



Revisiting the connection between progressive aspect and motion event cognition: Evidence from L1 Mandarin Chinese and L2 Swedish

Qiu-Jun Zhang

Department of Language Studies, Umeå University, Humanisthuset, Biblioteksgränd 3, 901 87 Umeå, Sweden

Received 5 April 2024; revised 25 September 2024; accepted in revised form 25 September 2024;

Abstract

This study investigates whether grammatical aspect influences the extent to which speakers attend to the endpoint (Goal) of motion events. Taking a typological perspective, the study examines the use of progressive markers in Mandarin Chinese through a progressive questionnaire. After identifying Mandarin's aspectual features, the study re-examines the relationship between progressive aspect and endpoint preferences by analyzing motion event descriptions made by native Mandarin speakers and comparing their performance with L1 Mandarin learners of L2 Swedish in a memory-based triad-matching task. The results of these comparisons reveal that Mandarin speakers mentioned the endpoint of an event as frequently as [+aspect] English speakers but significantly less frequently than [-aspect] Swedish and Afrikaans speakers, thereby indicating an 'endpoint preference' for Swedish and Afrikaans speakers. In a non-linguistic similarity judgment task, no significant differences were observed between the L1 and L2 groups in their frequency of pairing the target clip with the endpoint-highlighted alternative clip. However, a positive correlation between the participants' length of stay in Sweden and endpoint preferences was identified, implying a cognitive shift from 'progress salience' in [+aspect] L1 to 'endpoint salience' in [-aspect] L2. This suggests that immersion in an L2 context influences language acquisition and cognitive restructuring.

© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Motion event; Grammatical aspect; Endpoint behavior; Similarity judgment; Cognitive restructuring

1. INTRODUCTION

Motion is a complex and dynamic domain that is essential to our perception and daily interaction with the world around us. Whether it is the simple act of walking, cars moving on highways, or rockets launching into outer space, we can either visually perceive motion or actively experience motion. In linguistics, a 'motion event' is commonly defined as any physical movement that changes an entity's spatial location over time (Casati and Varzi, 2008). Consider the final stretch of a marathon: an athlete appears around the street corner, runs along the road, and dashes toward the finish line. Within ten seconds, the athlete changes their position from the corner to the finish line, completing a motion

E-mail address: qiujun.zhang@umu.se

<https://doi.org/10.1016/j.lingua.2024.103827>

0016-7037/© 2024 The Author(s). Published by Elsevier B.V.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

event. This motion event includes several components: a figure (e.g., *the athlete*), the manner of motion (e.g., *run*, *dash*), the path of motion (e.g., *the corner* as Source, *along the road* as Trajectory, *toward the finish line* as Goal or Endpoint), and a temporal contour (e.g., whether the race is ongoing or the finish line has been reached) (Comrie, 1976; Talmy, 1985, 2000).

Languages across the world provide rich linguistic categories to describe the semantic features of motion events. Since Talmy (1991, 2000) noticed that different languages exhibit different verb lexicalization patterns when expressing path and manner of motion, the domain of motion has been a fruitful research area for exploring the interplay between language and cognition (Wolff and Holmes, 2010). A significant amount of research has examined cross-linguistic differences in encoding the manner and path of motion events and their linguistic effects on task performances related to memory and the categorization of motion events (e.g., Papafragou et al., 2008; Ji and Hohenstein, 2017). However, less attention has been paid to the grammatical category of aspect, which specifies the temporal properties of an event, including whether a specified motion is ongoing or completed. (e.g., von Stutterheim et al., 2012; Athanasopoulos and Bylund, 2013). Some experimental studies in this research area have observed connections between grammatical aspect and how speakers perceive and conceptualize endpoints in motion events.

Imagine a goal-oriented motion event where a car moves rapidly along the road toward a gas station in the distance. To describe this scene, speakers can adopt either a maximal viewing frame (of the event) with endpoints included (i.e., a car driving toward the gas station) or a minimal viewing frame that focuses only on the ongoing action (i.e., a car driving) (see Langacker, 2008). Previous research on this topic has found that speakers of languages without obligatory grammatical markers to convey progressive meaning (i.e., so-called non-aspect languages, e.g., German and Swedish) were more likely to adopt a maximal viewing frame of an event and mention endpoints in their verbal descriptions of real-life motion events compared to speakers whose language display aspectual markers for 'ongoingness' (e.g., English). In addition to the linguistic evidence, the preference of speakers of non-aspect languages to focus on an event's endpoint (relative to speakers of aspect languages) has also been attested by their more frequent and prolonged fixation on potential endpoints (von Stutterheim et al., 2012). Furthermore, speakers of non-aspect languages more frequently select a reached endpoint as the criterion for categorizing a goal-oriented motion event in similarity judgment tasks (Bylund and Athanasopoulos, 2015).

However, recent studies using similar paradigms but involving speakers of different language pairs have not replicated the previously observed cross-linguistic differences in endpoint preferences (Liao et al., 2020; Misersky et al., 2022). This experimental inconsistency casts doubt on the hypothesis whether grammatical aspect does influence a speaker's endpoint preference. One potential explanation for these inconsistent findings may be related to how languages with typological differences in grammatical aspect are classified and compared. Several languages have been classified traditionally as either aspect languages (e.g., English, Chinese) or non-aspect languages (e.g., German, Dutch, Swedish) (Bylund et al., 2013; Misersky et al., 2022). This binary classification is primarily based on the absence or presence of obligatory grammatical markers to encode 'progressivity'. However, it is important to note that not all so-called aspect languages *obligatorily* mark progressive aspect. For example, Mandarin Chinese does not require explicit verb marking to convey temporal information about an event (e.g., the sentence without any aspectual markers *ta pao xiang hubian* 'he runs toward the riverbank' conveys a progressive meaning implicitly). Instead, the relevant temporal information can be inferred from the meaning of the verb and the context (Li and Thompson, 1981). Thus far, only a few empirical studies have investigated the optional and flexible usage of progressive markers in Mandarin Chinese. Consequently, its precise status among 'aspect languages' remains undetermined. One aim of the present study is to assess how obligatory progressive markers in Mandarin Chinese are. This is achieved by examining their occurrence in different contexts and providing an account of alternative ways to express 'progressivity' in Mandarin Chinese besides using designated progressive markers.

Given the inconsistent research findings regarding endpoint preferences in speakers of typologically different languages, this study revisits the hypothesis that grammatical aspect affects how speakers conceptualize and perceive the endpoint of a motion event by including a new set of research participants, namely, native Mandarin speakers and L1 Mandarin learners of L2 Swedish. I first examine how Mandarin speakers, who have a complex and flexible aspectual system at their disposal, encode and organize endpoint information of motion events during online speech production. Second, I identify how they categorize endpoints in a non-linguistic cognitive task. After establishing a baseline with respect to the above with L1 Mandarin speakers, I investigate L1 [+aspect] Mandarin speakers with advanced knowledge of L2 [-aspect] Swedish. This second group of participants allows for a bilingual perspective of the relationship between language and cognition in the domain of motion. The different linguistic typologies regarding the progressive aspect in Mandarin Chinese and Swedish enable us to investigate whether the cognitive patterns of Chinese learners of Swedish (advanced L2) differ from L1 Mandarin Chinese speakers because of the influence of an additional language at a relatively high level of proficiency.

2. BACKGROUND

2.1. Grammatical aspect and motion event endpoint

The grammatical category of aspect provides speakers with different perspectives on “the internal temporal constituency of a situation” (Comrie, 1976: 3). The temporal information of an event is traditionally divided into completed or ceased states (perfective aspect, e.g., *he has reached the goal*) and ongoing situations (imperfective aspect, e.g., *he is cycling*). The perfective aspect offers a holistic perspective of an event and presents it as a whole situation, including its beginning, intermediate, and end phases. In contrast, the imperfective aspect views an event in terms of its internal structure and focuses on the ongoing phase of the event (Comrie, 1976: 24). Across languages, the primary distinction between the perfective and the imperfective has been grammaticized and lexicalized, but there are differences in the prominence given to aspect encoding in different languages. Broadly speaking, ‘aspect languages’ have grammaticized aspectual systems to denote ‘imperfectivity’ and its subcategories. For example, in English, the obligatory usage of the auxiliary *BE* and a gerund (V-ing) in a wide range of contexts makes progressivity a fully grammaticalized feature of English (Smith, 1997: 67). In contrast, languages like German, which are referred to as ‘non-aspect languages’, do not possess verbal morphological markers to convey progressive meanings.

By explicitly linking linguistic grammatical aspect with descriptions of motion events, previous research has addressed Whorfian linguistic relativity, namely, whether the linguistic structures present in the language we speak influence the way we perceive and think about the world around us (Whorf, 1956). Some studies have provided support for this hypothesis by demonstrating that the presence or absence of progressive markers does affect (i) how frequently speakers encode endpoints and (ii) to what extent endpoints are attended to (e.g., von Stutterheim and Nüse, 2003; von Stutterheim et al., 2012). For example, Athanasopoulos and Bylund (2013) found that [-aspect] Swedish speakers mentioned endpoints more frequently than [+aspect] English speakers when describing real-life scenes where a vehicle or a person moved along a path toward an unreached endpoint. These language-specific patterns in describing motion event endpoints align with other classes of evidence. For example, in von Stutterheim et al.’s (2012) eye-tracking study, speakers of non-aspect languages (Czech, Dutch, and German) were found to allocate more visual attention to the endpoint region of a goal-oriented motion than speakers of aspect languages (English, Spanish, Russian, and Arabic). Furthermore, this increased attention to a motion event’s endpoint region was echoed in the descriptions provided by non-aspect language speakers. Cross-linguistic differences that are revealed during the description of motion events can be explained by the ‘thinking-for-speaking’ hypothesis that claims that speakers, while preparing information for verbalization, tend to focus on aspects of reality that are closely related to the linguistic categories available to them (Slobin, 1996, 2003). It has been argued that the grammaticized concept of ‘ongoingness’ in aspect languages makes their speakers more likely to take notice of the internal temporal constituency of an event and that this can lead to less frequent reference to endpoints or even no reference to an endpoint in their verbal descriptions (von Stutterheim et al., 2012; Bylund et al., 2013).

While real-time language use during speech provides insight into how speakers select and structure information for explicit verbalizations of motion events, such experiments primarily capture linguistic representations of visual stimuli and may only partially reflect the underlying thought processes. Therefore, to more thoroughly interrogate the Whorfian hypothesis, there is a growing call among researchers to examine nonlinguistic cognitive performance beyond the conscious act of speaking, especially performances involving higher-level cognitive processes such as ‘event categorization’ (see Casasanto and Boroditsky, 2008; Feinmann, 2020). In a study by Athanasopoulos and Bylund (2013), a triad-matching paradigm was implemented to assess whether the encoding of grammatical aspect affects a speaker’s similarity judgments of goal-oriented motion events and, if so, under what conditions can such effect(s) be observed. In their triad-matching task, participants were asked to judge whether a target scene (in a video clip) showing a possible but unreached endpoint (an intermediate degree of goal orientation) was more similar to an alternative scene without a visible endpoint (a low degree of orientation) or a scene that depicted a reached endpoint (a high degree of orientation). The video clips (sets of three) were presented to native speakers of English and Swedish under two conditions. The first was a memory-based condition where the clips in each triad appeared one after another with no pause, requiring the participants to recall what they had watched before making their judgment. The second condition was an online condition where a triad of clips was played synchronously in a loop until the participant provided a response. Despite displaying an overall preference for the low-degree of orientation alternative, the Swedish-speaking participants matched the target scene with the high-degree of orientation alternative significantly more often than the English speakers did in the memory-based matching task. This result aligns with the linguistic observation that Swedish speakers are more inclined to encode endpoints in verbal descriptions of motion event scenes than English speakers. However, no statistically significant cross-linguistic differences under the online condition were observed. Under the online condition, the participants were not expected to verbalize event scenes for later judgments since they could view the target clip and the

two alternatives playing simultaneously when they made their categorical decision. Given the varied results under the different conditions, one may argue that the effect of grammatical aspect on motion event cognition may be triggered when language is recruited online to solve a memory-based task (Athanasopoulos and Bylund, 2013: 302).

