for Mandarin) showed a mutual reduction in the facilitative effect of the other. Future studies could examine whether our conclusions apply to other temporal connectives (e.g., sequential clauses with "when"), thereby enhancing the generalisability of research on adverbial sentences.

4.2. Bilinguals' sentence comprehension correlates with their English proficiency

The correlation results on individual differences among bilinguals show that as their English proficiency increased, their comprehension of English temporal adverbial sentences improved in terms of both accuracy and RTs, though the RT results should be interpreted with caution. In terms of accuracy, this improvement was particularly pronounced for sentences in the main-adv order, a consistent finding observed with *after*-sentences in Experiment 1 and *before*-sentences in Experiment 2 (Hypothesis (v)). These findings suggest that as bilinguals become more proficient in English, they gain more exposure to English adverbial sentences, especially those in the main-adv order, which are rarely heard and used in Mandarin, leading to more accurate comprehension of these sentences. This aligns with the corpus-based study by Li and Long (2022), which also reported that the proportion of main-adv orders increased in Chinese learners of English's writing of temporal adverbial sentences (e.g., *when*, *before*, *after*, *until*) as their English proficiency improved. However, it is worth noting that while the correlations suggested a proficiency-related gain specifically for main-adv orders, the exploratory inclusion of English test scores in the final accuracy model (Experiment 2) did not yield a significant interaction between proficiency and clause order. This lack of significance is likely due to limited statistical power or the relatively homogeneous proficiency levels within our UK university sample. We acknowledge this as a limitation and suggest that future research investigate larger groups of Mandarin-English bilinguals across a broader range of English proficiency.

Regarding the lack of a correlation between bilinguals' English language ability and their pupil size changes during sentence comprehension, this is likely due to the large pupil size changes observed across the board among bilinguals, with some possibly reaching a ceiling effect. This is consistent with previous studies showing larger pupil sizes in general when processing sentences in a second language (Borghini & Hazan, 2018), which may have masked more subtle processing differences across the conditions. In light of this, we suggest that future studies could refine the stimulus setup to reduce the risk of ceiling effects. Pupil size naturally fluctuates due to various factors, including cognitive load (the primary focus of this study), as well as ambient room lighting and the brightness of the visual display. In this study, we followed standard guidelines of intermediate ambient lighting to ensure our pupil measures stayed within the physiological lower and upper limits (Mathot & Vilotijevic, 2023). However, our results suggest that we may have approached the upper limit due to the intensive cognitive demands associated with bilinguals' processing of complex sentences. Future work could consider using higher levels of ambient lighting and visual display brightness, allowing pupils to begin from smaller baseline sizes and thereby enabling larger measurable dilations and increased sensitivity to cognitive load effects.

5. Conclusions

In two studies, we examined how Mandarin-English bilingual and English monolingual adults comprehended English temporal *before*- and *after*-sentences, to determine whether they employ the same comprehension strategies—semantic and processing-based accounts, or exhibit differences reflecting crosslinguistic influence from the bilinguals' dominant language, Mandarin. Using both offline and online pupillometry sentence-picture matching tasks, we found that, unlike English monolinguals who rely on the semantic factor of iconicity to comprehend both types of sentences, Mandarin-English bilinguals showed the effect of iconicity and/or clause order only in *after*-sentences, but not in *before*-sentences. This is likely because the competition resulting from crosslinguistic influence from the preferred clause order in Mandarin diminishes any facilitative effect of iconicity, while iconicity diminishes any facilitative effect of the preferred clause order in

Mandarin on their comprehension of *before*-sentences. As bilinguals' English proficiency increased, their comprehension accuracy of adverbial sentences improved, particularly for those with a clause order rarely heard and used in Mandarin. Our results therefore contribute to current theories and literature by showing that iconicity does not universally facilitate the processing of adverbial clauses. Language-specific usage patterns, such as the preferred order of main and adverbial clauses, also influence how individuals construct mental representations of events during sentence comprehension and the cognitive load required.

Replication package

All research materials, data, and analysis code are available in the OSF repository (https://osf.io/7vmu4/).

Conflict of interest disclosure: None.

References

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239-273. https://doi.org/10.1017/S030500091400049X.

Ayasse, N. D., Hodson, A. J., & Wingfield, A. (2021). The principle of least effort and comprehension of spoken sentences by younger and older adults. *Frontiers in Psychology*, *12*, 629464. https://doi.org/10.3389/fpsyg.2021.629464.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005.

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.

Blything, L. P., & Cain, K. (2016). Children's processing and comprehension of complex sentences containing temporal connectives: The influence of memory on the time course of accurate responses. *Developmental Psychology*, *52*(10), 1517. https://doi.org/10.1037/dev0000201.

Blything, L. P., & Cain, K. (2019). The role of memory and language ability in children's production of two-clause sentences containing before and after. *Journal of Experimental Child Psychology*, *182*, 61-85. https://doi.org/10.1016/j.jecp.2019.01.011.

Blything, L. P., Davies, R., & Cain, K. (2015). Young children's comprehension of temporal relations in complex sentences: The influence of memory on performance. *Child Development*, *86*(6), 1922-1934. https://doi.org/10.1111/cdev.12412.

Borghini, G., & Hazan, V. (2018). Listening effort during sentence processing is increased for non-native listeners: A pupillometry study. *Frontiers in Neuroscience*, *12*, 152. https://doi.org/10.3389/fnins.2018.00152.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. University of California Press.

