ARTICLE

A self-paced reading study of context effects in the processing of aspectual verbs in Mandarin

Ye Ma [ID], Brian Buccola [ID], Shannon Cousins and Alan Beretta

Department of Linguistics, Languages, and Cultures, Michigan State University, East Lansing, MI, USA
Corresponding author: Ye Ma; Email: maye2@msu.edu

Abstract

Research in the past few years has investigated the semantic complexity of expressions with aspectual verbs followed by entity-denoting complements such as *finish the book* that led to processing costs cross-linguistically. The Structured Individual Hypothesis (SIH) proposes that aspectual verbs lexically encode a function whose value (dimension) must be resolved. This ambiguity resolution is hypothesized to occur at the verb's complement, where a specific dimension is selected based on context (Piñango & Deo, 2016). In light of the critical role of the context in SIH, recent research (Lai et al., 2023; Lai & Piñango, 2019) has investigated how the interpretations of sentences with aspectual verbs were affected by biased contexts in an offline sentence acceptability judgment study and an online eye-tracking study. However, results of the two studies showed that biased contexts disambiguated the interpretations of aspectual verb expressions offline while processing costs in biased contexts were not found to attenuate costs in real time. The reason why conflicting results were found offline versus online and the timecourse of context effects remain unclear, but in our view it may be due to pragmatic contexts, i.e., descriptions of the utterance context used to infer the salient reading of the utterance. We used *grammatical* contexts – two classes of adverbs – in a self-paced reading study to examine context effects for sentences with aspectual verbs in Mandarin. We found that biased grammatical contexts not only affected the interpretations in the offline task, but crucially facilitated processing in the online experiment as well. We conclude that biased grammatical contexts predetermine the interpretations of aspectual verb expressions immediately in real time.

1. Introduction

Aspectual verbs represent relations between a reference time and event type of the complement in the verb phrase, which could introduce the initiation of an event (*start, begin*), the cessation of an event (*finish, complete, end*), or the continuation

of an event (*continue*) (ter Meulen, 1990). Aspectual verbs often take event-denoting complements, as in *John finished the speech*, where the verb *finished* introduces the end of a speech event denoted by the complement of the verb phrase (VP).

However, aspectual verbs do not exclusively select for event-denoting complements, as we see in *John finished the book*, where *book* is an entity, but an adequate paraphrase of this sentence is *John finished reading/writing the book*. To arrive at this interpretation, semantic composition must add a reading/writing event which does not overtly exist in the meanings of the individual words or in the syntax that combines them. The aspectual verb *finished* refers to the temporal end of the event *reading/writing the book*, of which *John* is the agent and *the book* is the theme. This interpretation is called the *agentive* reading of a sentence containing an aspectual verb and an entity-denoting complement.

Interestingly, the agentive reading is not the only interpretation that is available to native English speakers. Another acceptable interpretation of *John finished the book* is that *John is the last character in the book*. In this reading, *John* is not an agent of an event but a subpart of the content of the book; thus, there is no need to add an event like *reading/writing* in the interpretation. In the second interpretation, the informational content about the specifier of the VP is a small part of the informational content of the complement, so it has been called a *constitutive* reading (Lai et al., 2017; 2023; Lai & Piñango, 2019).

This phenomenon of more than one interpretation being available suggests that sentences containing aspectual verbs have underspecified representations in their composition under certain circumstances. To investigate why and how people can access multiple interpretations in these kinds of sentences, Piñango and Deo (2016) proposed the *Structured Individual Analysis* of aspectual verbs, which claims that aspectual verbs select for structured individuals as their complement rather than events. A structured individual refers to an entity, for example *a book*, that can be mapped onto a one-dimensional directed path structure, along various dimensions such as the *informational* dimension, so that the content of *each chapter* would be the adjacent subparts of the content of *the book* (Lai et al., 2017; Piñango & Deo, 2016). Aspectual verbs are lexically encoded for functions corresponding to these dimensions, such as $f_{spatial}$, $f_{temporal}$, or $f_{informational}$, which realize the mapping process mentioned above (Piñango & Deo, 2016). Furthermore, Piñango and Deo (2016) proposed the *Structured Individual Hypothesis* (SIH) which takes a perspective on the real-time processing of sentences with aspectual verbs.

1.1. Structured individual hypothesis (SIH)

Based on the findings of processing costs of sentences with aspectual verbs (DiNardo, 2015; Katsika et al., 2012), Piñango and Deo (2016) proposed two sources of the processing costs in sentences with aspectual verbs in SIH. The first source is the *exhaustive activation* of all dimension functions at the aspectual verb. Under such an assumption, no matter which specific dimension a complement and a subject might be mapped onto, all of the functions like $f_{spatial}$, $f_{temporal}$, $f_{informational}$ and so forth must be exhaustively retrieved at the verb. The second source of processing costs is the *ambiguity resolution* at the complement. Exhaustive activation of all dimensions at

the verb sets up an ambiguity that has to be resolved at the complement, where a specific dimension has to be selected based on the context. The processing costs of ambiguity resolution may be sustained to a post-complement region, given that previous studies of complement coercion have reported subsequent effects at words following the complement (Traxler et al., 2002; Traxler et al., 2005). The hypothesis was supported by the experimental results of a self-paced reading study and an fMRI study in English which indicated aspectual-verb sentences were harder to be processed comparing to psychological-verb sentences (Lai et al., 2017). Lai et al. (2017) found significant longer reading times for the aspectual verb condition compared with the psych verb condition and the control condition at the two Regions of Interest (ROIs) following the complement in the self-paced reading study, as well as significant activations for the aspectual verb condition over the psych verb condition and the control condition in the fMRI study. In addition, cross-linguistic evidence supporting the SIH was found in an eye-tracking study on verb class in Mandarin (Ma et al., 2022). Processing costs were found in sentences with aspectual verbs compared with sentences with psychological verbs and control verbs at the verb, the complement, and a post-complement region comprising four Chinese characters following the complement.

If the SIH correctly predicts the processing of sentences with aspectual verbs, one crucial aspect of the analysis is the role of context. In the Structured Individual Analysis, aspectual verbs have a set of lexically encoded functions but only in context could an appropriate function be determined that can be applied to the denotation of the complement (Piñango & Deo, 2016). In light of this crucial role of context, the same research group investigated how context affects the processing of sentences with aspectual verbs. They carried out an offline acceptability rating task (Lai & Piñango, 2019) and an online eye-tracking experiment (Lai et al., 2023) to investigate how biased and non-biased contexts affect the acceptability of sentences with aspectual verbs. The prediction was that if context mattered, expressions with aspectual verbs in a context biased to a single interpretation would be easier to process than in a non-biased context with multiple interpretations. For example, Lai and Piñango (2019) created three types of context sentences prior to the target sentences with aspectual verbs, namely, the neutral context which did not bias the readings of the target sentence, the agentive-biasing context, and the constitutive-biasing context. Context effects were found in the offline acceptability rating study as predicted. For sentences with aspectual verbs in the agentive-biasing context, participants were found to choose the agentive reading more often; in the constitutive-biasing context, participants were found to choose the constitutive reading more often; while in the neutral context, both readings were chosen with equal percentage (Lai & Piñango, 2019, p. 598). However, no significant effects were found in the online eye-tracking study, in which sentences with aspectual verbs in a biased context were found to be as costly as in a non-biased context (Lai et al., 2023).

It is not clear why a biasing context facilitated the offline judgments of sentences with aspectual verbs but did not facilitate the online processing of these sentences. Given that the offline and online studies yielded conflicting results, Lai et al. (2023) claimed that the null results in the eye-tracking study suggested that the exhaustive activation of multiple dimensions should occur regardless of context. Context only *privileged* rather than *predetermined* the lexically encoded dimension functions.

1.2. Motivation for the present study

The SIH studies on context mentioned above raise a couple of related questions. First, the mismatch between the results of the online study and the offline study has to be considered. The offline results indicated that a biased context facilitated comprehension of sentences with aspectual verbs, whereas no such context effects facilitated processing in the online study. It is unclear to what extent offline findings reflect online processing, partly because offline results could possibly be affected by a different set of cognitive processes.