The same verbal description and memory-based triad-matching tasks have been performed by speakers of aspect languages, such as English and Spanish, and by speakers of non-aspect languages, such as Afrikaans, German, and Swedish (e.g., Bylund and Jarvis, 2011; Bylund et al., 2013). It was concluded that speakers of non-aspect languages were more inclined to mention endpoints of goal-oriented motion events and to base categorical decisions on endpoints to a greater extent than speakers of aspect languages (Athanasopoulos and Bylund, 2013; Bylund et al., 2013). Nevertheless, more recent research has not been able to replicate this endpoint bias in speakers of non-aspect languages, including Liao et al. (2020) and Misersky et al. (2022). In Liao et al.'s (2020) study, speakers of [+aspect] Mandarin Chinese and [-aspect] Dutch did not differ in terms of their frequency of mentioning endpoints or in their performance in endpoint recognition memory tasks. According to Liao et al. (2020: 513), the absence of an apparent linguistic effect on endpoint behavior can be attributed to the linguistic fact that Mandarin and Dutch do not neatly align with either the aspect or non-aspect language binary classification. While aspectual markers are not consistently used in Mandarin Chinese, periphrastic constructions are available in Dutch to encode atelic activities in progress. However, Liao et al.'s (2020) discussion stops short of analyzing the specific options encoding progressive aspect in each language that do not conform to its assigned language classification. For instance, while Mandarin Chinese is recognized as an aspect language, sentences without any aspectual markers on the verb can be readily produced. The lack of research on non-aspectual features in aspect languages, and vice versa, underscores a need for further investigation. It is crucial not to rely solely on the traditional binary classification of languages as aspectual or non-aspectual; instead, one should consider how the grammaticalization of progressive aspect affects speakers' attention to event endpoints.

For example, Misersky et al. (2022) do not report on any robust or replicable linguistic effects on the perception of motion event endpoints. Their study used immersive virtual reality (VR) combined with eye-tracking techniques, which provided a virtual environment where the participants could actively experience motion events by being the motion figure. This experimental approach differs from traditional computer-based experimental setups used in previous studies, where the participants passively watched videos and provided responses while seated in front of a computer. No significant differences between [-aspect] German and [+aspect] English speakers were found regarding their endpoint fixation patterns in this novel VR setting. During the experiment, the participants walked on a treadmill, encountered landmarks, and approached endpoint objects along a straight path. Misersky et al. (2022) argue that a setting that mimics the real world may eliminate any weak linguistic effects on motion event perception. However, this explanation is not entirely convincing. Knowing whether the new experimental setup accurately simulates real-world experiences is difficult to establish. Furthermore, the implementation of VR seems unnecessary since it offers little difference from traditional computer-based paradigms. For example, the participants are still focused on two-dimensional screens, with the main distinction being their movement on a treadmill. Therefore, a more compelling explanation for the absence of linguistic effects may lie in their participants' linguistic backgrounds. The group of German speakers did not consist of functional monolinguals. Most of them were immersed in an L2 environment, with many reporting good proficiency and frequent use of Dutch or English. A possible long-term L2 effect could explain the "English-like" pattern in the German group" (Misersky et al., 2022: 10).

2.2. Cognitive restructuring in the domain of motion

As previously mentioned, linguistic relativity studies in the domain of motion usually begin with questioning whether speakers of different L1s talk and think about motion events differently. This question can be extended to a bilingual context, where learners acquire an additional language with typological properties distinct from those of their L1. Many bilingual studies of motion events have examined whether L2 learners' lexicalization patterns of encoding manner and path in motion expressions differ from those of L1 and L2 monolinguals (e.g., Hendriks and Hickmann, 2015; Lewandowski and Özçalışkan, 2022). Essentially, these approaches still address the thinking-for-speaking phenomenon by questioning whether L2 learners stick to L1-based verb lexicalization patterns or shift to L2-based patterns during real-time language use. However, what does L2 learners' cognitive behavior look beyond the linguistic level? Specifically, there are three possibilities: L2 speakers might (i) behave similarly to native speakers of the target language (Montero-Melis et al., 2016); (ii) demonstrate a cognitive pattern along a continuum ranging from L1 patterns to L2 patterns (Ji, 2017); or (iii) retain a cognitive pattern consistent with their L1 (Filipović, 2018). Some studies have observed a long-term language effect where bilinguals switch their cognitive pattern from L1-like to L2-like as a function of their increasing L2 proficiency (Fuhrman et al., 2011), their amount of contact with L2 (Athanasopoulos et al., 2010), or age of acquisition (Boroditsky, 2001). Other research suggests that two patterns can be retained separately in a bilingual's mind, thus casting doubt on the existence of conceptual restructuring altogether (Pavlenko, 2011). Given the

ongoing debate in bilingual cognition, more studies are needed to investigate whether, and if so, to what degree L1-based cognitive patterns are recalibrated toward L2-based patterns.

In the domain of goal-oriented motion events, some studies have shown that learning another language can result in changes in L1-specific cognitive patterns (Athanasopoulos et al., 2015). Bylund and Athanasopoulos (2015) explored whether [-aspect] Swedish native speakers with advanced knowledge of L2 [+aspect] English recalibrate their categorization behavior. Operationalized by a memory-based triad-matching task, the categorization pattern of goal-oriented motion events by L1 Swedish learners of L2 English differed significantly from that of English monolinguals. Within the Swedish group, there was a significant negative correlation between endpoint preference and the amount of daily exposure to English through watching television. This finding suggests that the more English-language television they watched, the less likely the participants were to base similarity judgments on motion event endpoints (Bylund and Athanasopoulos, 2015: 129–130). In their interpretation of the role of watching television in restructuring the categorization of motion played, Bylund and Athanasopoulos (2015) referred to the multimodal nature of television, which simultaneously presents viewers with dynamic scenes and verbal descriptions. More specifically, they argued that viewers would develop an immediate viewing frame with event goals excluded in their minds after repeated exposure to the more visual aspects of audiovisual media in English, which emphasizes ongoing processes and provides less frequent reference to endpoints. The effect of L2 contact was also reported in a baseline study on Afrikaans speakers using the same triad-matching task (Bylund et al., 2013). Speakers of [-aspect] Afrikaans paired the target scene with the alternative showing a reached endpoint to a similar extent as [-aspect] Swedish speakers but more frequently than [+aspect] English speakers. Despite a significant difference in endpoint focus between Afrikaans and English speakers, variations among the Afrikaans-speaking participants can be partly predicted by how frequently they use English. The significant correlation between daily English use and endpoint preferences indicated that Afrikaans speakers who used English frequently were more likely to approximate the categorization preferences of English monolinguals.

This raises the question: Would L2 learners restructure their L1-specific cognitive patterns if they acquire L1 linguistic categories that have no corresponding units in L2? Athanasopoulos et al. (2015) address this question in L1 [+aspect] English learners of L2 [-aspect] German. Their results reveal that L2 learners become more inclined to make categorical decisions based on reached endpoints as they improve their L2 proficiency and have studied German at university for several years. Moreover, they also observed that long-term exposure L2 users (with an average exposure of 10.8 years) shifted toward the L2 behavior pattern in terms of endpoint saliency. Medium-exposure L2 users (with an average exposure of 7.5 years) seemed to approximate their L1 categorization pattern, suggesting a non-linear relationship between the length of L2 exposure and endpoint preferences. Based on an associative learning account, the absence of progressive markers in their L2 provides more cues for boundedness, leading bilinguals to readjust the weighting of endpoint saliency in motion event cognition after long-term exposure to L2. However, their study seems to be the only study that has examined the existence of cognitive restructuring in the context of an L1 with obligatory aspectual markers and an L2 without default aspectual markers.

2.3. Progressive aspect in Mandarin Chinese and Swedish

Most, if not all, human languages can express events in progress via lexical means, for example, by means of temporal adverbials. However, not every language employs grammatical devices to indicate progressivity. The semantic ‘progressivity’ is often associated with a timeframe during which a dynamic event unfolds (Chung and Timberlake, 1985). This timeframe can also be understood as a “focalization point”, a single point in time at which an event is ongoing (Bertinetto et al., 2000: 527). To express the meaning of a progressive situation over a specified timeframe, speakers may have a variety of linguistic tools at their disposal, including (i) the use of temporal adverbials such as *now* and *right now*, which is a common way to convey progressivity across languages; (ii) morphological marking on the verb through affixes; or (iii) periphrastic verbal constructions (Bertinetto et al., 2000; Mair, 2012). For example, the fully grammaticalized construction of BE + V-*ing* in English falls into both the second type (the suffix *-ing* on the verb) and the third type (the verb phrase led by a copula verb).

The two languages in the present study, Mandarin Chinese and Swedish, have different aspectual systems. Mandarin distinguishes between perfective and imperfective viewpoints by means of the imperfective markers *zai* and *zhe* and the perfective markers *le* and *guo*. These aspectual markers are considered “the only kind of inflectional morphology-like devices” in Mandarin, hence Mandarin’s classification as an aspect language (Klein, 2018: 10). In contrast, Swedish lacks a dedicated morphological category for the expression of progressivity and is, therefore, classified as a non-aspect language in most of the literature (Athanasopoulos and Bylund, 2013). Swedish speakers frequently use verbs in the simple present tense to refer to an ongoing situation at the moment of speaking (Hinchliffe and Holmes, 2008).

To convey an imperfective meaning, speakers of Mandarin Chinese frequently use the preverbal *zai* and the postverbal *zhe* in discourse. The morpheme *zai* is typically regarded as a progressive marker that adds focus to the ongoing activity. As Smith (1997: 358) observes, the progressive *zai* encodes an internal interval within a durative situation, while such an interval is by no means an initial point or a final point. Given this, we note that *zai* is compatible with an unbounded event that is still unfolding. Consequently, *zai* frequently appears with activity or accomplishment verbs that imply durative processes but never with verbs denoting instant changes, for example, ‘die’ or ‘win’ (Liu, 2015: 285). Additionally, it is noteworthy that the meaning of *zai* is influenced by both spatial and temporal domains (Xiao and McEnergy, 2004). From a diachronic perspective, *zai* was originally a locative verb and preposition but has since been grammaticalized as a progressive marker in Mandarin (Klein et al., 2000). Given that *zai* still retains its spatial meaning in (1a) but has also evolved into a grammatical marker in modern Mandarin Chinese in (1b), *zai* might serve a dual role within a single sentence like (1c), functioning either as a locative preposition and a progressive marker.