Chapman, L. R., & Hallowell, B. (2021). Expecting questions modulates cognitive effort in a syntactic processing task: Evidence from pupillometry. *Journal of Speech, Language, and Hearing Research*, *64*(1), 121-133. https://doi.org/10.1044/2020_JSLHR-20-00071.

Clark, E. V. (1971). On the acquisition of the meaning of before and after. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 266-275. https://doi.org/10.1016/S0022-5371(71)80054-3.

Croft, W. (2004). *Typology and Universals* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511840579.

Diessel, H. (2001). The ordering distribution of main and adverbial clauses: A typological study. *Language*, *77*(3), 433-455. https://doi.org/10.1353/lan.2001.0152.

Diessel, H. (2005). Competing motivations for the ordering of main and adverbial clauses. *Linguistics*, *43*(3), 449-470. https://doi.org/10.1515/ling.2005.43.3.449.

Diessel, H. (2008). Iconicity of sequence: A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics*, *19*(3), 465-490. http://dx.doi.org/10.1515/COGL.2008.018.

de Gee, J. W., Knapen, T., & Donner, T. H. (2014). Decision-related pupil dilation reflects upcoming choice and individual bias. *Proceedings of the National Academy of Sciences*, *111*(5), E618-E625. https://doi.org/10.1073/pnas.1317557111.

De Ruiter, L. E., Lemen, H. C., Lieven, E. V., Brandt, S., & Theakston, A. L. (2021). Structural and interactional aspects of adverbial sentences in English mother-child interactions: an analysis of two dense corpora. *Journal of Child Language*, *48*(6), 1150-1184.

De Ruiter, L. E., Lieven, E. V., Brandt, S., & Theakston, A. L. (2020). Interactions between givenness and clause order in children's processing of complex sentences. *Cognition*, *198*, 104130. https://doi.org/10.1016/j.cognition.2019.104130.

De Ruiter, L. E., Theakston, A. L., Brandt, S., & Lieven, E. V. (2018). Iconicity affects children's comprehension of complex sentences: The role of semantics, clause order, input and individual differences. *Cognition*, *171*, 202-224. https://doi.org/10.1016/j.cognition.2017.10.015.

Fernandez, L., Höhle, B., Brock, J., & Nickels, L. (2018). Investigating auditory processing of syntactic gaps with L2 speakers using pupillometry. *Second Language Research*, *34*(2), 201-227. https://doi.org/10.1177/0267658317722386.

Fang, Z. (2009). *Zhongguo EFL Xuexizhe Shijian Zhuangyu Congju Weizhi Yanjiu* [A study on the positioning of temporal adverbial clauses by Chinese EFL learners]. *Foreign Languages Research*, (6), 56-60.

Givón, T. (1985). Iconicity, isomorphism, and non-arbitrary coding in syntax. In: Haiman, J. (Ed), *Iconicity in Syntax: Proceedings of a symposium on iconicity in syntax, Stanford, June 24–26, 1983 (Typological Studies in Language)* (pp. 187-220). John Benjamins Publishing. https://doi.org/10.1075/tsl.6.10giv.

Halliday, M. A., & McDonald, E. (2004). Metafunctional profile of the grammar of Chinese. In: Caffarel, A., Martin, J. R., & Matthiessen, C. M (Eds), *Language Typology: A functional*

perspective (pp. 305-396). John Benjamins Publishing Company. https://doi.org/10.1075/cilt.253.

Hawkins, J. A. (1990). A parsing theory of word order universals. *Linguistic Inquiry*, *21*(2), 223-261.

Hawkins, J.A. (1992). Syntactic Weight Versus Information Structure in Word Order Variation. In: Jacobs, J. (Ed), *Informationsstruktur und Grammatik* (pp. 196-219). VS Verlag für Sozialwissenschaften Wiesbaden. https://doi.org/10.1007/978-3-663-12176-3_7.

Hawkins, J. A. (1994). *A performance theory of order and constituency* (No. 73). Cambridge University Press.

Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, *3*(3), 227-244. https://doi.org/10.1017/S1366728900000353.

Lenth, R. V., Bolker, B., Buerkner, P., Giné-Vázquez, I., Herve, M., Jung, M., Love, J., Miguez, F., Riebl, H., & Singmann, H. (2023). Emmeans: Estimated Marginal Means, aka Least-Squares Means. R Package Version 1.8.8.

Li, W., & Long, Y. (2022). A Development study on the ordering distribution of temporal adverbial clauses by Chinese EFL learners based on dependency treebank. *Chinese Journal of Applied Linguistics*, *45*(4), 551-565. https://doi.org/10.1515/CJAL-2022-0404.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. University of California Press. https://doi.org/10.1525/9780520352858.

MacWhinney, B. (1987). Applying the competition model to bilingualism. *Applied Psycholinguistics*, *8*(4), 315-327. https://doi.org/10.1017/S0142716400000357.

MacWhinney, B. (2018). A unified model of first and second language learning. In Hickmann, M., Jisa, H., & Veneziano, E. (Eds.), *Sources of Variation in First Language Acquisition* (pp. 287-312). John Benjamins Publishing Company. https://doi.org/10.1075/tilar.22.15mac.