To address the problem of why context failed to affect their online results, as alluded to above, Lai et al. (2023) argued that context merely privileged, but did not predetermine a particular dimension. The range of dimension functions, they concluded, is underspecified, and remains so unless participants are forced to make a choice, as they had to in the offline study.

This argument raises questions about what counts as context. The stimuli in Lai and Piñango (2019) and Lai et al. (2023) were composed of two parts: (i) context sentences, and (ii) target sentences. For example, from Lai et al. (2023), a constitutive-biasing context is provided by the following sentence: *Kevin owns numerous CDs by different jazz musicians.* And the target sentence is: *Dave Brubeck starts this CD of classic Jazz hits.* The context sentence to bias the constitutive reading is one in which 'the individual with salient subpart structure… is mentioned'.

In this example, a comprehender is expected to infer from the target sentence that the salient subpart is part of a compilation, not part of 'numerous CDs'. But this is not the only inference that is possible. Perhaps a more likely inference from the context sentence is that Kevin owns maybe a CD by Miles Davis, a CD by Art Tatum, a CD by Dave Brubeck, and so forth. If that is what a comprehender infers, then the salient subpart is not mentioned in the context, and arriving at an interpretation of the target sentence is potentially vexed. Given the good performance in offline measures, it appears that it can eventually be figured out that Dave Brubeck must constitute a part of a compilation album, but the non-binding nature of the relation between the context sentence and the target sentence could well account for the failure to facilitate processing.

Given that inconclusive, or 'insufficiently biasing' contexts failed to elicit an online effect in the studies mentioned above, the present study sought to create grammatical contexts that unambiguously bias agentive readings and contrasted them with grammatical contexts that unambiguously allow both agentive and constitutive readings. That is the main goal of the present study.

Based on Ma et al. (2022), in which effects related to aspectual verbs were found in Mandarin Chinese and provided partial support for the SIH, the present study investigated whether context played a role in processing sentences with aspectual verbs in Mandarin. The grammatical contexts were created by using predicational adverbs, since adverbs with different properties can either reduce the dimension space to a single dimension or not, and therefore could either facilitate processing or not.

1.3. Adverbs

Predicational adverbs in English are usually composed of an adjective and an *-ly* suffix, taking events or propositions as arguments (Ernst, 2001). Ernst (2001) analyzed two classes of adverbs, subject-oriented adverbs and speaker-oriented

adverbs. Subject-oriented adverbs include two subclasses, agent-oriented adverbs which take an agent of the event as the subject and mental-attitude adverbs which take an experiencer of the event as the subject. Speaker-oriented adverbs firstly appeared in Jackendoff (1972), referring to adverbs that express the emotion or evaluation of the proposition asserted by the speaker. Ernst (2001) furthermore investigated such adverbs and found that speaker-oriented adverbs did not make reference to the subject of the verb, and they took proposition, fact, or speech-act as argument. For example, a sentence like *Unbelievably, she decided to buy a camel* could be paraphrased as *The fact that she decided to buy a camel is unbelievable* (Ernst, 2001, p. 69).

According to Strand et al. (2014), grammatical class (e.g., verb, noun, etc.) as grammatical context can have strong influences on language-processing domains. So, intuitively, adverbs were used as grammatical context, providing the contextual information within the target sentence, which distinguishes from the extra-sentential context, i.e., the pragmatic context used in previous studies (Lai et al., 2023; Lai & Piñango, 2019) to infer the salient reading of the following target sentence.

In the present study, mental-attitude adverbs and speaker-oriented adverbs, as shown in (1), were used to construct contrasting contexts for sentences with aspectual verbs.

(1)   mental-attitude adverbs: calmly,   anxiously,   reluctantly,
      absent-mindedly etc.
      speaker-oriented adverbs: surprisingly, appropriately, fortunately,
      absurdly etc.
      (Ernst, 2001)

Our rationale for using such adverbs was as follows. Taking a sentence with an aspectual verb like (2) as an example, the most salient semantic representation of this sentence is the agentive reading: *John began reading/writing the book.* Another possible representation is a constitutive reading: *John is the first character in the book.*

(2)   John began the book.

Interestingly, if the sentence in (2) is combined with different types of adverb, the number of readings available to a native speaker is quite distinct. Consider the sentences in (3):

(3)   a. John surprisingly began the book.
      b. John reluctantly began the book.

In (3a), we add a speaker-oriented adverb to (2) and we can still get more than one reading, such as *It was surprising that John began reading/writing the book* (agentive reading) or *It was surprising that John is the first character in the book* (constitutive reading). But in (3b), when we add a mental-attitude adverb, the ambiguity disappears, because we can only get the agentive reading, *John reluctantly began reading/ writing the book*, while the constitutive reading is unavailable. It seems that mental-attitude adverbs limit the number of semantic representations denoted by aspectual verbs to one.

Table 1. Conditions: adverb type * verb type

		Adverb Type		
		no adverb	speaker-oriented adverb	mental-attitude adverb
Verb Type	**aspectual verb**	John began the book.	John surprisingly began the book.	John reluctantly began the book.
	control verb	John read the book.	John surprisingly read the book.	John reluctantly read the book.

Therefore, in the present study, two types of adverbs were used to provide context in order to investigate whether context affects real-time processing of sentences with aspectual verbs. The distinction observed in (3) illustrates that sentences with mental-attitude adverbs, such as (3b), generate only one reading, whereas sentences with speaker-oriented adverbs, such as (3a), and sentences without adverbs such as (2), do not. Based on this distinction, the prediction is that sentences with mental-attitude adverbs, like (3b), will incur fewer processing costs than sentences with speaker-oriented adverbs, like (3a), and sentences without adverbs like (2).

1.4. Research design

The present study aims to investigate the effects of context in sentences with aspectual verbs in Mandarin Chinese. The design of the study is factorial 2 × 3, with the factors of verb type and adverb type, as shown in Table 1. In this study, we compared sentences with aspectual verbs and sentences with neutral verbs. In addition, we compared sentences without adverbs, sentences with speaker-oriented adverbs, and sentences with mental-attitude adverbs. The hypothesis is that the processing of sentences with mental-attitude adverbs is less costly because only the agentive reading is available (and constitutive readings are not), and so in this condition there would be no ambiguity to resolve. By contrast, the processing of sentences with speaker-oriented adverbs and of those with no adverb could be more costly than sentences with mental-attitude adverbs because both agentive and constitutive readings are available, and so there would be an ambiguity that has to be resolved. By comparing mental-attitude adverb conditions with speaker-oriented conditions, we investigated the sentence processing in a biasing context versus a non-biasing context; and by comparing mental-attitude adverb conditions with no adverb conditions, we investigated the sentence processing given a biasing context versus no context. The incorporation of both speaker-oriented adverb and mental-attitude adverb in the experimental design could allow us to accurately test the effects caused by different contexts.

1.5. A puzzle in Mandarin Chinese

A subtle difference between English and Mandarin Chinese was discovered that not all of the aspectual verbs used in English studies had corresponding aspectual verbs in Mandarin. That is, not all of them allow an entity-denoting complement. An example is *kaishi* (*begin*) in (4).

(4) a. * Yuehan kaishi zhe ben   xiaoshuo
       John begin   the CLASSIFIER novel
       'John begins the novel'.
    b. Yuehan kaishi xie   zhe ben   xiaoshuo
       John begin   write the CLASSIFIER novel
       'John begins writing the novel'.

In (4a), *kaishi* ('begin') is followed by an entity *zhe ben xiaoshuo* ('the novel') and it gives rise to an ungrammatical sentence. Similar observations were reported in the literature that some aspectual verb does not always allow entity-complements in simple declarative sentences in Mandarin (Lin & Liu, 2005; Ma et al., 2022) To render the sentence acceptable, the event needs to be fully explicit, as in (4b). However, it happens that if the sentence structure is changed to (5), *kaishi* ('begin') does allow an entity-denoting complement. With a sentence structure *SHI…LAI…, kaishi zhe ben xiaoshuo* ('begin the novel') is grammatical.

