Expressions of Motion Events by German, Chinese, and English Native Speakers and German Learners of Chinese and English

A Dissertation
Submitted to the Faculty of Modern Languages
of the Ruprecht-Karls-Universität Heidelberg
for the Degree
Doctor of Philosophy

Can Zhang
Heidelberg, Sept. 24, 2020

Advisors: Prof. Dr. Christiane von Stutterheim
Prof. Dr. Wolfgang Klein
PD Dr. Sandra Pappert

Acknowledgements

During my journey to get a Ph.D. degree, there are many people who have inspired and helped me along the way. I want to express my gratitude to them for their support and guidance.

First and foremost, I want to express my deepest gratitude to Prof. Christiane von Stutterheim for taking me as her PhD student, so I had a chance to further my study. I was super fortunate to have her as my advisor. I enjoyed every lecture, workshop, and colloquium she held. I benefited a lot from her profound knowledge and exceptional vision of linguistics. She carefully read each line of my earlier drafts and gave me insightful comments. Without her challenging and productive discussions with me, my dissertation would not have reached this stage. Whenever I had difficulties, she used her 'magic power' to make the rather difficult thing quickly understood and solved.

I want to thank Prof. Wolfgang Klein for his interests in my work and Dr. Sandra Pappert for her critical feedback, which made a considerable contribution to my work. My heart-felt thanks also go to Prof. Giulio Pagonis and Prof. Gertrud Rösch for attending my oral defense. I am grateful for the insightful questions they raised. I want to thank Prof. Gertrud Rösch for chairing my oral defense.

My heart-felt thanks also go to Dr. Johannes Gerwien. Without his support my experiment would not have had a good start. I benefited greatly from the productive discussions with him. His insightful comments and critiques of some examples enriched the content of my dissertation.

I would like to thank Abbassia Bouhaous for showing me the video clips she used for the experiment and giving me valuable advice about my experiment. I am particularly indebted to Takara Baumbach. Whenever I had technical problems during my experiment, he was always ready to help. With the help of the scripts he wrote I could pace up the analysis of my experiment.

I would like to extend my sincere thanks to the participants who attended my experiment. All participants showed their enthusiasm and support for my study. Recruiting participants to attend the experiment was difficult, while finding participants who were advanced German learners of Chinese was even more difficult. I want to thank Mr. Gu Wen for allowing me to go to his class to recruit German learners of Chinese. I would also like to express my deepest gratitude to the academic staff and

students in the Department of Chinese Studies at the University of Heidelberg, among whom are Egas, Mariana, Marina, Emily Mae, Stephan, Kaja, Milena, Giulia, Sara, Rafael, Rudi, Jakob, Sophia, Chris, Joachim, Neil, Valentin, Michelle, Franziska, Odila, Fabienne, Becker, Lukas, Duncan, Sophie, Stanley. Some of them asked their friends to attend my experiment after they themselves finished the experiment. Some of them even took the train to Heidelberg to attend my experiment. I was so moved by them. My appreciation for them is much deeper than the words can express. Without them my dissertation would never have been completed.

I would also wish to express my gratitude to Lara Schmidt and Alexandra Schatz for their help to check the German transcriptions and Alyssa Patterson for the English transcriptions.

Last but not the least I am tremendously indebted to my parents and my parents-in-law. My parents always have faith in me. They have encouraged me to follow my heart. I am deeply grateful for my parents-in-law for their unconditional support and taking care of my daughter. My sincere and profound thanks also go to my sisters, for their constant spiritual support. Their positive attitudes have sustained me through the most difficult time. I also want to thank them to take care of our parents while I was away over these years.

I would like to express my sincere thanks to my husband for his understanding, encouragement, and support. I have lost my way at some point and he has helped me to refind my own path. I want to thank my lovely daughter. To watch her growing up has brought so much happiness to my life.

Abstract

Over the past few decades, the field of motion events has received much attention and has been studied extensively. Much of this research is founded on the typological contrasts between verb-, satellite- and equipollently-framed languages (Talmy, 1985, 2000; Slobin, 2004, 2006). This study goes deeper into this typology by looking at the encoded spatial and aspectual concepts as well as the interaction between those two categories in L1s and L2s, with the purpose of finding out whether L1 language-specific differences have an impact upon motion event conceptualization in L2s. The L2 learners in this analysis were advanced L1 German-L2 Chinese and L1 German-L2 English speakers. All speakers (including L1 speakers) watched the same video stimuli showing motion events with different degrees of endpoint orientation and answered the question "What is happening?" During the verbal task, the attention they paid to the area of interest was recorded via eye tracking. The findings revealed that although L2 speakers had generally learned to use the linguistic forms and their corresponding functions, they appeared to be unable to map these forms onto the underlying principles (perspective taking) for the event construal in their L2. Rather, more often, their L1 habitual conceptual patterns had an effect when L2 speakers described motion events. This can be seen in relation to the spatial concepts that L2 speakers preferred to encode (directional and boundary-crossing spatial concepts) as well as in the fixation patterns that L2 speakers tended to use. This suggests that the underlying principles for information organization might be subtle and partly immutable. Therefore, even for advanced L2 learners, the ability to restructure their *thinking and seeing for speaking* in the direction of the L2 was found to be limited.

Table of Contents

List of Figures

List of Tables

Abbreviations

ACC	Accusative
BC	Boundary-crossing spatial concept
CL	Classifier
CSC	Complex stative construction (*de*)
DAT	Dative
Deic	Deictic spatial concept
Dir	Directional spatial concept
GUO	Perfective aspect marker *guo*
Impf.	Imperfective
LE	Perfective aspect marker *le*
Loc	Locative
MOD	Modifier marker
NOM	Nominalizer (*de*)
Obj	Object-related spatial concept
Perf.	Perfective
Prosp.	Prospective
SA	Solicit Agreement (*ba*)
SOT	Speech onset time
TSit	The time of the situation
TT	The topic time
TU	The time of utterance
YAO	Prospective aspect marker *yào*
ZAI	Progressive aspect marker *zài*
ZHE	Durative aspect maker *zhe*

Chapter 1 Introduction

Does the language one speaks shape how one thinks about the world? This question has been discussed for centuries. Researchers in the field of anthropology, linguistics, and psychology have addressed the question of whether language has an impact on the way people view and take in the world they experience (Gentner & Goldin-Meadow, 2003; Gumperz & Levinson, 1996). However, until now, there is no consensus. Wilhelm von Humboldt was among the first to give a clear answer to this question. Humboldt (1963, p. 16) pointed out that thought is not merely dependent on the language in general, but is, to a certain degree, determined by language. One of the most influential hypotheses is the Sapir-Whorf hypothesis, which argues that speakers of different languages can be distinguished with respect to their different views of the world (Whorf, 1941). In recent discussions, two positions have been advocated with respect to the relationship between language and thought. One is based on the idea of universal conceptual structures. It claims that "linguistic systems are merely the formal and expressive medium that speakers devise to describe their mental representations and manipulations of their reference world" (Jackendoff, 1990; Li & Gleitman, 2002, p. 290); the other one highlights the role of grammatical form in relation to cognition, claiming that the "set of grammatically specified notions collectively constitutes the fundamental conceptual structuring system of language" (Talmy, 1988, p. 166). The latter position was further elaborated by Slobin, who proposed the hypothesis of *thinking for speaking*, which says that linguistic categories in one language have an impact on the way speakers of this language think in the process of preparing content for speaking (Slobin, 1996; for more details, see Chapter 2).

Today we can use experimental methods to advance the research in this field to gain more insight into this open question. Modern research examines the impact of language on cognition in various perceptual domains, like color (e.g., Thierry et al., 2009), time (e.g., Boroditsky, 2001; Casasanto & Boroditsky, 2008), space (e.g., Majid et al., 2004), and motion (e.g., von Stutterheim, 2003; von Stutterheim & Carroll, 2006; von Stutterheim et al., 2012). With the help of new tools, such as eye tracking, we can obtain more details about the relationship between language and cognitive processes. Von Stutterheim et al. (2012) took visual attention into account and proposed the hypothesis *seeing for speaking*, which indicates that linguistic categories direct

speakers' visual attention to certain aspects of an event in the process of preparing speech to describe it (for more details, see Chapter 2).

Within this context, this study aims to examine the influence of language-specific structures on event conceptualization in adult second language acquisition. To achieve this goal, this dissertation proceeds as follows: Chapter 1 introduces the background and the aim of this study. Chapter 2 presents the theoretical framework underlying this study. In this chapter, the processes of event conceptualization will be illustrated first, followed by the presentation of Talmy's (1985, 2000) lexicalization patterns. Next, two hypotheses relating to the linguistic categories and cognitive processes or visual attention, i.e., the hypotheses of *thinking for speaking* (Slobin 1996) and *seeing for speaking* (von Stutterheim et al., 2012), will be presented. Given that motion events are related to a figure's change in location over time, in addition to the spatial categories, temporal categories also play an essential role in event construal. Hence, Klein's (1994, 2009) framework of temporality will be delineated, followed by a review of the major studies on second language acquisition in the domain of motion events. Chapter 3 deals with the methodology used for the experiment. In Chapter 4, the spatial concepts encoded in L1s and L2s will be analyzed. More specifically, the use of verb types and adjuncts between L1 and L2 speakers as well as between two groups of L2 speakers will be compared in this chapter. In addition, the speech onset times and fixation patterns of L1 and L2 speakers will be analyzed in Chapter 4. In Chapter 5, the aspectual concepts encoded in L1s and L2s will be delineated. In this chapter, the use of aspectual markings by L1 and L2 Chinese as well as by L1 and L2 English speakers in relation to different video types will be compared. As a further step towards identifying the difference in the fixation patterns shown in Chapter 4 occurring among different language groups as time goes on, the attention paid to the areas of interest along the time axis will be compared between L1 and L2 speakers. In Chapter 6, the interrelationship between the spatial and aspectual categories in motion event descriptions in L1s and L2s will be described and analyzed. Finally, Chapter 7 is dedicated to the discussion of the findings.

Chapter 2 Theoretical Part

This chapter introduces the theoretical framework adopted in this dissertation. This study centers on the expression of motion events in L1s and L2s. Motion events involve a moving entity's change in location along a specified path and are among the most basic and pervasive events in our daily lives (Johnson, 1987). To begin, the processes of event conceptualization will be introduced briefly, and then the semantic components encoded in a motion event and the language types proposed by Talmy (1985, 2000) and Slobin (2004, 2006), respectively, will be presented. Following that, two hypotheses will be introduced. One is the hypothesis of *thinking for speaking* proposed by Slobin (1996) and the other one is the hypothesis of *seeing for speaking*, proposed by von Stutterheim and her research team (2012). Previous studies have shown that in addition to understanding spatiality, temporality is also indispensable in construing motion events (e.g., Bylund & Jarvis, 2011; von Stutterheim & Carroll 2006; von Stutterheim et al., 2017). Accordingly, section 2.4 deals with Klein's (1994, 2009) theoretical framework on temporal analysis. Subsequently, an overview of empirical studies on adult second language acquisition in the domain of motion events will be given. Finally, the major findings in this section will be summarized. Note that no introduction about the specific languages at issue (Chinese, English, and German) will be given in this chapter, but will rather be presented in subsequent chapters.

2.1 Event conceptualization

Language users can transform their intentions into speech. According to Levelt (1989, 1999), the process of transformation mainly involves three processing components, which can be illustrated in Figure 2.1. The three processing components are conceptualizer, formulator, and articulator. Information processing begins at the conceptualizer. The output of this process is a *preverbal message*, which is also the input for the formulator. A *preverbal message* is a conceptual structure, which is changed into a *surface structure* through *grammatical encoding*; by means of *phonological encoding*, this *surface structure* is turned into a *phonetic plan*, which is then processed in the articulator; after that, the final *overt speech* is produced.

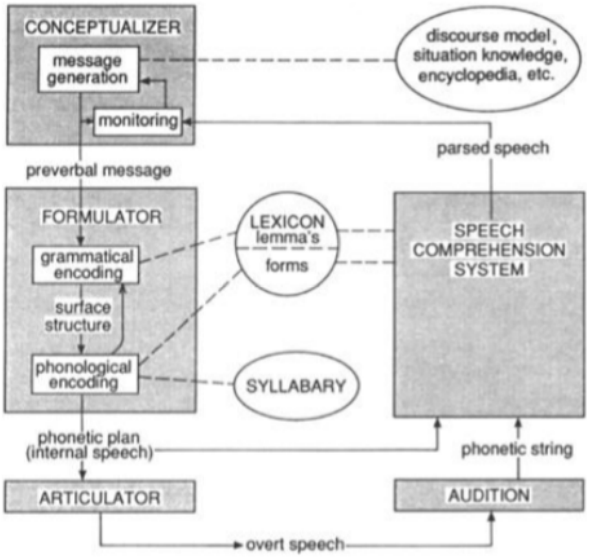


Figure 2.1: Levelt's 'A blueprint for the speaker' (1989, p. 9)

Given that this study is mainly concerned with the conceptualization of motion events in second language use, conceptualization is central to the *thinking for speaking* hypothesis. This hypothesis says that linguistic categories in a language exert an impact on the way the speakers of this language think in the process of preparing content for speech (for more details, see section 2.3). The process of preparing content for speech is the process of conceptualization in Levelt's (1989) terms. Therefore, only one of the above-mentioned processing components, the conceptualizer, is relevant. The other two processing components, i.e., the formulator and the articulator, will not be discussed further.

Before we introduce the processes of conceptualization, it is necessary to define the two terms *event* and *situation*, since they will be used throughout this dissertation. The term 'situation' refers to "what takes place in the external world", whereas the term 'event' means "a self-contained segment in a conceptual representation of a network of interrelated situations, conceptualized as a time-substance relation" (von Stutterheim & Nüse, 2003, p. 855).

Regarding the modeling of event conceptualization in language production, Habel and Tappe (1999) and von Stutterheim and Nüse (2003) proposed four planning processes: segmentation, selection, structuring, and linearization. In order to gain a better understanding of those four processes, an example of *a woman with a bag walking on the road towards some trash cans* will be used throughout the explanation of these four processes.

Segmentation refers to the decomposition of complex static situations into states or property predictions and the decomposition of complex dynamic situations into events or processes (von Stutterheim & Nüse, 2003, p. 853), meaning speakers decompose a situation into reportable units. Taking the above-mentioned woman walking towards trash cans as an example, speakers could represent the situation as *a woman is walking*. In this way, the situation is broken down into one dynamic action ('walking'). Speakers could also represent the situation as *a woman is holding a bag*, *she is walking towards trash cans*. If this is done, the situation is divided into static ('holding') and dynamic entities ('walking').

Through a selection process, speakers choose those components of the situation that could represent the event units to be verbalized. These components refer to spaces, times, properties, actions, etc. Therefore, it is the speaker's choice, for instance, whether or not to mention the endpoint in the above example. The speaker could either say *a woman is walking towards trash cans* or *a woman is walking across a yard*.

Structuring involves organizing the selected components in accordance with the spatial and temporal frame of reference. According to von Stutterheim and Nüse (2003, p. 865), the temporal frame of reference includes three conceptual components: the event, the timeline, and the observer or speaker. An event can be represented by a predicate (dynamic in nature) and its arguments. A situation denoted by a predicate can be characterized as having a beginning, a middle, and an end. The timeline is an abstract sequence of intervals. The observer or speaker decides where to set the anchor point and how to relate one event to another. In the above example, the speaker may take a temporal perspective by using both aspectual means and temporal adverbials; for example, *a woman is walking, she is looking at the trash cans in front of her and then she stops suddenly*. The events 'walking' and 'looking at the trash cans' take place within the same temporal interval. In contrast, the event 'stop' occurs after the previous interval, as indicated by the temporal adverbial *then*.

During linearization, the selected units or components are ordered in order to be encoded in the one-dimensional medium of language (Carroll et al., 2004; cf. Levelt, 1982). The processes of segmentation and selection are related to the decision of *what* to say, which corresponds to *macroplanning* in Levelt's terms (1989), while the processes of structuring and linearization are related to the decision of *how* to say it (i.e., perspective taking), which is in line with *microplanning* in Levelt's terms.

In order to gain insights into the cognitive processes in language production,

motion events have become an attractive research area in recent years. Numerous studies have examined the relationship between language-specific structures and their impact upon cognitive processes. These empirical studies are based on Talmy's lexicalization[1] patterns (see section 2.2) and Slobin's (1994, 1996, 2000, 2003, 2004, 2006, 2008) *thinking for speaking* hypothesis (see section 2.3). In the following, Talmy's and Slobin's frameworks will be presented.

2.2 Lexicalization patterns

The typological framework proposed by Leonard Talmy (1985, 2000) is one of the most influential works in cognitive semantics. It provides the basis for many fruitful studies in the domain of motion events. In his work, Talmy systematically investigated how conceptual structures are conveyed in different languages and more specifically, how the semantic components are encoded in different linguistic forms, including both lexical and grammatical categories. The former are open-class items, such as motion verbs (e.g., to walk), while the latter refer to closed-class syntactic devices, which are the elements associated with the verb (e.g., verb particles). Among all the semantic components of a motion event, Path[2] plays a critical role in categorizing language groups, based on which a universal typology can be proposed. Path is typically shown in the verb root in *verb-framed languages* (abbreviated as "V-languages") and in satellites[3] in *satellite-framed languages* (abbreviated as "S-languages"). The world's languages can be categorized accordingly into these two types. In this section, the semantic components of a motion event will be introduced first, based on which the typological dichotomy will be presented.

2.2.1 Brief introduction of semantic categories

Talmy (1985, 2000) investigated the relationship between linguistic form and meaning in order to identify typological and universal principles across different languages. In his work, the semantic components include Motion Event, Figure, Ground, Path, Motion, Manner, and Cause. The definitions are as follows:

[1] Lexicalization is a term used to refer to the relationship between a linguistic form and its meaning. It is defined as "the direct association of certain semantic components with a particular morpheme" (Talmy, 2000, p. 23).

[2] The first letter of the semantic categories will be written in capitals in this dissertation in order to distinguish them from the identical surface forms.

[3] Definition will be provided in the next part.

> The basic Motion event consists of one object (the **Figure**) moving or located with respect to another object (the reference object or **Ground**). The **Path** (with a capital P) is the path followed or site occupied by the Figure object with respect to the Ground object. The component of **Motion** (with a capital M) refers to the presence per se of motion or locatedness in the event. (Talmy, 2000, p. 25)

The terms Figure and Ground originate from Gestalt psychology. Figure is "the salient moving or stationary object in a Motion event whose path or site is the relevant issue" and Ground is "the reference object in a Motion event, with respect to which the Figure's path/site is characterized" (Talmy, 2000, pp. 153-54). Figure is the concept that "needs anchoring", while Ground is the concept that "does the anchoring" (Talmy, 2000, p. 311). So the relationship between Figure and Ground can be described as one entity serving as the reference object to anchor the other entity in space. Figure and Ground can either be two objects realized as nominals in a sentence or two events represented as main or subordinate clauses in a complex sentence. The Ground object can serve as the source, the landmark, or the endpoint in a motion event, and its geometric features such as the form of the object (point, line, surface, and two or three dimensions) are also taken into account for anchoring the Figure.

Manner and Cause are external semantic components and belong to the "Co-event" in Talmy's work. Talmy (2000, p. 152) gives the following definitions for Manner and Cause, respectively: Manner refers to "a subsidiary action or state that a Patient manifests concurrently with its main action or state" and Cause refers to "the qualitatively different kinds of causing events such as can be expressed by an English subordinate *from*- or *by*-clause." In most Indo-European languages, Manner or Cause is encoded in the verb, as seen in the English verbs "drive" (*the man is driving the car into the garage*), which means 'moving by means of driving', or "throw" (*a child is throwing a football*), which means 'moving by means of (due to) a child throwing it'.

Path refers to "the variety of paths followed or sites occupied by the Figure object in a Motion event"[4] (Talmy, 2000, p. 154). The semantic category Motion includes two motive states: motion and locatedness, which refers to "the occurrence (MOVE) or nonoccurrence (BELOC) specifically of translational motion" (Talmy, 2000, p. 25). So motion is defined by Talmy in a broader way, which includes, on the one hand, a Figure actually moving and changing places and, on the other hand, a Figure

[4] In fact, Path is a conceptual complex, which mainly includes three components: the Vector, the Conformation, and the Deictic (see more details in Talmy, 2000, pp. 53-57).

only located in a certain place without moving.

Motion is mainly encoded in verbs. There are three lexicalization patterns of verbs in terms of conflating patterns across different languages. The first type is the conflation of Motion with Manner/Cause on the verb root, which occurs, for example, in English, German, and Chinese.

Motion + Co-Event (Talmy, 2000, pp. 27-28)

a. The lamp *stood*/*lay*/*leaned* on the table.
b. The rock *slid*/*rolled*/*bounced* down the hill.
c. The napkin *blew* off the table.

The second type is the conflation of Motion with Path on the verb root. The information of Manner or Cause is expressed in adverbial or gerundive constructions, or is not mentioned in utterances. Examples of this language type are Romance languages (except for Latin), Japanese, Korean, etc. The following examples are in Spanish.

Motion + Path (Talmy, 2000, pp. 49-51)

a. La botella entró a la cueva (flotando)
   the bottle MOVED-in to the cave (floating)
   The bottle floated into the cave.
b. Tumbé el árbol serruchándolo// a hachazos/ con una hacha
   I-felled the tree sawing-it// by ax-chops/ with an ax
   I sawed// chopped the tree down.

In the third language type, the verb root encodes both Motion and Figure. Atsugewi and Hokan are examples of this language type.

Motion + Figure (Talmy, 2000, p. 57)

a. It *rained* in through the bedroom window.
b. I *spat* into the cuspidor.

Talmy (2000) further pointed out that a language usually uses one of the above-

mentioned three conflation types in the verb in the most characteristic[5] way in motion events.

2.2.2 Language types: verb-framed, satellite-framed, and equipollently-framed languages

To answer the question of which syntactic constituents express the Path information, a typological distinction can be made between *verb-framed languages* and *satellite-framed languages*. The syntactic constituents are verb root[6] and satellite, in Talmy's terms. As illustrated above, the verb root can express Motion and Manner/Cause, Motion and Path, or Motion and Figure per se. Satellite is "the grammatical category of any constituent other than a noun-phrase or prepositional-phrase complement that is in a sister relation to the verb root" (Talmy, 2000, p. 102), which overlaps partially but not completely with prepositions in English (for more details, see Talmy, 2000, pp. 103-09). Many empirical studies have expanded this notion of satellite to include all the non-verbal constituents referring to spatial concepts and call these constituents adjuncts (e.g., Carroll et al., 2012; Flecken et al. 2015). For instance, the prepositional phrase *into the cave* is an adjunct that expresses boundary-crossing Path information[7].

In *satellite-framed languages*, Path is expressed in the satellites. Grammatical forms such as verb particles in English, separable and inseparable verb prefixes in German, and verb prefixes in Latin or Russian are all considered satellites. Besides encoding the information of Motion, in S-languages the verb root also conveys the meaning of Manner or Cause, but not Path. Germanic, Slavic, Celtic, and Finno-Ugric languages as well as Chinese belong to this language type. For example,

English:

The bottle floated [Motion+ Manner] into [Path Satellite] the cave.

[5] According to Talmy (2000, p. 27), the meaning of "characteristic" is as follows: "(1) It is *colloquial* in style, rather than literary, stilted, and so on; (2) it is *frequent* in occurrence in speech, rather than only occasional; (3) it is *pervasive*, rather than limited - that is, a wide range of semantic notions are expressed in this type."

[6] Talmy only examined the verb root but not the verb, because he aimed at comparing different lexicalization patterns across languages with different word structures. In German, for example, in the particle verb *hineingehen*, meaning go-into-walk 'walk into (a place) away from the speaker', the verb root is *gehen* and the verb particle *hinein* is the satellite, whereas in Chinese, the verb root is the entire word and each can stand alone. German and Chinese will be compared later in terms of verb root as well as satellite.

[7] Since this section introduces Talmy's framework, the term satellites will continue to be used to indicate Path information in this section. However, for the spatial analysis in Chapter 4, the term adjuncts will be adopted to refer to various constituents that encode spatial concepts.

German:

Die Flasche schwebte [Motion+ Manner] in [Path Satellite] die Höhle.

the bottle float in the cave

The bottle floated into the cave.

The verbs *float* in English and *schweben* ('float') in German encode both Motion and Manner, whereas Path is encoded in the satellites *into* in English and *in* in German, respectively.

In *verb-framed languages*, Path is encoded in the verb root and Manner is either expressed in the gerundive or in an adverbial phrase, or is not expressed at all. Romance, Semitic, and Turkic languages, Japanese, Korean, Tamil, Polynesian, Nez Perce, and Caddo belong to this language type. Spanish is a typical verb-framed language. Take an example from Talmy (2000, p. 49) as below,

Spanish:

La botella entró [Motion+ Path] a la cueva (flotando) [Motion+ Manner]

the bottle MOVED-in to the cave (floating)

The bottle floated into the cave.

In this example, the verb *entró* encodes both motion and path, while the manner component *flotando* shows up in the gerundive phrase.

There are also distinctions which in a way concepts are across these two language types. This is related to the sources from which the spatial categories arise. Carroll et al. (2012) found that there are different types of spatial systems depending on whether the spatial concepts are derived from Figure in V-languages or Ground in S-languages. For example, the expression like 'along a river' is ground-oriented, while the expression like 'you turned' is figure-oriented. This study showed that the advanced French learners of German used in some circumstance figure-related concepts. But they didn't use ground-related concepts. They never said 'along' or something like this.

The pattern of Chinese is a matter of dispute. Chinese has serial verb constructions. Serial verb constructions refer to the co-occurrence of two or more verb phrases in the same clause without any morphological markings on each verb; each verb in such a serial verb construction can also appear alone in a clause[8]. It is thus not

[8] For more details about serial verb constructions in Chinese, see Chapter 4.

easy to identify which verb is the main verb[9] in a bipartite verb or serial verb construction. Talmy (1985, 2000) categorized Chinese as a type of S-language because he believes that the verb in Chinese expresses Manner information, rather than Path information, which in turn is encoded in path satellites. However, in contrast to path satellites such as verb particles and prefixes in S-languages, path verbs in Chinese can actually appear independently in utterance (for examples, see Chapter 4). Thus, Chinese poses a problem for Talmy's typological dichotomy. Chinese differs from S-languages in its capability to append different path segments on manner verbs within one utterance[10] (for more details, see Chapter 4). Furthermore, Chinese diverges from V-languages with respect to the frequent encoding of Manner information in the verb. It has been found that Manner information is not typically expressed in the main verb for boundary-crossing scenes in V-languages (e.g., Slobin, 2004, 2006; Slobin & Hoiting, 1994). By contrast, in Chinese one can freely express the Manner of a motion as well as Path components in boundary-crossing scenes[11].

To solve the problem of categorizing languages like Chinese with serial verb constructions, Slobin (1994, p. 492) in his early work suggested that serial verb languages and sign languages can be characterized as complex *verb-framed languages*. They are grouped as *verb-framed languages* because the core concept of Path is conveyed in the verbs, not in satellites; they are "complex" because there is no clear line between verb categories and satellite categories. Therefore, it is difficult to tell which verb in a verb compound is the main verb and which one is the subordinated one. Later, Slobin (2004, 2006) revised Talmy's typology and proposed that there is a third

[9] There is no consensus on the identification of the main verb in serial verb constructions in Chinese. A dominant view is that the first constituent in a serial verb construction that expresses the action is the main verb (e.g., Chao, 1968; Chirkova & Lamarre, 2005; Huang, 1988; Li & Thompson, 1981). A different opinion is that the second constituent in a serial verb construction, the so-called 'complement', which expresses the Path of the motion in a motion event, is the main verb (e.g., Hsueh, 1989; Tai, 1973, 2003). Talmy (2009) proposed criteria for judging the status of the main verb in a sentence. These criteria include phonological, morphological, syntactic, and semantic factors as well as the factors of co-occurrence patterns and class size (Talmy, 2009, pp. 391-92). Based on these criteria, the author further proposed criteria for identifying the status of the main verb in serial verb constructions in Mandarin Chinese (Talmy, 2009, pp. 396-400). The author argued that "the semantic and syntactic properties of a morpheme…seem largely to determine a native speaker's sense of the lexical category of that morpheme in its V2 usage – specifically, whether the morpheme is functioning as a verb or as a satellite there. And this assignment in turn determines whether the V1-V2 construction exhibits equipollent framing or satellite framing" (Talmy, 2009, p. 399). That is, in serial verb constructions (V1, V2, and V3 representing the co-event, Path, and deictic information, respectively, in Chinese), if the meaning of a morpheme in V2 is identical to the meaning of the same morpheme when this morpheme occurs alone in the sentence, then V2 is co-equal with V1: both V1 and V2 are the main verbs; in contrast, if the meaning of a morpheme in V2 diverges from the meaning of the same morpheme standing alone in the sentence, then V2 is subordinate to V1, which has main verb status.

[10] Chapter 4 will provide examples by L1 German-L2 Chinese speakers, who combined different kinds of path segments within one utterance. However, some combinations are not acceptable in Chinese.

[11] See more details in Chapter 4. In Chapter 4, the actual use of manner verbs + path verbs in L1 and L2 Chinese speakers when describing boundary-crossing scenes will be illustrated.

language type: *equipollently-framed languages* (abbreviated as E-languages), which are situated between V-languages and S-languages. In this type of language, Manner and Path are equal both in terms of grammatical status and significance or force.[12] Serial verb languages (Niger-Congo, Hmong-Mien, Sino-Tibetan, Tai-Kadai, Mon-Khmer, Austronesian), bipartite verb languages (Algonquian, Athabaskan, Hokan, Klamath-Takelman), and generic verb languages (Jaminjungan) are categorized as E-languages (Slobin, 2006, p. 65). Chinese is a type of equipollently-framed languages[13]. For example,

Chinese:

一个人走进屋去。

yī-ge rén zǒu [Manner verb] jìn [Path verb] le wū qù [Path verb].

one-CL person walk enter LE room go

A person walked into room (away from the speaker).

However, this tripartite typology fails to show the different degrees of attention to Path and Manner components in different languages. In order to show cross-linguistic differences from a cognitive perspective, Slobin (2004) further suggested that languages are better placed on a cline of manner salience, rather than categorizing them as S-languages, V-languages, or E-languages.

Manner salience refers to "the level of attention paid to the manner in describing events" (Slobin, 2006, p. 64). This cline of manner salience is a continuum of manner, and consists of high-manner-salient languages and low-manner-salient languages. Slobin (2004, p. 26) illustrated cline of manner salience as follows:

[12] Zlatev and Yangklang (2004) drew similar conclusions about Thai, which is also characterized as a serial verb language.

[13] In an empirical study on the categorization of Chinese, Chen (2007) examined the expressions of motion events from frog stories told by Chinese children and adults. In a subsequent study, Chen and Guo (2009) investigated the use of expressions of motion events in nine Chinese novels. These two studies showed that the structural or discourse features of Chinese do not completely match the characteristics of S-languages or V-languages. Instead, Chinese displays mixed characteristics of both S- and V-languages. The authors thus suggested categorizing Chinese as a type of E-language. However, to categorize Chinese as a type of E-language is not totally satisfactory. A recent psycholinguistic study on the morphosyntactic status of serial verb constructions in Chinese has shown that the verbs in serial verb constructions do not have equal grammatical weight (see Li, 2018). However, in comparison with German and English, where Manner is encoded in verbs and Path in satellites, Manner and Path are indeed frequently encoded in verbs in Chinese. This is indeed the typological difference between Chinese, German, and English. Basically speaking, the categorization of Chinese as a type of E-language captures such typological differences. So in this study, we will continue to categorize Chinese as an E-language, keeping in mind that this categorization is not totally satisfactory.

High-manner-salient languages: There is an accessible slot for manner in the language, made available in various ways:

- main verb in S-languages,
- manner verb in serial-verb languages,
- manner morpheme in bipartite verbs,
- manner preverb in Jaminjungan languages,
- ideophone.

Low-manner-salient languages: Manner is subordinated to path.

In high-manner-salient languages, speakers express Manner information frequently and easily. In low-manner-salient languages, in contrast, speakers express Manner information only when it is contextually required. When taking language types and cline of manner salience into consideration together, the correlation can be seen in Figure 2.2.

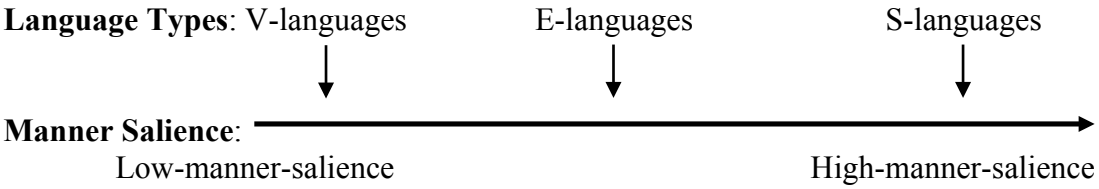


Figure 2.2: Correlation between language types and corresponding manner salience

As shown in Figure 2.2, we have attempted to place different language types on the scale of manner salience. There are two dimensions in Figure 2.2: One is the formal level (different language types) and the other is the conceptual level (different degrees of manner salience). Speakers of S-languages prefer to express Manner information because in S-languages, the conceptual component Manner is highly accessible and easily processable. It is the simplicity of the expressions and the ease of processing that make Manner encoding in S-languages a habitual pattern. So speakers of S-languages prefer focusing on Manner when describing motion events. That is the reason why S-languages belong to the high-manner-salient languages. It is important to note that speakers of different S-languages may have different degrees of preference for encoding Manner. Hasko (2010) and Pavlenko and Volynsky (2015) showed that Russian speakers prefer using more Manner verbs than English speakers in describing frog stories[14]. The degree of preference for encoding Manner by Russian speakers might be higher than that of English speakers. However, it should be pointed out that

[14] Frog stories are the retellings of a wordless book *Frog, where are you*. The author is Mayer (1969).

the preference of speakers of S-languages for encoding Manner does not exclude the possibility to encode Path in the main verb in motion events. There must be some restrictions to obey, either formal or conceptual, in order to encode Path in the main verb in S-languages. Slobin's proposal of manner salience does not show a fixed connection between language types nor different degrees of manner salience, but rather a preference for Manner encoding. As a consequence, speakers of V-languages can encode Manner in the main verb for some reasons and similarly, speakers of S-languages like English can encode Path in the main verb, given the fact that vocabulary from Romance languages (e.g., exit, enter) is used in English.

Talmy's classification was developed on the basis of numerous studies. One of the most important studies is Slobin's *thinking for speaking* hypothesis, which will be introduced in the following section. As a further step within this line of research, von Stutterheim et al. (2012) took visual attention into account and proposed the *seeing for speaking* hypothesis, which will also be delineated in the following section.

2.3 The hypotheses of *thinking for speaking* and *seeing for speaking*

In order to understand the relationship between cognition and language, this section introduces two hypotheses: the *thinking for speaking* hypothesis and the *seeing for speaking* hypothesis. The former was proposed by Slobin (1996). The hypothesis of *thinking for speaking* describes the dynamic thinking processes during the process of conceptualization. The latter was proposed by von Stutterheim and her colleagues (2012). With the hypothesis of *seeing for speaking*, von Sutterheim's research team extended event conceptualization research from lexicalized concepts to grammatical concepts and opened up a new window for examining the correlation between cognitive processes and grammatical concepts.

2.3.1 The hypothesis of *thinking for speaking*

Based on Talmy's *lexicalization patterns*, Slobin and Hoiting (1994) and Slobin (1996, 2000, 2003, 2004, 2006, 2008) further investigated the descriptions of motion events and the underlying cognitive processes in narratives from typological and cross-linguistic perspectives.

Unlike the two abstract and static entities "language" and "thought," the terms

"thinking" and "speaking" proposed by Slobin in the hypothesis of *thinking for speaking* are two dynamic concepts that focus on the dynamic mental process when speakers are talking. In this hypothesis, thinking is special because it is carried out during the process of speaking. Using Slobin's (1996, p. 76) own words, *thinking for speaking* would "involve picking those characteristics of objects and events that (a) fit some conceptualization of the event, and (b) are readily encodable in the language." In other words, while a speaker is verbalizing an event, they adapt their thoughts to the available linguistic forms. Those processes occur in the conceptualizer in Levelt's model (see Figure 2.1). The way in which speakers of different languages describe motion events is determined by those available lexical and grammatical resources. This is because, as Slobin (1996) pointed out, the world experiences of the speakers are filtered not only through the individual perspectives, but also through the particular linguistic options available in that language when they are speaking. That is, the morphosyntactic and lexical structures influence the mental process in the course of formulating utterances. Note that this hypothesis makes no claims about the impact of the language on the mental process which is not involved in the speaking process. *Thinking for speaking* takes place during the conceptualization phase of language production. In addition, it needs to be pointed out that this hypothesis applies to various forms of language production (not only speaking, but also translating, writing, etc.), reception (e.g., listening), and mental processes (e.g., remembering) (Slobin, 2000).

The impact of morphosyntax and lexicon on the mental process is reflected in the language-specific patterns of information organization in different languages. According to Slobin (2004), there are two factors that play a major role in determining the language-specific patterns of information organization: ease of processing and cultural practices and preferences. The author used the term *codability* to cover the factors for ease of processing, which are as follows[15] (Slobin, 2004, p. 16):

- expression by a finite rather than nonfinite verb
- expression by a high frequency rather than low frequency
- expression by a single word rather than a phrase or clause

The simpler a linguistic construction, the more often it will be used. A simple

[15] Note that Slobin (2006, p. 68) replaced the second factor listed above, "expression by a high frequency rather than low frequency" with "expression by an uninflected coverbal element rather than an inflected coverbal form". The latter means that it takes more effort to produce an expression with inflected forms than one without it.

expression is highly accessible and can be easily understood. As a consequence, speakers use this expression frequently. Choosing expressions that make it easier to be understood is a plausible explanation for the presence of different language-specific structures. It is assumed that speakers' preferences for easily understood expressions also have an effect on their perception. As a result of this, speakers will pay more attention to those semantic components that are easily understood and thus express them frequently. Speakers of S-languages have been trained to distinguish the fine-grained concepts of Manner, such as motor pattern and rate of the movement, whereas speakers of V-languages have not been trained to make such fine distinctions in terms of manner of motion. As Berman and Slobin (1994, p. 640) proposed:

> …that frequent use of forms directs attention to their functions, perhaps even making those functions (semantic and discursive) especially salient on the conceptual level. That is, by accessing a form frequently, one is also directed to the conceptual content expressed by that form.

However, this is not to say that speakers of V-languages do not pay attention to the manner of motion at all. On the contrary, when the manner of motion is significant, speakers of all different language types tend to express it explicitly.

2.3.2 The hypothesis of *seeing for speaking*

Based on empirical studies of seven languages (English, Spanish, Russian, Standard Arabic, German, Dutch, and Czech), von Stutterheim and her colleagues (2012) proposed the *seeing for speaking* hypothesis. In this hypothesis, a relationship is established between grammaticalized concepts and visual attention. According to the hypothesis of *seeing for speaking*, when a language has a particular grammaticalized concept and speakers of this language frequently express it, then they tend to pay attention to the visual features that are related to this grammaticalized concept. In contrast, if a language does not have the same grammaticalized concept and the same concept must be encoded in lexical, phrasal constituents or in other complex constructions and speakers of this language do not express this concept frequently, they will not pay attention to the related visual features in that situation, or do so to a lesser degree.

In order to test this hypothesis, speakers of the above-mentioned seven languages watched 60 video clips, of which ten were critical items that showed that the

endpoint was not reached (for example, a car driving down a road), ten were control items that showed that the endpoint was reached (for example, a man walking into a church), and 40 were filler items that showed different types of events (for example, a woman knitting). Speakers were asked to answer the question *What is happening?* Subjects were encouraged to start talking as soon as they recognized what happened in the video clips. This experiment used eye tracking as its main measurement tool. After the verbal task, a memory test was conducted. In the memory test, ten critical items and five filler items were selected from the stimulus set in the verbal task. Speakers were shown printed screenshots in which a specific part was cut out (the endpoint in the critical items and a particular object in the filler items) and asked to write down the missing part in the picture.

This study looked at the correlation between aspect and spatial concept and showed that aspect plays a role in construing motion events. The findings corroborated the *seeing for speaking* hypothesis. Speakers of aspect languages (Arabic, English, Russian, and Spanish) expressed the progressive or imperfective aspect in all the scenes. For the critical items, they mentioned fewer endpoints in the verbal task than speakers of non-aspect-languages (languages that do not have a grammaticized aspect of this type, for example, German, Dutch, and Czech). In the eye-tracking study, speakers of aspect languages had fewer fixation counts and the duration of the fixation on the endpoint was shorter for the critical items than among speakers of non-aspect languages. By contrast, there were no differences to be found in the control items between these two groups of speakers. In the memory task, it was hypothesized that speakers of non-aspect languages were better at remembering the object at the endpoint than speakers of aspect languages. The results validated this hypothesis.

Von Stutterheim et al. (2012) argued that the cross-linguistic differences could be ascribed to the structural system in each language and summarized the results as follows: In the verbal task, speakers of aspect languages and speakers of non-aspect languages differed in their preference for selecting and structuring information for verbalization. Speakers of aspect languages preferred decomposing the situation into different phrases and focused on the ongoingness by using the viewpoint aspect (imperfective/ progressive), while speakers of non-aspect languages preferred taking a holistic view in event construal and thus mentioned the endpoint in the descriptions. In the eye-tracking study, speakers of aspect languages and speakers of non-aspect languages differed in the allocation of attention to the endpoint during the language

production tasks. In the non-verbal task, that is, the memory task, the two language groups differed in performance with regard to memorizing specific components, i.e., the endpoint, in the motion events.

Von Stutterheim and her colleagues (2003, 2006, 2012, 2013, 2017) extended event conceptualization research from lexicalized concepts (Manner and Path encoding) to grammatical concepts and opened up a new window for examining the correlation between cognitive processes and grammatical concepts. Grammatical concepts may provide a highly automatized or most familiar routine when preparing content for speech, given their speed of delivery. In this sense, grammatical categories might form the basis of language-specific effects (Flecken et al., 2014). It is interesting to examine to which extent language-specific effects can affect cognitive processes. This study looks at the extent to which language effects impact motion event cognition in second language acquisition.

Based on Talmy's and Slobin's framework, Manner and Path encoding in different languages have been investigated extensively. However, neither Talmy's nor Slobin's framework deals with the temporal property of motion events. Given that space and time are two essential categories in motion events, if we limit ourselves to the spatial category, then we might get an incomplete picture of motion events conceptualization. In a recent study, von Stutterheim et al. (2017) examined the interaction between spatial and temporal concepts encoded in the expressions of motion events in Tunisian Arabic (TA) and Modern Standard Arabic (MSA). TA and MSA are two closely related varieties of Arabic. However, they differ in their repertoire of linguistic forms for expressing spatial concepts and their verbal aspectual categories. In comparison with MSA, TA has fewer forms (path verbs, prepositions) for encoding spatial concepts when describing directed motion events. However, the aspectual system in TA is richer than in MSA. The authors found that the expressions of directed motion events are conveyed in spatial concepts in MSA, whereas the same information is conveyed in temporal aspects (progression) in TA. Von Stutterheim and her colleagues' study unveiled the necessity to focus on the temporal domain in the typology of motion events "in order to gain further insight into the systematicity in the diversity of encoding options" (von Stutterheim et al., 2017, p. 245). Given the importance of taking temporal concepts into account when analyzing motion event expressions, this study examines the language products in verbal tasks from temporal perspectives between L1 and L2 groups. To this end, section 2.4 will present the basic

framework proposed by Klein (1994, 2009) for temporal analysis.

2.4 Temporal analysis

Klein's (1994, 2009) theoretical framework about temporality is time-relational and provides a precise definition of tense, aspect, and lexical content. This framework can be applied in different languages and thus serves as the theoretical basis for comparing temporality between different languages in this study.

Time plays an essential role in our daily life. But in different cultures, people may have different concepts of time. Among these different time concepts, there is a basic structure regarding time, the 'Basic Time Concept', which underlies the temporal relations in natural languages and has seven characteristics (Klein, 1994, pp. 59-63; 2009, pp. 27-29):

- *Segmentability*: Time can be separated into smaller units, like time spans or time intervals;
- *Inclusion*: There are different relations between those time spans or time intervals; a time span may fully or partially include another one;
- *Linear order*: The time spans do not include each other; rather, a time span precedes another one;
- *Proximity*: A time span may be close to or far away from another one;
- *Lack of quality*: A time span does not have any qualitative features;
- *Duration*: A time span may have different lengths of time, short or long;
- *Origo*: A time span is distinguished and is called 'the time of present experience'.

According to the 'Basic Time Concept,' time can be regarded as a structure that is composed of different time spans. A time span can precede, overlap, partially overlap, or follow another time span. Accordingly, the traditional three temporal relations (before, after, and during) can be enriched to include more refined temporal relations such as long before (or after), shortly before (or after), just before (or after), partly before (or after), etc. There are different linguistic means to express these different temporal relations in different languages. Some languages have grammatical means, like tense and aspect marking; other languages might not have such grammatical categories and instead have different lexical elements to express temporality. These devices include: temporal adverbials (e.g., *in an hour*, *two days ago*), temporal particles

(e.g., the Chinese aspect particles *zài*, *zhe*, *le*), compound expressions (e.g., compound verbs like *to continue studying*). To summarize, the devices frequently used to encode time in different languages are: tense, aspect, Aktionsart, temporal adverbials, temporal particles, and discourse principles (Klein, 2009).

There are three important time spans in Klein's (1994, 2009) framework: topic time (hereafter abbreviated as TT), time of the situation (hereafter Tsit), and time of utterance (hereafter TU), of which TT plays a critical role in temporal analysis because it is TT that hooks the lexical contents up to the time axis. In addition, based on the relation between TT and the other two time spans (Tsit and TU), both tense and aspect can be defined precisely. In the following, we will introduce the definitions of each of these three time spans, based on which tense and aspect will then be defined.

2.4.1 Topic Time, Time of Situation, and Time of Utterance

Klein (2009, p. 46) gave the following definitions for time of utterance, topic time, and time of the situation:

> the time of utterance; this is the time at which the utterance is expressed
> the topic time; this is the time about which something is asserted (or asked)
> the time of the situation; this is the time at which the situation obtains or occurs

Tsit is the time interval for which a situation holds. It is objectively fixed. TU is the time at which an utterance is made. It is usually but not necessarily related to *origo* (here and now). TT is the time for which an assertion is made, which plays an essential role in defining tense and aspect.

Tense: the relation between TT and TU

Tense can be defined according to the temporal relations between TT and TU. For a temporal relation to exist, two time spans are generally involved. One is *relatum*, which is given in different ways; the other one is *theme*, which is related to *relatum*. A *relatum* can be deictic (e.g., it can be TU), or it can be anaphoric (e.g., the time reference in the preceding text); it can also be given by knowledge of the world (e.g., some historical moment). According to the different features of a *relatum*, a temporal relation can be described as a deictic, anaphoric, or calendric temporal relation. Tense is a deictic temporal relation in which TT can be located in the present, past, or future, depending on the time at which an utterance is made.

Aspect: the relation between TT and Tsit

Aspect can be defined as the temporal relation between TT and Tsit. Note that TU cannot be directly related to Tsit. Only through TT can TU and Tsit be related to each other. Tsit is the time span in which the situation is derived from the lexical content. Klein (1994, pp. 99-100) further distinguished three types of temporal relations between Tsit and TT:

A Tsit is interpreted as fully including TT (abbreviated TT INCL Tsit)
B Tsit is interpreted as partly including TT (abbreviated TT AT Tsit)
C Tsit is interpreted as excluding TT (abbreviated TT EX Tsit)

Here INCL means TT is completely included in Tsit, whereas AT means that TT is partially before or partially after Tsit and EX shows that TT is completely outside Tsit, either long after or long before Tsit.

Languages differ in the linguistic means used to express the relations between Tsit and TT. If a language has grammatical means to specify the relations between Tsit and TT, it typically marks the verbs morphologically; if a language does not have grammatical means to specify the relations between Tsit and TT, this does not mean that said language cannot express those relations; rather, the speakers of this language may choose lexical items or provide contextual information to specify the relations between Tsit and TT. Based on the relation between qualitative and temporal properties of a situation characterized in the lexical contents, Klein (1994) distinguished three types of lexical contents: 0-state, 1-state, and 2-state lexical contents.

2.4.2 0-state, 1-state, and 2-state lexical contents

Lexical contents are the selective description of a situation. That is to say, speakers may choose some parts of the situation to express and leave other parts implicit. Lexical contents could be words, phrases, clauses, or full sentences. Also, they do not have any temporal anchors by themselves, but can be hooked up to the time axis via TT. For example, the lexical content <John read a novel> can be linked to the past when one is asked about what they saw at a specific time: *What did John do when you arrived at home?* The answer would be *John was reading a novel*. Lexical contents can be classified according to their behavior towards TT-contrast, which is the contrast between the TT about the assertion made and the other possible TTs.

0-state lexical contents

When lexical contents do not show any TT-contrast, they are called 0-state lexical contents. Take the utterance *Beijing is in China* as an example. In this example, Tsit occupies the whole time and always includes TT. There are no other possible time spans as TTs that could contrast the TT in this utterance.

1-state lexical contents

When lexical contents show an external TT-contrast, they are called 1-state lexical contents. Here 'external' means that the contrast is not within the time span for which the specified situation holds, but rather outside such a time span. Take the utterance *John is studying in Heidelberg* as an example. In this example, TT1 could be the time at which a friend of John is talking about John's current situation to other people in a party. There are many time spans before this TT1 (TT-) and after this TT (TT+) for which the same situation holds, that is, <John study in Heidelberg>. TT-, TT1, and TT+ constitute the whole Tsit. There are also time spans before Tsit and after Tsit, which Klein (1994, p. 84) calls PRETIME of Tsit and POSTTIME of Tsit, respectively. Accordingly, PRETIME of Tsit, Tsit, and POSTTIME of Tsit make up the 'entire time'. For 1-state lexical contents, the TT-contrast is either in PRETIME or POSTTIME or in both time spans.

2-state lexical contents

When lexical contents show an internal TT-contrast, they are called 2-state lexical contents. Here 'internal' means the contrast is located in the lexical contents themselves. The 2-state refers to the source state (SS) and the target state (TS), respectively. The time span occupied by SS and TS can be abbreviated as T-SS and T-TS, respectively. Take the utterance *a woman entered the supermarket* as an example. The 2-state lexical content *enter* in this example includes a SS when this woman was not in the supermarket and a TS when she was in the supermarket. There are various ways to link lexical contents to TT in this example. More specifically, SS can be linked to TT; TS can be linked to TT or part of SS; and part of TT can be linked to TT, which may lead to different choices of aspectual markings in different languages if a language has such an aspectual grammatical category.

2.4.3 Summary

Motion events can be characterized as a Figure moving over time in a certain manner with respect to a Ground object. That is to say, apart from the concept of space,

the concept of time is also indispensable when describing motion events. Previous studies have shown the importance of relating spatial categories to temporal categories in the domain of motion events (von Stutterheim et al., 2017). Hence, in addition to the spatial analysis, it is necessary to conduct a temporal analysis and consider the correlation between these two categories when comparing motion event descriptions between speakers of L1s and L2s in this study. Based on the frameworks of Talmy (1985, 2000), Slobin (1996, 2004, 2006), and Klein (1994), numerous empirical studies have been carried out. In the following section, the major studies on adult second language acquisition in the domain of motion events will be reviewed from both a spatial and a temporal perspective.

2.5 Studies on second language acquisition in the domain of motion events

Based on studies on first language acquisition, Berman and Slobin (1994) argued that children develop a specific way of *thinking for speaking* when they acquire their first language. Children learn "to attend to particular aspects of experience and to relate them verbally in ways that are characteristic of that language" (Berman & Slobin, 1994, p. 611). Although the *thinking for speaking* hypothesis is derived from first language acquisition, it also has implications for second language acquisition (SLA). Learning a second language means learning another way of *thinking for speaking*. According to Cadierno (2012), learning to express motion events in a second language involves two difficult aspects: first, how to direct attention to particular details in the input required by L2 and express them (such as attention to the trajectory of the movement vs. attention to a static scene) and second, how to acquire the meaning-form mappings in an L2.

Based on the typological framework proposed by Talmy (1985, 2000), recent studies about SLA in motion event encoding mainly concern how L2 speakers actually express Path and Manner information and whether L1 language-specific *thinking for speaking* has an effect on SLA. This line of research has one variable: the typological similarity or difference between source and target language (e.g., L1 speakers of S-languages acquiring an L2 that is a type of V-language or S-language, or vice versa). Besides this line of research, there is another important line of research within the

thinking for speaking paradigm, which examines the interplay of the grammaticalized aspect and endpoint encoding in the domain of motion events (e.g., Bylund, 2009; Bylund & Jarvis, 2011; Carroll & von Stutterheim, 2003; Carroll et al., 2004; Schmiedtová & Flecken, 2008; von Stutterheim, 2003; von Stutterheim & Nüse, 2003; von Stutterheim et al., 2012; von Stutterheim et al., 2017; more details below).

Overall, the empirical studies performed on SLA in motion events do not provide a clear and consistent picture with respect to the role of L1 in an L2. The research results show that L1 language-specific patterns either play a limited or no role, or a major role in acquiring a second language. In addition, language convergence can be found between L1 and L2. In the following, major studies will be reviewed with respect to these three positions.

L1 language-specific thinking for speaking plays a limited role in SLA

Some empirical studies show that L1 language-specific *thinking for speaking* plays a limited role when L2 learners talk about motion events. Studies such as those by Cadierno (2004), Cadierno and Ruiz (2006), Navarro and Nicoladis (2005), Stam (2001), and Kellerman and Van Hoof (2003) show this position.

Cadierno (2004) investigated how speakers of Danish (an S-language) with two levels of proficiency, intermediate and advanced, learned to express spatial concepts in Spanish (a V-language) when describing motion events. Based on retellings of a frog story, this study showed no consistent impact of L1 on L2. On the one hand, L1-specific patterns played an important role in the degree of complexity and elaboration of Path information by L1 Danish-L2 Spanish speakers[16]. On the other hand, L1 played a limited or no role in event conflation and the attention given to movement and setting by L2 speakers.

In a follow-up study, Cadierno and Ruiz (2006) investigated how advanced L1 Danish-L2 Spanish and L1 Italian-L2 Spanish speakers expressed Manner information. As mentioned above, Danish and Spanish belong to different language types, while both Italian and Spanish are V-languages and thus share the same typological features. Like the previous study, this study is also based on frog story narratives. The results showed that there were no significant differences between learner groups and native speaker group with respect to the number of manner verbs used and the alternative

[16] When we talk about L1 Danish-L2 Spanish speakers, we mean that they are Danish native speakers and learn Spanish as a foreign language. Likewise, in this dissertation, when we talk about L1 German-L2 Chinese or L1 German-L2 English speakers, we mean that those speakers are German native speakers and learn Chinese or English as a foreign language.

expressions of Manner information (e.g., gerundive or prepositional phrases). However, L1 Danish-L2 Spanish speakers differed from L1 Italian-L2 Spanish and Spanish native speakers in that they used manner verbs as main verbs in describing boundary-crossing events, which led to unacceptable expressions because in Spanish manner verbs are not allowed to appear as main verbs in boundary-crossing situations. In conclusion, the findings revealed that L1 *thinking for speaking* plays a limited role in L2 in the domain of motion events.

Navarro and Nicoladis (2005) provided further evidence to support the findings of Cadierno (2004) and Cadierno and Ruiz (2006). The study examined whether advanced L1 English-L2 Spanish speakers acquired the use of path verbs in Spanish. In Spanish, Path information is packed into path verbs, which contrasts with the way in which Path information is characteristically packed into verb particles and prepositions in English. This study was based on oral narratives from two video films. The findings showed that there were no significant differences between L1 English-L2 Spanish speakers and Spanish native speakers with respect to the number of path verbs used. However, L2 speakers used fewer bare path verbs than Spanish native speakers. In addition, the results illustrated that there were some expressions that included motion verbs with locative phrases. The authors concluded that the L2 speakers had not fully acquired the lexicalization pattern of the target language and there were still traces of L1 in L2 production. In the authors' own words, these advanced L2 speakers "had almost fully achieved the L1 Spanish patterns for the description of motion events in oral narratives" (Navarro & Nicoladis, 2005, p. 106).

Another line of research concerns the interface between speech and gesture. Stam (2001) examined how intermediate and advanced L1 Spanish-L2 English speakers expressed motion events and placed their gestures. The findings showed that when L2 speakers expressed Path information, some speakers placed the path gestures on the satellites, but some of them failed to do so; when L2 speakers expressed Manner information, they would place path gestures on the motion verbs. In short, L1 *thinking for speaking* played a partial role in L2.

Kellerman and Van Hoof (2003) investigated how high-intermediate and advanced L1 Spanish-L2 English and L1 Dutch-L2 English speakers expressed motion events and placed their path gestures. The results were complex. L1 Dutch-L2 English speakers displayed unexpected results in that they placed most of the path gestures on the motion verbs, which followed the patterns of V-languages. The authors did not give

a definite explanation for this. L1 Spanish-L2 English speakers tended to put path gestures on the motion verbs, which showed their reliance on L1 *thinking for speaking*.

L1 language-specific thinking for speaking plays an important role in SLA

Some empirical studies show that L1 language-specific *thinking for speaking* plays an important role when L2 learners talk about motion events. Representative studies for this position are Cadierno (2010), Carroll et al. (2012), Flecken et al. (2015), Hendriks et al. (2008), Negueruela et al. (2004), von Stutterheim (2003), von Stutterheim and Carroll (2006), and Schmiedtová and Sahonenko (2008).

Cadierno (2010) extended the learner groups to include advanced German, Russian, and Spanish learners of Danish and examined how L2 speaker expressed boundary-crossing events. The results showed L1 transfer in the learner groups. The German and Russian learners of Danish preferred using the construction "manner verbs + path satellites", while Spanish learners of Danish showed a tendency to use non-manner verbs + path information, which displayed their respective L1 patterns. The author argued that this L1 transfer was not a transfer of linguistic structure because German and Russian have different linguistic forms to encode Path information: verb particles, prepositions, and adverbs in German and prefixes in Russian. Rather, it was a transfer at the conceptual level. L2 speakers tended to select the conceptually salient components that were in line with their L1 and expressed them in L2.

Von Stutterheim (2003) and von Stutterheim and Carroll (2006) examined the impact of the grammatical category on the event construal in SLA. The starting point of their research was based on cross-linguistic observations (Carroll, 1993; Carroll & von Stutterheim, 2003; von Stutterheim et al., 2009; von Stutterheim & Lambert, 2005; von Stutterheim & Nüse, 2003): Speakers of a language such as English, where the grammatical aspect is obligatorily encoded in the predicate, tended to decompose the situation into different phases (initial, intermediate, and terminative). In contrast, speakers of a language such as German, where there are no distinctions between the imperfective and the perfective aspect, tended to take a holistic view. Both studies showed that even highly proficient L2 speakers were affected by their L1-specific conceptualization patterns.

The results reported in von Stutterheim (2003) showed that in the encoding of directed motion events, the frequency with which the endpoint was mentioned and the speech onset times in advanced L1 German-L2 English and L1 English-L2 German conformed with their respective L1 languages. Thus, both speakers of L1 English-L2

German and L1 English tended not to mention the endpoint, while both speakers of L1 German-L2 English and L1 German tended to mention the endpoint. In some of the scenes in which the endpoint was not actually shown, some speakers of L1 German-L2 English and L1 German imagined or even created an endpoint (for example, speakers saw a boy playing in the sand, but some of them would make statements like 'a boy was building a sandcastle'). However, L1 German-L2 English speakers were more target-like than L1 English-L2 German speakers in describing motion events. Von Stutterheim (2003, p. 202) speculated that "in English there is a salient grammatical form which encodes the specific perspective and by acquiring this form the learner is pointed, so to speak, to the function it serves." However, there are no overt grammatical devices to package the concept of a holistic view in German. It is thus not easy to discern how the concept is expressed for L2 speakers. In the author's own words, speakers have to "identify this concept and its function through inference from different informational components (arguments, locatives + verb)" (von Stutterheim, 2003, p. 202).

In a subsequent study, von Stutterheim and Carroll (2006) adopted the same methodology as the study in von Stutterheim (2003) and provided eye-tracking data in L1 and speech onset times in L1 and L2. The findings showed that L1 German speakers had more endpoint fixations than L1 English speakers either before or after the speech onset times, while L1 English speakers had relatively more fixations after the speech onset times. The eye-tracking data conformed with the tendencies of L1 German and English speakers shown in previous findings. The reason for the differences of fixations before and after speech onset in L1 English might lie in the fact that L1 English speakers might not pay attention to the endpoint initially because they focus on the ongoingness of the event. After they start talking, they can also encode the endpoint in the already uttered sentence (e.g., a car is driving along the road…to a village). So there are relatively more fixations on the endpoint after the speech onset times in L1 English. This evidence also showed that it is not that mentioning the endpoint is unimportant in English; rather, it appears that speakers of this language are not inclined to pay attention to the endpoint in describing a goal-oriented motion event. As for the speech onset times in L1 and L2, the findings showed that L1 German-L2 English speakers moved towards the target language (L1 German-L2 English speakers: 3.0 seconds; L1 English speakers: 3.6 seconds), while L1 English-L2 German still conformed with L1 English patterns (L1 English-L2 German speakers: 3.8 seconds; L1 German speakers: 4.3

seconds).

In line with the above-mentioned studies (von Stutterheim, 2003; von Stutterheim & Carroll, 2006) on SLA, Schmiedtová and Sahonenko (2008) examined the role of the grammatical aspect in the motion event construal in L1 Czech-L2 German and L1 Russian-L2 German speakers. They found that although Czech and Russian have a similar aspectual system, L1 Czech and L1 Russian speakers took different perspectives to encode goal-oriented motion events[17]. L1 Czech speakers tended to take a holistic view to conceptualize an event as completed, while L1 Russian speakers preferred to conceptualize an event as ongoing, without considering the endpoint in describing the motion events. The analysis showed that L1 Czech speakers mentioned the endpoint more frequently than L1 Russian speakers. L1 Czech-L2 German speakers displayed the same pattern as L1 Czech speakers, both of whom mentioned significantly more endpoints than L1 German speakers, while L1 Russian-L2 German speakers showed a similar tendency as L1 Russian speakers, both of whom mentioned significantly fewer endpoints than L1 German speakers.

To summarize, the three studies von Stutterheim (2003), von Stutterheim and Carroll (2006), and Schmiedtová and Sahonenko (2008) showed that the grammaticalized aspectual category plays an important role not only in information organization, but also in directing the speaker's attention to certain aspects of a situation. The empirical analysis showed that L1 German-L2 English speakers moved towards the target language, while L1 English-L2 German speakers retained their L1 pattern. These differences are ascribed to "the transparency of encoding of the new conceptual category" (Schmiedtová et al., 2011, p. 94). Therefore, the progressive in English is perceptually and formally salient, whereas the holistic concept in German is covert. The findings showed that L2 speakers have difficulties in applying the principles of event construal which are entailed by a grammatical category (such as aspect) in an L2, although they have formally mastered the target language to a nearly perfect level.

Carroll et al. (2012) conducted a fine-grained analysis of conceptual components in motion events used by advanced L1 French-L2 English and L1 French-L2 German speakers. In order to shape a trajectory, speakers of L1 French (a V-

[17] Schmiedtová and Sahonenko (2008) argued that the differences between Czech and Russian speakers lie in the fact that Czech aspectual system has changed as a result from the contact with German. Furthermore, the different perspectives taken in Czech and Russian may lie in restrictions on the combinations between motion verbs and the perspective and the secondary aspect. In addition, the differences in aspectual use are caused by the semantic change of the perfective in Czech (see Schmiedtová & Sahonenko, 2008; Schmiedtová et al., 2011).

language) preferred spatial concepts related to the entity, while speakers of L1 German and L1 English (both S-languages) preferred spatial concepts derived from ground features, i.e., the contours of the ground. It was found that L1 French-L2 English and L1 French-L2 German speakers had mastered the use of manner verbs, but used fewer ground-based adjuncts and particles than L1 English and L1 German speakers. In turn, in line with L1 French speakers, they preferred depicting the location of the entity in motion with a manner verb even in scenes where the ground features were prominent. In fact, the ground-based concept was present in L1 French and those advanced L2 speakers who had acquired ground-based linguistic expressions in English and German. In other words, L2 speakers had acquired the meaning-form mappings, but still displayed L1 patterns. The authors pointed out that besides the acquisition of meaning-form mappings in a second language, attentional patterns (such as the attention paid to entity- or ground-based spatial concepts), which profiled the concept and its linguistic forms, played an important role for ultimate attainment in L2 learning.

In a follow-up study, Flecken et al. (2015) supported these findings from Carroll et al. (2012). They carried out further research on the different allocation of attention patterns with time course measured via eye-tracking before speech onset times by L1 German, L1 French, and advanced L1 French-L2 German speakers. The findings showed increased fixations on the moving entity within the first peak among L1 French and L1 French-L2 German speakers in comparison with L1 German speakers. In line with the studies (Carroll et al., 2012; Daller et al., 2011; Schmiedtová, 2013), the authors found the L1 effect to be deeply entrenched in L2 and argued that L1 transfer occurred at the conceptual and information selection level in that unlike L1 German speakers. L1 French-L2 German speakers tended not to structure the path based on the ground contours.

Besides studies on French learners of English and German, there are also studies about English learners of French regarding the motion event construal. Hendriks et al. (2008) examined how adult low-intermediate- and advanced-level L1 English-L2 French learners expressed caused motion compared to English and French native speakers. The findings showed that advanced-level L2 speakers attempted to produce utterances with semantic density (i.e., the number of information components expressed), which were, however, less dense than those of native speakers. Concerning the semantic components expressed, both English and French native speakers tended to express three semantic components: Cause, Manner, and Path, while particularly the

low-intermediate learner group tended to express only Manner and Cause. In addition, the advanced learner group used either target-like or non-target-like devices (such as using particles to express Path information) to express the integration of Manner, Cause, and Path in one utterance, while the low-level learner group tended to separate Path from Manner and Cause and used non-target-like devices to express Path. This analysis demonstrated the occurrence of the L1 effect at both levels of proficiency.

Concerning the interface between speech and gesture, Negueruela et al. (2004) conducted a bi-directional analysis about the extent to which advanced L1 Spanish-L2 English and L1 English-L2 Spanish speakers displayed L2 *thinking for speaking*. The findings showed that both groups of L2 speakers had difficulties with the L2 pattern of *thinking for speaking*. This was particularly reflected in L1 English-L2 Spanish speakers, who tended to use manner gestures. This may be because, as there is a vast number of manner verbs in English, speakers of English tend to express Manner information.

L1 and L2 convergence[18]

Some empirical studies have shown that the restructuring (partially or wholly) of L1s is possible. This transfer seems to be common in bilinguals. For example, Brown and Gullberg (2008), Bylund (2009), Bylund and Jarvis (2011), and Hohenstein et al. (2006) hold this point of view.

Bylund (2009) investigated the role of L2 onset age in L1 conceptual patterns with respect to endpoint encoding and temporal perspectivation in describing goal-oriented events. It was found that when the age of L2 onset was lower than 12 years, L1 Spanish-L2 Swedish bilinguals tended to show L2 conceptual patterns (in line with L1 Swedish speakers, e.g., frequently mentioning the endpoint and showing a preference for simple present tense). In contrast, when the age of onset was higher than 12 years, the bilinguals tended to display L1 conceptual patterns (e.g., mentioning the endpoint less often and not using the simple present tense as much). The author ascribed this to L1 contact and socio-psychological factors (e.g., motivation and attitude towards maintaining L1). These two factors together with the knowledge of L2 conceptualization patterns resulted in different outcomes between these two bilingual groups.

As a follow-up analysis, Bylund and Jarvis (2011) further examined the

[18] Convergence means that the merged event conceptualization patterns differ from L1- and L2-based patterns (Pavlenko, 2005).

interplay between endpoint encoding and sensibility to aspectual distinctions as well as the factors that may influence this relation in L1 Spanish-L2 Swedish bilinguals. In this analysis, a grammaticality judgment test (e.g., verb agreement and aspectual contrasts) was carried out. The results showed that bilinguals who performed badly in the grammaticality test were inclined to encode endpoints and less inclined to express ongoingness in comparison with L1 Spanish speakers. Therefore, the correlation between endpoint encoding and sensibility to L1 aspectual contrasts was negative. The author proposed two factors as reasons for this finding: reduced contact with L1 Spanish and proficiency in L2 Swedish.

The reasons for the different results between the studies (Carroll et al., 2012; Flecken et al., 2015) and the studies (Bylund, 2009; Bylund & Jarvis, 2011) may lie in the fact that the L2 speakers in the latter study had lived in the target language country for a long time and therefore had been exposed to that language for a longer time. Complete immersion in the target language environment is regarded as a prerequisite for the likelihood of reconstructing L1 patterns (see the overview in Athanasopoulos, 2011).

2.6 Summary

In this chapter, we discussed the theoretical fundamentals surrounding the cognitive processes in language production, especially the spatial and temporal concepts in the domain of motion events[19]. In addition, in this chapter we also reviewed the major studies about SLA within this framework. The major points from this chapter are as follows:

Three processing components are involved in the process of language production: conceptualizer, formulator, and articulator (Levelt, 1989). In this study, we focus on the processing component conceptualizer. According to Habel and Tappe (1999), and von Stutterheim and Nüse (2003), there are four processes in event conceptualization: segmentation, selection, structuring, and linearization. When speakers conceptualize situations, they first segment complex situations into smaller dynamic or static situations; then these sub-situations or components of the sub-situations (such as mentioning or not mentioning the endpoint) are selected in order to

[19] However, in the theoretical part, we have not conducted any analysis in the specific languages (English, German, and Chinese) at issue. This will be done in the subsequent chapters.

verbalize; next, the the selected situations or components of the situations are structured in accordance with the perspective taken and the spatial and temporal frame of reference; and finally, the structured information is ordered.

Talmy (1985, 2000) investigated how semantic components are encoded in different linguistic forms. These semantic components include Figure, Ground, Motion, Path, Manner, and Cause. Among these, Path plays a critical role in categorizing language groups, based on which a universal typology can be proposed. When Path is characteristically encoded in the verb root in a language, then this language belongs to the V-languages, whereas when Path typically shows up in the satellites in a language, then this language is a type of S-language. Based on this typological dichotomy, numerous studies have been carried out. One of the most important works has been contributed by Slobin (1994, 1996, 2000, 2003, 2004, 2006, 2008), who developed this typological framework systematically and with a focus on cognition. Slobin (1996) proposed the hypothesis of *thinking for speaking*. That is to say, when a speaker is verbalizing an event, they tend to adapt their thoughts to the available linguistic forms. Furthermore, based on the typological dichotomy proposed by Talmy (1985, 2000), Slobin (2004, 2006) proposed a third type of languages, *equipollently-framed languages*, which includes languages with serial verb constructions (such as Chinese) and languages with bipartite verbs or preverbs. In order to capture the different degrees of attention to Path and Manner components in different languages, Slobin (2004) further suggested that languages are better placed on a cline of manner salience. S-languages are representatives of high-manner-salient languages and V-languages are low-manner-salient languages. Besides the hypothesis of *thinking for speaking*, the hypothesis of *seeing for speaking* proposed by von Stutterheim et al. (2012) addresses the relation between the grammaticalized concept and visual attention. According to the hypothesis of *seeing for speaking*, when a language has a particular grammaticalized concept and speakers of this language frequently express it, then attention will be more focused on the visual features related to this grammaticalized concept. In contrast, when a language does not have the same grammaticalized concept and the concept must be encoded in lexical or phrasal constituents and speakers of this language do not express this concept frequently, they will not pay attention to the visual features in that situation, or do so to a lesser degree.

Motion events are involved in a Figure's change of location as time goes on. Hence, besides spatiality, temporality also plays a fundamental role in construing

motion events. The concept of time in Klein's (1994, 2009) framework is structural in nature. There are different time spans, and the relations between these time spans may be such that a time span can occur before/after another one or partially (or wholly) overlap another one. There are three important time spans in temporal analysis. They are TU, Tsit, and TT, of which TT plays an essential role in linking lexical contents to the time axis and defining aspect and tense. Aspect is the relation between TT and Tsit, while tense is the relation between TT and TU. According to the presence or absence and the place of TT-contrasts, lexical contents can be classified into 0-state, 1-state, and 2-state lexical contents. The different types of lexical contents interact with the aspect, expressed either morphologically or lexically, in different languages.

Many empirical studies on motion events are based on the frameworks proposed by Talmy (1985, 2000), Slobin (1994, 2004, 2006), and Klein (1994, 2009). As for the studies on SLA in the domain of motion events, the results do not show a clear and consistent picture. There are studies showing that L1 language-specific *thinking for speaking* plays a limited role in L2 in describing motion events (Cadierno, 2004; Cadierno & Ruiz, 2006; Navarro & Nicoladis, 2005), between speech and gesture (Kellerman & Van Hoof, 2003; Stam, 2001), and in the domain of categorizing objects (Athanasopoulos, 2006, 2007). However, there are also studies showing that L1 language-specific *thinking for speaking* plays an important role in L2 (Cadierno, 2010; Carroll et al., 2012; Daller et al., 2011; Flecken et al., 2015; Hendriks et al., 2008; Hendriks & Hickmann, 2015; Negueruela et al., 2004; Schmiedtová, 2013; Schmiedtová et al., 2011; Schmiedtová & Sahonenko, 2008; von Stutterheim, 2003; von Stutterheim & Carroll, 2006). Moreover, there are studies showing the convergence between L1 and L2, especially in bilingual subjects (Brown & Gullberg, 2008; Bylund, 2009; Bylund & Jarvis, 2011; Hohenstein et al., 2006).

Chapter 3 Introduction of the Experiment

This chapter introduces the methodology of the experiment. First, the linguistic and social background of the participants in each language group will be presented, followed by information about the apparatus and the materials used to elicit linguistic descriptions. After that, the experiment procedure and the encoding of the linguistic and eye-tracking data as well as the speech onset times will be described. Finally, the statistical methods used for the analysis will be introduced.

3.1 Participants

The participating L1 German speakers (N=20) were students at the University of Heidelberg (Germany). They were between the age of 20 and 32. Eleven female and nine male German speakers participated in the experiment. The L1 English speakers (N=21) were participants in a summer school at the University of Heidelberg. Their age ranged from 18 to 35. Thirteen females and eight males took part in the experiment, of whom ten were from the UK, seven from the USA, three from Canada, and one from Ireland. Five participants reported that they had an intermediate level of German, while the remaining sixteen participants had only basic or no knowledge of German. There were twenty L1 Chinese participants, sixteen of whom were participants in a summer school at the University of Heidelberg, three were students who had just enrolled in the university, and one was a visiting scholar from China on her first day in Heidelberg. Six participants reported having an intermediate level of German, while the remaining fourteen participants had limited or no knowledge of German. Twelve Chinese females and eight males participated in the experiment; their ages were between 19 and 32. The recordings for the L1 English and L1 Chinese participants who attended the summer school were carried out during their first few days in Heidelberg. We wanted to make sure that the selected native speakers of German and English were as monolingual as possible or at least their foreign language proficiency was limited.

The participating L1 German-L2 Chinese speakers (N=24) were students at the University of Heidelberg. They were between the age of 22 and 38. Thirteen females and eleven males took part in the experiment. Twenty-three of the participants were majoring in Chinese Studies and one was majoring in Physics. Twenty-one participants

had attended HSK [20] and three participants [21] reported no HSK scores. Nine participants had achieved HSK Level V; nine participants had attained HSK Level VI; two participants had reached HSK 8; one participant had HSK 7[22]. The average time spent learning Chinese was 7.6 years and their average stay in China was 1.9 years. Six participants had attended Chinese courses since high school. Two participants had finished their Bachelor's studies in China and one participant was undertaking his Master's studies in China at the time of the recording. They were senior Bachelor's and Master's students, academic staff in Chinese Studies, a teacher for German-Chinese Translation, and a German-Chinese tourist guide.

The participating L1 German-L2 English speakers (N=20) were students at the University of Heidelberg. Their ages ranged from 21 to 30. Eleven females and nine males participated in the experiment. Nineteen of them were majoring in English and one of them in Conference Interpreting. Their average time learning English was 8.7 years and their average stay in English-speaking countries was 10.3 months. They reported their English levels to be from very good to excellent. They were Bachelor's and Master's students, a lecturer in the English Department, an English course teacher, and a conference interpreter.

3.2 Apparatus

The apparatus included a RED 500 eye tracker, a webcam, and a monitor. The RED 500 eye tracker developed by SMI (SensoMotoric Instruments GmbH) was used to record eye movements. It was attached to the monitor. The gaze point sampling rate

[20] Hanyu Shuiping Kaoshi (HSK) is the Chinese proficiency test, which is an international standardized language proficiency test for non-native Chinese speakers. It is administered by Hanban (汉办), affiliated institutions of the Ministry of Education of the People's Republic of China. The old HSK (before 2009) had 11 levels; HSK 1-3 were regarded as Basic level, HSK 3-5 as Elementary level, HSK 6-8 as Intermediate level, and HSK 9-11 as Advanced level. Because the old HSK had too many levels and it was hard to compare it with the standard exams in other countries, the HSK was changed in 2009. The new HSK is composed of six levels: HSK Level I, HSK Level II, HSK Level III, HSK Level IV, HSK Level V, and HSK Level VI, of which HSK Level V and VI correspond to the C1 and C2 Level of the Common European Framework of Reference (CEFR), respectively (see http://www.chinesetest.cn).

[21] These three participants had not taken part in the HSK in the past years. Two of them were PhD students in Chinese Studies and were also academic staff who actively spoke Chinese as lecturers or researchers. One participant had learned Chinese for over ten years and stayed in China for more than one year. These three participants spoke perfect Chinese. Based on these observations, they were considered as advanced learners in this study, even though they had no HSK scores.

[22] These three participants had taken the old HSK before 2009. According to the old HSK, HSK 7 and HSK 8 were regarded as Intermediate level. But, in fact, two participants were in the last year of their PhD studies in the Chinese Faculty at the time of the recording. In addition, they were also academic staff who worked as lecturers or researchers in the field of Chinese Studies. One participant was a Master's student. Since high school, she had been learning Chinese for over nine years. All of them spoke perfect Chinese. So these three participants were also qualified as advanced learners.

was 500 Hz and the gaze point tracking accuracy was 0.4 degrees. The TFT monitor was 20" in size and the participants sat at a distance of about 50 cm to 80 cm from the screen. The eye tracker recorded the participants' binocular eye movements and took head movements into account during the experiment. The audio data was recorded by a webcam. The eye-tracking data and the audio data was synced automatically.

3.3 Materials

The materials used were video clips[23] (N=70), of which 30 involved motion events and the remaining 40 served as filler items. Each video clip was about six to eight seconds long. Each video clip was followed by a blank screen, which lasted about eight seconds. The blank screen served to provide the participants with enough time to make sure they could complete their descriptions. All of the video clips showed daily events that were not connected with each other contextually.