Makrodimitris, C., & Schulz, P. (2025a). Comprehension of complex sentences containing temporal connectives: How children are led down the event-semantic kindergarten-path. *Journal of Child Language*, *52*(3), 615-647. https://doi.org/10.1017/S0305000924000205.

Makrodimitris, C., & Schulz, P. (2025b). Age of onset does not matter for bilingual children's understanding of late-acquired phenomena: The case of temporal connectives. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.24096.mak.

Mathôt, S., & Vilotijević, A. (2023). Methods in cognitive pupillometry: Design, preprocessing, and statistical analysis. *Behavior Research Methods*, *55*(6), 3055-3077. https://doi.org/10.3758/s13428-022-01957-7.

Müller, N., & Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition*, *4*(1), 1-21. https://doi.org/10.1017/S1366728901000116.

Münte, T. F., Schiltz, K., & Kutas, M. (1998). When temporal terms belie conceptual order. *Nature*, *395*(6697), 71-73.

Nyström, M., Hansen, D. W., Andersson, R., & Hooge, I. (2016). Why have microsaccades become larger? Investigating eye deformations and detection algorithms. *Vision Research*, *118*, 17–24. https:// doi.org/10.1016/j.visres.2014.11.007.

Oxford University Press. (2001). *Quick Placement Test*. Oxford University Press.

Politzer-Ahles, S., Xiang, M., & Almeida, D. (2017). " Before" and" after": Investigating the relationship between temporal connectives and chronological ordering using event-related potentials. *PloS one*, *12*(4), e0175199. https://doi.org/10.1371/journal.pone.0175199.

Satterthwaite, T. D., Green, L., Myerson, J., Parker, J., Ramaratnam, M., & Buckner, R. L. (2007). Dissociable but inter-related systems of cognitive control and reward during decision making: Evidence from pupillometry and event-related fMRI. *NeuroImage*, *37*, 1017–1031. https://doi.org/10.1016/j.neuroimage.2007.04.066.

Scherger, A. L., Urbanczik, G., Ludwigs, T., & Kizilirmak, J. M. (2021). The bilingual native speaker competence: evidence from explicit and implicit language knowledge using elicited production, sentence-picture matching, and Pupillometry. *Frontiers in Psychology*, *12*, 717379. https://doi.org/10.3389/fpsyg.2021.717379.

Schiffman, H. R. (2000). *Sensation and Perception: An Integrated Approach (*5th ed.). John Wiley & Sons.

Theakston, A. (2020). Where form meets meaning in the acquisition of grammatical constructions. In C. Rowland, A. Theakston, B. Ambridge, & K. Twomey (Eds.), *Current Perspectives on Child Language Acquisition: How children use their environment to learn* (Vol. 27, pp. 131-154). John Benjamins. https://doi.org/10.1075/tilar.27.

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.

Van Rijn, H., Dalenberg, J. R., Borst, J. P., & Sprenger, S. A. (2012). Pupil dilation co-varies with memory strength of individual traces in a delayed response paired-associate task. *PLoS One*, *7*(12), e51134. https://doi.org/10.1371/journal.pone.0051134.

Vos, S. H., Gunter, T. C., Kolk, H. H., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, *38*(1), 41-63. https://doi.org/10.1111/1469-8986.3810041.

Wagner, L., & Holt, R. F. (2023). Time after time: Factors influencing children's comprehension of Before and After. *Journal of Child Language*, 1-9. https://doi.org/10.1017/S0305000923000612.

Wang, Y. F. (2002). The preferred information sequences of adverbial linking in Mandarin Chinese discourse. *Text & Talk*, *22*(1), 141-172. https://doi.org/10.1515/text.2002.002.

Ye, Z., Kutas, M., George, M. S., Sereno, M. I., Ling, F., & Münte, T. F. (2012). Rearranging the world: Neural network supporting the processing of temporal connectives. *NeuroImage*, *59*(4), 3662-3667. https://doi.org/10.1016/j.neuroimage.2011.11.039.

Yeh, H-C., 2000. Temporal and conditional clauses in Chinese spoken discourse: a function-based study. In GIkeya, A., & Kawamori, M. (Eds.), *Proceedings of the 14th Pacific Asia Conference on Language, Information and Computation* (pp.365-376).

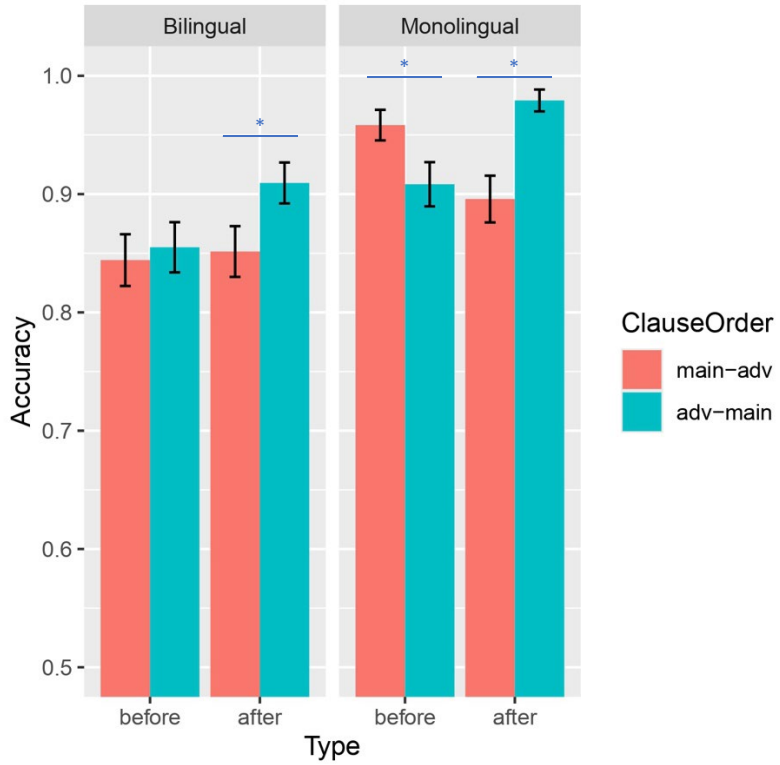
Zhang, S., Brandt, S., & Theakston, A. (2025). The role of iconicity in children's production of adverbial clauses. *Cognition*, *259*, 106119. https://doi.org/10.1016/j.cognition.2025.106119.