(1) a.	他在公园里 <i>ta zai gongyuan li</i> he <i>zai</i> _{PREP} park inside ‘He is in the park.’	b.	他在走路 <i>ta zai zoulu</i> he <i>zai</i> _{PROG} walk ‘He is walking.’	c.	他在公园里走路 <i>ta zai gongyuan li zoulu</i> He <i>zai</i> _{PREP/PROG} park inside walk ‘He is walking in the park.’
--------	---	----	---	----	---

The other imperfective marker, *zhe*, is close to *zai* in its semantics but emphasizes an event's durative, stative state. Consequently, it is referred to as the marker of the durative/continuous aspect (Smith, 1997; Xiao and McEnergy, 2004). While one primary function of the durative marker *zhe* is to “convert the predicate into a state”, it does not conflict with its frequent co-occurrence with the progressive marker *zai* (Liu, 2015: 286). As illustrated in (2) below, the progressive *zai* focuses on the dynamic, ongoing process of singing songs, whereas the durative *zhe* encodes the resultative state of singing songs. It is also noteworthy that *zhe* can contribute to “a backgrounding effect” and “freeze” a situation or a state in front of speakers when it marks a verb in a subordinate clause (Smith, 1997: 360–361).

(2)	他在唱着歌。 <i>ta zai chang zhe ge</i> He <i>zai</i> _{PROG} sing <i>zhe</i> _{DURA} songs ‘He is singing songs.’
-----	---

While Swedish lacks a dedicated, obligatory marker for indicating ongoing situations, Swedish speakers can optionally resort to using a variety of periphrastic progressive constructions, including serial verbs led by a motion or postural verb as an auxiliary (e.g., *sitta och läsa* ‘sit and read’; *gå och sjunga*, ‘go and sing’) and the *hålla på*-construction (lit. ‘hold on’) (Kvist-Darnell, 2005). However, these periphrastic constructions are not as highly grammaticalized as the progressive aspect is in English. For instance, the *hålla på*-construction, using *hålla* ‘hold’ as a special auxiliary verb, exhibits a relatively low degree of grammaticalization in the sense that it is infrequent and typically occurs with accomplishment and achievement verbs that encode an ongoing but incomplete situation (Blenselius, 2015).

To conclude, Mandarin Chinese encodes progressivity through aspectual markers, while Swedish lacks such verb marking, indicating differences in the degree of grammaticalization of the progressive aspect. Moreover, note that Mandarin has a flexible aspectual system: verbs are not obligatorily marked to express aspectual contrasts. One can often infer aspectual information from the speaker's pragmatic context or the inherent meaning of the predicate (whether it is telic or atelic), not necessarily from grammatical markers (Xiao and McEnergy, 2004; Klein, 2018).

3. SCOPE OF THE STUDY

Situated within the domain of goal-oriented motion, the present study re-examines the hypothesis that speakers of typologically different languages (in terms of whether progressive aspect is encoded) have different biases toward motion event endpoints. Starting from a typological perspective, the study centers on Mandarin Chinese, an aspect language that has not received as much attention as other aspect languages in research on the role of aspect in event construal. Unlike English, which is known for its grammaticized progressive, Mandarin allows for the optional and flexible use of aspectual markers. This means these markers are optional for describing ongoing or completed situations, and their omission is not considered a grammatical error, depending on specific pragmatic and contextual factors. While some linguists have provided theoretical accounts of aspectually unmarked sentences in Mandarin (see Smith, 1997, for

a discussion of ‘neutral aspect’), empirical research on the absence of aspectual marking in Mandarin remains scarce and limited solely to examining corpus data. The present study focuses on the factors that invoke the use of progressive markers. A progressive questionnaire (PROGQ) from Bertinetto, Ebert, and de Groot (2000) was adapted to examine the frequency and conditions under which Mandarin speakers tend to omit progressive markers and to identify alternative linguistic devices for conveying progressivity when these markers are absent.

Given that Mandarin Chinese does not require obligatory marking on verbs as other aspect languages do, I pose the following research question: Is the hypothesis still valid that speakers of aspect languages are less likely to exhibit endpoint biases than speakers of non-aspect languages? This study tests this hypothesis by measuring Mandarin speakers’ endpoint behavior at the linguistic and cognitive levels. Using a series of short video clips for speech elicitation, the verbal experiment investigated how often Mandarin speakers encode event endpoints when they describe goal-oriented motion events under an online condition. In addition, the participants’ speech data was analyzed to assess the extent to which aspectual markers were combined with different path elements (i.e., source, trajectory, endpoint) within a single motion event. At a cognitive level, the Mandarin speakers’ cognitive behavior with regard to endpoints was operationalized by a categorization task that required the participants to make judgments based on the degree of event similarity. Categorization, as a fundamental element of human cognition, is believed to involve higher-level cognitive processing (Nosofsky, 1986). The categorization task utilized a triad-matching paradigm, which has been widely used to evaluate how speakers perceive event similarity in the domain of motion (Athanasopoulos and Bylund, 2013; Ji and Hohenstein, 2017). The verbalization and categorization tasks were adopted from Athanasopoulos and Bylund’s (2013) study, and the same experimental materials have been used with English, German, Swedish, and Afrikaans speakers. The encoding and categorization patterns of the Mandarin speakers were compared with those of previous participants to investigate whether Mandarin speakers think about motion event endpoints more like speakers of aspect languages or non-aspect languages.

After establishing a baseline with the native Mandarin-speaking group, this study explored the relationship between progressive aspect and motion event cognition through a bilingual lens. The same categorization task was conducted by Mandarin-speaking learners of L2 Swedish who have learned and acquired Swedish in a Swedish-speaking environment. While these individuals typically attach progressive markers to verbs in their L1 to denote ongoing actions (as evidenced in the baseline data), they also used an additional language (Swedish) that lacks morphological resources to express progressivity. Supposing that the presence or absence of progressive grammatical markers is closely linked with the entrenchment of specific temporal frames (maximal vs immediate) (Langacker, 2000, 2008), I ask: How do these L1 Mandarin learners of L2 Swedish develop their L1-specific endpoint preferences after acquiring a non-aspect language? Finally, I gathered data on these L2 learners’ learning trajectories to determine which linguistic or extra-linguistic variables affected their endpoint behavior.

4. METHOD

In the following section, I present the participants’ language background, the design and materials, and the step-by-step procedures for the progressive questionnaire and the two experiments.

4.1. The progressive questionnaire

4.1.1. Participants

Twenty-five native speakers of Mandarin Chinese were recruited to complete a ‘progressive questionnaire’ (see next section) in standard Mandarin (adapted from Bertinetto et al., 2000).¹ The respondents were all born in mainland China (15 females and 10 males, with an average age of 29.8 years). The respondents originate from the following provinces: Shanghai (10 respondents), Hebei (3), Sichuan (3), Northeast (2), Anhui (1), Zhejiang (1), Jiangxi (1), Guangxi (1), Hubei (1), Gansu (1) and Xinjiang (1). According to their regions of origin, their native dialects are labelled as follows: ‘Wu dialect’ (spoken in Shanghai, Zhejiang), ‘Northern Mandarin’ (Hebei, Northeast, Gansu, Xinjiang), ‘South-Western Mandarin’ (Sichuan, Hubei, Guangxi), and ‘Jiang-Huai Mandarin’ (Anhui).² All of the respondents currently live in Shanghai and use Standard Mandarin in their daily work and studies.

¹ Standard Mandarin is the official language and the *lingua franca* of mainland China (Weng, 2018) and is based on the dialects of Mandarin spoken in northern China. In addition to Northern Mandarin, South-Western Mandarin and Jiang-Huai Mandarin are considered as two sub-dialects of Mandarin (Simmons, 1999).

² Alongside Mandarin, the Wu dialect is one of the major subdivisions of the Chinese language spoken in eastern China.

4.1.2. PROGQ design and materials

The original PROGQ ('progressive questionnaire') (Bertinetto et al., 2000) was designed to elicit progressive forms in a language. It was originally written in English and distributed to speakers of various European languages to establish a typological overview of the morpho-syntactic tools used to express progressivity across different European languages. PROGQ consisted of 83 sentences, each accompanied by specified linguistic contexts indicated by slashes and square brackets. Respondents to the questionnaire were asked to focus on the verb (presented in capital letters) and to translate English sentences into their mother tongue while also considering the provided context. In addition to contexts where a situation was ongoing at the moment of speech, some contexts presented combinations of several linguistic categories such as aspect (imperfective vs perfective), tense (present, past, and future), mood (indicative, imperative, speculative), the semantic properties of verbs (transitive, intransitive, motion, phasal, postural, non-durative, non-intentional, stative verb), or other properties relevant to the progressive (backgrounding, temporal reference).

The PROGQ used in the present study was adapted so that all of the sentences, accompanying contextual information, and questionnaire instructions were translated into Mandarin Chinese (see Appendix A). Instead of using slashes or square brackets, the contextual information was presented separately and marked with a number. The essential information, including Agent, Predicate, and Object, was listed in a column called 'content.' The respondents were instructed to construct complete sentences based on the provided content prompts, contextual information, and their daily speech habits. They were allowed to delete, change, or add any individual words as needed to create a natural-sounding sentence. Each question was constructed to elicit natural speech. For example, the first PROGQ question presented a dialogue in a context where someone wondered what Ann was doing at the moment of speech (see Table 1). In the description of the provided contexts, no aspectual markers were used to avoid any potential priming effects. Respondents were expected to add to the provided keywords in the 'content' column, for example, with a progressive marker *zai*, a time adverbial, or a particle.

4.2. Linguistic encoding task

4.2.1. Participants

Twenty native speakers of Mandarin Chinese participated in the linguistic encoding task after completing a basic language background questionnaire based on Marian's (2007) Language Experience and Proficiency Questionnaire (LEAP-Q). All of the participants (15 females and 5 males, with an average age of 25.7) were born and live in mainland China.

4.2.2. Design and materials

The experimental materials consisted of twelve video clips and six fillers in the form of PowerPoint slides created by Athanasopoulos and Bylund (2013). These materials have been previously used to elicit verbal endpoint encoding in similar studies (Athanasopoulos and Bylund, 2013; Bylund et al., 2013). The video clips were filmed and compiled by a research group led by Christiane von Stutterheim at the University of Heidelberg. Each clip presents a motion event where an entity moves toward a visible endpoint that is not reached by the end of the clip. For example, a clip may consist of a vehicle driving on a highway toward a gas station, some trees, or a mountain, as shown in Fig. 1). In contrast to the critical stimuli that present an intermediate degree of goal-orientation, the six filler clips depict non-motion dynamic events such as potato peeling and coffee drinking. In a manner identical to the Experiment 1 setup in Athanasopoulos and Bylund's (2013) study, each consecutive six-second clip was played automatically in full-screen mode using PowerPoint. An 8-second blue screen appeared between each clip, followed by a yellow star to indicate the start of the following clip. The filler clips were inserted between every two video clips depicting motion events.

Table 1

Illustration of how questions were presented in the adapted Chinese version of PROGQ (English translation added here).

场景一	电话里, 有人想了解你身边小安目前的情况。	Context	Someone on the phone wants to know about Ann. She is next to you.
内容提示	小安工作	Content	Ann WORK
你的回答	小安现在在工作	Answer	Ann is now working.



Fig. 1. Example of a motion event with an intermediate degree of goal-orientation in the linguistic task (critical stimulus).

4.2.3. Procedure

Before the task, the participants were informed that they would view 18 video clips depicting daily events. They were instructed to focus only on the event and briefly describe what was happening in the scene. After hearing a beep, the participants were asked to speak about what they had seen in each clip. They were given 8 s to complete each verbal description. The entire task took around 5–7 min to complete.