The 30 motion events consisted of three types of self-propelled locomotion: Type A (N=10) showed an entity in motion (for example, a person or a bus) moving along a short trajectory towards a highly evident endpoint marked by an object (for example, a car or a bus stop); Type B (N=9)[24] showed an entity in motion (such as a car) moving along a long trajectory towards a possible endpoint (such as a village); Type C (N=10) showed an entity in motion (such as a dog) crossing a boundary defined by an object (such as a house or a garage). For Types A and B, the video clips stopped playing before the moving entity reached the endpoint, while the Type C video clips showed the moving entity reaching the endpoint; they actually crossed the boundary.

The 40 filler items were composed of video clips showing static scenes (for example, a dog sitting on the ground), activities (for example, a woman playing the piano), causative events (for example, a man making a paper airplane), and the slow process of a change of state (for example, a smoking cigarette). The filler items served to distract the speakers' attention away from the goal-oriented motion. They were not

[23] The video clips were filmed and edited by the research group at the Institute of German as a Foreign Language at the University of Heidelberg. Some of the related researches are Beyer (2016), Carroll et al. (2012), Flecken et al. (2014, 2015), von Stutterheim & Carroll (2006), and von Stutterheim et al. (2012, 2017).

[24] In fact, ten video clips were prepared for Type B. But due to technical problems, one video clip (a woman rollerblading on the road probably towards a white building) could not be displayed during the experiment for four participants in the L1 German-L2 Chinese group. In addition, ten participants described it as a non-motion event (e.g., a woman playing outside, or a woman doing exercise) and two participants did not know how to describe it. So the data for this video clip was not sufficient for analysis in L1 German-L2 Chinese. In order to avoid item differences among different language groups, the language and eye-tracking recordings of this video clip were not analyzed.

analyzed in the present study, since they were not relevant to the research questions of this dissertation.

The 70 video clips were grouped into four pseudo-randomized lists. In each list, two filler items followed one critical item (of Type A, B, or C). Therefore, the sequence in each list was: critical item 1, filler 1, filler 2, critical item 2, filler 3, filler 4, critical item 3, filler 5, filler 6, etc. The purpose of this sequence was to avoid any possible influence of previous critical items on subsequent critical items. In each language group, five participants each (or six participants in the L1 German-L2 Chinese language group) were assigned the same list. For example, five participants in a language group saw list 1, the next five participants saw list 2, the next five saw list 3, and the last five participants saw list 4. This way of assigning different lists to the participants was intended to reduce possible bias effects that might be caused by the order of event types displayed among the speakers.

3.4 Procedure

Before the experiment, the participants were told to read the following instructions:

> *You will see a set of 70 video clips showing everyday events which are not in any way connected to each other. Before each clip starts, a blank screen with a cross will appear. Please focus on this cross, since this allows us to proceed to the next video clip. Your task is to tell "what is happening", and you may begin as soon as you recognize what is happening in the clip. It is not necessary to describe the video clips in detail (such as 'the sky is blue'). Please focus on the event only.*

Different language groups read the instructions in different languages. Instructions for L1 speakers were written in their native language, while instructions for L2 speakers were written in the respective L2.

After reading the instructions, there was a training session in the form of a presentation. The experiment instructor presented the experiment process in Chinese, English, or German for the corresponding language groups. Therefore, Chinese was the only language during the experiment for the L1 and L2 Chinese groups, while English was the only language for the L1 and L2 English groups and German for the L1 German speakers. The purpose of speaking the testing languages for the different groups was to

attempt to avoid activating any knowledge of an additional language[25]. In this session, the participants were told that a calibration would be performed at the beginning. After that, a cross would appear on the blank screen. They were asked to fixate on this cross and were informed that after the fixation, each video clip would be displayed automatically. The video shown during the training session included six video clips, which covered three types of motion event and filler items. These six video clips did not appear again in the experiment. During the training session, the participants could ask questions and the instructor could give feedback. In order to achieve optimal results from the eye tracking, the participants were asked to keep their eyes on the screen and avoid body movement.

After the training session, but prior to the experiment, a calibration was conducted. The participants were asked to follow the movement of a white dot on the screen so that the eye tracker could measure their gaze. If the initial calibration was not successful, the recordings would be excluded. After the calibration, the video clips were shown and the participants began to describe "what is happening" either in L1 or L2. They had to fixate on the cross shown on the screen in order for the next video clip to be displayed, and they had no means of controlling the sequence of the video clips (i.e., they could not pause or replay the video clips). No recalibration was necessary during the experiment since the eye tracker adapted automatically to the eye positions. The experiment lasted about twenty minutes. After the eye-tracking experiment, the participants were told to fill out a questionnaire about their linguistic and educational background, which took about five minutes to complete.

3.5 Data coding

The linguistic recordings, the eye-tracking data, and the speech onset times (SOTs) were evaluated for each language group and compared across language groups in this work.

The linguistic recordings were transcribed and checked by native speakers. The checked transcriptions were then categorized according to each video clip and coded with respect to the spatial and temporal category. The spatial category included manner

[25] Kersten et al. (2010) showed that in performing the task of motion event categorization (i.e., categorizing novel, animated objects and events according to the manner of motion), the behavior of Spanish-English bilinguals moved towards English patterns when they were instructed in English, while the behavior moved towards Spanish patterns when they were instructed in Spanish.

verbs, motion verbs, as well as the fine-refined spatial concepts derived from Figure, Ground, or speaker for the verbs and adjuncts (see Chapter 4), while the temporal category included the aspectual markings used (such as imperfective or perfective aspect in grammaticalized or lexical items, see Chapter 5). Since this work is only concerned with motion events, non-motion event descriptions were marked and excluded from the data. The examples that appear in this dissertation are specified with respect to language groups (L1 Ger-L2 Chn means German learners of Chinese and L1 Ger-L2 Eng means German learners of English), participants in this group (e.g., P2 means the second participant in this language group), and video types (e.g., A06 means the sixth video clip of Type A).

Gaze movement was recorded during the time each video clip was displayed. In order to extract the fixation data on the areas of interest (AoI) from the eye tracking, two AoIs (the moving entity and the endpoint) were defined for this analysis. The definition of the AoI for the moving entity was conducted on a frame-by-frame basis since the entity was moving along a trajectory, which resulted in a change of size over time. As a consequence, the AoI for the moving entity changed slightly over time. For example, when an entity was moving away, its AoIs would become smaller over the course of time. In contrast, the AoI for the endpoint was fixed in each video clip since the endpoint was usually still and unmovable. However, the size of the AoIs differed slightly in different video clips because the area of the endpoint varied in different video clips (for example, a village vs. a car). As an example, Figure 3.1 shows the AoIs for the moving entity and the endpoint, respectively:

Figure 3.1: Screenshot with AoIs for entity and endpoint in a Type A video clip (a woman walking towards a car)

Fixations within the AoIs were calculated using the NYAN[26] software, which was developed specifically for the analysis of dynamic stimuli and language production. NYAN uses an area-based algorithm for calculation, which only identifies the gaze points within the specified areas (AoIs) as fixation points and those points outside the specified areas as saccade points. A fixation can be recognized when a sample is composed of a minimum of six screen pixels and has a maximum deviation of 25 screen pixels. In this analysis, the total fixation counts and durations in the AoI, and the fixation counts and durations before the SOTs in the AoIs will be analyzed and compared among the speakers of the different language groups.

For each video clip, the recordings for the language data, the eye-tracking data, and the display of stimulus were synchronized. That is, the SOTs were time-locked to the stimulus onset. With the help of the scripts written by Dipl.-Ing. Takara Baumbach[27], the video data was converted into audio data, which was time-locked to the stimulus onset. The audio data was then imported into WavePad. According to the sound waves presented in WavePad, the SOTs were determined for each video clip and each participant. It is important to note that the participants themselves decided when they would start to talk, so the SOTs were not fixed.

3.6 Statistical methods

To compare the spatial and temporal concepts encoded among speakers of different language groups, i.e., for the comparisons between L1 and L2 speakers as well as between the two L2 groups, the Chi-square test was applied. The Chi-square test aims to test whether two categorical variables are related. The results of Chi-square analysis show the differences between the data in the sample and the independence hypothesis. Our analysis adopted the following convention: If the p-value is lower than 0.05, the difference is significant; if the p-value is higher than 0.05, the difference is not statistically significant.

To comparea the SOTs, fixation counts, and durations for the moving entity and the endpoint among the different language groups, one-way ANOVA was conducted in order to find out whether the means in the different groups were equal. If the means were equal ($p>0.05$), this indicated that the language groups did not differ from each

[26] NYAN® Eye Tracking Data Analysis Suite 2.0 is developed by Interactive Minds GmbH in Germany.

[27] Mr. Baumbach was the technical support in the Language and Cognition Lab of Heidelberg University. We would like to thank him for providing the scripts.

other in this regard; if the means were not equal ($p<0.05$), this indicated that there was a significant effect. Post-hoc tests were used to find out which means exactly differed from others (i.e., which language group differed from others). As for the statistical analysis of attention given to the areas of interests, univariate ANOVA was conducted to test the main and the interaction effect of time window and language groups.

Chapter 4 Spatial Analysis

4.1 Introduction

This chapter presents the analysis of the spatial concepts encoded in two learner groups: L1 German-L2 Chinese and L1 German-L2 English. First, empirical studies on motion events and cognition will be reviewed and then the spatial categories in the presented work will be defined. In what follows, we will first illustrate linguistic expressions encoding the conceptual components in Chinese, English, and German, respectively. Next, spatial analysis (regarding linguistic expressions used and motion events cognition) in L1 German-L2 Chinese and L1 German-L2 English speakers, respectively, will be conducted. Then we will compare the linguistic forms and the meaning in those two learner groups and finally provide a summary of this chapter.

Based on the frameworks of Talmy (1985, 2000) and Slobin (1994, 1996, 2000, 2003, 2004, 2006, 2008), numerous studies on motion events and cognition have been carried out. Generally speaking, cross-linguistic studies on motion events mainly center on two topics: Studies focus either on the lexicalized concepts, i.e., Manner or Path encoding (e.g., Berman & Slobin, 1994; Brown & Gullberg, 2008; Cadierno, 2004; Cadierno & Ruiz, 2006; Carroll et al., 2012; Hendriks & Hickmann, 2011; Hickmann & Hendriks, 2010) or on the grammaticalized concepts, i.e., the grammatical category of aspect (e.g., Athanasopoulos & Bylund, 2013; Bylund, 2009; Bylund & Jarvis, 2011; Bylund et al., 2013; Schmiedtová & Flecken, 2008; Schmiedtová et al., 2011; von Stutterheim et al., 2012, 2017; von Stutterheim & Nüse, 2003). Each topic can be approached following three lines of research: namely, studies on verbal, co-verbal, and non-verbal behaviors (see the overview in Bylund & Athanasopoulos, 2015a). Verbal tasks are overt speech production (e.g., to describe a situation), co-verbal tasks focus on investigating gestures or visual attention (which are concurrent with speaking), and non-verbal tasks include similarity judgments, remembering, reasoning, and attempting to examine the cognitive processes without overt speech production.

In reference to the studies on lexical concepts (Manner and Path encoding), verbal tasks have shown that speakers of S-languages tend to express Manner information since Manner is readily encoded in main verbs, while speakers of V-languages tend not to express Manner information. This is because the information about Manner in V-languages is encoded in peripheral constructions. Speakers of S-

languages tend to use a diverse manner lexicon and interpret unknown verbs as manner verbs, more so than speakers of V-languages such as Spanish, Japanese, and Greek (e.g., Berman & Slobin, 1994; Brown & Gullberg, 2008; Hendriks & Hickmann, 2011; Papafragou & Selimis, 2010; Slobin, 2004). Studies on co-verbal tasks have shown that speakers of S-languages might place Path gestures on the satellites, whereas speakers of V-languages may express Manner information through gestures (e.g., Brown & Gullberg, 2008). Studies on the expression of motion events and the placement of gestures of L1 Spanish-L2 English speakers have shown that some L2 speakers put the path gestures on the satellites, while others placed path gestures on the motion verb (e.g., Kellerman & Van Hoof, 2003; Stam, 2001). In reference to non-verbal tasks, Gennari et al. (2002) investigated how English and Spanish speakers performed recognition and similarity judgment tasks under three conditions in motion events (the speakers verbalized the scene; the speakers watched the scene without verbalization; the speakers watched the scene while uttering a nonsense syllable). Differences were found between groups in the similarity judgment task when speakers formerly verbalized the scene: Spanish speakers were prone to choose same-path alternates compared to English speakers, who showed no preference. The authors argued that cross-linguistic differences in cognition occurred when speakers verbalized the scene before categorization. In contrast, the non-verbal memory task in Papafragou et al. (2008) revealed no language-related preference between speakers of English (an S-language) and Greek (a V-language).

Regarding studies on grammatical concepts, there are several studies on the relationship between event conceptualization and grammaticalized aspect (Athanasopoulos & Bylund, 2013; Bylund, 2009; Bylund & Jarvis, 2011; Bylund et al., 2013; Flecken et al., 2014; Schmiedtová & Flecken, 2008; Schmiedtová et al., 2011; von Stutterheim et al., 2012; von Stutterheim & Nüse, 2003). These studies show that speakers of aspect languages[28] (e.g., English, Arabic, Russian, Spanish) tend to segment the situation into different phases and defocus the endpoint in the verbalization, while speakers of non-aspect languages (e.g., German, Swedish, Afrikaans) tend to take a holistic perspective when viewing the situation and focus more on the endpoint. Note

[28] Aspect languages are those languages in which the imperfective or progressive aspect is grammaticized (for more details, see Chapter 5). Speakers of these languages use this grammatical concept frequently, while non-aspect languages are those languages in which the imperfective or progressive aspect is not grammaticized. There are lexical or periphrastic forms that encode an event as ongoing in non-aspect languages. However, speakers of these languages use them rarely (von Stutterheim et al., 2012).

that cross-linguistic studies on endpoint encoding only reflect the preference of speakers of different languages. This is not an absolute pattern. It is grammatically acceptable for speakers of aspect languages to encode the endpoint and for speakers of non-aspect languages not to encode the endpoint. The reason why speakers of aspect languages focus on the ongoing nature of an event resides in the grammar. This grammatical concept indicates to speakers of those languages that they should pay attention to the event's internal temporal constituency, which is prominent in conceptualization (Carroll et al., 2004; von Stutterheim & Nüse, 2003). So speakers of aspect languages are sensitive to ongoingness. By contrast, speakers of non-aspect languages do not focus on the event's internal temporal constituency via their grammar, but are rather prone to taking a holistic, endpoint-oriented perspective. From this endpoint-oriented perspective, the event presented seems to have a point of completion. Additionally, the presence of verb particles in some of the non-aspect languages such as German, Dutch, and Swedish seems to make holistic perspectives more prominent, since verb particles can be applied on the verb stems in a productive way and together with the verb stems, they convey result-driven meanings (Bylund & Jarvis, 2011; von Stutterheim & Nüse, 2003). Second language studies within this approach have shown that when advanced L2 speakers use a foreign language that differs from their L1 in terms of the presence or absence of grammatical aspects, those L2 speakers are more likely to turn to their L1's conceptualization pattern (e.g., Carroll et al., 2012; Schmiedtová et al., 2011; von Stutterheim, 2003).

Studies on co-verbal tasks on endpoint encoding mainly investigate the allocation of attention patterns measured by eye-tracking across different languages. Visual attention serves as a window that allows us to see the processes of conceptualization. The findings show that speakers of non-aspect languages (e.g., Dutch, German) differ from speakers of aspect languages (e.g., English, Arabic) in their visual attention to different entities in scenes that require them to describe motion events. Speakers of non-aspect languages (e.g., German) tend to have more fixations on the endpoint than speakers of aspect languages (e.g., English, Arabic), whereas speakers of aspect languages tend to fixate more on the moving entity in motion and pay significantly less or no attention to the endpoint (Schmiedtová et al., 2011; von Stutterheim & Carroll, 2006; von Stutterheim et al., 2012). In addition, the duration of the fixation on the endpoint in a non-aspect language is longer than that in aspect languages (von Stutterheim et al., 2012). The results of eye-tracking studies have shown

that the tendency of speakers of aspect languages not to mention the endpoint cannot be interpreted as an appraisal of the endpoint's insignificance, but it rather seems that those speakers tend not to look at the endpoint.

Studies on endpoint encoding have also been extended to non-verbal tasks as a way to underscore the impact of grammar on event conceptualization (Athanasopoulos & Bylund, 2013; Bylund & Athanasopoulos, 2014; Bylund et al., 2013; Flecken et al., 2014; von Stutterheim et al., 2012). Bylund et al. (2013) examined this non-verbal endpoint preference by creating a memory-based triad matching task. The findings from this task showed that the behavior matching of the target with an endpoint alternate by speakers of Afrikaans (a non-aspect language) was on par with that of speakers of Swedish (also a non-aspect language) and differed from that of speakers of English (an aspect language). In addition, it has been found that speakers of Afrikaans who spoke English frequently were prone to selecting the alternate without an endpoint, whereas speakers of Afrikaans who spoke English less frequently were more likely to choose the alternate with an endpoint. Furthermore, Athanasopoulos and Bylund (2013) conducted experiments that included verbal descriptions and two similarity judgment tasks. These experiments centered on memory-based (with and without verbal interference) and online triad matching tasks of goal-oriented motion events by speakers of Swedish and English in order to examine whether the grammatical aspect affects cognition. The findings showed a difference between the two language groups in solving the memory-based (without verbal interference) triad matching task, while no group difference was found for the online triad matching task. The difference in the memory-based task disappeared under the condition of verbal interference. The authors thus argued that language might attenuate the perceptual processes rather than shape cognition. Flecken et al. (2014) compared the fixation patterns obtained in an event description task and a non-verbal task between speakers of Modern Standard Arabic (an aspect language) and German (a non-aspect language). The results showed the same tendency under both conditions: In comparison with speakers of Modern Standard Arabic, speakers of German tended to encode the endpoint linguistically and fixated more on the endpoint in verbal and non-verbal tasks.

This study analyzes both the lexicalized and grammatical concepts in L1 German-L2 Chinese and L1 German-L2 English. Particularly, this section focuses on the lexicalized concepts, while the grammatical concepts will be discussed in Chapter 5. More specifically, in this section motion event descriptions (i.e., verbal tasks) and

visual attention patterns (i.e., co-verbal tasks) will be analyzed in L1 German-L2 Chinese and L1 German-L2 English speakers, leaving non-verbal tasks such as similarity judgment and memory-based tasks (i.e., non-verbal tasks) to future research. The following research questions were developed:

(1) Whether and to what extent have L2 speakers adjusted their L1's habitual encoding patterns of Manner and Path to the constraints given in the L2?

(1a) Whether and to what extent have L1 German-L2 Chinese speakers learned to express the information of Manner and Path in Chinese?

(1b) Whether and to what extent have L1 German-L2 English speakers learned to express the information of Manner and Path in English?

(1c) Given that the source language in both learner groups is German, whether and to what extent do L1 German-L2 Chinese speakers differ from L1 German-L2 English speakers in Manner and Path encoding?

(2) Whether and to what extent have L2 speakers adjusted their L1's habitual patterns in visual attention to the L2?

(2a) In comparison with L1 Chinese speakers, whether and to what extent have L1 German-L2 Chinese speakers adjusted their L1's habitual patterns in visual attention to Chinese?

(2b) In comparison with L1 English speakers, whether and to what extent have L1 German-L2 English speakers adjusted their L1's habitual patterns in visual attention to English?

(2c) Given that the source language in both learner groups is German, whether and to what extent do L1 German-L2 Chinese speakers differ from L1 German-L2 English speakers in the habitual patterns in visual attention?

In order to address these questions, we will first define the spatial categories regarding the spatial concepts and then introduce the linguistic expressions encoding these spatial concepts in each language. Following that, we will report on the results of the spatial analyses conducted for L1 German-L2 Chinese and L1 German-L2 English, respectively. The spatial analyses mainly deal with the comparison of the linguistic expressions used, event unit formation, and the allocation of visual attention in L1s and L2s. Given that the source language in both learner groups was the same, i.e., German, a comparison of the linguistic forms used, the spatial concepts encoded, and the fixation patterns was conducted between speakers of L1 German-L2 Chinese and L1 German-L2 English. Finally, the findings from this section will be summarized.

4.2 Spatial categories

For the conceptualization of motion events, the conceptual component Manner is always accompanied by motion. Thus, the term *manner verbs* in this work indicates verbs that encode both pieces of information: Manner and Motion (e.g., to run), while the term *motion verbs* indicates verbs that only encode Motion information (e.g., to move). Based on the categorization of Beyer (2016), Carroll et al. (2012), and Flecken et al. (2015), the path of motion is characterized as being derived from the Figure (i.e., the moving entity), the Ground objects, and the observer. English examples are given for each spatial category in order to facilitate understanding. Detailed linguistic expressions in Chinese, English, and German that are encoded as the Path of motion will be presented in sections 4.3.1, 4.3.2, and 4.3.3, respectively.

1) Path of motion derived from the Figure

Figure-related concept: The direction is defined in terms of the body axis of the Figure. Example: 'turn' in the sentence 'a woman is turning around a street corner'.

2) Path of motion derived from the observer

Deictic concept: The direction is defined in terms of the observer or the speaker, i.e., the direction of the motion towards or away from the speaker. The observer or the speaker is an external anchor point, which could be described as Origo (Becker, 1994). Examples: 'come' in the sentence 'a bus is coming towards a bus stop' and 'down' in the sentence 'a woman is walking down the street'[29].

3) Path of motion derived from Ground objects

a. *Objected-related* concept: This relates to two subcategories of conceptualization patterns for a Ground object at the reference point. One subcategory is that the contours of the Ground object play a role in encoding this concept. That is, this concept is characterized according to the contours of the Ground object, which could be further distinguished with regard to the sphere of the contours between bounded and unbounded contours. Examples of bounded contours are 'through' in the sentence 'a lady is walking through a park', 'cross' in 'a lady is crossing a square towards a telephone booth', while an example of unbounded contours is 'along' in the sentence 'a car is driving along a road'. The other subcategory is that the contours of

[29] Apart from the use of the prepositions *up* and *down* denoting motion in the vertical direction, the English preposition *down* can also be used when the road is long and there is no motion in the vertical direction. In this case, *down* indicates a motion away from the speaker. In this study, the video clips used did not show any obvious motion in the vertical direction; rather, they showed roads with a long trajectory. Therefore, *down* expresses the deictic meaning, with its direction defined with respect to the Origo, but not the meaning related to the vertical direction.

the Ground object play no role in encoding this concept. The Ground object is then conceptualized as a reference point, such as ‘pass’ in the example ‘a car is passing a village’.

b. *Axial* concept: This describes motion in the vertical direction, e.g., ‘up’ in the sentence ‘there is a man climbing up a ladder’.

c. *Direction*: This is defined as the motion direction related to the place at the endpoint, such as ‘towards’ in the sentence ‘someone is walking towards a bus stop’.

d. *Goal-oriented* concept: This indicates that the Figure reaches the endpoint, such as ‘to’ in the example ‘a woman is walking to a green car’.

e. *Boundary-crossing* concept: This indicates that the Figure reaches the endpoint and crosses the boundary defined by the Ground object, such as ‘enter’ in the example ‘a woman is entering a building’.

4) Location

Location indicates the region where a motion takes place. As Flecken et al. (2015, p. 105) pointed out, a motion event can be conceptualized as “successive points of displacement, as changes of location of the moving entity”. Example: ‘on’ in the sentence ‘there is a truck driving on the road.’

These different categories of spatial cognition are used as a basis for the data analysis, which will be presented in sections 4.4, 4.5, and 4.6.

4.3 Linguistic expressions encoding conceptual components in Chinese, English, and German

The conceptual components, i.e., Figure, Ground, Motion, Path, Manner, and Cause, appear in different linguistic expressions in different languages. Generally speaking, Figure is typically encoded in a nominal phrase and is usually the subject of the sentence, while Ground is encoded in nominal phrases or adpositions. Motion and Manner are mainly packaged in verbs. Besides being encoded in verbs, Manner is also encoded in adverbs and adjectives, while Path is expressed in verbs, particles, case markings, and some adpositions. Since there were no caused motion events in the data used for this study, the conceptual component Cause is not considered. In the following, the morphological and lexical elements that encode these conceptual components in Chinese, English, and German, respectively, will be presented.

4.3.1 Chinese

Chinese lacks morphological inflections and case markings to indicate grammatical relations between linguistic items. Instead, the meanings and functions expressed by case markings in other languages are expressed via analytic means or unmarked items in Chinese (Liu, 2008, p. 42). Before we introduce the linguistic expressions that encode different conceptual components in Chinese, we will first look at the syntactic and semantic features of serial verb constructions (hereafter called SVCs), since SVCs are language-specific in Chinese. L2 Chinese speakers have to unravel these when they speak Chinese.

4.3.1.1 Serial verb constructions

In Chinese, the conceptual components Motion, Manner, and Path are frequently encoded in verbs. These verbs can appear simultaneously in one sentence, which forms a language-specific structure in Chinese, i.e., a serial verb construction. SVCs occur not only in Chinese, but can also be found in languages in four other areas of the world, e.g., in West Africa, Central America, South-East Asia, and Oceania (Müller & Lipenkova, 2009, p. 236). According to Li and Thompson (1981, p. 594), SVCs in Chinese refer to "a sentence that contains two or more verb phrases or clauses juxtaposed without any marker indicating what the relationship is between them." SVCs in Chinese are represented as follows (Li & Thompson, 1981, p. 594) (V and NP represent verb and noun phase respectively, and NP in parentheses means they can be omitted):

(NP) V (NP) (NP) V (NP)

SVCs: Syntactic features

Syntactically, the VPs (i.e., verb phrases) in a sentence share the same subject and the syntactic relations between VPs, such as coordination or subordination, are not marked explicitly. Example (VPs are underlined):

(1) 一个人骑着自行车拐进了一个走道里。(L1Chn P16, C09)

yī-ge rén <u>qí-zhe zìxíngchē</u> <u>guǎi-jìn-le yī-ge zǒudào- lǐ</u>.

one-CL person ride-ZHE bicycle turn-enter-LE one-CL corridor-inside

A person is riding a bicycle and turning into a corridor.

In this example, there is no overt syntactic marker to mark the relations of the two VPs, i.e., *qi-zhe zìxíngchē* 'lit. ride-zhe bicycle' and *guǎi-jìn-le yī-ge zǒudào-li* 'lit. turn-enter-LE one-CL corridor-inside'.

Furthermore, VPs in SVCs may share the same direct object. Example:

(2) 他买报看。

ta mái bàozhǐ kàn.

he buy newspapers read

He buys a newspaper to read it.

In this example, *bàozhǐ* 'newspaper' is the object of the verbs *mái* 'buy' and *kàn* 'read'. The second sub-event (i.e., *kàn bàozhǐ* 'read newspaper') is the purpose of the first sub-event (i.e., *mái bàozhǐ* 'buy newspaper').

SVCs: Semantic relations

The meanings of SVCs are not merely a combination of the meanings indicated by different VPs. Semantic relations and context play an important role in determining the meaning of an SVC[30]. Following Li and Thompson (1981), there are four possible relations between sub-events in SVCs, which are summarized in Table 4.1.

Table 4.1: Possible semantic relations between sub-events in SVCs

Semantic Relation	Definition
(i) Consecutive	One event occurs after the other
(ii) Purpose	The first event is done for the purpose of achieving the second
(iii) Alternating	The subject alternates between two actions
(iv) Circumstance	The first verb phrase describes the circumstances under which the event in the second verb phrase or clause occurs

Note. Adapted from Li and Thompson (1981, p. 595)

One sentence may have more than one interpretation. It depends on the context

[30] According to Li and Thompson (1981), SVCs can be categorized as four groups: 1) two or more separate events; 2) one verb phrase or clause serving as the subject or direct object of another verb (i.e., the first VP may serve as the subject of the second VP, or vice versa, the second VP may serve as the direct object of the first VP), e.g., *wǒ xiǎng xué kāi chē* 'lit. I want-learn-drive car'. In this example, the second VP is the direct object of the first VP; 3) pivotal constructions (i.e., the NP serves as the direct object of the first verb and the subject of the second VP simultaneously), e.g., *wǒ jiànyì tā xué déyǔ* 'lit. I-advise-he-learn-German'; and 4) descriptive clauses (i.e., the direct object of the first VP is described by the following clause. Semantically, this descriptive clause could be transformed into the relative clause, both of which have the same English translation), e.g., *wǒ rènshi yī-ge déguó-rén huì jiǎng hànyǔ* 'lit. I know one-CL German-person can speak Chinese'. In this example, the first underlined phrase is the direct object of the VP and the second phrase further describes the first phrase.

which specific reading is intended. Example:

(3) 他开车听广播。

tā kāi chē tīng guǎngbō.

he drive car listen radio

He drives a car and listens to the radio. (i) and (iii)

He drives a car in order to listen to the radio. (ii)

He drives a car while listening to the radio. (iv)

In example (3), there are three different readings, which can be disambiguated according to the context.

In this study, direction verb compounds, a specific type of SVCs, are of great importance[31]. These are widely used in describing motion events in L1 and L2 Chinese. Due to their structural complexity, L2 Chinese speakers are susceptible to making grammatical mistakes. Hence, in the following, we will elaborate on their structural properties.

Directional verb compounds consist of consecutive verb stems, with the first component denoting the information on Manner or Cause and the second component denoting the direction of the motion, which can be schematized as follows[32]:

$V_{1\ manner/cause}$ - $V_{2\ direction}$

In directional verb compounds, the first component is typically a manner verb or a causative verb, whereas the second component is a path verb. As for the path verb, it can be one of the following three types:

The first type includes two deictic verbs, *lái* 'come' and *qù* 'go'. *lái* 'come'

[31] In Li and Thompson's (1981) work, directional verb compounds are a sub-type of resultative verb compounds in Chinese. There are two elements in resultative verb compounds: the first element indicating the action or process and the second element indicating the result caused by the first element. Based on the meaning of the second element, resultative verb compounds can be categorized into four groups: cause, achievement, direction, and phase (see also Chao, 1968; Packard, 2000). Directional verb compounds belong to the third group, with the second element expressing the direction caused by the first element. We agree with the opinion that directional verb compounds are a sub-type of resultative verb compounds. However, in order to highlight the function of the second verb in describing motion events, that is, expressing the direction of the motion, we use the term "directional verb compounds" instead of the superordinate term "resultative verb compounds".

[32] Li and Thompson (1981, p. 58) schematized directional verb compounds as: V1 - V2
displacement direction.

indicates a motion towards the speaker, whereas *qù* ‘go’ indicates a motion away from the speaker. Each of these two deictic verbs can occur alone or with manner verbs.

(4) The deictic verb *lái* ‘come’ occurs alone:

公交车来了。

gōngjiāochē lái le.

bus come LE

The bus came.

(5) The deictic verb *qù* ‘go’ occurs with a manner verb:

公交车开去了车站。

gōngjiāochē kāi-qù le chēzhàn.

bus drive-go LE bus stop

A bus drove towards the bus stop (away from the speaker).

The second type includes the following non-deictic path verbs[33], all of which have directional meaning. In the glossary, the independent meaning of each path verb and the corresponding directional meaning are listed.

shàng ‘ascend-moving upward’

xià ‘descend-moving downward’

jìn ‘enter-moving from outside to inside’

chū ‘exit-moving from inside to outside’

guò ‘pass/cross-passing through a point or moving from one side to the other side’

huí ‘return-moving to the original place’

qǐ ‘rise-moving upward’[34]

dào ‘arrive-moving to a place at goal’

[33] There is no consensus on the total number of path verbs in this type (see Chao, 1968; Liu, 1998; Zhu, 1982). This study considers the path verbs on which most linguists agree, as listed above, since these path verbs also occur in the collected data. In addition, path verbs like *shàng* ‘ascend’, *jìn* ‘enter’, etc. are labeled as non-deictic path verbs in order to distinguish them from the deictic path verbs *lái* ‘come’ and *qù* ‘go’.

[34] There are two path verbs that correspond to the meaning of ‘moving upward’ in English: *shàng* and *qǐ*. The meaning of *shàng* focuses on the goal, while *qǐ* focuses on the source.

Each of the above-listed non-deictic path verbs can function as a full verb in a sentence or occur with manner verbs. Take *jìn* 'enter' as an example:

(6) Non-deictic path verb occurs alone:

他进了房间。

tā jìn le fángjiān.

he enter LE room

He entered the room.

(7) Non-deictic path verb occurs with manner verb:

他走进了房间。

tā zǒu-jìn le fángjiān.

he walk-enter LE room

He walked into the room.

The third type is a combination of non-deictic path verbs of the second type and deictic path verbs of the first type. The word order is that non-deictic path verbs of the second type precede deictic verbs; this order is irreversible[35]. Compared to the meaning of non-deictic path verbs of the second type, this type characterizes the motion direction with respect to the speaker. Table 4.2 shows the combinations.

Table 4.2: Co-occurrence of non-deictic and deictic path verbs

	shàng 'ascend'	*xià* 'descend'	*jìn* 'enter'	*chū* 'exit'	*guò* 'pass, cross'	*huí* 'return'	*qǐ*[36] 'rise'	*dào*[37] 'arrive'
lái 'come'	*shàng lái* 'ascend come'	*xià lái* 'descend come'	*jìn lái* 'enter come'	*chū lái* 'exit come'	*guò lái* 'pass come'	*huí lái* 'return come'	*qǐ lái* 'rise come'	*dào.. lái* 'arrive.. come'
qù 'go'	*shàng qù* 'ascend go'	*xià qù* 'descend go'	*jìn qù* 'enter go'	*chū qù* 'exit go'	*guò qù* 'pass go'	*huí qù* 'return go'		*dào.. qù* 'arrive.. go'

[35] There is an exception with *dào* 'arrive'. The verb serials *lái dào* 'arrive towards the speaker' and *qù dào* 'arrive away from the speaker' are also possible in Chinese.

[36] *qǐ* 'rise' could not occur with *qù* 'go'.

[37] In contrast with other path verbs, *dào* 'arrive' has special syntactic features. This path verb requires nominal phrase immediately follow it. That is, when there are no deictic verbs, the construction '*dào* + Ground NP' is correct. And when there are deictic verbs followed, only the construction '*dào* + Ground NP + deictic verbs' is permissible. The construction without Ground NP is not acceptable (e.g. **dào* + deictic verbs), which is possible in other path verbs.

Each combination of non-deictic and deictic path verbs can occur alone or with manner verbs.

(8) Non-deictic and deictic path verbs occur alone:

他进房间去了。

tā jìn fángjiān qù le.

he enter room go LE

He entered the room (away from the speaker).

(9) Non-deictic and deictic path verbs occur with manner verbs:

他走进房间去了。

tā zǒu-jìn fángjiān qù le.

he walk-enter room go LE

He walked into the room (away from the speaker).

The combinations of manner verbs with these three types of directional verb compounds can be schematized as follows. Note that the first verb (i.e., the manner or causative verb) can be omitted in these three types.

Type I: $V_{1\ manner/cause}$ - $V_{2\ deictic\ path}$

Type II: $V_{1\ manner/cause}$ - $V_{2\ non\text{-}deictic\ path}$

Type III: $V_{1\ manner/cause}$ - $V_{2\ non\text{-}deictic\ path}$ - $V_{3\ deictic\ path}$

Depending on the property of the direct object, that is, on whether the direct object is an object NP (e.g., a book) or a place NP (e.g., a square), there are different positions where the direct object can be inserted in directional verb compounds, as shown by the following examples (object or place NPs are underlined):

Type I: $V_{1\ manner/cause}$ - $V_{2\ deictic\ path}$

a. Position of object NPs in directional verb compounds

(10a) $V_{1\ manner/cause}$ + object NPs + $V_{2\ deictic\ path}$

他拿了一本书来。

tā ná le yī-běn shū lái.

he take LE one-CL book come

He brought a book (towards the speaker).

(10b) $V_{1\ manner/cause}$ + $V_{2\ deictic\ path}$ + object NPs

他拿来了一本书。

tā ná-lái le yī-běn shū.

he take-come LE one-CL book

He brought a book (towards the speaker).

b. Position of place NPs in directional verb compounds

(10c) $V_{1\ manner/cause}$ + $V_{2\ deictic\ path}$ + place NPs (example 5 is repeated here)

公交车开去了车站。

gōngjiāochē kāi-qù le chēzhàn.

bus drive-go LE bus stop

A bus drove towards the bus stop (away from the speaker).

Type II: $V_{1\ manner/cause}$ - $V_{2\ non\text{-}deictic\ path}$

a. Position of object NPs in directional verb compounds

(11a) $V_{1\ manner/cause}$ + $V_{2\ non\text{-}deictic\ path}$ + object NPs

他拿出了一本书。

tā ná-chū le yī-běn shū.

he take-exit LE one-CL book

He took out a book.

b. Position of place NPs in directional verb compounds

(11b) $V_{1\ manner/cause}$ + $V_{2\ non\text{-}deictic\ path}$ + place NPs (example 7 is repeated here)

他走进了房间。

tā zǒu-jìn le fángjiān.

he walk-enter LE room

He walked into the room.

Type III: $V_{1\ manner/cause}$ - $V_{2\ non\text{-}deictic\ path}$ - $V_{3\ deictic\ path}$

a. Position of object NPs in directional verb compounds

(12a) $V_{1\ manner/cause}$ + object NPs + $V_{2\ non\text{-}deictic\ path}$ - $V_{3\ deictic\ path}$

他拿了一本书出来。

tā ná le yī-běn shū chū-lái.

he take LE one-CL book exit-come

He took out a book (towards the speaker).

(12b) $V_{1\ manner/cause}$ + $V_{2\ non\text{-}deictic\ path}$ + object NPs + $V_{3\ deictic\ path}$

他拿出了一本书来。

tā ná-chū le yī-běn shū lái.

he take-exit LE one-CL book come

He took out a book (towards the speaker).

(12c) $V_{1\ manner/cause}$ + $V_{2\ non\text{-}deictic\ path}$ + $V_{3\ deictic\ path}$ + object NPs

他拿出来了一本书。

tā ná-chū-lái le yī-běn shū.

he take-exit-come LE one-CL book

He took out a book (towards the speaker).

b. Position of place NPs in directional verb compounds

(12d) $V_{1\ manner/cause}$ + $V_{2\ non\text{-}deictic\ path}$ + place NPs + $V_{3\ deictic\ path}$ (example 9 is repeated here)

他走进房间去了。

tā zǒu-jìn fángjiān qù le.

he walk-enter room go LE

He walked into the room (away from the speaker).

There are two dimensions that make the syntactic features of directional verb compounds complicated: its three sub-types and its associated word order rules. When speakers express self-propelled motion events, they are prone to encoding place NPs as underway source, goal, or reference objects (see examples 10c, 11b, 12d). In contrast, when speakers express a caused motion event, they are more likely to encode object NPs (see examples 10a, 10b, 11a, 12a, 12b, 12c). Speakers can also choose not to

mention place NPs or object NPs (e.g., without mentioning the place NP *fángjiān* 'room', they could only use the serial verbs *zǒu-jìn-qù* 'lit. walk-enter-go' in example 12d). If speakers use place NPs or object NPs, it seems that the positions of object NPs are more flexible than those of place NPs in directional verb compounds. Object NPs can be inserted between manner verbs (V_1) and deictic verbs (V_2) (example 10a) or occur after deictic verbs (V_2) (example 10b); or, more flexibly, speakers can split manner verbs (V_1) and non-deictic verbs (V_2) (example 12a) or non-deictic (V_2) and deictic verbs (V_3) (example 12b), or leave the directional verb compound as an intact unit and use it after deictic verbs (example 12c) in Type III. There are no semantic distinctions in terms of the different positions of the object NPs in examples 12a, 12b, and 12c. In contrast, the positions of place NPs are restricted; they can only occur after deictic verbs in Type I (example 10c) and after non-deictic path verbs in Type II (example 11b). In addition, place NPs must be inserted between non-deictic path verbs and deictic path verbs in Type III (example 12d) (Chao, 1968; Kang, 2001, p. 287; Li & Thompson, 1981, p. 64; Liu, 1998, p. 34; Wu, 2011). Any other position would lead to grammatically unacceptable sentences. L1 German-L2 Chinese speakers frequently make such mistakes (for more details, see 4.4.1). The following examples are some grammatically incorrect sentences:

*这只猫慢慢地走进去了房间。(L1 Ger-L2 Chn P01, C06)

(13a) *zhè-zhī māo màn-màn-de zǒu-jìn-qù le fángjiān[38].
this-CL cat slowly walk-enter-go LE room
This cat walked slowly into the room.

corrected as:

(13b) zhè-zhī māo màn-màn-de zǒu-jìn le fángjiān qù.
this-CL cat slowly walk-enter LE room go
This cat walked slowly into the room.

*她拉着行李跑那个火车站进去。(L1 Ger-L2 Chn P03, C02)

(14a) *tā lā-zhe xínglǐ pǎo nà-ge huǒchēzhàn jìn-qù.
she pull-ZHE luggage run that-CL train station enter-go

[38] The mark * in this dissertation means that the expression is unacceptable.

She ran into the train station with luggage.

corrected as:

(14b) tā lā-zhe xínglǐ pǎo jìn nà-ge huǒchēzhàn qù.

she pull-ZHE luggage run enter that-CL train station go

She ran into the train station with luggage.

In the above examples, the place NP (i.e., *fángjiān* 'room', *huǒchēzhàn* 'train station') must be inserted between the non-deictic path verb (*jìn* 'enter') and the deictic path verb (*qù* 'go'). As shown, this means that examples 13a and 14a are grammatically unacceptable, whereas examples 13b and 14b are acceptable.

In describing a motion event, these three types of directional verb compounds are at the speaker's disposal. This is a matter of choice. However, there is one grammatical rule that must be obeyed: If non-deictic path verbs occur alone, that is, if no place NPs or object NPs are used, then deictic verbs must be added in order to form a grammatically correct construction (Lamarre, 2008, p. 79; Liu, 1998, p. 37; Qi, 1998, p. 180; Wu, 2011, p. 422; Ying, 1988, p. 260). In other words, non-deictic path verbs must be followed either by place NPs or object NPs, or by deictic verbs (see examples 15b and 15c). This is illustrated in the following (the manner verb *zǒu* 'walk' could be omitted in the examples):

(15a) *一个人走进。

*yī-ge rén zǒu-jìn.

one-CL person walk-enter

(15b) 一个人走进房间。

yī-ge rén zǒu-jìn fángjiān.

one-CL person walk-enter room

A person walks into the room.

(15c) 一个人走进来。

yī-ge rén zǒu-jìn-lái

one-CL person walk-enter-come

A person walks into it (towards the speaker).

The features of SVCs and the syntactic complexity of directional verb compounds in Chinese may pose challenges for L2 speakers (for more details, see 4.1.1). In the next part, the linguistic expressions used to encode different conceptual components in Chinese will be presented. These expressions include verbs, adpositions (including prepositions, postpositions, and circumpositions), and adverbs. We will take an in-depth look at these linguistic expressions used in Chinese. Note that we will only analyze those linguistic expressions that occur in the data used for this study and disregard other options that do not appear.

4.3.1.2 Verbs

The verbs used in Chinese are divided into four categories: motion verbs, manner verbs, path verbs, and other verbs:

Motion verbs

Without specifying the Manner and Path of a motion, motion verbs describe continuous movement. Chinese verbs such as 行 *xíng* 'move', 走 *zǒu* 'move', and 移动 *yí-dòng* 'move-move' describe such movement.

Manner verbs

Both Manner and Motion are expressed by manner verbs. Examples are 走 *zǒu* 'walk', 赶 *gǎn* 'hurry', 跨 *kuà* 'step' 驶 *shǐ* 'drive', 开(车) *kāi* (*chē*) 'drive (car)', 骑 *qí* 'ride', 跑 *pǎo* 'run', 散步 *sànbù* 'walk', 停(车) *tíng* (*chē*) 'park (car)', and 爬 *pá* 'climb'. As Beyer (2016) pointed out, 走 *zǒu* has different meanings in different contexts. First, 走 *zǒu* means a way of moving, i.e., moving on foot.

(16a) 一个女生往前走。(L1 Ger-L2 Chn P17, A04)

yī-ge nǚshēng wǎng qián zǒu.
one-CL girl towards forward walk
A woman walks forwards.

Second, 走 *zǒu* means moving continuously, without indicating the manner of moving.

(16b) 在外面有一辆车走得越来越远。(L1 Ger-L2 Chn P22, B08)

zài wàimian yǒu yī-liàng chē zǒu de yuè lai yuè yuǎn.
on outside exist one-CL car move MOD more come more far
There is a car outside which moves further and further.

Third, 走 *zǒu* means a motion away from the starting point (see also Lamarre, 2008, p. 73). It occurs frequently with manner verbs. Example:

(16c) 一辆蓝色的车开走了。(L1 Chn P09, B02)

yī-liàng lánsè de chē kāi zǒu le.
one-CL blue MOD car drive leave LE
A blue car drove away.

Fourth, *zǒu* in the compound 走路 *zǒu-lù* 'walk-road' indicates the manner of motion. Meanwhile, *lù* in this compound has the meaning of location. Thus, the verb compound 走路 *zǒu-lù* packages information about both Manner and Ground. Example:

(16d) 有女人在走路。(L1 Ger-L2 Chn P24, A02)

yǒu nǚ-rén zài zǒu-lù.
exist female-person ZAI walk-road
There is a woman walking on the road.

Path verbs

According to the encoded path segments, path verbs can be further distinguished as follows:

• Object-related path verbs:

These path verbs are characterized according to the contours of the Ground objects. Examples are 过 *guò* 'pass, cross', 经过 *jīng-guò* 'pass-pass', 穿过 *chuān-guò* 'cross-cross', 路过 *lù-guò* 'road-cross/pass', and 越过 *yuè-guò* 'cross-cross'.

过 *guò* has different meanings in different contexts.

First, 过 *guò* indicates a movement from one side to the other side. Example:

(17a) 一个人经过一个停车场。(L1 Ger-L2 Chn P21, A10)

yī-ge rén jīng-guò yī-ge tíngchēchǎng.
one-CL person cross-cross one-CL parking lot
A person walks across a parking lot.

Second, the Ground object is conceptualized as a reference point. 过 *guò* indicates a movement passing the reference point. Example:

(17b) 车路过一座大山。(L1 Ger-L2 Chn P13, B02)

chē lù-guò yī-zuò dà-shān.
car road-pass one-CL big-mountain
A car passes a big mountain.

Third, the speaker or observer serves as the reference point and 过 *guò* indicates a movement from one point to another with respect to the speaker or observer (in the sense that the speaker sees the figure moving towards/away from themselves). Without mentioning the Ground object, *guò* with this meaning occurs frequently with the deictic verbs 来 *lái* 'come' and 去 *qù* 'go' (see the deictic path verbs below). Since the moving direction indicated by *guò* is related to the speaker or observer, i.e., the Origo, *guò* expresses the deictic meaning in this case, not the object-related meaning. Example:

(17c) 白色的车子开过去。(L1 Ger-L2 Chn P23, B05)

báisè de chēzi kāi guò qù.
white MOD car drive move over go
A white car drives away (from the speaker).

Note that the first two meanings of 过 *guò* fall into the category of object-related concepts, while the last meaning of 过 *guò* belongs to the deictic concept, which is shown below.

- Deictic path verbs

These are defined according to the moving direction with regard to the Origo: either towards or away from the Origo. 来 *lái* 'come', 去 *qù* 'go', and 过 *guò* (see example 17c) belong to this category.

- Figure-related path verbs

The moving direction is defined according to the body axis of the moving entity. In comparison with deictic path verbs, figure-related path verbs typically occur without any Ground object. Examples are 拐 *guǎi* 'turn', 转 *zhuǎn* 'turn'，转弯 *zhuǎn-wān* 'turn a corner', and 拐弯 *guǎi-wān* 'turn a corner'.

- Direction-related path verbs

These path verbs describe directed movement towards a Ground object. Thus, the the Ground object must be encoded in this category. Examples are the verbs 近 *jìn* 'approach'，靠近 *kào-jìn* 'approach', 前往 *qián-wǎng* 'head for', and 接近 *jiē-jìn* 'approach'.

- Goal-oriented path verbs

These path verbs indicate that the moving entity reaches the endpoint. Examples are 到 *dào* 'arrive', 来到 *lái-dào* 'arrive', and 到达 *dào-dá* 'reach, arrive'

- Boundary-crossing path verbs

These path verbs show the crossing of a boundary defined by the Ground object. Examples are 进 *jìn* 'enter'，出 *chū* 'exit'，and 入 *rù* 'enter'.

- Axial path verbs:

These path verbs show movement in the vertical direction. 上 shàng ‘ascend’ is such an example.

- Source-related path verbs

These path verbs refer to movement away from the starting point. Examples are 离 *lí* ‘leave’ and 走 *zǒu* ‘leave’ (see example 16c).

- Source and goal

These path verbs refer to a Figure moving away from the starting point and moving back to the starting point (which is also the endpoint). That is, the Figure reaches the endpoint, which is also the starting point. 回 *huí* ‘return’ is such an example.

Other verbs

Unlike manner, motion, and path verbs, the fourth category ‘Other verbs’ does not convey any change in place. However, they can also be used to make up SVCs in Chinese, so they are listed and counted as a separate category in this analysis. These verbs include the ‘carry’ verbs in Chinese. ‘carry’ verbs are used because one video clip in Type C shows a woman running into a train station with her luggage. The ‘carry’ verbs in Chinese are 拎 *līn* ‘carry’, 拿 *ná* ‘carry’, and 带 *dài* ‘carry’.

4.3.1.3 Adpositions

In Chinese, adpositions include prepositions, postpositions, and circumpositions (see below for definitions and examples). They all specify a specific Ground property. In the following, the semantic components encoded in adpositions in Chinese will be presented.

Prepositions

- Object-related prepositions

This category shows a motion along the contours of a Ground object. The preposition 沿 *yán* ‘along’ is an example of this.

- Direction

This category shows an explicit motion direction. The prepositions in this category are 往 *wǎng* ‘towards’, 向 *xiàng* ‘towards’, and 朝 *cháo* ‘towards’, which

occur with Ground objects. The Ground object is typically the object at the endpoint. When the prepositions in this category occur with the postposition 里 *lǐ* 'inside', then this construction 'preposition + Ground object + postposition *lǐ*' expresses the meaning of boundary-crossing.

- Goal-oriented prepositions

An example of this category is the preposition 到 *dào* 'to'[39], which shows the Figure reaching the goal. *dào* may or may not occur with postpositions. Depending on the postpositions used, the construction 'preposition + Ground object + postposition' may convey the meaning of boundary-crossing.

- Source

This category shows that the motion takes place from a starting point. The preposition 从 *cóng* 'from' is an example of this. Depending on the occurrence of postpositions, the preposition *cóng* with postpositions can express the meaning of boundary-crossing.

- Location

The preposition 在 *zài*[40] 'at' indicates the place where a motion takes place and may or may not occur with postpositions (e.g., *shàng* 'top'; see the next section below).

Postpositions

Postpositions are also called localizers (Chappell & Peyraube, 2008; Li & Thompson, 1981) or locative terms (Wu, 2008) in Chinese. They can specify the concrete spatial relations between the Figure and the Ground (see the following examples 16, 17, 18, and 19). Based on their phonological features, postpositions can be distinguished into monosyllabic and disyllabic postpositions. Monosyllabic postpositions include 上 *shàng* 'top', 下 xià 'under', 里 *lǐ* 'inside', 外 *wài* 'outside', 左 *zuǒ* 'left', 右 *yòu* 'right', etc. By contrast, disyllabic postpositions are combinations

[39] According to Li and Thompson (1981), 到 *dào* 'arrive' is a coverb in Chinese. It can function both as a verb and as a preposition with the same meaning. So it is no wonder that *dào* is categorized both as a goal-oriented path verb and a goal-oriented preposition in Chinese.

[40] According to Li and Thompson (1981), 在 *zài* is a locative coverb. It can act as a locative verb or as a locative preposition, both of which introduce locative phrases.

of monosyllabic postpositions with one of three suffixes: -*bian* 'surface', -*mian* 'side', -*tou* 'head'. The meanings of these three suffixes are related to parts of objects, e.g., the surface or side of an object, as shown in Table 4.3.

Table 4.3: Monosyllabic and disyllabic postpositions in Chinese

Monosyllabic Postpositions	**Disyllabic Postpositions** (monosyllabic postpositions with suffixes 面 *bian* 'surface', 边 *mian* 'side', 头 *tou* 'head')	**Gloss**
上 *shàng*	上边 *shàngbian*, 上面 *shàngmian*, 上头 *shàngtou*	'on top of, above'
下 *xià*	下边 *xiàbian*, 下面 *xiàmian*, 下头 *xiàtou*	'under, below'
里 *lǐ*	里边 *lǐbian*, 里面 *lǐmian*, 里头 *lǐtou*	'in, inside'
外 *wài*	外边 *wàibian*, 外面 *wàimian*, 外头 *wàitou*	'outside'
前 *qián*	前边 *qiánbian*, 前面 *qiánmian*, 前头 *qiántou*	'in front of'
后 *hòu*	后边 *hòubian*, 后面 *hòumian*, 后头 *hòutou*	'in back of, behind'
旁 *páng*	旁边 *pángbian*	'beside'
中间 *zhōngjian*		'in the center of'
左 *zuǒ*	左边 *zuǒbian*, 左面 *zuǒmian*	'left of'
右 *yòu*	右边 *yòubian*, 右面 *yòumian*	'right of'
东 *dōng*	东边 *dōngbian*, 东面 *dōngmian*, 东头 *dōngtou*	'east of'
西 *xī*	西边 *xībian*, 西面 *xīmian*, 西头 *xītou*	'west of'
南 *nán*	南边 *nánbian*, 南面 *nánmian*, 南头 *ántou*	'south of'
北 *běi*	北边 *běibian*, 北面 *běimian*, 北头 *běitou*	'north of'
这儿 *zhèr*/ 这里 *zhèli*	这边 *zhèbian*, 这面 *zhèmian*，这头 *zhètou*	'this side of'
那儿 *nàr*/ 那里 *nàli*	那边 *nàbian*, 那面 *nàmian*	'that side of'

Note. Adapted from Li and Thompson (1981, p. 391)

Syntactically, both disyllabic and monosyllabic postpositions work with nominal phrases to form a postpositional phrase. However, disyllabic postpositions differ from monosyllabic ones in two ways: The former can either occur alone without the Ground object being specified (compare examples 18a and 18b) or together with other disyllabic postpositions through conjunction (example 19a) or modification (example 19b), while the latter do not have these functions.

(18a) 玩具在上边。

wánjù zài shàngbian.
toy at top
The toy is at the top.

(18b) *玩具在上。

*wánjù zài shàng
toy at top
The toy is at the top.

(19a) 箱子的里面和外面都擦干净了。

xiāngzi de lǐmian hé wàimian dōu cā gānjìng le.
box MOD inside and outside already wipe clean LE
(I) have wiped the inside and outside of the box.

(19b) 玩具在箱子里面的右边。

wánjù zài xiāngzi lǐmian de yòubian.
toy at box inside MOD right side
The toy is in the right part of the box.

Semantically, monosyllabic and disyllabic postpositions generally have the same meaning (examples 20a and 20b). However, depending on the actual context, the meaning may differ or be distorted to some degree (examples 20c and 20d).

(20a) Monosyllabic postposition 上 *shàng*

一个车子在街上开着。(L1 Ger-L2 Chn P15, B02)

yī-ge chēzi zài jiē shàng kāi zhe.
one-CL car at street top drive ZHE
A car is driving on the street.

(20b) Disyllabic postposition 上边 *shàngbian/* 上面 *shàngmian/* 上头 *shàngtou*

一个车子在街上边 / 面 / 头开着。

yī-ge chēzi zài jiē shàngmian/bian/tou kāi zhe.
one-CL car at street top side/surface/end drive ZHE
A car is driving on the street.

(20c) Monosyllabic postposition 上 *shàng*

? 冷冻库在冰箱上。

lěngdòngkù zài bīngxiāng shàng.
freezer at refrigerator top
The freezer is on/above the refrigerator. (Chuang, 2010, p. 22)

(20d) Disyllabic postposition 上面 *shàngmian*

冷冻库在冰箱的上面。

lěngdòngkù zài bīngxiāng de shàngmian.
freezer at refrigerator MOD top surface
The freezer is on top of the refrigerator. (Chuang, 2010, p. 22)

In addition, it is important to mention that postpositions with the meaning 'in, inside', e.g., 里 *lǐ* 里边 *lǐbian,* 里面 *lǐmian*, and 里头 *lǐtou*, play a decisive role in changing the meaning of source, direction, and goal-oriented prepositions to convey boundary-crossing. That is, the prepositions of source (从 *cóng* 'from') / direction (往 *wǎng* 'towards', 向 *xiàng* 'towards', 朝 *cháo* 'towards') / goal orientation (到 *dào* 'reach, arrive') + Ground object + postpositions 里 *lǐ*/ 里边 *lǐbian* / 里面 *lǐmian*/ 里头 *lǐtou* express a sense of boundary-crossing. Prepositions and postpositions together form circumpositions in Chinese.

Circumpositions

As mentioned above, a circumposition typically consists of a preposition such as 在 *zài* 'at' or 往 *wǎng* 'towards', which establishes a spatial relation, and a postposition such as 上 *shàng* 'above', 前 *qián* 'front', or 旁边 *pángbian* 'inside', which further specifies the spatial location (Liu, 2002, 2008). A typical circumposition can be represented like this: preposition + NP + postposition. Since circumpositions, especially the preposition 在 *zài* + NP + postposition, frequently occurs in the data used in this study and L2 speakers often make mistakes in its use (for more details, see section 4.4.1.2), it is necessary to unveil the syntactic and semantic features or restrictions regarding this type of circumposition. According to Chuang (2010), the preposition *zài* 'at' and postpositions play different roles in establishing spatial relations: The preposition *zài* first forms a neutral spatial relation and then, through the postposition, a concrete spatial relation (such as an above-below relation in the vertical direction) can be specified.

Circumpositions: syntactic restrictions on postpositions

Depending on the features of the Ground object in the NP and its syntactic environment, the postpositions can be omitted. According to Li and Thompson (1981) and Liu (2008), when the NP is a place name like Germany or Shanghai, for example, the postposition must be omitted:

(21a) 在德国

zài déguó
at Germany
in Germany

(21b) *在德国里

* zài déguó-lǐ
at Germany-inside

When the NP refers to familiar places, i.e., buildings, rooms, or organizations such as house, train station, village, school, church, etc., or when the prepositional

phrase precedes the verb in the sentence, then the postposition can also be omitted. However, when the prepositional phrase follows the verb, even if the NP indicates one of the familiar places mentioned above, the postposition must be added. The following examples are selected from the descriptions in one of the video clips, which depicts a man walking in an underground parking lot towards a red car.

(22a) Prepositional phrase before verb: without postposition

一个男的在停车场走着。(L1 Ger-L2 Chn P15, A05)

yī-ge nán de zài tíngchēchǎng zǒu zhe.
one-CL male MOD at parking lot walk ZHE
A man is walking in a parking lot.

(22b) Prepositional phrase before verb: with postposition 里 *lǐ* 'inside'

一个男的在地下停车库里走。(L1 Ger-L2 Chn P16, A05)

yī-ge nán de zài dìxià-tíngchēkù-lǐ zǒu.
one-CL male MOD at underground-parking garage-inside walk
A man is walking in an underground garage.

(22c) Prepositional phrase after verb: without postposition

* 一个人走在地下车库。

* yī-ge rén zǒu zài dìxià-chēkù.
one-CL person walk at underground-parking garage

(22d) Prepositional phrase after verb: with postposition 里 *lǐ* 'inside'

一个人走在地下车库里。(L1 Chn P20, A05)

yī-ge rén zǒu zài dìxià-chēkù-lǐ.
one-CL person walk at underground-parking garage-inside
A person walks in the underground garage.

The meaning of the preposition 在 *zài* in examples 22a, 22b, and 22d is equal

to the meaning of the English preposition *at*. Furthermore, when the Ground object is two-dimensional, i.e., a flat surface, depending on the context and the features of the Ground object, a postposition is necessary. Example:

(23a) 一个巴士在路上开着。(L1 Ger-L2 Chn P17, B06)

yī-ge bāshì zài lù- shàng kāi zhe.
one-CL bus at road-top drive ZHE
A bus is driving on the road.

(23b) * 白色的车在路开。(L1 Ger-L2 Chn P13, B05)

* bái-sè de chē zài lù kāi.
white-color MOD car at road drive

Example 23b is grammatically unacceptable because the postposition 上 *shàng* 'top' is required. Semantically, 在 *zài* in example 23a is equal to the English preposition *on*.

Circumpositions: syntactic restrictions on prepositions

When the place NP is placed in the first position of a presentative sentence[41], the preposition 在 *zài* can be omitted (example 24a), whereas when a place NP follows a verb, the preposition 在 *zài* cannot be omitted (Li & Thompson, 1981, pp. 396-97). The following examples show this (the place NP is underlined):

[41] According to Li and Thompson (1981, p. 509), a presentative sentence "performs the function of introducing into a discourse a noun phrase naming an entity". In Chinese, such a noun phrase could either be introduced in the locative or existential phase; for example, (noun phrase is underlined)

树上有一只鸟。
shù-shàng yǒu yī-zhī niǎo.
tree-top exist one-CL bird
There is a bird on the tree.

An alternative are motion verbs; for example:

班里来了一个新同学。
bān-lǐ lái le yī-ge xīn tóngxué.
class-inside come LE one-CL new classmate
A new classmate came into the class.

Ellipsis of preposition 在 *zài*

(24a) 郊外有一辆车在移动。(L1 Ger-L2 Chn P06, B08)

jiāowài yǒu yī-liàng chē zài yí-dòng.
suburbs exist one-CL car ZAI move-move
There is a car moving in the suburbs.

(24b) *人房间前面走路。(L1 Ger-L2 Chn P12, A02)

* rén fángjiān qiánmian zǒu-lù.
person room front walk-road

(24c) *绿色的汽车开农村的一个马路。(L1 Ger-L2 Chn P20, B02)

* lǜsè de qìchē kāi nóngcūn de yī-ge mǎlù.
green MOD car drive countryside MOD one-CL road

The preposition 在 *zài* can be omitted in 24a since the place NP is at the beginning of the sentence, while it is not correct to omit the preposition 在 *zài* in 24b and 24c because in 24b, the place NP *fángjiān qiánmian* 'room front' is not in the initial position of a presentative construction and in 24c, the place NP *nóngcūn de yī-ge mǎlù* 'a road in the countryside' follows the verb. In these two examples, the preposition 在 *zài* must be added. Besides the addition of the preposition 在 *zài*, similar to example 23b, the postposition 上 *shàng* 'on top of' must be added in example 24c. As the above examples illustrate, the locative prepositional phrase, i.e., canonically, *zài* + NP + postposition, can precede or follow the verb, which can be represented as follows:

a) *zài* + NP + postposition + verb
b) verb + *zài* + NP + postposition

The preverbal locative prepositional phrase indicates the location where an activity takes place, while the postverbal one indicates the location of the subject or

direct object as a result of an activity[42] (Li & Thompson, 1981). The following examples 25a and 25b are taken from Li and Thompson (1981, p. 399):

(25a) 他在桌子上跳。

tā zài zhuōzi-shàng tiào.
he at table-top jump
He is jumping on the table.

(25b) 他跳在桌子上。

tā tiào zài zhuōzi-shàng.
he jump at table-top
He jumped onto the table.

However, since all movements take place in a certain location, the preverbal prepositional phrases are relatively unrestricted with respect to their occurrence with verbs, while postverbal ones are restricted in terms of syntactic requirements: They must be placed immediately after the verb. In other words, no other elements are allowed to be placed between the verb and the postverbal prepositional phrase, not even the direct object in a verb-object component (Li & Thompson, 1981, pp. 406-09). This is shown in the following examples (where the verb-object compounds are underlined once, while locative prepositional phrases are underlined twice):

* verb + direct object + postverbal locative phrase

(26a) *又有人开车在比较偏僻的地方。 (L1 Ger-L2 Chn P04, B04)

* yòu yǒu rén kāi-chē zài bǐjiào piānpì de dìfāng.
again exist person drive-car at relative remote MOD place

[42] Li and Thompson (1981, pp. 398-406) illustrated four verb types that basically allow the occurrence of preverbal and postverbal locative phrases: a) verbs of displacement that indicate the displacement of the subject or the direct object, e.g., 跳 *tiào* 'jump', 扔 *rēng* 'throw', etc.; b) verbs of posture that depict the posture of the subject, e.g., 站 *zhàn* 'stand', 坐 *zuò* 'sit', etc.; c) verbs of appearance that depict the subject's appearance and disappearance, e.g., 发生 *fāshēng* 'happen'; and d) verbs of placement that describe an activity that makes the object move to a certain place, e.g., 放 *fàng* 'put, place', 藏 *cáng* 'hide'.

(26b) *一个女人骑自行车在一个农村。(L1 Ger-L2 Chn P20, B09)

* yī-ge nǚ-rén qí zìxíngchē zài yī-ge nóngcūn.
one-CL female-person ride bicycle at one-CL countryside

26a and 26b are grammatically unacceptable since the postverbal locative phrase in the two examples does not follow the verb immediately. Instead, it follows the verb-object compound *kāi-chē* 'drive-car' in 26a and the direct object in the verbal phrase *qí zìxíngchē* 'ride bicycle' in 26b, which goes against the restriction that the locative phrase should follow the verb without any intervening elements.

4.3.1.4 Adverbs

Adverbs such as 快速地 kuàisù-de 'quickly', 很快地 hěnkuài-de 'quickly', and 慢慢地 mànmàn-de 'slowly' also describe Manner information in Chinese.

4.3.2 English

In English, spatial concepts are mainly encoded in verbs and adjuncts. In addition, adverbs also encode spatial concepts, but this happens rather rarely in the given data (see, e.g., Carroll, 2000; Carroll & Lambert, 2006). In the following, spatial concepts encoded in verbs, adjuncts, and adverbs, respectively, will be presented.

4.3.2.1 Verbs

Like Chinese, English verbs are also categorized into four groups: motion verbs, manner verbs, path verbs, and other verbs.

Motion verbs

Motion verbs do not include any Manner or Path information. In other words, apart from the movement component, the lack of Manner or Path information makes motion verbs neutral. Only two types of motion verbs were found in this study in L1 and L2 English: *move* and *go*. The verb *go* is a motion verb only when it occurs in the phrase *go by bicycle*. Here *go* is not a deictic verb; rather it implies a movement that is further specified by the means of transport. For example, one can also say *go by plane* or *go by ship*.

Manner verbs

Manner verbs imply the way of moving. So when we speak of manner verbs,

we always understand that they also include Motion information. Verbs like *walk*, *drive*, *run* are manner verbs.

Path verbs

Based on the encoded path segments, path verbs can be further categorized as follows:

- Object-related path verbs:

This type of path verbs can be defined based on the contours of the Ground object at the reference point. Examples are *pass* and *cross*.

- Deictic path verbs

In these verbs, the moving direction is defined with respect to the Origo. That is, the direction is either towards or away from the Origo. *come* and *go* are deictic verbs.

- Figure-related path verbs

The body axis of the moving entity is taken as the reference point to decide the moving direction. The verb *turn* is an example.

- Direction-related path verbs

These path verbs describe directed movement towards a Ground object. Thus, the Ground object must be encoded in this category. The English verbs or phrasal verbs *approach*, *head towards*, and *draw to* belong to this category of path verbs.

- Goal-oriented path verbs

These path verbs can be described as a moving entity arriving at an endpoint. Examples are the verbs *reach* and *arrive at.*

- Boundary-crossing path verbs

In these path verbs, a boundary defined by a Ground object is traversed and the moving entity is either inside or outside the Ground object after the movement. The verbs *enter* and *exit* belong to this category.

- Axial path verbs

Information about movement in the vertical direction is encoded in this group of path verbs. There is only one instance of this type of path verbs that occurs in L1 and L2 English in this study. An example is the phrasal verb *taking the steps* in the sentence *there is someone taking the steps to a building*. Apart from being an axial path verb, *taking the steps* also encodes Manner information in the sense that it expresses a way of moving.

Other verbs

Apart from motion verbs, manner verbs, and path verbs, other verbs also occur

in L1 and L2 English. For example, verbs like *search* in the sentence *a man is walking in a parking garage, probably searching for his car* and verbs like *carry* in the sentence *a woman is walking across a square carrying a bag*.

4.3.2.2 Prepositions

As previously explained in section 4.3.1.3, prepositions state a specific Ground property, which is fundamental for locating a Figure. The conceptual components encoded in each preposition are unveiled in the following categories:

- Object-related prepositions

The contours of the Ground object play an important role in determining this category. Examples are *along the road*, *around the corner*, *alongside a building*, *through a parking lot*, and *across a square*.

- Deictic prepositions

This category describes that the moving direction is related to the Origo, i.e., the speaker. The preposition *down* in the sentence *a car is driving down the street*[43] is such an example.

- Directional prepositions

An explicit motion direction is expressed by this category of prepositions. Some examples are *towards her car* and *in the direction of a phone booth*.

- Goal-oriented prepositions

This category of prepositions refers to a Figure reaching the goal. Examples of this category are the preposition *to* in the sentence *a car is driving to a village* and *at* in *a bus is arriving at a bus stop*.

- Boundary-crossing prepositions

This category includes the prepositions *into the garage*, *in a train station*, *out of a court*, and *inside* in the sentence *a dog ran through a door inside a garage*.

- Prepositions of source

This category shows movement that starts at a place defined by a Ground object. The preposition *from* in the sentence *a cat enters a room from the corridor* is such an example.

- Axial prepositions

The prepositions in this category describe movement in the vertical direction.

[43] The preposition *down* may also indicate moving downwards. But the video clips at issue showed the road was long without any slopes. So the meaning of moving downwards is excluded in this sentence.

For example, *up* in the sentence *there is a person walking up the stairs to a bench*.

- Prepositions of location

This category specifies the place where a motion event takes place. Examples are *on the street, besides/ next to a building, near some houses, in a rural area, between two fields*.

4.3.2.3 Adverbs

Manner information can be expressed both in manner verbs and in adverbs. Adverbs such as *unsteadily*, *slowly*, and *carefully* express Manner information. Adverbs like *upstairs* express movement in the vertical direction.

4.3.3. German

Like speakers of Chinese and English, German speakers also use verbs, adjuncts, and adverbs to describe motion events. Besides these, they also use case and verb particles[44] when describing motion events. In German, some prepositions govern one specific case, whereas other prepositions govern two cases. Depending on the cases used in the prepositional phrase, its meaning varies: the meaning of direction or the meaning of location (more details see 4.3.3.2). In addition, verb particles play an important role in encoding Path information. In the following, the spatial encodings in these linguistic devices in German will be presented.

4.3.3.1 Verbs

Like Chinese and English verbs, German verbs are also divided into four groups with respect to the conceptual components they encode: motion verbs, manner verbs, path verbs, and other verbs.

Motion verbs

This category refers to verbs that encode only one conceptual component, that is, Motion (the movement). The German verb *sich bewegen* 'move' belongs to this category.

Manner verbs

This category refers to verbs that encode two conceptual components simultaneously: Motion and Manner; that is, the movement itself and the way of moving. The manner verbs that occur in the German data in this study are *laufen* 'walk',

[44] English and German differ from each other in the use of verb particles. Due to the grammatical structures, verb particles in English are used rather restrictedly (Olsen, 1996; Stiebels, 2015).

rennen ‘run’, *fahren* ‘drive’, *klettern* ‘climb’, *spazieren* ‘take a walk’, *spazieren gehen* ‘take a walk’, *schlendern* ‘stroll, wander’, *parken* ‘park’, *schleichen* ‘sneak, creep’, *reiten* ‘ride’. It is also worth noting that *gehen* ‘go by walking’ is neither a pure manner verb nor a pure deictic verb in German; rather it is a manner verb with a deictic component. Furthermore, the deictic meaning in this verb is faded. In this sense, *gehen* is treated as a separate category in this study.

Path verbs

According to the path segments packaged in the verbs, path verbs can be further divided into the following categories:

- Object-related path verbs:

Information on the contours of a Ground object is packaged into the verbs in this category. The Ground object can be a two- or three-dimensional object, or it can also be conceptualized as a point. Examples are *überqueren* ‘cross’, *passieren* ‘pass’, *durchfahren* ‘drive through’, *durchlaufen* ‘go through’.

- Deictic path verbs

In these verbs, the moving direction is determined with respect to the speaker, that is, the direction towards or away from the speaker. *kommen* ‘come’ is a deictic verb. As mentioned above, *gehen* ‘go by walking’ is neither a pure manner verb nor a pure deictic verb. It is a manner verb with a deictic component.

- Figure-related path verbs

The body axis of the moving entity is taken as a reference point for the path verbs in this category. The difference to deictic path verbs is that the motion direction is derived from the moving entity itself, not from the external reference point. The German verbs *biegen* ‘turn’, and *abbiegen* ‘turn off’ are in this category.

- Directional path verbs

This category describes verbs where the movement is directed towards a place that ends in a goal. The German verb (*zu*)*steuern* ‘head for’ occurs only once in the data used in this study.

- Boundary-crossing path verbs

A Figure traverses a boundary line and the spatial relation between the Figure and the Ground changes after the movement: The Figure is either inside or outside the Ground object compared to the constellation before the movement. The German verb *betreten* ‘enter’ belongs to this category.

- Axial path verbs

The path verbs in this group package information on vertical movement. The German verb *steigen* 'ascend' belongs to this category.

Other verbs

This category is composed of verbs that have neither Manner nor Path information. However, they still make up a portion of the verbs recorded in this study, so they are counted separately here. Example: *anlegen* in the sentence *Ein Mann steigt eine Leiter hinauf, die an den Balkon im ersten Stock angelegt ist* 'A man climbs a ladder, which is attached to the balcony on the first floor'. Verbs like *stehen* 'stand', *umringen* 'surround', and *sitzen* 'sit' also belong in this category.

4.3.3.2 Prepositions

Unlike in Chinese and English, case is a language-specific category in German. In the collected data, prepositions in German govern either the accusative or dative case[45], or both. According to Carroll (2000), the accusative denotes motion towards a place, while the dative denotes motion within a place. A place marked by the accusative is divided into different sub-spaces, e.g., inner space, surface space, and adjacent space, whereas a place marked by the dative is understood as a point or a place. Examples:

Accusative: motion to a place (inner space)
(27a) Ein Auto fährt in eine Garage
'A car drives into a garage'

Dative: motion within a place
(27b) Ein Hund läuft im Garten.
'A dog runs in the garden'

However, there is an exception for the German preposition *zu* 'to', which can only govern the dative case. Although the dative denotes the place where a motion takes place, in this case, *zu* + dative conveys a goal-oriented movement. In the following, we will elaborate on the conceptual components conveyed in the prepositions and their interplay with the accusative or dative case.

- Object-related prepositions

[45] Prepositions or postpositions in German can also govern the genitive case, but this did occur in the data in this study, so the genitive will not be discussed further.

The prepositions *um* ‘around’, *an* ‘past’, *über* ‘across’, *durch* ‘across, through’, *entlang* ‘along’ belong to this category. The meaning conveyed by these prepositions is related to the contours of the Ground object. According to the cases they govern, these prepositions can be further divided into three groups:

i. Some prepositions can only govern the accusative. *um* ‘around’ and *durch* ‘across’ are such examples. *um* ‘around’ is used when the Ground object is in a round-like shape (e.g., *um eine Kurve* ‘around a curve’), whereas *durch* ‘across, through’ is applied when the Figure moves from one side to another, e.g., *durch eine Tür* ‘through a door’ or when the Figure moves across a distance (either long or short) in a certain area, e.g., *durch ein Parkhaus* ‘through a parking garage’. The exact meaning denoted by *durch* must be inferred from the context.

ii. Some prepositions can govern two cases, the accusative and the dative. Examples are the prepositions *an* and *über*. These two prepositions convey meaning that is related to the features of the Ground object. In this case, *an* governs the dative case and occurs frequently together with the particle *vorbei;* e.g., *ein Mann geht an einem Gebäude vorbei* ‘a man walks past a building’. When *über* governs the accusative case, it has a similar meaning to *across* in the sentence *eine Frau geht über einen Platz* ‘a woman walks across a square’.

iii. The preposition *entlang* ‘along’ has two positions with respect to the nominal phrase. It can either precede or follow a nominal phrase. When *entlang* precedes a nominal phrase, it governs the dative case; however, when *entlang* follows a nominal phrase, it governs the accusative case. Regardless of its position, the meaning denoted by *entlang* ‘along’ is typically related to the contours of the Ground object.

- Directional prepositions

This group of prepositions expresses an explicit motion direction with respect to a Ground object. Examples are *auf … zu* ‘towards’, *in Richtung* ‘towards’, *Richtung* ‘towards’, and *nach* ‘towards’. The preposition *auf* occurs frequently with the particle *zu* to express the direction (e.g., *auf eine Bushaltestelle zu* ‘towards a bus stop’). *in Richtung* and *Richtung* have the same meaning. The latter is a short form of the former and is mainly used in spoken German. Both of them precede the nominal phrase.

- Goal-oriented prepositions

This group of prepositions shows a Figure arriving at the goal. Examples are the prepositions *zu* ‘to’ and *an* ‘to’. As mentioned above, *zu* only governs the dative case (e.g., *zu einer Bushaltestelle* ‘towards a bus stop’), while *an* in this case governs

the accusative case[46] (e.g., *ein Bus fährt an eine Bushaltestelle* 'a bus drives towards a bus stop'.)

- Boundary-crossing prepositions

This category shows that the boundary defined by a Ground object is traversed. The preposition *in* is an example of this. In this case, *in* governs the accusative case (e.g., *in die Garage* 'into the garage').

- Prepositions of source

This group of prepositions shows that movement begins at a place. The preposition *aus* 'from' (e.g., *aus einem dunkleren Flur* 'from a darker corridor') belongs in this category. *aus* governs the dative case.

- Prepositions of location

The prepositions in this group specify a place where a specific movement happens. Examples are *auf* 'on' (e.g., *auf der Straße* 'on the street'), *neben* 'next to' (e.g., *neben einer Straße* 'next to a street'), *vor* 'before' (e.g., *vor einem Gebäude* 'before a building'), *hinter* 'behind' (e.g., *hinter einem großen Gebäude* 'behind a big building'), *in* 'in' (e.g., *in einem Parkhaus* 'in a parking garage'), and *zwischen* 'between' (e.g., *zwischen Bäumen* 'between trees'). The prepositions in this category govern the dative case.

4.3.3.3 Particles

As mentioned above, the accusative marks motion to a place, whereas the dative marks motion within a place (Carroll, 2000). In example 27a, repeated here, *ein Auto fährt in eine Garage* 'a car drives into a garage', it is not explicitly marked whether the moving entity reaches the place or not. When this example is used in conjunction with the verb particle *hinein* 'thither-in', then it explicitly expresses that the moving entity reaches the goal. An example of this appears in 28a:

preposition + accusative + verb particle (indicating boundary-crossing):

(28a) Ein Auto fährt in eine Garage hinein.

a car drives in the garage hither-in

A car drives in the garage (away from the speaker).

[46] The preposition *an* can govern both the accusative and the dative case. As already mentioned above, when it governs the dative case, it conveys object-related meaning; whereas then it governs the accusative case, it conveys goal-oriented meaning.