Zhang, S., Brandt, S., Warren, E., and Theakston, A. (submitted). *Greater Influence of Iconicity in Children with Developmental Language Disorder than in Typically Developing Children: Evidence from Adverbial Sentence Comprehension.*

Zhang, S., Junge, B., Lieven, E., Brandt, S., & Theakston, A. (2023). The Competition Between Processing and Discourse-Pragmatic Factors in Children's and Adults' Production of Adverbial When-Clauses. *Journal of Speech, Language, and Hearing Research*, *66*(12), 5048-5060. https://doi.org/10.1044/2023_JSLHR-23-00238.

Figure 1

Proportion of correct responses for bilinguals' and monolinguals' comprehension of adverbial sentences (Experiment 1)



Note: The asterisk above the comparison of two orders for one connective indicates a significant difference.

Figure 2

RTs for bilinguals' and monolinguals' comprehension of adverbial sentences (Experiment 1)

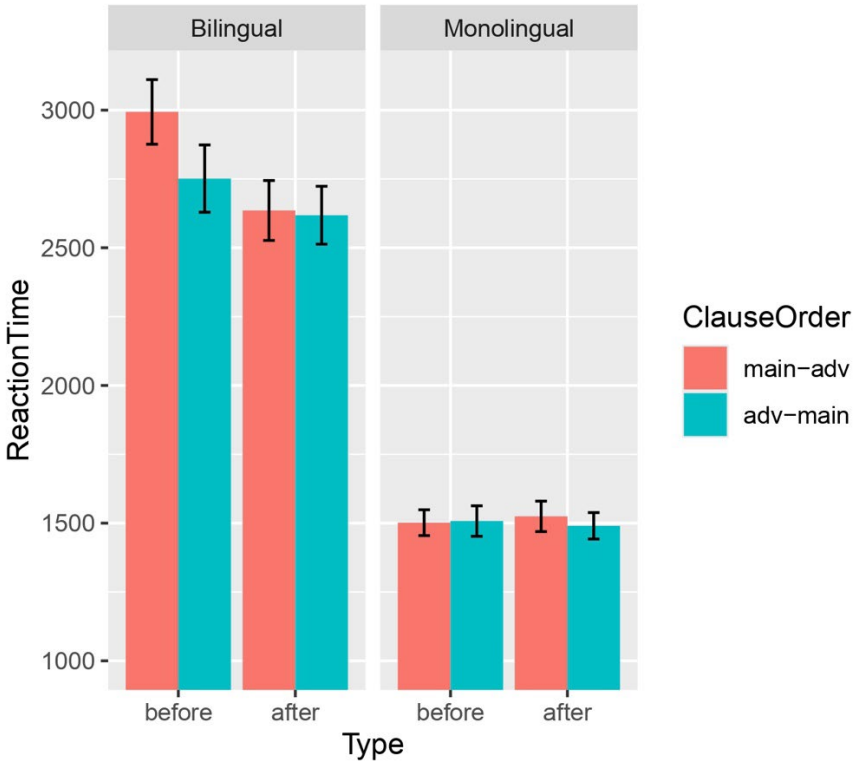


Figure 3

Example of the time course of each trial

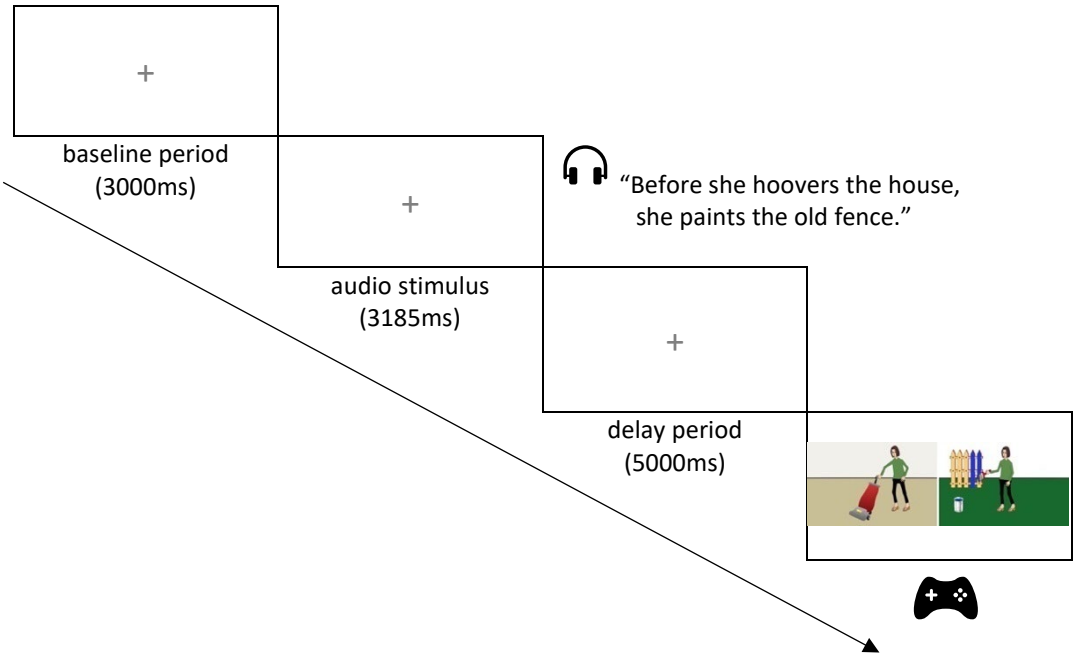
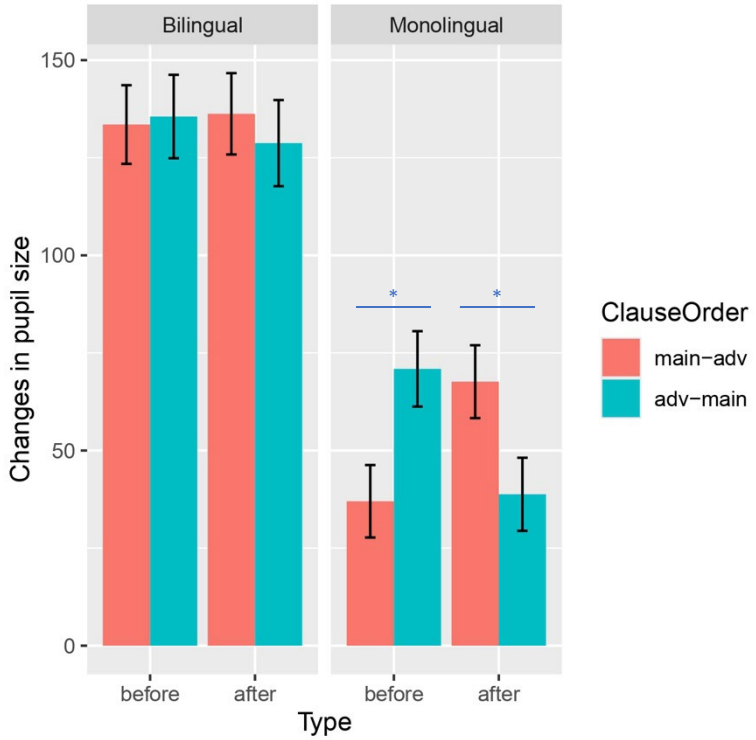


Figure 4

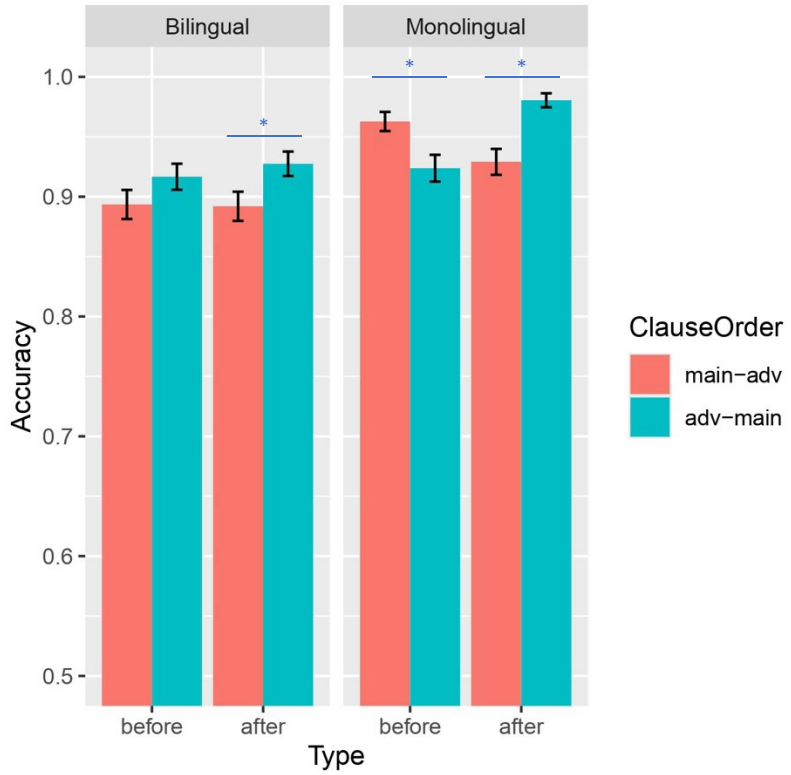
Increases in pupil size (arbitrary units) during bilinguals' and monolinguals' comprehension of adverbial sentences (Experiment 2)



Note: The asterisk above the comparison of two orders for one connective indicates a significant difference.

Figure 5

Proportion of correct responses for bilinguals' and monolinguals' comprehension of adverbial sentences (Experiment 2)



Note: The asterisk above the comparison of two orders for one connective indicates a significant difference.