(5) SHI Yuehan LAI   kaishi zhe ben   xiaoshuo
    SHI John   PARTICLE begin the CLASSIFIER novel
    'It is john who begins the novel'.

Thus, the first observation is that a sentence structure with *SHI…LAI…* increases the acceptability of a verb phrase that consists of an aspectual verb and an entity-denoting complement in Mandarin Chinese.

    A second observation is that with the sentence structure *SHI…LAI…,* constitutive readings become more salient than constitutive readings in non-*SHI…LAI* structures in Mandarin. For example, in (6), a constitutive reading could be *John is the last character in the novel*, which is much easier to access in (6b) than (6a).

(6) a. Yuehan jieshu zhe ben   xiaoshuo
       John   finish the CLASSIFIER novel
       'John finishes the novel'
    b. SHI Yuehan LAI   jieshu zhe ben   xiaoshuo
       SHI John   PARTICLE finish the CLASSIFIER novel
       'It is John who finishes the novel'.

Shi in Mandarin Chinese was classified as a verb or a focus marker, and the sentence structure *SHI…* was named *the bare shi construction* (Liu & Kempson, 2018). This construction was regarded similar to the cleft structure in English and the function of *shi* was claimed to put an emphasis on the focused part following it (Tang, 1983; Xu, 2003). For *lai* in Mandarin, it was usually translated as *come* in English. But in this SHI… LAI… structure, it has nothing to do with the meaning of *come*. Other than being a verb, *lai* was also defined as a focus marker when it was followed by a VP (Li & Li, 2014; Lu, 2006). Lu (2006) claimed that '*Lai does not show the time course of the action. It does not indicate the result of the action, either. The focus of the sentence generally falls on the method, means, and way to achieve the VP*'. However, it is not clear to us yet why the *SHI…LAI…*structure increases the acceptability of sentences with aspectual verb expressions and makes it much easier for native Mandarin speakers to access constitutive readings. This is a puzzle for future inquiry. But after

failing to identify any reason to be concerned that the structure involves some property that would militate against using it in the present study, we opted for this sentence structure across all conditions. By using the *SHI…LAI…*structure, we are able to test sentences with *kaishi* ('begin'), as well as *jixu* ('continue'), *wancheng* ('finish'), and *jieshu* ('finish').

2. Methods

2.1. Participants

A total of 160 participants (ages 23–49, 98 females) were recruited for the pre-tests, and 65 for the self-paced reading experiment (ages 25–54, 22 females). All of the participants recruited for this study fulfilled the following requirements: (1) right-handed, and (2) native Mandarin speaker.

2.2. Materials

A total of 59 sets of stimuli were initially created. Thirty sets of stimuli remained after 7 pre-tests and these were used in the experiments. Each set contained six sentences, one for each of six conditions. Sample stimuli are shown in Table 2. All the sentences were randomly distributed to six different lists to make sure that participants only read one sentence in each set. The ratio of experimental and filler sentences was 1:2 in the self-paced reading experiment. Each list contained 90 sentences, consisting of 30 experimental sentences and 60 fillers. Half of the fillers were plausible and half implausible. Comprehension questions were simple Yes/No questions. Each sentence was followed by a comprehension question.

2.3. Pre-tests

A total of 7 pre-tests were carried out, in order to collect matrix verbs for the neutral conditions, to control for the potential influences of frequency, word length of both verbs and adverbs, the predictability of the upcoming noun phrase (NP) complement, the plausibility of stimuli sentences, and the acceptability of different readings across conditions. Data of pre-tests were analyzed in R (R Core Team, 2021).

2.3.1. Fill-in-the-blank test

The first pre-test aimed at collecting target verbs in control conditions. Ten participants were recruited in total to take part in an online survey. Sixty sentences with aspectual verbs and without adverbs, such as *SHI John LAI finish the novel with a perfect ending*, were created prior to the pre-test. These sentences were randomized and participants were asked to fill in the blank with a suitable verb, as in *SHI John LAI finish _____ the novel with a perfect ending*.

Following the procedure that was used in three studies in the field carried out by McElree et al. (2001), Baggio et al. (2010), and Ma et al. (2022), synonyms and near-synonyms among the responses of each sentence were first conflated (Baggio et al., 2010; Ma et al., 2022). And then, 59 verbs were selected because the dominant response occurred more than twice as often as the next most frequent response

Table 2. Sample stimuli in Mandarin

Conditions	Examples
no adverb- aspectual (NoAdv-Asp)	SHI Yuehan LAI kaishi zhe bu wenxuezuopin keshi qingjie bingbu jingcai. SHI John PARTICLE begin the CLASSIFIER literature but story not interesting 'It is John who begins the literature but the story is not interesting'.
no adverb- control (NoAdv control)	SHI Yuehan LAI zhuanxie zhe bu wenxuezuopin keshi qingjie bingbu jingcai. SHI John PARTICLE write the CLASSIFIER literature but story not interesting 'It is John who writes the literature but the story is not interesting'.
mental-attitude adverb- aspectual (MA-Asp)	SHI Yuehan buqingbuyuande LAI kaishi zhe bu wenxuezuopin keshi qingjie bingbu jingcai. SHI John reluctantly PARTICLE begin the CLASSIFIER literature but story not interesting 'It is John who reluctantly begins the literature but the story is not interesting'.
mental-attitude adverb- control (MA control)	SHI Yuehan buqingbuyuande LAI zhuanxie zhe bu wenxuezuopin keshi qingjie bingbu jingcai. SHI John reluctantly PARTICLE write the CLASSIFIER literature but story not interesting 'It is John who reluctantly writes the literature but the story is not interesting'.
speaker-oriented adverb- aspectual (SO-Asp)	SHI Yuehan lingrenjingyade LAI kaishi zhe bu wenxuezuopin keshi qingjie bingbu jingcai. SHI John surprisingly PARTICLE begin the CLASSIFIER literature but story not interesting 'It is John who surprisingly begins the literature but the story is not interesting'.
speaker-oriented adverb- control (SO control)	SHI Yuehan lingrenjingyade LAI zhuanxie zhe bu wenxuezuopin keshi qingjie bingbu jingcai. SHI John surprisingly PARTICLE write the CLASSIFIER literature but story not interesting 'It is John who surprisingly writes the literature but the story is not interesting'.

(McElree et al., 2001; Traxler et al., 2002). The selected verbs occurred on average 8.5 times (out of 10), ranging from 6 to 10. As a result, 59 sentences with aspectual verbs without adverbs were retained, and their corresponding control sentences were created by substituting the aspectual verbs with the neutral verbs collected in this pre-test. Based on the 59 pairs of no-adverb sentences, a pair of speaker-oriented adverb conditions and a pair of mental-attitude adverb conditions were also created.

2.3.2. Verb frequency and verb length test

Word frequency and word length of aspectual and control verbs were examined to rule out potential alternative explanations for any observed results. Verb frequencies (raw token counts) were collected from BLCU (Beijing Language and Culture University) Corpus Center, which contained 15 billion Chinese characters. Mean verb frequencies were 369615.2 (standard deviation 'SD' = 198254.6) for aspectual verbs, and 407628.5 (SD = 539226.8) for control verbs. Mean word lengths were 2.0 (SD = 0) for aspectual verbs, and 1.97 (SD = 0.18) for control verbs. Since neither verb frequency data nor verb length data were normally distributed (examined by a Shapiro–Wilk test with $p < .001$), a logarithmic transformation (X' = log(X)) was performed on the data. A two-sample t-test revealed no significant difference in verb frequency between conditions ($t_{(116)} = 1.10$, $p = .28$). The mean verb length between conditions was not significantly different, either ($t_{(116)} = 1.43$, $p = .16$).