4.3. Memory-based triad-matching task

4.3.1. Participants

A different group of twenty-five native Mandarin speakers and twenty L1 Mandarin learners of L2 Swedish were recruited to complete the triad-matching task after consenting to participate in the study. Before the matching task, the L1 Mandarin participants completed the same questionnaire used in the LEAP-Q verbal task. The L2 Swedish learners completed a more detailed questionnaire (Marian, 2007), which collected linguistic and non-linguistic variables, including the participants' age when they began to learn Swedish, length of stay in Sweden, language exposure to Swedish media, daily interactions with native speakers, Swedish self-learning activities, and frequency of use and proficiency. The L1 Mandarin participants (14 females and 11 males, with an average age of 26.2) were born and raised in mainland China and reported no knowledge of any non-aspect languages. The Chinese learners of Swedish acquired Mandarin Chinese as their first language and began learning Swedish after they moved to Sweden (11 females and 9 males, with an average age of 28.7). This group consisted of intermediate Swedish learners who held at least a B1 certificate in the Common European Framework of Reference (CEFR) or had passed Swedish courses equivalent to B1 or higher at a Swedish university.

4.3.2. Design and materials

Data on motion event cognition was elicited by a non-linguistic categorization task. The task, originally designed by Athanasopoulos and Bylund (2013), used a memory-based triad-matching paradigm. The stimuli consisted of 19 triads, each including a target video clip and two alternatives with varying degrees of goal-orientation. The test items were selected from a pool of stimulus items created by von Stutterheim et al. (2012). Each clip lasted for 6 s. In each triad, the target clip showed a scene with an intermediate degree of orientation. The alternative clip with a low degree of orientation, labeled as [–endpoint], depicted an entity (e.g., a person or a vehicle) moving along a trajectory with no immediate endpoint(s) in sight. In contrast, the high-degree of orientation alternative, labeled as [+endpoint], contained a moving entity that arrived at a goal by the end of the clip. As illustrated in Fig. 2, the clips in each triad differed in terms of the Agent's relationship with an endpoint. The clips were matched and grouped to ensure consistency in the manner and direction of motion (e.g., riding a bicycle forward), as well as the number of Agents within each triad. The video clips were normalized for visual similarity to ensure that the participants based their judgments on endpoints or 'ongoingness' instead of other visual features. Thus, regardless of the participants' language background, [+endpoint] alternatives were perceived as visually similar to the [–endpoint] alternatives.

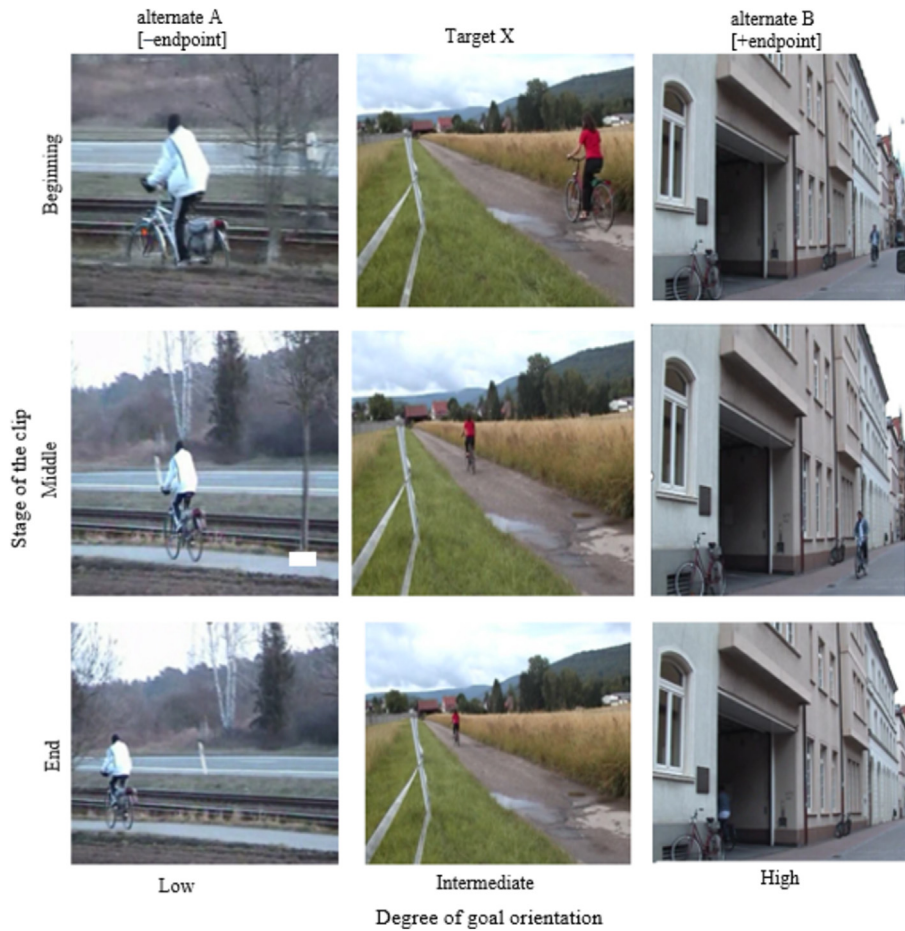


Fig. 2. Example of a triad of video clips.

The participants watched three types of clips in a triad: one alternative labeled 'A', another alternative labeled 'B', and a target clip labeled 'X'. After watching each clip, the participants judged whether clip 'A' or clip 'B' looked more like clip 'X'. With the sequence of the [+endpoint] and [-endpoint] alternative counterbalanced, 19 triads appeared twice, resulting in 38 triads. The display order of the triads was pseudo-randomized into 15 series by the PowerPoint software, and each series was randomly assigned to the participants.

4.3.3. Procedure

The experimental stimuli were presented on PowerPoint slides. All of the participants were informed beforehand in Mandarin Chinese that they would view 38 triads of video clips in total, each presented in an ABX format. After watching clip 'X' in its entirety, they were asked to tell the experimenter which clip, 'A' or 'B', was more similar to clip 'X'. Given that the participants might need extra time to recall or deliberate on their responses, they were encouraged to make judgments at their own pace and informed that there was no correct answer for each triad. After hearing the participants' oral responses, the experimenter circled their choice on the answer sheet and then pressed the 'enter' button to initiate the next test item. At the end of the test, the participants answered a follow-up question about their selection criteria. Each test session lasted for 15 to 20 min.

5. DATA ENCODING AND ANALYSIS

5.1. PROGQ

For the progressive questionnaire, a total of 2075 sentences were collected from twenty-five native speakers of Mandarin Chinese. As expected, not every given context elicited a linguistic encoding for progressivity or continuousness.

For example, none of the respondents used aspectual markers when describing a future event (Q66-Q69) or giving orders in imperative sentences (Q73-Q74). Such sentences were excluded from the analysis since the analysis focused on two areas, namely, (i) conditions where the progressive marker *zai* and durative marker *zhe* are used and (ii) alternative ways of conveying progressivity instead of using grammatical markers.

In one of the most frequent scenarios (Q1, Q6-Q28, Q35-38, Q44-47), the respondents were required to imagine themselves on a phone conversation and simultaneously answer questions about what someone nearby was doing at that moment. When tasked to focus on a single moment in a 'here-and-now' context, the respondents used their linguistic repertoire (grammatical markers, time adverbials, or other particles) to describe what was taking place at the time of speech. As summarized in Table 2 below, various linguistic means to express progressivity can combine and co-exist in Mandarin. In addition to the use of the progressive *zai* (on its own in the sentence), the co-occurrence of time adverbial *xianzai* (lit. 'now') and *zai*, and the co-occurrence of the adverbial *zheng* (lit. 'just') and *zai*, constituted two frequent linguistic patterns observed in the data.

The progressive marker *zai* was used in most cases when the Agent was engaged in an ongoing activity denoted by atelic verbs such as *read* and *sing*. In contrast, *zai* never occurred with phasal verbs such as *begin* and *finish* (Q23-Q26) and non-durative verbs implying a state of completion (Q30-32). For example, in the verb phrase *begin to peel potatoes* (Q23), a great deal of variation was shown in the response sentences. Some respondents interpreted the phrase as referring to an event that was happening and only added adverbials *zheng* or *gang* (lit. 'just'). However, some used the particle *yao* to indicate an immediate future where a person is about to peel potatoes, while other respondents did not use aspectual markers or any time adverbials by providing an answer based solely on the given content prompts. Note that the verb *begin* denotes the very outset of an event and lacks an internal interval. Despite the durative process conveyed by the second verb, *peel*, the serial verb phrase is incompatible with *zai*. The use of the progressive *zai* depends on the semantic properties of the verb(s) in the sentence. When the verb or verb phrase encodes instantaneous achievement, such as *reaching the mountain top* (see Vendler, 1957, for the classification of verb types), the respondents used either time adverbials meaning 'just now,' the postverbal perfective marker *le*, or a combination of both. These results confirmed the semantic restrictions of aspectual markers in Mandarin Chinese (Smith, 1997; Liu, 2015).

However, the occurrence of the progressive *zai* is also influenced by the sentence's meaning. Continuing with the activity verb *peel* as an example, the occurrence of *zai* depends on the specificity or quantity of potatoes (*peel* ∅/ *these/three kilos of/all the potatoes*). While respondents were more likely to use the progressive *zai* when describing an unspecified quantity of potatoes, they were more likely to use the perfective *le* when referring to a particular quantity of potatoes. As a telic verb, *peel* implies a natural endpoint, in this example, peeled potatoes. With quantifiers like *all* and *three kilos* preceding *potatoes*, the resultative state of a set of peeled potatoes can be highlighted. Consequently, the perfective *le* was frequently used to indicate the event of peeling potatoes as bounded. This contrasts with sentences where the noun *potato* is unspecified. The use of *zai* marks temporary engagement in the ongoing (but still telic) activity of peeling potatoes.

The second-most common scenario in the test asked the respondents to describe a past event based on a backgrounded event provided in a subordinate clause (Q48-Q55, Q70-72). To illustrate, the clause *when I was sleeping/slept* serves as a context in Q48 where the respondents were asked to modify the content of the main clause *she plays for two hours all by herself* by adding words as needed. In this case, most of the respondents attached the perfective *le* to the verb *play*, while only one respondent used the progressive *zai* without mentioning the duration *for two hours*. The lexical content of *for two hours* sets a boundary for the activity of playing and prompts the respondents to view the situation in its entirety, as denoted by *le*. While PROGQ was designed to investigate the complete range of progressive markers, it has been observed that some response sentences use the perfective aspect or remain aspectually unmarked. In Q54 and Q55, a specific moment of the Subject's arrival was focused on, as in *when I arrived*). In the main clause, 25% of the respondents used the progressive *zai* when describing a gradually improving situation. The occurrence of *zai* was even less frequent when the respondents described the gradual covering of the land by snow in Q55.

Table 2
Overt expressions for progressivity in the PROGQ data.

Progressive marker	<i>zai</i>
Time adverbial + progressive marker	<i>xianzai + zai</i>
Adverbial + progressive marker	<i>zheng + zai</i>
Time adverbial + adverbial + progressive marker	<i>xianzai + zheng + zai</i>
Progressive marker. + modal particle	<i>zai. + ne</i>
Adverbial. + particle	<i>zheng. + ne</i>

Table 3
Frequencies of endpoint encoding per language.

	Chinese [+aspect]	English [+aspect]	Swedish [–aspect]	Afrikaans [–aspect]
Number of participants	20	20	20	60
Average endpoint frequencies	35%	42.8%	61.7%	58.3%
Standard deviation	22.6%	17.2%	14.2%	17.7%
Common SD				17.84%

Note. The data on English and Swedish speakers was taken from Athanasopoulos and Bylund's (2013) study. Bylund et al.'s (2013) study provided data on Afrikaans speakers.