Figure 6

RTs for bilinguals' and monolinguals' comprehension of adverbial sentences (Experiment 2)

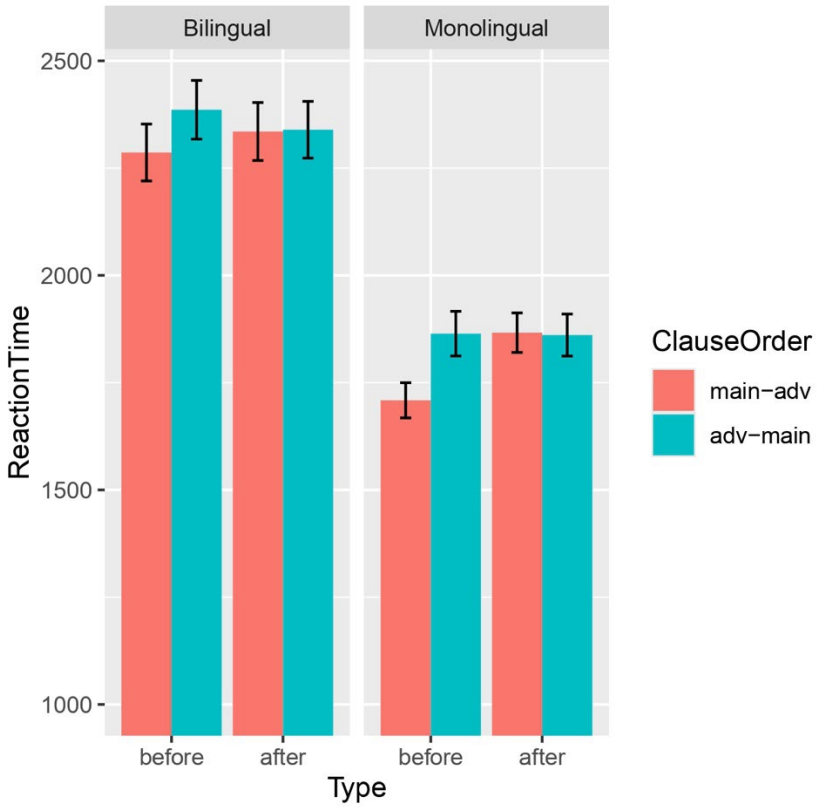


Table 1

The predictions from theoretical accounts on the comprehension of temporal adverbial sentences

Connective	Clause order	Sentence	Semantic account	Processing-based account	Crosslinguistic influence and usage patterns
Before	adverbial-main	(3) ***Before*** *she paints the old fence, she hoovers the house.*			✓
	main-adverbial	(4) *She hoovers the house* ***before*** *she paints the old fence.*	✓	✓	
After	adverbial-main	(5) ***After*** *she hoovers the house, she paints the old fence.*	✓		✓
	main-adverbial	(6) *She paints the old fence* ***after*** *she hoovers the house.*		✓	

Note. Tick mark refers to the clause order that the account would predict to be easiest to comprehend.

Table 2

Structure of the experimental trials

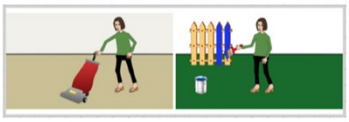
	Visual presentation	Auditory presentation
	(Blank screen)	"Before she hoovers the house, she paints the old fence." *beep*
top picture sequence →		
bottom picture sequence →		

Table 3

Final model for bilinguals' and monolinguals' comprehension of adverbial sentences in terms of behavioural accuracy (Experiment 1)

	β	*SE(β)*	*z*	*p*
(Intercept)	3.03	0.28	10.62	< 0.001
LanguageGroup [bilingual]	**-1.21**	**0.32**	**-3.77**	**< 0.001**
ClauseOrder [main-adv]	**-0.40**	**0.19**	**-2.13**	**0.033**
Type [*before*]	-0.32	0.18	-1.73	0.084
ClauseOrder [main-adv]: Type [*before*]	**1.73**	**0.37**	**4.65**	**< 0.001**
LanguageGroup [bilingual]: ClauseOrder [main-adv]	0.11	0.37	0.30	0.767
LanguageGroup [bilingual]: Type [*before*]	-0.05	0.37	-0.13	0.900
LanguageGroup [bilingual]: ClauseOrder [main-adv]: Type [*before*]	**-2.36**	**0.74**	**-3.18**	**0.001**

Note. Reference (or omitted) level: LanguageGroup = monolingual; ClauseOrder = adv-main; Type = *after*. Number of observations = 2064. Significant effects and interactions are highlighted in bold.

Table 4

Final model for bilinguals' and monolinguals' comprehension of adverbial sentences in terms of behavioural RTs (Experiment 1)

	β	*SE(β)*	*t*	*p*
(Intercept)	2204.55	137.99	15.98	< 0.001
LanguageGroup [bilingual]	**1318.14**	**183.88**	**7.17**	**< 0.001**
Type [*before*]	**136.25**	**49.96**	**2.73**	**0.006**
LanguageGroup [bilingual]: Type [*before*]	**227.32**	**99.92**	**2.28**	**0.023**

Note. Reference (or omitted) level: LanguageGroup = monolingual; Type = *after*. Number of observations = 1824. Significant effects and interaction are highlighted in bold.