Similarly, when example (27b) is used with the verb particle *hinein* 'thither-in', then this example implicates the continuation of the path across some boundary (Carroll, 2000). Example (28b) depicts this:

preposition + dative + particle: motion within space at a boundary

(28b) Ein Hund läuft zur Haustür hinein.

a dog runs to a house door hither-in

A dog runs in the house door (away from the speaker).

From these two examples, we can see that verb particles play an important role in encoding Path information. In the following we will scrutinize the path components encoded in the verb particles in this study.

- Object-related verb particles:

This category maps information on the contours of a Ground object into verb particles. The verb particles *vorbei-* 'past' and *entlang-* 'along' belong to this category. When the Ground object is conceptualized as a point, then *vorbei-* 'past' is used; in contrast, when the contours of the Ground object are unbounded, then *entlang-* 'along' is used.

- Deictic verb particles

There are two basic deictic verb particles in German: *hin-* 'thither' and *her-* 'hither'. They can occur with prepositions to form verb particles with an axial meaning, such as *hinauf-* 'thither-on', *hinunter-* 'thither-under'; with a boundary-crossing meaning, such as *hinein-* 'thither-in'; with goal-oriented meaning, such as *heran-* 'hither-to'; with an object-related meaning, such ase *hindurch-* 'thither-across'. These combined forms also convey deictic meaning related to the speaker.

- Directional verb particles

This category describes movement directed to the place at goal. The German verb particle *zu-* 'towards' belongs in this category. This verb particle typically occurs with manner verbs and the preposition *auf* 'towards' to express a directional meaning.

- Goal-oriented verb particles

This category describes a movement conveyed by the verb to the place at goal and the final position of the Figure is in close proximity to the place at goal. The verb particle *an-* 'at' belongs in this category.

- Boundary-crossing verb particles

The Figure reaches the place at goal and traverses the boundary. The German verb particles *ein-* 'in' and *rein-* 'in' belong in this category.

- Axial verb particles

The verb particles in this group describe movement in the vertical direction. The German verb particles *hoch-* 'up', *empor-* 'up', and *rauf-* 'up' describe an upward orientation.

4.3.3.4 Adverbs

Adverbs such as *schnell* 'quickly' and *eilig* 'quickly' encode Manner information. However, these verbs are found to be used only occasionally in German in this study.

Overall, the specific concepts are more differentiated across Chinese, English, and German. The specific dimensions are encoded in the verb lexicon or in the adjuncts in different languages. Both Manner and Path information are encoded in the serial verb constructions in Chinese. There are syntactic restrictions by using Chinese serial verb constructions. This may pose challenges for L2 Chinese acquisition.

4.4 Spatial analyses in L1 German-L2 Chinese

In section 3, we elaborated on how the linguistic expressions in this study can be categorized with respect to the conceptual components encoded in different languages. Based on the conceptual meanings encoded in the lexical items, this section shows the frequency of verbs and adjuncts in L2 Chinese and compares them with L1 Chinese and L1 German. The structural anomalies in L2 Chinese will also be noted. Following that, the combination of semantic components and segmentation of sub-events in L1 and L2 will be compared. Finally, speech onset times and fixation patterns in L1 and L2 will be presented in order to shed light on the correlation between linguistic structure and motion event cognition.

4.4.1 Linguistic expressions used in L1 and L2 Chinese

Our findings will be presented in the following way: First, an overview of the occurrence of verb types and adjuncts in the descriptions of Type A, B, and C videos will be provided in full, so as to get a general impression of the linguistic means used in different language groups (i.e., overall across all video types within each language). Following that, the occurrence of verb types and adjuncts for each video type with

respect to different language groups will be compared in order to show whether the features of different video types have any impact upon the choice of verb types and adjuncts.

4.4.1.1 Verb types

The percentage of verb types used in each language group is calculated as the verb type used over all verbs found among all responses within that language group. For the SVCs (serial verb constructions) in Chinese, each verb in such a construction is counted once. For example, in an SVC like *zǒu jìn qù* 'walk enter go', there are three verbs in total: one manner verb, one boundary-crossing path verb and one deictic verb. The percentage of manner verbs in this construction is calculated with the following formula: number of manner verbs used / total number of verbs used. So the percentage of manner verbs used in this example is 33.33%. Similarly, the percentages of the manner verbs used in L2 Chinese for Type A videos equals the number of manner verbs used in L2 Chinese for Type A videos divided by the total number of verbs used in L2 Chinese for Type A. Descriptions of non-motion events were excluded from the analysis, so verbs appearing in those descriptions are not counted.

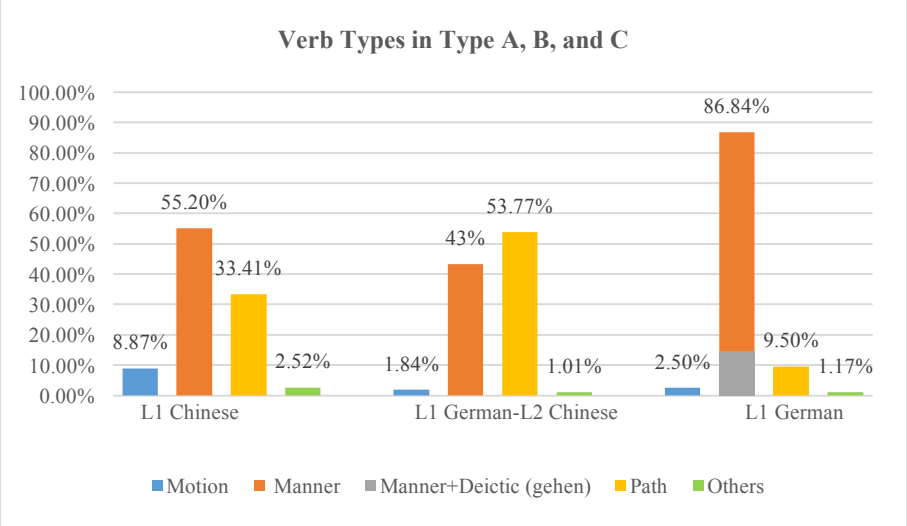


Figure 4.1: Verb types used by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos

Figure 4.1 shows an overview of the distribution of motion verbs, manner verbs, path verbs, and other verbs in the descriptions of Type A, B, and C videos in the three language groups. As already explained in section 4.3.3.1, the German verb *gehen* 'go by walking' is a manner verb with a deictic component, so it is labeled separately as

Manner + Deictic in the manner verb category in L1 German. When comparing the motion verbs used in the L1 and L2 groups, it was found that L1 German-L2 Chinese speakers used them significantly less than L1 Chinese speakers (L2 Chinese 1.84% vs. L1 Chinese 8.87%; $\chi 2$ (1, N = 101) =36.842, $p < .001$). There were no significant differences in the use of motion verbs between L2 Chinese and L1 German speakers (L2 Chinese 1.84% vs. L1 Chinese 2.50%; $\chi 2$ (1, N = 35) =0.714, $p = 0.398$). The following example shows the motion verb *yídòng* 'move' in L2 Chinese (the motion verb *yídòng* 'move' is underlined),

(29) 有一个人在移动。(L1 Ger-L2 Chn P06, A06)

yǒu yī-ge rén zài yí-dòng.
exist one-CL person ZAI move-move
There is a person moving.

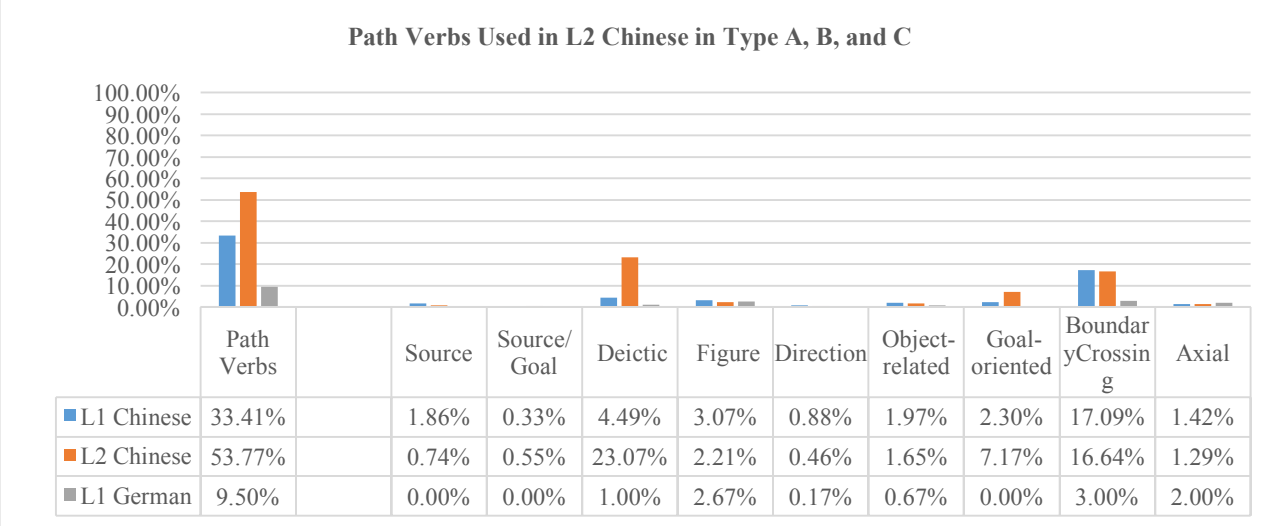
In general, L1 and L2 Chinese speakers differed in the use of manner verbs ($\chi 2$ (1, N = 2001) = 27.761, $p < .001$): L1 Chinese speakers used significantly more manner verbs than L2 Chinese speakers. Likewise, L2 Chinese speakers differed from L1 German speakers in the use of manner verbs ($\chi 2$ (1, N = 1688) = 301.452, $p < .001$): L1 German speakers used more manner verbs than L2 Chinese speakers. One difference in the use of manner verbs was also found between speakers of L1 Chinese and L1 German ($\chi 2$ (1, N = 1513) = 165.781, $p < .001$): L1 German speakers preferred using more manner verbs than L2 Chinese speakers. The following is an example of the use of the manner verb *zǒu* 'walk' in L2 Chinese (the manner verb *zǒu* 'walk' is underlined):

(30) 一个女人在城里走。(L1 Ger-L2 Chn P24, A07)

yī-ge nǚ-rén zài chéng-lǐ zǒu.
one-CL female-person at town-inside walk.
A woman is walking in the town.

The use of path verbs by L2 Chinese speakers was significantly higher than that by L1 Chinese speakers (L2 Chinese 53.77% vs. L1 Chinese 33.41%; $\chi 2$ (1, N = 890) = 88.090, $p < .001$) and L1 German speakers (L2 Chinese 53.77% vs. L1 German 9.50%;

$\chi 2$ (1, N = 642) = 434.243, p < .001). The question then is why path verbs are more likely to be used in L2 Chinese. Figure 4.2 illustrates the detailed path verbs used for the description of Type A, B, and C videos.



	Path Verbs		Source	Source/ Goal	Deictic	Figure	Direction	Object-related	Goal-oriented	BoundaryCrossing	Axial
L1 Chinese	33.41%		1.86%	0.33%	4.49%	3.07%	0.88%	1.97%	2.30%	17.09%	1.42%
L2 Chinese	53.77%		0.74%	0.55%	23.07%	2.21%	0.46%	1.65%	7.17%	16.64%	1.29%
L1 German	9.50%		0.00%	0.00%	1.00%	2.67%	0.17%	0.67%	0.00%	3.00%	2.00%

Figure 4.2: Detailed path verbs used by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos

Taking a closer look at the path verbs used in L2 Chinese, it was found that L2 Chinese speakers used significantly more deictic path verbs than L1 Chinese speakers (L1 Chinese 4.49% vs. L2 Chinese 23.07%, $\chi 2$ (1, N = 292) = 151.027, p < .001) and L1 German speakers (L2 Chinese 23.07% vs. L1 German 1.00%, $\chi 2$ (1, N = 257) = 233.560, p < .001) (see Figure 4.2). The following example illustrates this (the deictic verb *qù* 'go' is underlined):

(31) 有人去车子那边。(L1 Ger-L2 Chn P07, A01)

yǒu rén qù chēzi nà-bian.
exist person go car that-side
Someone goes to a car.

In addition, the use of goal-oriented path verbs by L2 Chinese speakers was significantly higher than that by L1 Chinese speakers (L1 Chinese 2.30% vs. L2 Chinese 7.17%, $\chi 2$ (1, N = 99) = 32.818, p < .001) and L1 German speakers (L2 Chinese 7.17% vs. L1 German 0%), as the following example shows (the goal-oriented path verb *dào* 'arrive' is underlined):

(32) 公交车开到公交车站。(L1 Ger-L2 Chn P04, A09)

gōngjiāochē kāi-dào gōngjiāochēzhàn.
bus drive-arrive bus station
The bus drove to the bus station.

Furthermore, the use of boundary-crossing path verbs in L2 Chinese was significantly higher than in L1 German (L2 Chinese 16.64% vs. L1 German 3.00%, $\chi 2$ (1, N = 199) = 133.513, p < .001). However, no statistical significance was found between L1 and L2 Chinese speakers ($\chi 2$ (1, N = 337) = 1.855, p = 0.173). Example:

(33) 有一个绿色的车开进一个车库。(L1 Ger-L2 Chn P05, C01)

yǒu yī-ge lǜsè-de chē kai-jìn yī-ge chēkù.
exist one-CL green-NOM car drive-enter one-CL garage
There was a green car that drove into a garage.

Concerning the use of object-related path verbs, there were no differences between L2 Chinese and L1 Chinese speakers ($\chi 2$ (1, N = 36) = .000, p = 1.000); however, L2 Chinese speakers differed greatly from L1 German speakers ($\chi 2$ (1, N = 22) = 8.909, p = .003), as the following example shows (the object-related path verb *jīng guò* 'pass' is underlined):

(34) 一个巴士经过一个村子。(L1 Ger-L2 Chn P21, B06)

yī-ge bāshì jīng-guò yī-ge cūnzi.
one-CL bus pass-pass one-CL village
A bus passed a village.

There were no significant differences in the use of figure-related path verbs ($\chi 2$ (2, N = 68) = 3.294, p = 0.193), direction related path verbs ($\chi 2$ (2, N = 14) = 5.286, p = 0.071), and axial path verbs ($\chi 2$ (2, N = 39) = 0.154, p = 0.926) among the three language groups. In the following, we will show whether the different features of the various types of video clips exerted an influence on the use of verbs.

Type A: Short trajectory with obvious endpoint

As shown in Figure 4.3, the use of motion verbs in L2 Chinese was significantly lower than in L1 Chinese (L2 Chinese 0.86% vs. L1 Chinese 6.69%, $\chi 2$ (1, N = 21) = 10.714, p = 0.001) and L1 German (L2 Chinese 0.86% vs. L1 German 5.26%, $\chi 2$ (1, N = 14) = 4.571, p = 0.033). (for the use of motion verbs in L2 Chinese please see example 29)

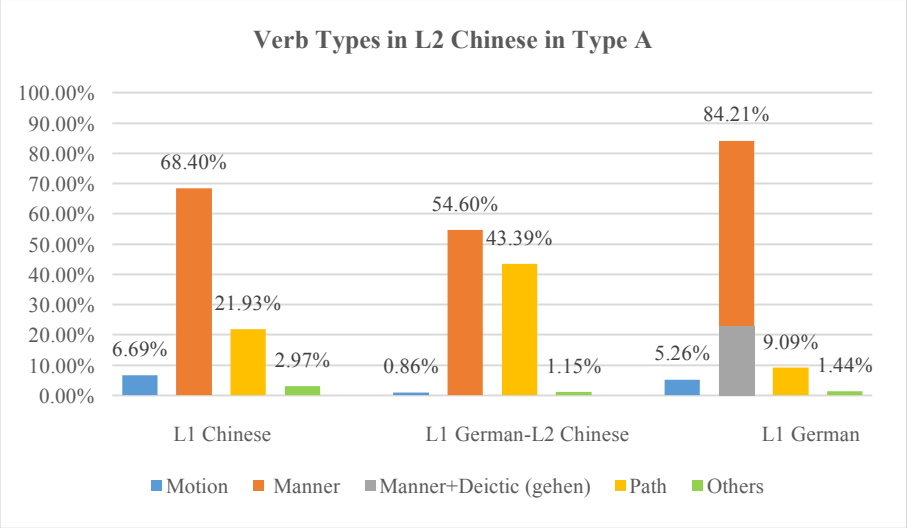


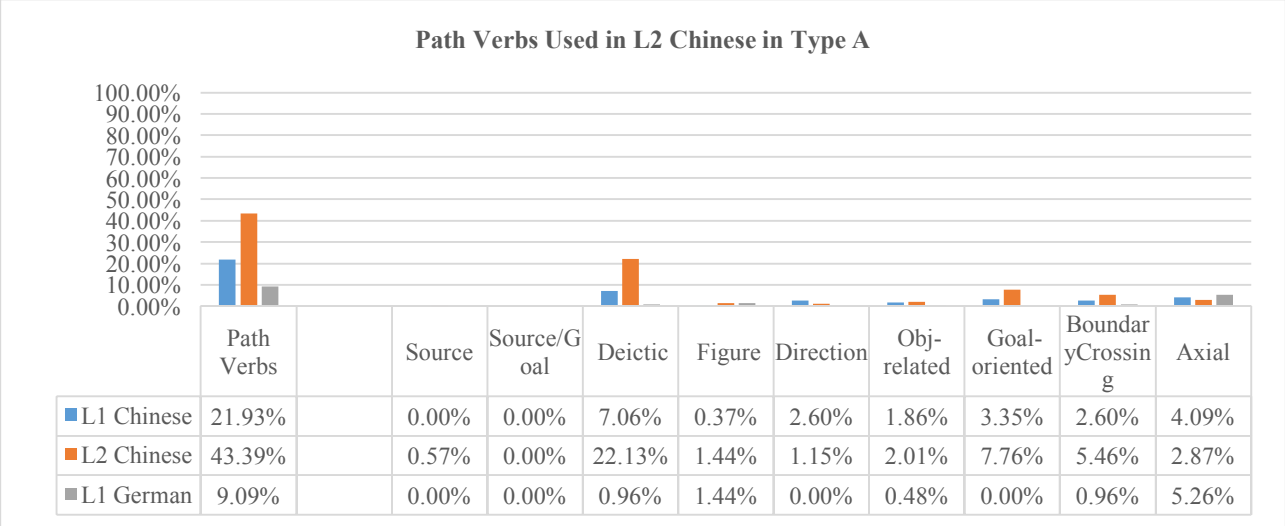
Figure 4.3: Verb types used by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type A videos

In the verbal task of Type A videos, L1 Chinese speakers used significantly more manner verbs than L2 Chinese speakers ($\chi 2$ (1, N = 617) = 12.110, p = .001). The same was true for L1 German speakers, who used significantly more manner verbs than L2 Chinese speakers ($\chi 2$ (1, N = 557) = 50.819, p < .001). In addition, in comparison with L1 German speakers, L1 Chinese speakers used significantly less manner verbs ($\chi 2$ (1, N = 478) = 15.811, p < .001).

The use of path verbs by L2 speakers was significantly higher than that by speakers of L1 German (L2 Chinese 43.39% vs. L1 German 9.09%, $\chi 2$ (1, N = 170) = 102.494, p < .001) and speakers of L1 Chinese (L2 Chinese 43.39% vs. L1 Chinese 21.93%, $\chi 2$ (1, N = 210) = 40.305, p < .001). In addition, there were no differences in the use of other verbs among the three language groups ($\chi 2$ (2, N = 15) = 2.800, p = 0.247).

As shown in Figure 4.4, an in-depth look at the path verbs revealed that L2 Chinese speakers differed remarkably in the use of deictic verbs from speakers of L1 Chinese (L2 Chinese 22.13% vs. L1 Chinese 7.06%, $\chi 2$ (1, N = 96) = 35.042, p < .001) and speakers of L1 German (L2 Chinese 43.39% vs. L1 German 9.09%, $\chi 2$ (1, N = 170)

= 102.494, p < .001). Example 31 illustrates the use of deictic verbs in L2 Chinese, so it will not be repeated here.



	Path Verbs		Source	Source/Goal	Deictic	Figure	Direction	Obj-related	Goal-oriented	BoundaryCrossing	Axial
L1 Chinese	21.93%		0.00%	0.00%	7.06%	0.37%	2.60%	1.86%	3.35%	2.60%	4.09%
L2 Chinese	43.39%		0.57%	0.00%	22.13%	1.44%	1.15%	2.01%	7.76%	5.46%	2.87%
L1 German	9.09%		0.00%	0.00%	0.96%	1.44%	0.00%	0.48%	0.00%	0.96%	5.26%

Figure 4.4: Detailed path verbs used by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type A videos

Furthermore, the use of goal-oriented path verbs was significantly higher in speakers of L2 Chinese than in speakers of L1 Chinese (L2 Chinese 7.76% vs. L1 Chinese 3.35%, $\chi 2$ (1, N = 36) = 9.000, p = 0.003) and speakers of L1 German (L2 Chinese 7.76% vs. L1 German 0%) (see example 32).

Concerning the use of boundary-crossing path verbs, L2 Chinese speakers used them significantly more often than L1 Chinese speakers (L2 Chinese 5.46% vs. L1 Chinese 2.60%, $\chi 2$ (1, N = 26) = 5.538, p = 0.019) and L1 German speakers (L2 Chinese 5.46% vs. L1 German 0.96%, $\chi 2$ (1, N = 21) = 13.762, p < .001). It appears strange to use boundary-crossing path verbs to describe Type A videos, since all of the Type A video clips do not actually depict any boundary-crossing events. There is only one Type A video clip where a higher number of boundary-crossing path verbs were used for the description. This video clip shows a man walking up the stairs towards the entrance of a building. The man in the video does not actually arrive at the entrance; rather he merely approaches it. For this situation, 15 of the 27 L2 Chinese speakers used the boundary-crossing verb *jìn* 'enter' or deictic verbs to form a serial verb construction *jìn qù* 'enter go'. In L2 Chinese, *jìn* or *jìn qù* occurred with modal verbs like *xiǎng* 'want to', temporal adverbs like *kuài* 'almost', the prospective aspect marker[47] *yào*, the progressive aspect marker *zài*, or zero marking, which are underlined in the following

[47] For more details about the aspect markers in Chinese, see Chapter 5.

examples.

With the modal verbs *xiǎng*/ *dǎsuan*/*zhǔnbèi* ‘want to/are going to/intend’

(35a)这位先生在走楼梯，* 想进去房子。(L1 Ger-L2 Chn P01, A06)

zhè-wèi xiānsheng zài zǒu lóutī, *xiǎng jìn- qù fángzi.
this-CL sir ZAI walk stairs, want enter-go house
This man is walking up the stairs and wants to go into the house.

(35b)一个男生打算进入大学的办公楼。(L1 Ger-L2 Chn P09, A06)

yī-ge nánshēng dǎsuan jìn-rù dàxué-de bàngōng-lóu.
one-CL boy be going to enter-enter university-DE office-building
A boy is going to enter the university’s office building.

(35c)*有一个男人他走上去一个楼梯，然后准备进去一个比较大的房子。

(L1 Ger-L2 Chn P22, A06)
* yǒu yī-ge nán-rén tā zǒu shàng- qù yī-ge lóutī,
exist one-CL male-person he walk ascend-go one-CL stairs,
There is a man who walks up the stairs (away from the speaker),

* ránhòu zhǔnbèi jìn-qù yī-ge bǐjiào dà-de fángzi.
then intend enter-go one-CL relatively big-DE house
then he intends to go into a relatively big house.

(36) With the temporal adverb *kuài* ‘almost’

*有一个男生他快进去海德堡大学的办公室。(L1 Ger-L2 Chn P02, A06)

* yǒu yī-ge nánshēng tā kuài jìn- qù hǎidébǎo dàxué-de bàngōngshì.
exist one-CL boy he almost enter-go Heidelberg university-DE office
A boy almost entered the office of Heidelberg University.

(37) With the prospective aspect marker *yào*

这是好像要进一个大门。(L1 Ger-L2 Chn P13, A06)

zhe shì hǎoxiàng yào jìn yī-ge dà-mén.
this be perhaps YAO enter one-CL big-gate
It seems that (he) wants to go into a big gate.

(38) With the progressive aspect marker *zài*

* 一个人在进我们大学的行政楼。(L1 Ger-L2 Chn P16, A06)

* yī-ge rén zài jìn wǒmen dàxué-de xíngzhèng-lóu.
one-CL person ZAI enter we university-DE administration-building
A person is entering our university's administration building.

(39) Zero marking

*有一个男人进去一个楼。(L1 Ger-L2 Chn P20, A06)

* you yī-ge nán-rén jìn- qù yī-ge lóu.
exist one-CL male-person enter-go one-CL building.
There is a man who enters a building.

It should be noted that there were word order problems with the placement of the spatial adjunct in the serial verb constructions *jìn qù* ('enter go') or *shàng qù* ('ascend go'), as can be seen in examples 35a, 35c, 36, and 39. When a path verb, a locative nominal phrase, and a deictic verb occur together, the locative nominal phrase must be put between the path verb *jìn* and the deictic verb *qù* (Kang 2001: 287, Li & Thompson 1981: 64, Liu 1998: 34; for more details, see section 3.1.1). L2 Chinese speakers tended to conceptualize this situation as a boundary crossing scenario about to take place, in progress, or undetermined (i.e., it must be inferred from the context). In line with previous findings (e.g., von Stutterheim & Carroll, 2006, von Stutterheim 2003, von Stutterheim & Nüse 2003, von Stutterheim et al. 2012), this may reflect L2 Chinese speakers' focus on the endpoint encoding, which is affected by their L1 patterns of conceptualization (see section 4.4.1.2 about the use of boundary-crossing adjuncts in L1 German for describing Type A videos), since the endpoint is highly profiled in boundary-crossing situations. L2 speakers perceived the endpoint in this situation as particularly significant and encoded it as the goal of the boundary-crossing event, although the man in the video actually never reaches the endpoint.

There were no great differences in the use of other path verbs between the L1 and L2 groups. (no differences between L1 and L2 Chinese speakers in the use of directional path verbs $\chi 2$ (1, N = 11) = 0.818, p = 0.366 and object-related path verbs $\chi 2$ (1, N = 12) = 0.333, p = 0.564; no differences among the three groups in the use of axial path verbs $\chi 2$ (2, N = 32) = 0.063, p = 0.969).

Type B: Long trajectory with possible endpoint

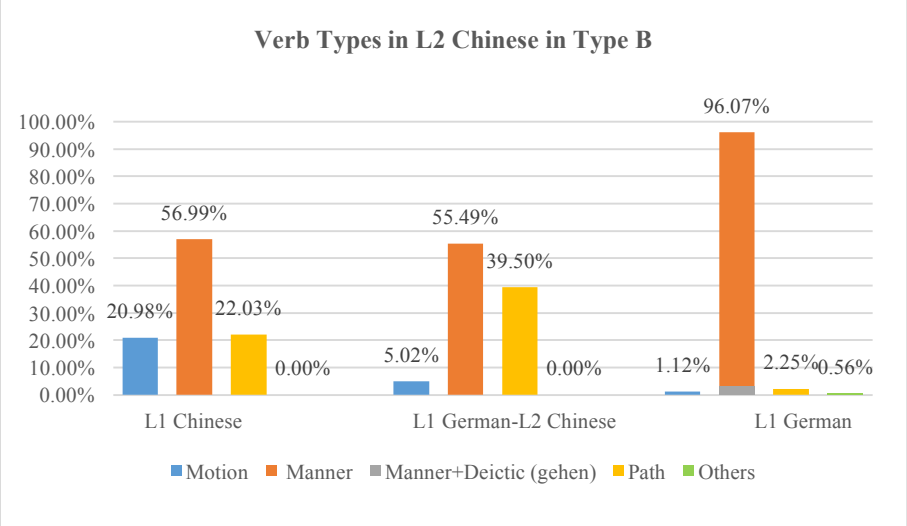


Figure 4.5: Verb types used by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type B videos

As shown in Figure 4.5, in the verbal task of Type B videos, the use of bare motion verbs in L2 Chinese was significantly lower than that in L1 Chinese (L2 Chinese 5.02% vs. L1 Chinese 20.98%, $\chi 2$ (1, N = 76) = 25.474, p < .001). However, its use was significantly higher than that in L1 German (L2 Chinese 5.02% vs. L1 German 1.12%, $\chi 2$ (1, N = 18) = 10.889, p = 0.001). Example:

(40) 车在路上走。(L1 Ger-L2 Chn P13, B01)

chē zài lù-shàng zǒu.

car on road-top move

A car is moving on the road.

zǒu 'move' in this example is a motion verb and encodes no Manner information (for more details, see section 3.1.2).

L1 and L2 Chinese speakers did not differ in the use of manner verbs to describe

Type B videos ($\chi 2$ (1, N = 605) = 0.139, p = 0.709). In contrast, L1 German speakers used significantly more manner verbs than L2 Chinese speakers ($\chi 2$ (1, N = 497) = 89.631, p < .001) and L1 Chinese speakers, respectively ($\chi 2$ (1, N = 464) = 83.061, p < .001). In L1 German, manner verbs occurred frequently with adjuncts either derived from the Ground features or, more specifically, the contours of the Ground object (e.g., *entlangfahren*, 'drive along') or relating to the endpoint (e.g., *fahren zu* / *in Richtung*, 'drive to/towards'). These ways of conceptualization were seen in L2 German-Chinese speaker responses, which appeared in their use of adjuncts to describe Type B videos.

The relative frequency of occurrence of path verbs in L2 Chinese was 39.50%, while it was lower in L1 Chinese (22.03%) and L1 German (2.25%). Statistically, there was a significant difference between L1 and L2 Chinese ($\chi 2$ (1, N = 189) = 21.000, p < .001) and L2 Chinese and L1 German ($\chi 2$ (1, N = 130) = 114.492, p < .001).

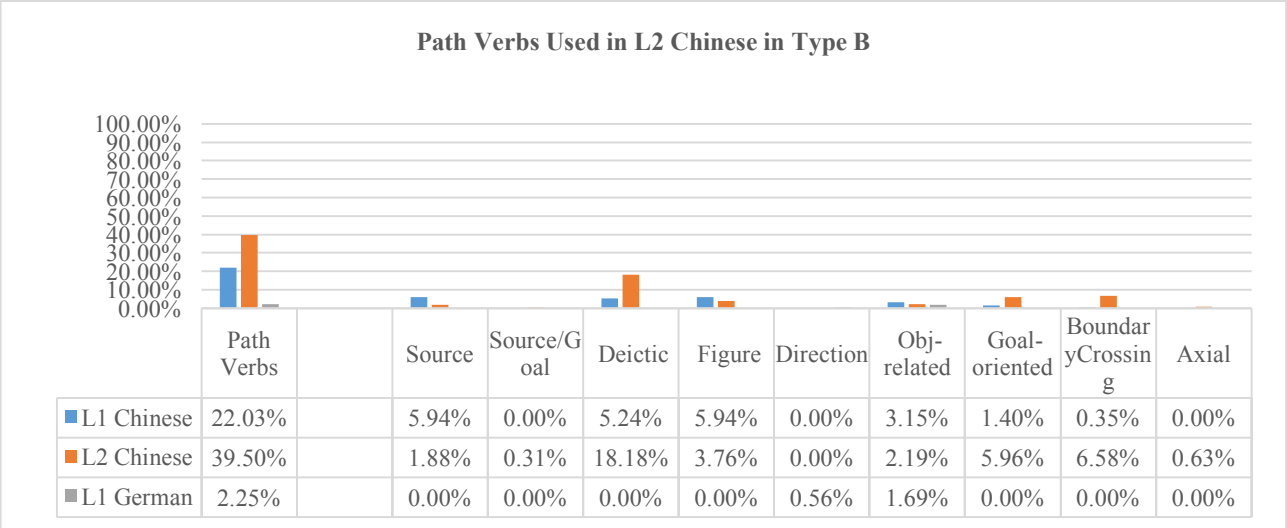
As shown below in Figure 4.6, in general, very few path verbs were used in L1 German in the verbal task of Type B, which is on par with the typological features of German as a type of S-language (see Figure 4.6). When taking a closer look at the path verbs used, it was found that the use of deictic verbs was significantly higher in L2 Chinese than in L1 Chinese (L2 Chinese 18.18% vs. L1 Chinese 5.24%, $\chi 2$ (1, N = 73) = 25.329, p < .001) and in L1 German (L2 Chinese 18.18% vs. L1 German 0%). An example of deictic verbs used in L2 Chinese and L1 German is as follows:

(41) 白色的车子开过去。(L1 Ger-L2 Chn P23, B05)

báisè-de chēzi kāi-guò-qù
white-NOM car drive-go over-go
A white car drives away from the speaker.

In this example, the deictic verb *qù* 'go' is part of the serial verb construction *kāi-guò-qù* 'lit. drive-move over-go'. In this construction, *guò* is also treated as a deictic verb. In fact, depending on the context, the meaning of the path verb *guò* varies. When the Ground object is uttered explicitly, the meaning of the path verb *guò* is considered as deriving from the contours of the Ground object; i.e., it equals the meaning of *cross* or *pass* in English, e.g., *guò* in the sentence *jīng-guò yī-ge cūnzi* 'lit. pass-pass a village'. When the Ground object is not uttered explicitly, the meaning of this path verb is related to the speaker. That is, speakers see a Figure moving from one point to another point

with respect to themselves, as example 41 shows.



	Path Verbs		Source	Source/Goal	Deictic	Figure	Direction	Obj-related	Goal-oriented	BoundaryCrossing	Axial
L1 Chinese	22.03%		5.94%	0.00%	5.24%	5.94%	0.00%	3.15%	1.40%	0.35%	0.00%
L2 Chinese	39.50%		1.88%	0.31%	18.18%	3.76%	0.00%	2.19%	5.96%	6.58%	0.63%
L1 German	2.25%		0.00%	0.00%	0.00%	0.00%	0.56%	1.69%	0.00%	0.00%	0.00%

Figure 4.6: Detailed path verbs used by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type B videos

In addition, there were significant differences in the use of goal-oriented path verbs between L1 and L2 Chinese speakers (L2 Chinese 5.96% vs. L1 Chinese 1.40%, $\chi 2$ (1, N = 23) = 9.783, p = 0.002). Consider example 42:

(42) 一辆汽车开到村子。(L1 Ger-L2 Chn P24, B05)

yī-liàng qìchē kāi-dào cūnzi.

one-CL car drive-arrive village

A car drives to a village.

Additionally, L2 Chinese speakers used significantly more boundary-crossing path verbs than L1 Chinese speakers (L2 Chinese 6.58% vs. L1 Chinese 0.35%, $\chi 2$ (1, N = 22) = 18.182, p < .001). Note that there were no boundary-crossing events depicted in Type B videos; rather the Figure was moving towards a possible Ground object at the endpoint. That is, the distance between the Figure and the Ground object at goal was long. Nevertheless, L2 Chinese speakers used more boundary-crossing path verbs than L1 Chinese speakers. This may be ascribed to the influence exerted by their L1 conceptualization patterns. L1 German speakers tended to take a holistic perspective when observing the situation and then tended to encode the endpoint of that situation (e.g., von Stutterheim & Nüse, 2003; von Stutterheim et al., 2012). In a boundary-crossing event, the endpoint is assumed to be conceptually highly salient and L2

Chinese speakers may perceive the endpoint as highly salient even in Type B videos. Therefore, they used more boundary-crossing path verbs than L1 speakers. As already shown for the descriptions of Type A videos, those verbs occurred together with the modal verbs *dăsuan* 'be going to', the prospective aspect marker *yào*, or zero marking[48] in L2 Chinese. The following examples show the boundary-crossing path verb *jìn rù* 'enter' in L2 Chinese.

(43a) With the modal verb *dăsuan* 'be going to'

汽车往右拐，打算进入城市。(L1 Ger-L2 Chn P09, B01)

qìchē wăng you guăi, dăsuan jìn-rù chéngshì.
car towards right turn, be going to enter-enter city
A car turns right and is going to enter the city.

(43b) With the temporal adverb *kuài* 'almost' and the prospective aspect marker *yào*

一辆车快要进入一个村子。(L1 Ger-L2 Chn P14, B02)

yī-liàng chē kuài yào jìn-rù yī-ge cūnzi.
one-CL car almost YAO enter-enter one-CL village
A car is about to enter a village.

(43c) With zero marking

车子往前走，进一个地方去。(L1 Ger-L2 Chn P17, B01)

chēzi wăng qián zŏu, jìn yī-ge dìfāng qù.
car towards forwards move, enter one-CL place go
A car moves forward and enters a place (away from the speaker).

Statistical significance was also found in the use of source-related path verbs between L1 and L2 Chinese speakers (L2 Chinese 1.88% vs. L1 Chinese 5.94%, $\chi 2$ (1, N = 23) = 5.261, p = 0.022). Example:

[48] See more details about aspect markers in Chinese in Chapter 5.

(44a) ？一辆车在路上开走。(L1 Ger-L2 Chn P03, B04)

yī-liàng chē zài lù-shàng kāi zǒu.
one-CL car on road-top drive leave
A car drives away.

(44b) 一辆白色的汽车开走了。(L1 Chn P06, B04)

yī-liàng báisè de qìchē kāi zǒu le.
one-CL white MOD car drive leave LE
A white car drove away.

zǒu means a motion away from the starting point (see also Lamarre, 2008, p. 73). It occurs frequently with manner verbs. In the above two examples, *zǒu* 'leave' occurs with the manner verb *kāi* 'drive' and requires the perfective aspect marker *le* as in the second example.

There were no significant differences in the use of path verbs related to the orientation of the Figure between L1 and L2 Chinese speakers ($\chi 2$ (1, N = 29) = 0.862, p = 0.353) and in the use of object-related path verbs among the three language groups ($\chi 2$ (2, N = 19) = 2.947, p = 0.229).

Type C: Boundary-crossing events

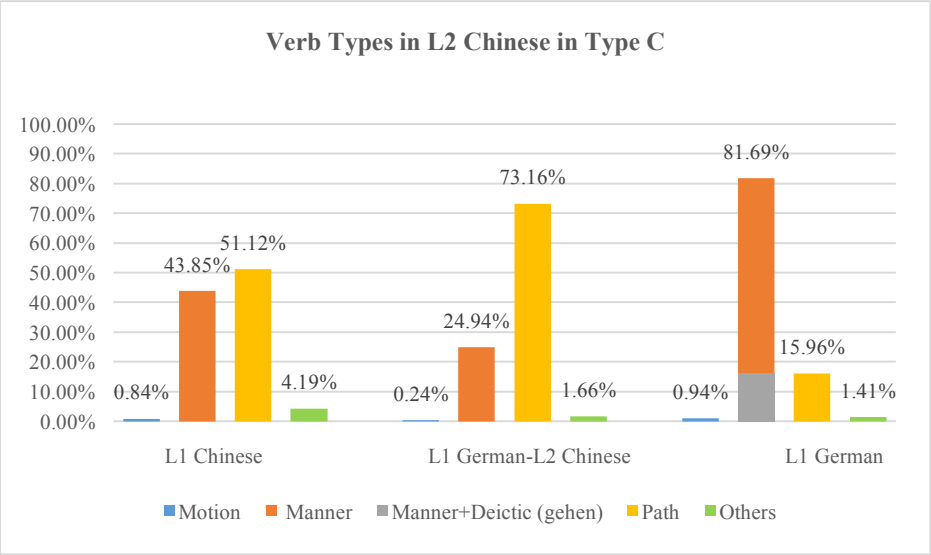
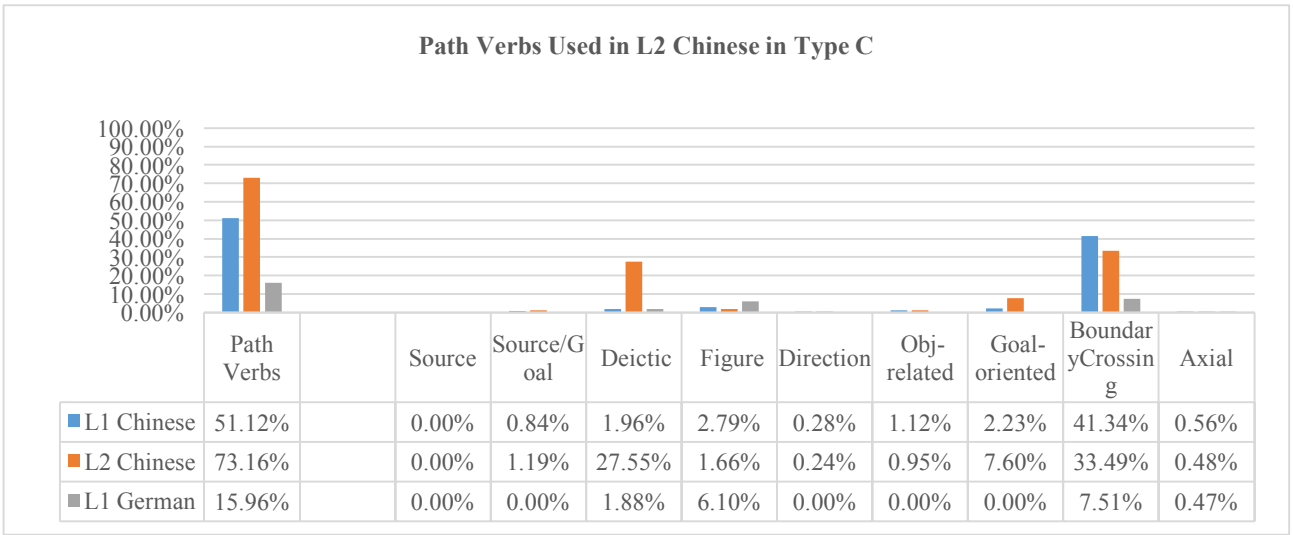


Figure 4.7: Verb types used by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type C videos

To describe Type C videos, L1 Chinese speakers used significantly more manner verbs than L2 Chinese speakers (1, N = 779) = 31.009, p < .001) (see Figure 4.7). The same holds true for L1 German speakers, who used significantly more manner verbs than L2 Chinese speakers (1, N = 634) = 184.860, p < .001). Likewise, in comparison with L1 Chinese speakers, L1 German speakers used more manner verbs (1, N = 571) = 78.461, p < .001). The reason for this is that L2 Chinese speakers tend to use path verbs to express boundary-crossing events, which is evident by the surge in the use of path verbs for Type C videos (73.16%) in comparison to path verbs used for Type A (39.50%) and Type B (43.39%). The use of path verbs by L2 Chinese speakers for Type A, B, and C videos is statistically significant ($\chi 2$ (2, N = 585) = 99.826, p < .001).

Concerning the path verbs used by the L1 and L2 groups, L2 Chinese speakers used them significantly more often than L1 Chinese speakers (L2 Chinese 73.16% vs. L1 Chinese 51.12%, $\chi 2$ (1, N = 491) = 31.823, p < .001) and L1 German speakers (L2 Chinese 73.16% vs. L1 German 15.96%, $\chi 2$ (1, N = 342) = 219.520, p < .001). Figure 4.8 shows the detailed path verbs used by L1 and L2 speakers. Furthermore, no differences were found in the use of motion verbs among the three language groups. However, L1 Chinese speakers differed from L2 Chinese and L1 German speakers in the use of other verbs ($\chi 2$ (2, N = 25) = 8.960, p = 0.011).



	Path Verbs		Source	Source/Goal	Deictic	Figure	Direction	Obj-related	Goal-oriented	BoundaryCrossing	Axial
L1 Chinese	51.12%		0.00%	0.84%	1.96%	2.79%	0.28%	1.12%	2.23%	41.34%	0.56%
L2 Chinese	73.16%		0.00%	1.19%	27.55%	1.66%	0.24%	0.95%	7.60%	33.49%	0.48%
L1 German	15.96%		0.00%	0.00%	1.88%	6.10%	0.00%	0.00%	0.00%	7.51%	0.47%

Figure 4.8: Detailed path verbs used by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type C videos

As shown in Figure 4.8, in regard to the deictic path verbs used, L2 Chinese speakers differed remarkably from L1 Chinese speakers (L2 Chinese 27.55% vs. L1

Chinese 1.96%, $\chi 2$ (1, N = 123) = 96.593, p < .001) and L1 German speakers (L2 Chinese 27.55% vs. L1 German 1.88%, $\chi 2$ (1, N = 120) = 104.533, p < .001). An example of the use of deictic verbs to describe Type C videos follows below:

(45) *一个女人进去超市。(L1 Ger-L2 Chn P20, C04)

*yī-ge nǚ-rén jìn-qù chāoshì.
one-CL female-person enter-go supermarket
A woman entered the supermarket.

As already explained above in examples 35a, 35c, 36, and 39, when path verbs, locative nominal phrases, and deictic verbs occur together, the locative nominal phrase must be put between the path verb *jìn* and the deictic verb *qu* (Kang, 2001, p. 287; Li & Thompson, 1981, p. 64; Liu, 1998, p. 34; for more details, see section 3.1.1).

What is more, the use of goal-oriented path verbs was significantly higher in L2 Chinese than in L1 Chinese (L2 Chinese 7.60% vs. L1 Chinese 2.23%, $\chi 2$ (1, N = 40) = 14.400, p < .001). As example (46) shows, *dào* 'arrive' is a goal-oriented path verb.

(46) 一只猫从走廊走到房间里面。(L1 Ger-L2 Chn P23, C06)

yī-zhī māo cóng zǒuláng zǒu-dào fángjiān-lǐmian.
one-CL cat from corridor walk-arrive room-inside
A cat walked from the corridor into the room.

There were no significant differences in the use of boundary-crossing path verbs between L1 and L2 Chinese ($\chi 2$ (1, N = 289) = 0.170, p = 0.681) (for a boundary-crossing path verb, see example 45). In contrast, L2 Chinese speakers used significantly more boundary-crossing path verbs than L1 German speakers ($\chi 2$ (1, N = 157) = 99.522, p < .001), since the information on boundary-crossing is typically encoded in path verbs in Chinese, whereas the same information is encoded in adjuncts and particles in German. There were no significant differences in the use of path verbs related to the orientation of the figure ($\chi 2$ (2, N = 30) = 1.800, p = 0.407), object-related verbs (L2 Chinese 0.95%, L1 Chinese 1.12%, and L1 German 0%), and axial path verbs (L2 Chinese 0.48%, L1 Chinese 0.56%, and L1 German 0.47%) among the three language

groups.

Summary

To summarize, L1 German speakers used more manner verbs than L1 and L2 Chinese speakers, respectively, to describe different video types. L1 Chinese speakers used more manner verbs than L2 Chinese speakers for Type A and C videos, while L1 and L2 Chinese speakers did not differ in this regard for Type B videos. Concerning the detailed path verbs used, L2 Chinese speakers used significantly more deictic verbs and goal-oriented path verbs than L1 Chinese speakers to describe Type A, B, and C videos, respectively. Furthermore, the occurrence of boundary-crossing path verbs was significantly higher in L2 Chinese than in L1 Chinese for Type A and B videos, while no significant differences were found between L1 and L2 Chinese in the use of boundary-crossing path verbs for Type C videos. The greater use of goal-oriented path verbs for all video types and of boundary-crossing path verbs for Type A and B videos in L2 Chinese reflects their L1s' conceptualization pattern, that is, the preference for a holistic view to describe motion events. This preference is especially evident for Type B videos, which show a long trajectory with a possible endpoint: These speakers chose to use goal-oriented and boundary-crossing path verbs to encode the endpoint, even though the endpoint was far away and not evident (see section 4.4.2.2 for the frequency of the endpoint mentioned by L2 Chinese). This raises the question why L2 Chinese speakers preferred using deictic path verbs[49] throughout different types of videos. In order to shed light on the reasons for their frequent use in L2 Chinese, we must first find out when and where deictic verbs occur in the data used in this study.

In the collected data, there is a total of 29 utterances in L1 Chinese and 206 utterances in L2 Chinese with *lái* and *qù* that describe motion events in Type A, B, and C videos. Table 4.4 shows the distribution of *lái*/*qù* for the three video types. *lái*/*qù* can occur alone, with path verbs (manner verbs may or may not appear in this case), or with manner verbs. In Table 4.4, apart from the percentages, the number of utterances with deictic verbs in the corresponding category (first number in parentheses) and the total number of utterances with deictic verbs for each video type (second number in parentheses) are also provided.

[49] Although deictic verbs in this study include *lái* 'come'/*qù* 'go', and *guò* 'come/ go over', the occurrence of *lái*/*qù* is significantly higher than that of *guò* in L2 Chinese. As already explained, when the Ground object is not explicitly uttered, then *guò* has a deictic meaning and *lái*/*qù* is typically added after it. In this sense, *lái*/*qù* and *guò* are closely related when both have a deictic meaning. Therefore, although Table 4.4 shows the distribution of *lái*/*qù* for the description of different video types, it also takes into account the path verb *guò* with a deictic meaning, so that we can get a complete picture of the deictic verbs used in L1 and L2 Chinese.

Table 4.4: Distribution of *lái* and *qù* in the verbal task of Type A, B, and C videos in L1 and L2 Chinese

	L1 Chinese			L2 Chinese		
	Type A	Type B	Type C	Type A	Type B	Type C
Total number of utterances with *lái/qù*	14	9	6	60	38	108
***lái/qù* as full verbs occur alone**	14.29% (2/14)	0% (0/9)	16.67% (1/6)	35% (21/60)	21.05% (8/38)	12.04% (13/108)
***lái/qù* occur with the following path verbs:** 过 *guò* 'cross, pass, move over'	35.71% (5/14)	11.11% (1/9)	0% (0/6)	30% (18/60)	50% (19/38)	6.48% (7/108)
进 *jìn* 'enter'	0% (0/14)	0% (0/9)	66.67%[50] (4/6)	15% (9/60)	13.16% (5/38)	69.44% (75/108)
出 *chū* 'exit'	7.14% (1/14)	11.11% (1/9)	0% (0/6)	1.67% (1/60)	0% (0/38)	0.93% (1/108)
回 *huí* 'return'	0% (0/14)	0% (0/9)	0% (0/6)	0% (0/60)	0% (0/38)	1.85% (2/108)
到 *dào* 'arrive'	0% (0/14)	0% (0/9)	0% (0/6)	5% (3/60)	5.26% (2/38)	3.70% (4/108)
上 *shàng* 'ascend'	0% (0/14)	0% (0/9)	0% (0/6)	11.67% (7/60)	5.26% (2/38)	1.85% (2/108)
***lái/qù* occur with manner verbs:** manner verbs+ *qù*	42.86% (6/14)	77.78% (7/9)	16.67% (1/6)	1.67% (1/60)	5.26% (2/38)	3.70% (4/108)

As shown in Table 4.4, an in-depth look at the utterances with deictic verbs for Type A videos reveals that L2 speakers used *lái*/*qù* significantly more often as stand-alone main verbs in their utterances (L2 Chinese 35% vs. L1 Chinese 14.29%, $\chi 2$ (1, N = 23) = 15.696, p < .001) or together with the path verb *guò* 'cross, pass' than L1 Chinese speakers ($\chi 2$ (1, N = 23) = 7.348, p = 0.007). In addition, *lái*/*qù* also occurred with *jìn* 'enter' (15%) and *shàng* 'ascend' (11.67%) in L2 Chinese. In contrast, there are no occurrences of *jìn* 'enter' (15%) (see examples 35a, 35c, 36, and 39) and *shàng* 'ascend' in conjunction with *lái*/*qù* in L1 Chinese for Type A videos. Although the total number of utterances with *lái*/*qù* was low in L1 Chinese, 6 out of 14 utterances (42.85%) show the co-occurrence of deictic verbs with manner verbs in L1 Chinese, while only 1 out of 60 utterances (1.67%) shows the same combination pattern in L2 Chinese.

The occurrence of *lái*/*qù* in Type A is exemplified in L2 Chinese as follows:

(47a) Type A: *lái*/*qù* occurs alone as a main verb:

这里有一个公车来了。(L1 Ger-L2 Chn P02, A09)

[50] Note that there are only 6 utterances with *lái*/*qù* in Type C in L1 Chinese, of which 4 utterances include *jìn* and *lái*/*qù*. So the percentage is high here.

zhèli yǒu yī-ge gōngchē lái le.
here exist one-CL bus come LE
Here came a bus.

有人去公车站。(L1 Ger-L2 Chn P07, A02)

yǒu rén qù gōngchēzhàn .
exist person go bus station
Someone goes to a bus station.

(47b) Type A: *lái*/*qù* occurs with the path verb *guò* 'cross, move over, pass':

一个人沿着一个老房子或者教堂走过去。(L1 Ger-L2 Chn P23, A02)

yī-ge rén yán-zhe yī-ge lǎo fángzi huòzhě jiàotáng zǒu-guò-qù.
one-CL person along-ZHE one-CL old house or church walk-move over-go
A person walks past an old house or church (away from the speaker).

(47c) Type A: *lái*/*qù* occurs with the path verb *shàng* 'ascend':

有一个男人在一个梯子爬上去，爬到一个阳台上吧。(L1 Ger-L2 Chn P11, A03)

yǒu yī-ge nán-rén zài yī-ge tīzi pá-shàng-qù,
exist one-CL male-person at one-CL ladder climb-ascend-go
There is a man who climbs a ladder,

pá-dào yī-ge yángtái-shàng ba.
climb-arrive one-CL balcony-top SA[51]
(he) climbs onto a balcony.

(47d) Type A: *lái*/*qù* occurs with a manner verbs:

一个女生向汽车走去。 (L1 Chn P19, A04)

yī-ge nǚshēng xiàng qìchē zǒu-qù.

[51] SA is a short form for solicit agreement. According to Li and Thompson (1981, p. 307), *ba* has the function of soliciting the agreement or the approval of the listener to an utterance to which *ba* is attached. It equals English expressions like 'don't you think so' or 'wouldn't you agree'.

one-CL girl towards car walk-go

A girl walks towards a car (away from the speaker).

In the verbal task for Type B videos, the co-occurrence of the path verb *guò* 'cross, pass' in conjunction with the deictic verbs *lái*/*qù* in L2 Chinese was significantly higher than in L1 Chinese ($\chi 2$ (1, N = 20) = 16.200, p < .001). Compared to L1 Chinese, there 5 out of 38 utterances (13.16%) show the co-occurrence of *jìn* 'enter' with the deictic verbs *lái*/*qù*, 8 out of 38 utterances (21.05%) show the occurrence of the deictic verbs *lái*/*qù* alone, and 2 out of 38 utterances (5.26%) show the co-occurrence of *dào* 'arrive', *shàng* 'ascend', and manner verbs with *lái*/*qù* in L2 Chinese. In contrast, there is no occurrence of these combinations in L1 Chinese. Instead, in L1 Chinese, 7 out of 9 phrases include deictic verbs with manner verbs. This is shown in the following examples (see example 43c for the co-occurrence of *jìn* 'enter' with the deictic verbs *lái*/*qù*):

(48a) Type B: *lái*/*qù* occurs alone as a main verb:

有一个女生在外面骑自行车，她是去一个农村的地方。(L1 Ger-L2 Chn P22, B09)

yǒu yī-ge nǚshēng zài wàimian qí zìxíngchē, tā shì qù yī-ge nóngcūn-de dìfāng.
exist one-CL girl at outside ride a bicycle, she be go one-CL village-NOM place
There is a girl who is riding a bicycle outside and going to a village-like place.

(48b) Type B: *lái*/*qù* occurs with the path verb *guò* 'cross, come/go over, pass':

也是在农村，有一个小卡车开过去。 (L1 Ger-L2 Chn P11, B08)

yě shì zài nóngcūn, yǒu yī-ge xiǎo kǎchē kāi-guò-qù
also be at village, exist one-CL little truck drive-go over-go
There is a little truck that also drives to a village (away from the speaker).

(48c) Type B: *lái*/*qù* occurd with the path verb *shàng* 'ascend':

*女生上去路那边。(L1 Ger-L2 Chn P12, B03)

*nǚshēng shàng-qù lù nà-bian.
girl ascend-go road that-side

A girl goes up to the road over there.

(48d) Type B: *lái*/*qù* occurs with a manner verb:

有公车开去那个村子。(L1 Ger-L2 Chn P07, B06)

yǒu gōngchē kāi-qù nà-ge cūnzi.
exist bus drive-go that-CL village
There is a bus that drives towards that village (away from the speaker).

In the verbal task for Type C videos, despite the extreme low occurrence of deictic verbs in L1 Chinese, both L1 and L2 speakers tended to use *lái*/*qù* in conjunction with the path verb *jìn* 'enter' because *jìn* 'enter' implies the crossing of a boundary, which corresponds to the characteristics of Type C videos. However, L2 Chinese speakers used *lái*/*qù* significantly more often in conjunction with the path verb *jìn* 'enter' than L1 Chinese speakers ($\chi 2$ (1, N = 79) = 63.810, p < .001), while for Type C videos, L1 Chinese speakers preferred using a construction manner verb + the boundary-crossing path verb *jìn* + the Ground object (for more details, see the use of SVCs in L1 and L2 Chinese below). In addition, L2 Chinese speakers used *lái*/*qù* significantly more often alone as a main verb in their utterances than L1 Chinese speakers ($\chi 2$ (1, N = 14) = 10.286, p = 0.001). 7 out of 108 utterances (6.48%) contain the path verb *guò* 'cross, pass' and 4 out of 108 utterances (3.70%) the goal-oriented path verb *dào* 'arrive', whereas no corresponding data were registered in L1 Chinese.

To sum up, L2 speakers tended to use the deictic verbs *lái*/*qù* in the following circumstances:

1) as a main stand-alone verb for Type A, B, and C videos;
2) in conjunction with the path verb *guò* 'cross, pass' for Type A and B videos;
3) in conjunction with the boundary-crossing path verb *jìn* 'enter' in Type C videos.

We can explain L2 speakers' preference for the deictic verbs *lái*/*qù* with respect to the directionality that these two deictic verbs indicate, their tendency to take a holistic perspective (focus on the endpoint), and the phonological and morphological features of modern Chinese.

The reason for L2 Chinese speakers' preference to use deictic verbs as stand-

alone verbs may stem from the internal meaning of the deictic verbs *lái*/*qù*. Both of the deictic verbs *lái* 'come' and *qù* 'go' are directional in nature: *lái* indicates the direction towards the speaker and *qù* the direction away from the speaker. When speakers use *lái*, it implies that the position of the speaker is the endpoint if there is no other explicit endpoint; when speakers use *qù*, the position of the speaker serves as the source and the endpoint should be mentioned in order to form an informative expression[52] (see examples 47a, 48a). In this sense, the use of *lái*/*qù* reflects L2 speakers' preference for a holistic perspective and their tendency to view the situation as an entirety[53]. In this study, most of the video clips showed a motion away from the speaker, with only one video clip showing a motion towards the speaker (this video showed a bus coming towards a bus stop). Accordingly, there are only two statements in which the deictic verb *lái* occurs alone (see *lái* in example 47a). In contrast, there are 40 statements in which the deictic verb *qù* functions as a full verb. L2 Chinese speakers used the deictic verb *qù* significantly more often than L1 Chinese speakers ($\chi 2$ (1, N = 43) = 31.837, p < .001). In addition, it is assumed that the relatively more frequent use of the deictic verb *qù* in L2 Chinese might correspond to the use of *gehen* 'go by walking' in L1 German. Throughout the description of the video clips, the German verb *gehen* occurred at a rate of 14.67%. Although it is a manner verb, it has a deictic meaning, which denotes a motion away from the speaker. The use of *gehen* might trigger the use of *qù* to express a movement away from the speaker in L2 Chinese.

Regarding the tendency of L2 Chinese speakers to use the deictic verbs *lái*/*qù* in conjunction with the path verb *guò* 'cross, pass' to describe Type A and B videos, and especially Type B, we could interpret this as a reflection of their preference for taking a holistic viewpoint as in their L1s. What is different with stand-alone deictic verbs is that it is not syntactically required to add an endpoint after the serial verb construction. Path verbs like *guò* 'cross, pass' or *jìn* 'enter' syntactically require either a spatial adjunct (i.e., a locative noun phrase in this case) or the deictic verb *lái* or *qù* after them in order to form a grammatically acceptable sentence (Lamarre, 2008, p. 79;

[52] One can also say *tā lái*/*qù le*. ('lit. he come/go LE'). It is grammatically acceptable, but the context is necessary to get a full understanding of this sentence.

[53] Slobin (2006, p. 67) pointed out that "[t]ypically, in verb-framed languages, a neutral verb of motion is used to designate a creature's normal manner of movement: owls 'go', fish 'go', people 'go', cats 'go', and so forth." It is speculated that the "neutral verb of motion" in Slobin's terms might be relatively easy for learners to use, at least for L1 German-L2 Chinese speakers in this study, because it does not include any manner information, that is, the semantic meaning in these "neutral verbs of motion" is not as dense as in manner verbs like 'flap' or 'plunge'. But more investigation is needed on the use of "neutral verbs of motion" and manner verbs with a dense meaning in second language acquisition.

Liu, 1998, p. 37; Qi, 1998, p. 180; Ying, 1988, p. 260). That is to say, the spatial adjunct, i.e., the endpoint, could be omitted, but *lái*/*qù* must be added after the path verb to replace the position of the endpoint if the endpoint is not in focus, has already been mentioned and can be inferred from the context, or the speaker does not want to mention it explicitly. In other words, *lái*/*qù* can represent an implicit endpoint that can be inferred from the context. In this sense, the use of the serial verb construction *guò qù* [54] 'pass/cross go', especially in the verbal task for Type B videos by L2 Chinese (at a rate of 50%), may reflect that L2 speakers tend to view the situation in its entirety. Since in this compact construction, – i.e., manner verb + path verb *guò* + deictic verb *qù* or manner verb + path verb *guò* + Ground object + deictic verb *qù* – depending on the context, *guò* is either an object-related path verb whose meaning is derived from the features of the Ground object or a deictic verb whose meaning is related to the speaker, *qù* represents the implicit endpoint in this construction. By using *qù* L2 speakers do not need to mention the endpoint explicitly. Therefore, the frequent use of *qù* reflects L2 speakers' tendency to view a situation holistically. In addition, L1 German speakers preferred encoding object-related, directional, and goal-oriented concepts in their descriptions of Type B videos (see section 4.4.1.2 about the use of adjuncts in German), while in Chinese, the conceptual component *guò* is either related to the features of the Ground object or the direction characterized by the speaker, and *qù* implies the deictic direction and is related to the implicit endpoint. The conceptual components indicated by *guò* and *qù* are the ones preferred in L1 German, so L1 German-L2 Chinese speakers prefer encoding these components when they speak Chinese. Accordingly, L2 Chinese speakers tend to conform to their L1 conceptual patterns and encode them frequently by different means.

Furthermore, L2 speakers tend to use the serial verb construction *jìn qù* ('enter go') to describe boundary-crossing situations. The difference in the frequent omission of the endpoint in the verb serial *guò qù* 'cross/pass go' for Type A and B videos lies in the fact that the verb serial *jìn qù* 'enter go' occurs frequently together with the endpoint in L2 Chinese. This may be explained by the fact that the endpoint is evident and speakers actually see the moving entities crossing the boundary defined by the Ground object at the goal. Therefore, the endpoint is so salient compared with the

[54] All the Type B videos showed the moving direction away from the speaker, so all the speakers used *qù* when they used a deictic verb.

endpoint in Type A and B videos that speakers cannot neglect it. Although it is syntactically allowed to omit the endpoint in the expression *jìn qù*, L2 speakers tend not to omit it. The frequent use of the deictic verb *qù* in the verb serial *jìn qù* might be a strategy to "compensate" for the adjunct function in German, like *Eine Frau geht in das Haus* 'a woman walks into the house'. This is because this German example implies that this woman is on her way to the house and into the house. In other words, this example indicates the direction to the house and the change of place from the outside to the inside of the house. The corresponding expression in Chinese without the deictic verb *qù* – for example, *yī-ge nǚrén zǒu jìn zhè-ge fángzi* 'a woman walks into the house' – only implies the change of position from the outside to the inside of the house. L2 Chinese speakers preferred additionally using the deictic verb *qù* in the statement *yī-ge nǚrén zǒu jìn zhè-ge fángzi qù* or *yī-ge nǚrén zǒu jìn qù* to express direction. This type of expression shows L1 German speakers' preference for encoding directionality. Furthermore, it should be pointed out that L1 Chinese speakers tended to use the construction "manner verb + boundary-crossing path verb + endpoint" (see the frequency of occurrence of SVCs in the following), while L1 German-L2 Chinese speakers tended to use the construction "boundary-crossing path verb + deictic verb + endpoint"[55], leaving the manner information implicit.

Besides the aforementioned two reasons (i.e., directionality indicated by a deictic verb and tendency to take a holistic perspective), the phonological and morphological status of modern Chinese may also contribute to the frequent use of deictic verbs in L2 Chinese. The old Chinese is monosyllabic. However, many monosyllabic words are pronounced in the same way. In order to distinguish them, those monosyllabic words have changed into two- or polysyllabic words over time. This shows a tendency in modern Chinese: Chinese tends to be disyllable or bipartite (Li & Thompson, 1981). Because of the prevalence of disyllable or bipartite words in Chinese, such as *jìn qù* or *guò qù*, L2 speakers may treat them as one lexical unit[56], so they do not want to insert any constituents between non-deictic path verbs and *lái*/*qù*. This may explain why L2 speakers make mistakes regarding the placement of locative NPs: they

[55] Note that this is a construction actually used in L2 Chinese that has a word order problem. The right order should be "boundary-crossing path verb + endpoint + deictic verb".

[56] Treating verb serials such as *jìn qù* or *guò qù* as one lexical unit is also possible in other equipollently-framed languages (e.g., Thai) and bipartite-verb languages (e.g., Tsou). As Slobin (2006, p. 69) observed, constructions such as 'fly exit come' "can probably be treated as quasi-lexical units in such languages" because such a construction could easily provide manner, path, and deictic information together.

treat assumedly non-deictic path verbs and deictic verbs together as one lexical unit and tend to put the direct object immediately after this lexical unit, rather than insert it in the middle. In a recent study about the morphological status of serial verb constructions, Li (2018) showed that when Chinese native speakers were asked to decide whether the second word in a serial verb construction was a Chinese word and were then asked to pronounce this word, 63.4% of the Chinese native speakers said that in the construction "manner verb + non-deictic path verb + deictic verb", the combination of the non-deictic path verb and the deictic verb was a Chinese word: for example, *jìn qù* 'enter go' in the serial verb construction *zǒu jìn qù* 'walk enter go'. Li's findings support our explanation: L2 Chinese speakers assumedly treat the combination of non-deictic path verb and deictic verb as one lexical unit. Table 4.5 displays the word order problems encountered by L2 Chinese speakers.

Table 4.5: Total number of utterances with the construction '(manner verb) + non-deictic path verb + place NP+ deictic path verb' and number of statements with appropriate and inappropriate word order in L1 and L2 Chinese, respectively

	L1 Chinese			L2 Chinese		
	Total	Appropriate word order	Inappropriate word order	Total	Appropriate word order	Inappropriate word order
(V_{manner})+*jìn*+placeNP+ *lái/qù* 'enter come/go'	1	1	0	87	12	75
(V_{manner})+*shàng*+place NP+ *lái/qù* 'ascend come/go'	0	0	0	7	2	5
(V_{manner})+*dào*+placeNP + *lái/qù* 'arrive come/go'	0	0	0	5	5	0

In Table 4.5, three types of serial verb constructions are listed, more specifically, directional verb compounds: (V_{manner}) + *jìn* + placeNP + *lái/qù;* (V_{manner}) + *shàng* + placeNP + *lái/qù;* (V_{manner}) + *dào* + placeNP + *lái/qù.* Manner verbs in parentheses mean that they could be omitted. Directional verb compounds that include the non-path verbs *guò* 'cross/pass', *huí* 'return', and *chū* 'exit' do not appear in this table because these path verbs either occurred without deictic verbs (e.g., *jīng-guò cūnzi* 'pass-pass village') or with deictic verbs but without mention of the Ground NP (e.g., *zǒu-guò-qù* 'walk-pass-go', *chū-lái* 'exit-come', *huí-lái* 'return-come') in L1 and L2 Chinese. L2 speakers exhibited high accuracy in using these expressions, as opposed to the constructions listed in Table 4.5. Therefore, directional verb compounds with *guò* 'cross/pass', *huí* 'return', and *chū* 'exit' are not included in Table 4.5. In addition, the

total number of statements with the listed constructions in Table 4.5 in L1 Chinese is extremely low because there are few statements with deictic verbs in L1 Chinese (29 statements with deictic verbs in total; for more details, see Table 4.4) and the directional verb compounds in those statements tend to appear in the following constructions in L1 Chinese: manner verb + deictic verb and (manner verb) + non-deictic path verb + deictic path verb.

Table 4.5 shows that L2 Chinese speakers had difficulties inserting place NPs into directional verb compounds, especially in this construction: (V_{manner}) + *jìn* + placeNP + *lái/qù*. The total number of statements with this construction amounts to 87, of which 75 statements have word order problems. Example 45 is repeated here as follows:

(49) *一个女人进去超市。(L1 Ger-L2 Chn P20, C04)

*yī-ge nǚ-rén jìn-qù chāoshì.
one-CL female-person enter-go supermarket
A woman enters a supermarket.

To make this sentence correct, the place NP *chāoshì* 'supermarket' should be inserted between the non-deictic path verb *jìn* 'enter' and the deictic verb *qù* 'go'.

In addition, L2 Chinese speakers had word order problems with the following construction: (V_{manner}) + *shàng* + placeNP + *lái/qù*. In 5 out of 7 statements, the word order is inappropriate. Example 48c is repeated here.

(50)*女生上去路那边。(L1 Ger-L2 Chn P12, B03)

*nǚshēng shàng-qù lù nà-bian.
girl ascend-go road that-side
A girl goes up to the road over there.

To make this sentence correct, the place NP *lù nà-bian* 'lit. road that-side' should be put between *shàng* 'ascend' and *qù* 'go'. The non-target-like word order problems with place NPs might be caused by the overgeneralization of the positions of object NPs. The placement of object NPs is more flexible than the positions of place

NPs in the sentence (for more details, see section 3.1.1). When L2 Chinese speakers do not heed the type of NPs, they are susceptible to producing unacceptable sentences. In contrast, if the position for place NPs or object NPs in a construction is fixed, they might not make similar mistakes. This is demonstrated in the construction (V_{manner}) + *dào* + placeNP + *lái/qù*. In this construction, NPs must follow *dào* and precede *lái/qù*. Any other positions are not allowed. For such a fixed position, it is relatively easy for L2 Chinese speakers to master the construction. 5 out of 5 statements have the appropriate word orders for this construction.

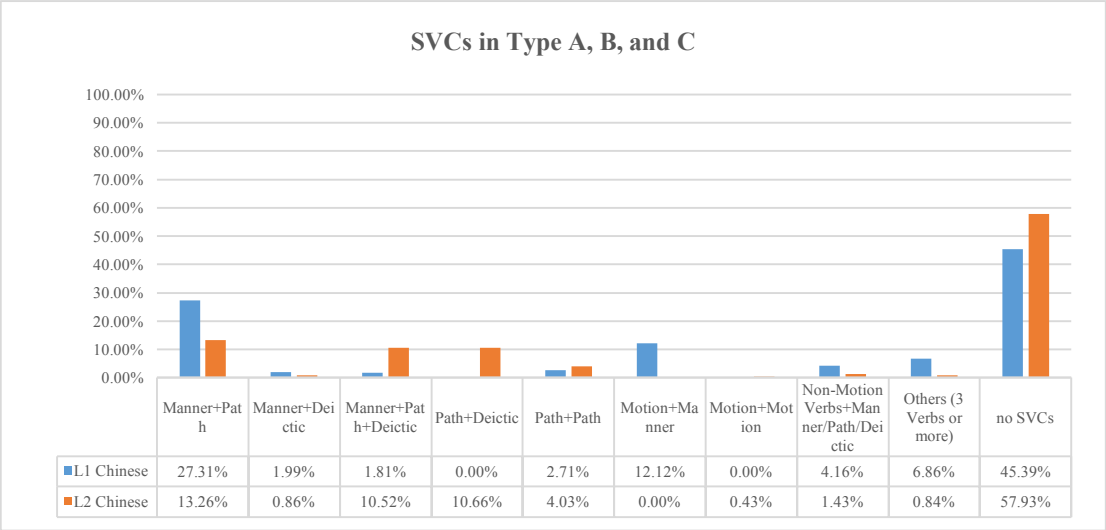
The same word order problems have also been found in L1 Japanese-L2 Chinese (Qian, 1997) and L1 English-L2 Chinese (Wu, 2011, 2014; Yang, 2003). Like L1 German-L2 Chinese speakers, those L2 speakers might also treat the co-occurrence of non-deictic path verbs and deictic verbs in serial verb constructions (e.g., *jìn qù* 'enter go' or *guò qù* 'pass/cross go') as one lexical unit, thus making the same mistakes regarding the placement of place NPs. Additionally, the syntactic complexity of directional verb compounds and the word order rules associated with different types of NPs further complicate the picture and make it extremely difficult for L2 speakers to learn them correctly (for more details, see section 3.1.1).

The deictic verbs discussed above constitute serial verb constructions (SVCs) in Chinese[57]. However, SVCs do not always include deictic path verbs. Instead, SVCs may contain two or more verbs that encode Manner and Path, Motion and Path, Motion and Motion, Path and Path, etc. Figure 4.9 provides an overview of the SVCs used in L1 and L2 Chinese.

In general, the occurrence of SVCs in L1 Chinese was found to amount to 54.61%, while in L2 Chinese, its use amounted to 42.07% (see Figure 4.9). No significant difference was found in the use of SVCs between the L1 and L2 Chinese groups ($\chi 2$ (1, N = 588) = 0.027, p = 0.869). In Figure 4.9, no SVCs means that manner, path, or deictic verbs occurred alone with or without adjuncts (for adjuncts, see section 4.4.1.2) in L1 and L2 Chinese. Since they are single verbs in the sentence, they are not counted as SVCs. An in-depth look at the SVCs used shows that L1 and L2 speakers had different preferences for the different types of SVCs. L1 Chinese speakers used significantly more SVCs encoding Manner and Path than L2 speakers ($\chi 2$ (1, N = 243)

[57] There is only one instance where a SVC was used in German in the data underlying this study: *Ein Bus kommt eine Straße entlang gefahren, auf eine Haltestelle zu*. 'A bus comes and drives along the street, towards a bus stop.' (L1 Ger P01, A09).

= 14.325, p < .001; see example 51), while the use of SVCs encoding Manner, Path, and Deictic in L2 Chinese (example 52) was significantly higher than in L1 Chinese ($\chi 2$ (1, N = 83) = 47.819, p < .001). In addition, SVCs with Path and Deictic encodings occurred at a rate of 10.66% in L2 Chinese (example 53), whereas no such data is found in L1 Chinese.



	Manner+Path	Manner+Deictic	Manner+Path+Deictic	Path+Deictic	Path+Path	Motion+Manner	Motion+Motion	Non-Motion Verbs+Manner/Path/Deictic	Others (3 Verbs or more)	no SVCs
L1 Chinese	27.31%	1.99%	1.81%	0.00%	2.71%	12.12%	0.00%	4.16%	6.86%	45.39%
L2 Chinese	13.26%	0.86%	10.52%	10.66%	4.03%	0.00%	0.43%	1.43%	0.84%	57.93%

Figure 4.9: Total number of SVCs used by L1 and L2 Chinese speakers in the verbal task of Type A, B, and C videos (in this Figure, Path means the non-deictic Path)

This is shown in the following examples:

(51) SVCs: Manner + Path

一条狗跑进了一个房子。(L1 Chn P15, C05)

yī-tiáo gǒu pǎo-jìn le yī-ge fángzi.

one-CL dog run-enter LE one-CL house

A dog ran into a house.

(52) SVCs: Manner + Path + Deictic

一只狗跑进房子里去。(L1 Ger-L2 Chn P03, C05)

yī-zhī gǒu pǎo-jìn fángzi-lǐ qù.

one-CL dog run-enter house-inside go

A dog runs into the house (away from the speaker).

(53) SVCs: Path + Deictic

*小孩进去玩的地方。(L1 Ger-L2 Chn P12, C07)

*xiǎohái jìn-qù wán-de dìfāng.
child enter-go play-NOM place
A child enters a playground.

Example 53 has word order problems. *wán-de dìfāng* 'lit. play-NOM place' should be put between the path verb *jìn* and the deictic verb *qù*.

Furthermore, L2 speakers used significantly more SVCs with Path and Path encodings than L1 Chinese speakers ($\chi 2$ (1, N = 43) = 3.930, p = 0.047) (example 54). In contrast, no SVCs with motion and manner verbs were found in L2 Chinese, whereas L1 Chinese speakers tended to use them (12.12%) (example 55).

(54) SVCs: Path + Path

一个比较旧的车进入一个村子。(L1 Ger-L2 Chn P14, B08)

yī-ge bǐjiào jiù-de chē jìn-rù yī-ge cūnzi.
one-CL relatively old-MOD car enter-enter one-CL village
A rather old car enters a village.

(55) SVCs: Motion + Manner

公交车正在行驶。(L2 Chn P20, B06)

gōngjiāochē zhèng zài xíng-shǐ.
bus ZHENG ZAI move-drive
A bus is driving along.

Despite the low occurrence of other SVCs, it was found that L1 Chinese speakers used more SVCs consisting of three or more verbs ($\chi 2$ (1, N = 25) = 6.760, p = 0.009; see example 56) and non-motion verbs with manner or path verbs ($\chi 2$ (1, N = 33) = 5.121, p = 0.024; see example 57) than L2 Chinese speakers. Non-motion verbs are mainly 'carry' verbs in Chinese. Two video clips were responsible for this use. One video clip showed a woman walking towards a trash can with a bag and the other showed a woman running into a train station with luggage. This is shown by the

following examples:

(56) SVCs: Manner + Path + Path

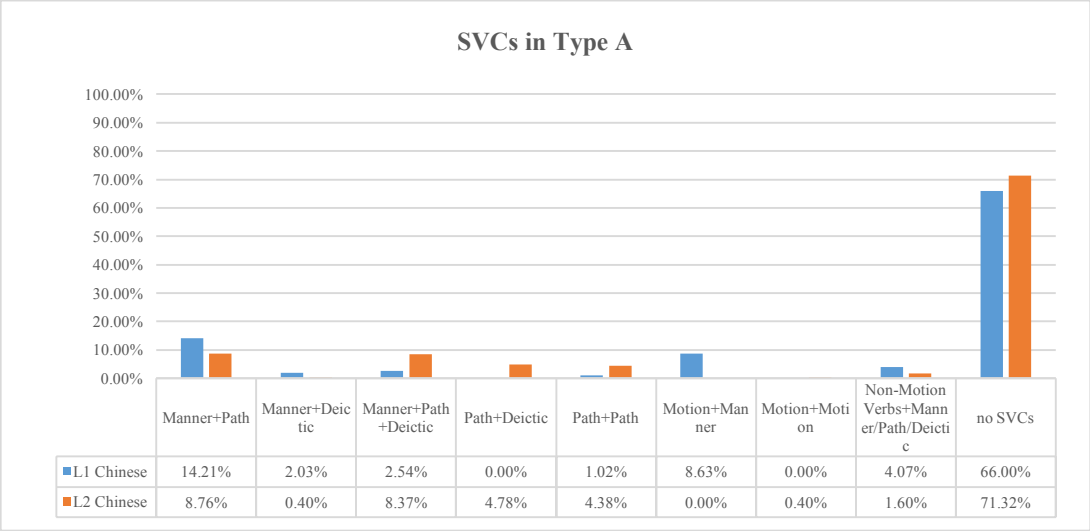
一个女孩骑着马进入了一个房子里。(L1 Chn P11, C10)

yī-ge nǚhái qí-zhe mǎ jìn-rù le yī-ge fángzi-lǐ.
one-CL girl ride-ZHE horse enter-enter LE one-CL house-inside
A girl riding a horse entered a house.