2.3.3. Adverb frequency and adverb length test

Based on the 59 pairs of sentences arrived at after the fill-in-the-blank pre-test, 4 adverb conditions were created as mentioned above. A pre-test on word frequency and a pre-test on word length of adverbs were carried out to make sure that these were not different between conditions. Adverb frequencies (raw token counts) were collected from BLCU Corpus. The mean adverb frequencies were 1189.6 (SD = 2055.1) for mental-attitude adverbs, and 3331.7 (SD = 12357.0) for speaker-oriented adverbs. The mean word lengths of adverbs were 3.90 (SD = 0.99) for mental-attitude adverbs, and 3.69 (SD = 0.99) for speaker-oriented adverbs. Neither adverb frequency data nor word length data were normally distributed (examined by a Shapiro–Wilk test with $p < .001$), so a logarithmic transformation (X' = log(X)) was performed on the data. A two-sample t-test revealed no significant difference in adverb frequency between conditions ($t_{(116)} = .22$, $p = .83$). The mean adverb length between conditions was not significantly different ($t_{(116)} = 1.18$, $p = .24$).

2.3.4. Plausibility test

A plausibility judgment test was carried out to control for possible differences in sentence plausibility across conditions. A total of 59 sets of stimuli across 6 conditions were divided into 6 lists. A total of 60 participants were recruited, and 10 participants were assigned to each list. Each participant was required to rate sentences on a Likert scale from 1 (makes no sense) to 5 (makes perfect sense). There was a significant difference between conditions in the plausibility ratings ($p < .001$). As a result, sentence sets with low plausibility (rating < 2) were deleted, and 30 sets of sentences were retained. The mean plausibility ratings of the 30 sets of sentences in the six conditions were 4.2 (SD = 0.94) for the no-adverb aspectual condition, 4.2 (SD = 0.98) for the no-adverb control conditions, 4.1 (SD = 1.0) for the mental-attitude adverb

aspectual condition, 3.0 (SD = 1.0) for the mental-attitude adverb control condition, 3.9 (SD = 1.1) for the speaker-oriented adverb aspectual condition, and 4.1 (SD = 1.0) for the speaker-oriented adverb control condition. A repeated-measures ANOVA with condition as within-subjects factor did not show significant difference ($F_{(5,\ 1128)}$ = 1.57, p = .17). To further confirm that this non-significance was not due to lack of statistical power, we performed a power analysis. We found that with the *Effect Size (Cohen's F)* = .25 and *alpha* = .05, the calculated power for the pre-test was 0.999*, so there is over 99.9% chance of correctly rejecting the null hypothesis of no difference across the conditions with a total of 60 participants, suggesting that the statistical analysis had enough power.

2.3.5. *Cloze probability test*

To control for possible differences in the predictability of the NP complement across conditions, a cloze probability test was carried out. Stimuli were divided into 6 lists. Sixty different participants were recruited. They were asked to fill in the blank with a classifier and an NP, as in *SHI John LAI finish ___(for classifier) ___(for NP) with a perfect ending*. The reason why the classifier after the determiner was also kept blank was that, in Mandarin, NPs require particular classifiers, so the presence of a classifier will provide a clue as to what the upcoming word might be (Ma et al., 2022). The mean cloze probabilities were 0.53 (NoAdv-Asp), 0.65 (NoAdv control), 0.54 (MA-Asp), 0.53 (MA control), 0.51 (SO-Asp), and 0.65 (SO control). A one-way ANOVA on the cloze probabilities across conditions revealed no difference ($F_{(5,\ 174)}$ = 1.107, p = .36).

3. Experiment 1: sentence interpretation task

A sentence interpretation task was carried out to investigate the acceptability of agentive and constitutive readings of sentences with aspectual verbs in various contexts. This offline experiment aimed at investigating whether the different adverb conditions changed the accessibility of readings generated by aspectual verbs.

A forced-choice paradigm was used in which three conditions were examined: (i) a mental-attitude adverb aspectual condition (MA-Asp), (ii) a speaker-oriented adverb aspectual condition (SO-Asp), and (iii) a no-adverb aspectual condition (NoAdv-Asp). The hypothesis was that when participants read sentences in the MA-Asp condition, the choice of 'constitutive' readings or 'both' readings should occur less frequently in the MA-Asp condition compared with the other two conditions. And the choice of the 'agentive' reading in the MA-Asp condition should occur more frequently than in the other two conditions.

Stimuli were randomized into 3 lists so that none of the participants saw sentences from the same set. Each list contained 30 sentences. A total of 30 participants (ages 25–33, 14 females) were recruited, all native Mandarin speakers, right-handed. For each sentence, participants were asked to perform a forced-choice task asking what reading(s) they could accept from the original sentences as *Shi John Lai begins the literature but the story is not interesting*. Choices included an 'agentive' reading like *John begins writing the literature but the story is not interesting*; a 'constitutive' reading like *John is the first character mentioned in the literature but the story is not interesting*, and 'both' readings like *Both paraphrases mentioned above are correct*.

3.1. Data analysis

Participants' choices were encoded by *0* and *1* for data analysis, *0* being not choosing a reading and *1* being choosing a reading. For example, in a case that a participant chose 'both' reading, *1* was assigned to the response of 'both' reading, while *0* was assigned to the response of 'constitutive' reading and 'agentive' reading. Since each response was encoded as a binary dataset, Logistic Mixed-Effects Models were conducted in R (R Core Team, 2021), with glmer() function in *lme4* package (Bates et al., 2015). Given the complex random-effect structure, selection was performed among a series of models increasing the complexity of random effects step by step so as to determine the best-fitting model (Baayen et al., 2008). The goodness-of-fit of the models was compared using the Akaike Information Criterion (AIC) and likelihood ratio test. Model selection was conducted for all three responses. Results showed that Model 3 with condition as fixed effect, and by-subject and by-item random intercepts, was the best-fitting model for all three responses. Model selection results for choice of 'both' reading were reported in Table 3 as an example. The syntax of the model was specified below:

Response   ~ condition   + (1|pariticipant)   + (1|item)

To determine whether condition had a significant effect on the model, likelihood ratio tests were used to compare models with and without the fixed effect at *alpha* = 0.05. Pairwise comparisons were conducted with Bonferroni adjustment using *lsmeans* package (Lenth, 2016).

3.2. Results of the sentence interpretation task

Results of Experiment 1 were shown in Figure 1 in which the y-axis showed the average per person counts of times each reading being chosen for each condition. For the choice of 'both' readings, a main effect of condition was found ($\chi^2_{(2)}$ = 19.41, $p <$ .0001). A pairwise comparison indicated that the choice of both readings for the NoAdv-Asp condition occurred more frequently than for the MA-Asp condition ($z =$ 4.23, $p = $ .0001). In addition, the choice of 'both' readings for the SO-Asp condition occurred more frequently than for the MA-Asp condition ($z = 3.51$, $p = $ .0014). There was no significant difference between the NoAdv-Asp condition and the SO-Asp condition ($z = 0.80$, $p > $ .99).

For the choice of 'constitutive' reading, a main effect of condition was found ($\chi^2_{(2)}$ = 31.92, $p < $ .0001). A pairwise comparison indicated that the choice of 'constitutive' reading for the NoAdv-Asp condition occurred more frequently than for the MA-Asp condition ($z = 5.02$, $p < $ .0001). In addition, the choice of 'constitutive' reading for the SO-Asp condition occurred more frequently than for the MA-Asp condition ($z = 5.11$, $p < $ .0001). No significant difference between the NoAdv-Asp condition and the SO-Asp condition was found ($z = -0.05$, $p > $ .99).