Instead, one respondent used the durative *zhe*. For the same test item, one-third of the respondents used the perfective *le*, and the rest did not use any aspectual markers.³ The aspectually unmarked sentences rendered imperfective or perfective readings. Note that the verb *cover* can denote either an ongoing process of snow accumulating on the land or a resultant state of the land being covered by snow. Due to the ambiguity of this verb, more contextual information was needed to determine the aspect expressed by these sentences.

5.2. Linguistic encoding task

The participants' verbalizations were audio-recorded, transcribed by a native Mandarin speaker, and coded for the frequency of endpoint focus. Any reference to an endpoint object or Goal, regardless of whether the moving entity reached the goal (e.g., *ta zoujin yige huashi* 'she has walked into a flower market) or just had the intention of approaching the goal (e.g., *ta zheng chaozhe qiche zou qu* 'she is walking to a car'), was considered as endpoint focus. However, descriptions of a reference object as a ground (e.g., *ta zai guang huashi* 'she is scrolling in the flower market') were not coded as endpoint focus. If more than one endpoint object was presented for one clip (e.g., *nvzi zouxiang yiliang che he yuanchu de shanmai* 'the girl is walking toward the car and the mountains in the distance'), only one instance of endpoint focus was recorded. A total of 240 sentences were collected, but 4 of them did not include any information about motion events (e.g., *zhege nvhaizi xiaban le* 'this girl is off duty'; *ta haoxiang yaoqu hekafei* 'it seems she is going to drink coffee.') and were thus excluded from the subsequent linguistic analysis. The analysis revealed that, on average, the L1 Mandarin speakers included endpoints in verbal descriptions 4.2 times (i.e., 35% of the 12 motion event scenes, $SD = 22.6$). A one-sample *t*-test revealed that the respondents' overall performance was significantly below the level of change (50%) ($t(19) = -2.974, p = 0.008$). For example, when describing a scene where a car approaches the gas station ahead, the respondents were more likely to omit mentioning the gas station from their descriptions than to mention it.

Given that the same linguistic task had been performed previously by native speakers of English, Swedish, and Afrikaans (Athanasopoulos and Bylund, 2013; Bylund et al., 2013), it was possible to compare the Mandarin Chinese speakers' frequency endpoint focus with those of other language groups.⁴ As presented in Table 3, the endpoint-encoding frequency in Mandarin approximated that of English but was lower than those of the two aspect languages. Using an online calculator for one-way ANOVA (Soper, 2024), the statistical analysis revealed a significant effect of language groups ($F(3,116) = 12.036, p < 0.01$). Post hoc *t*-tests showed that Mandarin speakers mentioned significantly fewer endpoints than Swedish and Afrikaans speakers (both *p* values less than 0.01). However, there was no significant difference in endpoint encoding between Mandarin and English speakers ($p = 0.227$). A post hoc power analysis was conducted with G*Power 3.1 (Faul et al., 2009) to determine the achieved power with $\alpha = 0.05$. The results revealed a large effect size ($f = 0.55$) according to Cohen's criteria (1988) and a high-power value ($1-\beta \text{ err prob} = 0.99$), suggesting a statistically robust and reliable influence of the language on the frequency of endpoint encoding.

Close examination of the frequency of endpoint encoding for each video clip, as reported in Table 4, shows that the participants referenced nearby endpoints in clips depicting a short path more frequently than endpoints situated further

³ A distinction between two uses of the word *le* usage should be made: the perfective *le*, as a grammatical suffix, and the discourse marker *le* in the sentence-final position (see Liu, 2015). Several respondents used the sentence-final *le* to relate the described event to the provided context (*when I arrived*), while *le* as a perfective marker was not used.

⁴ The criteria for participant selection in this study may differ from the selection used in the referenced studies. Therefore, further research is needed to obtain a larger and more representative sample of Mandarin speakers to support the generalization made in this study. The same consideration applies to the similarity judgment task discussed later in this paper.

Table 4
Frequencies of endpoint encoding per video clip.

long path	Endpoints in the distance	Total mentions	Frequency
Clip 2	Villages/trees	2	10%
Clip 5	Gas station/trees/building	2	10%
Clip 10	Houses/mountains	3	15%
Clip 16	A gas station/trees	4	20%
			Average = 13.8%
short path	Nearby endpoints	Total mentions	Frequency
Clip 1	A car/ a house	9	45%
Clip 4	A café	8	40%
Clip 7	An outdoor market	10	50%
Clip 8	A parking lot/car	11	55%
Clip 11	A river/trees	12	60%
Clip 13	A house/trees	7	35%
Clip 14	A slide	6	30%
Clip 17	Several shops	10	50%
			Average = 45.6%

away (e.g., almost at the end of a long path). An independent samples *t*-test confirmed a statistically significant difference in the frequency of encoding endpoints between these two types of video clips ($t(10) = 3.068$, $p = 0.012$).

5.3. Memory-based triad-matching task

The total number of times each participant paired the target clip with the [+endpoint] alternative was counted and then divided by 38. This was converted into a percentage indicating endpoint preferences. The categorization results showed that L1 Mandarin speakers selected the [+endpoint] alternative 28.3% ($SD = 13.25\%$) of the time, while L2 Swedish learners exhibited a high preference for endpoints at 33.4% ($SD = 11.64\%$) on average. A two-sample *t*-test revealed that the L2 Swedish learners did not differ significantly from the L1 Mandarin speakers in terms of their similarity judgments ($t(46) = -1.374$, $p = 0.177$). A one-sample *t*-test was conducted against chance level within each group to examine the respondents' general categorization patterns. Both groups (L1 Mandarin speakers and Chinese L2 Swedish learners) matched the target clip with the [+endpoint] alternative below the chance frequency (50%) to a statistically significant degree, ($t(24) = -8.18$, $p < 0.001$, in L1speakers; $t(19) = -6.37$, $p < 0.001$, in L2 learners). These results are consistent with previous findings that show, regardless of their language background, speakers tend to categorize goal-oriented motion events based on 'ongoingness' rather than endpoint saliency (Athanasopoulos and Bylund, 2013; Bylund and Athanasopoulos, 2015).

Like the linguistic task, the triad-matching task has been successfully completed by speakers of different languages, thus allowing for a horizontal comparison of their endpoint preferences (see Fig. 3). A one-way ANOVA with the language group as the independent variable and the percentage of selecting the [+aspect] alternative as the dependent variable yielded a significant effect regarding the respondents' language ($F(5, 106) = 4.309$, $p = 0.001$). Following this, post hoc *t*-tests were conducted between L1 Mandarin speakers and other languages. Tukey HSD post hoc tests showed that the categorization pattern of Mandarin speakers did not differ from those of English monolinguals ($p = 0.35$) but differed significantly from the categorization patterns of speakers of [-aspect] languages ($p < 0.05$ in all other cases). Compared to L1 Mandarin and L1 Swedish groups, L1 Mandarin learners of L2 Swedish exhibited a pattern distinct from both L1 groups ($p > 0.3$ in both cases).

Although no direct effect of learning an additional language on cognitive restructuring was revealed by Experiment 2, I continued to investigate whether these learners' experiences with the Swedish language might explain individual variations in endpoint preferences. A series of Pearson's correlation tests were thus performed for every Swedish language variable. A moderate but significantly positive correlation was found between the respondents' length of stay in Sweden and their endpoint preferences, suggesting that the longer the respondent had lived in Sweden, the more likely they were to choose the [+endpoint] alternative ($r = 0.46$, $p = 0.044$). To further explore their endpoint preferences according to how many years of immersion they had experienced, the L2 Swedish learners were divided into two groups: the first group had lived in Sweden for less than or equal to three years ($n = 10$), and the other had lived in Sweden for more than

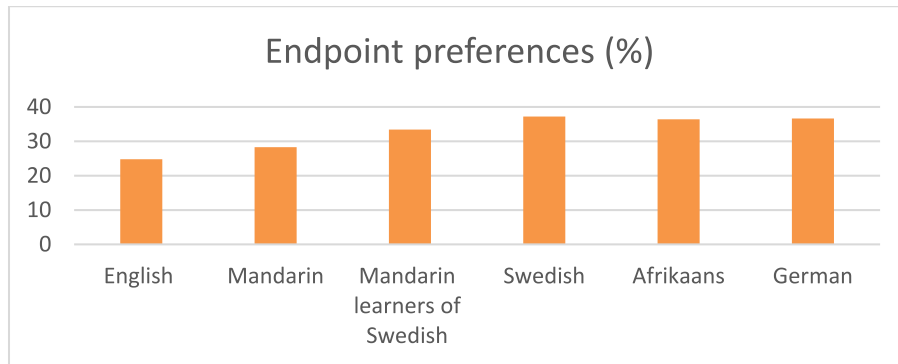


Fig. 3. Endpoint preferences (%). Note. The data on English, Swedish, and Afrikaans speakers' endpoint preferences was extracted from [Bylund et al.'s \(2013\)](#) study. The data on German speakers is based on [Athanasopoulos et al.'s \(2015\)](#) study.

three years ($n = 10$). A one-way ANOVA test conducted on the group of L1 Mandarin speakers and the two groups of L2 Swedish learners revealed a significant main effect of group ($F(2, 40) = 3.36, p = 0.045 < 0.05$). Post hoc t -tests showed that participants who had lived in Sweden for more than three years selected [+endpoint] alternatives significantly more frequently than L1 Mandarin speakers ($p = 0.026$) and more frequently than L2 learners who had less than three years' Swedish language immersion ($p = 0.030$). However, note that the L2 learners with less than three years of language immersion did not differ significantly from L1 Mandarin speakers in terms of their endpoint preferences. Similarly, proficiency in Swedish turned out to be moderately correlated with endpoint preferences and was only marginally significant ($r = 0.44, p = 0.051$). Moreover, a relatively strong and significant correlation between length of stay and Swedish proficiency ($r = 0.53, p = 0.02$) was identified, as was a moderate correlation between length of stay and frequency of use ($r = 0.42, p = 0.062$). Nonetheless, no statistically significant relationship was found between endpoint bias and other variables, including age of learning onset, frequency of L2 use, or L2 exposure.

6. GENERAL DISCUSSION

6.1. Progressive aspect from a typological perspective

The primary aim of this study was to question whether typological differences in the progressive aspect cause differences in motion event cognition. While some studies have confirmed the hypothesis that speakers of aspect languages are more likely to attend to endpoints in motion events than speakers of non-aspect languages, more recent studies have not found any replicable cross-linguistic effects of linguistic aspect on endpoint behavior. Ignoring the language skills of monolingual participants, who may not be strictly monolingual and may have acquired knowledge of both aspect and non-aspect languages, a reasonable explanation for the conflicting findings in the literature can be related to typological differences in the progressive aspect between different languages. Not every language fits neatly into the binary categories of aspect and non-aspect languages. Furthermore, some languages within one category may show a great deal of variation in their respective aspect systems. I claim that such a simple, binary classification overlooks the fact that languages display variation in the grammaticalization and routinization of how progressive aspect is expressed.