(57) SVCs: non-motion verb + Manner + Path + Deictic

一个女人拎着蓝色的手提袋朝垃圾桶走过去。(L1 Chn P20, A10)

yī-ge nǚ-rén līn zhe lánsè-de shǒutídài cháo lājītǒng zǒu-guò-qù.
one-CL female-person carry ZHE blue-NOM bag towards trash can walk-pass-go
A woman walks towards a trash can with a blue bag (away from the speaker).

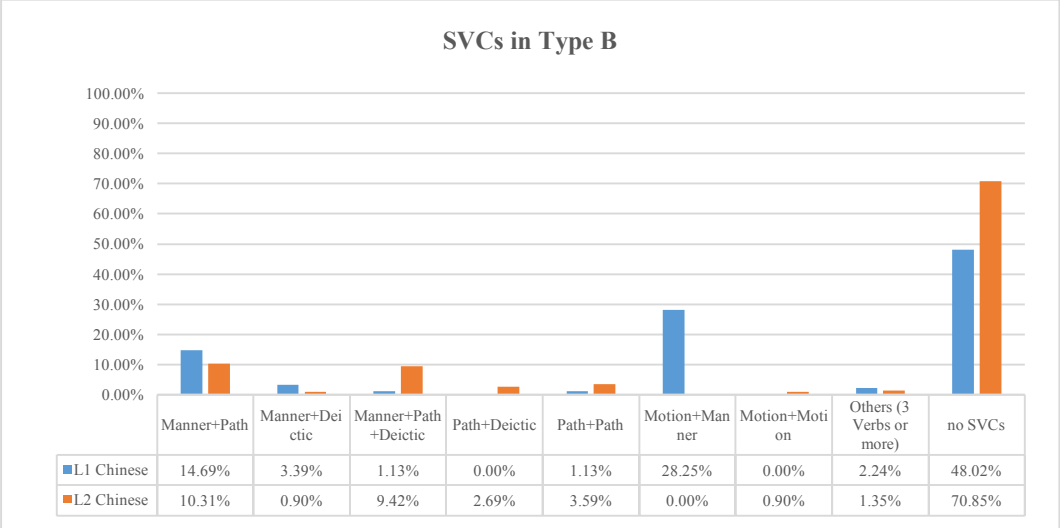


	Manner+Path	Manner+Deictic	Manner+Path+Deictic	Path+Deictic	Path+Path	Motion+Manner	Motion+Motion	Non-Motion Verbs+Manner/Path/Deictic	no SVCs
L1 Chinese	14.21%	2.03%	2.54%	0.00%	1.02%	8.63%	0.00%	4.07%	66.00%
L2 Chinese	8.76%	0.40%	8.37%	4.78%	4.38%	0.00%	0.40%	1.60%	71.32%

Figure 4.10: SVCs used in L1 and L2 Chinese in the verbal task of Type A videos

As shown in Figure 4.10, in the descriptions of Type A videos, no significant difference in the use of SVCs was found between L1 and L2 Chinese speakers (L2 Chinese 28.68% vs. L1 Chinese 34%, $\chi 2$ (1, N = 136) = 0.471, p = 0.493). Taking a closer look at the SVCs used in L1 and L2 Chinese, it is found that there was no significant difference in the use of SVCs encoding Manner and Path between the two groups ($\chi 2$ (1, N = 50) = 0.720, p = 0.396). However, in contrast to L1 Chinese speakers, L2 Chinese speakers used significantly more SVCs encoding Manner + Path + Deictic

($\chi 2$ (1, N = 26) = 9.846, p = 0.002), Path + Path ($\chi 2$ (1, N = 13) = 6.231, p = 0.013), and Path + Deictic (L2 Chinese 4.78% vs. L1 Chinese 0%). L1 Chinese speakers, in turn, used more SVCs with Motion and Manner encodings (L2 Chinese 0% vs. L1 Chinese 8.63%). As mentioned previously, the non-mention verbs are 'carry' verbs, which can occur together with other manner or path verbs (example 57).



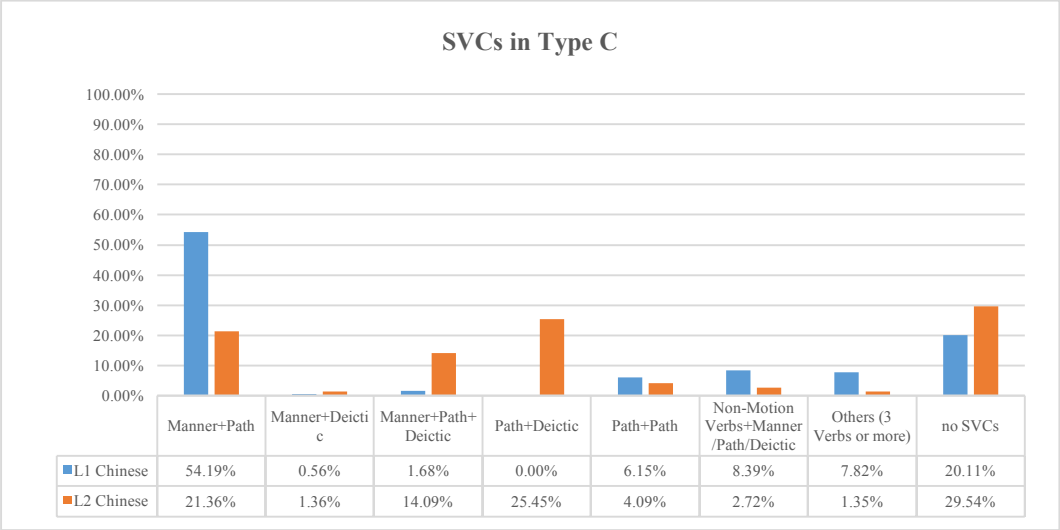
	Manner+Path	Manner+Deictic	Manner+Path+Deictic	Path+Deictic	Path+Path	Motion+Manner	Motion+Motion	Others (3 Verbs or more)	no SVCs
L1 Chinese	14.69%	3.39%	1.13%	0.00%	1.13%	28.25%	0.00%	2.24%	48.02%
L2 Chinese	10.31%	0.90%	9.42%	2.69%	3.59%	0.00%	0.90%	1.35%	70.85%

Figure 4.11: SVCs used in L1 and L2 Chinese in the verbal task of Type B videos

In the descriptions of Type B videos, as shown in Figure 4.11, statistically speaking, L1 Chinese speakers used significantly fewer SVCs than L2 Chinese speakers (L2 Chinese 51.98% vs. L1 Chinese 29.15%, $\chi 2$ (1, N = 155) = 4.032, p = 0.045). When examining each type of SVC used, it was found that there were no significant differences in the use of SVCs with Manner + Path ($\chi 2$ (1, N = 49) = 0.184, p = 0.668) and Path + Path ($\chi 2$ (1, N = 10) = 3.600, p = 0.058) between the two groups. However, L2 speakers used more SVCs with Manner + Path + Deictic ($\chi 2$ (1, N = 23) = 15.696, p < .001) and Path + Deictic (L2 Chinese 2.69% vs. L1 Chinese 0%) than L1 Chinese speakers. The use of SVCs with Motion + Manner by L1 Chinese speakers, however, was significantly higher than that of L2 Chinese speakers (L1 Chinese 28.25% vs. L2 Chinese 0%; see example 55).

In describing boundary-crossing events, as shown in Figure 4.12 below, L1 and L2 Chinese speakers tended to use SVCs (L1 Chinese 79.89% vs. L2 Chinese 70.46%); there was no significant difference between these two groups ($\chi 2$ (1, N = 297) = 0.569, p = 0.451). However, L1 Chinese speakers tended to use a combination of manner verbs and path verbs (L1 Chinese 54.19% vs. L2 Chinese 21.36%), and a significant difference was found between L1 and L2 Chinese speakers ($\chi 2$ (1, N = 144) = 17.361,

p < .001). Unlike L1 Chinese speakers, L2 Chinese speakers preferred using constructions that include Manner + Path + Deictic (L2 Chinese 14.09% vs. L1 Chinese 1.68%, $\chi 2$ (1, N = 34) = 23.059, p < .001) and Path + Deictic (L2 Chinese 25.45% vs. L1 Chinese 0%). There was no significant difference in the use of SVCs with Path + Path between the two groups of speakers ($\chi 2$ (1, N = 20) = 0.200, p = 0.655); however, a statistical significance was found in the use of SVCs with non-motion verbs + manner/path verbs ($\chi 2$ (1, N = 18) = 8.000, p = 0.005) and other SVCs ($\chi 2$ (1, N = 21) = 3.857, p = 0.050). L1 Chinese speakers used those SVCs more frequently than L2 Chinese speakers.



	Manner+Path	Manner+Deictic	Manner+Path+Deictic	Path+Deictic	Path+Path	Non-Motion Verbs+Manner/Path/Deictic	Others (3 Verbs or more)	no SVCs
L1 Chinese	54.19%	0.56%	1.68%	0.00%	6.15%	8.39%	7.82%	20.11%
L2 Chinese	21.36%	1.36%	14.09%	25.45%	4.09%	2.72%	1.35%	29.54%

Figure 4.12: SVCs used in L1 and L2 Chinese in the verbal task of Type C videos

Summary

When comparing the SVCs used to describe Type A, B, and C videos, it was found that their occurrence increased from Type A and B to C in L1 and L2 Chinese speakers. This might be related to the characteristics of the different video types. Type A showed a short trajectory and an obvious endpoint. In describing the situation in this type of videos, L1 and L2 Chinese speakers could choose a manner verb with a directional adjunct or with a locative adjunct (for adjuncts, see section 4.4.1.2), which already fulfills the communicative task. Type B showed a long trajectory with a possible endpoint, in which a Figure typically passes by a mountain, field, or meadow. The information about 'passing or crossing' a referential object is encoded in the path verb *guò* in Chinese. The path verb *guò* and manner verbs like *kāi* 'drive' or *zǒu* 'walk' or deictic verbs can easily form an SVC to express such a situation. In addition, L1

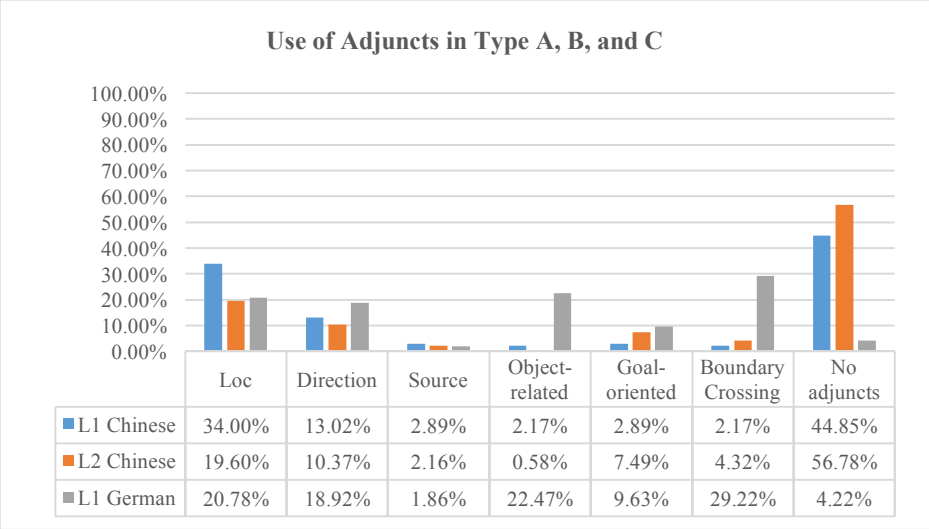
Chinese speakers tended to use the serial verb *xíng-shǐ* 'move-drive' alone or with locative adjuncts to express the meaning of "A Figure moving or a Figure moving in a certain manner at a certain place", while L2 Chinese speakers did this to a lesser degree. The use of SVCs reached its peak in the verbal task for Type C videos because Type C videos showed boundary-crossing events. In order to express information about boundary-crossing, the Chinese path verb *jìn* 'enter' is typically applied. The path verb *jìn* 'enter' occurs with manner verbs or deictic verbs to express a boundary-crossing situation. So the number of SVCs rose dramatically for Type C videos in both groups. However, L1 and L2 Chinese speakers differed remarkably in the use of specific types of SVCs. It seems that L2 Chinese speakers have a preference for SVCs with Manner + Path + Deictic and Path + Deictic. This preference was found consistently throughout the descriptions of Type A, B, and C videos. As already explained above, the preference for deictic verbs in L2 Chinese could be interpreted in terms of the phonological and morphosyntactic status of modern Chinese, as L2 Chinese speakers' tendency to take a holistic perspective, and as being the result of the directional meaning indicated by deictic verbs. In contrast, L1 Chinese speakers used fewer deictic verbs (see Table 4.4) and tended to use SVCs with Manner + Path and Motion + Manner.

4.4.1.2 Adjuncts

Apart from the spatial information encoded in verbs, adjuncts also encode spatial relations between the Figure and the Ground. In this section, we will first compare the spatial concepts encoded in adjuncts in L1 Chinese, L1 German-L2 Chinese, and L1 German. Furthermore, according to the presence or absence of postpositions in Chinese, the use of adjuncts in L1 and L2 Chinese will be analyzed once more. The reason for doing so is based on two observations. First, the postpositions *lǐ/lǐbian*/*lǐmian*/*lǐtou* 'inside' play a crucial role in determining a boundary-crossing situation, and second, L2 Chinese speakers in our study omitted postpositions where they are necessary. So L2 Chinese speakers' inappropriate productions will also be presented. The percentages were obtained by using the following mathematical formula: number of occurrences of a specific adjunct divided by total number of utterances. For example, the percentages of locative adjuncts in the descriptions of Type A videos were obtained using the following formula: number of locative adjuncts used to describe Type A videos divided by total number of utterances for Type A videos. Given that one statement may include two or more adjuncts, the

total percentages of all adjuncts may slightly surpass 100%.

As shown in Figure 4.13, the overall picture regarding the use of adjuncts in the three groups shows that L1 German speakers used significantly more adjuncts than speakers of L1 Chinese (L1 German 95.78% vs. L1 Chinese 55.15%, $\chi 2$ (1, N = 925) = 92.810, p < .001) and L2 Chinese (L1 German 95.78% vs. L2 Chinese 43.22%, $\chi 2$ (1, N = 918) = 98.039, p < .001). These results are in conformance with the typological features of the S-language German, that is, that Path information is typically encoded in adjuncts, but not in path verbs (see section 4.4.1.1 for the use of path verbs in the three language groups). Between L1 and L2 Chinese speakers, the use of adjuncts shows no significant difference ($\chi 2$ (1, N = 625) = 0.078, p = 0.779).



	Loc	Direction	Source	Object-related	Goal-oriented	Boundary Crossing	No adjuncts
L1 Chinese	34.00%	13.02%	2.89%	2.17%	2.89%	2.17%	44.85%
L2 Chinese	19.60%	10.37%	2.16%	0.58%	7.49%	4.32%	56.78%
L1 German	20.78%	18.92%	1.86%	22.47%	9.63%	29.22%	4.22%

Figure 4.13: Overview of adjuncts used in L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos

The following examples show no adjuncts used in Chinese and German (i.e., when a single manner verb or serial verbs are used).

No adjuncts used

(58a) 有一个女生在走路。(L1 Ger-L2 Chn P16, A01)

yǒu yī-wèi nǚshēng zài zǒu-lù.

exist one-CL girl ZAI walk-road

There is a girl walking on the road[58].

[58] *zǒulù* is translated as 'walking on the road' in English. In this translation the adjunct 'on the road' is used. But in Chinese *zǒulù* is a verb compound (Packard, 2000, p. 40), so it's no wonder no adjunct is used in example

(58b) 有人走过去，她右转。(L1 Ger-L2 Chn P23, A01)

yǒu rén zǒu-guò-qù, tā yòu zhuǎn.
exist person walk-pass-go, she right turn
Someone walks over (away from the speaker) and she turns right.

(58c) Eine Frau geht spazieren. (L1 Ger P05, B03)
one woman go walk
A woman goes for a walk.

An in-depth look at the use of adjuncts shows that L1 Chinese speakers used significantly more locative adjuncts than L2 Chinese speakers (L1 Chinese 34.00% vs. L2 Chinese 19.60%, $\chi 2$ (1, N = 324) = 8.346, p = 0.004) and L1 German speakers (L1 Chinese 34.00% vs. L1 German 20.78%, $\chi 2$ (1, N = 311) = 13.585, p < .001). This is shown in the following examples (the relevant adjuncts are underlined):

Locative adjunct

(59a) 一位女士在一个广场上走着。(L1 Chn P04, A07)

yī-wèi nǚshì <u>zài yī-ge guǎngchǎng-shàng</u> zǒu-zhe.
one-CL lady at one-CL square-top walk-ZHE
A lady is walking in a square.

(59b) Ein Mann läuft <u>im Park</u>. (L1 Ger P10, B03)
one man walk in the park
A man walks in the park.

In contrast, the use of directional adjuncts was significantly higher in L1 German than in L1 Chinese (L1 German 18.92% vs. L1 Chinese 13.02%, $\chi 2$ (1, N = 184) = 8.696, p = 0.003) and L2 Chinese (L1 German 18.92% vs. L2 Chinese 10.37%, $\chi 2$ (1, N = 189) = 6.481, p = 0.011). There was no significant difference in the use of directional adjuncts between L1 and L2 Chinese speakers ($\chi 2$ (1, N = 144) = .000, p =

(58a) in Chinese.

1.000).

Directional adjunct

(60a) 公共汽车往公共汽车站开。(L1 Ger-L2 Chn P09, A09)

gōnggòngqìchē wǎng gōnggòngqìchēzhàn kāi.
bus towards bus station drive
A bus drives towards the bus station.

(60b) Ein Bus fährt eine Straße entlang in Richtung einer Haltestelle. (L1 Ger P20, A09)
one bus drive one (ACC) road along in the direction of one (GEN) stop
A bus drives along a road towards a stop.

In addition, no significant difference was found regarding the use of adjuncts encoding source information in the three language groups ($\chi 2$ (2, N = 42) = 1.000, p = 0.607), whereas the use of object-related adjuncts was significantly higher in L1 German than in L1 Chinese (L1 German 22.47% vs. L1 Chinese 2.17%, $\chi 2$ (1, N = 145) = 100.972, p < .001) and L2 Chinese (L1 German 22.47% vs. L2 Chinese 0.58%, $\chi 2$ (1, N = 137) = 121.467, p < .001). Although the total number of occurrences of object-related adjuncts was low in L1 and L2 Chinese, its use was found significantly more often in L1 Chinese than in L2 Chinese (L1 Chinese 2.17% vs. L2 Chinese 0.58%, $\chi 2$ (1, N = 16) = 4.000, p = 0.046).

Adjuncts encoding source information

(61a) 可能一个人要从外面进入自己的房间。(L1 Ger-L2 Chn P14, A03)

kěnéng yī-ge rén yào cóng wàimian jìn-rù zìjǐ-de fángjiān.
perhaps one-CL person YAO from outside enter-enter self-NOM room
Perhaps a person wanted to enter his/her own room from outside.

(61b) Ein Mann steigt auf einer Leiter hoch vom Erdgeschoss in den ersten Stock. (L1 Ger P20, A03)
one man climb on one (DAT) ladder up from the (DAT) ground floor to the

(ACC) first floor

A man climbs up on a ladder from the ground floor to the first floor.

Object-related adjuncts

(62a) 一辆车沿着路行驶。(L1 Chn P16, B02)

yī-liàng chē yán zhe lù xíng-shǐ.
one-CL car along ZHE road move-drive
A car is driving along the road.

(62b) Ein Auto fährt auf einer Straße entlang einer Kurve. (L1 Ger P01, B02)
one car drive on one (DAT) street along one (DAT) bend
A car drives on a street along a bend.

Concerning goal-oriented adjuncts, it was found that L1 German-L2 Chinese speakers and L1 German speakers used them significantly more often than L1 Chinese speakers (L2 Chinese 7.49% vs. L1 Chinese 2.89%, $\chi 2$ (1, N = 68) = 19.059, p < .001; L1 German 9.63% vs. L1 Chinese 2.89%, $\chi 2$ (1, N = 73) = 23.027, p < .001), while there was no significant difference between speakers of L2 Chinese and L1 German ($\chi 2$ (1, N = 109) = 0.229, p = 0.632).

Goal-oriented adjuncts

(63a) 一辆车开到小城市。(L1 Ger-L2 Chn P12, B05)

yī-liàng chē kāi-dào[59] xiǎo chéngshì.
one-CL car drive-arrive small city
A car drove to a small city.

(63b) Hier fährt ein Auto die Straße entlang zu einem Dorf. (L1 Ger P06, B02)
here drive one car the (ACC) street along to one (DAT) village
Here a car drives along the street to a village.

Furthermore, the use of boundary-crossing adjuncts was significantly higher in

[59] *dào* 'arrive' is a coverb. That is, it is a verb as well as a preposition. So *dào* 'arrive' is counted twice.

L1 German than in L1 Chinese (L1 German 29.22% vs. L1 Chinese 2.17%, $\chi 2$ (1, N = 185) = 140.114, p < .001) and L2 Chinese (L1 German 29.22% vs. L2 Chinese 4.32%, $\chi 2$ (1, N = 203) = 100.734, p < .001). This is because boundary-crossing information is typically encoded in adjuncts in German, as opposed to the same information being encoded in path verbs in Chinese. The comparison of the use of boundary-crossing adjuncts between L1 and L2 Chinese speakers shows that L2 Chinese speakers used them significantly more often than L1 Chinese speakers (L2 Chinese 4.32% vs. L1 Chinese 2.17%, $\chi 2$ (1, N = 42) = 7.714, p = 0.005).

Boundary-crossing adjuncts

(64a) 有车子开到村子里面了。(L1 Ger-L2 Chn P07, B01)

yǒu chēzi kāi-dào cūnzi-lǐmian le.

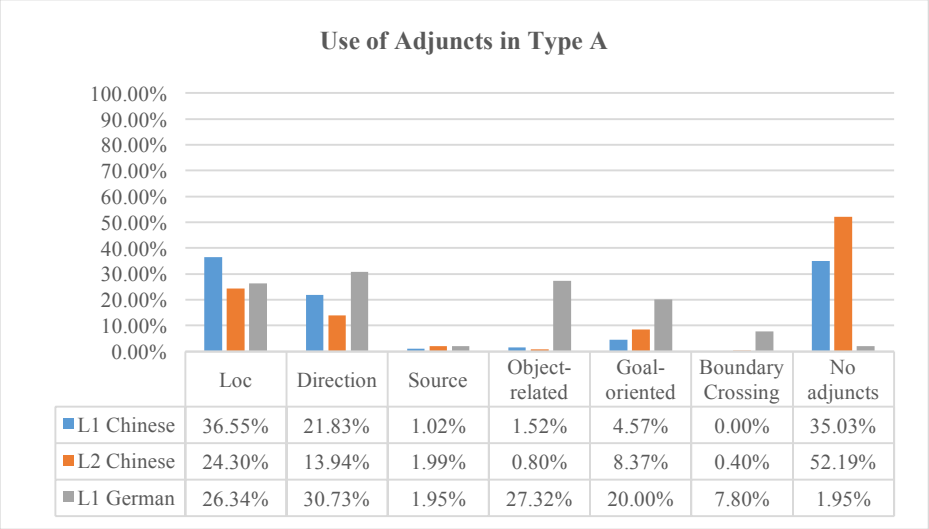
exist car drive-arrive village-inside LE

There was a car that drove into a village.

(64b) Ein Auto fährt in ein Dorf. (L1 Ger P02, B02)

one car drive in one (ACC) village

A car drives into a village.

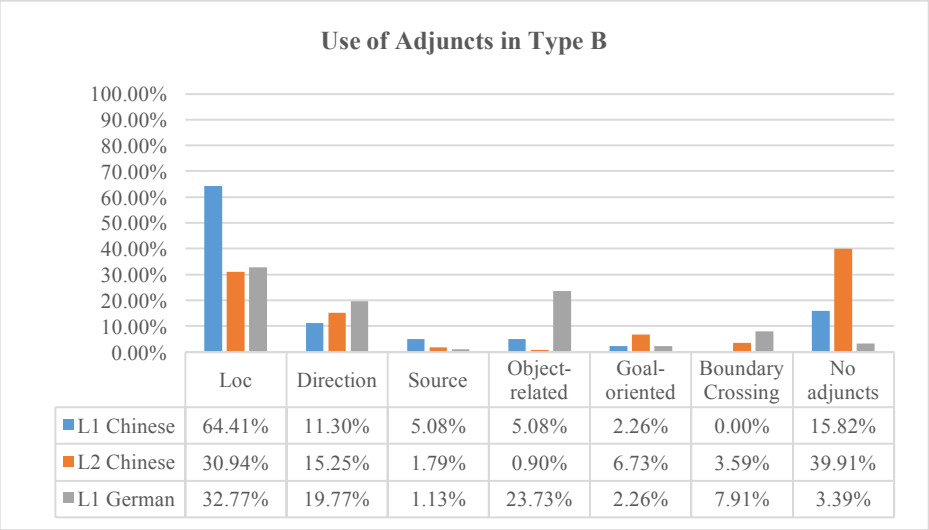


	Loc	Direction	Source	Object-related	Goal-oriented	Boundary Crossing	No adjuncts
L1 Chinese	36.55%	21.83%	1.02%	1.52%	4.57%	0.00%	35.03%
L2 Chinese	24.30%	13.94%	1.99%	0.80%	8.37%	0.40%	52.19%
L1 German	26.34%	30.73%	1.95%	27.32%	20.00%	7.80%	1.95%

Figure 4.14: Percentages of adjuncts used in L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type A videos

In the verbal task for Type A videos, as shown in Figure 4.14, the use of adjuncts in L1 German was significantly higher than in L1 Chinese ($\chi 2$ (1, N = 363) = 30.372, p

< .001) and in L2 Chinese ($\chi 2$ (1, N = 359) = 33.095, p < .001). However, no significant effect was found in the use of locative adjuncts among the three language groups ($\chi 2$ (1, N = 187) = 2.642, p = 0.267), but the use of directional adjuncts was found to be higher in L1 German than in L2 Chinese ($\chi 2$ (1, N = 98) = 8.000, p = 0.005). In contrast, no significant difference was found between L1 German and L1 Chinese speakers ($\chi 2$ (1, N = 106) = 3.774, p = 0.052) and between L1 Chinese and L2 Chinese speakers ($\chi 2$ (1, N = 78) = 0.821, p = 0.365). There was a higher frequency of adjuncts encoding object-related information in L1 German than in L1 and L2 Chinese, respectively (L1 German 27.32% vs. L1 Chinese 1.52%, $\chi 2$ (1, N = 59) = 47.610, p < .001 and L1 German 27.32% vs. L2 Chinese 0.80%, $\chi 2$ (1, N = 58) = 50.276, p < .001). Regarding the use of goal-oriented adjuncts, it was found that L1 German speakers used them significantly more often than L1 Chinese speakers (L1 German 20.00% vs. L1 Chinese 4.57%, $\chi 2$ (1, N = 50) = 20.480, p < .001) and L2 Chinese speakers (L1 German 20.00% vs. L2 Chinese 8.37%, $\chi 2$ (1, N = 62) = 6.452, p = 0.011). In addition, L2 Chinese speakers also tended to use them more often than L1 Chinese speakers did (L2 Chinese 8.37% vs. L1 Chinese 4.57%, $\chi 2$ (1, N = 30) = 4.800, p = 0.028). The use of boundary-crossing adjuncts was found to be significantly higher in L1 German than in L2 Chinese (L1 German 7.80% vs. L2 Chinese 0.40%, $\chi 2$ (1, N = 17) = 13.235, p < .001), whereas no such data was found in L1 Chinese. The occurrence of boundary-crossing adjuncts in L1 German in the descriptions of Type A videos is mainly related to two video clips, one showing a man climbing a ladder to the first floor and the other one showing a man walking upstairs towards the entrance of a building. Both of these video clips ended before showing the man arriving at the place at goal. Nevertheless, L1 German speakers tended to encode boundary-crossing information in adjuncts. Section 4.4.1.1 showed that L2 Chinese speakers used significantly more boundary-crossing path verbs to describe the situation in the latter video clip (i.e., a man walking upstairs towards a building's entrance) than L1 Chinese speakers. L2 speakers' inclination for encoding boundary-crossing information in this video clip is in line with that of L1 German speakers, that is, both groups tended to focus on the endpoint, since in a boundary-crossing situation, the place at goal is so highly profiled that one cannot neglect it. The tendency to encode boundary-crossing information in L1 German and L1 German-L2 Chinese reflects that L1's conceptualization pattern has an impact on L2.

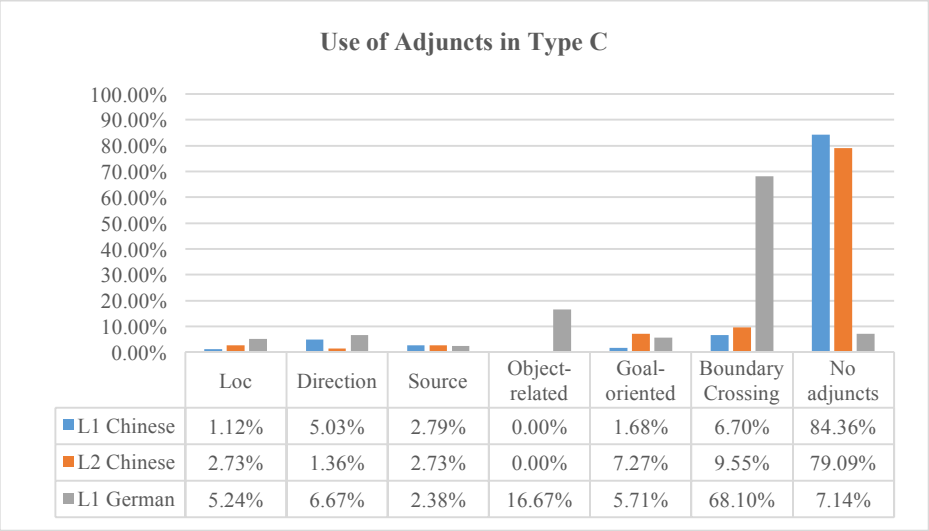


	Loc	Direction	Source	Object-related	Goal-oriented	Boundary Crossing	No adjuncts
L1 Chinese	64.41%	11.30%	5.08%	5.08%	2.26%	0.00%	15.82%
L2 Chinese	30.94%	15.25%	1.79%	0.90%	6.73%	3.59%	39.91%
L1 German	32.77%	19.77%	1.13%	23.73%	2.26%	7.91%	3.39%

Figure 4.15: Percentages of adjuncts used in L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type B videos

In the verbal task for Type B videos, as shown in Figure 4.15, there was no significant difference in the number of adjuncts used among the three language groups ($\chi 2$ (2, N = 443) = 2.497, p = 0.287). However, L1 and L2 speakers differed in their preferences for encoding different spatial concepts. L1 Chinese speakers tended to encode more locative information than L2 Chinese speakers ($\chi 2$ (1, N = 183) = 11.066, p = 0.001) and L1 German speakers ($\chi 2$ (1, N = 172) = 18.233, p < .001), while there was no significant difference in the use of locative adjuncts between L2 Chinese speakers and L1 German speakers ($\chi 2$ (1, N = 127) = 0.953, p = 0.329). Compared to the locative adjuncts used to describe Type A videos, it was found that L1 Chinese speakers used them significantly more often to describe Type B videos than to describe Type A videos ($\chi 2$ (1, N = 186) = 9.484, p = 0.002), while there was no significant difference in the use of locative adjuncts by speakers of L2 Chinese between Type A and Type B ($\chi 2$ (1, N = 130) = 0.492, p = 0.483) and by speakers of L1 German between Type A and Type B ($\chi 2$ (1, N = 112) = 0.143, p = 0.705). L1 Chinese speakers' tendency to express locative information is related to the characteristics shown in the Type B videos. Type B video clips showed a long trajectory with a possible endpoint. L1 Chinese speakers preferred describing the situation as a Figure moving in a certain manner at a certain place, while it seems that L2 Chinese speakers and L1 German speakers tended not to do so. L1 German speakers used more directional adjuncts than L1 Chinese speakers (L1 German 19.77% vs. L1 Chinese 11.30%, $\chi 2$ (1, N = 55) = 4.091, p = 0.043), while no significant difference was found between L1 and Chinese

speakers (L1 Chinese 11.30% vs. L2 Chinese 15.25%, $\chi 2$ (1, N = 54) = 3.630, p = 0.057) and between L1 German and L2 Chinese speakers (L1 German 19.77% vs. L2 Chinese 15.25%, $\chi 2$ (1, N = 69) = 0.014, p = 0.904). Furthermore, L1 German speakers tended to use more object-related adjuncts than speakers of L1 Chinese (L1 German 23.73% vs. L1 Chinese 5.08%, $\chi 2$ (1, N = 51) = 21.353, p < .001) and L2 Chinese (L1 German 23.73% vs. L2 Chinese 0.90%, $\chi 2$ (1, N = 44) = 36.364, p < .001). In addition, the use of object-related adjuncts was also significantly higher in L1 Chinese than in L2 Chinese ($\chi 2$ (1, N = 11) = 4.455, p = 0.035). In contrast, L2 Chinese speakers tended to use more goal-oriented adjuncts than L1 Chinese ($\chi 2$ (1, N = 19) = 6.368, p = 0.012) and L1 German speakers ($\chi 2$ (1, N = 19) = 6.368, p = 0.012). Concerning the use of boundary-crossing adjuncts, no significant difference was found between speakers of L2 Chinese and L1 German ($\chi 2$ (1, N = 22) = 1.636, p = 0.201), while no such data was found in L1 Chinese. Furthermore, no significant effect was found in the use of adjuncts encoding source information ($\chi 2$ (2, N = 15) = 5.200, p = 0.074) among the three groups.



	Loc	Direction	Source	Object-related	Goal-oriented	Boundary Crossing	No adjuncts
L1 Chinese	1.12%	5.03%	2.79%	0.00%	1.68%	6.70%	84.36%
L2 Chinese	2.73%	1.36%	2.73%	0.00%	7.27%	9.55%	79.09%
L1 German	5.24%	6.67%	2.38%	16.67%	5.71%	68.10%	7.14%

Figure 4.16: Percentages of adjuncts used in L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type C videos

In the verbal task for Type C videos, shown in Figure 4.16, the number of adjuncts used in L1 German was significantly higher than that in L1 Chinese ($\chi 2$ (1, N = 251) = 142.315, p = .021) and L2 Chinese ($\chi 2$ (1, N = 272) = 103.765, p < .001). In comparison with L1 Chinese speakers, L2 Chinese speakers used more adjuncts ($\chi 2$ (1, N = 83) = 5.313, p = 0.021). The difference in the number of adjuncts used in German and Chinese is attributable to the different typological features of these two languages.

In German, Path information, and more specifically in this case, boundary-crossing information in Type C videos, was encoded in adjuncts, while the same information was encoded in path verbs in Chinese. So the number of adjuncts used in L1 German was higher than that in L1 and L2 Chinese.

Moreover, compared to the use of locative adjuncts used to describe Type B videos, their use decreased sharply for Type C videos in all three groups (Type B: 64.41% vs. Type C: 1.12% in L1 Chinese, Type B: 30.94% vs. Type C: 2.73% in L2 Chinese and Type B: 32.77% vs. Type C: 5.24% in L1 German). The reason for this is that because the motion of boundary-crossing was more salient than, for example, specifying the location where a movement takes place in Type C videos. Therefore, the use of boundary-crossing adjuncts was significantly higher in L1 German than in L1 Chinese (L1 German 68.10% vs. L1 Chinese 6.70%, $\chi2$ (1, N = 155) = 110.716, p < .001) and L2 Chinese (L1 German 68.10% vs. L2 Chinese 9.55%, $\chi2$ (1, N = 164) = 90.756, p < .001), while L1 and L2 Chinese speakers encoded the same information in path verbs (see section 4.4.1.1). Also, no significant difference was found between L1 and L2 Chinese speakers regarding the use of boundary-crossing adjuncts ($\chi2$ (1, N = 33) = 2.455, p = 0.117). Besides the use of boundary crossing adjuncts, the use of directional adjuncts showed no significant difference between L1 and L2 Chinese speakers and between L1 Chinese and L1 German speakers, respectively (L1 Chinese vs. L2 Chinese: $\chi2$ (1, N = 12) = 3.000, p = 0.083, L1 Chinese vs. L1 German: $\chi2$ (1, N = 23) = 1.087, p = 0.297), while L1 German speakers used them more often than L2 Chinese speakers ($\chi2$ (1, N = 17) = 7.118, p = 0.008). Also, no significance was found in the use of adjuncts encoding source information among the three groups ($\chi2$ (2, N = 16) = 0.125, p = 0.939). L1 German speakers used more object-related adjuncts than L1 and L2 Chinese speakers, while the latter two groups did not express object-related information in adjuncts. Regarding the use of goal-oriented adjuncts, speakers of L2 Chinese and L1 German used them significantly more often than L1 Chinese speakers (L2 Chinese 7.27% vs. L1 Chinese 1.68%, $\chi2$ (1, N = 19) = 8.895, p = 0.003 and L1 German 5.71% vs. L1 Chinese 1.68%, $\chi2$ (1, N = 15) = 5.400, p = 0.020), while the use of goal-oriented adjuncts in L2 Chinese and L1 German showed no significant difference ($\chi2$ (1, N = 28) = 0.571, p = 0.450). Given that in German spatial information is also encoded in verb particles, it is necessary to present this use in different video types as well, as shown in Figure 4.17.

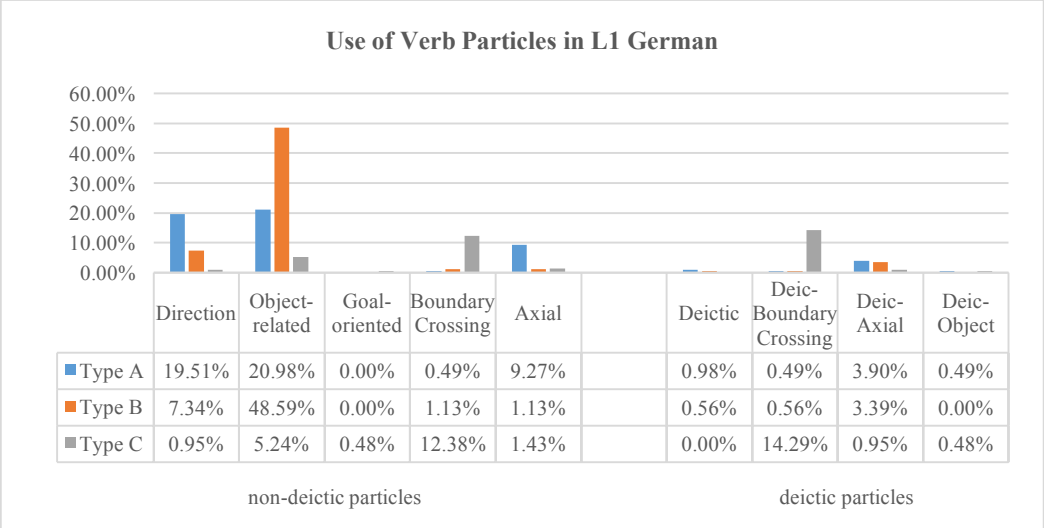
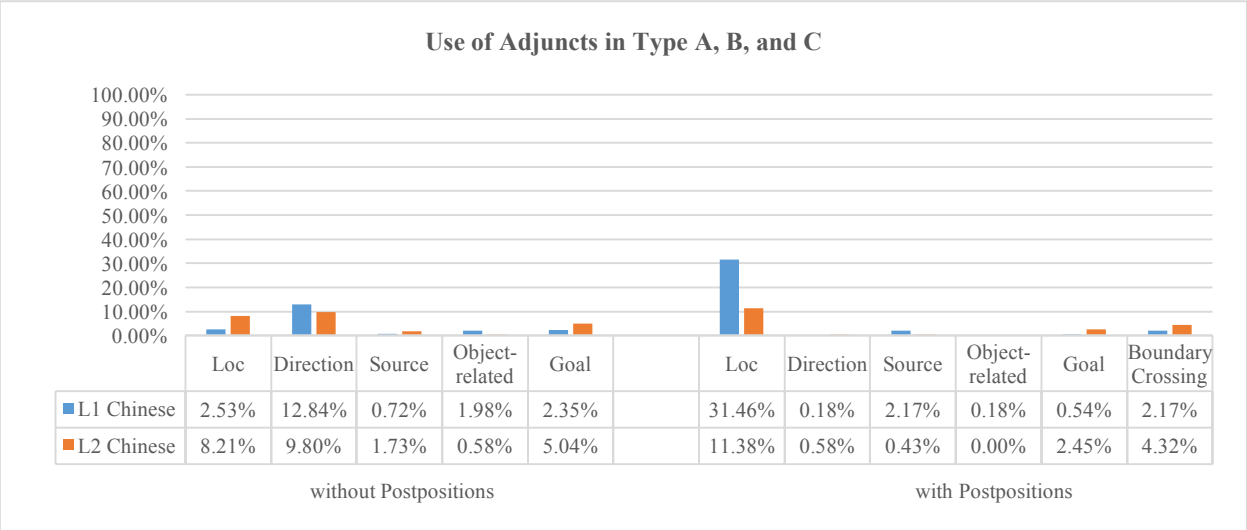


Figure 4.17: Use of particles in L1 German in the verbal task of Type A, B, and C videos

In general, particles used in German can be differentiated into non-deictic and deictic particles. In the verbal task for Type A videos, the use of particles mainly centered on encoding object-related information (e.g., *entlanglaufen* 'walk along' with a rate of 20.98%), direction (e.g., *auf ein Auto zu* 'towards a car' with a rate of 19.51%), axial information (e.g., *hochklettern* 'climb up' with a rate of 9.27%), and deictic-axial information (e.g., *hinaufklettern* 'climb up away from the speaker' with a rate of 3.90%), which is in conformance with the features of the video types in this category: a short trajectory with an obvious endpoint. In contrast, the rate of object-related particles used increased to 48.59%, while the rate of directional particles decreased to 7.34% in the verbal task for Type B videos. The choice of different particles reflects the characteristics of Type B videos: a long trajectory with a possible endpoint. L1 German speakers tended to encode the spatial concepts derived from the features of the Ground object in Type B videos. In addition, the use of deictic-axial particles (e.g., *hinaufgehen* 'walk up away from the speaker') in the verbal task for Type B videos occurred at a rate of 3.39%. In describing Type C videos, L1 German speakers chose either non-deictic boundary-crossing particles (e.g., *einfahren* 'drive into') with a rate of 12.38% or deictic boundary-crossing particles (e.g., *hineinfahren* 'drive into away from the speaker') with a rate of 14.29% to describe a boundary-crossing situation.

Given that the Chinese postpositions *lǐ/lǐbian/lǐmian/lǐtou* 'inside' play an essential role in determining boundary-crossing information and inappropriate omission of postpositions occurred in L2 Chinese, the following examples demonstrate the use of adjuncts in L1 and L2 Chinese with respect to the presence or absence of

postpositions (see Figure 4.18, Figure 4.19, Figure 4.20, and Figure 4.21). Furthermore, inappropriate expressions found in L2 Chinese will also be presented.



	Loc	Direction	Source	Object-related	Goal	Loc	Direction	Source	Object-related	Goal	Boundary Crossing
L1 Chinese	2.53%	12.84%	0.72%	1.98%	2.35%	31.46%	0.18%	2.17%	0.18%	0.54%	2.17%
L2 Chinese	8.21%	9.80%	1.73%	0.58%	5.04%	11.38%	0.58%	0.43%	0.00%	2.45%	4.32%

Figure 4.18: Total number of adjuncts used in L1 and L2 Chinese in the verbal task of Type A, B, and C videos

Without considering the presence or absence of postpositions in adjuncts, both L1 and L2 Chinese speakers tended to encode the concepts of location, direction, and goal (see Figure 4.18). However, if we take a closer look at the encoded concepts, differences between the two groups emerge. L2 speakers used significantly more locative adjuncts without postpositions than L1 speakers (L2 Chinese 8.21% vs. L1 Chinese 2.53%, $\chi 2$ (1, N = 71) = 26.042, $p < .001$), whereas L1 speakers used significantly more locative adjuncts with postpositions than L2 speakers (L2 Chinese 11.38% vs. L1 Chinese 31.46%, $\chi 2$ (1, N = 253) = 35.672, $p < .001$). This is shown in the following examples:

(65a) Locative adjunct with postposition (example 59a is repeated)

一位女士在一个广场上走着。(L1 Chn P04, A07)

yī-wèi nǚshì zài yī-ge guǎngchǎng-shàng zǒu-zhe.

one-CL lady at one-CL square-top walk-ZHE

A lady is walking in a square.

(65b) Locative adjunct without postposition

一个男人在公园散步。(L1 Ger-L2 Chn P24, B03)

yī-ge nán-rén zài gōngyuán sànbù.
one-CL male-person at park take a walk
A man is taking a walk in the park.

(65c) Locative adjunct without postposition

*一个人在田骑自行车。(L1 Ger-L2 Chn P19, B09)

*yī-ge rén zài tián qí zìxíngchē.
one-CL person at field ride bicycle
To be correct, this sentence should be constructed as follows:

(65d) Locative adjunct with postposition

一个人在田边骑自行车。

yī-ge rén zài tián-biān qí zìxíngchē.
one-CL person at field-side ride bicycle
A person is riding a bicycle beside a field.

Depending on the features of the Ground object and the actual context, a circumposition without postposition may occasionally result in a grammatically incorrect sentence. Examples 65b and 65c do not include postpositions in the sentences and are therefore inappropriate. As already explained in section 3.1.3, when an NP in a circumposition refers to a familiar place such as a house, a building, etc. and the adpositional phrase precedes the verb, then the postposition can be omitted. Note that these familiar places are usually three-dimensional objects. This is the case in example 65b. Although *tián* 'field' in example 65c is also a familiar place and the prepositional phrase precedes the verb, the postposition cannot be omitted. This can be ascribed to two facts. First, *tián* 'field' is not a three-dimensional object. Second, this sentence describes a video that shows a woman riding a bicycle between fields, possibly towards a house. The postposition *biān* 'side' should be further specified as in example 65d.

In cases without postpositions, L2 speakers used significantly more adjuncts encoding source ($\chi 2$ (1, N = 16) = 4.000, p = 0.046) and goal-oriented information ($\chi 2$ (1, N = 48) = 10.083, p = 0.001) than L1 speakers, despite the overall low number of occurrences. In contrast, no significant differences were found in the use of adjuncts

encoding direction ($\chi 2$ (1, N = 139) = 0.065, p = 0.799) and object-related information ($\chi 2$ (1, N = 15) = 3.267, p = 0.071). This is illustrated in the following examples:

(66) Adjuncts without postposition: encoding source information

可能一个人要从外面进入自己的房间。(L1 Ger-L2 Chn P14, A03) (example 61a is repeated)

kěnéng yī-ge rén yào cóng wàimian jìn-rù zìjǐ-de fángjiān.
perhaps one-CL person YAO from outside enter-enter self-NOM room
Perhaps a person wanted to enter his/her own room from the outside.

(67) Adjuncts without postposition: encoding goal-oriented information

一辆车开到小城市。(L1 Ger-L2 Chn P12, B05) (example 63a is repeated)

yī-liàng chē kāi-dào[60] xiǎo chéngshì.
one-CL car drive-arrive small city
A car drove to a small city.

(68) Adjuncts without postposition: encoding direction

公共汽车往公共汽车站开。(L1 Ger-L2 Chn P09, A09) (example 60a is repeated)

gōnggòngqìchē wǎng gōnggòngqìchēzhàn kāi.
bus towards bus station drive
A bus drives towards the bus station.

(69) Adjuncts without postposition: encoding object-related information

一辆车沿着路行驶。(L1 Chn P16, B02) (example 62a is repeated)

yī-liàng chē yán zhe lù xíng-shǐ.
one-CL car along ZHE road move-drive
A car is driving along the road.

When the use of postpositions is taken into account, it is found that the use of

[60] *dào* 'arrive' is a coverb. That is, it is a verb as well as a preposition. So *dào* 'arrive' is counted twice.

adjuncts with goal-oriented meaning was significantly higher in L2 Chinese than in L1 Chinese ($\chi 2$ (1, N = 20) = 9.800, p = 0.002) (example 67), despite the low number of occurrences overall. However, L1 speakers used more adjuncts with source information (example 66) than L2 speakers ($\chi 2$ (1, N = 15) = 5.400, p = 0.020). The frequency of the directional adjuncts used was rather low in L1 and L2 Chinese, at a rate of 0.18% and 0.58%, respectively (example 68). In addition, the use of boundary-crossing adjuncts was significantly higher in L2 Chinese than in L1 Chinese ($\chi 2$ (1, N = 42) = 7.714, p = 0.005). This is exemplified in the following:

(70) Adjuncts with postposition: encoding goal-oriented information

公共汽车开到房子那边。(L1 Ger-L2 Chn P12, B06)

gōnggòngqìchē kāi-dào fángzi-nàbian.
bus drive-arrive house-that side
A bus drives to the house over there.

(71) Adjuncts with postposition: encoding boundary-crossing information

有车子开到村子里面了。(L1 Ger-L2 Chn P07, B01) (example 64a is repeated)

yǒu chēzi kāi-dào cūnzi-lǐmian le.
exist car drive-arrive village-inside LE
There was a car that drove into the village.

(72) Adjuncts with postposition: encoding source information

一个人从路边的草坪上走到了路上。(L1 Chn P16, B03)

yī-ge rén cóng lùbian-de cǎopíng-shàng zǒu-dào le lù-shàng.
one-CL person from roadside-NOM grass-top walk-arrive LE road-top
A person walked from the meadow along the roadside to the road.

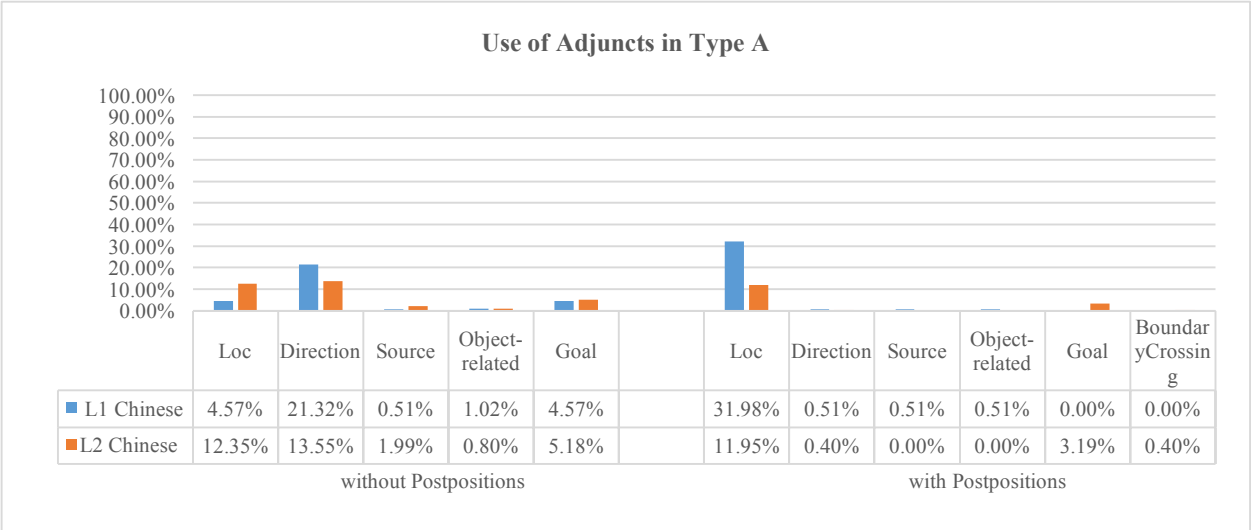
(73) Adjuncts with postposition: encoding boundary-crossing information

有一辆车开往一个院子里。(L1 Ger-L2 Chn P06, C03)

yǒu yī-liàng chē kāi wǎng yī-ge yuànzi -lǐ.
exist one-CL car drive towards one-CL yard-inside

There is a car that drives into a yard.

Examples 71 and 73 show the important role that the postpositions *lǐmian* and *lǐ* 'inside' play in changing the meaning, from a goal-oriented and directional meaning to a boundary-crossing meaning.



	Loc	Direction	Source	Object-related	Goal	Loc	Direction	Source	Object-related	Goal	BoundaryCrossing
L1 Chinese	4.57%	21.32%	0.51%	1.02%	4.57%	31.98%	0.51%	0.51%	0.51%	0.00%	0.00%
L2 Chinese	12.35%	13.55%	1.99%	0.80%	5.18%	11.95%	0.40%	0.00%	0.00%	3.19%	0.40%

Figure 4.19: Use of adjuncts in L1 and L2 Chinese in the verbal task of Type A videos

In the verbal task for Type A videos, as shown in Figure 4.19, L1 and L2 speakers tended to use adjuncts with a locative and directional meaning. This is in conformance with the characteristics of the video clips in this category. Speakers could select whether to encode the information on direction or the information on location. When speakers chose to encode directional information, they usually mentioned the endpoint; however, when they solely chose to express locative information, they usually did not mention the endpoint. As presented above, when taking all adjuncts with locative information into account – regardless of the presence or absence of postpositions – no significant difference was found between L1 and L2 Chinese speakers ($\chi 2$ (1, N = 133) = 0.910, p = 0.340). However, L1 speakers used significantly more locative adjuncts with postpositions than without postpositions ($\chi 2$ (1, N = 72) = 40.500, $p < .001$), whereas L2 speakers used locative adjuncts with and without postpositions on an equal basis ($\chi 2$ (1, N = 61) = 0.016, p = 0.898). That is, when no postpositions were used, L2 speakers used significantly more locative adjuncts than L1 speakers (L2 Chinese 12.35% vs. L1 Chinese 4.57%, $\chi 2$ (1, N = 40) = 12.100, p = 0.001). In turn, when postpositions were used, the use of locative adjuncts was

significantly lower in L2 Chinese than in L1 Chinese (L2 Chinese 11.95% vs. L1 Chinese 31.98%, $\chi 2$ (1, N = 93) = 11.710, p = 0.001). When the postpositions used in locative adjuncts in L1 Chinese were examined further, it was found that although the postpositions could be omitted on some occasions, L1 speakers tended not to do so (example 74a), while L2 speakers omitted the postpositions either appropriately or inappropriately (examples 74b and 74c)

(74a) Postposition can be omitted

一个人在停车场里走。(L1 Chn P15, A05)

yī-ge rén zài tíngchēchǎng-lǐ zǒu.
one-CL person at parking lot-inside walk
A person is walking in the parking lot.

(74b) Appropriate ellipsis of postposition

一个男的在停车场走着。(L1 Ger-L2 Chn P15, A05)

yī-ge nán-de zài tíngchēchǎng zǒu-zhe.
one-CL male-NOM at parking lot walk-ZHE
A man is walking in the parking lot.

(74c) Inappropriate ellipsis of postposition

*有个女孩子她在行人道就走路。(L1 Ger-L2 Chn P03, A01)

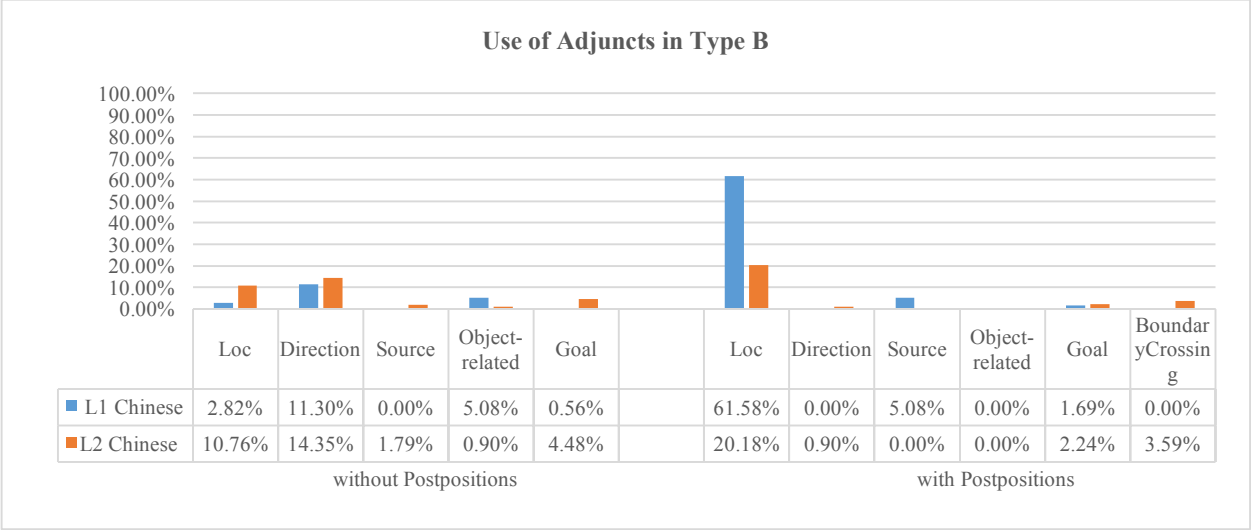
*yǒu ge nǚháizi tā zài xíngréndào jiù zǒu-lù.
exist CL girl she at sidewalk just walk-road
There is a girl who just walks on the sidewalk.

In example 74a, the postposition *lǐ* 'inside' could be omitted because *tíngchēchǎng* 'parking lot' is a daily familiar place and in this video clip[61] it was a three-dimensional object. However, L1 speakers tended to keep it (example 74a), while L2 speakers tended to omit it (example 74b). L2 speakers might overgeneralize the ellipsis to occasions where it is not possible to omit the postposition. This happened in

[61] This video clip shows a man walking in a parking lot towards a red car.

example 74c. The postposition *shàng* 'top' must be added after *xíngréndào* 'sidewalk'.

Furthermore, when postpositions were not used, no significant difference was found between the two groups in terms of the use of adjuncts with information on direction ($\chi 2$ (1, N = 76) = 0.842, p = 0.359) and goal ($\chi 2$ (1, N = 22) = 0.727, p = 0.394). However, as presented above, when taking all goal-oriented adjuncts into consideration – regardless of the presence or absence of postpositions – L2 speakers used significantly more goal-oriented adjuncts than L1 speakers ($\chi 2$ (1, N = 30) = 4.800, p = 0.028).



	Loc	Direction	Source	Object-related	Goal	Loc	Direction	Source	Object-related	Goal	BoundaryCrossing
L1 Chinese	2.82%	11.30%	0.00%	5.08%	0.56%	61.58%	0.00%	5.08%	0.00%	1.69%	0.00%
L2 Chinese	10.76%	14.35%	1.79%	0.90%	4.48%	20.18%	0.90%	0.00%	0.00%	2.24%	3.59%

Figure 4.20: Use of adjuncts in L1 and L2 Chinese in the verbal task of Type B videos

Different from the descriptions of Type A videos, the percentages of locative adjuncts used in both groups to describe Type B videos increased (see Figure 4.20). This increase was extremely remarkable in L1 Chinese (64.4% in Type B vs. 36.55% in Type A, $\chi 2$ (1, N = 186) = 9.484, p = 0.002). However, this increase had no significance in L2 Chinese (30.94% in Type B vs. 24.30% Type A, $\chi 2$ (1, N = 130) = 0.492, p = 0.483). The prevalent use of locative adjuncts is on par with the features of the video clips in this category, since this type of video showed a long trajectory with a possible endpoint. L1 speakers preferred describing the situation as a Figure moving in a certain manner at a certain place, while L2 speakers did so to a lesser degree.

(75a) Adjuncts with postposition: encoding information on location

一个车在高速公路上行驶。(L1 Chn P07, B01)

yī-ge chē zài gāosùgōnglù-shàng xíng-shǐ.

one-CL car at highway-top move-drive

A car is driving on the highway.

(75b) Adjuncts without postposition: encoding information on location

在外面有一辆车走得越来越远。(L1 Ger-L2 Chn P22, B08)

zài wàimian yǒu yī-liàng chē zǒu de yuè-lái-yuè-yuǎn.

at outside exist one-CL car move CSC[62] *more-come-more-far*

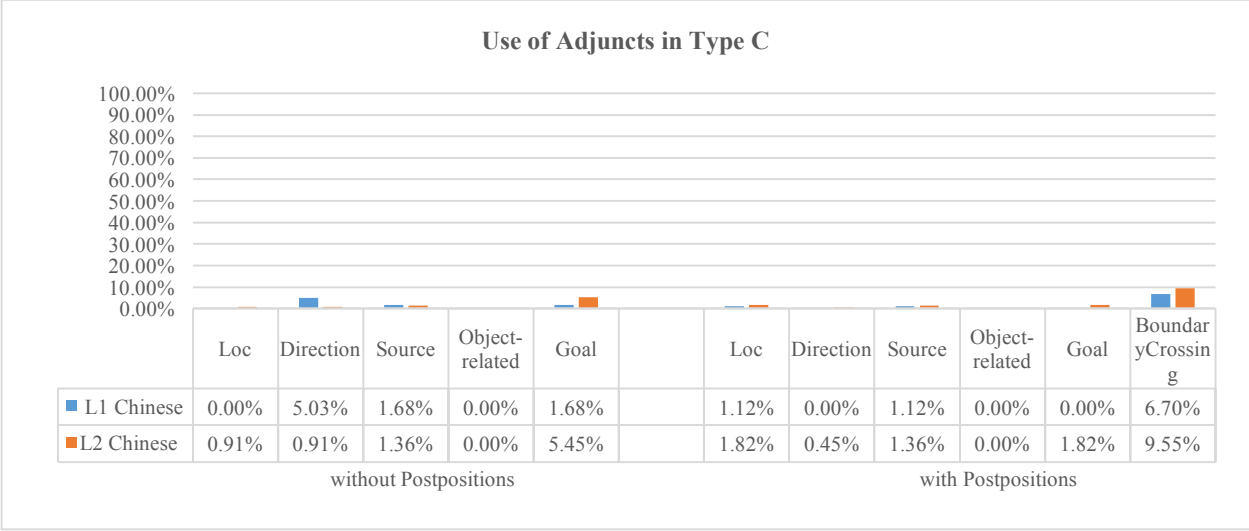
There is a car that moves further and further away.

In contrast, and irrespective of the presence or absence of postpositions, L2 speakers used significantly more goal-oriented adjuncts ($\chi 2$ (1, N = 19) = 6.368, p = 0.012) than L1 speakers, but no significant difference was found in the use of directional adjuncts in L2 Chinese ($\chi 2$ (1, N = 54) = 3.630, p = 0.057) compared to L1 Chinese. When postpositions were absent, it was found that L2 speakers also used significantly more goal-oriented adjuncts ($\chi 2$ (1, N = 11) = 7.364, p = 0.007) than L1 speakers, but there was no significant difference in the use of directional adjuncts between the two groups ($\chi 2$ (1, N = 52) = 2.769, p = 0.096). In addition, the use of locative adjuncts without postpositions was significantly higher in L2 Chinese than in L1 Chinese ($\chi 2$ (1, N = 29) = 12.448, p < .001), which is the opposite of the use of locative adjuncts with postpositions in L1 Chinese. That is, L1 speakers used significantly more locative adjuncts with postpositions than L2 speakers ($\chi 2$ (1, N = 154) = 26.597, p < .001). Concerning the use of object-related adjuncts without postpositions, it was found that L1 speakers used them significantly more often than L2 speakers ($\chi 2$ (1, N = 11) = 4.455, p = 0.035). In addition, the use of boundary-crossing adjuncts in L2 Chinese occurred at a rate of 3.59%, while no such data was found in L1 Chinese.

Type C videos showed boundary-crossing events. It has already been shown that the use of locative adjuncts decreased drastically in the descriptions of Type C videos in both language groups compared to the use of locative adjuncts in the

[62] CSC is short for 'complex stative construction', which consists of two clauses or one clause and a verb phrase. It could be schematized as 'clause de stative clause or verb phrase' (Li & Thompson, 1981, p. 623). The relationship between the two parts of this construction could be inferred depending on the meaning of the clauses. In example 75b, yuè-lái-yuè-yuǎn 'lit. more-come-more-far' further specifies the extent of the first clause.

descriptions of Type B videos. The reason for this marginal use resides in the property of the video types in this category. L1 and L2 speakers tended to depict the movement involved in crossing the boundary rather than the location where the movement took place. Concerning other adjuncts used in the verbal task of Type C videos, as shown in Figure 4.21, it was found that L2 speakers generally used significantly more goal-oriented adjuncts than L1 speakers ($\chi 2$ (1, N = 19) = 8.895, p = 0.003). This significant effect was also found if no postpositions were used ($\chi 2$ (1, N = 15) = 5.400, p = 0.020). The use of goal-oriented adjuncts with postpositions in L2 Chinese occurred only at a rate of 1.82%, whereas no such data was found in L1 Chinese. In addition, the use of directional adjuncts generally did not differ greatly between both groups ($\chi 2$ (1, N = 12) = 3.000, p = 0.083). However, the use of directional adjuncts without postpositions was significantly higher in L1 Chinese than in L2 Chinese ($\chi 2$ (1, N = 11) = 4.455, p = 0.035). The use of boundary-crossing adjuncts with postpositions did not differ in L1 and L2 Chinese ($\chi 2$ (1, N = 33) = 2.455, p = 0.117).



	Loc	Direction	Source	Object-related	Goal	Loc	Direction	Source	Object-related	Goal	BoundaryCrossing
L1 Chinese	0.00%	5.03%	1.68%	0.00%	1.68%	1.12%	0.00%	1.12%	0.00%	0.00%	6.70%
L2 Chinese	0.91%	0.91%	1.36%	0.00%	5.45%	1.82%	0.45%	1.36%	0.00%	1.82%	9.55%

Figure 4.21: Use of adjuncts in L1 and L2 Chinese in the verbal task of Type C videos

On the whole, apart from the differences in selecting different conceptual components encoded in adjuncts between L1 and L2 Chinese speakers, L2 speakers produced some inappropriate expressions, which can be summarized as problems regarding the omission of pre- and postpositions and the placement of postverbal locative phrases. The following examples illustrate these two problems encountered by L2 speakers.

Problems with preposition omission

There are two instances where prepositions can be omitted. As already stated in section 3.1.3, prepositions can generally not be omitted, unless an NP is placed in the initial position of presentative sentences. Example 24a is repeated here:

(76a) Appropriate omission of preposition

郊外有一辆车在移动。(L1 Ger-L2 Chn P06, B08)

jiāowài yǒu yī-liàng chē zài yí-dòng.
suburbs exist one-CL car ZAI move-move
There is a car moving in the suburbs.

(76b) Inappropriate omission of preposition

*人房间前面走路。(L1 Ger-L2 Chn P12, A02)

*rén fángjiān qiánmian zǒu-lù.
person room front walk-road

(76c) Inappropriate omission of preposition

*绿色的汽车开农村的一个马路。(L1 Ger-L2 Chn P20, B02)

*lǜsè de qìchē kāi nóngcūn-de yī-ge mǎlù.
green NOM car drive countryside-MOD one-CL road

The sentences in examples 76b and 76c lack the preposition *zài* 'at'. They should be corrected as *zài fángjiān-de qiánmian* 'lit. at room-MOD front' and *zài nóngcūn-de yī-ge mǎlù-shàng* 'lit. at countryside-MOD one-CL road-top', respectively. Note that the postposition *shàng* 'top, above' should also be added in example 76c.

Problems with postposition omission

Inappropriate expressions occur when the postpositions *shàng* 'top, above' in a locative adjunct and *nàbian* 'there' in a directional adjunct must be added, but L2 speakers fail to do so.

Inappropriate omission of the postposition *shàng* 'top, above'

(77) *白色的车在路开。(L1 Ger-L2 Chn P13, B05)

*báisè-de chē zài lù kāi.
white-MOD car at road drive

Inappropriate omission of a postposition like *nàbian* ‘that side’

(78a) *一个女生往车子走。(L1 Ger-L2 Chn P09, A04)

*yī-ge nǚshēng wǎng chēzi zǒu.
one-CL girl towards car walk

(78b) *有一个带袋子的女孩子往勒瑟桶走。(L1 Ger-L2 Chn P02, A10)

*yǒu yī-ge dài dàizi-de nǚháizi wǎng lèsètǒng zǒu
exist one-CL carry bag-MOD girl towards trash can walk

In example 77, the postposition *shàng* ‘top, above’ in the locative adjunct must be added. The correct version is *zài lù-shàng* ‘lit. at road-top’. Most L2 Chinese speakers added the postpositions where necessary. There are only six utterances with inappropriate omission of the postposition *shàng* ‘top, above’ in the locative adjunct. In example 78, L2 speakers had problems with changing a common noun to a place word in order to put it in the directional adjunct. The preposition *wǎng* ‘towards’ must be followed by a place word. *chēzi* ‘car’ and *lèsètǒng* ‘trash can’ are common nouns. By adding postpositions like *nàr*/ *nàli*/ *nàbian* ‘that side’ or nouns denoting position or direction after those common nouns, they can be used appropriately in the directional adjuncts, as shown in the following examples:

wǎng ‘towards’ + NP + postposition *nàr*/ *nàli*/ *nàbian* ‘that side’

(79) 一个女生往车子那边走。(L1 Ger-L2 Chn P12, A04)

yī-ge nǚshēng wǎng chēzi-nàbian zǒu.
one-CL girl towards car-that side walk
A girl walks towards a car.

wǎng ‘towards’ + NP + *fāngxiàng* ‘direction’ or dìfāng ‘place’

(80a) 一个女生往垃圾桶的方向走。(L1 Ger-L2 Chn P15, A10)

yī-ge nǚshēng wǎng lājītǒng-de fāngxiàng zǒu.
one-CL girl towards trash can-MOD direction walk
A girl walks towards a trash can.

(80b) 一个女生往可以打电话的地方走。(L1 Ger-L2 Chn P09, A07)

yī-ge nǚshēng wǎng kěyǐ dǎdiànhuà-de dìfāng zǒu.
one-CL girl towards be able to make a call-MOD place walk
A girl walks towards a place where she can make a call.

However, it seems that L2 Chinese speakers had not fully learned that *wǎng* 'towards' requires an NP that denotes a place or direction. Inappropriate expressions occurred mostly in the descriptions of Type A videos. There are 36 utterances including adjuncts with *wǎng* 'towards', of which 16 are inappropriate expressions. Two video clips are responsible for this: one showing a woman walking towards a green car (example 78a) and the other one showing a woman walking towards a trash can (examples 78b and 80a). As examples 78a and 78b show, L2 speakers used common nouns (e.g., *chēzi* 'car', *lājītǒng* 'trash can') instead of the required place words, so they produced idiosyncratic expressions. In the verbal task for Type B videos, there are 33 utterances that include adjuncts with *wǎng* 'towards'; however, all of them are appropriate expressions. These 33 utterances include directions such as *qián* 'front', *yòu* 'right', or *shàng* 'top' in the adjuncts with *wǎng* 'towards', as illustrated by examples 81 and 82:

(81) *wǎng* 'towards' + *qián* 'front' and *wǎng* 'towards' + *zuǒ* 'left'

车子往前开，慢慢往左走。(L1 Ger-L2 Chn P01, B01)

chēzi wǎng qián kāi, mànmàn wǎng zuǒ zǒu.
car towards front drive, slowly towards left move
A car drives forward and moves slowly towards the left.

(82) *wǎng* 'towards' + *shàng* 'top'

一个女孩在走路，在往上走。(L1 Ger-L2 Chn P01, B03)

yī-ge nǚhái zài zǒu-lù, zài wǎng shàng zǒu.
one-CL girl ZAI walk-road, ZAI towards top walk
A girl is walking up the road.

The expression of direction in adjuncts with *wǎng* 'towards' in the verbal task for Type B videos is in conformance with the characteristics of the type of videos in this category. There is no obvious endpoint, but a possible endpoint, which is far away from the Figure in Type B videos. So L2 Chinese speakers did not mention the endpoint explicitly as they did for Type A videos (e.g., car, trash can); rather, they tended to adopt a deictic perspective to describe the situation when using adjuncts with *wǎng* 'towards' to describe Type B videos. The total number of adjuncts used with *wǎng* 'towards' in the verbal task for Type C videos amounts to 8, all of which are used appropriately. 5 out of 8 utterances include the preposition *wǎng* 'towards' and the postposition *lǐ* 'inside' to indicate a boundary-crossing motion, which also corresponds to the property of this type of videos (see example 73). The remaining three utterances express the directional information similar to the use of adjunct with *wǎng* in example 81.

Problems with the placement of postverbal locative adjuncts

As stated in section 4.3.1.3, there is a strict syntactic restriction when locative adjuncts are placed after the verb. That is, postverbal locative adjuncts must immediately follow the verb. No other constituents are allowed to be inserted between verb and postverbal locative adjuncts, including the direct object of the verb (Li & Thompson, 1981). This restriction can be schematized as follows:

* verb + direct object + postverbal locative phrase
verb + postverbal locative adjunct

Example 26 is repeated below:

(83a) *又有人开车在比较偏僻的地方。 (L1 Ger-L2 Chn P04, B04)

*yòu yǒu rén kāi-chē zài bǐjiào piānpì-de dìfāng.
again exist person drive-car at relative remote MOD place

This should be corrected as follows:

yòu yǒu rén kāi zài bǐjiào piānpì-de dìfāng.
again exist person drive at relative remote MOD place
Someone again drives in a relatively remote place.

(83b) *一个女人骑自行车在一个农村。(L1 Ger-L2 Chn P20, B09)

*yī-ge nǚ-rén qí zìxíngchē zài yī-ge nóngcūn.
one-CL female-person ride bicycle at one-CL countryside

This should be corrected as follows:

yī-ge nǚ-rén qí zài yī-ge nóngcūn.
one-CL female-person ride at one-CL countryside
A woman rides in the countryside.

Most L2 Chinese speakers obeyed this rule. But there are still 12 out of 694 utterances that show problems with the placement of postverbal locative adjuncts.

Summary

In general, L1 German speakers used more adjuncts than L1 and L2 Chinese speakers, while L1 and L2 Chinese speakers did not differ in this regard. Concerning the use of locative adjuncts, no difference was found regarding its use among speakers of the three groups in the verbal task for Type A videos. However, in the verbal task for Type B videos, L1 Chinese speakers used them more frequently than speakers of L2 Chinese and L1 German, while speakers of L2 Chinese and L1 German did not differ in this regard. Regarding the use of directional adjuncts, L1 German speakers used them more than L1 and L2 Chinese speakers, while no significant effect was found between L1 and L2 Chinese speakers in the verbal task for the different video types. In addition, speakers of L2 Chinese and L1 German encoded more goal-oriented information in adjuncts overall than L1 Chinese speakers, while no difference was found between speakers of L2 Chinese and L1 German. In terms of encoding boundary-crossing information in adjuncts, L1 German speakers encoded this more often in their descriptions of Type A and Type C videos than L1 and L2 Chinese speakers did, while

no difference was found between L2 Chinese and L1 German in this regard regarding descriptions of Type B videos. Furthermore, speakers of L1 and L2 Chinese did not differ in the use of boundary-crossing adjuncts to describe Type C videos.

Given that postpositions like *lǐ/lǐbian/lǐmian/lǐtou* 'inside' play an important role in contributing boundary-crossing information in Chinese, an in-depth comparison with respect to the presence and absence of postpositions between L1 and L2 Chinese speakers was necessary. The results show that L2 Chinese speakers tended not to use postpositions, even in cases where postpositions are required, thus making grammatical mistakes.

4.4.1.3 Event unit formation: combination and segmentation

In describing motion events, conceptual components are encoded in a distributed way. They are encoded in verbs, particles, and adjuncts across different languages. Languages differ in how much path information they maximally package into an event unit (Bohnemeyer et al., 2007; Carroll et al., 2012; Gerwien & von Stutterheim, 2018). In principle, various path segments can be stacked after manner verbs in S-languages such as English or German (e.g., Hickmann & Hendriks, 2010; Slobin, 1996). In contrast, in E-languages such as Chinese, depending on the context, path verbs can only occur with certain path segments, so their use is rather restricted. This is because path verbs per se include the characteristics of the Ground object. For example, the path verb *jìn* 'enter' in Chinese is typically used when the Ground object is a three-dimensional object such as *fángzi* 'house', and *shàng* 'ascend' is used when the Ground object has a vertical feature, such as *lóutī* 'stairs'. That is, path verbs denote particular Path information. Thus, when path segments occur with path verbs, they must be compatible with the particular meaning denoted by the path verbs. If they are not compatible, then a second utterance that includes another manner verb, path verb, or adjunct must be produced. Hence, it is interesting to see how L1 German-L2 Chinese speakers cope with the differences found between German and Chinese with respect to packaging different semantic components into an event unit. To address this question, the following part provides an integrated overview of the combination of verbs, adjuncts, and particles that encode different conceptual components in describing different types of motion events among the three language groups.

One-clause-pattern

In general, speakers of the three language groups tended to utter one clause to

describe motion events. When calculating the frequency of occurrence of one clause in each group (collapsing them across descriptions of Type A, B, and C videos), it was found that the rate was 95.47% in L1 Chinese, 95.01% in L1 German, and 88.80% in L1 German-L2 Chinese. No significant difference was thus found among the three language groups ($\chi 2$ (2, N = 1586) = 1.643, p = 0.440). In the following, we will show the most typical way of combining semantic components in a one-clause-pattern among the three language groups. Note that the order of the semantic components represented in the combinations (from Figure 4.22 to Figure 4.26) is not fixed as shown; rather, it could be changed in actual use. For example, the combination of the semantic components Manner + Direction could be realized in actual use as manner verb + directional adjunct or directional adjunct + manner verb.