For the choice of 'agentive' reading, a main effect of condition was found ($\chi^2_{(2)}$ = 50.78, $p < $ .0001). A pairwise comparison indicated that the choice of 'agentive' reading for the NoAdv-Asp condition occurred less frequently than for the MA-Asp condition ($z = -6.17$, $p < $ .0001). In addition, the choice of 'agentive' reading for the SO-Asp condition occurred less frequently than for the MA-Asp condition ($z = $

Table 3. Model selection results suggested that Model 3 with by-subject and by-item random intercepts was the best-fitting model. Four models with condition as the fixed effect and different random-effects structure were compared for data of each choice. To determine the best model, the Akaike Information Criterion (AIC) value and likelihood ratio tests were used to estimate the goodness-of-fit. An example of model selection results for choice of 'both' readings is presented below

| Model	Outcome variable	Fixed effect	Random effects					d.f.	AIC	logLik	χ^2	d.f.	P
	Response	Condition	By-subject random intercept	By-item random intercept	By-list random intercept	By-subject random slope for condition	By-subject random intercept–slope correlation for condition						
Model 1	X	X	X	X	X	X	X	12	991.68	−483.84	8.96	6	0.18
Model 2	X	X	X	X	X			6	988.64	−488.32	7.75	1	1.00
Model 3	X	X	X	X				5	986.64	−488.32	6.78	1	0.0092**
Model 4	X	X	X					4	991.42	−491.71			
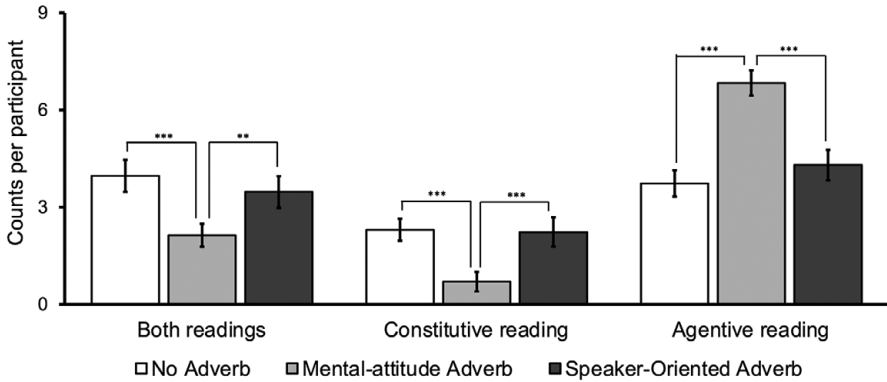
Figure 1. Results of the sentence interpretation task (**$p < .01$, ***$p < .001$). Error bars indicate standard error.

-6.24, $p < .0001$). No significant difference between the NoAdv-Asp condition and the SO-Asp condition was found ($z = -0.19$, $p > .99$).

4. Experiment 2: self-paced reading experiment

A moving window self-paced reading experiment was conducted to investigate how adverbs as context affected the real-time processing of aspectual verbs in Mandarin Chinese. The experiment was programmed in PsychoPy 3 (Peirce, 2019) and carried out on the Pavlovia platform. A total of 65 participants were recruited, all native Mandarin speakers, right-handed. Six lists of stimuli were created with 30 experimental sentences and 60 fillers (30 plausible and 30 implausible) in each list. We used a Latin square design to divide stimuli into the 6 lists, such that the 6 sentences in every set were distributed into different lists and each list contained all 6 conditions. All stimuli were randomized.

During the experiment, each sentence started with a series of dashes to hide the whole sentence. When participants pressed the space bar on the keyboard, the first set of dashes were replaced by the first Chinese character. By pressing the space bar again, the second chunk of characters appeared and the first character was hidden by dashes again. By pressing the space bar sequentially, participants read sentences chunk by chunk till the end of the sentence. A Yes/ No comprehension question followed every sentence. Participants pressed *A* or *L* on the keyboard to answer the comprehension questions. Data of 5 participants were not included in the analysis due to low accuracy of the comprehension questions (lower than 80%).

Each sentence was divided into chunks, as shown in Table 4. Some chunks contained 1 character while others contained multiple characters. Reading time (RT) was recorded for each chunk.

4.1. Data analysis

For the no-adverb conditions, 5 regions of interest were defined, and for the adverb conditions, 6 regions of interest were defined (Table 4). The verb region included the matrix verb. The NP region included the NP complement. All NP complements in all 6 conditions consisted of the same number of characters. The Post 1 region included the 1st and 2nd characters following the NP complement. The Post 2 region included

Table 4. Experimental paradigm and regions of interest (ROIs)

Experimental paradigm				RT3	RT4	RT5		RT6	RT7	RT8	
ROIs of no-adverb conditions	dashes	RT1	RT2		Verb	NP		Post1	Post2	Post3	RT9
no adverb- aspectual (NoAdv-Asp)	------	SHI	Yuehan	LAI	kaishi	zhe bu wenxuezuopin		keshi	qingjie	bingbu	jingcai.
		SHI	John	PARTICLE	begin	the CLASSIFIER literature		but	story	not	interesting
'It is John who begins the literature but the story is not interesting'.											
no adverb- control (NoAdv control)	------	SHI	Yuehan	LAI	zhuanxi	zhe bu wenxuezuopin		keshi	qingjie	bingbu	jingcai.
		SHI	John	PARTICLE	write	the CLASSIFIER literature		but	story	not	interesting
'It is John who writes the literature but the story is not interesting'.											

Experimental paradigm				RT3		RT5	RT6		RT7	RT8	RT9	
ROIs of adverb conditions	dashes	RT1	RT2	Adverb	RT4	Verb	NP		Post1	Post2	Post3	RT10
mental-attitude adverb- aspectual (MA-Asp)	------	SHI	Yuehan	buqingbuyuande	LAI	kaishi	zhe bu wenxuezuopin		keshi	qingjie	bingbu	jingcai.
		SHI	John	reluctantly	PARTICLE	begin	the CLASSIFIER literature		but	story	not	interesting
'It is John who reluctantly begins the literature but the story is not interesting'.												
mental-attitude adverb- control (MA control)	------	SHI	Yuehan	buqingbuyuande	LAI	zhuanxi	zhe bu wenxuezuopin		keshi	qingjie	bingbu	jingcai.
		SHI	John	reluctantly	PARTICLE	write	the CLASSIFIER literature		but	story	not	interesting
'It is John who reluctantly writes the literature but the story is not interesting'.												
speaker-oriented adverb- aspectual (SO-Asp)	------	SHI	Yuehan	lingrenjingyade	LAI	kaishi	zhe bu wenxuezuopin		keshi	qingjie	bingbu	jingcai.
		SHI	John	surprisingly	PARTICLE	begin	the CLASSIFIER literature		but	story	not	interesting
'It is John who surprisingly begins the literature but the story is not interesting'.												
speaker-oriented adverb- control (SO control)	------	SHI	Yuehan	lingrenjingyade	LAI	zhuanxi	zhe bu wenxuezuopin		keshi	qingjie	bingbu	jingcai.
		SHI	John	surprisingly	PARTICLE	write	the CLASSIFIER literature		but	story	not	interesting
'It is John who surprisingly writes the literature but the story is not interesting'.												

the 3rd and 4th characters following the NP. The Post 3 region included the 5th and 6th characters following the NP. The adverb region included the adverb. Data from all 6 conditions were analyzed together for the Verb, NP, Post 1, Post 2, and Post 3 regions. Regarding the adverb region, data from 4 adverb conditions were analyzed.

A logarithmic transformation ($X' = \log(X)$) was carried out due to the non-normal distribution (examined by a Shapiro–Wilk test with $p < .001$) in all regions of interest. Linear Mixed-Effects Models were used to analyze these data by regions in R (R Core Team, 2021), with the lmer() function in the *lme4* package (Bates et al., 2015). Given the complex random-effect structure, a backward model selection was performed among a series of models decreasing the complexity of random effects step by step so as to determine the best-fitting model (Baayen et al., 2008). All models have verb type, adverb type, and interaction between verb type and adverb type as fixed effects, and by-subject and by-item random intercepts. The maximal complex model also has by-list random intercept, the random slope, and the interaction of random intercepts and slopes. However, it failed to converge in all regions and singularity warnings were reported, suggesting high risks of overfitting. We then removed the random intercept–slope correlation but same issues occurred. Furthermore, we tried to use allFit() function in *afex* package (Singmann et al., 2023) to search for appropriate optimizers and used brm() function in *brms* package (Bürkner, 2017) to solve the problem. Unfortunately, these methods did not work. As discussed in Godfroid (2019, p. 281) that 'It is worth emphasizing that backward model selection, as promoted by Matuschek et al. (2017), does not offer a one- size-fits-all solution. Some models will have a maximal structure, others will have by-subject and by-item random intercepts only, and others still will end up with something in between. Psychologists nowadays, though, seem to agree that having just a by-subject random intercept is never a good idea (high Type I error rates), even if these models are still quite common in psychology and in L2 research. If random effects are added, the model should minimally have a by-item random intercept as well. The point is to engage in model selection to obtain a parsimonious, yet well-fitting model that is true to the data'. So, we decided to use the model with by-subject and by-item random intercepts which converged successfully and prevented overfitting:

reading time (RT)   ~ verb type   * adverb type   + (1|Subject)   + (1|Item)

The same model was applied to all regions.