Using Mandarin Chinese as a case in point, we note that the language provides abundant grammatical markers for expressing aspect but also allows for the absence of aspect marking in sentences under certain circumstances. In contrast to English, where verbs are obligatorily marked with the morpheme *-ing* to convey progressivity, Mandarin Chinese has 'syntactically optional' aspect morphemes and sentences without these morphemes are described as maintaining a neutral viewpoint ([Smith, 1997: 277–278](#)). Mandarin speakers have to base their temporal interpretation of sentences that are unmarked for aspect on the lexical content of the sentence and contextual information. In this study, the data elicited by the PROGQ confirmed that some test sentences lacked aspect marking. Nevertheless, the semantic meaning of the verbs used in the test sentences and contextual cues help render either a perfective or imperfective interpretation of the sentence. Furthermore, other lexical means, such as discourse particles and temporal adverbials, were frequently used to express progressive meaning explicitly with or without progressive markers. Notably, the co-occurrence of the adverbial *zheng* and sentence-final modal particle *ne*, typically found only in informal conversations, plays a similar role as the progressive *zai* in expressing ongoing activities. The adverbial *zheng* also plays a role in con-

veying progressivity, even though it is neither a grammatical marker nor a temporal adverbial. As the literal meaning of *zheng* 'just' suggests, *zheng* pinpoints a given specific time point when an event takes place. This function differs from the progressive *zai*, which denotes a time interval when a durative situation unfolds.

The progressive *zai* represents a highly grammaticized marker for progressivity in Mandarin Chinese. While the progressive *zai* is frequently used in everyday speech, lexical means such as the adverb *zheng*, the particle *ne*, or temporal adverbials, along with pragmatic factors like contextual cues, allow for the omission of the progressive marker. In comparison, English, as a well-known aspect language, has only one verb form (BE + *-ing*) that encodes progressivity. The lack of alternative forms to express progressivity in English is considered to necessitate the obligatory marking of verbs with the morpheme *-ing* in various contexts of use (Traugott, 2003). Despite different degrees of obligatoriness in using progressive aspect in English and Mandarin Chinese, the pre-verbal progressive *zai* serves a function similar to the morpheme *-ing* in English. The results from both the verbal and non-verbal tasks showed that L1 Mandarin speakers exhibited endpoint behavior similar to English speakers while also being significantly different from speakers of Swedish and Afrikaans, indicating a linguistic effect of progressive markers on endpoint behavior. In the field of cognitive linguistics, progressive markers, defined as a cognitively salient category, impose a restricted scope on verbs and direct the speaker's attention toward the ongoing phase of an event (Langacker, 2000; Kermer, 2020). From this perspective, it is argued that speakers are used to frequent combinations of grammaticized markers of progressivity and a minimal temporal viewing frame with event endpoints excluded. Similarly, lexical means for expressing progressivity, such as the adverb *zheng*, may also activate a minimal viewing frame in the speaker's mind. However, these lexical means in Mandarin are not considered the primary reason why Mandarin speakers' endpoint behavior differs from speakers of non-aspect languages. While speakers of non-aspect languages lack highly grammaticized progressive markers and often leave a sentence's progressivity implicit, they still can express progressivity explicitly by various lexical means, including adverbs and particles (e.g., *just nu* 'right now' in Swedish, *gerade* 'right now' in German). In this regard, the crucial cross-linguistic difference between speakers of aspect languages and non-aspect languages lies solely in the grammatical means for expressing progressivity, which is associated with the entrenchment of a minimal viewing frame.

Variations in periphrastic verbal constructions used to express progressivity have been identified in non-aspect languages. First, it is crucial to keep in mind that while non-aspect languages lack fully grammaticalized morphological markers for progressive aspect, this does not necessarily mean that they do not have any other grammaticalized ways of conveying progressivity. Take the Germanic languages as an example: prepositional constructions with different degrees of grammaticalization are available for encoding progressive aspect such as *am/beim* (lit., 'in/at the') in German, and *aan het* (lit., 'at the') in Dutch. The *am/beim*-form in German displays a very low degree of grammaticalization since it only appears in dialectal and informal speech (Andersson, 1989; Mair, 2012). The *aan het* construction in Dutch has become "highly sensitive to situations with a progressive component", and its use has gradually expanded from a 'here-and-now' context into a past-tense context (Flecken, 2011: 509). This suggests that *aan het* has undergone grammaticalization, exhibiting a low-to-medium level of grammaticalization and routinization. The absence of endpoint bias in [-aspect] Dutch speakers, as recorded in Liao et al.'s (2020) study, can be associated with the frequent usage of *aan het* in describing ongoing activities. Like other progressive markers, the *aan het* construction can invoke a progressive viewpoint with an emphasis on an event's 'ongoingness' instead of the entirety of the event. Although the *aan het* construction is far from attaining the same level of grammaticalization as the English BE + *V-ing* construction, its expanded use in a wide range of semantic contexts implies that Dutch speakers may be accustomed to both an immediate temporal viewing frame and a maximal frame that includes an event's totality. As illustrated above in the German and Dutch examples, the degree of grammaticalization varies across different progressive constructions in non-aspect languages, thereby influencing their frequency of use and the degree to which an immediate viewing frame is entrenched in the speaker's mind. This variation may explain inconsistent findings regarding endpoint behavior among speakers of non-aspect languages.

6.2. Ongoingness versus endpoint saliency

What does the endpoint behavior of Mandarin speakers look like? The results of the linguistic re-telling task suggest that Mandarin speakers tend to exclude endpoints in their verbal descriptions of ongoing motion events. Compared with speakers of different languages in studies using the same speech elicitation stimuli, Mandarin speakers were found to omit mentioning endpoints as frequently as English speakers. However, they also differed significantly from speakers of Swedish and Afrikaans, non-aspect languages, thereby indicating an effect of grammatical aspect on event construal at a linguistic level. This observation lends further empirical support to previous findings that show speakers of non-aspect languages are more likely to mention event endpoints than speakers of aspect languages (e.g., von Stutterheim and Nüse, 2003; Athanasopoulos and Bylund, 2013). Although Mandarin speakers showed a similar frequency as English speakers in terms of endpoint encoding, Mandarin speakers differ somewhat in terms of the frequencies with which they

use aspect markers. When describing the same video clips, English speakers used the present progressive construction (BE + V-ing) in all their utterances (Athanasopoulos and Bylund, 2013: 294), whereas Mandarin Chinese speakers switch between perfective and imperfective markers. Considering the optional usage of aspect markers in Mandarin Chinese, it was not unexpected that speakers of these two aspect languages vary in how often grammatical means were employed to convey 'ongoingness'.

Regarding the degree to which different aspectual markers were used in combination with endpoint encoding, I found that the progressive marker *zai* was frequently attached to verbs without any reference to an endpoint, while the perfective marker *le* was accompanied by descriptions of endpoints in nearly half of the cases. This result indicates that the speakers construed the scenes differently when using different aspect markers. When speakers were about to verbalize an ongoing dynamic motion event they had just seen, they paid attention to the motion in progress, such as walking and running at the expense of the components of the motion event, including its endpoint. This was the case, presumably, because the progressive marker highlighted 'ongoingness' and the speakers tended to zoom in on the moving figure and its action. In contrast, some speakers viewed the motion event in its entirety and selected a perfective marker when describing events they interpreted as completed. As reflected by the co-occurrence of *le* and the mention of an endpoint in their verbal descriptions, the perfective viewpoint required that participants take a maximal perspective on motion events, which gave rise to the verbal encoding of endpoints as an established final boundary of the event. This pattern in the use of aspect in Mandarin Chinese agrees with previous accounts of aspect (Comrie, 1997; Langacker, 2000, 2008).

Given that the task elicited the participants' verbal descriptions of motion events under an online condition, the findings from Experiment 1 were compatible with Slobin's (1996, 2003) 'thinking-for-speaking' framework. At the stage of formulating speech content, speakers are likely to consider the role of grammatical aspect in their linguistic system and conceptualize certain elements of motion events after selecting one type of event interpretation of the scene, for example, a reached or upcoming goal from a perfective viewpoint. When transforming their thought content into speech, the Mandarin speakers explicitly encoded the phases of the events they viewed in a way that was relevant to the aspectual marker used; for example, an intermediate ongoing phase would be associated with the progressive marker *zai*, or a whole situation, including its final points, would be expressed by using *le*. Consequently, I present two primary probabilistic cues, namely, (i) the progressive aspect is linked to infrequent endpoint encoding, and (ii) the perfective aspect is related to frequent mentions of endpoints. The relationship between grammatical aspect and endpoint encoding is thus not absolute but, instead, a matter of probability, which means that the possibility of mentioning endpoints with progressive markers without any aspect markers is not excluded.

The experiment participants generally did not show a strong predilection for encoding endpoints in their linguistic descriptions. However, they frequently did include nearby endpoints along a relatively short path. In similar linguistic encodings of real-life motion events, Mandarin and Dutch speakers referred to reached endpoints frequently when presented with scenes showing a high degree of goal orientation, regardless of their language background (Liao et al., 2020). Other empirical studies have shown that speakers are highly likely to verbally report reached goals because such goals indicate qualitative changes from one location into another and are thus highly salient as right-hand boundaries of motion events (Croft, 2012; von Stutterheim et al., 2012). Although the stimuli in our verbal task only showed an intermediate degree of goal orientation with endpoints visible but unreached, the participants construed nearby endpoints as reached or about to be reached. The contrast between speakers' mentions of nearby endpoints and endpoints in the distance is a function of the perceptual characteristics of path length. According to the theory of affordance (Gibson, 2014), a short path "affords" an Agent's arrival at a place. As the core schema of motion events, the path represents a salient environmental feature of motion events. While preparing their verbal responses, the participants likely noticed the path of the depicted motion event and perceived a short path as a necessary route toward the only goal ahead. On the other hand, a long extended path does not promise one's arrival at a destination since it provides a broader perspective, including one's current location, upcoming midpoints, and distant endpoints. As a result, the participants were more likely to perceive highly evident endpoints situated along a short path as either reached or about to be reached, and they mentioned such endpoints more frequently in their verbal descriptions than potential endpoints in the distance.

The effect of perceptually salient endpoints on motion event cognition was also evident in the non-linguistic task. When asked to pair the target scene with [-endpoint] or [+endpoint] alternatives, both groups selected more [-endpoint] alternatives than [+endpoint] alternatives, suggesting an overall bias toward engaging in low-degree goal orientation regardless of language background. As the designers of the triad-matching task have explained, the perceptual differences between two alternatives, the target scene and [-endpoint] alternatives, depict one-state and unbounded motion events, whereas the endpoint-reached motion events in [+endpoint] alternatives are bounded and involve a change of state (Athanasopoulos and Bylund, 2013: 301). From this perspective, the differences between target scenes and [+endpoint] alternatives are more distinct than those between target scenes and [-endpoint] alternatives. Consequently, [-endpoint] alternatives are more likely to be judged similar to target scenes than [+endpoint] alternatives. Thus

far, two main factors have been discussed that give rise to differences in linguistic and memory-based behavior regarding the 'ongoingness' and endpoints of motion events, namely, grammatical means of conveying aspectual contrasts in the native language and perceptual characteristics of scenes shown in video clips.