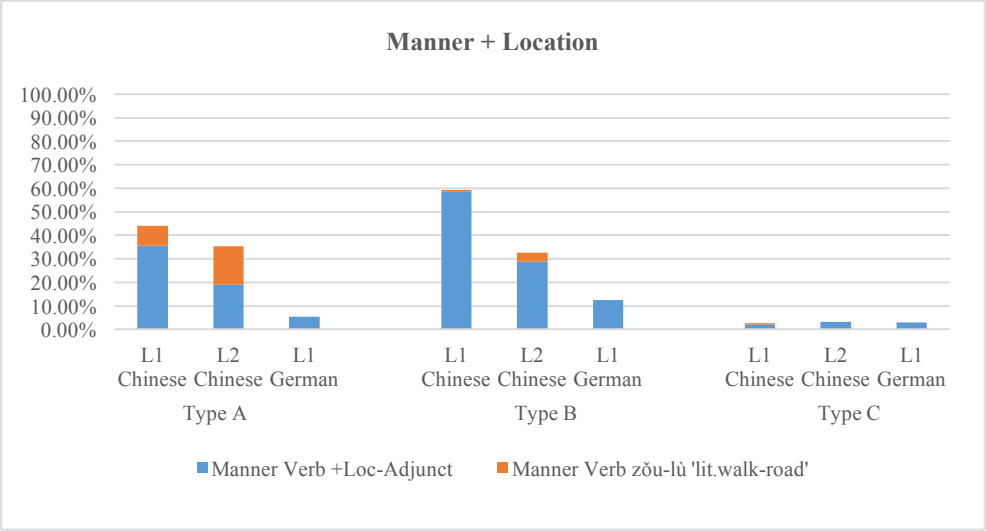


Figure 4.22: Encoding of Manner + Location by speakers of L1, L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos

In Chinese, *zǒu-lù* 'lit. walk-road' includes Manner as well as locative information, with the meaning of walking on the road. In general, speakers of L1 and L2 Chinese expressed Manner + Location more often than speakers of L1 German in describing Type A and B videos, while speakers of all three groups expressed them significantly less in describing Type C videos. In the verbal task for Type A videos, the frequency of occurrence of Manner + Location encodings is comparable in L1 and L2 Chinese ($\chi 2$ (1, N = 176) = 0.023, p = 0.880). When examining the use of the construction "manner verb + locative adjunct" (or "locative adjunct + manner verb") and the stand-alone use of the manner verb *zǒu-lù*, it was found that L2 Chinese speakers used the former construction significantly less often than L1 Chinese speakers

($\chi2$ (1, N = 118) = 4.102, p = 0.043), while the frequency of the use of the manner verb *zǒu-lù* in L2 Chinese was higher than in L1 Chinese ($\chi2$ (1, N = 58) = 9.931, p = 0.002). In contrast, the frequency of the occurrence of the construction "manner verb + locative adjunct" (or "locative adjunct + manner verb") was significantly lower in L1 German than in L2 Chinese ($\chi2$ (1, N = 100) = 60.840, p < .001) and in L1 Chinese ($\chi2$ (1, N = 98) = 58.939, p < .001), respectively. In the verbal task for Type B videos, the frequency of encoding Manner + Location was significantly higher in L1 Chinese than in L2 Chinese ($\chi2$ (1, N = 178) = 5.753, p = 0.016). When examining the construction "manner verb + locative adjunct" (or "locative adjunct + manner verb") and the manner verb *zǒu-lù* separately, it was found that L2 Chinese speakers used the construction "manner verb + locative adjunct" (or "locative adjunct + manner verb") significantly less often than L1 Chinese speakers ($\chi2$ (1, N = 168) = 9.524, p = 0.002), but the manner verb *zǒu-lù* more often than L1 Chinese speakers ($\chi2$ (1, N = 10) = 6.400, p = 0.011). Both L1 and L2 Chinese speakers encoded the combination of the concepts Manner + Location significantly more often than L1 German speakers (L1 Chinese vs. L1 German $\chi2$ (1, N = 127) = 54.244, p < .001 and L2 Chinese vs. L1 German $\chi2$ (1, N = 95) = 27.379, p < .001). In the verbal task for Type C videos, the frequency of Manner + Location encodings was significantly lower and no significant effect was found among the three groups ($\chi2$ (2, N = 18) = 0.333, p = 0.846).

As mentioned above, *zǒu-lù* 'lit. walk-road' conflates the information of Manner and Location. L2 Chinese speakers might neglect the locative meaning in this verb. Hence, they produced grammatically correct, but unnatural expressions. 8 out of 694 utterances in L2 Chinese sound unnatural, as the following examples illustrate:

? Directional adjunct + *zǒu-lù*

(84a) ? 有一个人在往一个公车站走路。(L1 Ger-L2 Chn P06, A02)

yǒu yī-ge rén zài <u>wǎng yī-ge gōngchēzhàn zǒu-lù</u>.
exist one-CL person ZAI towards one-CL bus station walk-road
There is a person walking towards a bus station.

(84b) ? 又有人在走路，向下走路。(L1 Ger-L2 Chn P13, B07)

yòu yǒu rén zài zǒu-lù, <u>xiàng xià zǒu-lù</u>.

again exist person ZAI walk-road, towards down walk-road

Again there is a person walking. He (or she) is walking down the road.

? Locative adjunct + *zǒu-lù*

(85) ? 一个人在路上走路。(L1 Ger-L2 Chn P25, B07)

yī-ge rén zài lù-shàng zǒu-lù.

one-CL person at road-top walk-road

A person is walking on the road.

If the compound verb *zǒu-lù* 'lit. walk-road' is reduced to *zǒu* 'walk' in examples 84 and 85, the sentences in the example sound quite natural.

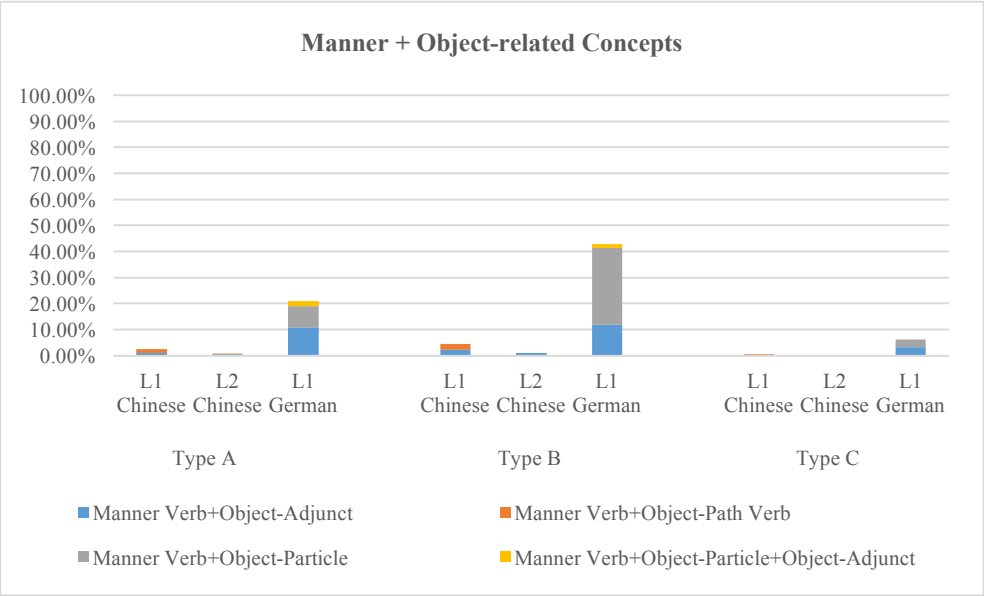


Figure 4.23: Encoding of Manner + Object-related concepts by speakers of L1 Chinese, L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos

L1 German speakers tended to encode more Manner + Object-related concepts than L1 and L2 Chinese speakers throughout their descriptions of Type A, B, and C videos. Statistical significance was found in the verbal task for Type A videos between L1 German and L1 Chinese ($\chi 2$ (1, N = 48) = 30.083, $p < .001$) and between L1 German and L2 Chinese ($\chi 2$ (1, N = 45) = 37.356, $p < .001$). No difference was found between L1 and L2 Chinese. Compared to the encoding of Manner + Object-related concepts in the verbal task for Type A videos, L1 German speakers encoded it more often in the verbal task for Type B videos ($\chi 2$ (1, N = 119) = 9.151, $p = 0.002$). It seems that L1

German speakers tended to conceptualize the long trajectory with respect to the features of the Ground object, which contrasts with L1 Chinese speakers, who preferred conceptualizing the situation as a Figure moving in a certain manner at a certain place. Unlike L1 Chinese speakers, L2 Chinese speakers did this to a lesser degree. They tended to encode Manner + Goal-oriented concepts (see Figure 4.25). No significance was found in encoding Manner + Object-related concepts between L1 and L2 Chinese in the verbal task for Type B videos ($\chi 2$ (1, N = 10) = 3.600, p = 0.058). In the verbal task for Type C videos, L1 German speakers again expressed more Manner + Object-related concepts than L1 and L2 Chinese (L1 German 6.19% vs. L1 Chinese 0.56%; L1 German 6.19% vs. L2 Chinese 0%). The following examples show the constructions used:

Manner verb + object-related adjunct

(86a) 一个女人沿着田地在走。(L1 Chn P03, B07)

yī-ge nǚ-rén yán-zhe tiándì zài zǒu.
one-CL female-person along-ZHE field ZAI walk
A woman is walking along the field.

(86b) Eine Frau geht um eine Ecke. (L1 Ger P02, A01)
one woman walk around one corner
A woman walks around a corner.

Manner verb + object-related path verb

(87) 一辆车驶过郊区。(L1 Chn P20, B01)

yī-liàng chē shǐ guò jiāoqū.
one-CL car drive pass suburb
A car drives past a suburb.

Manner verb + object-related particle

(88) Ein Auto fährt eine Straße entlang. (L1 Ger P08, B01)
one car drive one street along
A car drives along a street.

Manner verb + object-related particle + object-related adjunct

(89) Eine Person läuft an einem alten Gebäude vorbei. (L1 Ger P04, A02)

one person walk by one-DAT old building past

A person walks past an old building.

As shown in Figure 4.24, in the verbal task for Type A videos, there was no significant difference in the frequency of Manner + Direction encodings among the three groups ($\chi 2$ (1, N = 93) = 3.161, p = 0.206). The reason for this is found in the characteristics of Type A videos, which showed a short trajectory with an obvious endpoint. Because the place at goal was evident, speakers of different languages were likely to mention it by means of directional adjuncts or particles. However, the encoding of Manner + Direction differed greatly among the three language groups in the verbal task for Type B videos. L2 Chinese speakers expressed more Manner + Direction than L1 Chinese speakers ($\chi 2$ (1, N = 38) = 8.526, p = 0.004) and L1 German speakers ($\chi 2$ (1, N = 30) = 22.533, p < .001). In the verbal task for Type B videos, L1 German speakers tended to encode Manner + Object-related concepts (Figure 4.23) or Manner + Goal-oriented concepts (Figure 4.25), so the frequency of occurrence of Manner + Direction encodings was lower. In the verbal task for Type C videos, the frequency of occurrence of Manner + Direction was rather low among all three groups and no significant effect was found ($\chi 2$ (1, N = 19) = 2.000, p = 0.368).

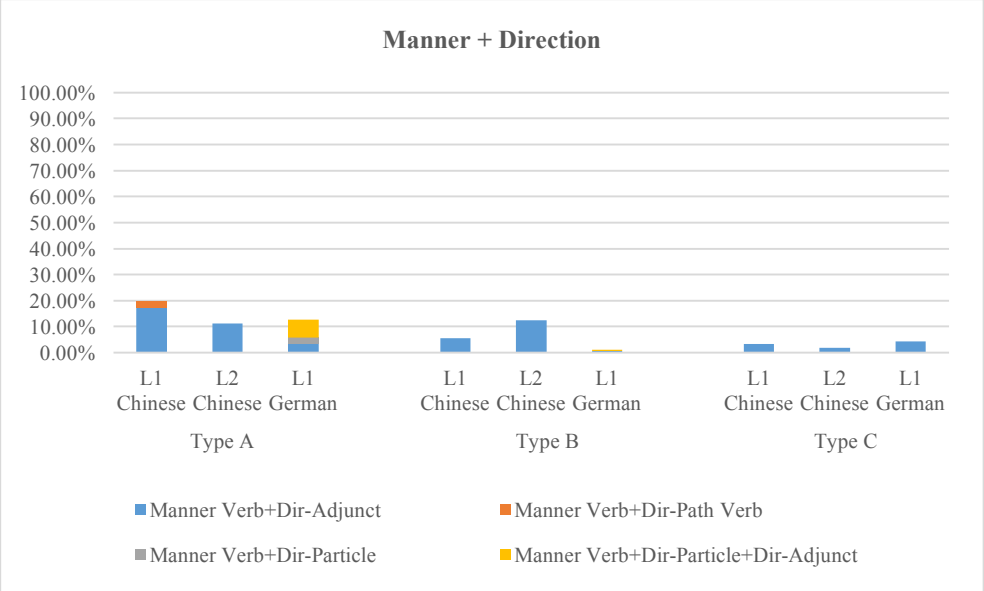


Figure 4.24: Encoding of Manner + Direction by speakers of L1 Chinese, L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos

Interestingly, when considering the encoding of Manner + Direction in the verbal task for Type A, B, and C videos overall, it was found that L2 Chinese speakers tended to put the directional adjuncts before the manner verbs (53 out of 56 utterances had the word order "directional adjunct + manner verb"; see example 91; 3 out of 56 utterances used the reverse word order; see example 90), while L1 Chinese speakers were inclined to do it in the reverse way, i.e., putting the manner verb before the directional adjuncts (44 out of 50 utterances used the word order "manner verb + directional adjunct"; see example 90; in contrast, 6 out of 50 utterances had the reverse order "directional adjunct + manner verb"; see example 91). There is no semantic difference between preverbal directional adjuncts and postverbal directional adjuncts. The preference for the preverbal position of directional adjuncts might reflect L2 speakers' focus on the endpoint, since they process the directional information earlier than L1 speakers. The following examples show the constructions that occurred.

(90) Manner verb + directional adjunct

这个人在走向公共电话亭。(L1 Chn P17, A07)

zhè-ge rén zài zǒu xiàng gōnggòngdiànhuàtíng.
this-CL person ZAI walk towards public telephone booth
A woman is walking towards a public telephone booth.

(91) Directional adjuncts + manner verbs

一个女生往可以打电话的地方走。(L1 Ger-L2 Chn P09, A07) (example 37b repeated)

yī-ge nǚshēng wǎng kěyǐ dǎdiànhuà-de dìfāng zǒu.
one-CL girl towards be able to make a call-MOD place walk
A girl walks towards a place where she can make a call.

(92) Eine Frau läuft in Richtung einer Telefonzelle. (L1 Ger P10, A07)
one woman walk in the direction of one-DAT telephone booth
A woman walks towards a telephone booth.

(93) Manner verb + directional path verb

一辆巴士正驶近车站。(L1 Chn P01, A09)

yī-liàng bāshì zhèng shǐ-jìn chēzhàn.

one-CL bus ZHENG drive-approach station

A bus is driving towards the station.

(94) Manner verb + directional particle

Ein Mann klettert eine Leiter hoch. (L1 Ger P08, A03)

one man climb one ladder up

A man climbs up a ladder.

(95) Manner verb + directional particle + directional adjunct

Ein Mann klettert vorsichtig eine Leiter hoch bis zu einem Balkon. (L1 Ger P06, A03)

one man climb carefully one ladder up to one-DAT balcony

A man carefully climbs up a ladder to a balcony.

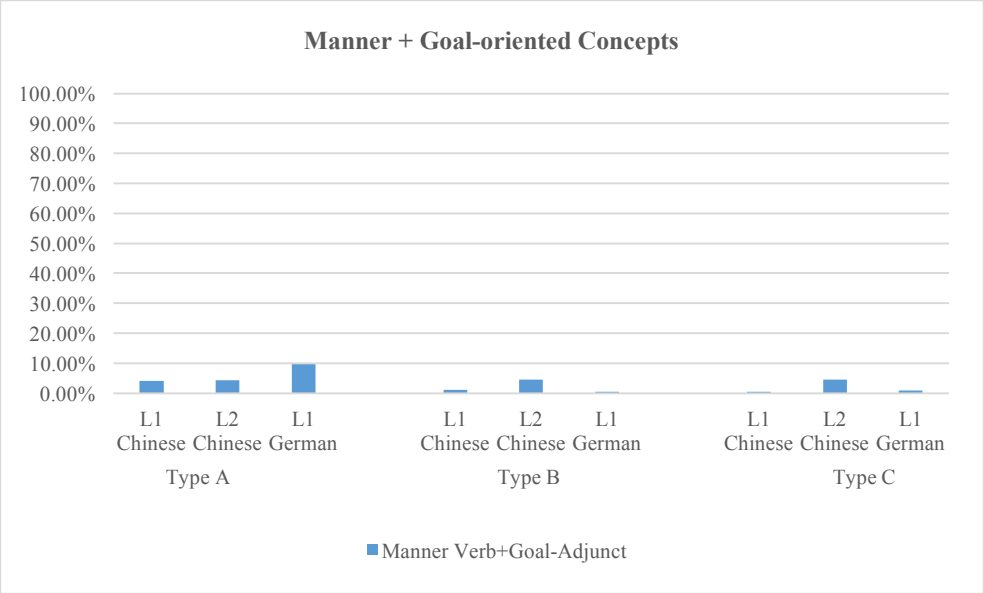


Figure 4.25: Encoding of Manner + Goal-oriented concepts by speakers of L1 Chinese, L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos

In general, as shown in Figure 4.25, the frequency of the encoding Manner + Goal-oriented[63] was low in the verbal task for Type A, B, and C videos in all three

[63] Note that the goal-oriented concept is encoded in the coverb *dào* 'arrive, to' in Chinese. Although in Figure 4.25 it is labeled as a goal-oriented adjunct, it is also a goal-oriented path verb.

groups. In the verbal task for Type A videos, L1 German speakers expressed it significantly more often than L1 Chinese speakers ($\chi 2$ (1, N = 28) = 5.143, p = 0.023) and L2 Chinese speakers ($\chi 2$ (1, N = 31) = 2.613, p = 0.106), respectively. However, no significant effect was found between L1 and L2 Chinese speakers ($\chi 2$ (1, N = 19) = 0.474, p = 0.491). Different from descriptions of Type A videos, in the verbal task for Type B videos, L2 Chinese speakers encoded Manner + Goal-oriented concepts more often than L1 Chinese speakers ($\chi 2$ (1, N = 12) = 5.333, p = 0.021) and L1 German speakers ($\chi 2$ (1, N = 11) = 7.364, p = 0.007). In the verbal task for Type C videos, L2 Chinese speakers also encoded it significantly more often than speakers of L1 Chinese ($\chi 2$ (1, N = 11) = 7.364, p = 0.007) and L1 German ($\chi 2$ (1, N = 12) = 5.333, p = 0.021), despite the low frequency of occurrence. The following examples show constructions used in L2 Chinese and L1 German, respectively.

(96) Manner verb + goal-oriented adjunct

公交车开到下一站。(L1 Ger-L2 Chn P24, B06)

gōngjiāochē kāi-dào xià yī-zhàn.

bus drive-arrive next one-station

The bus drives to the next station.

(97) Ein Bus fährt zur Bushaltestelle. (L1 Ger P05, A09)

one bus drives to-DAT bus station

A bus drives to a bus station.

In Figure 4.26, the concepts of Manner and Deictic in parentheses indicate that they may or may not appear in the constructions listed. The occurrence of encoding boundary-crossing with other concepts is mainly clustered in the verbal task for Type C videos, since Type C videos showed boundary-crossing events. Despite its low frequency, encoding of boundary-crossing information also appeared in the verbal task for Type A and B videos. In the verbal task for Type A videos, there was no significant effect in encoding the combination of boundary-crossing and other concepts between L2 Chinese and L1 German speakers ($\chi 2$ (1, N = 22) = 0.727, p = 0.394); however, a significant difference was found between L2 Chinese and L1 Chinese speakers ($\chi 2$ (1, N = 17) = 4.765, p = 0.029). In the verbal task for Type B videos, L2 Chinese speakers

encoded the combination of boundary-crossing and other concepts more often than L1 Chinese speakers ($\chi 2$ (1, N = 19) = 15.211, p < .001) and L1 German speakers ($\chi 2$ (1, N = 22) = 8.909, p = 0.003).

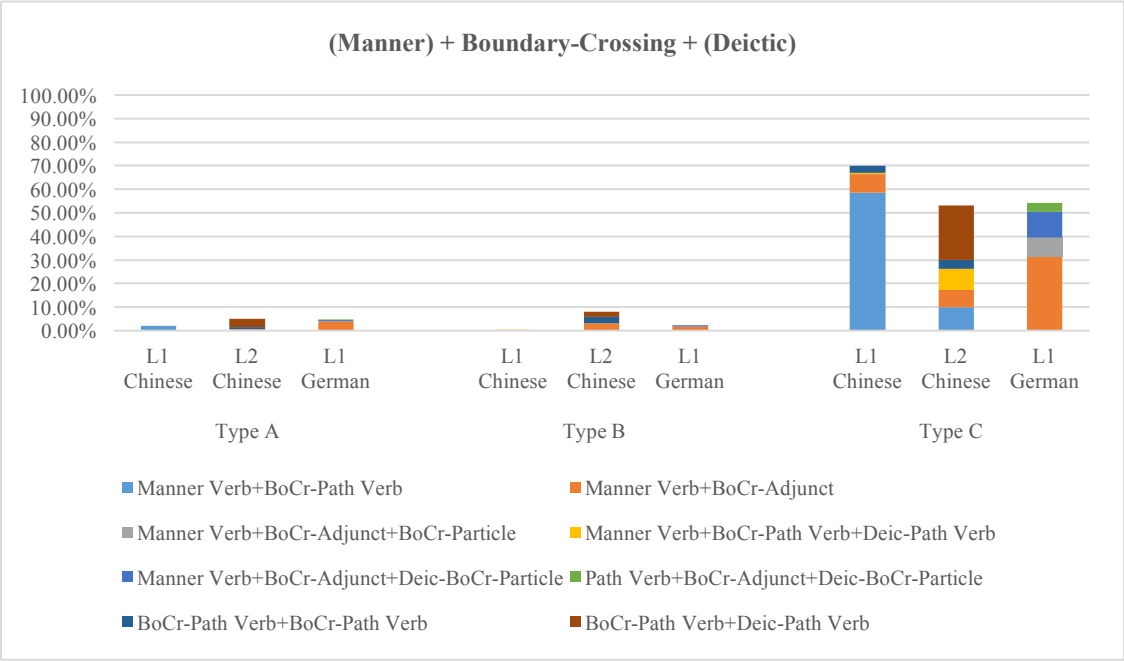


Figure 4.26: Encoding of (Manner) + Boundary-Crossing + (Deictic) by speakers of L1 Chinese, L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos

In the verbal task for Type C videos, the constructions used varied among the different language groups (see Figure 4.26). L1 Chinese speakers used the construction "manner verb + boundary-crossing path verb" significantly more often than L2 Chinese speakers (L1 Chinese 58.66% vs. L2 Chinese 10.00%, $\chi 2$ (1, N = 127) = 54.244, p < .001), whereas L2 Chinese speakers used the constructions "manner verb + boundary-crossing path verb + deictic verb" (9.09%) and "boundary-crossing path verb + deictic verb" (23.18%) more often than L1 Chinese speakers (the former construction 0.56% and the latter 0%). As discussed in section 4.4.1.1, the preference for the use of deictic verbs in L2 Chinese might stem from the directionality indicated by deictic verbs, the bimorphemic status of modern Chinese, and L2 Chinese speakers' inclination to take a holistic viewpoint. L1 German speakers used the construction "manner verb + boundary-crossing adjunct" more often than speakers of L1 Chinese ($\chi 2$ (1, N = 80) = 33.800, p < .001) and L2 Chinese ($\chi 2$ (1, N = 82) = 30.488, p < .001). There was no significant effect in the use of the construction "manner verb + boundary-crossing adjunct" between L1 and L2 Chinese speakers ($\chi 2$ (1, N = 30) = 0.133, p = 0.715).

Furthermore, L1 and L2 Chinese speakers used two boundary-crossing path verbs together: L2 Chinese speakers used them more often than L1 Chinese speakers in the verbal task for Type A and Type B videos, respectively (L1 Chinese 0% vs. L2 Chinese 0.80% in Type A and L1 Chinese 0% vs. L2 Chinese 2.69% in Type B), while there was no significant difference in the verbal task for Type C videos between L1 and L2 Chinese speakers ($\chi 2$ (1, N d= 13) = 0.692, p = 0.405). In contrast to L1 and L2 Chinese speakers, L1 German speakers additionally used boundary-crossing particles and deictic particles with a boundary-crossing meaning. The relative frequency of occurrence of the construction "manner verb + boundary-crossing adjunct + boundary-crossing particle", "manner verb + boundary-crossing adjunct + deictic-boundary-crossing particle", and "path verb + boundary-crossing adjunct + deictic boundary-crossing particle" occurred at a rate of 8.10%, 10.95%. and 3.81%, respectively, in the verbal task for Type C videos. The preference for different constructions for encoding Manner + Boundary-Crossing concepts in the three language groups reflects the typological features across different types of languages. Path information is typically encoded in path verbs in Chinese, while the same information is encoded in adjuncts or particles in German[64]. The following examples show the constructions used:

(98) Manner verb + boundary-crossing path verb

有人快速地跑进停车场或者车站。(L1 Ger-L2 Chn P23, C02)

yǒu rén kuàisù-de pǎo-jìn tíngchēchǎng huòzhě chēzhàn
exist person quick-CSC run-enter parking lot or station
Someone quickly runs into a parking lot or station.

(99) Manner verb + boundary-crossing path verb + deictic verb

*这只猫慢慢地走进去了房间。(L1 Ger-L2 Chn P01, C06)

*zhè-zhī māo màn-màn-de zǒu-jìn-qù le fángjiān.
this-CL cat slowly walk-enter-go LE room
This cat walked slowly into the room.

[64] German speakers also used boundary-crossing path verbs in the verbal task for Type C videos. But they occurred at a lower rate. 16 out of 210 utterances included the path verb *betreten* 'enter'.

(100) Boundary-crossing path verb + deictic verb

*一个女人进去超市。(L1 Ger-L2 Chn P20, C04)

*yī-ge nǚ-rén jìn-qù chāoshì.

one-CL female-person enter-go supermarket

A woman entered a supermarket.

Note that there are word order problems in examples 99 and 100. The place NPs *fángjiān* 'room' and *chāoshì* 'supermarket' should be inserted between the boundary-crossing path verb and the deictic verb, such as *zǒu-jìn le fángjiān qù* 'walk-enter LE room go' and *jìn chāoshì qù* 'enter supermarket go' in examples 99 and 100, respectively.

Boundary-crossing path verb + boundary-crossing path verb

(101) 汽车往右拐，打算进入城市。(L1 Ger-L2 Chn P09, B01)

qìchē wǎng you guǎi, dǎsuan jìn-rù chéngshì.

car towards right turn, be going to enter-enter city

A car turns right and is going to enter the city.

Manner verb + boundary-crossing adjunct

(102a) 一个人往火车站里跑。(L1 Ger-L2 Chn P05, C02)

yī-ge rén wǎng huǒchēzhàn-lǐ pǎo.

one-CL person towards train station-inside run

A person runs into the train station.

(102b) Jemand rennt ins Bahngebäude. (L1 Ger P05, C02)

someone run in-ACC train building

Someone runs into the train building.

Manner verb + boundary-crossing adjunct + boundary-crossing particle

(103) Hier fährt ein großes weißes Auto in eine Hofeinfahrt ein. (L1 Ger P06, C03)

here drive one big white car in one-ACC courtyard entrance in

Here drives a big white car into a courtyard entrance.

Manner verb + boundary crossing adjunct + deictic-boundary crossing particle

(104) Eine Katze schleicht in ein Zimmer hinein. (L1 Ger P03, C06)

one cat sneak in one-ACC room thither-in

A cat sneaks into a room (away from the speaker).

Path verb + boundary-crossing adjunct + deictic-boundary crossing particle

(105) Ein Auto biegt in eine Garage ein. (L1 Ger P19, C01)

one car turn in one-ACC garage in

A car turns into a garage.

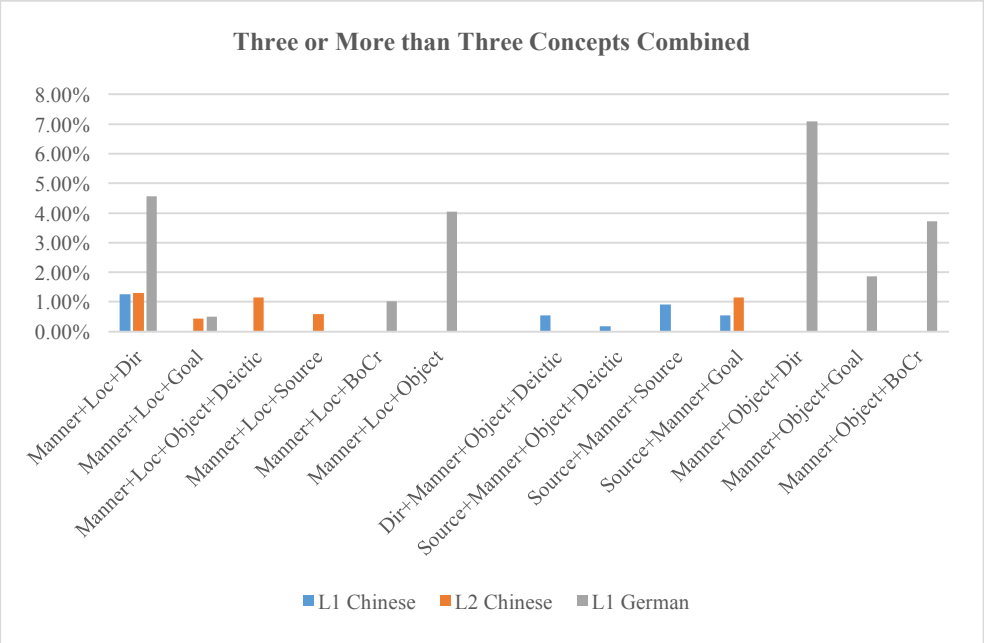


Figure 4.27: Encoding of three or more concepts by speakers of L1 Chinese, L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos overall

Figure 4.27 shows some combinations of three or more concepts in the L1 and L2 groups[65]. Note that the order of the spatial concepts in the combination is not fixed;

[65] There were more than 40 possible combinations of different concepts in the three groups. Most occurred rather rarely. Some possible combinations are shown as follows:
L1 Chinese: Object-Adjunct + Manner Verb + Source-Path Verb (0.90%), Source-Adjunct + Manner Verb + Goal-Path Verb (0.54%), Direction-Adjunct + Manner Verb + Boundary-Crossing-Path Verb + Deictic (0.18%), Object-Adjunct + Manner Verb + Object-Path Verb + Deictic (0.18%), etc.
L2 Chinese: Direction-Adjunct + Direction-Path Verb (0.14%), Direction-Adjunct + Figure-Path Verb (0.43%), Object-Adjunct + Manner Verb + Object-Path Verb + Deictic (0.14%), Direction-Adjunct + Manner Verb + Source-Path Verb (0.14%), etc.
L1 German: Motion-Verb + Direction-Adjunct + Loc (0.68%), Motion- Verb + Loc (0.17%), Manner Verb + Source + Boundary-Crossing-Adjunct (0.34%), Direction-Particle + Manner Verb + Goal-Adjunct (0.34%), Direction-

rather, it differs in different constructions across different languages. The frequency of encoding Manner + Loc + Direction was found to be higher in L1 German than in L1 Chinese ($\chi 2$ (1, N = 34) = 11.765, p = 0.001) and L2 Chinese ($\chi 2$ (1, N = 36) = 9.000, p = 0.003), while no significant effect was found between L1 and L2 Chinese ($\chi 2$ (1, N = 16) = 0.250, p = 0.617). The encoding of Manner + Location + Goal occurred extremely rarely in the L1 and L2 groups (L1 Chinese 0%, L2 Chinese 0.43%, and L1 German 0.51%). Anyway, the combination of Manner + Location + Goal sounds quite unnatural in Chinese. Consider the following example:

(106) ? Manner + Loc + Goal

? 有一个人在路上走到一个公交车站。(L1 Ger-L2 Chn P22, A02)

yǒu yī-ge rén zài lù-shàng zǒu-dào yī-ge gōngjiāochēzhàn.
exist one-CL person at road-top walk-arrive one-CL bus station
Someone walks on the road to a bus station.

In addition, combinations of the semantic components "Manner + Location + Object + Deictic" and "Manner + Location + Source" also sound unnatural in Chinese, but both occurred at a rather lower rate in L2 Chinese (1.15% for the former combination and 0.58 for the latter). For example:

(107) ? Manner + Location + Object + Deictic

? 一个人在地下停车场走过去。(L1 Ger-L2 Chn P08, A05)

yī-ge rén zài dìxià tíngchēchǎng zǒu-guò-qù.
one-CL person at underground parking lot walk-move over-go
A person walks across the underground parking lot (away from the speaker)

(108) ? Manner + Location + Source

? 一辆车在路上开走。(L1 Ger-L2 Chn P03, B04)

yī-liàng chē zài lù-shàng kāi-zǒu.
one-CL car at road-top drive-leave

Particle + Manner Verb + Boundary-Crossing-Adjunct (0.84%), Boundary-Crossing Deictic-Particle + Manner Verb + Goal-Adjunct (0.34%), etc.

A car drives away on the road.

The unnatural expressions in examples 106, 107, and 108 are a result of the co-occurrence of a locative adjunct with an SVC, i.e., manner verb + path verb. The reason for the unacceptable combination does not lie in the manner verb, since manner verbs can typically be combined freely with various adjuncts (Figures 4.23- 4.26). Rather, the reason is grounded in the use of the path verb (*dào* 'arrive' in example 106, *guò* 'pass, cross, move over' in example 107, and *zǒu* 'leave' in example 108) in SVCs. It seems that path verbs cannot occur with locative adjuncts in Chinese. The reason might be related to two facts. First, path verbs implicate the orientation of the Figure in space. That is, path verbs encode directional information, which is compatible with an adpositional phrase encoding directional or dynamic information. Locative adjuncts indicate the place where a motion takes place, which might not be compatible with the directional requirement of path verbs. Second, the Ground object following path verbs has a vector-like structure (e.g., verticality, border property, etc.), while the Ground object in a locative adjunct is considered to have a set of locations that include the Figure as it moves through space in a motion event (Flecken, Carroll, et al., 2015). The features of the Ground object required by path verbs and locative adpositions differ greatly. Accordingly, there are three options for improving examples 106, 107, and 108, respectively.

The first option is to omit the locative adjunct.

(109) 有人走到那个公共汽车站。(L1 Ger-L2 Chn P04, A02)

yǒu rén zǒu-dào nà-ge gōnggòngqìchēzhàn.
exist person walk-arrive that-CL bus station
Someone walks to the bus station.

The second option is to divide the utterance into two utterances.

(110) 一个女人走在停车场上，直着走过去。(L1 Ger-L2 Chn P03, A10)

yī-ge nǚrén zǒu zài tíngchēchǎng, zhí-zhe zǒu-guò-qù.
one-CL woman walk at parking lot, straight-ZHE walk-move over-go
A woman walks on the parking lot; she walks straight across (away from the

speaker).

The last option is to use adjuncts with a directional or dynamic meaning.

(111a) 一个身着黑衣的人从草坪上走到路上。(L1 Chn P20, B03)

yī-ge shēnzhuó hēiyī-de rén cóng cǎopíng-shàng zǒu-dào lù-shàng.

one-CL wear black clothes-NOM person from lawn-top walk-arrive road-top

A person with black clothes walks from the lawn to the road.

(111b) 一辆公交车正朝着我们开过来。(L1 Chn P04, A09)

yī-liàng gōngjiāochē zhèng cháo-zhe wǒmen kāi-guò-lái.

one-CL bus ZHENG towards-ZHE we drive-move over-come

A bus is driving towards us.

(111c) 一辆车沿着路开走了。(L1 Chn P20, B04)

yī-liàng chē yán-zhe lù kāi-zǒu le.

one-CL car along-ZHE drive-leave LE

A car drove away along the road.

(111d) 一辆车从公路上开走了。(L1 Chn P11, B01)

yī-liàng chē cóng gōnglù-shàng kāi-zǒu le.

one-CL car from road-top drive-leave LE

A car drove away from a road.

In addition, Figure 4.27 shows that despite the low numbers of occurrence, L1 Chinese speakers did use the combination of Direction + Manner + Object + Deictic (0.58%, example 111b) and Source + Manner + Source (0.90%, example 111d). The occurrence of the combination of Source + Manner + Goal (example 111a) was found more often in L2 Chinese (1.15%) than in L1 Chinese (0.54%).

In contrast to the less frequent occurrence of combinations of three or more semantic components in L1 and L2 Chinese, the combination of different components

was found to be more frequent in L1 German: Manner + Location + Boundary-Crossing (1.01%), Manner + Location + Object (4.05%), Manner + Object + Direction (7.09%), Manner + Object + Goal (1.86%), and Manner + Object + Boundary-Crossing (3.72%). The discrepancy in the capacity for combining different semantic components between Chinese and German resides in the different typological features. As previous studies (e.g., Hickmann & Hendriks 2010, Slobin 1996) have suggested, conceptual components are easier to stack in Path satellites in S-languages such as German, liberating the verb to encode Manner information. In contrast, in E-languages such as Chinese, verbs are reserved for expressing Manner and Path information. Due to the syntactic and semantic constraints imposed by the path verbs, certain semantic components are incompatible in Chinese. Typologically speaking, the restrictions imposed by the path verbs on the combination of different path segments in the same construction are also found in French (a V-language), so one cannot say **Il traverse la rue au magasin* 'He crosses the street to a shop' (Gerwien & von Stutterheim, 2018), because the path verb *traverse* 'to cross' limits the combination of other path segments. As a consequence, German speakers can provide denser information within one utterance. In contrast, if L1 and L2 Chinese speakers want to express the same amount of information, they have to express it in two or more utterances.

Two-clause pattern

In comparison with the occurrence of one clause in the three language groups, the occurrence of two clauses was low[66]; the rate was 4.53% in L1 Chinese, 11.20% in L2 Chinese, and 4.99% in L1 German (averaging across the descriptions of Type A, B, and C videos). Use of two clauses was found significantly more often in L2 Chinese than in L1 Chinese ($\chi 2$ (1, N = 93) = 21.774, $p < .001$) and L1 German ($\chi 2$ (1, N = 97) = 17.330, $p < .001$), while there was no significant difference between L1 Chinese and L1 German speakers ($\chi 2$ (1, N = 52) = 0.308, $p = 0.579$).

[66] Three criteria were applied to determine whether a description had one utterance or more than two utterances: (1) whether there was an obvious pause between two subevents; (2) whether there was a linking conjunction such as *and or but* between two subevents; (3) whether there was a MACRO-EVENT PROPERTY (MEP) in an expression. An MEP is "a property of constructions that assesses the event construal they convey – specifically, the 'tightness of packaging' of subevents in the construction. A construction has the MEP if temporal operations such as time adverbials, temporal clauses, and tenses necessarily have scope over all subevents encoded by the construction" (Bohnemeyer, 2007, p. 497).

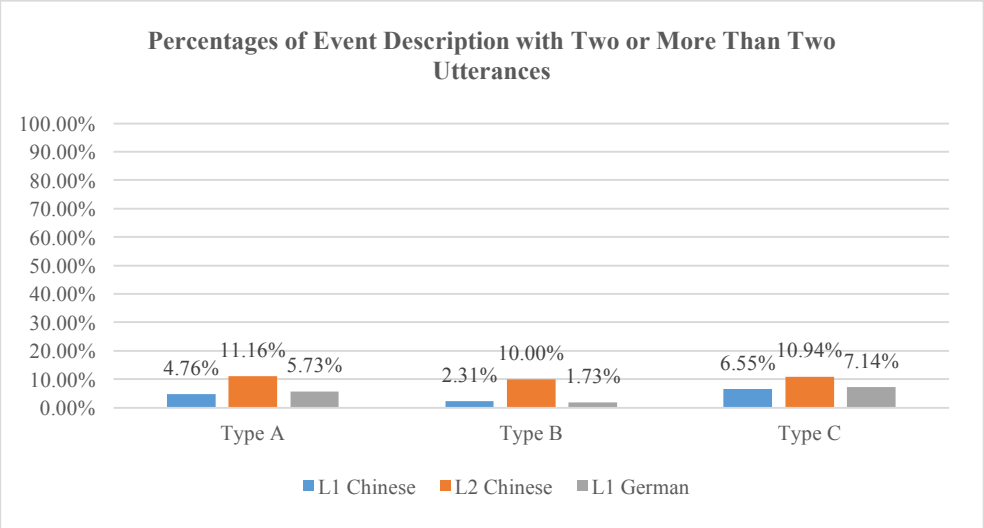


Figure 4.28: Percentages of event descriptions with two or more utterances by speakers of L1 Chinese, L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos, respectively

When considering two or more utterances produced in the descriptions of Type A, B, and C videos overall, as shown in Figure 4.28, it was found that there was no significant difference across different video types in L1 Chinese ($\chi 2$ (2, N = 24) = 3.250, p = 0.197) and in L2 Chinese ($\chi 2$ (2, N = 69) = 0.609, p = 0.738), while in L1 German, a significant effect was found between Type A and Type B videos ($\chi 2$ (1, N = 14) = 4.571, p = 0.033) and between Type B and Type C videos ($\chi 2$ (1, N = 17) = 7.118, p = 0.008), but no significance between Type A and Type C videos ($\chi 2$ (1, N = 25) = 0.360, p = 0.549). When considering two or more utterances produced in the verbal task for Type A and Type B videos, respectively, the same tendency was found among the three language groups: the frequency of occurrence of two or more utterances was higher in L2 Chinese than in L1 Chinese (Type A: $\chi 2$ (1, N = 34) = 7.529, p = 0.006; Type B: $\chi 2$ (1, N = 24) = 10.667, p = 0.001) and in L1 German (Type A: $\chi 2$ (1, N = 36) = 5.444, p = 0.020; Type B: $\chi 2$ (1, N = 23) = 12.565, p < .001). In the verbal task for Type C videos, L2 Chinese speakers produced two utterances more often than L1 Chinese speakers ($\chi 2$ (1, N = 35) = 4.829, p = 0.028), whereas no significant effect was found between speakers of L2 Chinese and L1 German ($\chi 2$ (1, N = 38) = 2.632, p = 0.105).

When checking for which situation two utterances were most likely to be produced in L2 Chinese, it was found that when the situation involved a change of the Figure's orientation or direction, L2 speakers tended to use two utterances. This is in line with the findings of Gerwien and von Stutterheim (2018). Particularly, two Type A video clips appeared to have 'triggered' L2 Chinese speakers to produce two utterances: one showing a woman walking around a corner towards a car and the other

one showing a man walking upstairs towards the entrance of a building. It is interesting to note that 6 out of 24 L2 Chinese speakers and 6 out of 20 L1 German speakers produced two or more utterances in describing the first situation, and 4 out of 24 L2 Chinese speakers and 3 out of 20 L1 German speakers did so in describing the second situation. By contrast, only 1 out of 20 L1 Chinese speakers produced two utterances in describing each of the two situations mentioned above. L2 Chinese speakers tended to express locative meaning in the first clause and directional meaning (either mentioning or not mentionin g the endpoint) in the second clause. This is illustrated by the following examples:

Type A: a woman walking around a corner towards a car

(112a) 女孩子在走路，她在右转。(L1 Ger-L2 Chn P01, A01)

Clause 1: nǚháizi zài zǒu-lù,

girl ZAI walk-road

A girl is walking on the road.

Clause 2: tā zài yòu zhuǎn.

she ZAI right turn

She is turning right.

(112b) Eine Frau geht den Bürgersteig entlang. Sie biegt nach rechts ab. (L1 Ger P20, A01)

Clause 1: Eine Frau geht den Bürgersteig entlang.

one woman walk the-ACC pavement along

A woman walks along the pavement.

Clause 2: Sie biegt nach rechts ab.

she turn towards right off

She turns right.

Type A: a man walking upstairs towards the entrance of a building

(113a) *有一个男人他走上去一个楼梯，然后准备进去一个比较大的房子。(L1 Ger-L2 Chn P22, A06)

Clause 1: *yǒu yī-ge nán-rén tā zǒu-shàng-qù yī-ge lóutī,

exist one-CL male-person he walk-ascend-go one-CL stairs

There is a man who walks upstairs (away from the speaker)

Clause 2: *ránhòu zhǔnbèi jìn-qù yī-ge bǐjiào dà-de fángzi.

then prepare enter-go one-CL relatively big-NOM house

then he is about to enter a relatively big house.

Note that the place NPs *lóutī* 'stairs' and *fángzi* 'house' should be inserted between the non-deictic path verb and the deictic verb in the SVCs *shàng qù* 'ascend go' in the first clause and *jìn qù* 'enter go' in the second clause (for more details, see section 4.4.1.1).

(113b) Ein Mann bewegt sich auf ein Gebäude zu und steigt die Treppen zu dem Gebäude hoch. (L1 Ger P16, A06)

Clause 1: Ein Mann bewegt sich auf ein Gebäude zu

one man move himself towards one building towards

A man moves towards a building

Clause 2: steigt die Treppen zu dem Gebäude hoch.

ascend the stairs to the-DAT building up

(he) climbs up the stairs to the building.

In the verbal task for Type B videos, as discussed above, both L1 Chinese and L1 German speakers rarely produced two utterances, while L2 speakers were more likely to express direction in the first clause and then mention the endpoint in the second clause. Type B videos showed a long trajectory with a distant endpoint. The long trajectory might have a big curve. L2 Chinese speakers preferred expressing the direction related to this characteristic. That is, they expressed directional information related to the features of the Ground object. This is illustrated in the following examples found in the verbal task for Type B videos:

(114a) 汽车往右拐，打算进入城市。(L1 Ger-L2 Chn P09, B01)

Clause 1: qìchē wǎng you guǎi,

car towards right turn

A car turns right.

Clause 2: dǎsuan jìn-rù chéngshì.

be going to enter-enter city

(it) is going to enter the city.

(114b) 一个巴士转弯，开到一个村庄里面。(L1 Ger-L2 Chn P23, B06)

Clause 1: yī-ge bāshì zhuǎn-wān,

one-CL bus turn-bend

A bus turns,

Clause 2: kāi-dào yī-ge cūnzhuāng-lǐmian

drive-arrive one-CL village-inside

(it) drives into a village.

In the verbal task for Type C videos, it seems that three video clips were more likely to induce L2 speakers to utter two or more clauses: one showing a dog running across a square into a house; one showing a man walking up the stairs into a church, and another one showing a man riding a bicycle and turning into an entrance. In the dog scene, 5 out of 24 L2 speakers uttered two sentences: 4 out of 5 speakers expressed the directional information in the first clause and the information about the arrival at the endpoint in the second clause (example 115a), while 1 out of 5 speakers expressed the locative information in the first clause and the arrival at the endpoint in the second clause (example 115b). In sharp contrast, none of the L1 Chinese speakers produced two utterances in this dog scene. All of them focused on the final phase of the situation, that is, entering the house, since this was extremely salient in this situation. There was one L1 German participant who uttered two clauses in this dog scene. Consider the following examples:

Type C: a dog running across a square into a house

(115a) 有一只狗从左边跑到右边，*进去一个房子。(L1 Ger-L2 Chn P22, C05)

Clause 1: yǒu yī-zhī gǒu cóng zuǒ-bian pǎo-dào yòu-bian,

exist one-CL dog from left-side run-arrive right-side

There is a dog that runs from left to right,

Clause 2: *jìn-qù yī-ge fángzi.

enter-go one-CL house

(it) enters a house.

Note that in the second clause, the place NP *fángzi* 'house' should be put between jìn 'enter' and qù 'go'.

(115b) 一个停车的地方，有一条狗在上面跑，进房间了。(L1 Ger-L2 Chn P22, C05)

Clause 1: yī-ge tíngchē-de dìfāng, yǒu yī-tiáo gǒu zài shàng-mian pǎo,

one-CL parking-NOM place, exist one-CL dog at top-face run,

There is a dog that runs on a parking lot,

Clause 2: jìn fángjiān le.

enter room LE

(it) entered the room.

In describing the situation 'a man walking upstairs into a church', 3 out of 24 L2 Chinese speakers and 2 out of 20 L1 Chinese speakers chose to express the 'walking upstairs' information in the first clause and the boundary-crossing information 'into a church' in the second clause (example 116). In contrast, none of the L1 German speakers produced two utterances for this situation, since in German the two subevents, namely, 'upstairs' and 'into a church' can be easily integrated into one utterance.

Type C: a man walking upstairs into a church

(116) 一个男人爬着一个楼梯，然后走进一个房子。(L1 Ger-L2 Chn P05, C08)

Clause 1: yī-ge nán-rén pá-zhe yī-ge lóutī,

one-CL male-person climb-ZHE one-CL stairs

A man is climbing stairs,

Clause 2: ránhòu zǒu-jìn yī-ge fángzi.

then walk-enter one-CL house

then walks into a house.

When considering the descriptions of the situation ‘a man is riding a bicycle and turning into an entrance’, it was found that only 3 out of 24 L2 Chinese speakers produced two utterances, while the production of two utterances was more frequent in L1 Chinese (7 out of 20 participants) and L1 German (9 out of 20 participants). There are two ways of describing this situation in Chinese using a two-clause pattern: (1) The first clause expresses the manner of the motion with or without locative information and the second clause expresses the boundary-crossing information (example 117a); (2) the first clause conveys the directional information and the second clause the boundary-crossing information (example 117b). All of the L1 Chinese speakers who produced two clauses chose the first option, while 1 out of 3 L2 Chinese speakers who uttered two clauses chose the first option and the other two speakers chose the second option. Although in L1 German, the subevents ‘riding a bicycle’, ‘along the street’, and ‘into an entrance’ could be integrated into one utterance, when L1 German speakers expressed the motion of ‘turning’, they chose to use the path verb *einbiegen* ‘turn in’ or *abbiegen* ‘turn off’. When these path verbs are used, a second utterance is needed, since there is a change in the Figure’s orientation or direction (example 117c). This is illustrated in the following examples:

Type C: a man riding a bicycle and turning into an entrance

(117a) 一个人在街上骑自行车，并且骑进了一栋建筑。(L1 Chn P08, C09)

Clause 1: yī-ge rén zài jiē-shàng qí zìxíngchē,
one-CL person at street-top ride bicycle
A person is riding a bicycle on the street,
Clause 2: bìngqiě qí-jìn le yī-dòng jiànzhù.
and ride-enter LE one-CL building
and riding into a building.

(117b) 一个男生骑自行车往右拐，进入中文系的院子。(L1 Ger-L2 Chn P09, C09)

Clause 1: yī-ge nánshēng qí zìxíngchē wǎng yòu guǎi,
one-CL boy ride bicycle towards right turn
A boy turns right on his bicycle,

Clause 2: jìn-rù zhōngwénxì-de yuànzi.
enter-enter Chinese studies department-NOM yard
enters the yard of the department of Chinese Studies.

(117c) Ein Mann fährt auf seinem Fahrrad die Straße entlang und biegt in einen Hof ab. (L1 Ger P13, C09)

Clause 1: Ein Mann fährt auf seinem Fahrrad die Straße entlang
one man ride on his-DAT bicycle the street along
A man rides his bicycle along the street

Clause 2: und biegt in einen Hof ab.
and turn in one-ACC yard off
and turns in a yard.

Summary

In sum, speakers of all three language groups were more likely to produce one utterance than two utterances. When considering the combination of semantic components in the one-clause pattern, it was found that L2 Chinese speakers tended to combine Manner with Direction, or Goal or Boundary-Crossing with Deictic information, while L1 Chinese speakers tended to combine Manner with Location, Manner with Direction, and Manner with Boundary-Crossing information, L1 German speakers preferred combining Manner with Object-related information and Manner with Boundary-Crossing information (realized in adjuncts or particles in actual use). When three or more semantic components were encoded in one utterance, L1 German-L2 Chinese speakers faced a challenge since German and Chinese differ remarkably in their ability to combine different concepts.

German is a type of S-language. This type of languages generally permits different path adjuncts and particles in a verb phrase. That is, it is allowed to integrate the subevents of departure from the source, passing of a Ground object, and arrival at the endpoint into one event unit since manner verbs do not impose restrictions on the combination of different path segments (e.g., walk from A along B over C to D). In contrast, due to the syntactic and semantic constraints imposed by path verbs in Chinese, only certain semantic components can be combined. The differences between German and Chinese regarding the ability to combine path segments may reside in the verb type chosen as the predicate, which determines the event type. In German, the event type is typically framed by manner verbs, e.g., *walk*, *drive*, *ride,* etc., while the event type in

Chinese is typically framed by a manner verb, e.g., *zǒu* 'walk' or *kāi* 'drive', or by a combination of manner verbs and path verbs, e.g., *zǒu-jìn* 'walk-enter' or *zǒu-jìn-qù* 'walk-enter-go' (see the use of SVCs in section 4.4.1.1). If an asserted event type is determined by manner verbs in Chinese (i.e., if manner verbs occur alone in the expression), then the combination of path segments with manner verbs is relatively free; e.g., *zǒu zài lù-shàng* 'lit. walk at road-top', *zǒu xiàng chēzi* 'lit. walk towards car'. In contrast, if the event type asserted is determined by SVCs, i.e., a combination of manner verbs and path verbs, then the co-occurrence of path segments is restricted. For example, the expression *zài lù-shàng zǒu-guò-qù* 'lit. at road-top walk-cross-go' sounds quite odd in Chinese, but the expression *xiàng chēzi zǒu-guò-qù* 'lit. towards car walk-cross-go' sounds natural. L2 Chinese speakers might not be used yet to this new type of event unit formation. Despite the low occurrence, some of them attached different path segments to a manner verb, which is a way of event unit formation in L1 German. However, this may result in odd expressions in Chinese.

It seems that L2 Chinese speakers wanted to express more information than L1 Chinese speakers. To this end, as discussed above, they stacked different path segments after one manner verb, which may result in unnatural expressions in Chinese. Alternatively, they chose to produce two or more utterances in order to express the same amount of information as their L1 German speakers did within one utterance. By doing so, L2 speakers "accommodated" the semantic restrictions imposed by the path verbs.

When considering the descriptions where two utterances were produced, it was found that L2 Chinese speakers produced two utterances more often than L1 Chinese and L1 German speakers across their descriptions of Type A, B, and C videos. Two utterances were more likely to be produced in L2 Chinese when the situation involved a change of the Figure's orientation or direction. This is on par with the findings by Gerwien and von Stutterheim (2018). In addition, it was found that although in the Type B videos, there was no obvious change of the Figure's orientation in the respective situation, L2 Chinese speakers paid attention to the features of the Ground (e.g., a big curve of the road) and drew on path verbs to encode such features in a separate clause, mentioning the endpoint in another clause. The information encoded in two clauses in the verbal task for Type B videos corresponds to L1 German speakers' habitual event construal pattern; that is, they tended to encode information derived from the features of the Ground object and mentioned the endpoint (e.g., Carroll et al., 2012; Flecken et

al., 2015; von Stutterheim et al., 2012).

When speakers utter two clauses, they can choose which information to encode in the first, respectively the second clause. In the first clause, they can choose to express the manner of motion with or without locative information, or the path of the motion with directional information; in the second clause, they can choose to express the moving direction or verbalize the endpoint. The analysis of the information encoded in the first clause showed differences between L1 and L2 Chinese: L1 Chinese speakers preferred the first option (i.e., encoding the manner of the motion with or without locative information in the first clause), while L2 Chinese speakers preferred the second option (i.e., encoding the path of the motion with directional information in the first clause). L1 and L2 Chinese speakers did not differ with regard to the information encoded in the second clause (i.e., encoding directional information or the endpoint) when speakers of both groups uttered two clauses. L1 Chinese speakers' two-clause pattern is in line with the findings of Carroll et al. (2012) and von Stutterheim et al. (2017), which showed that when speakers of verb-framed languages (e.g., French, Arabic) divide an event into two sub-events, the first event typically shows that an entity is moving in a certain manner at a certain location and the second event shows that an entity reaches the endpoint by referring to the path information, that is, using path verbs and mentioning the endpoint at goal. In comparison with L1 Chinese speakers, when L2 Chinese speakers produced two utterances, they preferred encoding directional or dynamic information rather than locative information in the first clause and verbalizing the endpoint in the second clause.

4.4.2 Motion event cognition in L1 and L2 Chinese

We compared the spatial concepts encoded in L1 Chinese, L1 German-L2 Chinese, and L1 German in section 4.4.1. The analysis showed that speakers of the different languages differed in their inclination towards encoding different spatial concepts. L1 German-L2 Chinese speakers tended to follow the conceptual pattern of L1 German, that is, they tended to package directional and goal-oriented information into linguistic expressions and preferred taking a holistic view to describe motion events. However, through these linguistic products alone we cannot get an insight into the language planning process, that is, the conceptualization phase in the process of language production. To unveil the difference in the conceptualization phase in the verbal tasks, speech onset times and fixation patterns (i.e., fixation on the moving entity

and the endpoint) will be analyzed in this section.

Speech onset times (SOTs) are an important indicator for cognitive processes, since at the level of visual attention they indicate the effort made to perceive and process visual information or the attention paid to different components, while at the linguistic level they indicate the cognitive cost for organizing and structuring linguistic information (Beyer, 2016; von Stutterheim & Carroll, 2006). In other words, SOTs indicate the time span required by the speakers to process the visual and linguistic information. Given the above-mentioned role that SOTs play in language production tasks, the total fixations and the fixations before SOTs in different language groups, respectively, will be examined in this section. Previous empirical studies have shown the correlation between the grammatical aspect and endpoint encoding (e.g., Flecken et al., 2014; von Stutterheim et al., 2012; von Stutterheim & Carroll, 2006). In this section, we will examine whether speakers of the three language groups differed in their cognitive processes, that is, speech onset times (SOTs) and fixation patterns, and to what extent L2 Chinese speakers exhibited their L1's conceptual patterns.

4.4.2.1 Speech onset times

As shown in Table 4.6, generally speaking, L1 German speakers started speaking earlier than L1 and L2 Chinese speakers, while L1 Chinese speakers began to speak earlier than L2 Chinese speakers. That is, in comparison with speakers of L1 Chinese and L1 German, L2 Chinese speakers had late SOTs. L2 speakers tend to take more time generally. This is because the conceptual patterns in L2 are not that automatized[67].

One-way ANOVA showed a significant effect among the three groups ($F(2, 1795) = 42.812$, $p < .001$). Post-hoc tests (Dunnett T3) revealed significant differences in average SOTs (averaging across Type A, B, and C videos) between speakers of L1 Chinese ($M = 3175$, $SD = 1832$) and L2 Chinese ($M = 3573$, $SD = 1647$) ($p < 0.001$), between speakers of L2 Chinese and L1 German ($M = 2709$, $SD = 1365$) ($p < 0.001$), and between speakers of L1 Chinese and L1 German ($p < 0.001$).

[67] We can not exclude the L2 effect, so I have to admit the limitation of this study in terms of the conclusions drawn from comparing the SOTs between L1 and L2 speakers.

Table 4.6: Average speech onset times in the verbal task of Type A, B, and C videos by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German

	L1 Chinese		L1 German-L2 Chinese		L1 German	
	Mean	SD	Mean	SD	Mean	SD
Type A	3394ms	1923ms	3745ms	1689ms	2792ms	1362ms
Type B	2996ms	1850ms	3521ms	1616ms	2596ms	1383ms
Type C	3117ms	1705ms	3446ms	1626ms	2728ms	1351ms
Total	3175ms	1832ms	3573ms	1647ms	2709ms	1365ms

When taking the impact of different types of videos on the SOTs into account, different language groups showed the same tendency, namely that the latest SOTs occurred in the verbal task for Type A videos. The reason for that may lie in the fact that although the endpoint in Type A videos was evident, the video stopped playing before it showed the moving entity actually reaching the endpoint. Therefore, the speakers had to wait to see whether the moving entity reached the endpoint or not. In the verbal task for Type A videos, L1 German speakers had earlier SOTs than L1 and L2 Chinese speakers. One-way ANOVA revealed significant differences in the SOTs in the verbal task for Type A videos among the three groups ($F(2, 617) = 17.236$, $p < 0.001$). Dunnett T3 post-hoc tests showed a significant effect between speakers of L2 Chinese ($M = 3745$, $SD = 1689$) and L1 German ($M = 2792$, $SD = 1362$) ($p < 0.001$), and between speakers of L1 Chinese ($M = 3394$, $SD = 1923$) and L1 German ($p < 0.01$), while no significance was found between speakers of L1 Chinese and L2 Chinese ($p = 0.137$).

In the verbal task for Type B videos, speakers of L1 Chinese and L1 German showed the earliest SOTs, compared to Type A and C videos. That is, the order for SOTs in L1 Chinese and L1 German was Type A > Type C > Type B. Speakers of L1 Chinese and L1 German tended to have late SOTs in the verbal task for Type A videos and early SOTs in those of Type B videos. This may be attributable to the characteristics of Type B videos, which showed an entity moving along a long trajectory towards a possible endpoint. Perhaps due to the long trajectory and the non-obvious endpoint, speakers did not need more time to get the relevant information than for Type A and C videos. In contrast, L2 Chinese speakers had late SOTs in the verbal task for Type B videos compared to L1 Chinese and L1 German speakers. The order for SOTs in L2 Chinese was Type A > Type B > Type C. This is because L2 Chinese speakers tended to encode directional, goal-oriented, and boundary-crossing information in the verbal

task for Type B videos. To do so, they had to wait until the endpoint became clear and then started to speak. From the temporal perspective, the relative frequency of occurrence of the progressive marker *zai* in the verbal task for Type B videos amounted to 37.89% in L2 Chinese, which is significantly less than in L1 Chinese (57.30%) (see more details in Chapter 5). That is, L2 Chinese speakers tended to focus on the endpoint (see the following discussion on fixation patterns), which required them to wait until the endpoint became identifiable as the scene unfolded. Those might be possible reasons why the SOTs in the verbal task for Type B videos were later in L2 Chinese than in L1 Chinese. The statistical results showed significant differences in the SOTs in the verbal task for Type B videos in the three groups ($F (2, 555) = 15.413$, $p < 0.001$). Dunnett T3 post-hoc tests revealed a significant effect between speakers of L2 Chinese ($M = 3521$, $SD = 1616$) and L1 Chinese ($M = 2996$, $SD = 1850$) ($p < .05$) and between speakers of L2 Chinese ($M = 3521$, $SD = 1616$) and L1 German ($M = 2596$, $SD = 1383$) ($p < .001$), while no significant difference was found between speakers of L1 Chinese ($M = 2996$, $SD = 1850$) and L1 German ($M = 2596$, $SD = 1383$) ($p = 0.061$).

In the verbal task for Type C videos, both L1 and L2 Chinese speakers had late SOTs compared with L1 German speakers. One-way ANOVA showed a significant effect among the three language groups ($F (2, 617) = 10.957$, $p < 0.001$). Significant differences were found between L2 Chinese and L1 German ($p < 0.001$) and between L1 Chinese and L1 German ($p < 0.05$), while no significance was found between L1 and L2 Chinese ($p = 0.127$).

In a nutshell, SOTs in L1 and L2 Chinese were found to be longer than in L1 German, while L2 Chinese speakers had later SOTs than L1 Chinese speakers. Given that the L2 Chinese participants were not native speakers, we have to admit that there are L2 effects that might have given rise to their later SOTs compared to L1 Chinese speakers. However, we do not attribute tremendous importance to the role of L2 proficiency when interpreting the SOTs in L2 Chinese since all the L1 German-L2 Chinese speakers had been carefully selected for this study based on their proficiency in Chinese, i.e., their HSK Levels. They were qualified as advanced Chinese learners (for more details, see Chapter 3 about the participants' social and educational background). Therefore, the variable L2 proficiency could be considered as high and constant during the language production tasks. Taking this characteristic of the L2 Chinese speakers into account, we assume that it was not L2 proficiency alone which led to later SOTs in L2; rather, we believe that there must be other factors responsible

for this. These factors may be the language-specific structures in Chinese (i.e., syntactic complexity of serial verb constructions, restrictions on the combination of semantic components, compatibility with aspect markers). Serial verb constructions, especially directional verb compounds, are typically used in the description of motion events. The word order differs according to the type of directional verb compounds and the type of NPs (see section 3.1.1). This syntactic complexity requires speakers to process the information about the type of directional verb compounds and NPs in order to make an appropriate decision to insert NPs, especially place NPs, into the correct position. Furthermore, L1 and L2 Chinese speakers were free to choose which phases of the situation to verbalize (i.e., the initial/ intermediate/terminative phase or the holistic situation), and they could choose to either use or not to use aspectual markers to mark their utterances when describing the above-mentioned phases. Once they had chosen an aspect marker, they then had to pay attention to the position of this aspect marker (e.g., the progressive aspect marker *zài* precedes the verb, whereas the perfective aspect marker *le* follows the verb). At the same time, they also needed to heed the compatibility of the aspect marker with the serial verb construction. For example, the progressive aspect marker *zài* is not compatible with serial verb constructions that are 2-state expressions (more details see Chapter 5). It is hypothesized that all of these language-specific features are so intricately intertwined that it takes longer for L1 and L2 Chinese speakers to select and structure the relevant information. Furthermore, given that L1 German-L2 Chinese speakers tended to take a holistic view to encode directional information, it is assumed that the habitual event construal patterns of L1 German speakers had an impact upon L2 Chinese speakers. This is because if L2 speakers paid attention to the endpoint, they had to wait until the endpoint became identifiable while the scene unfolded. Therefore, the cognitive load may have been heavier for L2 Chinese. To examine this assumption, in the following we will analyze the fixation patterns used by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German.

4.4.2.2 Fixation counts and durations

The eye-tracking data for the allocation of attention to the relevant AoIs, i.e., endpoint and moving entity (Figure), was analyzed based on the following measures: total number of fixation counts, number of fixation counts before SOTs, total duration of fixations, and duration of fixations before SOTs. The entire analysis was carried out

with average measures across speakers over different items in Type A, B, and C videos, respectively. In what follows, I will first present the relative frequency of the endpoint being mentioned at the linguistic level between L1 and L2 speakers. Then I will compare the number and duration of fixations on the endpoint between L1 and L2 speakers. I want to find out whether there is a correlation between linguistic encoding and the cognitive process related to endpoint encoding.

The linguistic expressions including the moving entity's arrival or intention to arrive at the goal (either specified or unspecified) were calculated as endpoint encoding. For example, driving to a village, walking towards a car, enterin g a shop, driving to an unknown place. The relative frequency of the endpoint being mentioned by speakers of the three groups is represented in Figure 4.29.

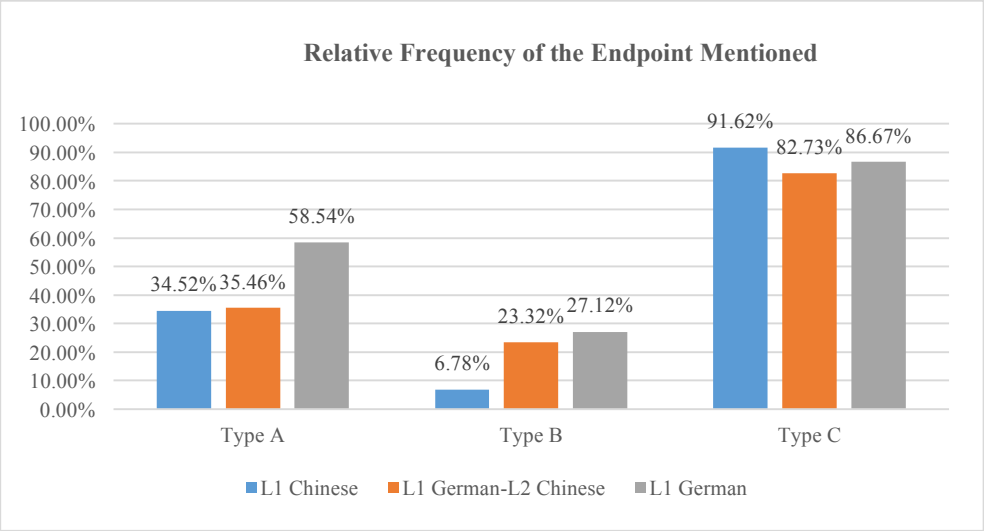


Figure 4.29: Bar charts showing the relative frequency of the endpoint being mentioned by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German in the verbal task of Type A, B, and C videos

The endpoint was most salient in Type C videos, followed by Type A videos, and was least salient in Type B videos. Accordingly, the endpoint was mentioned most frequently in the verbal task for Type C videos, followed by those of Type A, and least often in those of Type B by speakers of all three groups. However, the frequency of the endpoint being mentioned within each video type differed greatly among speakers of L1 and L2 Chinese and L1 German. In the verbal task for Type A videos, L1 German speakers mentioned the endpoint significantly more often than speakers of L1 Chinese ($\chi 2$ (1, N = 188) = 14.383, $p < .001$) and L2 Chinese ($\chi 2$ (1, N = 209) = 4.598, p = 0.032), whereas no significant difference was found between L1 and L2 Chinese ($\chi 2$ (1,

N = 157) = 2.809, p = 0.094). In the verbal task for Type B videos, L2 Chinese speakers mentioned the endpoint remarkably more often than L1 Chinese speakers ($\chi 2$ (1, N = 64) = 25.000, p < .001). The same was true for L1 German speakers, who mentioned the endpoint significantly more often than L1 Chinese speakers as well ($\chi 2$ (1, N = 60) = 21.600, p < .001). In contrast, no significant relationship is documented between L2 Chinese and L1 German ($\chi 2$ (1, N = 100) = 0.160, p = 0.689). In the verbal task for Type C videos, no significant effect was found among speakers of all three groups ($\chi 2$ (2, N = 528) = 1.227, p = 0.541).

Eye-tracking data analysis

Type A: Short trajectory with obvious endpoint

Fixation on the endpoint

Table 4.7: Mean value and standard deviation of fixation counts and duration of fixations on the endpoint in total and before SOTs for Type A videos

	L1 Chinese		L2 Chinese		L1 German	
Fixation on the Endpoint (Type A)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	2.82	2.67	2.58	2.33	3.05	2.32
Before SOTs	1.49	1.98	1.54	1.74	1.27	1.47
Duration of Fixations						
Total (in ms)	160	115	177	146	193	127
Before SOTs (in ms)	128	131	153	159	150	149

As shown in Table 4.7 above and Figure 4.30 below, in the verbal task for Type A, concerning the total fixation counts on the endpoint in the L1 and L2 groups, one-way ANOVA showed no significant effect among speakers of the three groups (F (2, 569) = 1.697, p = 0.184) (Tukey HSD post-hoc tests: L2 Chinese (M = 2.58, SD = 2.328) vs. L1 Chinese (M = 2.82, SD = 2.670), p = 0.609; L2 Chinese vs. L1 German (M = 3.05, SD = 2.320), p = 0.157; L1 Chinese vs. L1 German, p = 0.642). The same tendency was also found for the fixation counts before SOTs. That is, no significant difference was found among the three groups (F (2, 546) = 1.237, p = 0.291) (Dunnett T3 post-hoc tests: L2 Chinese (M = 1.54, SD = 1.74) vs. L1 Chinese (M = 1.49, SD = 1.98), p = 0.989; L2 Chinese vs. L1 German (M = 1.27, SD = 1.47), p = 0.312; L1 Chinese vs. L1 German, p = 0.542). However, when comparing the total duration of fixations among the three groups, a significant effect was only found between L1 Chinese and L1 German (p < 0.05), whereas no significance was found between L2 Chinese and L1 Chinese (p = 0.509) and between L2 Chinese and L1 German (p =

0.577). When comparing the duration of fixations before SOTs, no significant relationship was documented between L2 Chinese and L1 Chinese (p = 0.315), between L2 Chinese and L1 German (p = 0.998), and between L1 Chinese and L1 German (p = 0.315).

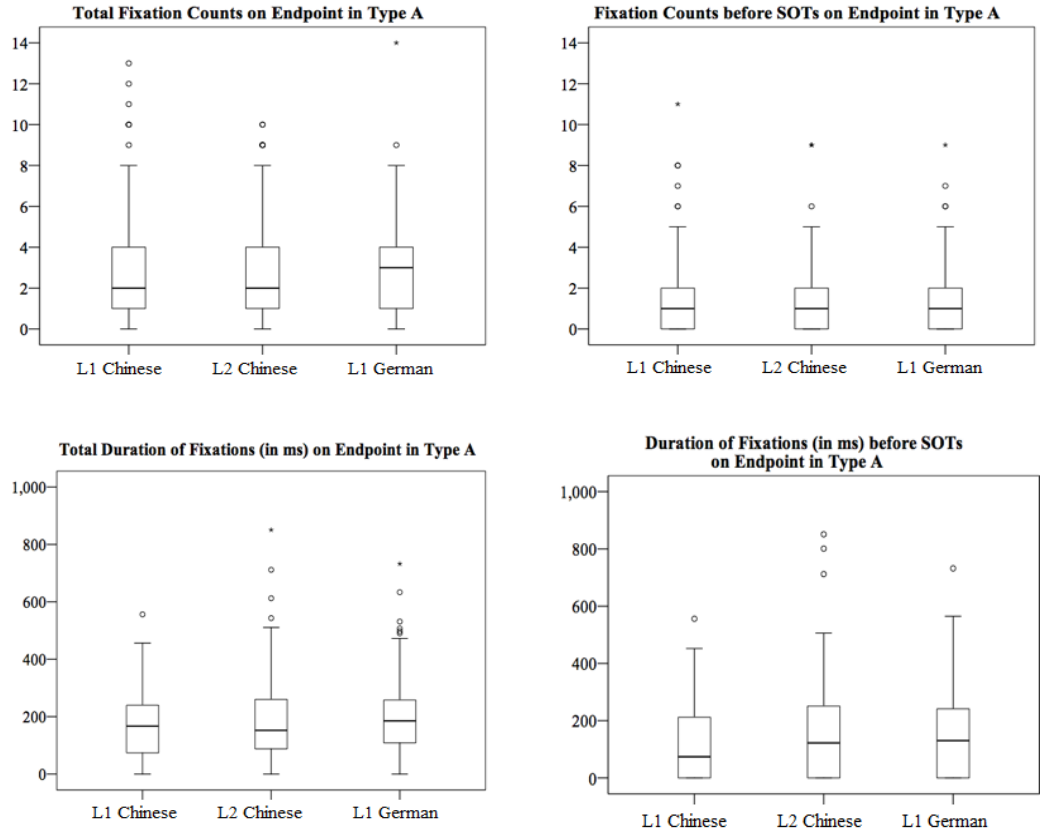


Figure 4.30: Number and duration of fixations on the endpoint in total and before SOTs by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German for Type A videos

Fixation on the moving entity (Figure)

There was no significant difference in the total number of fixations on the moving entity among all three groups in Type A videos (F (2, 569) = 0.705, p = 0.495; Dunnett T3 post-hoc tests: between L2 Chinese and L1 Chinese p = 0.593, between L2 Chinese and L1 German p = 0.888, and between L1 Chinese and L1 German p = 0.919) (see Table 4.8 and Figure 4.31). However, a significant effect was found when comparing the number of fixations before SOTs (F (2, 546) = 17.076, p < .001). L2 Chinese speakers fixated more on the moving entity before SOTs than speakers of L1 Chinese (p < .001) and L1 German (p < .001), whereas there was no significant difference between L1 Chinese and L1 German speakers (p = 0.363). Concerning the duration of the fixations on the moving entity in total and before SOTs, Dunnett T3 post-hoc tests did not show any significant differences in both regards between any two

groups (duration of fixations in total: between L2 Chinese and L1 Chinese p = 0.735, between L2 Chinese and L1 German p = 1.000, and between L1 Chinese and L1 German p = 0.593; duration of fixations before SOTs: between L2 Chinese and L1 Chinese p = 0.948, between L2 Chinese and L1 German p = 0.785, and between L1 Chinese and L1 German p = 0.993).