For data analyzed in Linear Mixed-Effects Models, a Type III analysis of variance was carried out for the fixed effects, using Satterthwaite's method. If a main effect of verb type and/or a main effect of adverb type was found, a pairwise comparison was conducted using *lsmeans* package with Bonferroni adjustment (Lenth, 2016). If a significant interaction between verb type and adverb type was found, custom contrasts were conducted with the Bonferroni method.

Generally speaking, what we are most interested in are the main effect of verb type and the interaction between verb type and adverb type. Regarding the main effect of verb type, we would like to know whether there is a significant difference between sentences with aspectual verbs and sentences with control verbs, so as to see whether the processing costs of aspectual verbs found in Mandarin in the previous literature (Ma et al., 2022) are replicated. Only when the verb type effect was found, can we be licensed to further investigate the context effects in expressions with aspectual verbs.

When the interaction between verb type and adverb type was found, 9 custom contrasts were conducted. The first 3 contrasts as in (7) look at verb effects in three adverb conditions.

(7)   Contrasts for verb effect:
      a. 'NoAdv-Asp *versus* NoAdv control'
      b. 'MA-Asp *versus* MA control'
      c. 'SO-Asp *versus* SO control'

The first contrast looks at aspectual versus control verb effects when there was no adverb at all in the sentence. The second contrast looks at the verb effect when there is a speaker-oriented adverb in the sentence. The third contrast looks at the verb effect when there is a mental-attitude adverb in the sentence.

Regarding interactions between verb type and adverb type, this is the key to answering the research question of whether adverbs as context can affect the processing of aspectual verbs. To be more specific, the goal is to know if a mental-attitude adverb as context can facilitate the processing of aspectual verbs, because sentences in the mental-attitude adverb aspectual condition only have an agentive reading. The other type of adverb, the speaker-oriented adverb, allows both agentive and constitutive readings of aspectual verb sentences; thus, it should be more difficult to process, as should sentences in the no-adverb aspectual condition. In this case, another three contrasts as in (8), were examined.

(8)   Contrasts for Context effect
      d. 'MA-Asp *versus* NoAdv-Asp'
      e. 'MA-Asp *versus* SO-Asp'
      f. 'SO-Asp *versus* NoAdv-Asp'

If the MA-Asp as context (permitting only one reading) facilitates the processing of aspectual verbs, compared with the SO-Asp condition (permitting two readings), there should be a significant difference in contrast 'e'. Since the NoAdv-Asp condition permits more than one reading, there should be a significant difference in contrast 'd'. Contrast 'f' is predicted to show no difference.

Moreover, three control conditions were compared as in (9), which could help examine the quality of control conditions and better interpret the results.

(9)   Contrast of control conditions
      g. 'MA control *versus* NoAdv control'
      h. 'MA control *versus* SO control'
      i. 'SO control *versus* NoAdv control'

The accuracy of comprehension questions was inspected. The accuracy of 5 participants (out of 65) was lower than 80%, suggesting that those 5 participants were not paying attention. So, in total, the data of 60 participants were analyzed.

4.2. Results of the self-paced reading experiment

Results are presented in Table 5 with means and standard deviations (SD) by region of interest (ROI).

Table 5. Reading Time (ms) in each ROI. C, MA-A, and NoAdv-C stand for control conditions, mental-attitude adverb aspectual condition and no adverb-control condition

Verb Region			
	No adverb	Speaker-oriented adverb	Mental-attitude adverb
Aspectual	433.53(222.49)	434.84(225.58)	444.40(191.17)
Control	422.27(182.76)	428.64(205.53)	431.41(222.92)
		NP Region	
	No adverb	Speaker-oriented adverb	Mental-attitude adverb
Aspectual	486.90(838.08)	401.74(183.38)	411.62(185.05)
Control	432.96(498.61)	437.44(328.86)	404.03(223.10)
		Post1 Region	
	No adverb	Speaker-oriented adverb	Mental-attitude adverb
Aspectual	707.26(683.99)	653.62(668.93)	669.68(644.56) >C
Control	618.41(550.11)	637.44(652.05)	605.53(551.45)
		Post2 Region	
	No adverb	Speaker-oriented adverb	Mental-attitude adverb
Aspectual	503.07(246.78) > MA-A, NoAdv-C	483.25(268.06)	460.35(201.75) >C
Control	462.39(231.33)	469.44(247.70)	474.44(266.10)
		Post3 Region	
	No adverb	Speaker-oriented adverb	Mental-attitude adverb
Aspectual	401.91(156.54)	389.19(135.16)	396.12(153.30)
Control	397.58(156.98)	392.39(135.34)	393.04(155.94)
		Adverb Region	
	No adverb	Speaker-oriented adverb	Mental-attitude adverb
Aspectual	no value	470.24(329.31)	503.64(465.68)
Control	no value	486.01(370.92)	505.28(583.35)

4.2.1. Region 1: Verb

There were no significant differences in verb type ($F_{(1, 3476)} = 3.28, p = .070$) or adverb type ($F_{(2, 3477)} = 1.63, p = .20$), and no interaction between verb type and adverb type ($F_{(2, 3479)} = 0.80, p = .45$).

4.2.2. Region 2: NP

There were no significant differences in verb type ($F_{(1, 3476)} = 0.74, p = .39$) or adverb type ($F_{(2, 3477)} = 0.33, p = .72$). A significant interaction of verb type and adverb type was found ($F_{(2, 3479)} = 5.63, p = .0036$), but no significant difference was found in the pairwise contrasts adjusted by the Bonferroni method.

4.2.3. Region 3: Post 1

A main effect of verb type ($F_{(1, 3476)} = 11.09, p = .0009$) was found in the Post 1 region, indicating that the reading times for the conditions with aspectual verbs were significantly longer than for the conditions with control verbs. No significant effects were found in adverb type ($F_{(2, 3476)} = 1.90, p = .15$) and the interaction of verb type and adverb type ($F_{(2, 3477)} = 1.57, p = .21$).

4.2.4. Region 4: Post 2

A main effect of verb type ($F_{(1, 3476)} = 9.45, p = .0021$) was found in the Post 2 region where the 3rd and 4th characters after NP complement appeared, indicating that the
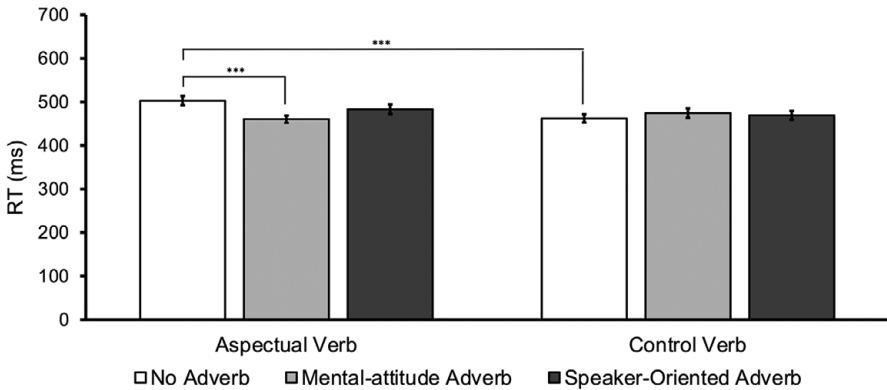
Figure 2. Reading times in the Post 2 region (*** p < .001). Error bars indicate standard error.

reading times for the conditions with aspectual verbs were significantly longer than for the conditions with control verbs.

In addition, an interaction between verb type and adverb type ($F_{(2, 3479)} = 6.47, p = .0016$) was found in the Post 2 region as shown in Figure 2. Contrasts indicated that the reading time for the MA-Asp condition was significantly shorter than for the NoAdv-Asp condition ($t_{(3478)} = -4.04, p = .0005$). It was also found that the reading time for the NoAdv-Asp condition was significantly longer than for the NoAdv control condition ($t_{(3478)} = 4.37, p = .0001$).