In addition to the above observations, it is worth noting whether culture also affects endpoint behavior. The study's participants are native Mandarin speakers from an Eastern cultural background, whereas the speakers of other languages (e.g., English, German, and Swedish) in previous investigations primarily represent Western cultures. Cross-cultural psychology studies have shown that Easterners have a holistic thinking style and pay more attention to context and background information, while Westerners are analytic thinkers, who tend to focus on a central object in isolation from its context (Nisbett, 2001; Miyamoto and Wilken, 2013). In a recent study on motion event construal, Park et al. (2022) discovered a cross-cultural effect where English speakers wrote down more focal information (i.e., direct references to the entity's movement, including its motion verb, path, and manner) but less peripheral information (information irrelevant to the entity's movement, including the entity's appearance and attributes, context) about motion events than Korean speakers. Participants were explicitly instructed to focus only on the motion event in the linguistic description task. Therefore, their verbal descriptions centered on the 'focal information' about the moving entity. Suppose the participants exhibited a holistic thinking pattern as suggested by cross-cultural studies. In that case, they should be expected to mention endpoints more frequently than Westerners because (it has been argued) holistic thinkers typically view an event in its entirety. However, the results of this study indicate that the Chinese Mandarin speakers did not encode endpoints as frequently as Western speakers of non-aspect languages. This raises an important question: Considering that the path of motion is the core schema of motion events,⁵ and the endpoint is the goal of the path (Talmy, 2000), is endpoint information treated as focal information in motion event construal compared to other types of focal information such as manner of motion? A tentative answer to this question would be negative. The endpoint may be a salient feature of a motion event but is not necessarily the central part of a motion event. Further research could employ eye-tracking experiments or neurolinguistic approaches to examine how speakers allocate their attention to the various components of motion events in real-time and what components can be counted as focal or most salient to speakers of different languages and cultures. Future research could also include comparisons between languages with varying degrees of grammaticalization in progressive aspect and between typologically similar languages from different cultural groups.

6.3. Additional language learning and cognitive restructuring

In support of Slobin's (1996, 2003) 'thinking-for-speaking' hypothesis, the linguistic encoding task reported on in this study provides evidence that the Mandarin speakers took a grammaticized view of visual motion events and attended to relevant elements when they used aspectual markers in their verbal descriptions. Going beyond the linguistic level, the triad-matching task investigated whether the presence or absence of specific grammatical categories, such as progressive aspect, exerted any long-term effects on motion event cognition as measured in a categorization task and whether learning a non-aspect language affects how learners think about motion events. At first glance, no significant differences in categorization patterns were documented between the L1 Mandarin speakers and L2 Swedish learners. However, the possibility of cognitive restructuring was not altogether excluded in the L2 Swedish learners. A correlational analysis revealed that their bias toward [+endpoint] alternatives was positively correlated with how many years of immersion in the target-language setting they had experienced. In other words, the percentage of selecting [+endpoint] alternatives increases as a function of the participant's length of stay in a Swedish-speaking environment.

Once immersed in the Swedish language environment, L2 learners are constantly exposed to Swedish typical usage patterns, where event endpoints are frequently encoded in descriptions of ongoing situations but without grammatical markers of progressivity. From a cognitive grammar perspective (Langacker, 2000, 2008), these language learners would abstract a maximal temporal viewing frame from relevant linguistic cues, i.e., from frequent mentions of endpoints. Increased immersion in an L2 setting offers the learner more exposure to this usage pattern and more opportunities to apply the corresponding time schema to the construal of similar situations. When sufficient repeated associations between the absence of progressive markers and the inclusion of endpoints are internalized as part of a cognitive routine, learners can overcome their L1-related bias toward a restricted viewing frame. Instead, they can adopt a strengthened viewing frame as a cue for making judgments in the categorization task. In addition, it should also

⁵ Languages across the world differ in encoding manner and path of motion events. According to Talmy's (1985, 2000) motion-event typology, *satellite-framed* languages express the path of motion via a preposition or verb particle while *verb-framed* languages express the path of motion via the verb root itself. Encoding manner and path and encoding grammatical aspect are two major avenues in current linguistic relativity research in the domain of motion events that are often studied separately. However, it would be beneficial for future research to integrate these two lines of inquiry to explore their combined effect on motion event cognition.

be noted that the L2 learners who participated in this study do not live in a purely Swedish monolingual environment but, instead, live in a multilingual context where they might communicate with their families in L1, work with their colleagues in L2, or use both languages in their Swedish language classes. In everyday communicative practices, L2 learners can draw on all linguistic resources from all the languages they know to make meaning and sense—a dynamic process known as ‘translanguaging’ (Baker, 2001; García and Li, 2014). However, their ability to innovate and adapt their language use while learning or using a second language may also hinder the complete internalization of a standard L2-specific cognitive framework. In this study, the L2 learners, as a group, exhibited a cognitive pattern that was neither entirely like their L1 nor their L2. This finding suggests that L2 learners undergo a cognitive shift that cannot be explained by any single language or a combination of separate languages (Li, 2024). With this in mind, the dynamic integration of multiple languages could be an additional factor influencing L2 learners’ cognitive process.

Although the L2 Swedish learners did not differ significantly from the L1 Mandarin and L1 Swedish speakers with regard to average endpoint preferences, the positive correlation between length of stay in an L2 environment and endpoint preferences provides indirect evidence of a developing cognitive pattern from progress-salience in the [+aspect] L1 to endpoint-salience in the [-aspect] L2. This finding resonates with the categorization pattern of L1 English learners of L2 German, as documented in Athanasopoulos et al.’s (2015) study, which used the same experimental paradigm as in this study. Athanasopoulos et al. confirmed that the lack of progressive aspect in L2 (German) affected motion event cognition in bilinguals who had acquired grammatical means to express ‘ongoingness’ in their L1. Cognitive restructuring in a specific domain might occur even if no relevant grammatical categories (relevant to that domain) are involved in learning an additional language. As Athanasopoulos et al. (2015: 148) explain, the lack of grammatical aspect in L2 German brought about an interpretation of events as bounded by their participants, and the weight of unboundedness as a categorization cue from L1 underwent readjustment as a function of their participants’ L2 proficiency and duration of study. What matters in cognitive restructuring is the weightings of specific conceptual categories, such as the boundedness and unboundedness of an ongoing motion event. These cognitive representations of goal-oriented motion are associated not only with grammatical means to express progressivity but also with the absence of grammatical markers.

Furthermore, the findings of this study also reveal that proficiency contributed to the learners’ higher tendency to focus on endpoints when making judgments. The L2 learners with a short period of language immersion came to Sweden from China for their graduate studies, whereas those with more extended periods of immersion had settled in Sweden and gradually integrated into Swedish society and culture. Considering such different backgrounds, the learners were divided into two groups. This division revealed a significant group effect: compared to L1 Mandarin speakers and L2 Swedish learners with a relatively short period of residence, the learners who had lived in Sweden for more than three years were more likely to select endpoints as a basis for categorization. This observation supports the idea that motion event cognition can be recalibrated through increased experience with the target language. In our samples, the participants who had lived in Sweden for a relatively short period (less than three years) still produced the pattern of L1 speakers, while those who had lived in Sweden for more than three years had shifted their cognitive preferences away from L1. The effect of length of immersion in a target language environment on cognitive restructuring has been documented in other perceptual domains. For example, Athanasopoulos et al.’s (2010) study of color perception investigated whether L1 Greek learners of L2 English altered their previously formed representations of light blue (*ghalazio*) and dark blue (*ble*) after sufficient cultural exposure to an L2-speaking environment where no lexical distinction between light blue and dark blue exists. The results from an oddball task measuring electrophysiological responses, a similarity judgment task, and a color-naming task all showed that long-stay bilinguals (range: 18–60 months), like English monolinguals, distinguished between their native blue contrasts (*ghalazio/ble*) to a significantly lesser extent than short-stay bilinguals (range: 5–12 months). In summary, Greek-English bilinguals exhibited a significant shift toward target-like pre-attentive color processing after at least 1.5 years of residence in the UK. In contrast, in the present study, the minimal period of L2 immersion needed for cognitive restructuring was three years. There thus appears to be no consensus over how long a person’s residence in an L2 environment should be if it is to influence bilinguals’ cognitive representations. Nonetheless, it is apparent that bilinguals living in a target language setting have more opportunities to associate new linguistic cues with relevant conceptual categories than learners who are exposed to a formal language classroom context only.

The duration of a learner’s stay in an L2 context is a broad variable that subsumes several related factors. Specifically, a period of cumulative experience in an L2 context usually involves frequent exposure to L2, decreased proficiency in L1, increased proficiency in L2, and greater cultural immersion, where L2 learners gradually learn and adopt cultural thinking patterns of the target language community. Given this, the role of length of stay in an L2 environment in cognitive restructuring has practical implications for language learning. This study presents a relatively unique case of language acquisition: L2 learners who possess a rich aspectual system in their native languages acquire an L2 that lacks grammatical markers to encode ‘ongoingness’. Instead of acquiring new complicated aspectual forms

and rules, they only need to use the simple form of the verb. During this process of learning and adaptation, cognitive resources might be saved from mastering new forms and understanding their grammaticalized meanings and be reallocated to observing other sentence elements, such as the co-occurrence of the simple form of a manner verb and an event endpoint. As a result, the longer an L2 learner stays in an L2 linguistic and cultural context, the stronger the association becomes between the simple verb form and boundedness of an ongoing event, and the more likely the learner is to implicitly adopt a maximal viewing frame similar to that of native speakers of the target language.

Last but not least, this study did not attest to the effects of other L2 factors, such as the age of learning onset, frequency of L2 use, and language exposure. As to their age when they began to learn Swedish, the participants in this study were somewhat disadvantaged because they started learning Swedish after the age of 14, i.e., they began learning an additional language after they had acquired their native language. Learning a language under this condition makes it more difficult for learners to display a convergent or L2-like cognitive pattern than early bilinguals. Furthermore, the lack of observed effects of use frequency and language exposure can be attributed to the self-rating measure, a somewhat coarse method of operationalizing these factors in this study. Future investigations of L2 learners' cognitive restructuring would do well to use more nuanced measures and collect longitudinal data on both L1 and L2 to track the development of learners' cognitive patterns over time.

7. CONCLUSION

This study has revisited a controversial hypothesis that claims speakers of non-aspect languages are more likely to show a preference for endpoints of motion events than speakers of aspect languages. This hypothesis was tested by including new groups of participants: L1 Mandarin speakers and L2 Swedish learners. Bearing in mind that maintaining a strict binary distinction between aspect and non-aspect languages is problematic and cannot capture every typological difference regarding progressive aspect across every language, the study first examined the optional and flexible usage of progressive markers in [+aspect] Mandarin through an adapted questionnaire. The mini corpus that was collected showed that *zai* serves as a prominent progressive marker that is fully grammaticalized. Other lexical means, including the adverb *zheng* and colloquial patterns ending with the particle *ne*, are also used to express progressivity. Furthermore, it was observed that the flexible selection between the progressive *zai* and the perfective *le* depends on the semantic context and the speaker's aspectual viewpoint.

The results of the two experiments showed that Mandarin speakers verbally encoded endpoints and selected the alternative with a low degree of goal orientation to a similar extent as English speakers but significantly less than speakers of non-aspect languages. These findings provide evidence for the hypothesis regarding the linguistic effect of the progressive aspect on endpoint behavior. Similarly, a comparison between L1 speakers and L2 learners revealed that L2 learners exhibited a categorization pattern between L1 and L2, suggesting cognitive restructuring had taken place. To conclude, linguistic effects were attested in linguistic and cognitive tasks across different language groups. However, it is essential to highlight that these linguistic cues, including progressive markers, cannot outweigh specific perceptual properties of motion events, such as length of path and endpoint saliency. For example, an endpoint defines the boundary of motion events and is a salient feature that guides a person's perception of motion events. A reached endpoint is even more salient as it involves a qualitative change from one location into another and a change from a state of activity to a state of completion. In summary, the way speakers attend to a motion event endpoint is influenced by various factors. While grammatical tools for conveying progressivity are indeed a factor, they are absolutely not the most crucial factor with regard to the perceptual and cognitive processing of motion events.