Table 4.8: Mean value and standard deviation of fixation counts and duration of fixations on the moving entity in total and before SOTs for Type A videos

	L1 Chinese		L2 Chinese		L1 German	
Fixation on the Moving Entity (Type A)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	7.60	3.85	8.09	4.50	7.82	3.60
Before SOTs	4.56	2.98	6.00	3.77	4.14	2.59
Duration of Fixations						
Total (in ms)	266	128	280	156	281	122
Before SOTs (in ms)	312	234	301	182	317	167

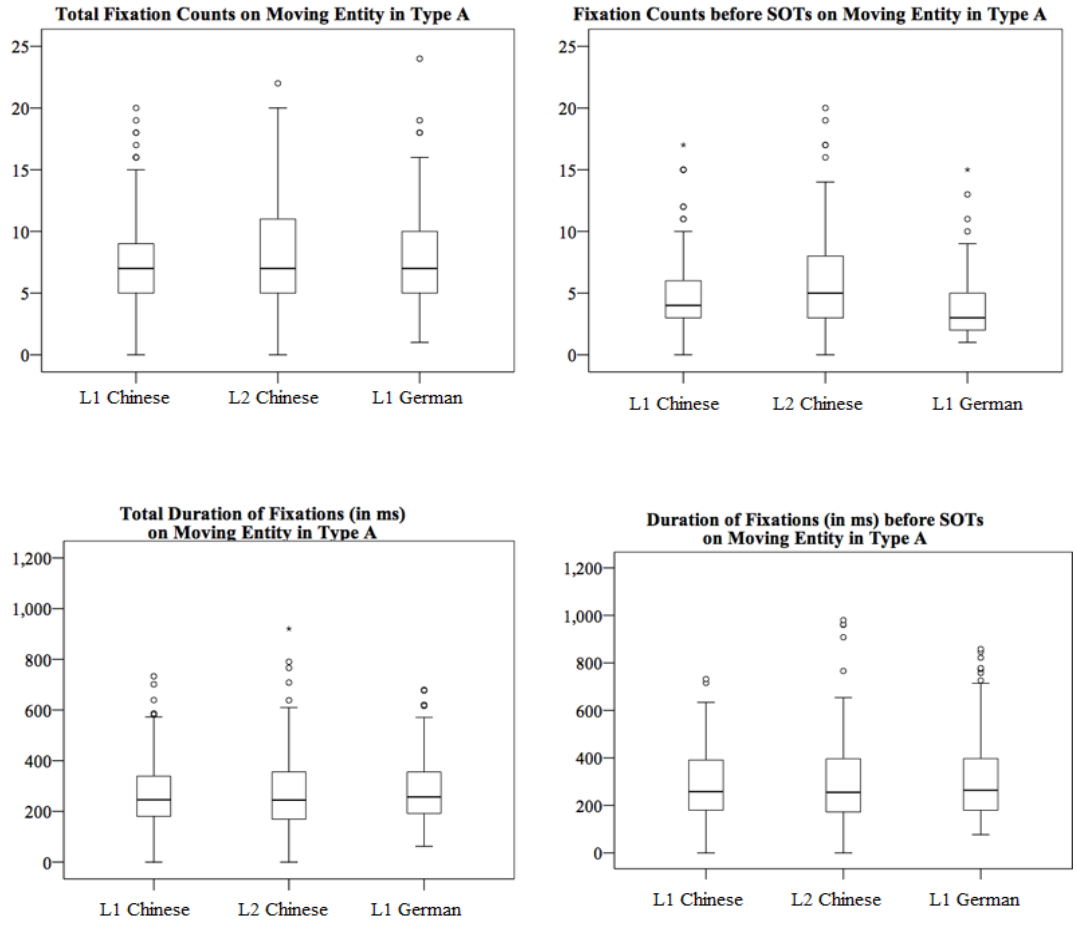


Figure 4.31: Number and duration of fixations on the moving entity in total and before SOTs by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German for Type A videos

Type B: Long trajectory with possible endpoint

Fixation on the endpoint

Table 4.9: Mean value and standard deviation of fixation counts and duration of fixations on the endpoint in total and before SOTs for Type B videos

	L1 Chinese		L2 Chinese		L1 German	
Fixation on the Endpoint (Type B)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	1.05	1.63	1.57	1.91	1.72	1.90
Before SOTs	0.48	1.15	0.95	1.41	0.88	1.31
Duration of Fixations						
Total (in ms)	94	118	122	131	143	143
Before SOTs (in ms)	57	106	86	136	109	155

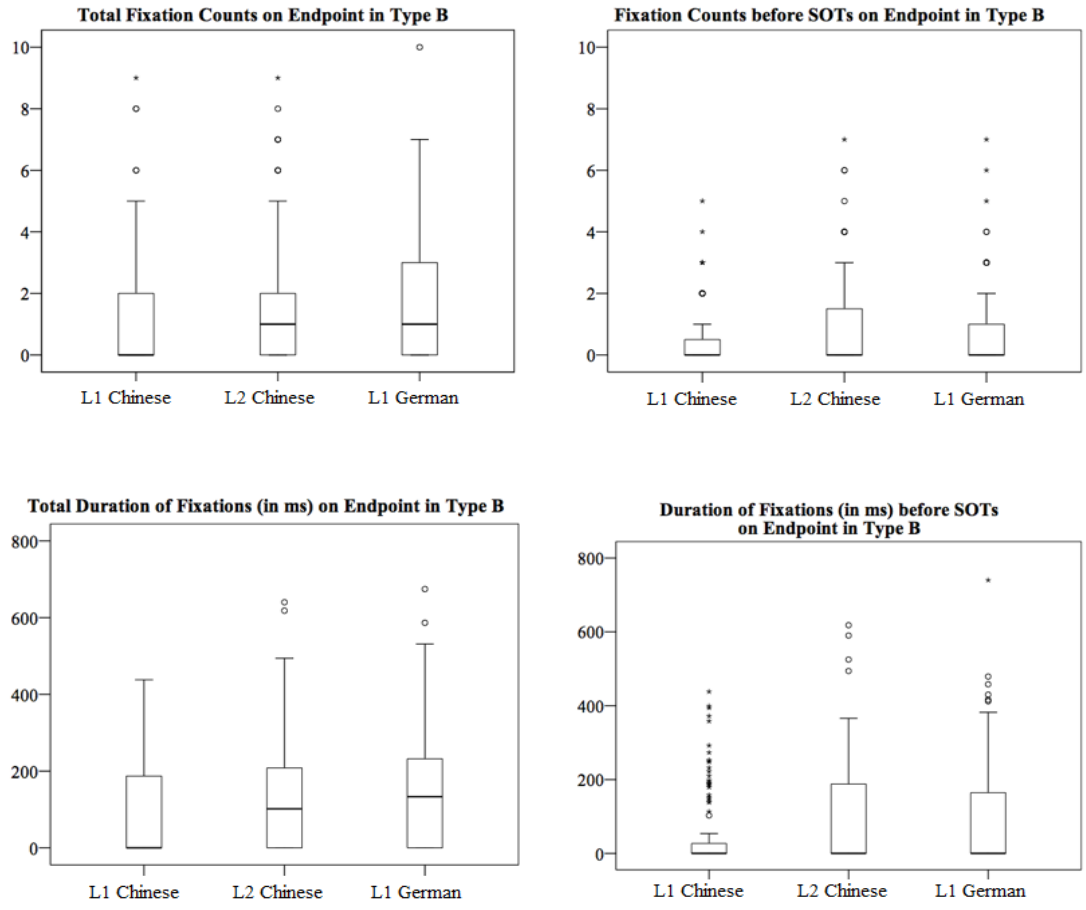


Figure 4.32: Number and duration of fixations on the endpoint in total and before SOTs by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German for Type B videos

In the verbal task for Type B videos, as shown in Table 4.9 and Figure 4.32, a significant difference in the total number of fixations on the endpoint was found among the three groups ($F(2, 517) = 6.695$, $p < 0.01$). Dunnett T3 post-hoc tests revealed that

L2 Chinese speakers (M = 1.57, SD = 1.91) fixated on the endpoint significantly more often than L1 Chinese speakers (M = 1.05, SD = 1.63) ($p < 0.05$). The same was true for L1 German speakers (M = 1.72, SD = 1.90), who had significantly more fixations than L1 Chinese speakers as well ($p < 0.05$), whereas no significant effect was found between L2 Chinese and L1 German ($p = 0.851$), although the average value of fixation counts in L1 German (M = 1.72) was higher than in L2 Chinese (M = 1.57). When comparing the fixation counts on the endpoint before SOTs, significance was found among speakers of the three groups as well ($F (2, 443) = 6.137$, $p < 0.01$). Dunnett T3 post-hoc tests showed that before SOTs, L2 Chinese speakers (M = 0.95, SD = 1.41) fixated on the endpoint significantly more often than L1 Chinese speakers (M = 0.48, SD = 1.15) ($p < 0.01$). The same was true for L1 German speakers (M = 0.88, SD = 1.31), who also fixated on the endpoint before SOTs significantly more frequently than L1 Chinese speakers ($p < 0.05$). In contrast, there was no significant difference between L2 Chinese and L1 German ($p = 0.962$). In addition, a significant difference in the total duration of fixations occurred among the three groups ($F (2, 517) = 6.163$, $p < 0.01$). Dunnett T3 post-hoc tests further showed that no significant effect was found between L1 and L2 Chinese ($p = 0.123$) and between L2 Chinese and L1 German ($p = 0.385$), although the average value of the duration of fixations in L2 Chinese was 122 ms, which was more than that in L1 Chinese (94 ms) and less than that in L1 German (143 ms) (see Table 4.9). In contrast, a significant difference was found between L1 Chinese and L1 German ($p < 0.01$). Significance was also found with respect to the duration of fixations before SOTs among the three groups ($F (2, 443) = 5.647$, $p < 0.01$). L1 German speakers fixated on the endpoint before SOTs significantly longer than L1 Chinese speakers ($p < 0.01$), while no significant relationship was documented between L2 Chinese and L1 German ($p = 0.495$). The average value of the duration of fixations before SOTs in L2 Chinese (M = 86 ms) was higher than that in L1 Chinese (M = 57 ms). Nevertheless, no significance was found between speakers of those two groups ($p = 0.086$).

Fixation on the moving entity (Figure)

Concerning the fixation on the moving entity in the verbal task for Type B videos, as shown in Table 4.10 and Figure 4.33, one-way ANOVA did not show any significant difference in the total number of fixations on the moving entity among the three groups in Type B videos ($F (2, 517) = 1.176$, $p = 0.309$) (Dunnett T3 post-hoc tests: L2 Chinese vs. L1 Chinese $p = 0.944$, L2 Chinese vs. L1 German $p = 0.308$, and

L1 Chinese vs. L1 German p = 0.686) (see Table 4.10 and Figure 4.33).

Table 4.10: Mean value and standard deviation of fixation counts and duration of fixations on the moving entity in total and before SOTs for Type B videos

	L1 Chinese		L2 Chinese		L1 German	
Fixation on the Moving Entity (Type B)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	7.48	4.29	7.70	4.00	7.07	3.47
Before SOTs	4.14	3.02	5.33	3.50	3.80	2.44
Duration of Fixations						
Total (in ms)	234	124	279	139	281	127
Before SOTs (in ms)	264	161	304	165	297	150

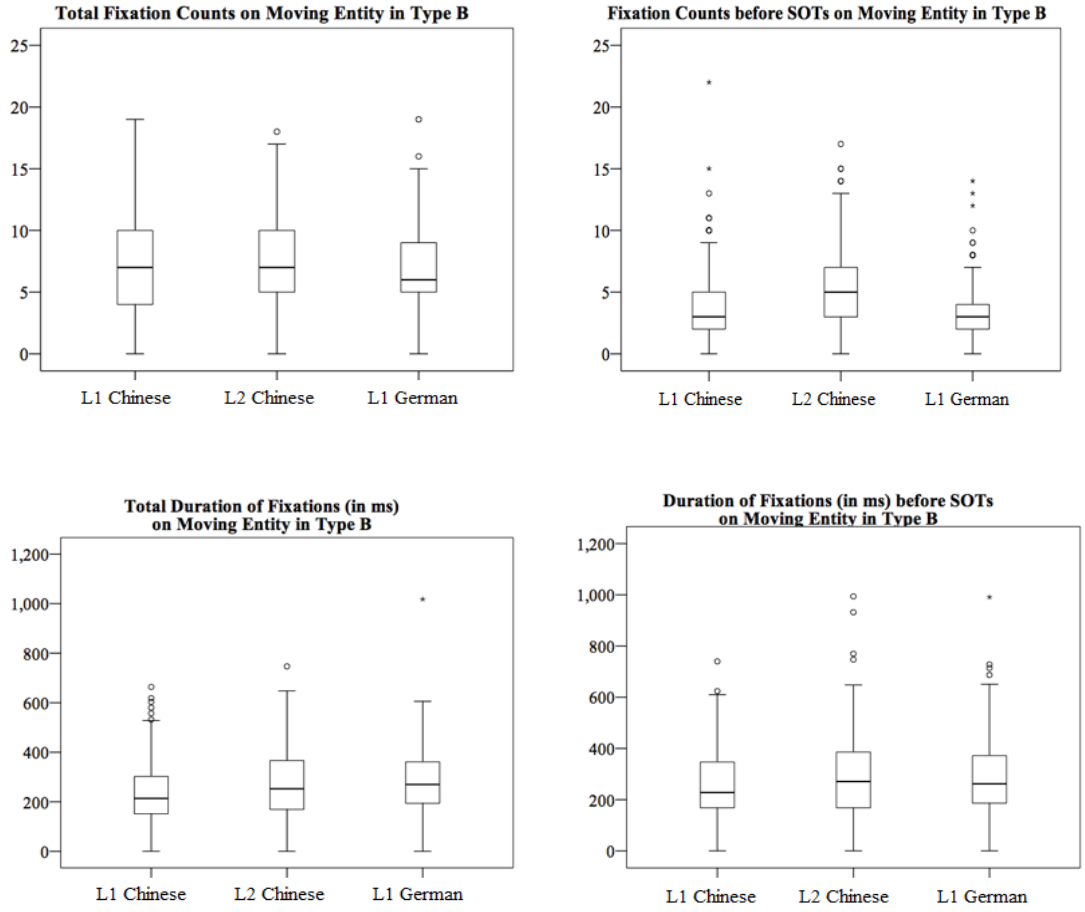


Figure 4.33: Number and duration of fixations on the moving entity in total and before SOTs by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German for Type B videos

By contrast, a significant relationship was documented when comparing the number of fixation counts before SOTs among speakers of the three groups (F (2, 483)

= 11.206, $p < 0.001$). L2 Chinese speakers fixated on the moving entity before SOTs significantly more often than L1 Chinese speakers ($p < 0.01$) and L1 German speakers ($p < 0.001$), while no significance was found between L1 Chinese and L1 German ($p = 0.586$), although the mean value in L1 Chinese ($M = 4.14$) was higher than that in L1 German ($M = 3.80$) (see Table 4.10). When considering the total duration of the fixations, a significant difference was found among the three groups ($F (2, 517) = 7.473$, $p < 0.01$). Dunnett T3 post-hoc tests revealed that speakers of L2 Chinese and L1 German fixated on the moving entity significantly longer than L1 Chinese speakers (both L2 Chinese vs. L1 Chinese and L1 German vs. L1 Chinese $p < 0.01$), while no significant effect was found between L2 Chinese and L1 German ($p = 0.999$).

Concerning the duration of fixations before SOTs, there was no significant difference among the speakers of the three groups ($F (2, 483) = 2.971$, $p = 0.052$) (Dunnett T3 post-hoc tests: L2 Chinese vs. L1 Chinese $p = 0.093$, L2 Chinese vs. L1 German $p = 0.978$, and L1 Chinese vs. L1 German $p = 0.135$).

Type C: Boundary-crossing events

Fixation on the endpoint

Note that the AoI for the endpoint and the AoI for the moving entity overlaps at some point in time as the scene unfolds in Type C videos. It is then hard to distinguish the fixation on the endpoint from the fixation on the moving entity when the two AoIs overlap. Therefore, only the fixations before the overlapping of the two AoIs were registered in this study.

Table 4.11: Mean value and standard deviation of fixation counts and duration of fixations on the endpoint in total and before SOTs for Type C videos

	L1 Chinese		L2 Chinese		L1 German	
Fixation on the Endpoint (Type C)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	1.17	1.48	0.83	1.13	1.54	1.95
Before SOTs	0.91	1.24	0.67	0.93	0.79	0.99
Duration of Fixations						
Total (in ms)	148	186	133	255	174	217
Before SOTs (in ms)	134	182	104	170	142	206

One-way ANOVA revealed a significant effect on the total number of endpoint fixation counts among the three speakers in Type C videos ($F (2, 505) = 8.511$, $p < 0.001$) (see Table 4.11 and Figure 4.34). Dunnett T3 post-hoc tests showed a significant

difference between L2 Chinese and L1 German ($p < 0.001$), whereas no significant relationship was documented between L1 and L2 Chinese ($p = 0.061$) and between L1 Chinese and L1 German ($p = 0.120$).

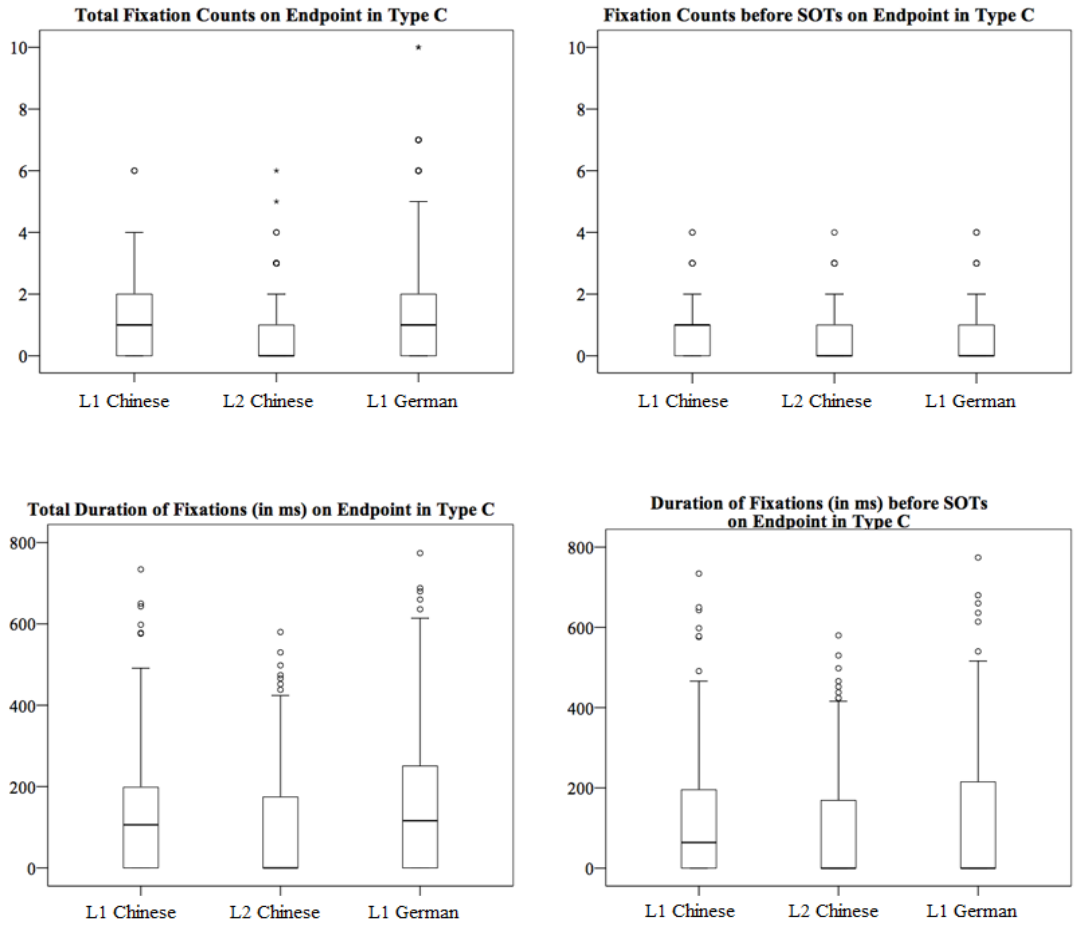


Figure 4.34: Number and duration of fixations on the endpoint in total and before SOTs by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German for Type C videos

When considering the fixation counts before SOTs, no significant effect was found among speakers of the three groups ($F (2, 505) = 2.107$, $p = 0.123$) (Dunnett T3 post-hoc tests: L2 Chinese vs. L1 Chinese $p = 0.130$, L2 Chinese vs. L1 German $p = 0.550$, and L1 Chinese vs. L1 German $p = 0.704$). Although the mean value of the duration of the fixations in L1 German ($M = 174$ ms) was higher than that in L2 Chinese ($M = 133$ ms) and L1 Chinese ($M = 148$ ms) (see Table 4.7), the statistical analysis did not show any significant effect among the three groups ($F (2, 506) = 1.524$, $p = 0.219$) (Dunnett T3 post-hoc tests: L2 Chinese vs. L1 Chinese $p = 0.912$, L2 Chinese vs. L1 German $p = 0.309$, and L1 Chinese vs. L1 German $p = 0.520$). The same was true when

comparing the duration of the fixations before SOTs among the three groups. Although its average value was higher in L1 German (M = 142) than in L2 Chinese (M = 104) and in L1 Chinese (M = 134), the statistical results did not reach a significant level (F (2, 505) = 1.808, p = 0.165) (Dunnett T3 post-hoc tests: L2 Chinese vs. L1 Chinese p = 0.340, L2 Chinese vs. L1 German p = 0.195, and L1 Chinese vs. L1 German p = 0.973).

Fixation on the moving entity (Figure)

Table 4.12: Mean value and standard deviation of fixation counts and duration of fixations on the moving entity in total and before SOTs for Type C videos

	L1 Chinese		L2 Chinese		L1 German	
Fixation on the Moving Entity (Type C)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	4.11	3.09	4.42	3.15	4.50	2.99
Before SOTs	3.38	2.31	4.09	2.91	3.44	2.34
Duration of Fixations						
Total (in ms)	236	133	266	187	236	134
Before SOTs (in ms)	245	144	271	206	236	151

As shown in Table 4.12 and Figure 4.35, one-way ANOVA revealed no significant difference in the total number of fixations on the moving entity among the three groups in Type C videos (F (2, 506) = 0.799, p = 0.451) (Dunnett T3 post-hoc tests: L2 Chinese vs. L1 Chinese p = 0.746, L2 Chinese vs. L1 German p = 0.993, and L1 Chinese vs. L1 German p = 0.531). However, when comparing the number of fixations before SOTs, a significant effect was found among speakers of the three groups (F (2, 505) = 3.971, p < 0.05). L2 Chinese speakers fixated on the moving entity before SOTs significantly more often than L1 Chinese speakers (p < 0.05), while no significant difference was found between L2 Chinese and L1 German (p = 0.076) and between L1 Chinese and L1 German (p = 0.993). Concerning the total duration of fixations, no significant difference was found among the three groups (F (2, 506) = 2.046, p = 0.130) (Dunnett T3 post-hoc tests: L2 Chinese vs. L1 Chinese p = 0.270, L2 Chinese vs. L1 German p = 0.285, and L1 Chinese vs. L1 German p = 1.000). Although the average duration of fixations before SOTs in L2 Chinese (M = 271 ms) was higher than that in L1 Chinese (M= 245 ms) and L1 German (M = 236 ms), the statistical results did not reveal any significance among the three groups (F (2, 505) = 1.933, p = 0.146) (Dunnett T3 post-hoc tests: L2 Chinese vs. L1 Chinese p = 0.474, L2 Chinese vs. L1 German p = 0.220, and L1 Chinese vs. L1 German p = 0.909).

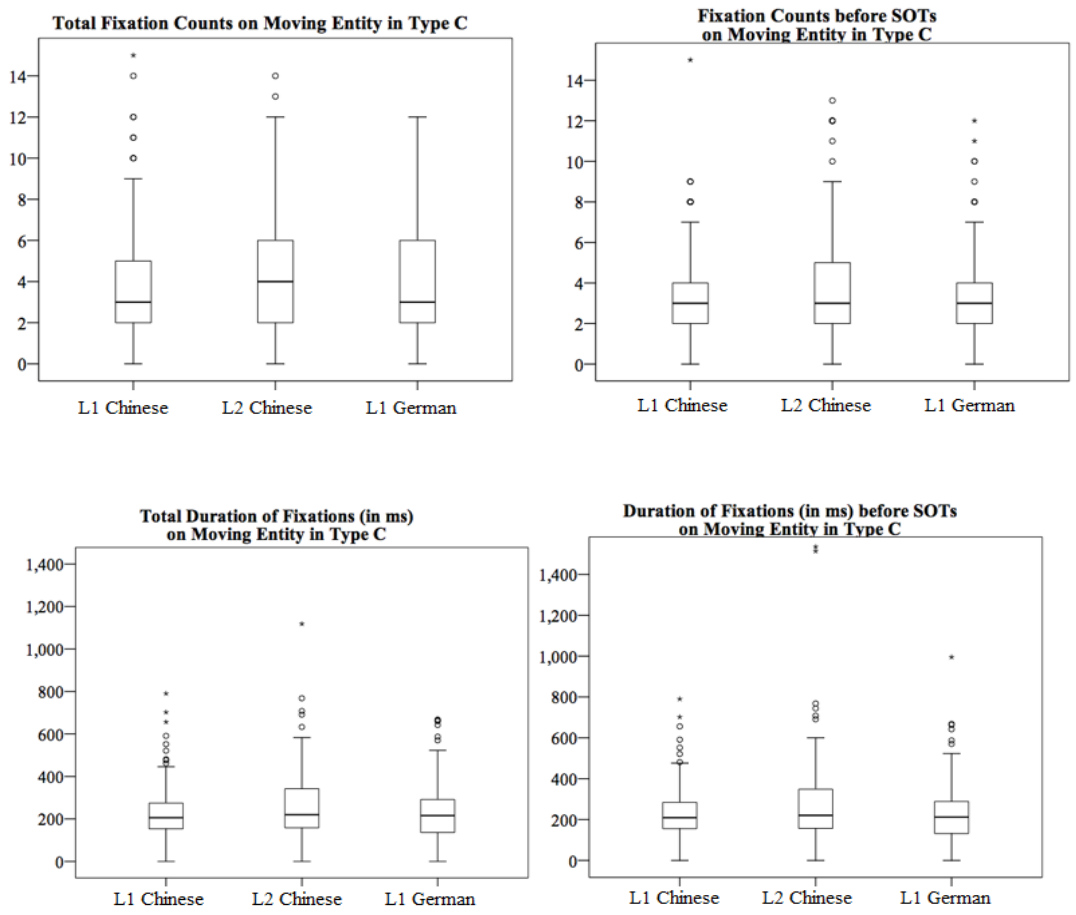


Figure 4.35: Number and duration of fixations on the moving entity in total and before SOTs by speakers of L1 Chinese, L1 German-L2 Chinese, and L1 German for Type C videos

Summary and interpretation

When comparing the fixation on the endpoint with the fixation on the moving entity (Figure) in different types of video clips, it was found that speakers of the different languages tended to have more fixations on the moving entity and these fixations were longer than fixations on the endpoint. This is because the moving entity, as its name suggests, was moving in the video clip, while the endpoint was still. So the moving entity was relatively more salient than the endpoint. On the other hand, the moving entity was typically the subject of the utterance and canonically the first element to be uttered (one cannot exclude the examples with the subject in other positions). Therefore, speakers tended to fixate more on the moving entity than on the endpoint before SOTs.

In Type A videos, the endpoint was obvious, but the video stopped playing before the moving entity actually reached the endpoint. Speakers of all languages

needed more time to determine whether the moving entity reached the endpoint or not. So the SOTs in Type A videos in all three language groups were the longest compared with the SOTs in Type B and Type C videos. The endpoint encoding at the linguistic level in the verbal task for Type A videos showed that L1 German speakers mentioned it significantly more often than L1 and L2 Chinese speakers, while L1 and L2 Chinese speakers did not differ in this regard (see Figure 4.30). The analysis of the allocation of attention in Type A videos showed that the mean value of total fixation counts and the duration on the endpoint was the highest in L1 German (total fixation count: $M = 3.05$; total duration of fixations: $M = 193$ ms) compared to L1 Chinese (total fixation count: $M = 2.82$; total duration of fixations: $M = 160$ ms), and to L2 Chinese (total fixation count: $M = 2.58$; total duration of fixations: $M = 177$ ms). However, only the difference in the total duration of fixations on the endpoint between L1 German and L1 Chinese reached a significant level (one-way ANOVA $p < 0.05$), while remarkably, speakers of all three groups did not differ in the total number of fixation counts. Following Griffin and Bock (2000), and Griffin and Spieler (2006), fixation duration reflects the processes of word retrieval. It seems that L1 German speakers took more time to ponder on how to describe the Ground object at goal from their knowledge base; after this pondering, they tended to encode it linguistically and utter it explicitly. On the other hand, the longer fixation on the endpoint in L1 German may also reflect their cognitive processes figuring out whether the moving entity was moving towards or entering into a reference object. By contrast, the total duration of fixations on the endpoint was relatively shorter in L1 and L2 Chinese. This might be attributable to the fact that L1 and L2 Chinese speakers had alternative options to conceptualize the situation in Type A videos; that is, they could choose not to pay attention to the endpoint, and rather conceptualize the situation as “The Figure is moving in a certain manner at certain place”, which is not a characteristic option in L1 German.

In comparison with the endpoint in Type A videos, the endpoint in Type B videos was far less evident and the distance between the endpoint and the moving entity was much longer. The analysis of the endpoint mentioned linguistically in the verbal task for Type B videos showed that speakers of L2 Chinese and L1 German encoded the endpoint significantly more often than L1 Chinese speakers, while speakers of L2 Chinese and L1 German did not differ in this regard (see Figure 4.30). The allocation of attention to the endpoint in Type B videos showed the same tendency: The total number of fixation counts and the number of fixation counts before SOTs in L2 Chinese

and in L1 German were significantly higher than those in L1 Chinese, while no significant difference was found between L2 Chinese and L1 German. That is, in comparison with L1 Chinese speakers, L2 Chinese speakers directed more attention to the endpoint in Type B, like L1 German speakers did. In addition, the duration of the attention to the endpoint in Type B videos showed unanimous results among all three groups: When considering the duration of the fixations in total and before SOTs, respectively, L1 German speakers showed longer fixations than L1 Chinese speakers in both cases. By contrast, no significant difference was found between L2 Chinese and L1 German and between L2 Chinese and L1 Chinese in both cases. However, the mean values of the total duration of the fixations and the duration of the fixations on the endpoint before SOTs were higher in L2 Chinese (total duration of fixations: M = 122 ms, duration of fixations before SOTs M = 86 ms) than in L1 Chinese (total duration of fixations: M = 94 ms, duration of fixations before SOTs M = 57 ms). In a nutshell, no significance was found between L2 Chinese and L1 German across all four measures (i.e., the number and duration of the fixations on the endpoint in total and before SOTs). Accordingly, L1 German-L2 Chinese speakers' allocation of attention to the endpoint conforms to that of L1 German speakers.

Particularly when considering the fixation patterns for Type B videos before SOTs, that is, during the conceptualization phase, it was found that L2 Chinese speakers directed more attention both to the endpoint and to the moving entity than L1 Chinese speakers. This might reflect that L2 speakers were "struggling" as to which information to select and how to structure this information, since the habitual conceptual patterns differ greatly in L1 Chinese and L1 German. L1 Chinese speakers tended to conceptualize the situation in Type B videos as "A Figure moving in a certain manner at a certain location". With this conceptual pattern, speakers do not need to pay attention to the endpoint. In contrast, L1 German speakers preferred conceptualizing the same situation as "A Figure moving in a certain manner along a Ground object" or "A Figure moving in a certain manner towards or to a Ground object". With this conceptual pattern, speakers may need to pay attention to the endpoint. These two conceptual patterns may compete in the process of L2 speakers' conceptualization to a certain degree. So it appears that L2 Chinese speakers pay much more attention both to the endpoint and the moving entity than L1 Chinese speakers. The different conceptual options for L2 speakers and the complexity of the language-specific structures (i.e., syntactic complexity of serial verb constructions, constraints on the combination of

semantic components, capability of co-occurrence with aspect markers) may lead to a heavier cognitive load for L2 Chinese speakers. Therefore, in comparison with L1 Chinese speakers, L2 speakers in this study had later SOTs in the verbal task for Type B videos[68].

In the verbal task for Type C videos, no significant effect was found among the three groups in terms of the endpoint being mentioned in Type C videos in the verbal task (see Figure 4.30). The eye-tracking data showed no significant difference in the number of fixation counts on the endpoint before SOTs and in the duration of the fixation on the endpoint in total and before SOTs among speakers of the three groups. A significant difference was only seen when the total number of fixation counts between L2 Chinese and L1 German speakers was compared: L2 Chinese speakers directed less attention to the endpoint than L1 German speakers. This might be attributable to the prevalent use of the deictic verb *qù* 'go' in the serial verb constructions used to describe Type C videos in L2 Chinese since *qù* in serial verb constructions could represent the endpoint implicitly when it is not uttered explicitly (see section 4.4.1.1 for more details). That is, speakers do not need to look at the endpoint when they use a serial verb construction that includes *qù* since *qù* represents the endpoint in general and can be specified according to the context.

4.4.3 Summary

In general, the relative frequency of occurrence of manner verbs was found to be higher in L1 German than in L1 and L2 Chinese. L1 Chinese speakers preferred using manner verbs in SVCs, while L2 German-Chinese speakers preferred using manner verbs alone. As for path verbs, both L1 and L2 Chinese speakers preferred using them in SVCs. In addition, the use of bare path verbs was found more often in L2 German-Chinese than in L1 Chinese. In particular, L2 speakers preferred the use of deictic verbs in different video types. It is assumed that the reason for their inclination to use deictic verbs is grounded in the directionality indicated by deictic verbs, the holistic perspective L2 speakers tended to take, the implicit endpoint that deictic verbs can represent, and the phonological and morphological status in modern Chinese (for more details, see section 4.4.1.1). In addition, L2 Chinese speakers preferred using goal-oriented path verbs to describe Type B videos and boundary-crossing path verbs

[68] I have to admit that the L2 effect is also responsible for the later SOTs by L2 speakers.

to describe in Type A and B videos, although Types A and B did not show any boundary-crossing situation. This might be related to their preference to focus on the endpoint, since in a boundary-crossing situation the endpoint is so prominent that one cannot neglect it. In comparison with L1 Chinese speakers, L2 Chinese speakers used fewer bare motion verbs. Regarding the use of SVCs, L2 Chinese speakers preferred using SVCs encoding Manner + Path + Deictic or Path + Deictic in all video types, while L1 Chinese speakers tended to use SVCs encoding Motion + Manner in Type A and B videos and Manner + Path in Type C videos. In other words, L2 Chinese speakers tended to add deictic verbs to the SVCs. We explained the reason for L2 speakers' tendency to use deictic verbs above.

The use of adjuncts was found more often in L1 German than in L1 and L2 Chinese, which is in conformance with the typological differences between S-languages and E-languages. In S-languages such as German, Path information is encoded in adjuncts, while in E-languages like Chinese, Path information is encoded in path verbs. So the use of adjuncts is found more frequently in German. When considering the spatial information encoded in adjuncts, it was found that L2 Chinese speakers tended to use adjuncts encoding goal-oriented and boundary-crossing information to describe Type A and B videos, while L1 Chinese speakers tended to use locative adjuncts for Type B videos. In comparison with L1 and L2 Chinese speakers, L1 German speakers tended to encode object-related, directional, and goal-oriented information in adjuncts, especially using object-related adjuncts to describe Type B videos. L1 and L2 Chinese speakers differed in the conceptualization patterns they applied in the description of Type B videos. L1 Chinese speakers tended to conceptualize the situation in Type B videos as "A Figure moving in a certain manner at a certain location", while L1 German speakers tended to conceptualize the same situation as "A Figure moving in a certain manner along a Ground object" or "A Figure moving in a certain manner towards or to a Ground object at goal". It is apparent that L1 German speakers' habitual conceptual patterns influenced their L2 Chinese, since L2 Chinese speakers tended to encode deictic, goal-oriented, and boundary-crossing concepts.

The majority of speakers of the three language groups encoded the different semantic components in one utterance in describing motion events. However, given the lexicalization differences and the availability of morphosyntactic constructions in different languages (Bohnemeyer et al., 2007), German and Chinese differ in their

capacity to combine different semantic components in one utterance. In German, different path segments can be stacked after a manner verb, while in Chinese, SVCs, i.e., manner verb + path verb, tend to be used in describing motion events, so path verbs in SVCs can only occur with certain path segments (depending on the context). Their use is therefore rather restricted. As far as other verbs are concerned, German and Chinese differ in the verb type chosen as predicate, which determines the event type. The event type in German is framed by manner verbs, while the event type in Chinese is framed by manner verbs and path verbs. The syntactic and semantic features of path verbs restrict the co-occurrence of other path segments. As already shown in section 4.4.1.3, the example *zài lù-shàng zǒu-guò-qù* 'lit. at road-top walk-cross-go' sounds quite unnatural in Chinese, while the expression *xiàng chēzi zǒu-guò-qù* 'lit. towards car walk-cross-go' sounds natural. Despite the low number of occurrences, it seems that some L2 Chinese speakers followed L1 German speakers' way of event formation, that is, integrating different path segments into one utterance. Some of the resulting expressions sound quite odd, as the above example shows. To solve the problem of incompatibility of different path segments in one utterance in Chinese, L2 Chinese speakers tended to use two clauses to express more information. The occurrence of two clauses was found more often in L2 Chinese than in L1 Chinese and L1 German in the verbal task for different video types. Concerning spatial information encoded in the respective two clauses, L2 Chinese speakers tended to encode directional information in the first clause and mention the endpoint in the second clause, while L1 Chinese speakers tended to encode locative information in the first clause and verbalize the endpoint in the second clause.

Speech onset times (SOTs) reflect the cognitive processes related to visual attention and to the selection and structuring of information for verbalization. In general, L2 Chinese speakers had later SOTs than speakers of L1 Chinese and L1 German. The order for SOTs from longest to shortest in L2 Chinese was Type A > Type B > Type C, while the order in L1 Chinese and L1 German was Type A > Type C > Type B. This means that, compared to L1 Chinese and L1 German, the SOTs for Type B videos were later in L2 Chinese. We argue that the late SOTs in L2 Chinese are attributable to the language-specific structures in Chinese (i.e., syntactic complexity of SVCs, restrictions on the combination of semantic components, choice of aspect markers and compatibility of SVCs with aspect markers). As previously mentioned in the analysis of spatial concept encodings, L2 Chinese speakers tended to encode deictic, goal-

oriented, and boundary-crossing information. To encode these spatial concepts, it is assumed that L2 Chinese speakers needed to pay earlier attention to the endpoint. They had to decide earlier whether an event was an ongoing event or a boundary-crossing event compared to the German speakers. In German, one can say *eine Frau läuft über die Brücke zu einem Haus* 'lit. a woman walks on the bridge towards a house'. The spatial components like *über die Brücke* 'lit. on the bridge' and *zu einem Haus* 'lit. towards a house' can all be combined with *läuft* 'lit. walk' in German. In Chinese, however, depending on the availability and visibility of endpoint, Chinese speakers had to make a earlier decision about the perspective to be taken in oder to decide the possible combination of semantic components. All these factors may lead to late SOTs by Chinese speakers.

The analysis of the endpoint mentioned above shows that in the verbal task for Type A videos, L1 German speakers mentioned more endpoints than L1 and L2 Chinese speakers, while no difference was found between L1 and L2 Chinese speakers; in the verbal task for Type B videos, both speakers of L2 Chinese and L1 German mentioned more endpoints than L1 Chinese speakers, whereas no difference was found between L2 Chinese and L1 German; in the verbal task for Type C videos, there was no difference in the frequency of mentioning the endpoint among the speakers of all three groups. This finding is in conformance with the results of the analysis of the fixation patterns in the three groups. For Type A videos, only the total duration of the fixations on the endpoint was longer in L1 German than in L1 and L2 Chinese; regarding other measures, speakers of the three groups did not differ, i.e., in terms of the total number of fixations on the endpoint, the number of fixations on the endpoint before SOTs, and the duration of the fixations on the endpoint before SOTs. According to Griffin and Bock (2000), and Griffin and Spieler (2006), the fixation duration reflects the word retrieval processes. The longer duration of the fixations on the endpoint in L1 German might reflect L1 German speakers' process of retrieving the appropriate word for the endpoint from their knowledge base. Furthermore, in Type B videos, speakers of L2 Chinese and L1 German had more fixations on the endpoint, both in total and before SOTs, while no difference was found between L2 Chinese and L1 German. In addition, the analysis of the duration of the fixations on the endpoint in total and before SOTs showed no significant difference between L1 and L2 Chinese and between L1 German and L2 Chinese in both regards, although the average value in L1 German was higher than in L2 Chinese and the average value in L2 Chinese was higher than L1

Chinese in both regards (see Table 4.9).

However, L1 German speakers had significantly longer fixations on the endpoint both during the whole language production tasks and before SOTs than L1 Chinese speakers. In other words, although no difference in the duration of the fixations on the endpoint was found between L1 and L2 Chinese during the conceptualization phase, L2 Chinese speakers had more fixations on the endpoint than L1 Chinese speakers. This is in conformance with the explanation that because L2 Chinese speakers tended to focus on the endpoint, they had later SOTs when describing Type B videos. It is also in conformance with the results related to the higher frequency with which the endpoint was mentioned verbally in the verbal task for Type B videos in L2 Chinese than in L1 Chinese. For Type C videos, L1 German speakers had more fixations on the endpoint during the whole language production tasks than L2 Chinese speakers, while no difference was found between L1 and L2 Chinese and between L1 German and L1 Chinese. Furthermore, speakers of the three groups did not differ in all other measures, i.e., the number of fixation counts before SOTs, the duration of the fixations in total, and the duration of the fixations before SOTs.

4.5 Spatial analysis in L1 German-L2 English

Like the spatial analysis in L1 German-L2 Chinese, the spatial analysis in L1 German-L2 English also centers on the comparison of verb types and adjuncts used among speakers of L1 English, L1 German-L2 English, and L1 German. The comparison will proceed as follows: First, an overview of the verb types and adjuncts used will be provided (overall across Type A, B, and C videos) in order to get a general impression of their distribution, and then detailed analyses aimed at discovering the influence of different video types upon the occurrence of verb types and adjuncts will be presented for each video type. Following that, we will compare the semantic components expressed in one or two clauses among speakers of the three groups. Finally, the speech onset times (SOTs) and fixation patterns in the three groups will be discussed.

4.5.1 Linguistic expressions used in L1 and L2 English

4.5.1.1 Verb types

Verb types are categorized as motion verbs, manner verbs, path verbs, and other

verbs in English and German. In addition, as already discussed in section 4.4.1.1, the German verb *gehen* is a manner verb with a deictic component, so it is listed as a separate category.

Generally speaking, speakers of L1 English, L1 German-L2 English, and L1 German frequently used manner verbs to describe motion events, while the use of path verbs was dramatically less frequent in all three language groups (see Figure 4.36). Statistically speaking, no significant difference was found in the use of manner verbs among the three language groups ($\chi 2$ (2, N = 1576) = 2.933, p = 0.231). In contrast, a significant difference did occur when the use of path verbs was compared among the three language groups ($\chi 2$ (2, N = 246) = 26.073, p < .001). Despite the low number of occurrences, L1 German-L2 English speakers used more path verbs than L1 English speakers ($\chi 2$ (1, N = 189) = 12.704, p < .001) and L1 German speakers ($\chi 2$ (1, N = 176) = 21.841, p < .001), while no significant effect was found between L1 English and L1 German speakers ($\chi 2$ (1, N = 127) = 1.331, p = 0.249). In addition, the use of motion verbs showed no significant difference between L2 English and L1 German speakers ($\chi 2$ (1, N = 22) = 2.909, p = 0.088), while none of the L1 English speakers used bare motion verbs in the collected data. Also, there was no significant difference in the use of other verbs among the three language groups ($\chi 2$ (1, N = 29) = 1.931, p = 0.381).

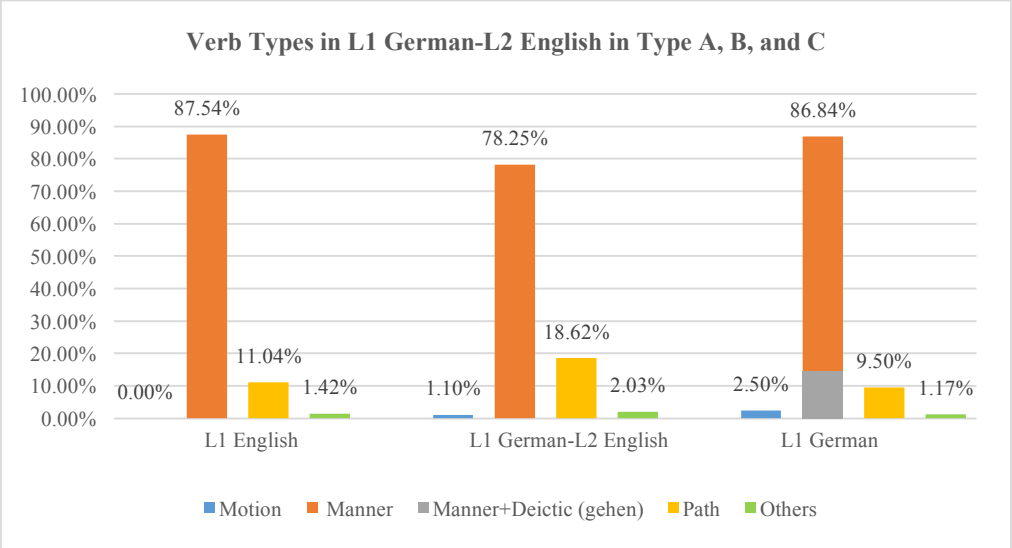


Figure 4.36: Verb types used by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos (overall)

The following examples show the various verb types used in L2 English.

(118) Use of manner verb:

A woman is walking along the sidewalk. (L1 Ger-L2 Eng P02, A01)

(119) Use of path verb:

Someone is approaching a car, his or her car probably. (L1 Ger-L2 Eng P09, A04)

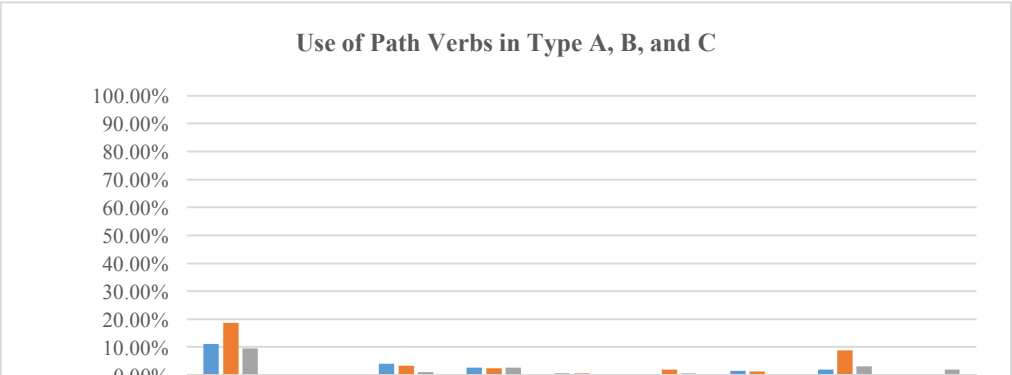
(120) Use of motion verb:

A car is moving on a rural road. (L1 Ger-L2 Eng P09, B02)

(121) Use of other verb:

A man is walking in a parking house, probably to search his car.[69] (L1 Ger-L2 Eng P09, A05)

As shown in Figure 4.37, despite the low number of occurrences, L2 English speakers used more path verbs than L1 English ($p < .001$) and L1 German speakers ($p < .001$) in the verbal tasks of Type A, B, and C videos.



	Path Verbs		Deictic	Figure	Direction	Object-related	Goal-oriented	Boundary Crossing	Axial
L1 English	11.04%		3.94%	2.68%	0.47%	0.00%	1.58%	2.05%	0.00%
L2 English	18.62%		3.44%	2.50%	0.63%	1.88%	1.25%	8.76%	0.16%
L1 German	9.50%		1.00%	2.67%	0.17%	0.67%	0.00%	3.00%	2.00%

Figure 4.37: Detailed path verbs used by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos (overall)

A detailed analysis of the path verbs used in Figure 4.37 revealed that both L1

[69] However, this expression is not appropriate, since the speaker was confused about the meaning of 'search' and 'search for'. It can be improved as 'a man is walking in a parking garage, probably to search for his car'. Thanks for the proofreader to point out this mistake.

English and L2 English speakers used significantly more deictic verbs than L1 German speakers (L1 English 3.94% vs. L1 German 1%, $\chi 2$ (1, N = 31) = 11.645, p = 0.001; L2 English 3.44% vs. L1 German 1%, $\chi 2$ (1, N = 28) = 9.143, p = 0.002), while there was no significant effect in the use of deictic verbs between L1 and L2 English speakers ($\chi 2$ (1, N = 47) = 0.191, p = 0.662). Furthermore, no significant difference was found in the use of path verbs encoding figure-related information among the three groups ($\chi 2$ (2, N = 49) = 0.041, p = 0.980). A marginal significance was found in the use of object-related path verbs between L2 English and L1 German speakers ($\chi 2$ (1, N = 16) = 4.000, p = 0.046), whereas no such data was found in this study in L1 English. The use of goal-oriented path verbs also showed no significant difference between L1 and L2 English speakers ($\chi 2$ (1, N = 18) = 0.222, p = 0.637). However, L2 English speakers used significantly more boundary-crossing path verbs than L1 English speakers ($\chi 2$ (1, N = 69) = 26.797, p < .001) and L1 German speakers ($\chi 2$ (1, N = 74) = 19.514, p < .001). Concerning the axial path verbs used, it was found that L1 German speakers used them significantly more often than L2 English speakers ($\chi 2$ (1, N = 13) = 9.308, p = 0.002), whereas no such data was found in L1 English. This is illustrated by the following examples.

(122) Use of deictic verb:

A car is <u>going</u> down a street. (L1 Ger-L2 Eng P15, B04)

(123) Path verb encoding figure-related information:

I see a woman on a crossroad and she is <u>turning</u> right on the sidewalk. (L1 Ger-L2 Eng P19, A01)

(124) Path verb encoding directional information:

The bus <u>headed</u> towards a bus stop. (L1 Ger-L2 Eng P02, A09)

(125) Path verb encoding object-related information:

A woman with a handbag is <u>crossing</u> a parking space. (L1 Ger-L2 Eng P18, A10)

(126) Path verb encoding goal-oriented information:

A man is climbing up the ladder to <u>reach</u> the first floor. (L1 Ger-L2 Eng P13, A03)

(127) Path verb encoding boundary-crossing information:

A little truck is driving down the road and entering the village. (L1 Ger-L2 Eng P07, B08)

(128) Path verb encoding axial information:

There is someone taking steps[70] to a building. (L1 Ger-L2 Eng P01, A06)

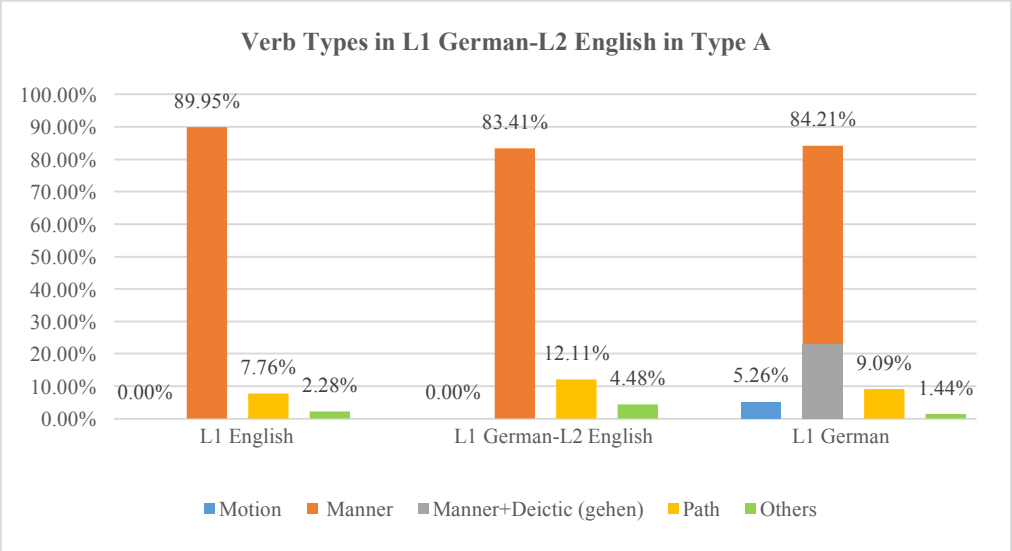


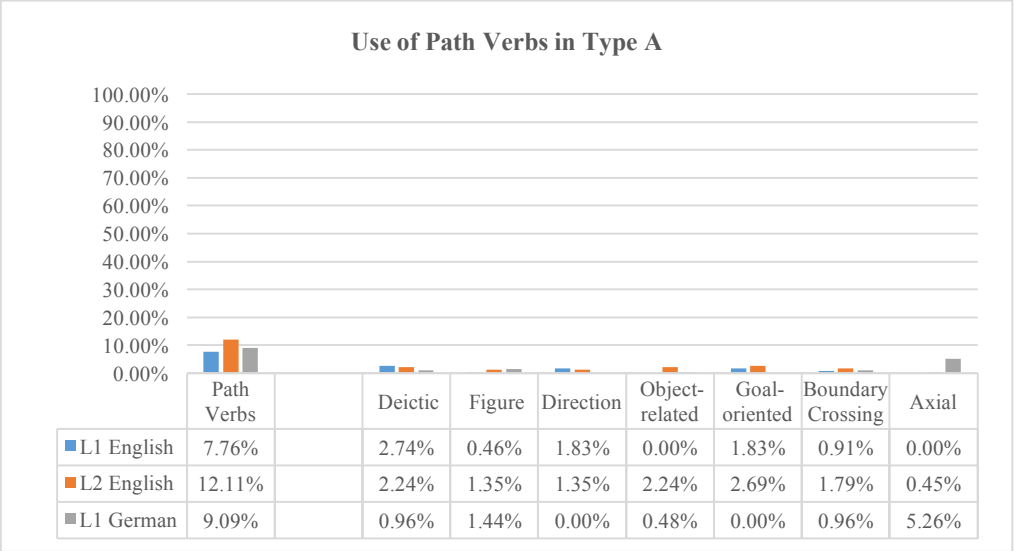
Figure 4.38: Verb types used by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A videos

In the verbal task for Type A videos, as shown in Figure 4.38, speakers of these three language groups did not differ significantly in the use of manner verbs ($\chi 2$ (2, N = 559) = 1.184, p = 0.553), nor in the use of path verbs ($\chi 2$ (2, N = 63) = 2.667, p = 0.264) and other verbs (2, N = 18) = 4.333, p = 0.115). However, L1 German speakers also used bare motion verbs (5.26%), whereas no such data was found in L1 and L2 English.

As demonstrated above, no significant difference was found in the frequency of occurrence of path verbs among the three language groups (p = 0.264). Taking a closer look at the path verbs used, as shown in Figure 4.39, it was found that L1 and L2 English speakers did not differ in the use of deictic verbs ($\chi 2$ (1, N = 11) = 0.091, p = 0.763) and goal-oriented path verbs ($\chi 2$ (1, N = 10) = 0.400, p = 0.527), respectively, both groups using more deictic verbs and goal-oriented path verbs than L1 German speakers (deictic verbs: L1 English 2.74% vs. L1 German 0.96% and L2 English 2.24% vs. L1

[70] Note that *taking a step* encodes both Manner and axial information.

German 0.96%; goal-oriented path verbs: L1 English 1.83% vs. L1 German 0% and L2 English 2.69% vs. L1 German 0%). Despite the low number of occurrences, it seems that L2 English speakers tended to use object-related path verbs (2.24%), whereas no such data was found in L1 English. In addition, L1 German speakers used significantly more axial path verbs than L2 English speakers ($\chi 2$ (1, N = 12) = 8.333, p = 0.004), whereas no use of axial path verbs by L1 English speakers was recorded in the data of this study. Also, the relative frequency of occurrence of path verbs encoding information about the figure, the direction, and crossing a boundary was rather low in all three groups, so they will not be discussed in detail.



	Path Verbs		Deictic	Figure	Direction	Object-related	Goal-oriented	Boundary Crossing	Axial
L1 English	7.76%		2.74%	0.46%	1.83%	0.00%	1.83%	0.91%	0.00%
L2 English	12.11%		2.24%	1.35%	1.35%	2.24%	2.69%	1.79%	0.45%
L1 German	9.09%		0.96%	1.44%	0.00%	0.48%	0.00%	0.96%	5.26%

Figure 4.39: Detailed path verbs used by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A videos

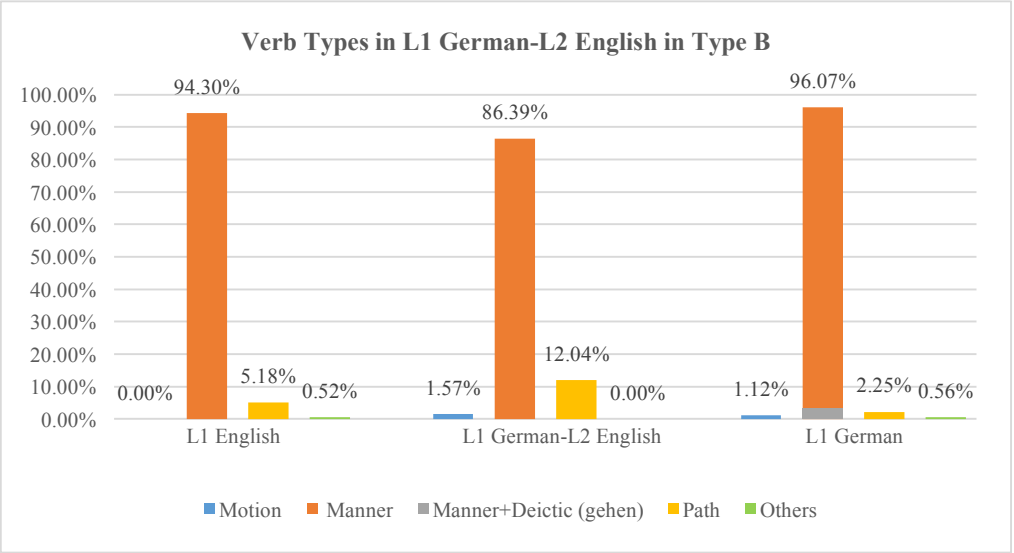
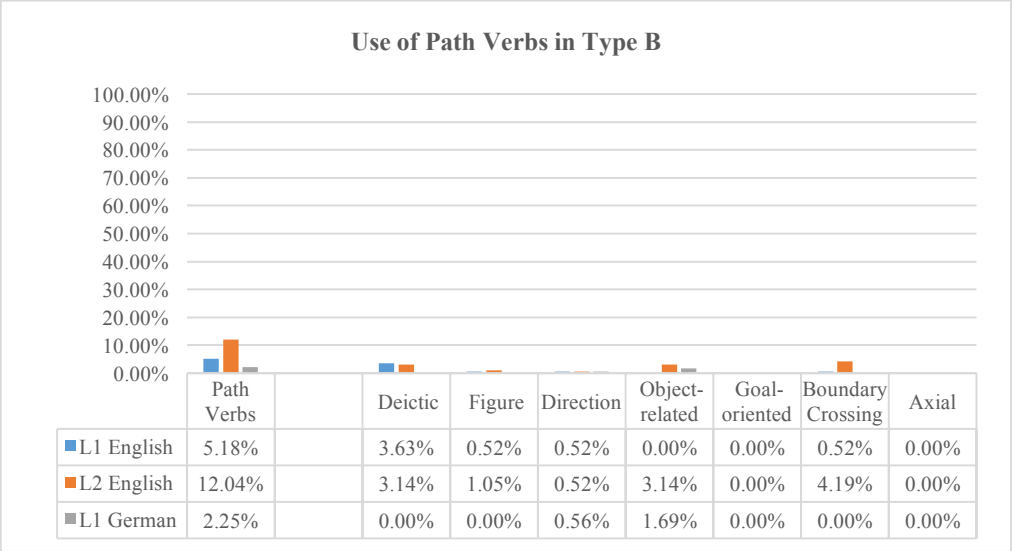


Figure 4.40: Verb types used by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type B videos

In the verbal task for Type B videos, as shown in Figure 4.40, speakers of L1 English, L1 German-L2 English, and L1 German did not differ in the use of manner verbs ($\chi 2$ (2, N = 518) = 0.861, p = 0.650). However, they differed remarkably in the use of path verbs ($\chi 2$ (2, N = 37) = 15.297, p < .001). Divergences resided in the higher occurrence of path verbs in L2 English. That is, L2 English speakers used them more often than L1 English speakers ($\chi 2$ (1, N = 33) = 5.121, p = 0.024) and L1 German speakers ($\chi 2$ (1, N = 27) = 13.370, p < .001). The detailed path verbs used are provided in Fig. 4.42. The frequency of use of motion verbs and other verbs was rather low in all three groups, so these will not be discussed any further.



	Path Verbs		Deictic	Figure	Direction	Object-related	Goal-oriented	Boundary Crossing	Axial
L1 English	5.18%		3.63%	0.52%	0.52%	0.00%	0.00%	0.52%	0.00%
L2 English	12.04%		3.14%	1.05%	0.52%	3.14%	0.00%	4.19%	0.00%
L1 German	2.25%		0.00%	0.00%	0.56%	1.69%	0.00%	0.00%	0.00%

Figure 4.41: Detailed path verbs used by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type B videos

Concerning the use of path verbs in the verbal task for Type B videos, as shown in Figure 4.41, the relative frequency of occurrence of path verbs was higher in L2 English than in L1 English (p = 0.024) and L1 German (p < .001). An in-depth look at the use of path verbs revealed that no significant relationship was documented in the use of deictic verbs between L1 and L2 English speakers ($\chi 2$ (1, N = 13) = 0.077, p = 0.782). However, the relative frequency of occurrence of object-related and boundary-crossing path verbs in L2 English was higher than in L1 English (object-related path verbs: L2 English 3.14% vs. L1 English 0%; boundary-crossing path verbs: L2 English 4.19% vs. L1 English 0.52%). L2 English speakers' inclination towards the use of boundary-crossing path verbs might reflect their focus on the endpoint, since the endpoint in a boundary-crossing event is so salient than one cannot neglect it. This

interpretation is further reinforced by the finding that L1 German-L2 Chinese speaker also tended to use boundary-crossing path verbs to describe Type B videos (see Figure 4.6). It is assumed that L1 German speakers' habitual conceptualization patterns (i.e., their tendency to take a holistic view and encode the endpoint) exerted an influence on the learners' conceptual patterns, even though they were speaking English or Chinese. Since the relative frequency of occurrence of path verbs encoding figure-related and directional information was rather rare, we will not further discuss them in full length.

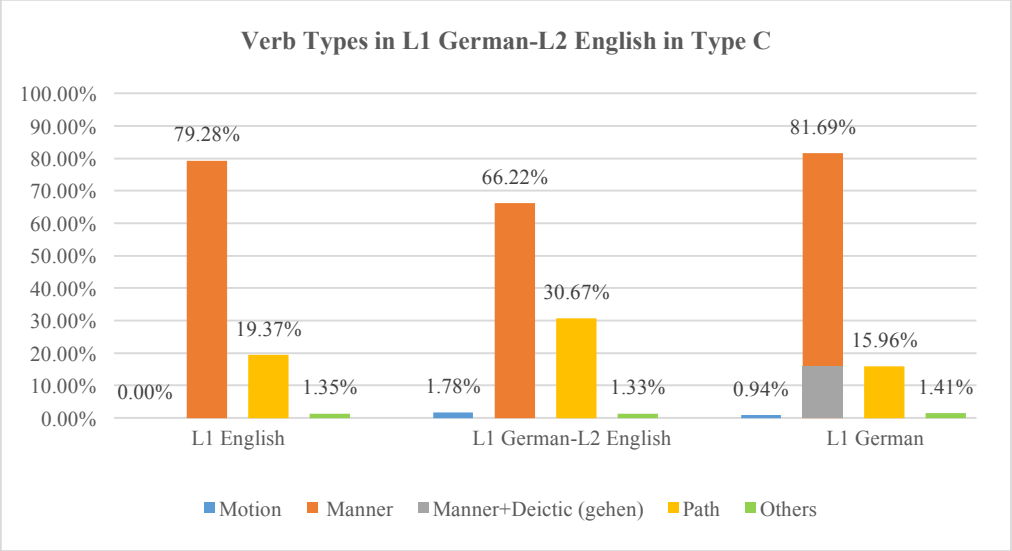
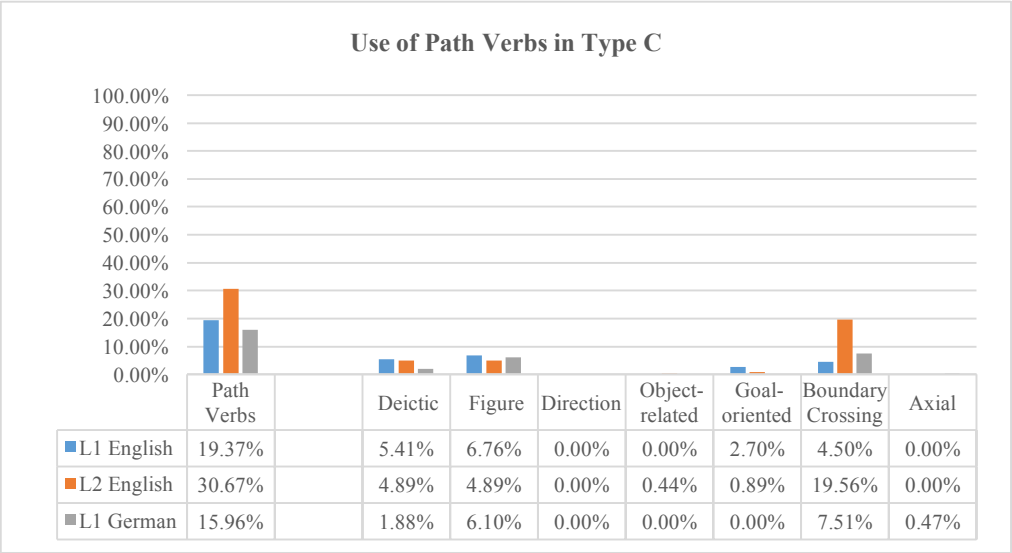


Figure 4.42: Verb types used by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type C videos

In the verbal task for Type C videos, as shown in Figure 4.42, speakers of the three language groups did not show differences in the use of manner verbs ($\chi 2$ (2, N = 499) = 2.721, p = 0.256). However, a notable difference was found in the use of path verbs by speakers of the three language groups ($\chi 2$ (2, N = 146) = 13.575, p = 0.001). This significant effect resulted from the higher numbers of occurrence of path verbs in L2 English: L2 English speakers used more path verbs than L1 English speakers ($\chi 2$ (1, N = 112) = 6.036, p = 0.014) and L1 German speakers ($\chi 2$ (1, N = 103) = 11.893, p = 0.001). Motion verbs only occurred in L2 English and in L1 German, and their occurrence was low in both languages (L2 English 1.78% and L1 German 0.94%).

As mentioned above, path verbs occurred more frequently in L2 English than in the other two groups, and these path verbs mainly encoded deictic, figure-related, and boundary-crossing information (see Figure 4.43). No significant relationship in the use of path verbs encoding deictic and figure-related information was documented among the three language groups (deictic verbs: $\chi 2$ (2, N = 27) = 4.222, p = 0.121; path

verbs encoding figure-related information: $\chi 2$ (2, N = 39) = 0.615, p = 0.735). However, the relative frequency of occurrence of boundary-crossing path verbs was higher in L2 English than in L1 English and L1 German, respectively (L2 English 19.56% vs. L1 English 4.50%, $\chi 2$ (1, N = 54) = 21.407, p < .001; L2 English 19.56% vs. L1 German 7.51%, $\chi 2$ (1, N = 60) = 13.067, p < .001).



	Path Verbs		Deictic	Figure	Direction	Object-related	Goal-oriented	Boundary Crossing	Axial
L1 English	19.37%		5.41%	6.76%	0.00%	0.00%	2.70%	4.50%	0.00%
L2 English	30.67%		4.89%	4.89%	0.00%	0.44%	0.89%	19.56%	0.00%
L1 German	15.96%		1.88%	6.10%	0.00%	0.00%	0.00%	7.51%	0.47%

Figure 4.43: Detailed path verbs used by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type C videos

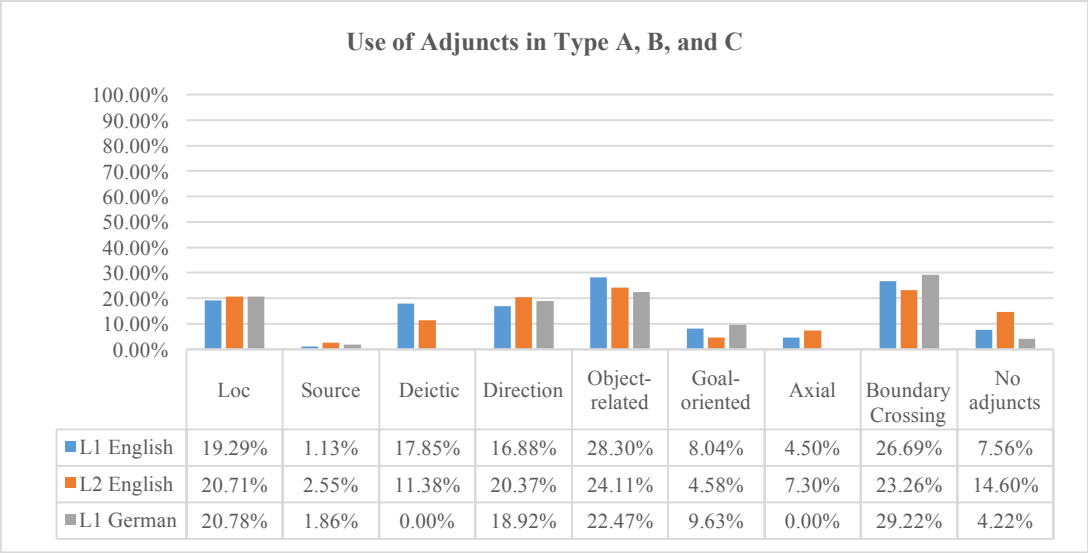
Summary

In general, speakers of the three groups did not differ in the use of manner verbs to describe different video types. However, despite the low number of occurrences, L2 English speakers used more path verbs than speakers of L1 English and L1 German. To be more specific, in the verbal task for Type A videos, L2 English speakers used more path verbs encoding deictic and goal-oriented information than L1 German speakers, while no difference was found between L1 and L2 English speakers in this regard. In the verbal task for Type B videos, L2 English speakers used more path verbs encoding object-related and boundary-crossing information than L1 English speakers. And in the verbal task for Type C videos, L2 English speakers used more boundary-crossing path verbs than speakers of L1 English and L1 German. In addition, speakers of L2 English and L1 German did not differ in the use of motion verbs, while L1 English speakers did not use any motion verbs in this study.

4.5.1.2 Adjuncts

Apart from encoding semantic components in verbs, spatial relations are also

encoded in adjuncts. Adjuncts can encode information about location, source, deictic, direction, object, goal, axial and boundary-crossing. In the following, we will present the use of adjuncts in L1 German-L2 English, as compared with L1 English and L1 German. The usage percentages of adjuncts in each language group were obtained through the following formula: number of specific adjuncts divided by total number of utterances in this language. Since one utterance can contain two or more adjuncts, the total percentages might slightly surpass 100%.



	Loc	Source	Deictic	Direction	Object-related	Goal-oriented	Axial	Boundary Crossing	No adjuncts
L1 English	19.29%	1.13%	17.85%	16.88%	28.30%	8.04%	4.50%	26.69%	7.56%
L2 English	20.71%	2.55%	11.38%	20.37%	24.11%	4.58%	7.30%	23.26%	14.60%
L1 German	20.78%	1.86%	0.00%	18.92%	22.47%	9.63%	0.00%	29.22%	4.22%

Figure 4.44: Use of adjuncts by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos

As shown in Figure 4.44, in general, no significant relationship was documented in encoding information about location ($\chi 2$ (2, N = 365) = 0.038, p = 0.981), source ($\chi 2$ (2, N = 33) = 2.909, p = 0.234), direction ($\chi 2$ (2, N = 358) = 3.291, p = 0.193), and boundary-crossing ($\chi 2$ (2, N = 476) = 4.592, p = 0.101) among the three language groups. However, a significant effect was found in encoding object-related ($\chi 2$ (2, N = 451) = 6.843, p = 0.033) and goal-oriented information ($\chi 2$ (2, N = 134) = 11.030, p = 0.004) among speakers of all three language groups. Concerning the object-related concepts encoded, there was no significant difference between L1 English and L2 English speakers nor between L2 English speakers and L1 German speakers (L1 English 28.30% vs. L2 English 24.11%, $\chi 2$ (1, N = 318) = 3.635, p = 0.057; L1 German 22.47% vs. L2 English 24.11% $\chi 2$ (1, N = 275) = 0.295, p = 0.587). However, statistical significance was found between L1 English and L1 German speakers: object-related concept encodings were found more often in L1 English than in L1 German ($\chi 2$ (1, N

= 309) = 5.984, p = 0.014). This is because in addition to encoding object-related information in adjuncts, German speakers also encode the same information in verb particles (see Figure 4.17). Regarding the encoding of goal-oriented concepts, it was found that speakers of L1 English and L1 German encoded them significantly more often than L2 English speakers (L1 English 8.04% vs. L2 English 4.58%, $\chi 2$ (1, N = 77) = 6.870, p = 0.009; L1 German 9.63% vs. L2 English 4.58%, $\chi 2$ (1, N = 84) = 10.714, p = 0.001), while no significant difference was found between L1 English and L1 German speakers ($\chi 2$ (1, N = 107) = 0.458, p = 0.499). In addition, L1 English speakers used significantly more deictic adjuncts than L2 English speakers ($\chi 2$ (1, N = 178) = 10.876, p = 0.001), while there was no significant difference in the use of axial adjuncts between L1 and L2 English speakers ($\chi 2$ (1, N = 71) = 3.169, p = 0.075). When considering the frequency of occurrence of utterances without adjuncts, it was shown that their occurrence was more frequent in L2 English than in L1 English ($\chi 2$ (1, N = 133) = 11.436, p = 0.001) and in L1 German ($\chi 2$ (1, N = 111) = 33.523, p < .001). The reason might lie in the fact that L2 English speakers used relatively more path verbs than speakers of L1 English and L1 German (see Figure 4.38). Among these path verbs, L2 English speakers used more boundary-crossing path verbs than the other two groups. When the boundary-crossing path verb *enter* is used, additional adjuncts might not be required (see example 137).

(129) Adjuncts encoding locative information

A woman is walking <u>on</u> the pavement. (L1 Ger-L2 Eng P11, A01)

(130) Adjuncts encoding source information

A man on the bike is turning into a court <u>from</u> the street. (L1 Ger-L2 Eng P15, C09)

(131) Adjuncts encoding deictic information

A car is driving <u>down</u> the street. (L1 Ger-L2 Eng P04, B04)

(132) Adjuncts encoding information about direction

A woman is walking <u>towards</u> garbage cans or garbage containers. (L1 Ger-L2 Eng P20, A10)

(133) Adjuncts encoding object-related information

A car is driving along the street. (L1 Ger-L2 Eng P10, B02)

(134) Adjuncts encoding goal-oriented information

A woman is walking on a road to a car. (L1 Ger-L2 Eng P17, A04)

(135) Adjuncts encoding axial information

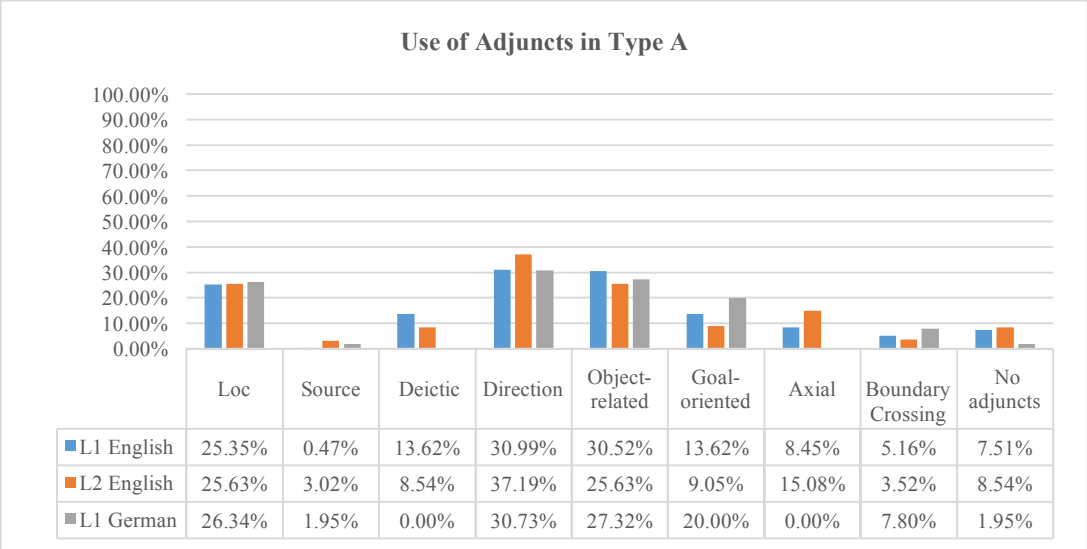
A guy is climbing up the ladder. (L1 Ger-L2 Eng P02, A03)

(136) Adjuncts encoding boundary-crossing information

A car is driving into garage. (L1 Ger-L2 Eng P19, C01)

(137) No adjuncts used

A car is entering the garage. (L1 Ger-L2 Eng P02, C01)

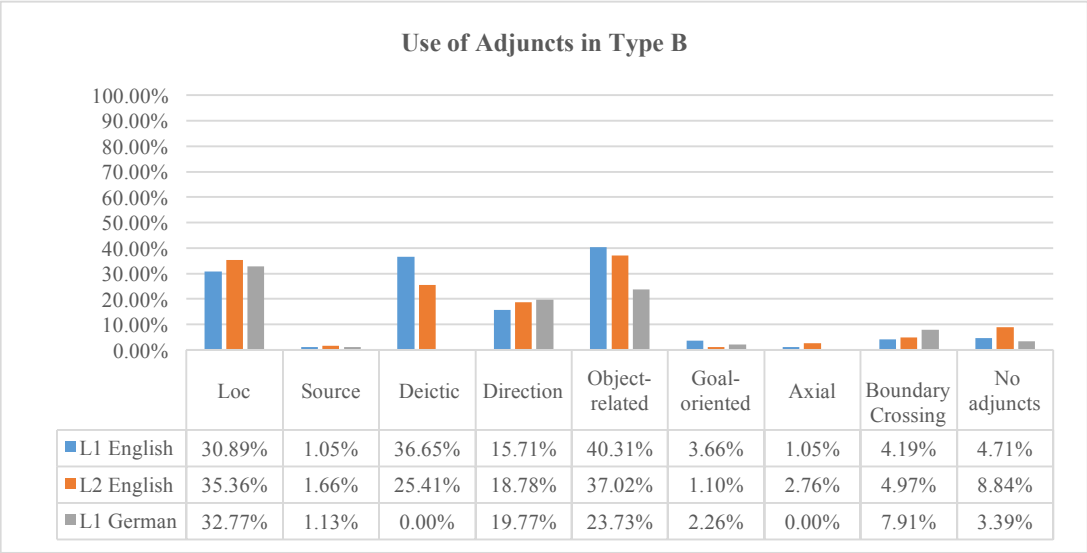


	Loc	Source	Deictic	Direction	Object-related	Goal-oriented	Axial	Boundary Crossing	No adjuncts
L1 English	25.35%	0.47%	13.62%	30.99%	30.52%	13.62%	8.45%	5.16%	7.51%
L2 English	25.63%	3.02%	8.54%	37.19%	25.63%	9.05%	15.08%	3.52%	8.54%
L1 German	26.34%	1.95%	0.00%	30.73%	27.32%	20.00%	0.00%	7.80%	1.95%

Figure 4.45: Use of adjuncts by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A videos

In the verbal task for Type A videos, as shown in Figure 4.45, statistical analysis showed no significant difference in the use of adjuncts encoding information on location ($\chi 2$ (2, N = 159) = 0.113, p = 0.945), direction ($\chi 2$ (2, N = 203) = 0.956, p = 0.620), object ($\chi 2$ (2, N = 172) = 1.756, p = 0.416), and boundary-crossing ($\chi 2$ (2, N = 34) = 3.588, p = 0.166) among speakers of the three language groups. In addition, no significant effect was found in the use of deictic ($\chi 2$ (1, N = 46) = 3.130, p = 0.077) and

axial adjuncts (χ2 (1, N = 48) = 3.000, p = 0.083), respectively, between L1 and L2 English speakers. Concerning the use of goal-oriented adjuncts, it was found that L2 English speakers used them significantly less often than L1 German speakers (χ2 (1, N = 59) = 8.966, p = 0.003), while no significance was found between L1 English and L2 English speakers (χ2 (1, N = 47) = 2.574, p = 0.109) nor between L1 English and L1 German speakers (χ2 (1, N = 70) = 2.057, p = 0.151). Regarding the occurrence of utterances without adjuncts, the statistical results showed that no significant difference was found between L1 and L2 English speakers (χ2 (1, N = 33) = 0.030, p = 0.862), while the frequency of its occurrence in L2 English was significantly higher than in L1 German (χ2 (1, N = 21) = 8.048, p = 0.005).