4.2.5. Region 5: Post 3

No main effect of verb type ($F_{(1, 3476)} = 0.82, p = .37$) or adverb type ($F_{(2, 3477)} = 0.98, p = .37$) was found in the Post 3 region. Also, there was no interaction between verb type and adverb type ($F_{(2, 3479)} = 0.89, p = .41$).

4.2.6. Region 6: Adverb

There were no significant adverb differences across the 4 adverb conditions ($F_{(3, 2269)} = 0.94, p = .42$).

Finally, no significant differences were found at any region between the 3 control conditions (the NoAdv control, the SO control, and the MA control).

In sum, a verb type effect of Asp conditions compared to control conditions was found in the Post 1 and the Post 2 regions, indicating that the processing of sentences with aspectual verbs was more costly than that of sentences with control verbs in Mandarin. A context effect of MA-Asp condition compared to NoAdv-Asp condition was found in the Post 2 region, indicating that by adding mental-attitude adverbs, the processing of sentences with aspectual verbs was facilitated.

5. Experiment 3: sentence interpretation post-test

A sentence interpretation post-test was conducted following the self-paced reading experiment. The purpose of the post-test was to help interpret results of the self-paced reading experiment. Participants for the post-test were the same participants as in the self-paced reading experiment. After finishing the self-paced reading experiment,
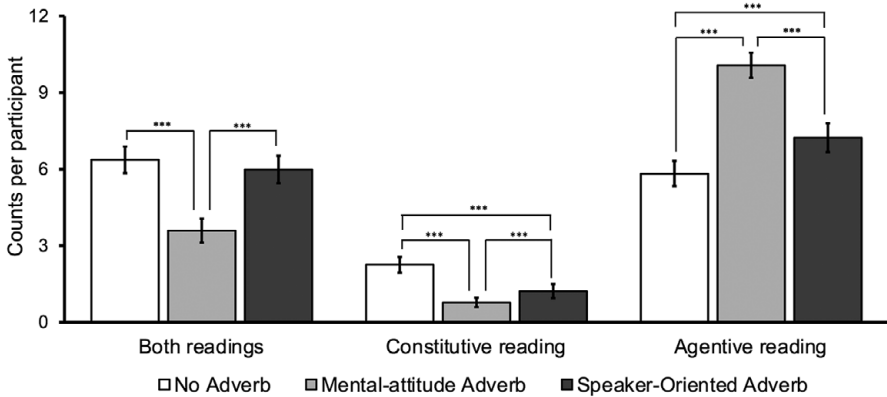
Figure 3. Results of the sentence interpretation post-test (***$p < .001$). Error bars indicate standard error.

participants were given time to take a break. Once they were ready for the following task, they pressed the space bar to start the sentence interpretation post-test.

The research design and hypotheses are the same as for Experiment 1. The post-test was also a forced-choice task. Stimuli included sentences of three conditions, namely, the MA-Asp condition, the SO-Asp condition, and the NoAdv-Asp condition, divided into 6 lists, as in Experiment 2. Each list contained 15 sentences for each of the three conditions. The post-test stimuli that participants read were the same stimuli that they read in the Experiment 2. The same data analysis was conducted as in Experiment 1.

5.1. Results of the sentence interpretation post-test

Results of Experiment 3 are shown in Figure 3. For the choice in which both readings are acceptable, a main effect of condition was found ($\chi^2_{(2)} = 110.31$, $p < .0001$). A pairwise comparison indicated that the choice of 'both' readings for the NoAdv-Asp condition occurred more frequently than for the MA-Asp condition ($z = 9.52$, $p < .0001$). In addition, the choice of 'both' readings for the SO-Asp condition occurred more frequently than for the MA-Asp condition ($z = 8.34$, $p < .0001$). There was no significant difference between the NoAdv-Asp condition and the SO-Asp condition ($z = 1.19$, $p = .70$).

For the choice of 'constitutive' reading, a main effect of condition was found ($\chi^2_{(2)} = 66.70$, $p < .0001$). A pairwise comparison indicated that the choice of 'constitutive' reading for the NoAdv-Asp condition occurred more frequently than for the MA-Asp condition ($z = 2126.52$, $p < .0001$). In addition, the choice of 'constitutive' reading for the SO-Asp condition occurred more frequently than for the MA-Asp condition ($z = 582.20$, $p < .0001$). Differing from the results of Experiment 1, a pairwise comparison also indicated that the choice of 'constitutive' reading occurred more frequently for the NoAdv-Asp condition than for the SO-Asp condition ($z = 1303.39$, $p < .0001$).

For the choice of 'agentive' reading, a main effect of condition was found ($\chi^2_{(2)} = 209.81$, $p < .0001$). A pairwise comparison indicated that choice of 'agentive' reading for the NoAdv-Asp condition occurred less frequently than for the MA-Asp

condition ($z = -13.65$, $p < .0001$). The choice of 'agentive' reading for the SO-Asp condition occurred less frequently than for the MA-Asp condition ($z = -9.36$, $p < .0001$). Moreover, the choice of 'agentive' reading occurred less frequently for the NoAdv-Asp condition than for the SO-Asp condition ($z = -4.72$, $p < .0001$).

6. Discussion

In the present study, the effect of context in the processing of expressions involving aspectual verbs was investigated in Mandarin. We sought to determine whether context affects the availability of agentive and constitutive readings in an offline behavioral test; and whether context that biases to an agentive reading facilitates the real-time processing of aspectual verb sentences compared to contexts that are non-biasing.

6.1. Aspectual conditions compared to controls

Significant processing costs were found in the self-paced reading experiment when comparing sentences with aspectual verbs with their controls. Effects were found in the Post 1 region, the 1st and the 2nd characters after the complement, and in the Post 2 region, the 3rd and 4th characters after the complement. The reading times for the aspectual conditions were significantly longer than for the control conditions at both Post 1 and Post 2 regions. The processing costs of expressions with aspectual verbs found in the previous literature (Ma et al., 2022) in Mandarin were successfully replicated, although a *SHI…LAI…* sentence structure was used in current study. The replicated effects suggest that the *SHI…LAI…* sentence structure did not affect the processing of aspectual verb expressions in an unexpected way.

6.2. NoAdv-Asp condition versus MA-Asp condition

Results of the offline sentence interpretation task (Experiment 1) and the post-test (Experiment 3) showed that both agentive and constitutive readings were available for the NoAdv-Asp condition, while interpretation was biased to the agentive reading for the MA-Asp condition. The offline results suggested that mental-attitude adverbs reduced the number of available readings in aspectual verb expressions.

Consistent with the offline results, the online self-paced reading experiment (Experiment 2) found attenuation of processing costs in the MA-Asp condition compared to the NoAdv-Asp condition. The significant difference occurred in the Post 2 region, where NoAdv-Asp sentences were read more slowly than sentences with MA-Asp adverbs.

Processing costs were found in the NoAdv-Asp condition compared to the NoAdv control, but were not observed in the MA-Asp condition compared to the MA control. The absence of processing costs suggests that the MA-Asp condition is as easy to process as its control. The attenuation of processing costs is thus interpreted as a context effect in real time, which is likely due to the fact that MA adverbs as context reduced the available dimensions, and therefore facilitated the processing of aspectual verb sentences. [Note: reducing the readings to only agentive entails reducing the dimensions to one, namely, the eventive dimension. A wide range of dimensions are available with the constitutive reading.]

Integrating the findings of the offline sentence interpretation experiments and the online self-paced reading experiments lends support to the view that grammatical context *predetermine*s the dimension rather than merely *privileges* them, i.e., leaves them all underspecified. Lai et al. (2023) proposed that if a biasing context *predetermined* one dimension, the other dimensions would be suppressed, and therefore processing would be facilitated; and if a biasing context merely *privileged* dimensions, all dimensions would still be available, and therefore the online processing of aspectual verb sentences would be as costly as in non-biasing contexts.

Lai and Piñango (2019) found context effects in their *offline* acceptability judgment experiments, but Lai et al. (2023) found no context effects in their online eye-movement experiment. In view of their findings, they argued that all dimensions remained underspecified in online processing, and they were only resolved in offline measures because participants were forced to make a choice.