DECLARATION OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Qiu-Jun Zhang: Writing – review & editing, Writing – original draft, Visualization, Project administration, Investigation, Formal analysis, Data curation, Conceptualization.

Data availability

Data will be made available on request.

ACKNOWLEDGEMENTS

I am grateful to Emmanuel Bylund and Panos Athanasopoulos for generously providing their experimental stimuli. I would also like to express my sincere thanks to Maria Rosenberg and Marlene Johansson Falck for their helpful comments on earlier versions of the manuscript.

APPENDIX A AND OTHER SUPPLEMENTARY MATERIAL

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.lingua.2024.103827>.

REFERENCES

- Andersson, S.-G., 1989. On the generalization of progressive constructions: "Ich bin (das Buch) am Lesen" – Status and usage in three varieties of German. In: Larsson, L.-G. (Ed.), *Proceedings of the Second Scandinavian Symposium on Aspectology*. Almqvist & Wiksell, pp. 95–106.
- Athanasopoulos, P., Bylund, E., 2013. Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognit. Sci.* 37 (2), 286–309.
- Athanasopoulos, P., Dering, B., Wiggett, A., Kuipers, J.R., Thierry, G., 2010. Perceptual shift in bilingualism: Brain potentials reveal plasticity in pre-attentive colour perception. *Cognition* 116 (3), 437–443.
- Athanasopoulos, P., Damjanovic, L., Burnand, J., Bylund, E., 2015. Learning to think in a second language: Effects of proficiency and length of exposure in English learners of German. *The Modern Language Journal* 99 (S1), 138–153.
- Baker, C., 2001. Foundations of bilingual education and bilingualism. *Multilingual Matters*.
- Bertinetto, P., Ebert, K., de Groot, C., 2000. The progressive in Europe. In: Dahl, Ö. (Ed.), *Tense and Aspect in the Languages of Europe*. De Gruyter Mouton, pp. 517–558.
- Blensienius, K., 2015. Progressive constructions in Swedish. Gothenburg University, PhD thesis.
- Boroditsky, L., 2001. Does language shape thought? Mandarin and English speakers' conceptions of time. *Cogn. Psychol.* 43 (1), 1–22.
- Bylund, E., Athanasopoulos, P., Oostendorp, M., 2013. Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics* 51 (5), 929–955.
- Bylund, E., Athanasopoulos, P., 2015. Televised Whorf: Cognitive restructuring in advanced foreign language learners as a function of audio-visual media exposure. *Mod. Lang. J.* 99 (S1), 123–137.
- Bylund, E., Jarvis, S., 2011. L2 effects on L1 event conceptualization. *Biling. Lang. Cogn.* 14, 47–59.
- Casasanto, D., Boroditsky, L., 2008. Time in the mind: Using space to think about time. *Cognition* 106 (2), 579–593.
- Casati, R., Varzi, A., 2008. Event concepts. In: Shipley, T., Zacks, J. (Eds.), *Understanding Events*. Oxford University Press, pp. 31–55.
- Chung, S., Timberlake, A., 1985. Aspect, tense, mood. In: Shopen, T. (Ed.), *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon*, Vol. 3. Cambridge University Press, pp. 202–258.
- Cohen, J., 1988. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Comrie, B., 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge University Press.
- Croft, W., 2012. *Verbs: Aspect and causal structure*. Oxford University Press.
- Faul, F., Erdfelder, E., Buchner, A., Lang, A.-G., 2009. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Res. Methods* 41 (4), 1149–1160.
- Feinmann, D., 2020. Language and thought in the motion domain: Methodological considerations and new empirical evidence. *J. Psycholinguist. Res.* 49 (1), 1–29.
- Filipović, L., 2018. Speaking in a second language but thinking in the first language: Language-specific effects on memory for causation events in English and Spanish. *Int. J. Biling.* 22 (2), 180–198.
- Flecken, M., 2011. What native speaker judgments tell us about the grammaticalization of a progressive aspectual marker in Dutch. *Linguistics: Interdiscipl. J. Lang. Sci.* 49 (3), 479–524.
- Fuhrman, O., McCormick, K., Chen, E., Jiang, H., Shu, D., Mao, S., Boroditsky, L., 2011. How linguistic and cultural forces shape conceptions of time: English and Mandarin time in 3D. *Cognit. Sci.* 35 (7), 1305–1328.
- García, O., Li, W., 2014. *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.
- Gibson, J.J., 2014. *The ecological approach to visual perception*. Psychology Press.
- Hendriks, H., Hickmann, M., 2015. Finding one's path into another language: On the expression of a boundary crossing by English learners of French. *The Modern Language Journal* 99 (S1), 14–31.
- Hinchliffe, I., Holmes, P., 2008. *Swedish: An essential grammar*. Routledge.
- Ji, Y., 2017. Motion event similarity judgments in one or two languages: An exploration of monolingual speakers of English and Chinese vs. L2 learners of English. *Front. Psychol.* 8, 909.
- Ji, Y., Hohenstein, J., 2017. Conceptualizing voluntary motion events beyond language use: A comparison of English and Chinese speakers' similarity judgments. *Lingua* 195, 57–71.

- Kermer, F., 2020. A cognitive grammar perspective on temporal conceptualization in SLA. *Stud. Univ. Babeş-Bolyai Inform. Anglica Posnaniensia* 55, 223–246.
- Klein, W., 2018. Looking at language. De Gruyter Mouton.
- Klein, W., Li, P., Hendriks, H., 2000. Aspect and assertion in Mandarin Chinese. *Nat. Lang. Linguist. Theory* 18 (4), 723–770.
- Kvist-Damell, U., 2005. Utvidgade och andra användningar av de svenska verben sitta, stå, och ligga. In: Byrman, G., Einarsson, J., Hammarbäck, S., Lindgren, M., Stille, P. (Eds.), *Svenskans Beskrivning 27*. Växjö University Press, pp. 148–160.
- Langacker, R.W., 2000. Grammar and conceptualization. De Gruyter Mouton.
- Langacker, R.W., 2008. *Cognitive grammar: A basic introduction*. Oxford University Press.
- Lewandowski, W., Özçalışkan, Ş., 2022. How language type influences patterns of motion expression in bilingual speakers. *Second. Lang. Res.* 37 (1), 27–49.
- Li, W., 2024. Translanguaging and intercultural communication: Rethinking cultural thought patterns. In: Murata, K. (Ed.), *ELF and Applied Linguistics*. Routledge, pp. 68–96.
- Li, C., Thompson, S.A., 1981. *Mandarin Chinese: A functional reference grammar*. University of California Press.
- Liao, Y., Flecken, M., Dijkstra, K., Zwaan, R.A., 2020. Going places in Dutch and Mandarin Chinese: Conceptualizing the path of motion cross-linguistically. *Language, Cognition and Neuroscience* 35 (4), 498–520.
- Liu, M., 2015. Tense and aspect in Mandarin Chinese. In: Wang, W., Sun, C. (Eds.), *The Oxford Handbook of Chinese Linguistics*. Oxford University Press, pp. 189–205.
- Mair, C., 2012. Progressive and continuous aspect. In: Binnick, R. (Ed.), *The Oxford Handbook of Tense and Aspect*. Oxford University Press, pp. 803–826.
- Marian, V., Blumenfeld, H.K., Kaushanskaya, M., 2007. The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *J. Speech Lang. Hear. Res.* 50 (4), 940–967.
- Misersky, J., Peeters, D., Flecken, M., 2022. The potential of immersive virtual reality for the study of event perception. *Frontiers in Virtual Reality* 3, 697934.
- Miyamoto, Y., Wilken, B., 2013. Cultural differences and their mechanisms. In: Reisberg, D. (Ed.), *The Oxford Handbook of Cognitive Psychology*. Oxford University Press, pp. 970–985.
- Montero-Melis, G., Jaeger, T.F., Bylund, E., 2016. Thinking is modulated by recent linguistic experience: Second language priming affects perceived event similarity. *Lang. Learn.* 66 (3), 636–665.
- Nisbett, R.E., Peng, K., Choi, I., Norenzayan, A., 2001. Culture and systems of thought: Holistic versus analytic cognition. *Psychol. Rev.* 108 (2), 291–310.
- Nosofsky, R.M., 1986. Attention, similarity, and the identification-categorization relationship. *J. Exp. Psychol. Gen.* 115 (1), 39–57.
- Papafragou, A., Hulbert, J., Trueswell, J., 2008. Does language guide event perception? Evidence from Eye Movements. *Cognition* 108 (1), 155–184.
- Park, I.H., Jarvis, S., Kim, J.-E., 2022. Exploring motion event construal: How much attention do speakers of different languages and cultures pay to context?. *Lingua* 165, 103164.
- Pavlenko, A., 2011. Thinking and speaking in two languages: An overview of the field. In: Pavlenko, A. (Ed.), *Thinking and Speaking in Two Languages*. Multilingual Matters, pp. 237–257.
- Simmons, R.V., 1999. Chinese dialect classification: A comparative approach to Hangzhou, Old Jintarn, and Common Northern Wu. In: Koerner, E.F.K. (Ed.), *Current Issues in Linguistic Theory*, Vol. 188. John Benjamins Publishing Company, pp. 239–260.
- Slobin, D.I., 1996. From “thought and language” to “thinking for speaking”. In: Gumperz, J.J., Levinson, S.C. (Eds.), *Rethinking Linguistic Relativity*. Cambridge University Press, pp. 70–96.
- Slobin, D.I., 2003. Language and thought online: Cognitive consequences of linguistic relativity. In: Gentner, D., Goldin-Meadow, S. (Eds.), *Language in Mind: Advances in the Investigation of Language and Thought*. MIT Press, pp. 157–191.
- Smith, C.S., 1997. *The parameter of aspect*. Springer.
- Soper, D.S. 2024. Analysis of variance (ANOVA) calculator - One-way ANOVA from summary data. <https://www.danielsoper.com/statcalc/calculator.aspx?id=43>.
- Talmy, L., 1985. Lexicalization patterns: Semantic structure in lexical forms. In: Shopen, T. (Ed.), *Language Typology and Syntactic Description*, Vol. 3. Cambridge University Press, pp. 36–149.
- Talmy, L., 1991. Path to realization: A typology of event conflation. In: *Proceedings of the Annual Meeting of the Berkeley Linguistics Society*, pp. 480–519.
- Talmy, L., 2000. *Toward a Cognitive Semantics: Concept Structuring Systems*, Vol. 1. MIT Press.
- Traugott, E.C., 2003. Constructions in grammaticalization. In: Joseph, B.D., Janda, R.D. (Eds.), *The Handbook of Historical Linguistics*. Blackwell, pp. 624–647.
- Vendler, Z., 1957. Verbs and times. *The Philosophical Review* 66 (2), 143–160.
- Von Stutterheim, C., Nüse, R., 2003. Processes of conceptualization in language production: Language-specific perspectives and event construal. *Linguistics* 41 (5), 851–881.
- Von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., Schmiedtová, B., 2012. How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analyses, eye tracking data, and memory performance. *Linguistics* 50 (4), 833–867.

- Weng, J., 2018. What is Mandarin? The social project of language standardization in early Republican China. *The Journal of Asian Studies* 59 (1), 611–633.
- Whorf, B.L., 1956. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. MIT Press.
- Wolff, P., Holmes, K.J., 2010. Linguistic relativity. *Wiley Interdiscip. Rev. Cogn. Sci.* 2 (3), 253–265.
- Xiao, R., McEnery, T., 2004. *Aspect in Chinese: A corpus-based study*. John Benjamins Publishing Company.