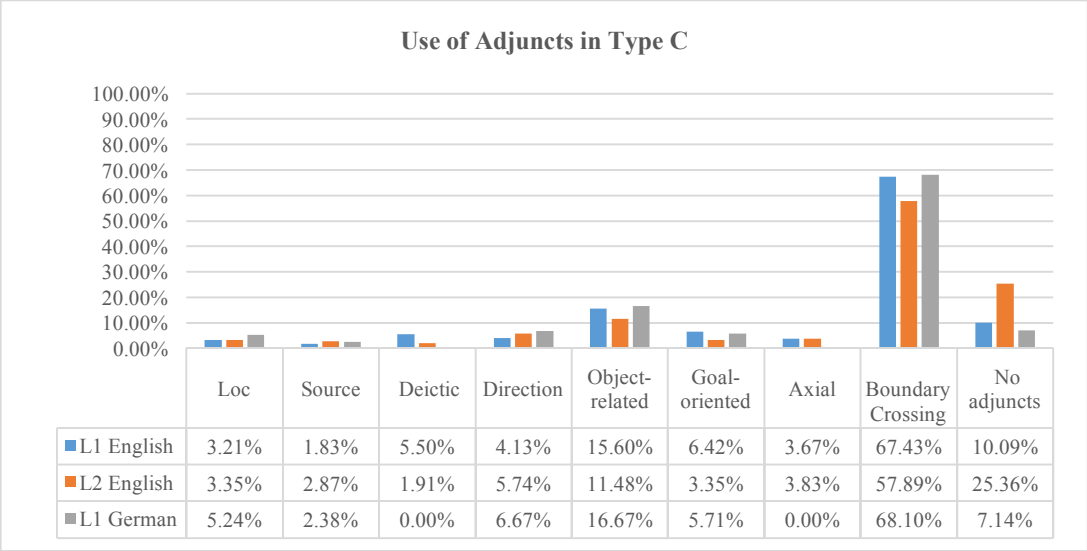


	Loc	Source	Deictic	Direction	Object-related	Goal-oriented	Axial	Boundary Crossing	No adjuncts
L1 English	30.89%	1.05%	36.65%	15.71%	40.31%	3.66%	1.05%	4.19%	4.71%
L2 English	35.36%	1.66%	25.41%	18.78%	37.02%	1.10%	2.76%	4.97%	8.84%
L1 German	32.77%	1.13%	0.00%	19.77%	23.73%	2.26%	0.00%	7.91%	3.39%

Figure 4.46: Use of adjuncts by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type B videos

In the verbal task for Type B videos, as shown in Figure 4.46, no statistical significance was found in the use of adjuncts encoding information on location (χ2 (2, N = 181) = 0.343, p = 0.843), direction (χ2 (2, N = 99) = 0.424, p = 0.809), and boundary-crossing (χ2 (2, N = 31) = 2.000, p = 0.368). Furthermore, no significant difference was found in the use of deictic adjuncts between L1 and L2 English speakers (χ2 (1, N = 116) = 4.966, p = 0.026). Concerning the use of object-related adjuncts, the statistical results showed that L2 English speakers used them significantly more often than L1 German speakers (χ2 (1, N = 109) = 5.734, p = 0.017), while no significant difference was found between L1 English and L2 English speakers (χ2 (1, N = 144) = 0.694, p = 0.405). The reason for the lesser use of object-related adjuncts in L1 German

is grounded in the fact that German speakers additionally encode object-related information in particles (see Figure 4.17). Concerning the occurrence of utterances without adjuncts, no significant difference was found between L1 and L2 English speakers ($\chi 2$ (1, N = 25) = 1.960, p = 0.162). However, its occurrence was significantly higher in L2 English than in L1 German ($\chi 2$ (1, N = 22) = 4.545, p = 0.033).



	Loc	Source	Deictic	Direction	Object-related	Goal-oriented	Axial	Boundary Crossing	No adjuncts
L1 English	3.21%	1.83%	5.50%	4.13%	15.60%	6.42%	3.67%	67.43%	10.09%
L2 English	3.35%	2.87%	1.91%	5.74%	11.48%	3.35%	3.83%	57.89%	25.36%
L1 German	5.24%	2.38%	0.00%	6.67%	16.67%	5.71%	0.00%	68.10%	7.14%

Figure 4.47: Use of adjuncts by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type C videos

Type C videos showed boundary-crossing events. Accordingly, the use of adjuncts encoding boundary-crossing information occurred most frequently in the three language groups, and no statistical significance was found in its occurrence among the three groups ($\chi 2$ (2, N = 411) = 2.861, p = 0.239) (see Figure 4.47). This statistical result is in conformance with the typological features of S-languages, to which both English and German belong. Path information, especially boundary-crossing information in Type C videos, was typically encoded in adjuncts. In addition, no significant relationship was documented in the occurrence of adjuncts encoding information about location ($\chi 2$ (2, N = 25) = 1.280, p = 0.527), direction ($\chi 2$ (2, N = 35) = 1.086, p = 0.581), object ($\chi 2$ (2, N = 93) = 2.387, p = 0.303), and goal ($\chi 2$ (2, N = 33) = 2.364, p = 0.307) among the three language groups. However, marginal significance was found in the use of deictic adjuncts between L1 and L2 English speakers ($\chi 2$ (1, N = 16) = 4.000, p = 0.046), while no significant difference occurred in the use of axial adjuncts ($\chi 2$ (1, N = 16) = .000, p = 1.000) in L1 and L2 English. Regarding the occurrence of utterances without adjuncts, the statistical results showed that this occurred more frequently in L2 English than in L1 English ($\chi 2$ (1, N = 75) = 12.813, p

< .001) and in L1 German ($\chi 2$ (1, N = 68) = 21.235, p < .001). This is because L2 English speakers used more boundary-crossing path verbs to describe Type C videos (see Figure 4.44), so additional adjuncts expressing the boundary-crossing situation might not have been required.

Summary

In general, L1 and L2 English speakers did not differ in the use of adjuncts encoding various spatial information in Type A and Type B videos. However, L2 English speakers used less goal-oriented adjuncts to describe Type A videos and more object-related adjuncts to describe Type B videos than L1 German speakers did. The finding that the use of object-related adjuncts occurred more often in L2 English than in L1 German in the verbal task for Type B videos is related to the fact that L1 German speakers also used verb particles to encode the same information (see Figure 4.17). In the verbal task for Type C videos, L1 English speakers used more deictic adjuncts than L2 English speakers, while there was no difference in the use of other adjuncts between L1 and L2 English speakers. In addition, in comparison with L1 English and L1 German, no adjuncts occurred more often in L2 English in the verbal task for Type C videos (see example 137). This is because L2 English speakers used the boundary-crossing path verb *enter* more frequently than the other two groups. When the path verb *enter* is used, an additional adjunct encoding the same information may not be necessary.

4.5.1.3 Event unit formation: combination and segmentation

Previous studies have shown that various path components can be stacked after manner verbs in S-languages such as English or German (e.g. Carroll et al., 2012; Hickmann & Hendriks, 2010; Slobin, 1996). In contrast, there are restrictions on the combinations of different semantic components in Chinese (see section 4.4.1.3). It is thus hypothesized that speakers of L1 and L2 English and L1 German tend to produce one utterance to encode different spatial concepts and that there is no difference in the frequency of occurrence of different semantic components being combined in one utterance or divided into two utterances among speakers of these three groups.

One-clause pattern

When describing motion events, speakers of L1 English, L1 German-L2 English, and L1 German preferred encoding different semantic components in one utterance. When comparing the frequency of occurrence of one clause in each language group (overall across Type A, B, and C videos), it was found that it occurred at a rate

of 95.29% in L1 English, 93.87% in L1 German-L2 English, and 95.01% in L1 German. The statistical results did not show any significant difference among the three language groups ($\chi 2$ (2, N = 1620) = 2.011, p = 0.366). In the following, the most typical combinations of semantic components within one utterance among speakers of the three language groups will be presented.

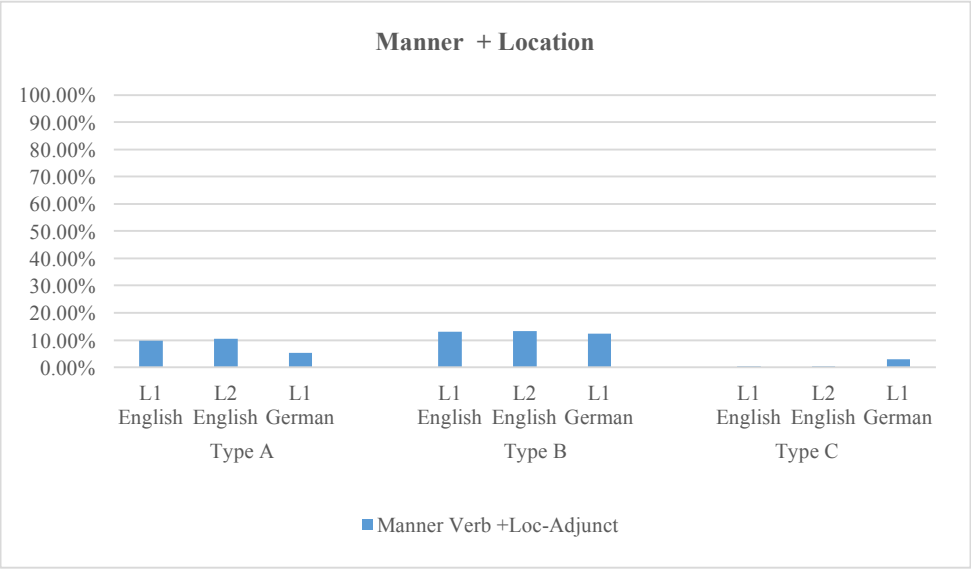


Figure 4.48: Encoding of Manner + Location by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos

As shown in Figure 4.48, the combination of the semantic components Manner and Location is realized in the construction "manner verb + locative adjunct" in English and German. No significant relationship was documented in the use of this construction in the verbal task for Type A videos ($\chi 2$ (2, N = 53) = 3.774, p = 0.152) and Type B videos ($\chi 2$ (2, N = 71) = 0.197, p = 0.906) among the three language groups. In the verbal task for Type C videos, the occurrence of this construction was rather low in all three groups (L1 English 0.46%, L2 English 0.48%, and L1 German 2.86%).

Figure 4.49 shows the combination of the semantic components Manner + Object is realized with the construction "manner verb + object adjunct" in English and German. In German, it can additionally be realized with the constructions "manner verb + object-particle" or "manner verb + object adjunct + object particle". The occurrence of these constructions showed no significant difference in the verbal task for Type A videos ($\chi 2$ (2, N = 105) = 3.257, p = 0.196) and Type C videos ($\chi 2$ (2, N = 37) = 0.703, p = 0.704) among speakers of the three language groups, while a significant difference was found in the verbal task for Type B videos ($\chi 2$ (2, N = 161) = 13.950, p = 0.001).

L1 German speakers used these constructions significantly more often than L1 English speakers ($\chi 2$ (1, N = 119) = 9.151, p = 0.002) and L2 English speakers ($\chi 2$ (1, N = 118) = 9.797, p = 0.002), while no significant difference was found between L1 and L2 English speakers ($\chi 2$ (1, N = 85) = 0.012, p = 0.914).

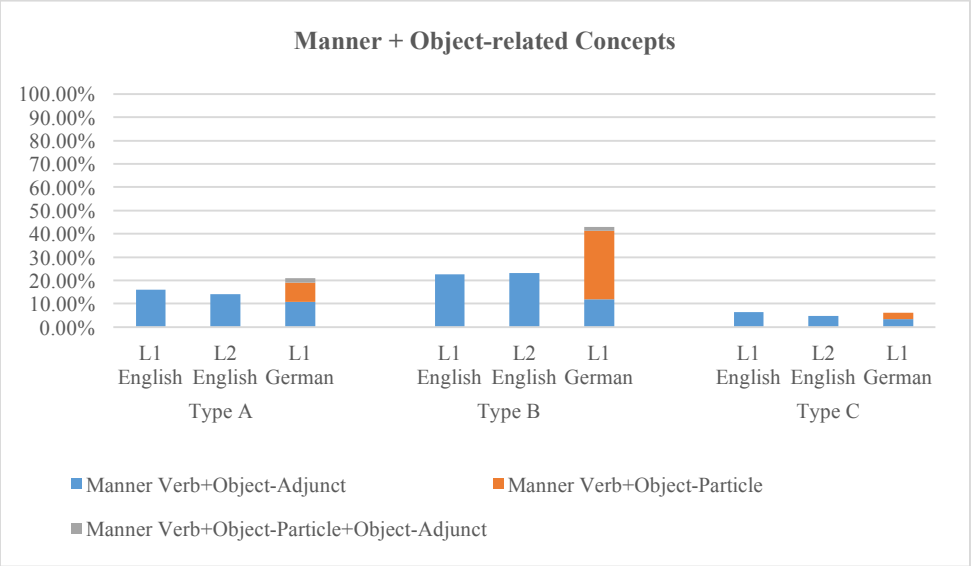


Figure 4.49: Encoding of Manner + Object-related concepts by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos

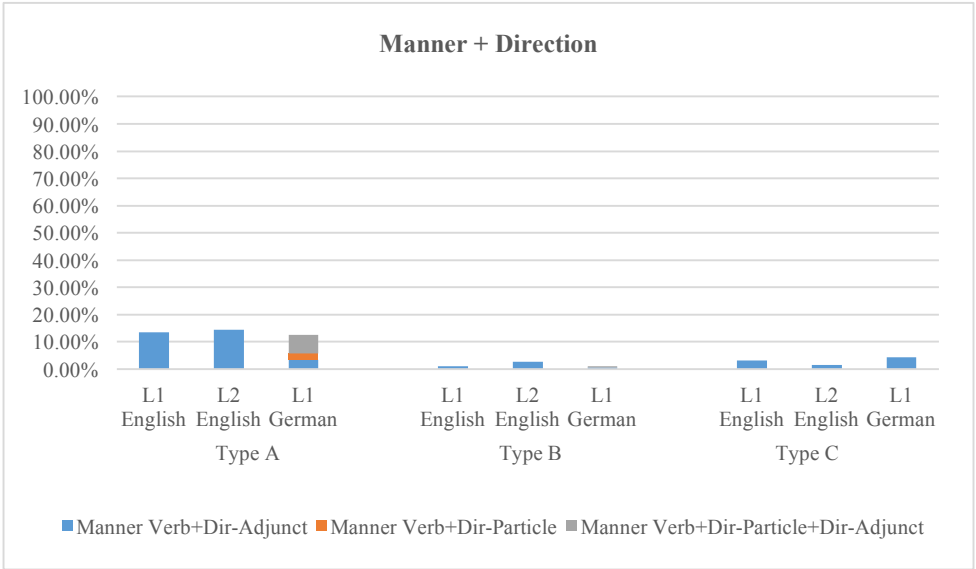


Figure 4.50: Encoding of Manner + Direction by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos

The spatial concept of direction can be realized as a directional particle or as an adjunct in German. As shown in Figure 4.50, there are three constructions that can be

found in the collected data in English and German: "manner verb + directional adjunct", "manner verb + directional particle", and "manner verb + directional adjunct + directional particle". No significant difference was found in the use of these constructions in the verbal task for Type A videos ($\chi 2$ (2, N = 84) = 0.214, p = 0.898) and Type C videos ($\chi 2$ (2, N = 19) = 2.947, p = 0.229) among the three language groups. In the verbal task for Type B videos, the overall occurrence was rather low in all three groups (L1 English 1.05%, L2 English 2.76%, and L1 German 0.56%).

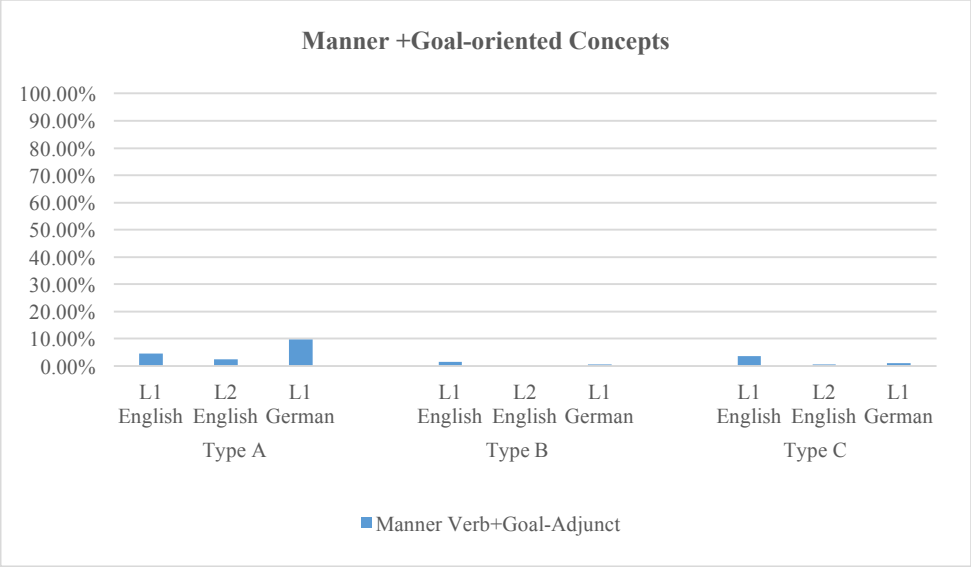


Figure 4.51: Encoding of Manner + Goal-oriented concepts by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos

In the verbal task for Type A videos, as shown in Figure 4.51, L1 German speakers encoded the combination of the semantic components Manner + Goal-oriented more often than L2 English speakers ($\chi 2$ (1, N = 25) = 9.000, p = 0.003), while no significant difference was found between L1 English and L2 English speakers ($\chi 2$ (1, N = 15) = 1.667, p = 0.197). In the verbal task for Type B and Type C videos, the occurrence of manner verb + goal-oriented adjunct was rather low in all three groups (Type B: L1 English 1.57%, L2 English 0%, and L1 German 0.56%; Type C: L1 English 3.67%, L2 English 0.48%, and L1 German 0.95%).

The spatial concepts of Manner and Deictic in parenthesis, as shown in Figure 4.52, indicate that they can be omitted in this combination. In the verbal task for Type A and Type C videos, no significant difference was documented in encoding the combination of (Manner) + Boundary-Crossing + (Deictic) in the verbal task for Type A videos ($\chi 2$ (2, N = 21) = 2.000, p = 0.368) and Type C videos ($\chi 2$ (2, N = 332) =

1.614, p = 0.446) among speakers of the three language groups. In the verbal task for Type B videos, its occurrence was rather low in all three language groups, so this will not be discussed any further.

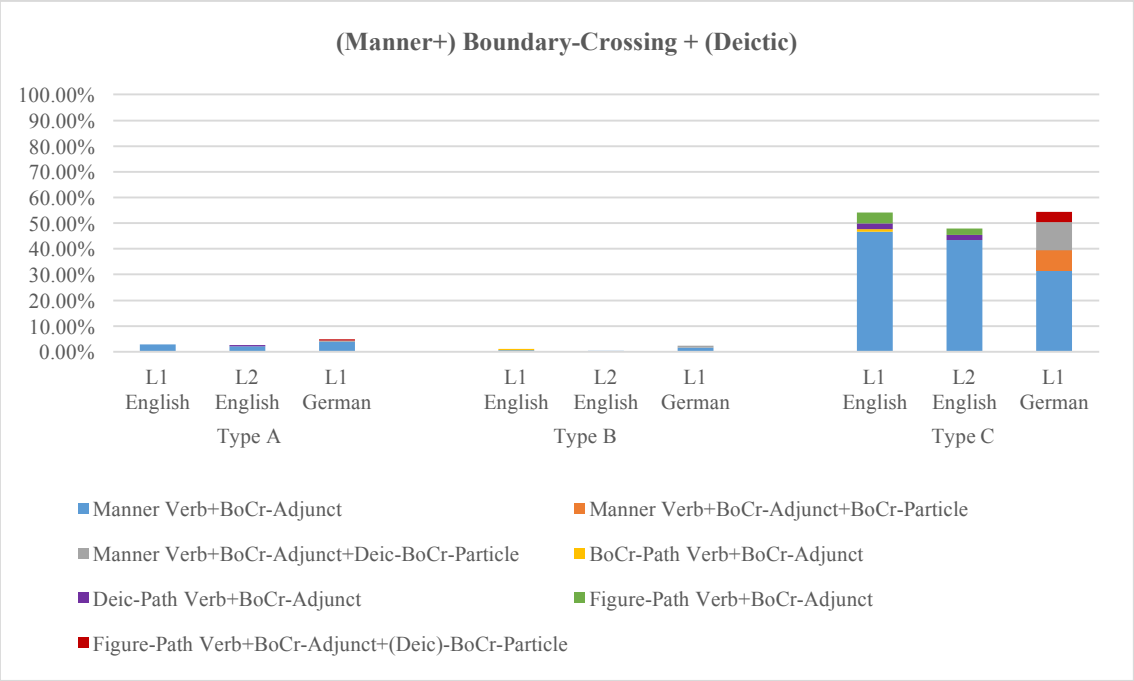


Figure 4.52: Encoding of (Manner) + Boundary-Crossing + (Deictic) by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos

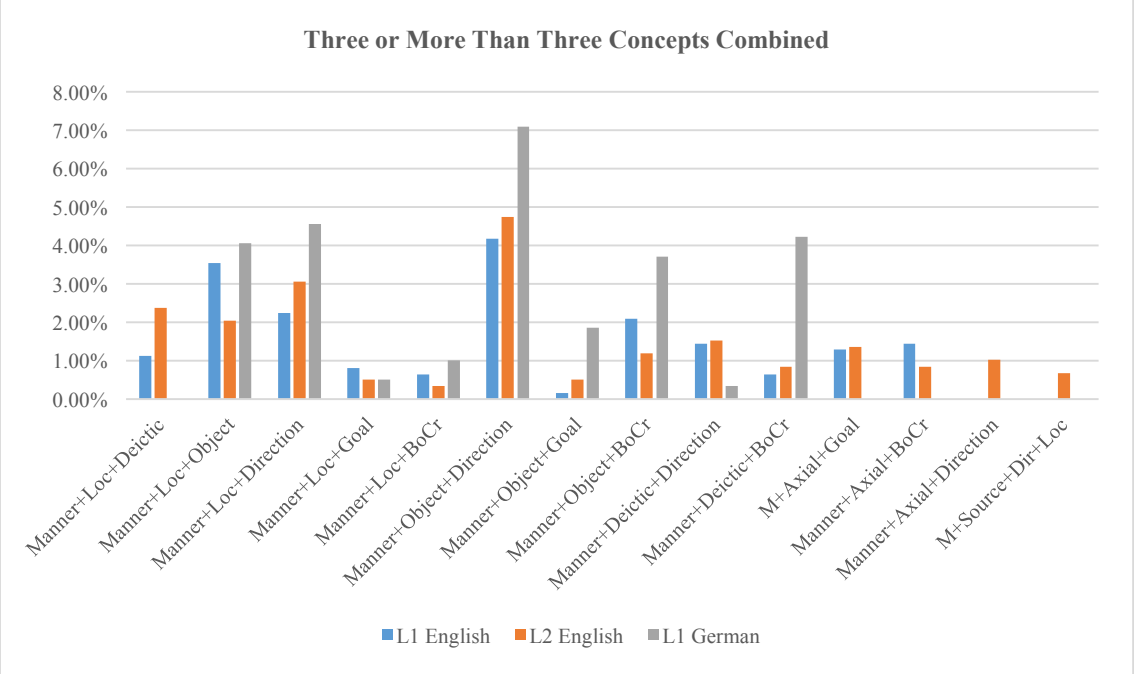


Figure 4.53: Encoding of three or more spatial concepts by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos overall

There are more than 50 different combinations of semantic components in L1 English, L1 German-L2 English, and L1 German, some of which had rather low numbers of occurrence. Figure 4.53 therefore only shows part of the combinations of three or more semantic components[71].

Unlike Chinese, there are no restrictions on the combination of the spatial concept Location with other concepts in English and German. The statistical results showed no significance in the combinations of Manner + Location + Deictic ($\chi 2$ (1, N = 21) = 2.333, p = 0.127), Manner + Location + Object ($\chi 2$ (2, N = 58) = 4.276, p = 0.118), and Manner + Location + Direction ($\chi 2$ (2, N = 59) = 4.508, p = 0.105) among speakers of these three language groups. The occurrence of the combinations Manner + Location + Goal and Manner + Location + Boundary-Crossing was lower than 1% in all three groups. Furthermore, the spatial concept of Object frequently occurred together with other concepts. Statistical analysis showed no significant difference in encoding Manner + Object + Direction ($\chi 2$ (2, N = 96) = 4.750, p = 0.093) among all three language groups. A significant relationship was documented, however, in encoding Manner + Object + Goal and Manner + Object + Boundary-Crossing among speakers of the three groups. L1 German speakers encoded Manner + Object + Goal more often than speakers of L1 English ($\chi 2$ (1, N = 12) = 8.333, p = 0.004) and L2 English ($\chi 2$ (1, N = 14) = 4.571, p = 0.033). As for the encoding of Manner + Object + Boundary-Crossing, L1 German speakers encoded them more often than L2 English speakers ($\chi 2$ (1, N = 29) = 7.759, p = 0.005), while no significant difference was documented between L1 and L2 English speakers ($\chi 2$ (1, N = 20) = 1.800, p = 0.180) and between L1 English and L1 German speakers ($\chi 2$ (1, N = 35) = 2.314, p = 0.128). Concerning the encoding of Manner + Deictic + Direction, no significant effect was found among speakers of the three groups ($\chi 2$ (2, N = 20) = 4.900, p = 0.086). In contrast, L1 German speakers encoded Manner + Deictic + Boundary-Crossing more

[71] Other possible combinations in L1 English, L1 German-L2 English, and L1 German are as follows:
L1 English: Manner Verb + Loc Adjunct + Obj Adjunct + Deictic Adjunct (0.32%), Manner Verb + Loc Adjunct + Obj Adjunct + BoCr Adjunct (0.32%), Manner Verb + Obj Adjunct + Dir Adjunct + BoCr Adjunct (0.16%), Manner Verb + Loc Adjunct + Obj Adjunct + Dir Adjunct (0.64%), Manner Verb + Dir Adjunct + BoCr Adjunct (0.32%), Deictic Path Verb + Deictic Adjunct (0.48%), Deictic Path Verb + Obj Adjunct (0.64%), Deictic Path Verb + Dir Adjunct (0.48%), etc.

L2 English: Manner Verb + Goal Adjunct + Obj Adjunct + Deictic Adjunct (0.17%), Manner Verb + Deictic Adjunct + Goal Adjunct (0.34%), Manner Verb + Source Adjunct + Dir Adjunct (0.34%), Manner Verb + Source Adjunct + Dir Adjunct + Loc (0.68%), Figure-Path Verb + Loc Adjunct (0.34%), Deictic-Path Verb + Deictic Adjunct (1.02%), Deictic-Path Verb + Goal Adjunct (0.34%), Goal-Path Verb + Goal Adjunct (0.34%), etc.

L1 German: Motion-Verb + Direction Adjunct + Loc (0.68%), Motion Verb + Loc (0.17%), Manner Verb + Source + Boundary-Crossing Adjunct (0.34%), Direction Particle + Manner Verb + Goal Adjunct (0.34%), Direction Particle + Manner Verb + Boundary-Crossing Adjunct (0.84%), Boundary-Crossing Deictic Particle + Manner Verb + Goal Adjunct (0.34%), etc.

often than speakers of L1 English ($\chi 2$ (1, N = 29) = 15.207, p < .001) and L2 English ($\chi 2$ (1, N = 30) = 13.333, p < .001). In addition, although they did so rarely, L2 English speakers also encoded Manner + Axial + Goal (1.36%), Manner + Axial + Boundary-Crossing (0.85%), Manner + Axial + Direction (1.02%), and Manner + Location + Source + Direction (0.68%).

Two-clause pattern

In contrast to the occurrence of one utterance, the occurrence of two utterances was found to be drastically lower in all three language groups: the rate was 4.71% in L1 English, 6.13% in L1 German-L2 English, and 4.99 % in L1 German (overall across Type A, B, and C videos). Statistical analysis showed no significant difference in the frequency of occurrence of two utterances among speakers of all three language groups ($\chi 2$ (2, N = 90) = 0.800, p = 0.670).

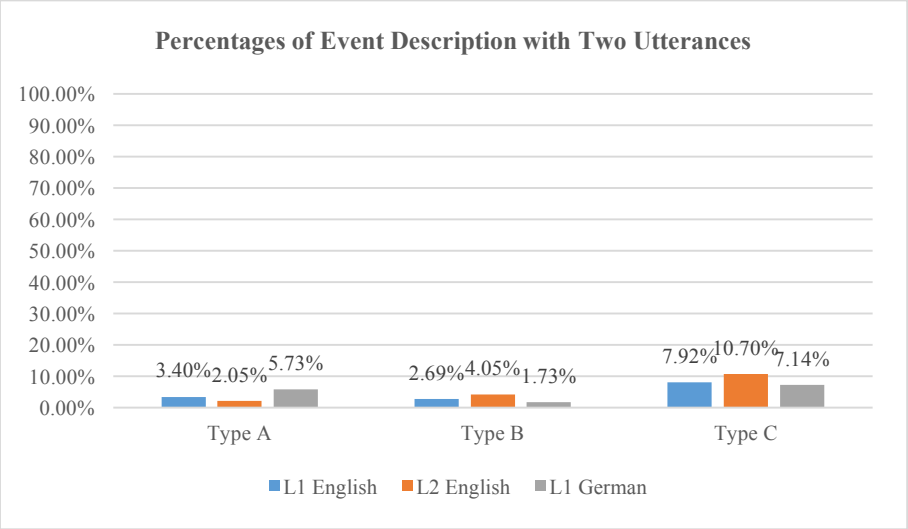


Figure 4.54: Percentages of event descriptions with two utterances by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos

When considering the occurrence of two utterances in different video types, as shown in Figure 4.54, statistically, the results showed no significant difference in the verbal task for Type A videos ($\chi 2$ (2, N = 22) = 3.364, p = 0.186), Type B videos ($\chi 2$ (2, N = 15) = 1.600, p = 0.449), and Type C videos ($\chi 2$ (2, N = 50) = 1.120, p = 0.571) among speakers of the three language groups. However, it appears that more two utterances were produced in the verbal task for Type C videos than in those of Type A and Type B videos. When comparing the occurrence of two utterances in the verbal task for Type C and Type A videos, it was found that two utterances were produced more frequently in the verbal task for Type C videos than in the verbal task for Type A

videos in L2 English ($\chi 2$ (1, N = 24) = 10.667, p = 0.001), while no significant difference in the occurrence of two utterances was found between descriptions of Type A and Type C videos in L1 English ($\chi 2$ (1, N = 23) = 3.522, p = 0.061) and L1 German ($\chi 2$ (1, N = 25) = 0.360, p = 0.549). When comparing the occurrence of two utterances in the verbal task for Type C and Type B videos, it was found that speakers of all three groups produced two utterances significantly more often in the verbal task for Type C than in those of Type B videos (L1 English $\chi 2$ (1, N = 21) = 5.762, p = 0.016, L2 English $\chi 2$ (1, N = 27) = 6.259, p = 0.012, and L1 German ($\chi 2$ (1, N = 17) = 7.118, p = 0.008).

In the verbal task for Type C videos, two utterances were mainly produced when speakers described the situation showing a man riding a bicycle and turning into an entrance. 11 out of 20 L2 English speakers uttered two utterances when describing this situation. Similarly, 10 out of 21 L1 English speakers and 9 out of 20 L1 German speakers did this in the same way. This is shown in the following examples:

One-clause pattern

(138) Here is a man on the bicycle driving down the street into a garage between different buildings. (L1 Ger-L2 Eng P03, C09)

Two-clause pattern

(139) Clause 1: A man is slowly riding his bike along the street
Clause 2: and turning into a garage. (L1 Ger-L2 Eng P10, C09)

The above two examples show that different path segments can be stacked after manner verbs within one utterance (example 138), while when path verbs are used, the combination with other semantic components is restricted (example 139). For example, when the path verb *turn* is used, one cannot say * *a man is turning along the street into the garage*, since the path verb *turn* implicates the moving of the body or a part of the body so that it is facing in a different direction, which is not compatible with the spatial information derived from the contours of the Ground object. Hence, speakers have to produce two utterances as in example 139. Among the L2 English speakers who produced two utterances, 8 out of 10 speakers produced the first clause with a manner verb such as *ride*, *cycle,* or *drive* and the second clause with a path verb such as *enter* or *turn*. This was true for those speakers of L1 English and L1 German who produced two utterances: 10 out of 10 L1 English speakers and 6 out of 9 L1 German speakers

uttered the first clause with a manner verb and the second clause with a path verb. This finding is in line with that by Gerwien and von Stutterheim (2018), which showed that when the Figure's orientation or direction changes, a second utterance is required. Although Gerwien & von Sutterheim's finding is based on the observation of French, a type of V-language, it can be applied to S-languages as well when speakers of S-languages use path verbs implicating a change of the Figure's orientation or direction.

4.5.2 Motion event cognition in L1 and L2 English

In section 4.5.1, the use of verb types and adjuncts by speakers of L1 English, L2 German-English, and L1 German was analyzed. It was shown that despite low numbers of occurrences, L2 English speakers tended to use more object-related and boundary-crossing path verbs than L1 English speakers to describe Type B videos. Concerning the use of adjuncts, L1 English speakers used deictic adjuncts more often than L2 English speakers, while there was no difference in the use of adjuncts encoding information such as location, direction, or boundary-crossing between L1 and L2 English in different video types. In order to gain insights into the cognitive processes involved in L1 and L2 English, two indicators were employed: speech onset times (SOTs) and fixation patterns. In this section, the SOTs and fixation patterns (i.e., the number of the fixation counts and the duration of the fixations on the endpoint and the moving entity, respectively) among speakers of L1 English, L1 German-L2 English, and L1 German will be analyzed.

4.5.2.1 Speech onset times

Table 4.13: Average speech onset times in the verbal task of Type A, B, and C videos by speakers of L1 English, L1 German-L2 English, and L1 German

	L1 English		L1 German-L2 English		L1 German	
	Mean	SD	Mean	SD	Mean	SD
Type A	2051ms	818ms	2705ms	1388ms	2792ms	1362ms
Type B	1893ms	680ms	2577ms	1423ms	2596ms	1383ms
Type C	2024ms	782ms	2662ms	1364ms	2728ms	1351ms
Total	1994ms	768ms	2651ms	1389ms	2709ms	1365ms

As shown in Table 4.13, generally speaking, L1 German-L2 English and L1 German speakers had later SOTs (overall across Type A, B, and C videos) than L1 English speakers. One-way ANOVA showed a significant difference among speakers

of the three groups ($F (2, 1747) = 63.615$, $p < 0.001$). Dunnett T3 post-hoc tests revealed a significant effect in average SOTs (overall across Type A, B, and C videos) between speakers of L2 English ($M = 2651$, $SD = 1389$) and L1 English ($M = 1994$, $SD = 768$) ($p < 0.001$) and between speakers of L1 German ($M = 2709$, $SD = 1365$) and L1 English ($M = 1994$, $SD = 768$) ($p < 0.001$), while no significant effect was found between speakers of L2 English ($M = 2651$, $SD = 1389$) and L1 German ($M = 2709$, $SD = 1365$) ($p = 0.850$).

When the characteristics of different video types were taken into consideration, it was found that the SOTs showed the same tendency with respect to the time span in all three groups; that is, the time span of SOTs in the verbal task for Type A videos was longer than in those of Type C, and the time span of SOTs in the verbal task for Type C videos was longer than in those of Type B, which can be represented as Type A > Type C > Type B. The reason for that might lie in the fact that the video clips of Type A stopped playing before the moving entity's arrival at the endpoint. Therefore, speakers had to wait until they could be sure whether the moving entity actually reached the endpoint or not. This longer processing time led to later SOTs in the verbal task for Type A videos. In contrast, Type B videos showed a long trajectory with a possible endpoint. Because the trajectory was long and the enpoint was not evident, so it did not take a long time for speakers to recognize whether the moving entity reached the endpoint or not. Consequently, speakers of all three groups had relatively early SOTs in the verbal task for Type B videos. However, statistical analysis did not show any significant differences in SOTs with respect to Type A, B, and C videos among speakers of L1 English ($F (2, 587) = 2.306$, $p = 0.101$), L2 English ($F (2, 577) = 0.410$, $p = 0.664$), and L1 German ($F (2, 577) = 1.003$, $p = 0.368$).

When comparing the SOTs in the verbal task for different video clips, it was found that in those of Type A, speakers of L2 English and L1 German had later SOTs than speakers of L1 English. One-way ANOVA revealed a significant effect among speakers of the three groups ($F (2, 607) = 23.041$, $p < 0.001$). Dunnett T3 post-hoc tests showed a significant difference between speakers of L2 English and L1 English ($p < 0.001$) and between speakers of L1 German and L1 English ($p < 0.001$), while no significant effect was found between speakers of L2 English and L1 German ($p = 0.895$). In the verbal task for Type B videos, speakers of L2 English and L1 German again had later SOTs than L1 English speakers. The statistical results showed a significant effect among speakers of the three groups ($F (2, 537) = 19.716$, $p < 0.001$). Dunnett T3 post-

hoc tests showed a significant difference between L2 English and L1 English ($p < 0.001$) and between L1 German and L1 English ($p < 0.001$), while no significant difference was documented between L2 English and L1 German ($p = 0.999$). The same was true for descriptions of Type C videos: SOTs were later in L2 English and L1 German than in L1 English. There was a significant effect among the three groups ($F (2, 597) = 21.112$, $p < 0.001$). Dunnett T3 post-hoc tests further showed a significant effect between L2 English and L1 English ($p < 0.001$) and between L1 German and L1 English ($p < 0.001$), whereas no significant relationship was documented between L2 English and L1 German ($p = 0.947$).

In sum, SOTs in L2 English were quite close to those in L1 German. In contrast to L1 English speakers, L2 English speakers had later SOTs. As already discussed in section 4.4.2.1, we cannot exclude an L2 effect, since all the L2 English participants were not native speakers. However, there may be other factors for the later SOTs in L2 English. We have found that L2 English speakers tended to take a holistic view and encode the endpoint, which contrasts with L1 English speakers, who tended to decompose the situation into different phases and select one phrase to describe. The former way of event construal requires speakers to process more information about the endpoint. Hence, L2 speakers have to wait until the endpoint becomes evident and identifiable as the motion scenes unfold, which leads to relatively late SOTs in L2 English.

4.5.2.2 Fixation counts and durations

Fixation counts reflect the attention directed to the moving entity or the endpoint, while the duration of the fixations reflects the word retrieval processes (cf. Griffin & Bock, 2000; Griffin & Spieler, 2006). So through these two aspects, we can gain insights into the cognitive processes of L2 speakers. Similar to the analysis in L1 and L2 Chinese, the eye-tracking data was analyzed based on the following measures: number of fixation counts in total, number of fixation counts before SOTs, duration of fixations in total, and duration of fixations before SOTs. The entire analysis was conducted with average measures across speakers over different items in Type A, B, and C videos, respectively. Before the analysis of the eye-tracking data, the relative frequency with which the endpoint was mentioned in the verbal task among speakers of the three groups was compared as presented in the following. Expressions like *towards a car*, *to a village*, *enter a shop*, *into a stable* are marked as expressions where

the endpoint is mentioned.

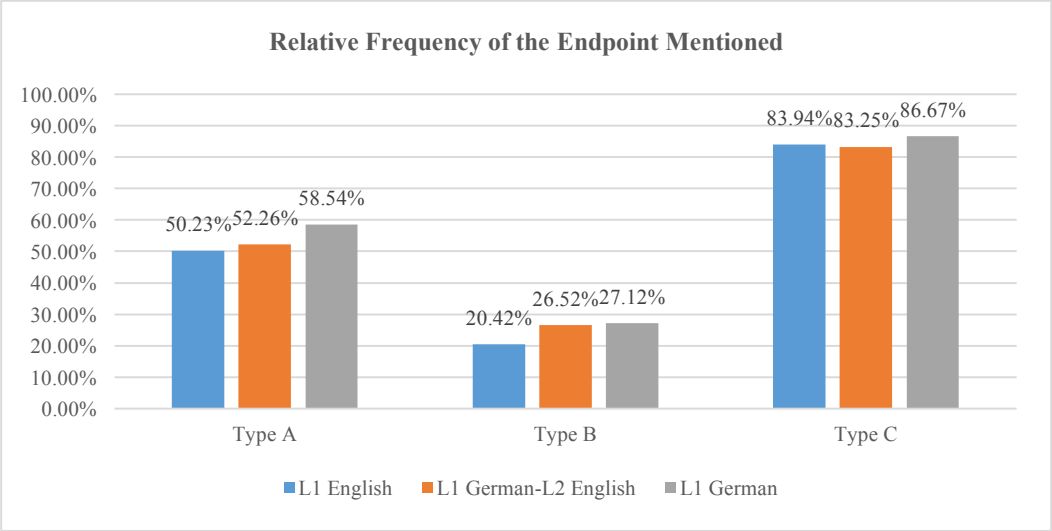


Figure 4.55: Bar charts showing the relative frequency of the endpoint being mentioned by speakers of L1 English, L1 German-L2 English, and L1 German in the verbal task of Type A, B, and C videos

As shown in Figure 4.55, in general, the endpoint was mentioned more frequently in the verbal task for Type C videos, followed by those of Type A videos, and was least mentioned in the verbal task for Type B videos among speakers of the three groups. This order is in conformance with the features of different video types. The endpoint was most salient in the boundary-crossing events in Type C videos, rather less evident in Type A videos, and least evident in Type B videos. Speakers could easily perceive the salient element in the situation and formulate it in utterances. When comparing the endpoint mentioned within each video type, no significant relationship was documented among speakers of the three groups in the verbal task for Type A ($\chi 2$ (2, N = 331) = 1.311, p = 0.519), Type B ($\chi 2$ (2, N = 135) = 1.200, p = 0.549), and Type C videos ($\chi 2$ (2, N = 539) = 0.271, p = 0.873).

Eye-tracking data analysis

Type A: Short trajectory with obvious endpoint

Fixation on the endpoint

For Type A videos, as shown in Table 4.14 and Figure 4.56, when the total number of fixation counts on the endpoint was compared among speakers of the three groups, one-way ANOVA showed no significant effect (F (2, 602) = 1.480, p = 0.228). Neither did the one-way ANOVA on the number of fixation counts before SOTs show any significance among the three groups (F (2, 586) = 0.771, p = 0.463).

Table 4.14: Mean value and standard deviation of fixation counts and duration of fixations on the endpoint in total and before SOTs in Type A videos

	L1 English		L2 English		L1 German	
Fixation on the Endpoint (Type A)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	3.46	2.69	3.15	2.59	3.05	2.32
Before SOTs	1.11	1.31	1.15	1.39	1.27	1.45
Duration of Fixations						
Total (in ms)	197	143	151	101	193	127
Before SOTs (in ms)	139	175	109	115	150	149

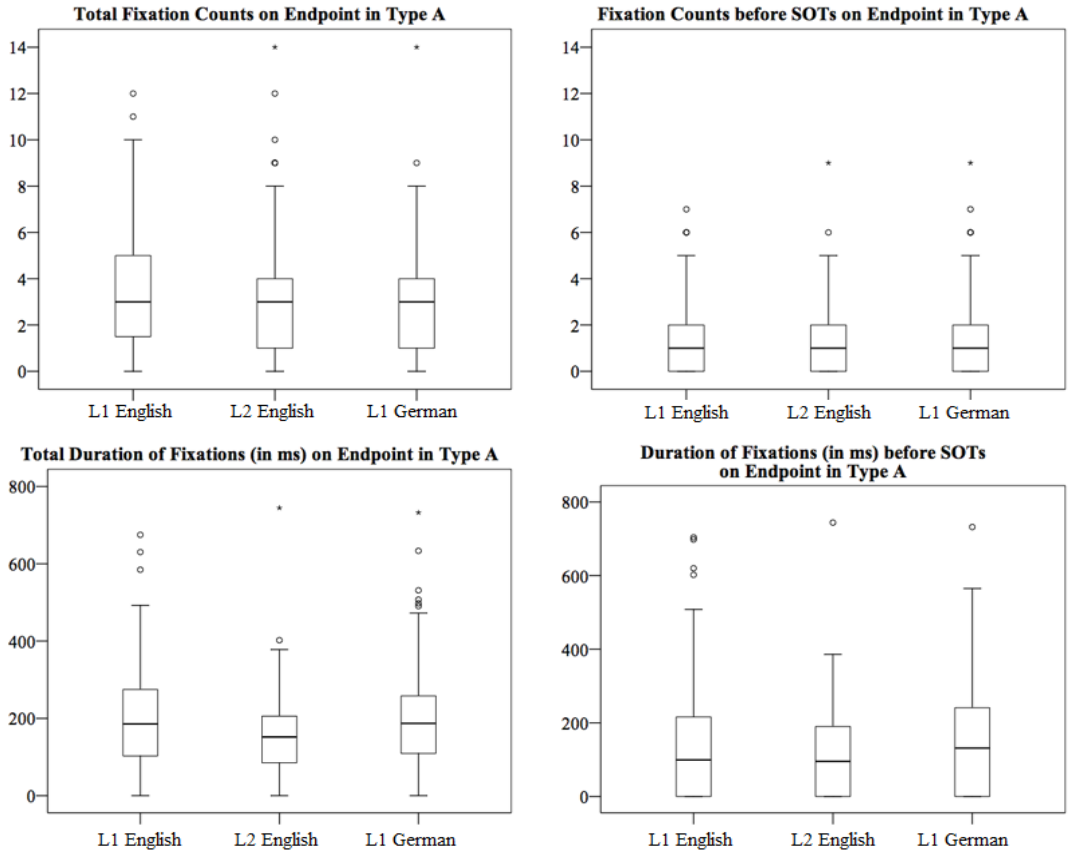


Figure 4.56: Number and duration of fixations on the endpoint in total and before SOTs by speakers of L1 English, L1 German-L2 English, and L1 German in Type A videos

However, a significant effect occurred when the total duration of the fixations was compared ($F(2, 602) = 8.125$, $p < 0.001$). Dunnett T3 post-hoc tests showed that L2 English speakers had significantly shorter fixations on the endpoint than speakers of L1 English ($p < 0.01$) and L1 German ($p < 0.01$), while no significant difference was found between L1 German and L1 English ($p = 0.993$). A significant difference was also found in terms of the duration of the fixations before SOTs among the three groups ($F(2, 586) = 4.040$, $p < 0.05$). Dunnett T3 post-hoc tests revealed a significance

between L2 English and L1 German ($p < 0.01$), while no significance was found between L2 English and L1 English ($p = 0.123$) nor between L1 English and L1 German ($p = 0.873$).

Fixation on the moving entity (Figure)

Table 4.15: Mean value and standard deviation of fixation counts and duration of fixations on the moving entity in total and before SOTs in Type A videos

	L1 English		L2 English		L1 German	
Fixation on the Moving Entity (Type A)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	6.50	3.75	8.63	4.21	7.82	3.60
Before SOTs	3.01	2.01	4.59	2.98	4.14	2.59
Duration of Fixations						
Total (in ms)	281	227	237	128	281	122
Before SOTs (in ms)	313	258	268	189	317	167

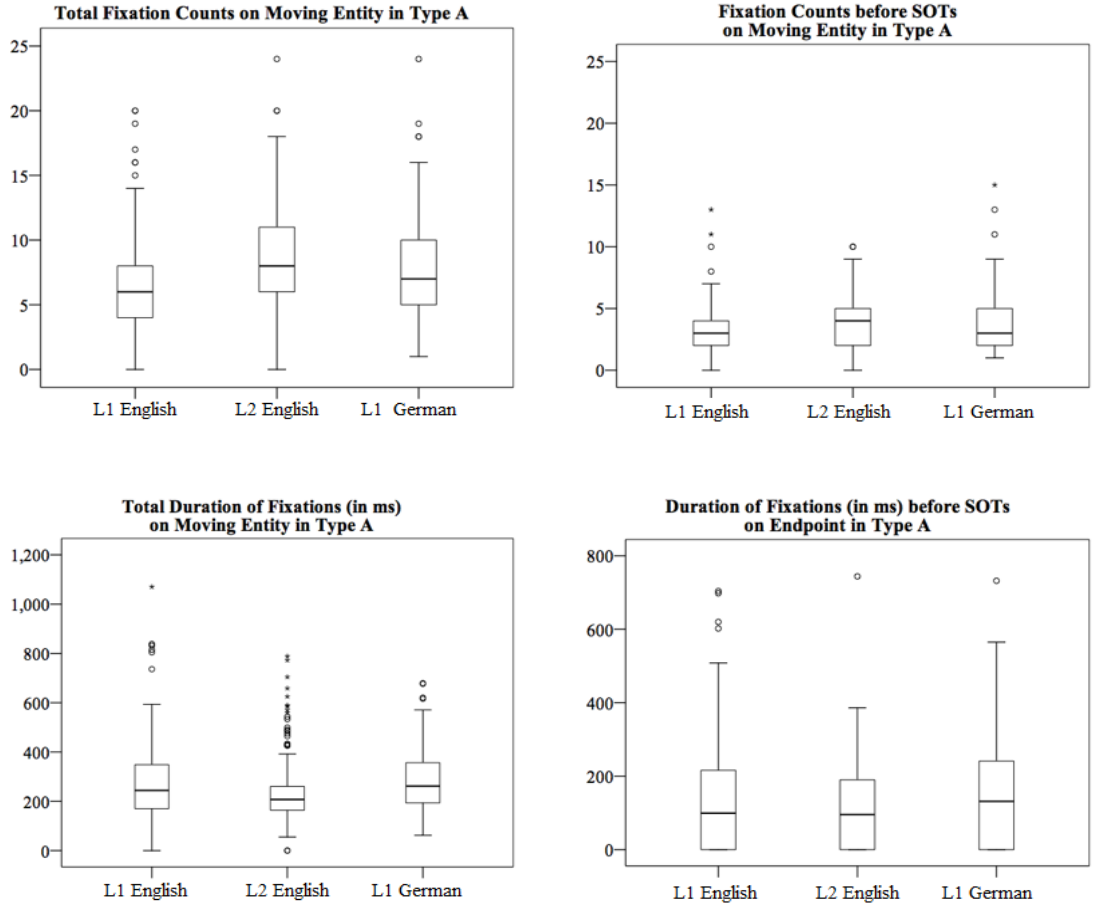


Figure 4.57: Number and duration of fixations on the moving entity in total and before SOTs by speakers of L1 English, L1 German-L2 English, and L1 German in Type A videos

As shown in Table 4.15 and Figure 4.57, one-way ANOVA revealed a significant difference in the total number of fixations on the moving entity among the three groups ($F(2, 602) = 15.852$, $p < 0.001$). Tukey HSD post-hoc tests showed that speakers of L2 English and L1 German fixated more frequently on the moving entity than L1 English speakers (both $p < 0.01$), while speakers of L2 English and L1 German did not differ in this regard ($p = 0.094$). The same was true for the statistical analysis of the fixation counts before SOTs ($F(2, 586) = 20.396$, $p < 0.001$). Tukey HSD post-hoc tests revealed that speakers of L2 English and L1 German had significantly more fixations on the moving entity than L1 English speakers (both $p < 0.001$), whereas no significance was found between L2 English and L1 German ($p = 0.290$). In addition, a significant effect was also found in terms of the total duration of the fixations $F(2, 602) = 4.519$, $p < 0.05$). L2 English speakers had significantly shorter fixations on the moving entity than L1 German speakers ($p < 0.01$). Marginal significance was found between L2 English and L1 English ($p = 0.047$), while no significant difference was found between L1 English and L1 German ($p = 1.000$). In addition, concerning the duration of the fixations before SOTs, L2 English speakers had shorter fixations on the moving entity than L1 German speakers ($p < 0.05$), while no significance was found between L2 English and L1 English ($p = 0.150$) nor between L1 German and L1 English ($p = 0.996$).

Type B: Long trajectory with possible endpoint

Fixation on the endpoint

Table 4.16: Mean value and standard deviation of fixation counts and duration of fixations on the endpoint in total and before SOTs in Type B videos

	L1 English		L2 English		L1 German	
Fixation on the Endpoint (Type B)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	2.19	2.26	1.39	1.93	1.72	1.90
Before SOTs	0.47	0.88	0.73	1.01	0.88	1.31
Duration of Fixations						
Total (in ms)	164	137	104	117	143	142
Before SOTs (in ms)	70	126	87	110	109	155

As for the total number of fixations on the endpoint among the three groups, as shown in Table 4.16 above and Figure 4.58 below, one-way ANOVA showed a significant effect ($F(2, 546) = 7.168$, $p < 0.01$). Dunnett T3 post-hoc tests showed that L1 English speakers fixated on the endpoint significantly more often than L2 English

speakers ($p < 0.01$), while no significant correlation was documented between L2 English and L1 German ($p = 0.282$) nor between L1 English and L1 German ($p = 0.090$).

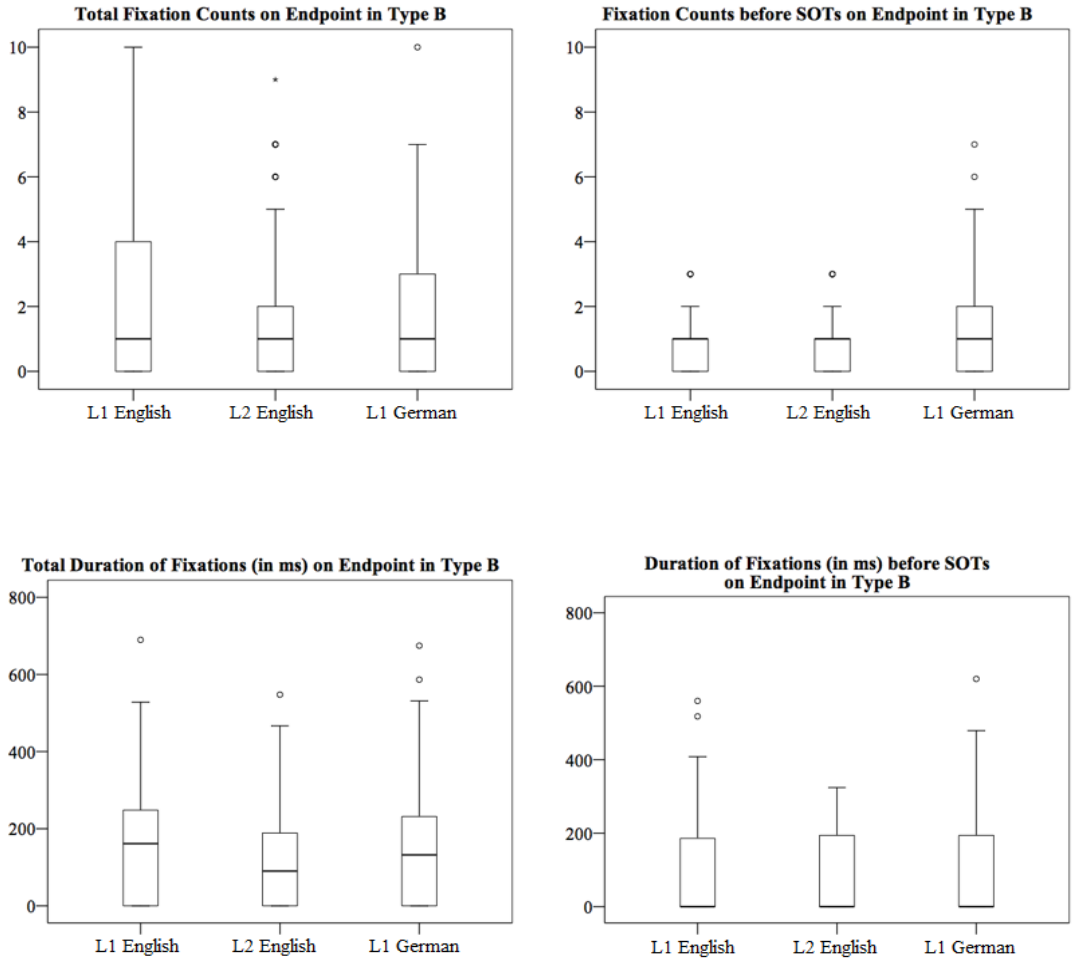


Figure 4.58: Number and duration of fixations on the endpoint in total and before SOTs by speakers of L1 English, L1 German-L2 English, and L1 German in Type B videos

However, when considering the number of fixation counts before SOTs in the three groups, a significant difference did occur ($F (2, 393) = 5.856$, $p < 0.01$). L1 German speakers had significantly more fixations on the endpoint than L1 English speakers ($p < 0.01$), while no significant difference was found between L2 English and L1 English ($p = 0.109$) and between L2 English and L1 German ($p = 0.713$). In addition, there was a significant difference in the total duration of the fixations on the endpoint among the three groups ($F (2, 545) = 9.611$, $p < 0.001$). Dunnett T3 post-hoc tests showed that speakers of L1 English and L1 German fixated on the endpoint longer than L2 English speakers ($p < 0.001$ and $p < 0.05$, respectively), while no significance was found between L1 English and L1 German ($p = 0.293$). Concerning the duration of the

fixations on the endpoint before SOTs, although the mean value in the three groups can be ordered as L1 German (M = 109 ms) > L2 English (M = 87 ms) > L1 English (M = 70 ms), the statistical results did not reach a significant level (F (2, 393) = 3.172, p = 0.069).

Fixation on the moving entity (Figure)

Table 4.17: Mean value and standard deviation of fixation counts and duration of fixations on the moving entity in total and before SOTs in Type B videos

	L1 English		L2 English		L1 German	
Fixation on the Moving Entity (Type B)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	5.87	3.68	7.79	3.70	7.07	3.47
Before SOTs	2.80	1.84	4.02	2.35	3.80	2.44
Duration of Fixations						
Total (in ms)	262	163	244	132	281	127
Before SOTs (in ms)	272	182	276	168	297	150

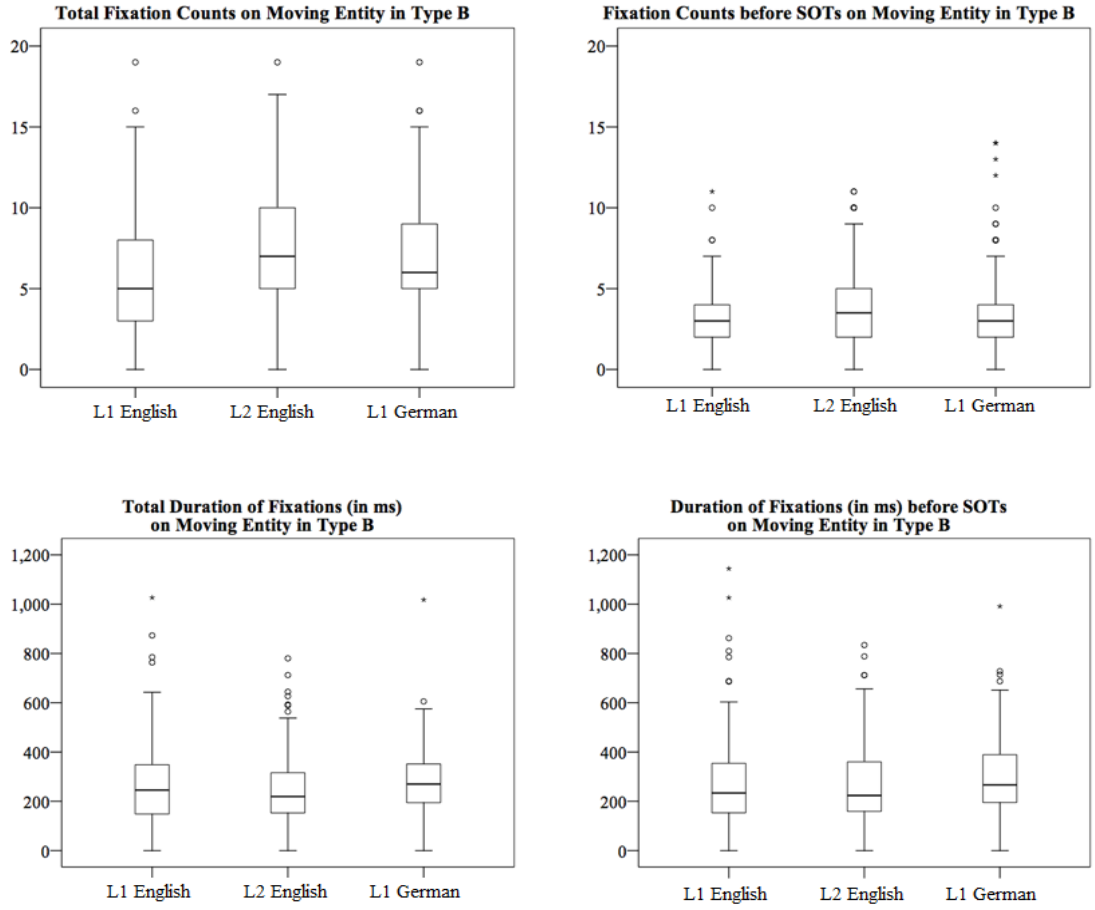


Figure 4.59: Number and duration of fixations on the moving entity in total and before SOTs by speakers of L1 English, L1 German-L2 English, and L1 German in Type B videos

As shown in Table 4.17 and Figure 4.59, one-way ANOVA showed a significant difference in the total number of fixations on the moving entity among the three groups ($F(2, 545) = 13.353$, $p < 0.001$). Tukey HSD post-hoc tests revealed that speakers of L2 English and L1 German fixated more frequently on the moving entity than L1 English speakers ($p < 0.001$ and $p < 0.01$, respectively), while there was no significant difference between L2 English and L1 German ($p = 0.139$). The same was true with respect to the number of fixation counts on the moving entity before SOTs. That is, speakers of L2 English and L1 German had more fixations on the moving entity than L1 English speakers (both $p < 0.001$), while no significant effect was documented between L2 English and L1 German ($p = 0.767$). Concerning the total duration of the fixations on the moving entity, significance was found among the three groups ($F(2, 545) = 3.070$, $p < 0.05$). L1 German speakers had longer fixations than L2 English speakers ($p < 0.05$), while no significance was found between L2 English and L1 English ($p = 0.564$) nor between L1 English and L1 German ($p = 0.507$). In contrast, there was no significant difference in the duration of the fixations on the moving entity before SOTs among all three groups ($F(2, 514) = 13.353$, $p = 0.314$).

Type C: Boundary-crossing events

Fixation on the endpoint

As already stated in section 4.4.2.2, the AoIs for the endpoint and for the moving entity overlapped at a certain time in Type C videos, and it was then difficult to differentiate the fixations on the endpoint from the fixations on the moving entity. So in this study, only the fixations that occurred before the overlapping were analyzed.

Table 4.18: Mean value and standard deviation of fixation counts and duration of fixations on the endpoint in total and before SOTs in Type C videos

	L1 English		L2 English		L1 German	
Fixation on the Endpoint (Type C)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	1.14	1.40	0.82	1.13	1.54	1.95
Before SOTs	0.62	0.90	0.60	0.82	0.79	0.99
Duration of Fixations						
Total (in ms)	152	210	114	182	174	217
Before SOTs (in ms)	123	210	103	179	142	206

The statistical analysis showed that L1 German speakers had significantly more fixations on the endpoint than L2 English speakers ($p < 0.001$), while no significant difference was found between L2 English and L1 English ($p = 0.052$) nor between L1

English and L1 German (p = 0.071) (see Table 4.18 and Figure 4.60). In contrast, speakers of the three groups did not differ in the number of fixations on the endpoint before SOTs (F (2, 533) = 2.386, p = 0.093). Concerning the total duration of the fixations on the endpoint, significance occurred among the three groups (F (2, 535) = 3.844, p < 0.05). Tukey HSD post-hoc tests showed that L1 German speakers had longer fixations on the endpoint than L2 English speakers (p < 0.05), while no significance was found between L2 English and L1 English (p = 0.195) nor between L1 English and L1 German (p = 0.540). By contrast, there was no significant difference in the duration of the fixations on the endpoint before SOTs among the three groups (F (2, 490) = 1.670, p = 0.189).

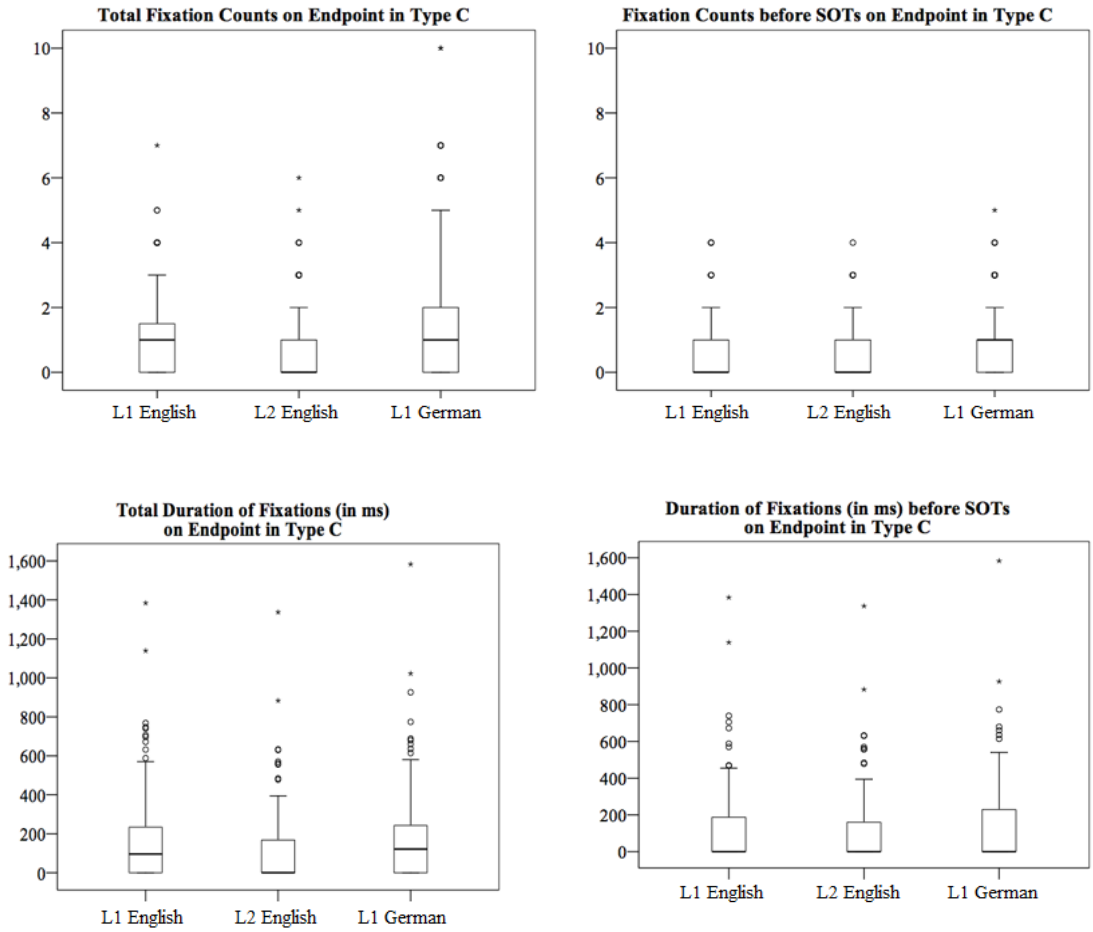


Figure 4.60: Number and duration of fixations on the endpoint in total and before SOTs by speakers of L1 English, L1 German-L2 English, and L1 German in Type C videos

Fixation on the moving entity (Figure)

Table 4.19: Mean value and standard deviation of fixation counts and duration of fixations on the moving entity in total and before SOTs in Type C videos

	L1 English		L2 English		L1 German	
Fixation on the Moving Entity (Type C)	Mean	SD	Mean	SD	Mean	SD
Fixation Counts						
Total	3.79	2.80	4.67	3.45	4.50	2.99
Before SOTs	2.97	2.11	3.77	2.58	3.44	2.34
Duration of Fixations						
Total (in ms)	254	159	227	144	236	134
Before SOTs (in ms)	256	171	228	155	236	151

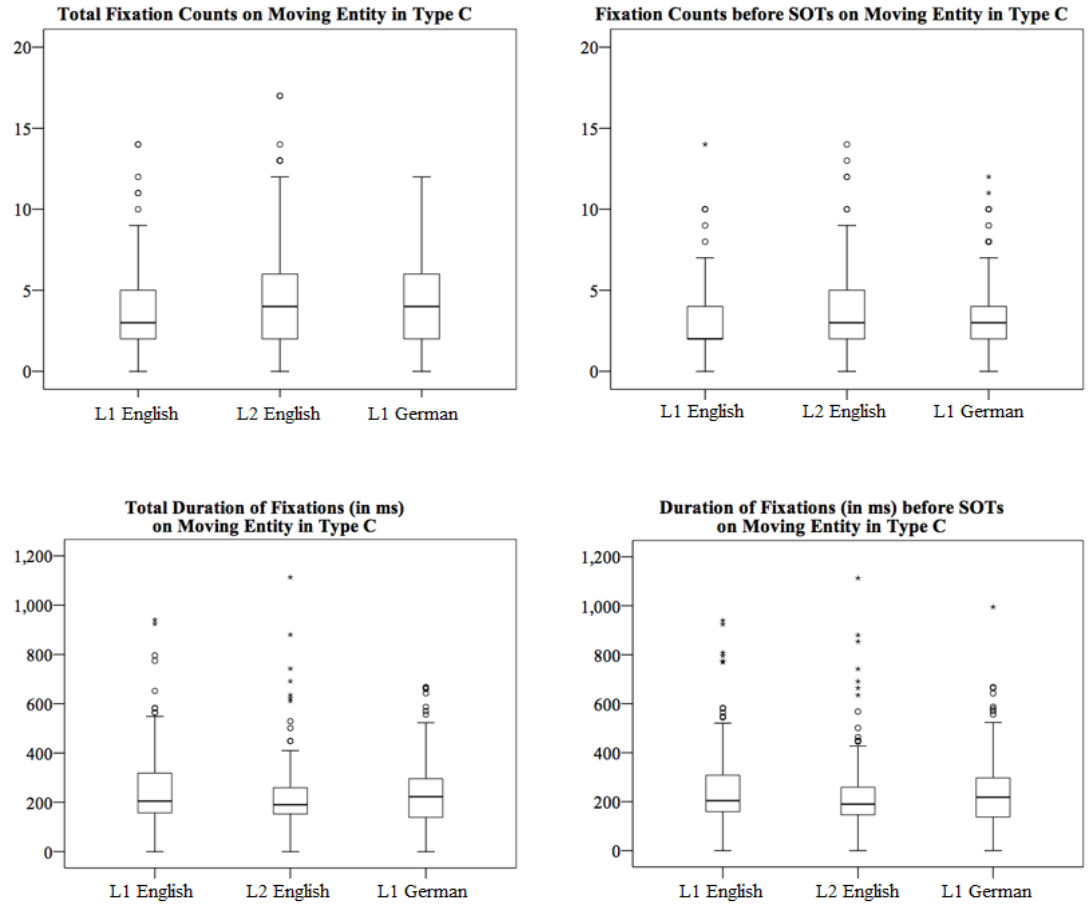


Figure 4.61: Number and duration of fixations on the moving entity in total and before SOTs by speakers of L1 English, L1 German-L2 English, and L1 German in Type C videos

As illustrated in Table 4.19 and Figure 4.61, the statistical results showed that L2 English speakers fixated more frequently on the moving entity than L1 English speakers ($p < 0.05$), while no significant difference was found between L2 English and L1 German ($p = 0.945$) nor between L1 English and L1 German ($p = 0.054$). This was

also true for the number of fixations before SOTs. L2 English speakers had significantly more fixations on the moving entity than L1 English speakers ($p < 0.01$), whereas no significance was found between L2 English and L1 German ($p = 0.492$) nor between L1 English and L1 German ($p = 0.130$). In contrast, speakers of the three groups did not differ in the total duration of the fixations ($F (2, 535) = 1.538$, $p = 0.216$) and the duration of the fixations before SOTs on the moving entity ($F (2, 533) = 1.526$, $p = 0.218$).

Summary and interpretation

To sum up, speakers of the three groups had more fixations and the fixations were longer on the moving entity than on the endpoint. As already discussed in section 4.4.2.2, this is because the moving entity, as its name indicates, was moving, while the endpoint was typically an unmovable object in the situation. According to the perceptual processing, the moving entity is more prominent compared to the static element. Therefore, all speakers tended to have more and longer fixations on the moving entity.

In the verbal task for Type A videos, speakers of the three groups did not differ significantly in mentioning the endpoint linguistically (see Figure 4.55). The eye-tracking data showed no significant difference in the number of fixations on the endpoint in total and before SOTs among the three groups either. In addition, L2 English speakers did not differ from L1 English speakers in the duration of the fixations on the endpoint before SOTs, while the duration of their fixations on the endpoint in total and before SOTs was shorter than that of L1 German speakers. That is, in comparison with L1 English speakers, L2 English speakers showed no difference in their attention to the endpoint. In particular, they did not differ from L1 English speakers with respect to the attention and the duration of the attention to the endpoint before SOTs, i.e., during the conceptualization phase. This may be attributable to the features of Type A videos, where the endpoint was evident and the distance between the endpoint and the moving entity was short. All speakers were drawn to the endpoint shown in Type A videos.

In the verbal task for Type B videos, no significant difference was found in mentioning the endpoint at the linguistic level among speakers of the three groups. However, the allocation of attention to the endpoint did not show the same tendency. L1 English speakers had significantly more fixations on the endpoint in total than L2 English speakers, while no significance was found between L2 English and L1 German

nor between L1 English and L1 German in this regard. However, when considering the number of fixations on the endpoint before SOTs, L1 German speakers fixated on the endpoint significantly more frequently than L1 English speakers, whereas no significant difference was found between L2 English and L1 German nor between L2 English and L1 English. This is to say, speakers of L1 English and L1 German differed in the conceptualization phase with respect to endpoint encoding. L1 German speakers tended to focus on the endpoint during their conceptualization process, while L1 English speakers did so to a remarkably lesser degree. When considering the duration of the fixations in total and before SOTs, it was found that both speakers of L1 German and L1 English had longer fixations on the endpoint in total than L2 English speakers. No significant difference in the duration of the fixations on the endpoint before SOTs was found among the three groups, although the mean value of the duration of the fixations before SOTs in L1 German (109 ms) was the highest and in L1 English (70 ms) it was the lowest (L1 German > L2 English > L1 English). In a nutshell, L2 English speakers showed a similar allocation of attention to the endpoint as L1 German speakers with respect to the number of fixations both in total and before SOTs as well as the duration of the fixations before SOTs. That is, they did not differ in their total attention to the endpoint and their attention before SOTs, i.e., in the conceptualization phase. That may be the reason why the SOTs did not differ greatly (L2 English: 2577 ms and L1 German 2596 ms) between the two groups for Type B videos. In contrast, L2 English speakers differed from L1 English speakers with respect to the total number of fixations and the total duration of the fixations on the endpoint, whereas they did not differ in the number of fixations and the duration of the fixations on the endpoint before SOTs. That is, it seems that L2 English speakers did not show significant differences in their attention to the endpoint during the conceptualization phase to speakers of L1 English and L1 German, while speakers of L1 English and L1 German differed significantly in the allocation of attention to the endpoint during the conceptualization phase. If we used a scale to measure the different degrees of inclination for endpoint encoding in the conceptualization phase, it would appear that at an early point, L2 English speakers would be located in the middle of the scale, with L1 German speakers, who tended to focus on the endpoint, located on one end of the scale and L1 English speakers, who paid less attention to the endpoint, on the other end. Thus, L1 German speakers' habitual event construal patterns were found to have influenced L2 English speakers when these L2 speakers described motion events in Type B videos.

For Type C videos, the eye-tracking data showed that the total number of fixations and the total duration of the fixations on the endpoint in L1 German were significantly higher than in L2 English, while no significant difference was found between L2 English and L1 English nor between L1 German and L1 English in this regard. In addition, no significant effects were found regarding the number and the duration of the fixations before SOTs among the three groups.

4.5.3 Summary

Typologically speaking, English and German belong to the S-language group. In this group of languages, Manner is typically encoded in manner verbs and Path in adjuncts and verb particles. In this study, speakers of L1, L2 English, and L1 German did not differ in the use of manner verbs. However, despite the low number of occurrences, L2 English speakers used more path verbs than speakers of L1 English and L1 German. In particular, they used more boundary-crossing path verbs than speakers of L1 English and L1 German to describe Type B and Type C videos. In addition, L2 English speakers used more object-related path verbs than L1 English speakers to describe Type A and Type B videos. That is to say, L2 English speakers used more path verbs encoding boundary-crossing and object-related information than L1 English speakers to describe Type B videos. The tendency to use these two types of path verbs in L2 English might reflect the influence of their L1's habitual conceptual patterns to describe Type B videos. L1 German speakers tended to conceptualize the situation in Type B videos as "A Figure moves in a certain manner along a Ground object" or "A Figure moves in a certain manner towards or to a Ground object at goal". The use of object-related path verbs reflects that L2 English speakers tried to draw on the features of the Ground object and the use of boundary-crossing path verbs showed L2 English speakers' focus on the endpoint, since in a boundary-crossing situation the endpoint is so salient that one cannot ignore it.

In general, L1 English speakers used more goal-oriented and deictic adjuncts than L2 English speakers, while no difference was found in the use of adjuncts encoding information on location, direction, axial, and boundary-crossing between L1 and L2 English speakers. In addition, the occurrence of no adjunct was found more in L2 English than in L1 English in the verbal task for Type C videos. This is because L2 English speakers used more boundary-crossing path verbs like *enter* than L1 English speakers did to describe Type C videos. When *enter* is used, additional adjuncts

encoding boundary-crossing information may no longer be required.

When taking an integrated view on the use of verbs, adjuncts, and verb particles in the expressions of motion events, it was found that most of the speakers tended to encode different spatial information in one utterance. In comparison with L1 and L2 English speakers, L1 German speakers tended to encode a combination of the semantic components Manner and goal in the verbal task for Type A videos and a combination of the semantic components Manner and object in the verbal task for Type B videos, while no difference was found in encoding combinations of semantic components such as Manner and location, Manner and object, Manner and direction, Manner and goal, as well as Manner and boundary-crossing within one utterance between L1 and L2 English speakers. The occurrence of the two-clause pattern was extremely low in all three groups. Also, there was no difference in the frequency of occurrence of two clauses among speakers of all three groups in the verbal task for Type A, B, and C videos. However, in comparison with descriptions of Type A and B videos, L2 English speakers tended to produce two utterances to describe Type C videos. In addition, the other two native speaker groups, i.e., speakers of L1 English and L1 German, also produced two utterances more often when describing Type C than Type B videos. One Type C video clip seemed to 'trigger' the occurrence of two clauses in L1 and L2. This video clip showed a man riding a bicycle and turning into an entrance. It was found that when the path verb *turn* in English, respectively *abbiegen* 'turn off' in German, is used to describe this situation, speakers of the three groups tended to produce two utterances. The path verb *turn* indicates a change of the Figure's orientation or direction. This finding is in line with that of Gerwien & von Stutterheim (2018), which showed that speakers of French (a V-language) tended to produce two utterances when a change of orientation or direction of the Figure was involved.

When comparing the average SOTs (overall across Type A, B, and C videos), it was found that L1 German speakers had the latest SOTs, followed by L2 English speakers. while L1 English speakers had the earliest SOTs. That is, L1 German speakers began to speak rather late, while L1 English speakers began to talk somewhat earlier. In contrast, L2 speakers began to talk later than L1 English and earlier than L1 German speakers. When the influence of the different types of videos on SOTs was taken into account, it was found that the order of SOTs showed the same tendency in all three groups: Type A > Type C > Type B. That is, speakers of all three groups tended to have late SOTs when describing Type A videos, while they tended to have early SOTs when

describing Type B videos. This phenomenon may be related to the characteristics of the different video types. Type A videos showed a short trajectory with an obvious endpoint. To process the information in this type of video, speakers must figure out whether the moving entity moves towards the endpoint or whether this is a boundary-crossing situation. So the cognitive processes are rather complex compared to Type B videos, where a long trajectory with a possible endpoint was shown. Because the endpoint was far away from the moving entity in Type B videos, speakers of the three groups could easily draw on the features of the Ground object to describe the situation. However, as mentioned above, despite the low number of occurrences, L2 English speakers used more boundary-crossing path verbs than L1 English speakers to describe Type B videos. By encoding boundary-crossing information, L2 English speakers have to pay attention to the endpoint since the endpoint is typically explicitly encoded when speakers encode boundary-crossing information. That is, L2 English speakers must wait until the endpoint becomes identifiable as the motion scene unfolds. This may be one reason why L2 English speakers had late SOTs in describing Type B videos compared with L1 English speakers.