However, in the present study, context effects *were* found in both sentence interpretation experiments and in the online self-paced reading experiment. These findings suggest that mental-attitude adverbs *predetermined* the agentive reading and its associated eventive dimension, and suppressed the range of dimensions associated with a constitutive reading. So, the number of available readings reduced from two to one, and the activation of multiple dimensions was reduced from many to one.

6.3. SO-Asp condition versus MA-Asp condition

Results of the offline sentence interpretation experiments showed that both agentive and constitutive readings were available in the SO-Asp condition, whereas interpretation was biased to the agentive reading in the MA-Asp condition.

However, in the online experiment, no difference was found between the SO-Asp condition and the MA-Asp condition. This could be due to a limitation of the self-paced reading method, which is less sensitive than other methods, such as eye-tracking, and a less natural form of reading. Conceivably, the tool was too coarse as a measure of reading to detect subtle differences between the two conditions.

6.4. NoAdv-Asp condition versus SO-Asp condition

No difference between the NoAdv-Asp condition and the SO-Asp condition was found in the sentence interpretation task or in the self-paced reading experiment when they were directly compared.

However, results of Experiment 3 regarding the two conditions were a little different from the results of Experiment 1. In Experiment 3 (the post-test), results showed that for the choice of the 'both' reading, the two conditions were the same, but regarding the choice of the 'agentive' reading, it occurred more frequently for the SO-Asp condition compared to the NoAdv-Asp condition. It is the other way around for the choice of the 'constitutive' reading: that reading was chosen less frequently for the SO-Asp condition compared to the NoAdv-Asp condition. The post-test of the self-paced reading results suggested that participants reacted differently to the NoAdv-Asp condition than they did to SO-Asp condition.

Regarding verb type effects in the self-paced reading experiment, a significant difference between the NoAdv-Asp condition and the NoAdv control condition occurred in the Post 2 region. However, verb type effects between the SO-Asp condition and the SO control condition were not found in the Post 2 region.

Given the observed verb type effects in the Post 2 region, it seems that the SO-Asp condition is as easy to process as its control. This suggests that SO adverbs may facilitate processing. By contrast, when we directly compare the NoAdv-Asp condition to the SO-Asp condition, there is no significant difference. Given this null effect of context, it seems that the SO-Asp condition is as hard to process as the NoAdv-Asp condition. Thus, SO adverbs do *not* seem to facilitate processing. Apparently, the two interpretations are contradictory.

So, the processing of aspectual verb sentences with SO adverbs is a puzzle that cannot be solved in this study, most likely due to the limitations of the self-paced reading method.

7. Conclusion

In the research reported here, we investigated the effects of context on the processing of aspectual verb expressions in Mandarin. Two sentence interpretation experiments and one self-paced reading experiment were carried out to give a thorough view of both offline behavioral effects and real-time processing.

The most striking and novel finding is of a real-time effect of context in the processing of aspectual verb sentences. Compared with aspectual verb sentences with no adverb, processing costs for aspectual verb sentences with MA adverbs attenuated in real time. This finding partially supports the idea that a grammatical context, such as the one provided by MA adverbs, *predetermines* the eventive dimension rather than keeping a broad range of dimensions open.

In this study, we made a preliminary and intuitive distinction between grammatical context and pragmatic context. Further research on the nature of the two types of contexts would be useful. The finding is partial because, for the claim to go through completely, sentences with MA adverbs would also take less time to read online than sentences with SO adverbs. But this was not observed. If the reason for this non-finding really was due to the coarse nature of self-paced reading as a tool for examining subtle grammatical distinctions, then an eye-tracking study should be sensitive to those distinctions.

References

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.

Baggio, G., Choma, T., Van Lambalgen, M., & Hagoort, P. (2010). Coercion and compositionality. *Journal of Cognitive Neuroscience*, 22(9), 2131–2140.

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.

Bürkner, P. (2017). brms: An R package for Bayesian multilevel models using stan. *Journal of Statistical Software*, 80(1), 1–28. https://doi.org/10.18637/jss.v080.i01.

DiNardo, L. (2015). Competing analyses of complement coercion: New evidence from behavioral and electropsychophysiological methods. *Senior thesis*.

Ernst, T. (2001). *The syntax of adjuncts* (Vol. 96). Cambridge University Press.

Godfroid, A. (2019). Eye Tracking in Second Language Acquisition and Bilingualism: A Research Synthesis and Methodological Guide (1st ed.). Routledge. https://doi.org/10.4324/97813157756162.

Jackendoff, R. S. (1972). *Semantic interpretation in generative grammar*. The MIT Press.

Katsika, A., Braze, D., Deo, A., & Piñango, M. M. (2012). Complement coercion: Distinguishing between type-shifting and pragmatic inferencing. *Mental Lexicon*, 7(1), 58–76.

Lai, Y. Y., Braze, D., & Piñango, M. M. (2023). The time-course of contextual modulation for underspecified meaning: An eye-movement study. *Mental Lexicon*, 18, 41–93.

Lai, Y.-Y., Lacadie, C., Constable, T., Deo, A., & Piñango, M. M. (2017). Complement coercion as the processing of aspectual verbs: Evidence from self-paced reading and fMRI. In Chungmin Lee (ed.), *Compositionality and concepts in linguistics and psychology* (pp. 191–222). Springer.

Lai, Y.-Y., & Piñango, M. M. (2019). Searching for specific sentence meaning in context: The conceptual relation between participants. *Language and Cognition*, 11(4), 582–620.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, 69, 1–33.

Li, X., & Li, X. (2014). Functions and distributions of 'lai' in Modern Chinese. *Journal of Hengyang Normal University*, 35(2), 85–88.

Lin, T. H. J., & Liu, C. Y. C. (2005). Coercion, event structure, and syntax. *Nanzan Linguistics*, 2, 9–31.

Liu, W., & Kempson, R. (2018). Chinese cleft structures and the dynamics of processing. *Transactions of the Philological Society*, 116(1), 91.

Lu, X. (2006). 'Lai' as a focus marker. *Chinese Teaching in the World*, 2, 20–30.

Ma, Y., Buccola, B., Wang, Z., Cousins, S., Godfroid, A., & Beretta, A. (2022). Expressions with aspectual verbs elicit slower reading times than those with psychological verbs: An eye-tracking study in Mandarin Chinese. *Journal of Psycholinguistic Research*, 52, 1–37.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. doi:10.1016/j.jml.2017.01.001

McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition*, 78(1), B17–B25.

Peirce, J. W. (2019). PsychoPy—psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1–2), 8–13.

Piñango, M. M., & Deo, A. (2016). Reanalyzing the complement coercion effect through a generalized lexical semantics for aspectual verbs. *Journal of Semantics*, 33(2), 359–408.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Singmann, H., Bolker, B., Westfall, J., Aust, F., & Ben-Shachar, M. (2023). *afex: Analysis of Factorial Experiments*. R package version 1.2-1. https://CRAN.R-project.org/package=afex.

Strand, J., Simenstad, A., Cooperman, A., & Rowe, J. (2014). Grammatical context constrains lexical competition in spoken word recognition. *Memory & Cognition*, 42, 676–687.

Tang, T. C. (1983). Focusing constructions in Chinese: Cleft sentences and pseudo-cleft sentences. In *Studies in Chinese Syntax and Semantics Universe and Scope: Presupposition and Quantification in Chinese* (pp. 127–226). Student Book Co. Ltd.

ter Meulen, A. (1990). English aspectual verbs as generalized quantifiers. In Aaron L. Halpern (ed.), *The ninth West Coast Conference on Formal Linguistics, Stanford, CA*. Center for the Study of Language and Information.

Traxler, M. J., McElree, B., Williams, R. S., & Pickering, M. J. (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory and Language*, 53(1), 1–25.

Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, 47(4), 530–547.

Xu, J. (2003, October). Focus-marking in Chinese and Malay: A comparative perspective. In Dong Hong Ji, Kim Teng Lua (eds.), *Proceedings of the 17th Pacific Asia Conference on Language, Information and Computation* (pp. 2–15).