Table 5

Correlations between bilinguals' IELTs scores and their mean accuracy and RTs in the offline sentence-picture matching task (Experiment 1)

Sentences	**Mean accuracy**	**Mean RTs**
main-*after*	0.46**	-0.44**
after-main	0.02	-0.77***
main-*before*	0.23	-0.40*
before-main	0.07	-0.44**

*Note. *p < 0.05. **p < 0.01. ***p < 0.001.*

Table 6

Examples of how the sequence of events in the test sentence and in the pictures match across four experimental lists

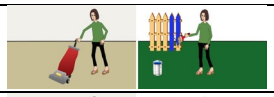
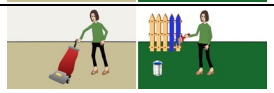
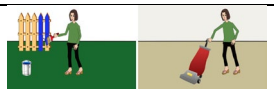
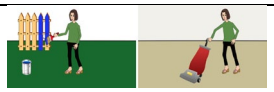
List	Test sentence	Picture sequence	Matching
List 1	*Before she hoovers the house, she paints the old fence.*		MISMATCH
List 2	*She paints the old fence before she hoovers the house.*		MISMATCH
List 3	*Before she hoovers the house, she paints the old fence.*		MATCH
List 4	*She paints the old fence before she hoovers the house.*		MATCH

Table 7

Exploratory model for bilinguals' and monolinguals' comprehension of adverbial sentences in terms of pupil size changes (Experiment 2)

	β	*SE(β)*	*t*	*p*
(Intercept)	94.80	9.62	9.86	< 0.001
LanguageGroup [bilingual]	**78.43**	**17.21**	**4.56**	**< 0.001**
ClauseOrder [main-adv]	-0.04	6.57	-0.01	0.995
Type [*before*]	1.17	6.57	0.18	0.859
ClauseOrder [main-adv]: Type [*before*]	**-35.32**	**13.13**	**-2.69**	**0.007**
LanguageGroup [bilingual]: ClauseOrder [main-adv]	7.05	13.14	0.54	0.591
LanguageGroup [bilingual]: Type [*before*]	1.58	13.15	0.12	0.904
LanguageGroup [bilingual]: ClauseOrder [main-adv]: Type [*before*]	**50.28**	**26.26**	**1.92**	**0.056**

Note. Reference (or omitted) level: LanguageGroup = monolingual; ClauseOrder = adv-main; Type = *after*. Number of observations = 3720. Significant and marginally significant effects are highlighted in bold.

Table 8

Significant fixed effect and marginal interaction for bilinguals' and monolinguals' comprehension of adverbial sentences in terms of behavioural RTs (Experiment 2)

	β	*SE(β)*	*t*	*p*
(Intercept)	2123.41	91.08	23.31	< 0.001
LanguageGroup [bilingual]	**546.54**	**147.81**	**3.70**	**< 0.001**
ClauseOrder [main-adv]	-56.88	35.95	-1.58	0.114
Type [*before*]	-41.09	35.97	-1.14	0.253
ClauseOrder [main-adv]: Type [*before*]	**-135.40**	**71.90**	**-1.88**	**0.060**

Note. Reference (or omitted) level: LanguageGroup = monolingual; ClauseOrder = adv-main; Type = *after*. Number of observations = 4419. Significant and marginally significant effects are highlighted in bold.

Table 9

Correlations between bilinguals' English language ability and their mean accuracy and RTs in the online sentence-picture matching task (Experiment 2)

Sentences	Mean accuracy	Mean RTs	Mean pupil size changes
main-*after*	0.30*	-0.30*	-0.14
after-main	0.30*	-0.26	-0.16
main-*before*	0.51***	-0.30*	-0.15
before-main	0.39**	-0.24	-0.05

*Note. *p < 0.05. **p < 0.01. ***p < 0.001.*

Table S1

Experimental sentences for the experimental List 1

Session	Trial No.	Item No.	Test sentence
1	1	1	After she paints the old fence, she hoovers the house.
	2	7	Before he reads his new book, he plays his big drum.
	3	9	She breaks her small train before she builds a tower.
	4	5	She hides over there after she runs over here.
	5	6	After she dances around, she bounces away.
	6	10	She hops up and down before she crawls on the floor.
	7	11	Before he shouts out loudly, he drives away fast.
	8	4	He laughs really hard after he coughs a few times.
	9	8	She takes a hot bath before she draws a picture.
	10	2	After he sweeps the new floor, he watches TV.
	11	12	Before he waves happily, he swims on his back.
	12	3	He drinks some water after he eats a green pear.
2	1	4	He laughs really hard before he coughs a few times.
	2	11	After he shouts out loudly, he drives away fast.
	3	9	She breaks her small train after she builds a tower.
	4	6	Before she dances around, she bounces away.
	5	7	After he reads his new book, he plays his big drum.
	6	8	She takes a hot bath after she draws a picture.
	7	1	Before she paints the old fence, she hoovers the house.
	8	5	She hides over there before she runs over here.
	9	12	After he waves happily, he swims on his back.
	10	10	She hops up and down after she crawls on the floor.
	11	2	Before he sweeps the new floor, he watches TV.
	12	3	He drinks some water before he eats a green pear.

Table S2

Pairwise comparisons of the effects of ClauseOrder and Type on bilinguals' and monolinguals' comprehension of adverbial sentences in terms of behavioural accuracy (Experiment 1)

Effects		Contrast	*β*	*SE(β)*	*z*	*p*
bilingual	*before*	main-adv – adv-main	-0.06	0.25	-0.26	0.798
	after		**-0.62**	**0.28**	**-2.19**	**0.028**
monolingual	*before*		**1.01**	**0.40**	**2.51**	**0.012**
	after		**-1.91**	**0.50**	**-3.83**	**< 0.001**
bilingual	main-adv	*before – after*	-0.07	0.25	-0.27	0.788
	adv-main		**-0.62**	**0.28**	**-2.20**	**0.028**
monolingual	main-adv		**1.16**	**0.40**	**2.93**	**0.003**
	adv-main		**-1.75**	**0.50**	**-3.50**	**< 0.001**
before	main-adv	bilingual – monolingual	**-1.77**	**0.46**	**-3.87**	**< 0.001**
	adv-main		-0.70	0.39	-1.79	0.074
after	main-adv		-0.54	0.38	-1.41	0.158
	adv-main		**-1.83**	**0.56**	**-3.26**	**0.001**

Note. Significant effects are highlighted in bold.

Table S3

Pairwise comparisons of bilinguals' and monolinguals' comprehension of two types of adverbial clauses in terms of behavioural RTs (Experiment 1)

Effects	Contrast	*β*	*SE(β)*	*t*	*p*
bilingual	*before – after*	**249.90**	**69.60**	**3.59**	**< 0.001**
monolingual		22.60	71.80	0.32	0.753
before	bilingual – monolingual	**1432.00**	**192.00**	**7.44**	**< 0.001**
after		**1204.00**	**192.00**	**6.28**	**< 0.001**

Note. Significant effects are highlighted in bold.

Table S4

Pairwise comparisons of the effects of ClauseOrder and Type on bilinguals' and monolinguals' comprehension of adverbial sentences in terms of pupil size changes (Experiment 2)

Effects		Contrast	*β*	*SE(β)*	*t*	*p*
bilingual	*before*	main-adv – adv-main	-1.61	13.30	-0.12	0.904
	after		8.57	13.30	0.64	0.521
monolingual	*before*		**-33.80**	**12.90**	**-2.61**	**0.009**
	after		**26.66**	**13.00**	**-2.06**	**0.040**
bilingual	main-adv	*before – after*	-3.13	13.40	-0.23	0.815
	adv-main		7.05	13.30	0.53	0.596
monolingual	main-adv		**-29.86**	**12.90**	**-2.32**	**0.020**
	adv-main		**30.61**	**13.10**	**2.34**	**0.019**
before	main-adv	bilingual – monolingual	**95.30**	**20.70**	**4.60**	**< 0.001**
	adv-main		**63.10**	**20.80**	**3.03**	**0.003**
after	main-adv		**68.60**	**20.80**	**3.30**	**0.001**
	adv-main		**86.70**	**20.80**	**4.17**	**< 0.001**

Note. Significant effects are highlighted in bold.

Table S5

Final model for bilinguals' and monolinguals' comprehension of adverbial sentences in terms of behavioural accuracy (Experiment 2)

	β	*SE(β)*	*z*	*p*
(Intercept)	3.17	0.18	17.48	< 0.001
LanguageGroup [bilingual]	**-0.88**	**0.23**	**-3.75**	**< 0.001**
ClauseOrder [main-adv]	**-0.31**	**0.13**	**-2.41**	**0.016**
Type [*before*]	-0.09	0.18	-0.48	0.633
ClauseOrder [main-adv]: Type [*before*]	**1.25**	**0.26**	**4.79**	**< 0.001**
LanguageGroup [bilingual]: ClauseOrder [main-adv]	-0.16	0.26	-0.62	0.537
LanguageGroup [bilingual]: Type [*before*]	0.22	0.31	0.70	0.486
LanguageGroup [bilingual]: ClauseOrder [main-adv]: Type [*before*]	**-2.14**	**0.52**	**-4.12**	**< 0.001**

Note. Reference (or omitted) level: LanguageGroup = monolingual; ClauseOrder = adv-main; Type = *after*. Number of observations = 4848. Significant effects are highlighted in bold.

Table S6

Pairwise comparisons of the effects of ClauseOrder and Type on bilinguals' and monolinguals' comprehension of adverbial sentences in terms of behavioural accuracy (Experiment 2)

Effects		Contrast	*β*	*SE(β)*	*z*	*p*
bilingual	*before*	main-adv – adv-main	-0.31	0.19	-1.59	0.113
	after		**-0.48**	**0.20**	**-2.41**	**0.016**
monolingual	*before*		**0.93**	**0.29**	**3.25**	**0.001**
	after		**-1.39**	**0.34**	**-4.15**	**< 0.001**
bilingual	main-adv	*before – after*	0.11	0.23	0.49	0.627
	adv-main		-0.06	0.25	-0.25	0.799
monolingual	main-adv		**0.97**	**0.33**	**2.95**	**0.003**
	adv-main		**-1.35**	**0.37**	**-3.64**	**< 0.001**
before	main-adv	bilingual – monolingual	**-1.39**	**0.35**	**-3.93**	**< 0.001**
	adv-main		-0.16	0.32	-0.49	0.622
after	main-adv		-0.53	0.28	-1.88	0.060
	adv-main		**-1.45**	**0.38**	**-3.77**	**< 0.001**

Note. Significant effects are highlighted in bold.