Speakers of all three groups did not differ in mentioning the endpoint in the verbal task of describing Type A, B, or C videos. The eye-tracking data showed that in Type A videos, speakers of the three groups did not differ in the number of fixation counts on the endpoint in total and before SOTs. However, L1 German speakers had longer fixations on the endpoint with respect to the total time span of the video clips and the time span before SOTs than L2 English speakers for Type A videos. Regarding Type B videos, L1 English speakers had more fixations on the endpoint than L2 English speakers in total, whereas speakers of L2 English and L1 German did not differ in this regard. In sharp contrast, the number of fixation counts before SOTs was found more often in L1 German than in L1 English, while no difference was found between L2 English and L1 English nor between L2 English and L1 German. That is to say, in the conceptualization phase, L1 German speakers paid more attention to the endpoint than L1 English speakers in describing Type B videos, whereas speakers of L2 English and L1 German as well as speakers of L2 English and L1 English did not differ in this regard. Concerning the duration of the fixations in Type B videos, speakers of L1 English and L1 German had longer fixations on the endpoint in total, whereas speakers of the three groups did not differ in the duration of the fixations on the endpoint before SOTs. To summarize the fixation patterns for Type B videos, it appears that L2 English

speakers kept their L1 German habitual fixation patterns (i.e., paying more attention to the endpoint, especially during the conceptualization phase) while acquiring L1 English speakers' fixation patterns (i.e., paying less attention to the endpoint during the conceptualization phase). This may explain why for Type B videos, the SOTs of L2 English speakers were later than those of L1 English speakers and earlier than those of L1 German speakers (see Table 4.13) and why the percentages of the endpoint being mentioned in the verbal task of Type B videos were higher in L2 English than in L1 English speakers and lower than in L1 German speakers (see Figure 4.55). In Type C videos, the number and duration of the fixations on the endpoint in total were found more often in L1 German than in L2 English, whereas no difference was found between L1 and L2 English in both regards. By contrast, speakers of the three groups did not differ in the number and duration of the fixations before SOTs. That is, speakers of the three groups did not differ in the amount of attention paid to the endpoint in the conceptualization phase for Type C videos. As a consequence, speakers of all three groups did not differ in mentioning the endpoint in the verbal task for Type C videos.

4.6 Comparison of linguistic forms and meaning in L1 German-L2 Chinese and L1 German-L2 English

The encoding of spatial concepts and motion event cognition in L1 German-L2 Chinese and L1 German-L2 English were already discussed in sections 4.4 and 4.5, respectively. Since the source language for these two learner groups was the same, i.e., German, it would be interesting to find out whether and to what extent L1 German-L2 Chinese and L1 German-L2 English differed in describing motion events. To address this question, the use of linguistic forms (i.e., verbs and adjuncts), the encoding of spatial concepts, and the fixation patterns will be compared between the two learner groups in this section.

4.6.1 Comparison of linguistic forms used

As shown in Figure 4.62, L1 German-L2 English speakers used more manner verbs than L1 German-L2 Chinese speakers ($\chi 2$ (1, N = 1727) = 198.878, $p < .001$), while L1 German-L2 Chinese speakers used more path verbs than L1 German-L2 English speakers ($\chi 2$ (1, N = 1727) = 205.927, $p < .001$). However, L1 German-L2 Chinese speakers did not differ from L1 German-L2 English speakers in the use of

motion verbs ($\chi 2$ (1, N = 1727) = 1.443, p = 0.230) and other verbs ($\chi 2$ (1, N = 1727) = 3.077, p = 0.079). This finding is in conformance with the typological features of S-languages and E-languages, respectively. Chinese is an E-language, where both Manner and Path are encoded in verbs. In contrast, English is an S-language, where Manner is typically encoded in verbs while Path is encoded in adjuncts. As shown in Figure 4.63, as expected, L1 German-L2 English speakers used extremely more adjuncts than L1 German-L2 Chinese speakers did ($\chi 2$ (1, N = 982) = 134.925, p < .001).

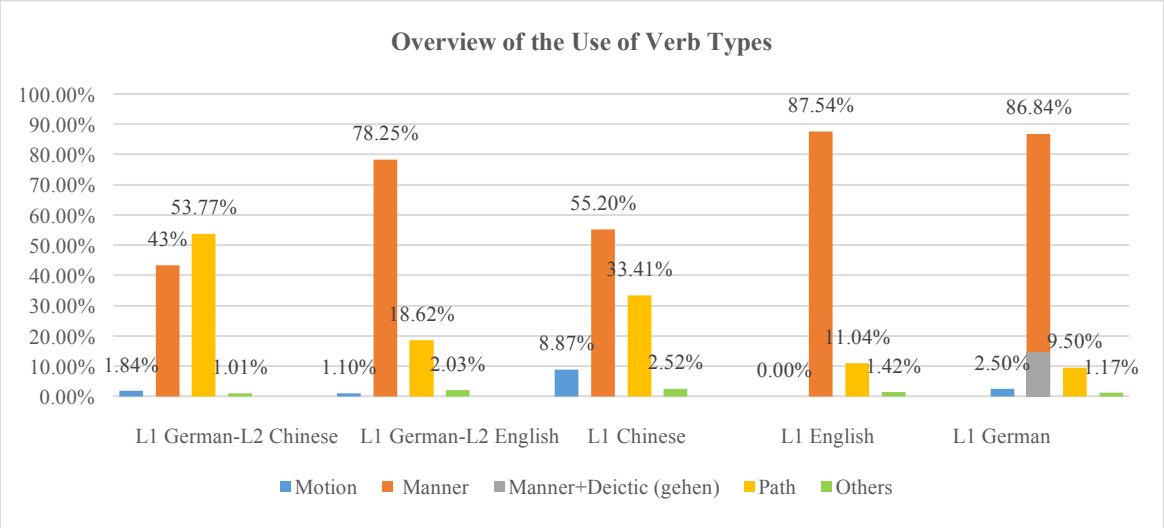


Figure 4.62: Percentages of the use of verb types in L1s and L2s

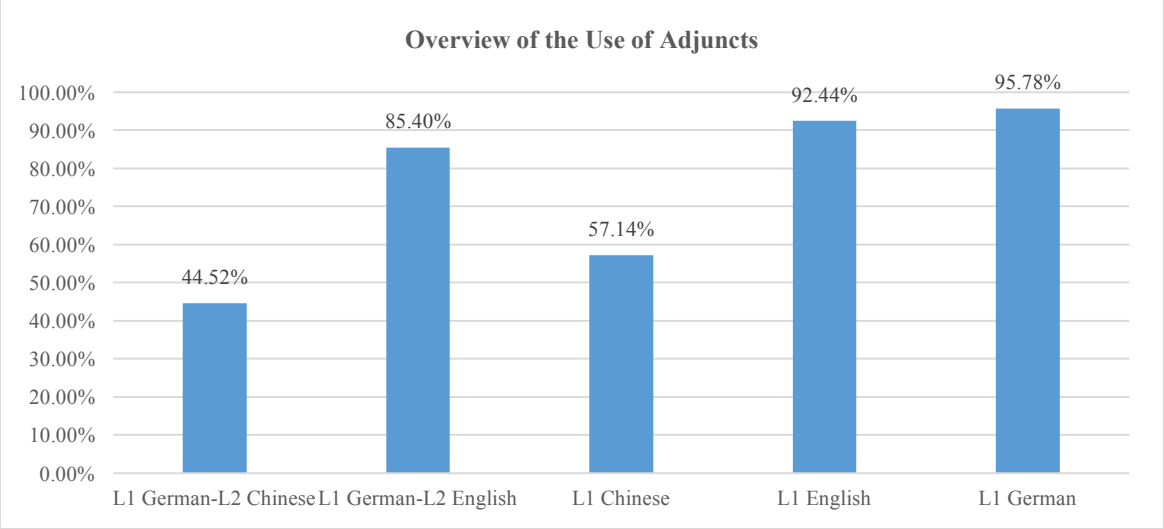


Figure 4.63: Percentages of the use of adjuncts in L1 and L2

4.6.2 Comparison of spatial concepts encoded

Depending on the language used, the same spatial concept can be encoded in

verbs, adjuncts, or verb particles. In the following, we will compare the encoding of spatial concepts in L1 German-L2 Chinese and L1 German-L2 English. The comparison will be done as follows: The spatial concepts encoded in the two learner groups will first be compared with each target language, respectively, and then the spatial encodings in these two learner groups will be compared.

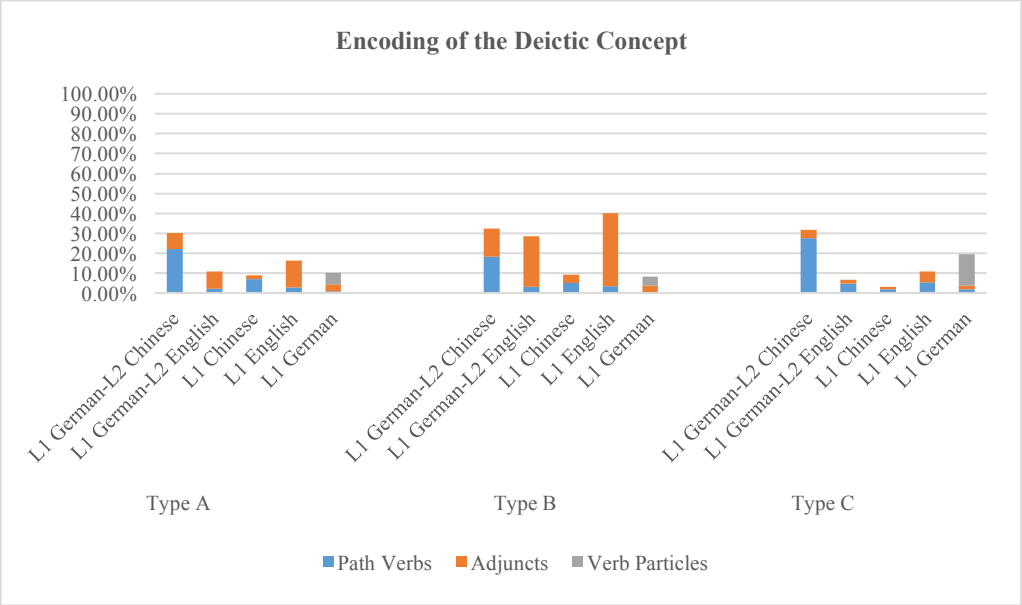


Figure 4.64: Encoding of the deictic concept in L1s and L2s in the verbal task of Type A, B, and C videos

As shown in Figure 4.64, Chi-square tests revealed that L2 Chinese speakers encoded the deictic concept overall (including in path verbs, adjuncts, and verb particles) significantly more often than speakers of L1 Chinese ($p < 0.001$ in all video types) and L1 German ($p < 0.001$ in all video types) in the verbal task of Type A, B, and C videos, respectively. This was true for its encoding in path verbs ($p < 0.001$ in all video types). In addition, L2 Chinese speakers used more adjuncts encoding deictic information than L1 Chinese speakers ($p < 0.05$ in all video types) and L1 German speakers in the verbal task of Type A ($p < 0.05$) and B ($p < 0.001$) videos, while no significant difference in the use of deictic adjuncts was found between L2 Chinese and L1 German in the verbal task of Type C videos ($p < 0.05$).

When comparing the deictic concept encoded in total between L1 and L2 English, it was found that in the verbal task of Type A videos, no significance occurred between L1 and L2 English (Chi-square tests $p = 0.085$), while L2 English speakers encoded it more often than L1 German speakers ($p < 0.05$). The same was true for the

use of adjuncts in the verbal task of Type A videos: no significance was found between L1 and L2 English (Chi-square test $p = 0.077$). Deictic adjuncts were used more often in L2 English than in L1 German (Chi-square test $p < 0.05$). In the verbal task of Type B videos, the total encoding of the deictic concept was found more often in L1 English than in L2 English (Chi-square test $p < 0.05$) and more often in L2 English than in L1 German (Chi-square test $p < 0.001$). No significant effect was found in the use of deictic path verbs between L1 and L2 English ($p = 0.782$), while the use of deictic adjuncts was found more often in L1 English than in L2 English (Chi-square test $p < 0.05$) and L1 German (Chi-square test $p < 0.001$). In the verbal task of Type C videos, significance was only found with respect to the use of deictic adjuncts. That is, the use of deictic adjuncts was found more often in L1 English than in L2 English (Chi-square test $p < 0.05$), while no significance was found between L2 English and L1 German (Chi-square test $p = 1.000$). In contrast, no significance was found in the total number of deictic encodings (L1 vs. L2 English $p = 0.150$; L2 English vs. L1 German $p = 0.144$) and in the use of deictic path verbs between any two groups (L1 English vs. L2 English $p = 0.835$; L2 English vs. L1 German $p = 0.071$).

It was interesting to find a common tendency between the two learner groups for different video types. That is, the total frequency of occurrence of deictic encodings (including path verbs and adjuncts) in L2 Chinese was significantly higher than in L2 English in the verbal task of Type A, B, and C videos (Chi-square test $p < 0.001$ in all types). More specifically, deictic path verbs were used more often in L2 Chinese than in L2 English in the verbal task of Type A, B, and C videos, respectively (Chi-square test $p < 0.001$ in all types), while there was no difference in the occurrence of deictic adjuncts between the two learner groups in all three video types (Chi-square test Type A: $p = 0.622$; Type B: $p = 0.113$ and Type C: $p = 0.166$).

Interestingly, Chi-square tests showed no significance between L1 and L2 Chinese for different video types with respect to the total number of encodings of the object-related concept (Type A: $p = 0.808$, Type B: $p = 0.083$, Type C: $p = 1.000$) and the use of object-related path verbs (Type A: $p = 0.564$, Type B: $p = 0.250$, Type C: $p = 1.000$), while the use of adjuncts was only found more often in the verbal task of Type B videos in L1 Chinese than in L2 Chinese ($p < 0.05$) (see Figure 4.65). L1 German speakers encoded the same concept more often overall than L2 Chinese speakers ($p < 0.001$ in all types). The same was true for the use of object-related adjuncts ($p < 0.001$ in all types). In addition, the occurrence of object-related path verbs was rather low in

L1 German. The statistical results showed no significance between L2 Chinese and L1 German in the verbal task of Type B videos (Chi-square test: $p = 0.206$).

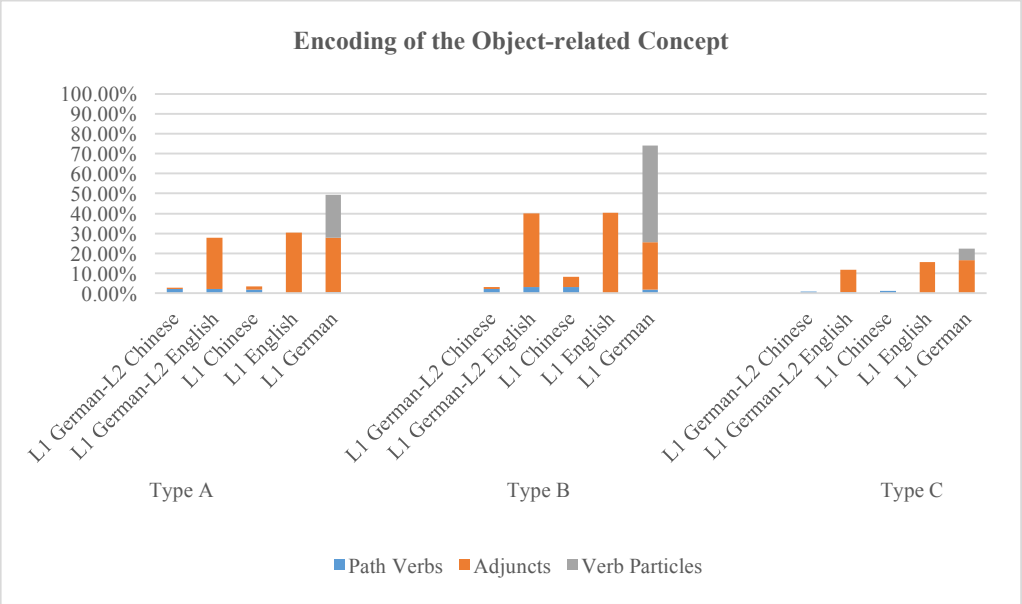


Figure 4.65: Encoding of the object-related concept in L1s and L2s in the verbal task of Type A, B, and C videos

When comparing the use of the object-related concept in L1 and L2 English and L1 German, no significant difference was documented between L1 and L2 English with respect to the total number of encodings (including path verbs, adjuncts, and verb particles) (Chi-square tests: Type A: $p = 0.413$, Type B: $p = 0.744$, and Type C: $p = 0.241$) and the occurrence of object-related adjuncts (Chi-square tests: Type A: $p = 0.194$, Type B: $p = 0.405$, and Type C: $p = 0.189$). In contrast, more encodings of the object-related concept were found overall in L1 German than in L2 English for all video types (Chi-square test: $p < 0.05$ for all types), while significance in the use of object-related adjuncts was only found more often in L2 English than in L1 German in the verbal task of Type B videos (Chi-square test: $p < 0.05$). By contrast, there was no difference in the same encodings in the verbal task of Type A videos (Chi-square test: $p = 0.413$) and Type C videos (Chi-square test: $p = 0.152$) between L2 English and L1 German. The occurrence of path verbs was rather low in all three groups, so it will not be analyzed further.

In general, the object-related concept was encoded more often in L2 English than in L2 Chinese in all three video types (Chi-square test: $p < 0.001$ in all types). The same was true for the use of object-related adjuncts: L2 English speakers used them

more often than L2 Chinese speakers in all types of videos (Chi-square test: p = 0.000 in all types), while there was no significant difference in the use of object-related path verbs between the two learner groups in the verbal task of Type A and Type B videos (Chi-square test: p = 0.413 for Type A and p = 0.782 for Type B).

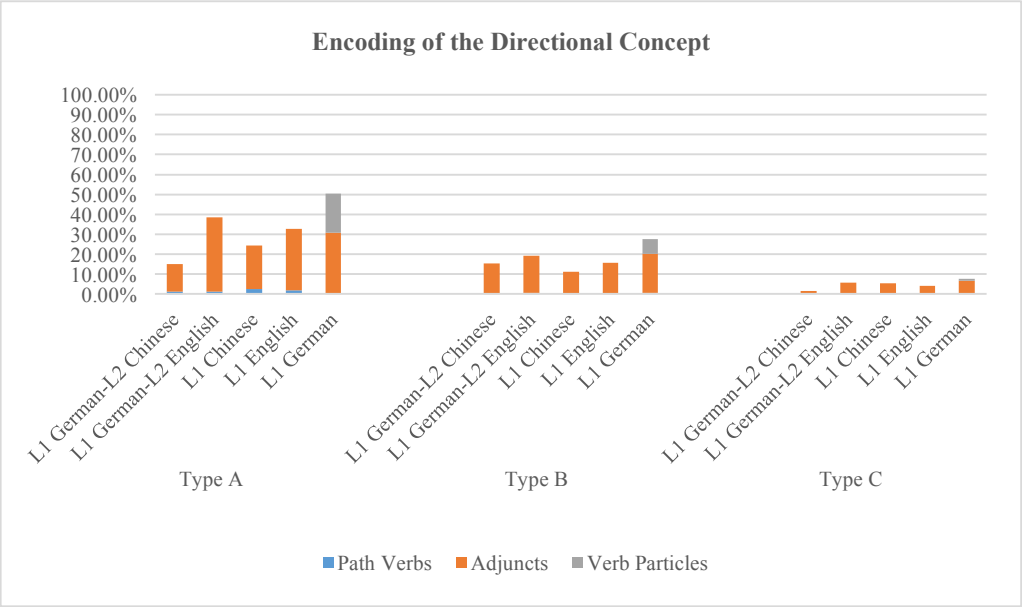


Figure 4.66: Encoding of the directional concept in L1s and L2s in the verbal task of Type A, B, and C videos

Concerning the concept of direction encoded in L1, L2 Chinese, and L1 German, as shown in Figure 4.66, no significance was found between L1 and L2 Chinese with respect to the total number of encodings (including path verbs, adjuncts, and verb particles) (Chi-square test: Type A: p = 0.244, Type B: p = 0.057, and Type C: p = 0.109) and the use of directional adjuncts (Chi-square test: Type A: p = 0.365, Type B: p = 0.057, and Type C: p = 0.083), while the total number of encodings (including path verbs, adjuncts, and verb particles) and the use of directional adjuncts was found more often in L1 German than in L2 Chinese in the verbal task of Type A and Type C videos, respectively (Chi-square test: all $p < 0.05$). By contrast, in the verbal task of Type B videos, no significance was found between L2 Chinese and L1 German with respect to the total number of encodings and the use of directional adjuncts (encodings overall: p = 0.100, use of directional adjuncts: p = 0.904). The use of directional path verbs was rather low in all three groups. Nevertheless, no significant difference was found between L1 and L2 Chinese in the verbal task of Type A and Type C videos (Chi-square test: Type A: p = 0.366 and Type C: p = 1.000).

When comparing the concept of direction encoded in L1, L2 English, and L1 German, no significance was found between L1 and L2 speakers with respect to the total number of encodings (Chi-square test: Type A: L2 English vs. L1 English $p = 0.564$, L2 English vs. L1 German $p = 0.053$; Type B: L2 English vs. L1 English $p = 0.535$, L2 English vs. L1 German $p = 0.127$, and Type C: L2 English vs. L1 English $p = 0.513$, L2 English vs. L1 German $p = 0.450$) and the use of directional adjuncts (Chi-square test: Type A: L2 English vs. L1 English $p = 0.499$, L2 English vs. L1 German $p = 0.347$; Type B: L2 English vs. L1 English $p = 0.617$, L2 English vs. L1 German $p = 0.904$, and Type C: L2 English vs. L1 English $p = 0.513$, L2 English vs. L1 German $p = 0.695$). Since the use of path verbs was rare in all three groups, it will not be discussed further.

In general, speakers of L2 Chinese and L2 English encoded the directional concept more often in the verbal task of Type A and B videos than that of Type C videos. Between the groups, the effect occurred in all video types. The total number of encodings was higher in L2 English than in L2 Chinese in the different video types (Chi-square test: $p < 0.001$ in all video types). Also, they used directional adjuncts more often than L2 Chinese speakers in the verbal task of Type A and Type C videos (Chi-square test: both $p < 0.05$), while in the verbal task of Type B videos, no difference was found between the two groups. In addition, the use of directional path verbs was rather rare in both groups.

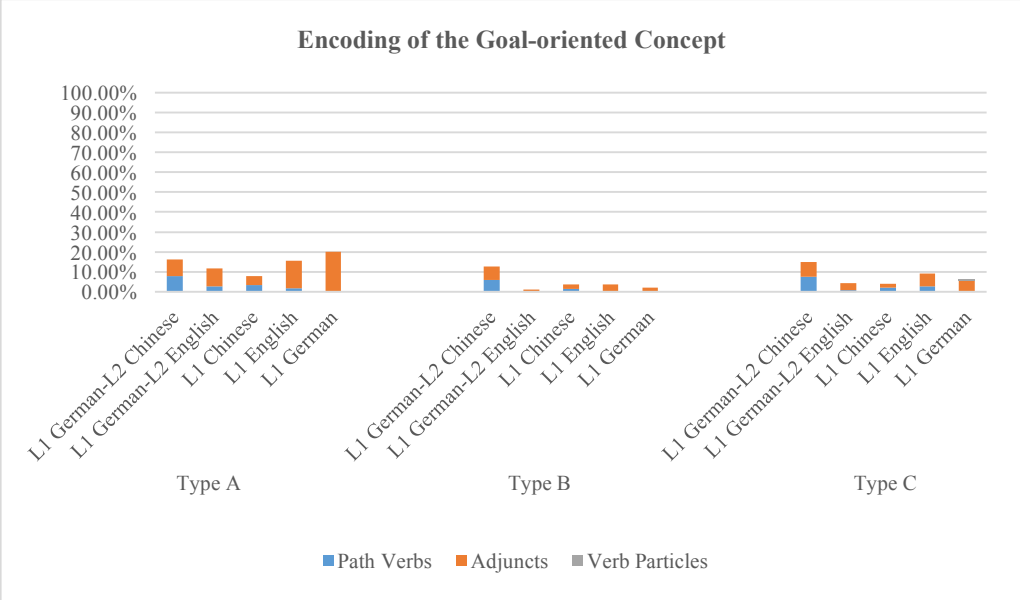


Figure 4.67: Encoding of the goal-oriented concept in L1s and L2s in the verbal task of Type A, B, and C videos

Concerning the encoding of the goal-oriented concept in L1, L2 Chinese, and L1 German, it was found that L2 Chinese speakers encoded it significantly more often than L1 Chinese speakers with respect to the total number of encodings (Chi-square test: $p < 0.001$ in all types), the use of path verbs (Chi-square test: $p < 0.05$ in all types), and the use of adjuncts (Chi-square test: $p < 0.05$ in all types) (see Figure 4.67). In addition, the total number of encodings was found to be higher in L2 Chinese than in L1 German in the verbal task of Type B and Type C videos (Chi-square test: $p < 0.001$ in both types), while no significance was found in the verbal task of Type A videos between L2 Chinese and L1 German (Chi-square test: $p = 0.458$). The use of goal-oriented adjuncts was found more often in L1 German than in L2 Chinese in the verbal task of Type A videos (Chi-square test: $p < 0.05$), while in the verbal task of Type B videos, it was found more often in L2 Chinese than in L1 German ($p < 0.05$) and no significance was found between these two groups in the verbal task of Type C videos ($p = 0.450$). None of the L1 German speakers used goal-oriented path verbs in the different video types.

Regarding the encoding of the goal-oriented concept in L1, L2 English, and L1 German, the total number of encodings was found to be higher in L1 English than in L2 English in the verbal task of Type C videos (Chi-square test: $p < 0.05$), while no significance was found between L1 and L2 English regarding the total number of encodings in the verbal task of Type A ($p = 0.233$) and Type B videos ($p = 0.535$) as well as the use of adjuncts in the different video types (Type A: $p = 0.109$, Type B: $p = 0.617$, and Type C: $p = 0.127$). Furthermore, the total number of encodings and the use of adjuncts was higher in L1 German than in L2 English in the verbal task of Type A videos (both $p < 0.05$), while no significant relationship was documented with respect to the total number of encodings and the use of adjuncts in the verbal task of Type B and Type C videos between these two groups (all $p > 0.05$).

In comparison with L2 English speakers, L2 Chinese speakers tended to encode the goal-oriented concept in all video types (all $p < 0.05$). Encodings of this concept in path verbs were found more often in L2 Chinese than in L2 English throughout all three video types (Chi-square test: $p < 0.001$ in all video types). In contrast, encodings of the goal-oriented concept in adjuncts were only found more often in the verbal task of Type B videos in L2 Chinese than in L2 English ($p < 0.05$), while no significance was found in the verbal task of Type A videos (Chi-square test: $p = 0.631$) and Type C videos ($p = 0.061$).

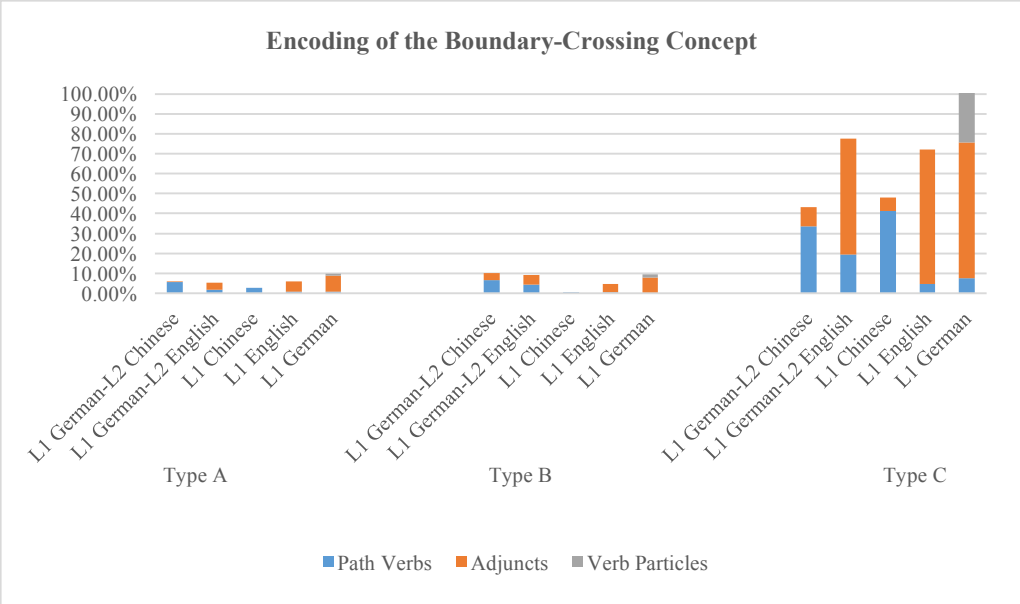


Figure 4.68: Encoding of the boundary-crossing concept in L1s and L2s in the verbal task of Type A, B, and C videos

The total number of encodings of the boundary-crossing concept was found to be higher in L2 Chinese than in L1 Chinese in the verbal task of Type A videos (Chi-square test: $p < 0.05$) and Type B videos (Chi-square test: $p < 0.001$), whereas no significant difference was found between L2 Chinese and L1 German in these two video types (Chi-square test: Type A: $p = 1.000$, Type B: $p = 0.077$) (see Figure 4.68). There was no difference in the occurrence of the total number of encodings between L1 and L2 Chinese in the verbal task of Type C videos (Chi-square test: $p = 0.911$), while the total number of encodings of the boundary-crossing concept was found to be higher in L1 German than in L2 Chinese ($p < 0.05$). This is attributable to the use of verb particles encoding boundary-crossing information in German. In addition, the encoding of the boundary-crossing concept in path verbs was found more often in L2 Chinese than in L1 Chinese in the verbal task of Type A videos (Chi-square test: $p < 0.05$) and Type B videos (Chi-square test: $p < 0.001$), whereas no significant effect was found in the verbal task of Type C videos (Chi-square test: $p = 0.681$). L2 Chinese speakers used more path verbs than L1 German speakers throughout the verbal tasks of Type A, B, and C videos ($p < 0.05$ in all types). The use of boundary-crossing adjuncts was rare in the verbal task of Type A videos in L1 and L2 Chinese, while in the verbal task of Type B videos, despite the low number of occurrences, L2 Chinese speakers used it more often (3.59%) than L1 Chinese speakers (0%); in the verbal task of Type C videos, no significant effect was found between these two groups (Chi-square test: $p = 0.117$).

Since boundary-crossing information is typically encoded in adjuncts in German, L1 German speakers used them significantly more often than L2 English speakers in the verbal task of Type A and Type C videos (both $p < 0.001$), while no significant difference was found between L2 Chinese and L1 German in the verbal task of Type B videos ($p = 0.201$).

When comparing the same concept encoded in L1, L2 English, and L1 German, no significant difference was found between L1 and L2 English and between L2 English and L1 German (except for Type C videos) with respect to the total number of encodings and the use of adjuncts (in the verbal task of Type A, B, and C videos) (Chi-square test: all $p > 0.05$). However, due to the encoding of boundary-crossing information in verb particles, the total number of encodings in the verbal task of Type C videos was found to be higher in L1 German than in L2 English ($p < 0.05$). In addition, in the verbal task of Type C videos, L2 English speakers used more path verbs than speakers of L1 English and L1 German, respectively (both $p < 0.001$).

Interestingly, there was no difference in the total number of encodings of the boundary-crossing concept between the two learner groups for the three video types (Chi-square test: Type A: $p = 0.106$, Type B: $p = 0.077$, and Type C: $p = 0.868$). When considering the use of path verbs and adjuncts, respectively, it was found that L2 Chinese speakers used more boundary-crossing path verbs than L2 English speakers in all video types (Chi-square test: $p = .002$ for Type A, $p = 0.016$ for Type B, and $p < .001$ for Type C), while L2 English speakers used more boundary-crossing adjuncts than L2 Chinese speakers in the verbal task of Type A and B videos, respectively (Chi-square test: $p < .001$). In comparison, there was no difference in the use of adjuncts encoding the boundary-crossing concept in the verbal task of Type B videos between the two groups (Chi-square test: $p = 0.808$).

The total number of encodings of location was found to be higher in L2 Chinese than in L1 Chinese in the verbal task of Type B videos (Chi-square test: $p < 0.05$), while no significance was found between L1 and L2 Chinese in the verbal task of Type A videos ($p = 0.106$) (see Figure 4.69). In contrast, L2 Chinese speakers encoded more information on location than L1 German speakers in the verbal task of Type A and B videos (both $p < 0.05$). Note that the Chinese compound verb *zǒu-lù* ('walk-road') includes Manner as well as location information, with the meaning *walk on the road*. Therefore, it is no wonder that some verbs encoding location were used in L1 Chinese and L1 German-L2 Chinese. It seems that L2 Chinese speakers used this compound

verb more often than speakers of L1 Chinese and L1 German in the verbal task of Type A and B videos ($p < 0.05$ in all cases). In addition, L1 Chinese speakers used remarkably more locative adjuncts to describe Type B videos than L2 Chinese speakers ($p < 0.001$), while no significant difference was found in the verbal task of Type A videos between L1 and L2 speakers (L1 vs. L2 Chinese $p = 0.340$, L2 Chinese vs. L1 German $p = 0.329$) and in that of Type B videos between L2 Chinese and L1 German ($p = 0.329$). In the verbal task of Type C videos, the occurrence of the locative concept was rare, so no statistical analysis will be conducted.

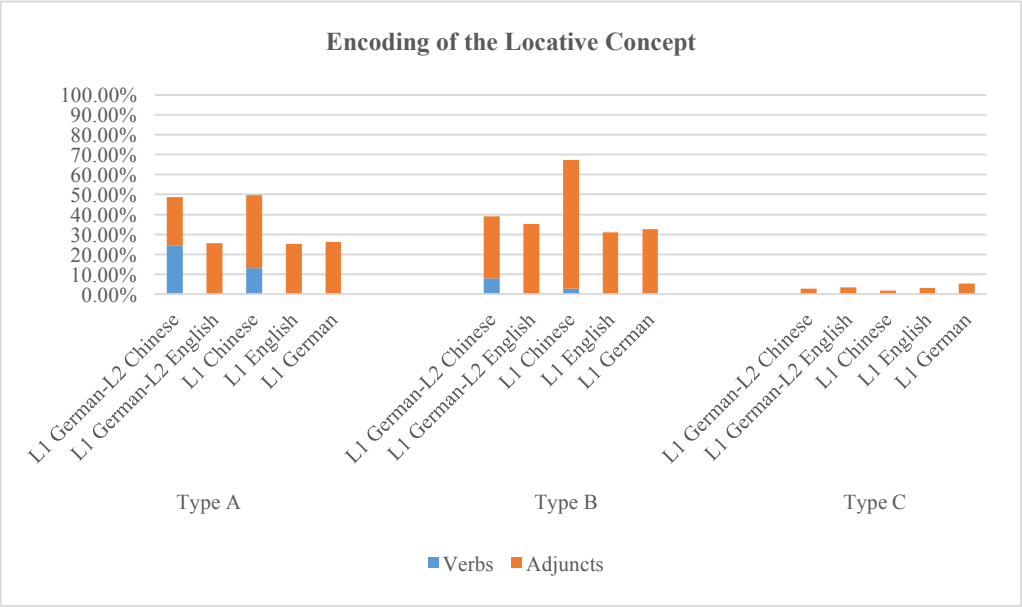


Figure 4.69: Encoding of the locative concept in L1s and L2s in the verbal task of Type A, B, and C videos

Regarding the concept of location encoded in L1, L2 English, and L1 German, no significant relationship was documented between speakers of L1 and L2 English and between speakers of L2 English and L1 German in terms of the total number of encodings in the different video types (all $p > 0.05$). Since no location verbs were used by speakers of these three groups, the total number of encodings actually refers to the use of locative adjuncts. Hence, there was no difference in the use of locative adjuncts between L1 and L2 speakers in the different video types (all $p > 0.05$).

The locative concept mainly appeared in the verbal task of Type A and Type B videos in the two learner groups. When comparing the locative adjuncts used by the two learner groups, it was found that they did not differ in all video types (all $p > 0.05$).

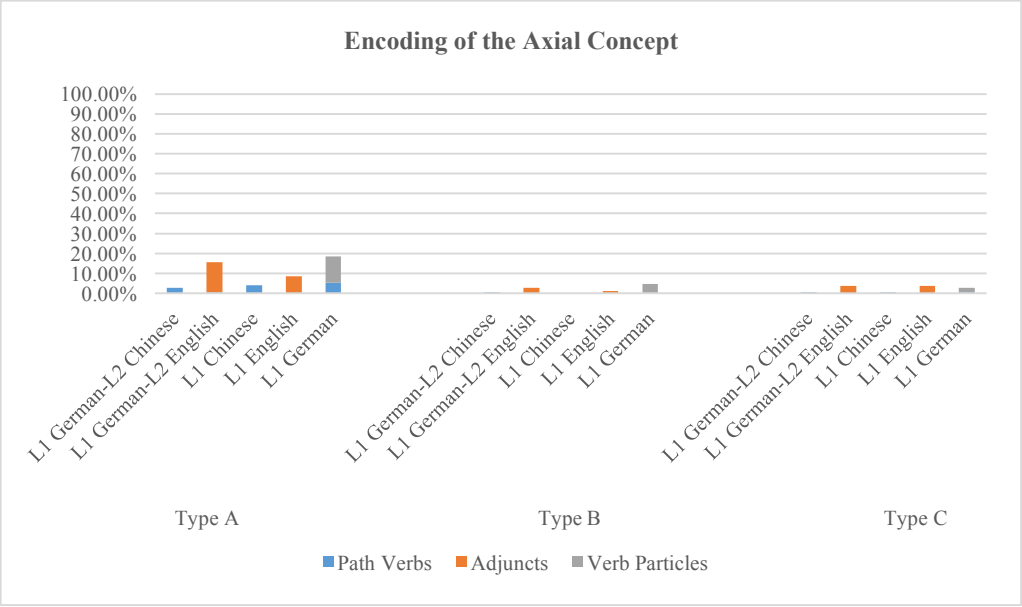


Figure 4.70: Encoding of the axial concept in L1s and L2s in the verbal task of Type A, B, and C videos

The encoding of the axial concept mainly occurred in the verbal task of Type A videos, where one video clip showed a man climbing up ladder to a balcony. No significance was found between L1 and L2 Chinese in encoding the axial concept in the verbal task of Type A videos (Chi-square test: $p = 0.827$), while due to the use of axial verb particles, L1 German speakers encoded it more often than L2 Chinese speakers ($p < 0.001$) (see Figure 4.70). In addition, L1 (i.e., L1 Chinese and L1 German) and L2 speakers did not differ in encoding this concept in path verbs in the verbal task of Type A videos (both $p > 0.05$).

When considering the axial concept encoded in L1, L2 English, and L1 German, L1 German speakers used more axial path verbs than L2 English speakers in the verbal task of Type A videos ($p < 0.05$). L1 (i.e., L1 English and L1 German) and L2 speakers did not differ in the total number of encodings and the use of axial adjuncts in the other cases ($p > 0.05$ in all cases).

The axial concept was mainly encoded in path verbs in L2 Chinese and in adjuncts in L2 English. The occurrence of the axial concept was low in both learner groups. Nevertheless, it was encoded more often in L2 English than in L2 Chinese in all three video types (Chi-square test: $p < 0.05$ in all types).

Figure-related information is packaged in path verbs and its occurrence was low in the different language groups, especially in the verbal task of Type A videos (see Figure 4.71). Nevertheless, a Chi-square test showed no significant effect between L1

and L2 Chinese in the verbal task of Type B videos, while more encodings were found in L2 Chinese (3.76%) than in L1 German (0%). In the verbal task of Type C videos, no significant difference was found between L1 and L2 Chinese ($p = 0.467$) nor between L2 Chinese and L1 German ($p = 0.180$).

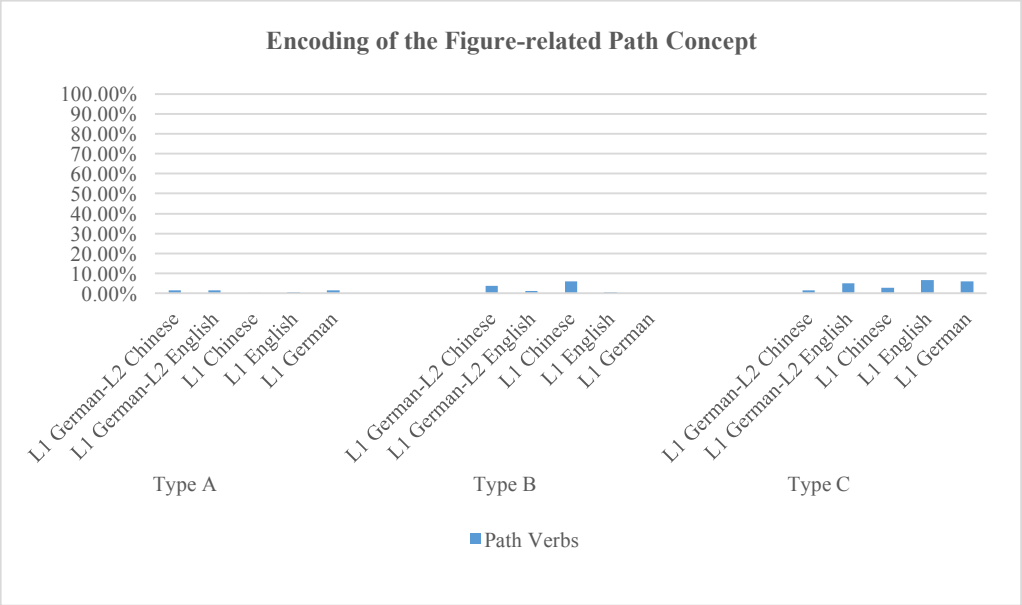


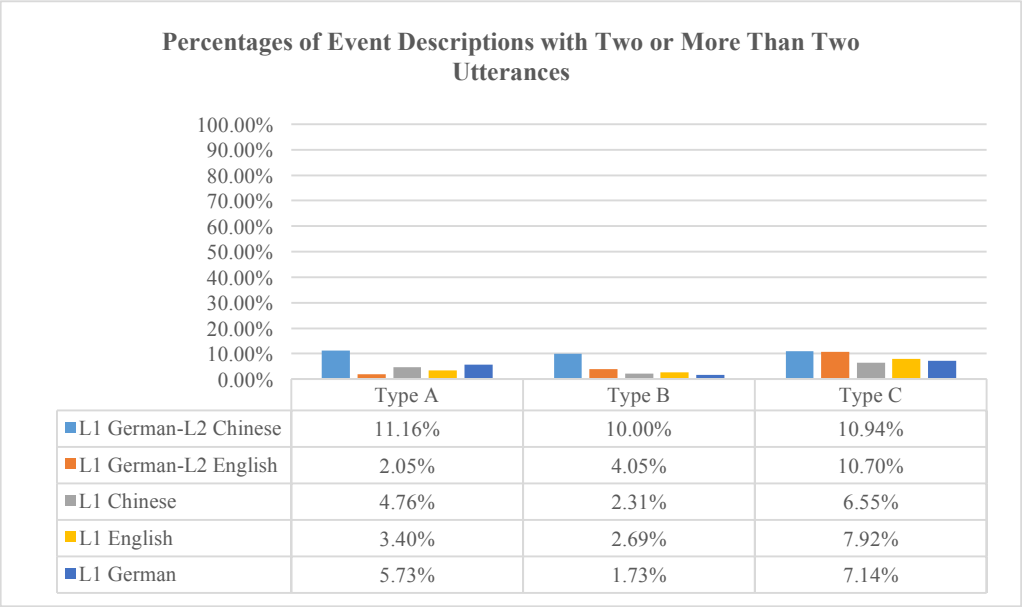
Figure 4.71: Encoding of the figure-related concept in L1s and L2s in the verbal task of Type A, B, and C videos

Regarding encodings of the figure-related concept in L1, L2 English, and L1 German, a Chi-square test revealed no significance between L1 and L2 speakers in the verbal task of Type C videos (both $p > 0.05$). Since its occurrence was rather low in the verbal task of Type A and B videos, no statistical analysis was conducted.

When comparing the use of the figure-related concept in the two learner groups, a significant difference was found in the verbal task of Type B videos. That is, L2 Chinese speakers used it significantly more often than L2 English speakers did ($p < 0.05$), while no difference was found in the verbal task of Type C videos ($p = 0.346$).

4.6.3 Comparison of event descriptions with two utterances

Speakers of L1 German-L2 Chinese and L1 German-L2 English did not differ in producing one utterance when they described motion events (Chi-square test: Type A: $p = 0.685$, Type B: $p = 0.452$, Type C: $p = 0.828$). However, the two learner groups did differ in producing two utterances, as Figure 4.72 shows.



	Type A	Type B	Type C
L1 German-L2 Chinese	11.16%	10.00%	10.94%
L1 German-L2 English	2.05%	4.05%	10.70%
L1 Chinese	4.76%	2.31%	6.55%
L1 English	3.40%	2.69%	7.92%
L1 German	5.73%	1.73%	7.14%

Figure 4.72: Percentages of event descriptions with two or more utterances in L1s and L2s in the verbal task of Type A, B, and C videos

L1 German-L2 Chinese speakers produced two utterances significantly more often than L1 German-L2 English speakers in the verbal task of Type A (Chi-square test: $p < 0.001$) and Type B videos ($p < 0.05$), whereas no significant relationship was documented in the verbal task of Type C videos ($p = 0.876$) (see Figure 4.72). As already discussed in section 4.4.1.3, two video clips of Type A seemed to "trigger" L2 Chinese speakers to produce two utterances: one showing a woman walking around a corner and towards a car and the other one showing a man climbing up the stairs towards the entrance of a building. To describe these two situations, path verbs such as *zhuǎn* 'turn' are typically used for the first situation and *shàng* 'ascend' for the second situation. When these path verbs are used, their combinations with other spatial concepts are syntactically or semantically restricted. So if L2 Chinese speakers want to convey this information, they have to divide the situation into two utterances. In contrast, Path is typically encoded in adjuncts in English, which can be stacked after manner verbs. Hence, L2 English speakers do not need to produce a second utterance.

Furthermore, L2 Chinese speakers produced two utterances more often than L2 English speakers in the verbal task of Type B videos. L2 Chinese speakers were prone to expressing direction in the first clause and then mention the endpoint in the second clause. The long trajectory shown in some video clips of Type B had a big curve. L2 Chinese speakers preferred expressing the direction related to this feature. That is, they expressed directional information related to the features of the Ground object. For example, clause 1: *qìchē wǎng yòu guǎi* 'lit. car towards right turn', clause 2: *dǎsuan*

jìn-rù chéngshì 'lit. be going to enter-enter city'.

In the verbal task of Type C videos, two utterances were likely to be produced when speakers of the two learner groups described the situation 'a dog running across square into a house' and 'a man riding a bicycle and turning into an entrance'. For the former situation, 5 out of 24 L2 Chinese speakers chose to produce two utterances, in which the first clause conveyed either information on Manner und Location (e.g., *yī-ge tíngchē-de dìfāng*, *yǒu yī-tiáo gǒu zài shàng-mian pǎo* 'lit. one-CL parking-MOD place, exist one-CL dog at top-surface run') or directional information (e.g., *yǒu yī-zhī gǒu cóng zuǒ-bian pǎo-dào yòu-bian* 'lit. exist one-CL dog from left-side run-arrive right-side') and the second clause conveyed boundary-crossing information (e.g.,*jìn fángjiān le* 'enter room LE'). By contrast, for this dog scene, only 1 out of 20 L2 English speakers produced two utterances. This is attributable to the fact that speakers of English are relatively free to add locative and directional information after manner verbs within one utterance. It is not necessary for English speakers to produce a second clause. However, when describing the latter situation (i.e., a man riding a bicycle and turning into an entrance), it was found that when the English path verb *turn* was used, L2 English speakers were more likely to utter two clauses. This is because the semantic implication of *turn* is incompatible with the spatial information indicated by some adjuncts such as * *a man is turning along a street*. So speakers have to describe it in two utterances: *a man is riding a bicycle along a street and turning into an entrance*. This finding is in conformance with the findings of Gerwien and von Stutterheim (2018), according to which a second clause is required if a motion involves a change of orientation or direction.

In sum, when path verbs are used, syntactic or semantic restrictions are imposed on their co-occurrence with other spatial concepts. In order to produce appropriate utterances, speakers have to segment the situation and utter two clauses. In this study, L1 German-L2 Chinese speakers were more likely to do so than L2 English speakers. This is because Path information is typically encoded in path verbs in Chinese, while the same information is encoded in adjuncts in English. Different adjuncts can be combined with manner verbs and integrated into one utterance in English, whereas only certain spatial adjuncts can be combined with path verbs in Chinese. However, when path verbs are used in English, especially when a change of orientation or direction occurs, speakers of English may have to utter two clauses, depending on the path verbs actually used.

4.6.4 Comparison of SOTs, fixation counts, and duration of fixations

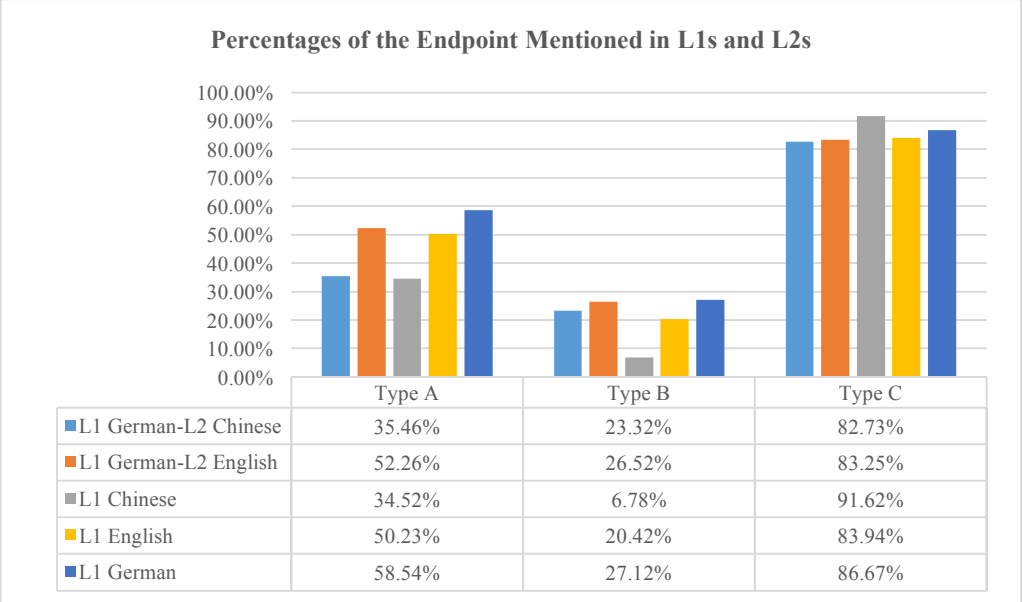
Table 4.20: Average speech onset times and standard deviations in L1s and L2s

	L2 Chinese	L2 English	L1 Chinese	L1 English	L1 German
Type A	3745 ms (SD 1689)	2705 ms (SD 1388)	3394 ms (SD 1923)	2051 ms (SD 818)	2792 ms (SD 1362)
Type B	3521 ms (SD 1616)	2577 ms (SD 1423)	2996 ms (SD 1850)	1893 ms (SD 680)	2596 ms (SD 1383)
Type C	3446 ms (SD 1626)	2662 ms (SD 1364)	3117 ms (SD 1705)	2024 ms (SD 782)	2728 ms (SD 1351)
Mean	3573 ms (SD 1647)	2651 ms (SD 1389)	3175 ms (SD 1832)	1994 ms (SD 768)	2709 ms (SD 1365)

In comparison with L1 German-L2 English speakers, L1 German-L2 Chinese speakers began to talk rather late (see Table 4.20). One-way ANOVA showed that the difference between L2 Chinese and L2 English in SOTs reached a significant level for Type A videos ($F(1, 418) = 46.992$, $p < 0.001$), Type B videos ($F(1, 376) = 36.038$, $p < 0.001$), and Type C videos ($F(1, 418) = 28.369$, $p < 0.001$). In addition, when considering the influence of different video types on SOTs, the two learner groups differed in the order of their SOTs in different video types. According to the duration of SOTs (from longest to shortest), the order of SOTs in L2 Chinese was Type A > Type B > Type C, while the order of SOTs in L2 English was Type A > Type C > Type B. All speakers took the longest to process the information for Type A videos. The reason for this is the short trajectory between the Figure and the Ground object at goal in Type A videos. Speakers had to wait until they could identify whether the Figure was moving towards an endpoint or crossing a boundary. By contrast, for the same situation in Type B videos, where the Ground object at goal was far away from the Figure, L2 Chinese and L2 English speakers behaved differently. In L2 Chinese, SOTs for Type B videos were longer than for Type C videos, while this was reversed in L2 English, i.e., SOTs for Type C videos were longer than for Type B videos. It seems that L2 Chinese speakers needed more time to process the information in Type B videos. As already discussed in section 4.4.2.1, the reason might be related to the language-specific structures in Chinese (i.e., syntactic complexity of serial verb constructions, constraints on the combination of semantic components, choice of aspect marker, and compatibility of serial verb constructions with aspect markers) and the influence of L1's event construal pattern (i.e., focus on the endpoint). Since there are no limitations

on the co-occurrence of manner verbs with different path segments within one utterance and no restrictions on the compatibility of verbs with imperfective or perfective markings in English, the first concern, i.e., language-specific rules as in Chinese, do not exist in English. It should be examined whether the second concern, i.e., L1's event construal pattern, plays the same role in L2 Chinese and L2 English; in other words, whether each learner group had later SOTs than the respective target language. In the following, we will first examine how the endpoint was mentioned linguistically by the two learner groups and then compare the fixation patterns.

As shown in Figure 4.73 below, concerning the endpoint being encoded linguistically, the statistical results did not show any significant difference between the two learner groups for different video types (Type A: $\chi 2$ (1, N = 193) = 1.166, p = 0.280, Type B: $\chi 2$ (1, N = 100) = 0.160, p = 0.689, and Type C: $\chi 2$ (1, N = 356) = 0.180, p = 0.672). As already discussed in section 4.4.2.2, L1 German-L2 Chinese speakers mentioned the endpoint more often than L1 Chinese speakers in the verbal task of Type B videos ($\chi 2$ (1, N = 64) = 25.000, p < .001), while no significant difference was found between L2 English and L1 English ($\chi 2$ (1, N = 87) = 0.931, p = 0.335).



	Type A	Type B	Type C
L1 German-L2 Chinese	35.46%	23.32%	82.73%
L1 German-L2 English	52.26%	26.52%	83.25%
L1 Chinese	34.52%	6.78%	91.62%
L1 English	50.23%	20.42%	83.94%
L1 German	58.54%	27.12%	86.67%

Figure 4.73: Bar charts showing the percentages of the endpoint mentioned in L1s and L2s in the verbal task of Type A, B, and C videos

Fixation on the endpoint

For Type A videos, as shown in Figure 4.74, one-way ANOVA showed that the total number of fixation counts on the endpoint was found to be higher in L2 English than in L2 Chinese (F (1, 383) = 5.059, p < 0.05). In contrast, the number of fixation

counts before SOTs (F (1, 357) = 5.677, p < 0.05), the total duration of the fixations (F (1, 383) = 4.103, p < 0.05), and the duration of the fixations before SOTs (F (1, 357) = 9.164, p < 0.05) were found to be higher in L2 Chinese than in L2 English, respectively, for Type A videos. Interestingly, no significant relationship was documented between the two learner groups in all respects for Type B and C videos, neither regarding the total number of fixation counts (Type B: F (1, 345) = 0.761, p = 0.384; Type C: F (1, 325) = 0.010, p = 0.920) and fixation counts before SOTs (Type B: F (1, 251) = 1.795, p = 0.182; Type C: F (1, 327) = 0.460, p = 0.498), nor in the total duration of the fixations (Type B: F (1, 344) = 1.772, p = 0.184; Type C: (F (1, 325) = 0.587, p = 0.444) and the duration of the fixations before SOTs (Type B: F (1, 251) = 0.001, p = 0.973; Type C: F (1, 327) = 0.004, p = 0.950).

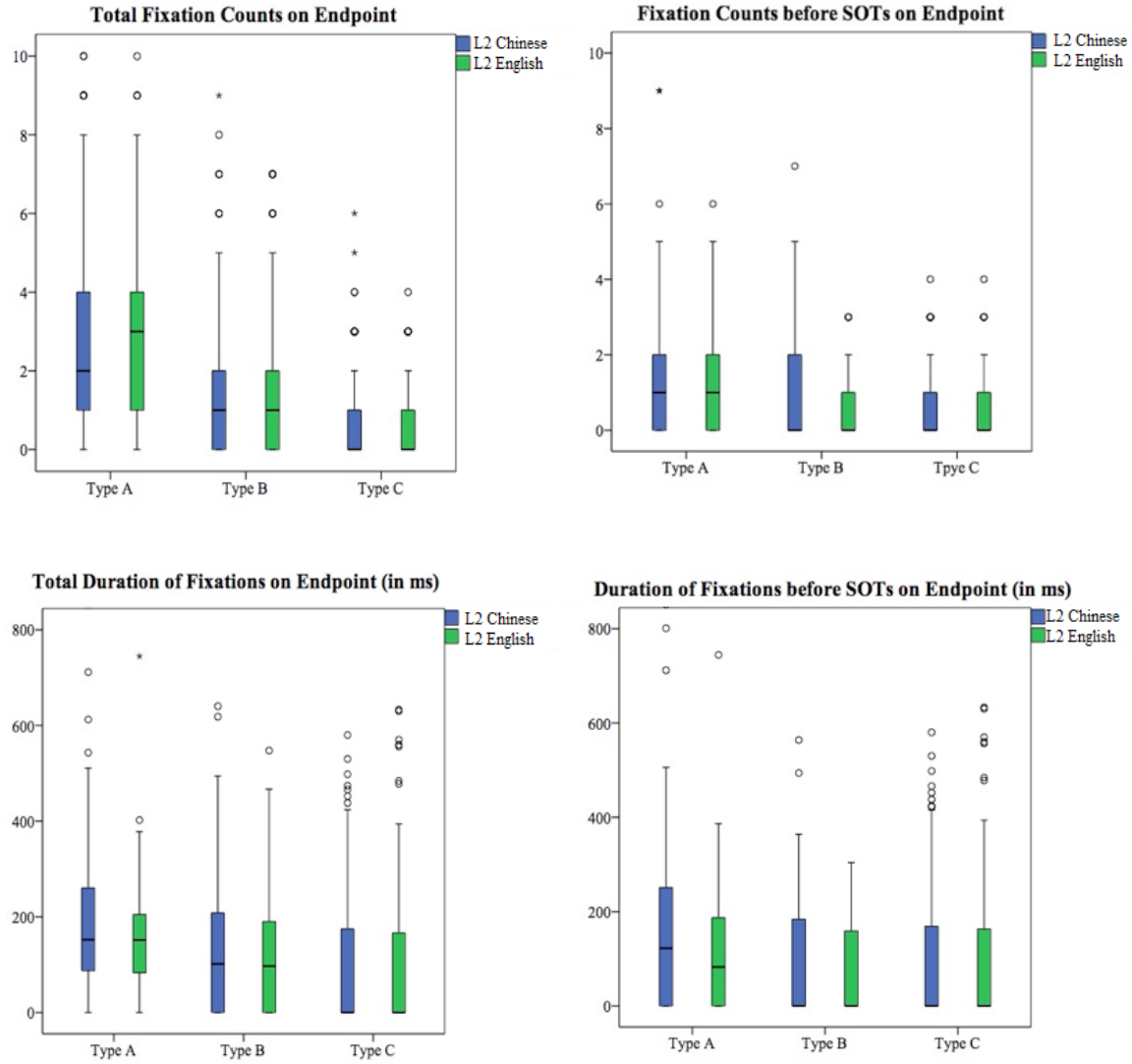


Figure 4.74: Number and duration of fixations on the endpoint in total and before SOTs in L1s and L2s in Type A, B, and C videos

Fixation on the moving entity (Figure)

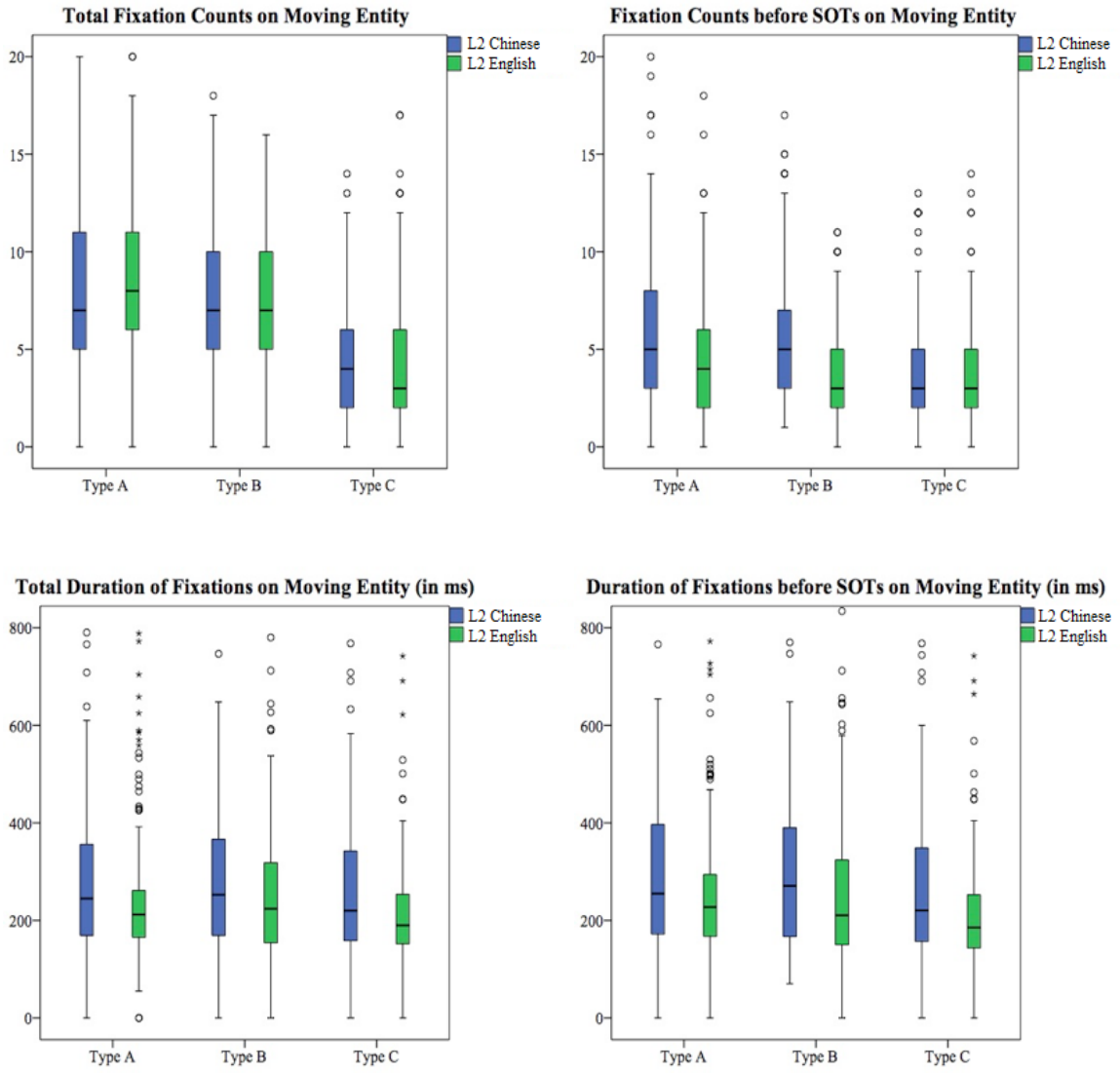


Figure 4.75: Number and duration of fixations on the moving entity in total and before SOTs in L1s and L2s in Type A, B, and C videos

When comparing the fixation on the moving entity (Figure) between the two learner groups, as shown in Figure 4.75, it was interesting to find that the fixation patterns (i.e., the number and duration of the fixations) for Type A and Type B videos showed the same tendency. That is, one-way ANOVA showed no significant difference between two learner groups with respect to the total number of fixation counts for Type A ($F (1, 383) = 1.474$, $p = 0.225$) and Type B videos ($F (1, 344) = 0.047$, $p = 0.829$), respectively, while the number of fixation counts before SOTs was found more often in L2 Chinese than in L2 English for Type A ($F (1, 357) = 15.592$, $p < 0.001$) and Type B videos ($F (1, 309) = 15.276$, $p < 0.001$), respectively. In contrast, the total duration

of fixations on the moving entity was found more often in L2 Chinese than in L2 English for Type A ($F (1, 383) = 8.586$, $p < 0.05$) and Type B videos ($F (1, 344) = 5.990$, $p < 0.05$), respectively, whereas with respect to the duration of the fixations before SOTs, no significance was found for Type A ($F (1, 357) = 2.793$, $p = 0.096$) nor for Type B videos ($F (1, 309) = 2.192$, $p = 0.140$). Furthermore, no significance was found with respect to the number of fixations in total ($F (1, 325) = 0.484$, $p = 0.487$) and before SOTs ($F (1, 327) = 1.084$, $p = 0.299$) for Type C videos. By contrast, the duration of the fixations in total and before SOTs was found to be longer in L2 Chinese than in L2 English (one-way ANOVA: $p < 0.05$ in both cases).

Summary and interpretation

In sum, the total number of fixation counts was found to be higher in L1 German-L2 English than in L1 German-L2 Chinese for Type A videos, which is in line with the percentages of the endpoint being encoded linguistically in the two learner groups in the verbal task of Type A videos (L2 English 52.26% vs. L2 Chinese 35.46%, see Figure 4.73). However, the difference in the endpoint being mentioned linguistically did not reach a significant level between the two learner groups (Chi-square test: $p = 0.280$). The statistical analysis also showed that in comparison with L2 English speakers, L2 Chinese speakers tended to fixate more on the endpoint and on the moving entity before SOTs than L2 English speakers did for Type A videos. This is to say, L2 Chinese speakers directed more attention both to the endpoint and to the moving entity in the conceptualization phase. This might reflect that L2 Chinese speakers were "struggling" to select which information to verbalize and how to structure it. Typically, speakers had options to conceptualize the situation. For Type A videos, they could conceptualize the situation as "A Figure is moving in a certain manner", "A Figure is moving in a certain manner at a certain location", "A Figure is moving in a certain manner towards a Ground object at goal", or "A Figure is moving in a certain manner at a certain location towards a Ground object at goal". L2 Chinese speakers had to be very careful when they chose the last option to conceptualize the situation in Type A videos, since the locative information might not be compatible with directional information in Chinese. In Chinese, locative information is encoded in adjuncts or in a verb compound like *zǒu-lù* (lit. 'walk-road'), which indicates both Manner and locative information (see Figure 4.69). Locative adjuncts might not occur with serial verb constructions (e.g., ? *zài lù-shàng zǒu-guò-qù* 'lit. at road-top walk-cross-go'), and the verb compound *zǒu-lù* cannot co-occur with directional adjuncts

(e.g., ? *wǎng yī-liàng chē zǒu-lù* 'lit. towards one-CL car walk-road'). Encoding locative information might require speakers to focus on the moving entity, while encoding directional information might require speakers to direct attention to the endpoint. Hence, the more frequent fixation on the endpoint and the moving entity in L2 Chinese might reflect that they had to think about which information to verbalize. Also, it took them longer to make this decision in the conceptualization phase, so the duration of the fixations on the endpoint before SOTs was longer than for L2 English speakers. These complex cognitive processes in L2 Chinese might have led to later SOTs for Type A videos compared to the SOTs in L2 English.

Regarding Type B and C videos, speakers of L1 German-L2 Chinese and L1 German-L2 English did not differ in encoding the endpoint at the linguistic level (see Figure 4.73), which is in line with the findings regarding the allocation of attention to the endpoint between these two learner groups for Type B and C videos. They did not differ in the number and duration of the fixations on the endpoint in total and before SOTs. However, in Type B videos, L1 German-L2 Chinese speakers fixated more often on the moving entity before SOTs than L1 German-L2 English speakers; also, the total duration of the fixation on the moving entity was longer in L2 Chinese than in L2 English. In addition, for Type C videos, the duration of the fixations on the moving entity in total and before SOTs was longer in L2 Chinese than in L2 English as well. That is, during the conceptualization phase, L2 Chinese speakers directed more attention to the moving entity in Type B videos and fixated on it longer, on average, in Type C videos.

As already discussed in sections 4.4.2.2 and 4.5.2.2, speakers of L2 Chinese and L2 English did not differ from L1 German speakers with respect to the number and duration of the fixations before SOTs for Type B videos. In contrast, L2 Chinese speakers had more fixations on the endpoint before SOTs than L1 Chinese speakers, while L2 English speakers did not differ from L1 English speakers in both regards for Type B videos. However, L1 German speakers had more fixations on the endpoint before SOTs than L1 Chinese and L1 English speakers, respectively. In other words, during the conceptualization phase in Type B videos, L1 German speakers' habitual way of event construal (i.e., focusing on the endpoint) played a role both in L2 Chinese and in L2 English. Since both learner groups focused on the endpoint, they tended to wait until the endpoint became identifiable. The SOTs in both learner groups were thus longer than in the target languages (i.e., Chinese and English), respectively (see Table

4.20). In addition, in comparison with SOTs in L2 Chinese, the SOTs in L2 English were shorter for Type B videos. There might be two reasons for this difference between the two learner groups. First, as already explained above, L2 Chinese speakers had to think about which information to verbalize because they must pay attention to the compatibility of different semantic components as well as the capacity of co-occurrence of aspect markers with serial verb constructions in Chinese, so their cognitive load was heavier in L2 Chinese, which led to rather late SOTs. Second, L2 English speakers did not differ greatly from L1 English speakers in their fixation and the duration of the fixation on the endpoint during the conceptualization phase, whereas this was not the case in L2 Chinese. In other words, L2 English speakers went further towards the target language than L2 Chinese speakers did. L1 English speakers tended to have early SOTs for Type B videos, as did L2 English speakers. Hence, the SOTs in L2 English were shorter than in L2 Chinese.

4.7 Summary

In this study, we have analyzed the linguistic means for encoding different semantic components and motion event cognition in L1s and L2s. The findings show the joint impact of typological and cognitive factors on adult second language acquisition.

Typologically speaking, Chinese and English belong to different language types. Chinese is an E-language. In this type of languages, both Manner and Path information are typically encoded in verbs. In contrast, English and German are S-languages, in which Manner information is typically encoded in verbs and Path in satellites, i.e., adjuncts in our analysis. In general, speakers of L1 German-L2 Chinese and L1 German-L2 English had learned the linguistic forms and their corresponding functions in Chinese and English, respectively. L2 Chinese speakers encoded both Manner and Path information in verbs, while L2 English speakers encoded Manner information in verbs and Path in adjuncts. This is in conformance with the typological contrasts between S-languages and E-languages. The statistical analysis showed that speakers of L2 Chinese and L2 English did not differ in the use of manner verbs, whereas speakers of the two groups differed in the use of path verbs. L2 Chinese speakers used more path verbs than L2 English speakers. In particular, L2 Chinese speakers used more path verbs encoding deictic, goal-oriented, and boundary-crossing information than L2

English speakers. By contrast, L2 English speakers used more adjuncts encoding object-related, directional, and boundary-crossing information than L2 Chinese speakers. In addition, the two learner groups did not differ in the total number of encodings of locative and boundary-crossing information (encoded in verbs, adjuncts, or verb particles) in all video types. However, the encoding of locative information was found more often in L1 Chinese than in L2 Chinese in the verbal task of Type B videos, while no difference was found between L1 and L2 English in this regard. L1 Chinese speakers tended to conceptualize the situation in Type B videos as "A Figure moving in a certain manner at a certain location", while L2 Chinese speakers tended to encode directional and dynamic information for the same situation in Type B videos. Furthermore, the encoding of boundary-crossing information was found more often in L2 Chinese than in L1 Chinese and also more often in L2 English than in L1 English for Type B videos. Although Type B videos showed a long trajectory with a possible endpoint rather than a boundary-crossing situation, it seems that the endpoint played an important role in construing the situation in Type B videos for the two learner groups because in order to encode boundary-crossing information, the endpoint has to be encoded. In other words, L1 German speakers' habitual conceptual patterns (i.e., the tendency to take a holistic perspective and focus on the endpoint) played a role when speakers of L1 German-L2 Chinese and L1 German-L2 English described motion events.

Regarding the frequency of occurrence of two utterances, L2 Chinese speakers produced two utterances more frequently than L2 English speakers in the verbal task of Type A and B videos, while no difference was found between the two learner groups in the verbal task of Type C videos. The reason for the increased occurrence of two utterances in L2 Chinese lies in the fact that Path information is typically encoded in path verbs in Chinese. The syntactic and semantic features of path verbs restrict the co-occurrence of path segments. When L2 Chinese speakers want to express the same amount of information as L2 English speakers, they have to divide the situation into two parts, especially when the situation shows a change of the Figure's orientation or direction (Gerwien & von Stutterheim, 2018). Further analysis of the use of two utterances in L2 Chinese showed that L2 Chinese speakers tended to encode directional information in the first clause and mention the endpoint in the second clause, whereas when L1 Chinese speakers uttered two clauses, they tended to encode Manner information or Manner with locative information in the first clause and mention the

endpoint in the second clause.

The statistical analysis of SOTs showed that both learner groups had later SOTs compared to the target languages. As already discussed in sections 4.4.2.1 and 4.5.2.1, given that the descriptions were not native-like in both learner groups, an L2 effect cannot be excluded. However, except L2 effect there may be other factors that led to late SOTs in the two learner groups. When the characteristics of the different video types were taken into account, it was found that the order of SOTs from longest to shortest in L2 Chinese was Type A > Type B > Type C, whereas the order of SOTs in L2 English was Type A > Type C > Type B. The two learner groups differed in their SOTs for Type B videos. L2 Chinese speakers had later SOTs than L2 English speakers for Type B videos. According to the analysis of the encodings of spatial concepts, L2 Chinese speakers tended to encode deictic, goal-oriented, and boundary-crossing concepts, while L2 English speakers encoded slightly more boundary-crossing concepts than L1 English speakers. To encode those concepts, learners of the two groups needed to pay attention to the endpoint to a certain degree, so when they watched the scenes, they had to wait until the endpoint became clear and then they began to talk. Hence, in comparison with each target language, both learner groups had late SOTs. This is to say, L1 German speakers' conceptualization patterns had an effect when L2 speakers described motion scenes. In addition, the reason that L2 Chinese speakers had longer SOTs than L2 English speakers might reside in the language-specific rules in Chinese: the syntactic complexity of serial verb constructions (see section 4.3.1.1), the compatibility of different path segments within one utterances (see section 4.4.1.3), the choice of aspect markers and the compatibility of aspect markers with serial verb constructions (see Chapter 5). All of these language-specific rules might have led to the late SOTs observed in L2 Chinese.

Concerning the endpoint mentioned in the verbal task and the fixation patterns, L1 and L2 speakers displayed great contrasts for Type B videos. Speakers of L2 Chinese and L2 English did not differ in mentioning the endpoint in the verbal task in all video types. However, L2 Chinese speakers mentioned the endpoint more often than L1 Chinese speakers in the verbal tasks of Type B videos. The eye-tracking data showed that for Type B videos, the number of fixation counts on the endpoint in total and before SOTs was found to be higher in L2 Chinese than in L1 Chinese, whereas no difference was found between L2 Chinese and L1 German in both regards. In addition, the duration of the fixations on the endpoint both in total and before SOTs for Type B

videos was found to be longer in L1 German than in L1 Chinese, whereas no difference was found between L2 Chinese and L1 German nor between L2 Chinese and L1 Chinese. That is, L2 Chinese speakers' duration of the fixations on the endpoint resides in the middle of the scale, with L1 German speakers, who fixated longer on the endpoint, on one end of this scale and L1 Chinese speakers, whose fixation on the endpoint was shorter, on the other end. To summarize, regarding the number and duration of the fixations for Type B videos, L2 Chinese speakers paid much more attention to the endpoint in Type B videos, especially during the conceptualization phase. This finding confirms the above explanation for the late SOTs in L2 Chinese and is in conformance with the results showing that L2 Chinese speakers mentioned the endpoint more often in the verbal task of Type B videos.

The eye-tracking data further showed that speakers of L2 English and L1 German did not differ in the number of fixation counts on the endpoint in total and before SOTs for Type B videos. By contrast, L1 German speakers had more fixations on the endpoint before SOTs than L1 English speakers, while no difference was found between L2 English and L1 English. In other words, if we have a ruler in hand, L1 German speakers, who paid more attention to the endpoint, are put on the right side of the ruler, while L1 English speakers, who paid relatively less attention to the endpoint, are put on the left side of the ruler. L2 English speakers would be located in the middle of the ruler: they did not differ from L1 German speakers and also did not differ from L1 English speakers in this regard during the conceptualization phase. That is, L2 English speakers kept the L1 German speakers' habitual conceptual patterns while at the same time trying to follow L1 English speakers' conceptualization patterns. It therefore appears that L1 German speakers' way of conceptualizing the situation in Type B videos played a role when L2 English speakers described motion events.

When comparing the fixation patterns between L2 Chinese and L2 English, it was found that the two learner groups did not differ in the number and duration of the fixations on the endpoint in total and before SOTs for Type B and C videos, while speakers of the two learner groups differed for Type A videos in both regards. For Type A videos, the total number of fixations was found to be higher in L2 English than in L2 Chinese. Accordingly, the percentages of the endpoint being mentioned in the verbal task were higher in L2 English than in L2 Chinese, although this difference did not reach a significant level. By contrast, the number of fixations on the endpoint before SOTs and the duration of the fixations in total and before SOTs for Type A videos were

found to be higher in L2 Chinese than in L2 English. In addition, the number of fixations on the moving entity before SOTs in Type A videos was found to be higher in L2 Chinese than in L2 English. That is, it seems that L2 Chinese speakers tended to fixate on the moving entity and on the endpoint in Type A videos. L2 Chinese speakers might have thought about which information to verbalize and how to structure this information in describing Type A videos because Type A videos could be conceptualized as "A Figure moving in a certain manner", "A Figure moving in a certain manner at a certain location", or "A Figure moving in a certain manner towards a Ground object at goal". Unlike English and German, path segments cannot be freely combined with manner verbs in Chinese. So L2 Chinese speakers must first decide on which information to verbalize. The fact that they fixated more often on the endpoint and on the moving entity might reflect that they were 'struggling' with the selection and structuring of information.

L1 event conceptualization patterns are language-specific. They are habitual and, to some extent, automated in grammaticalized categories (von Stutterheim et al., 2012; von Stutterheim & Nüse, 2003). The L1 event conceptualization patterns indicate the principles of information organization, i.e., perspective-taking, in different languages. These principles might be subtle and difficult to change. When L2 learners speak an L2 that differs from their L1's conceptualization patterns, then they may have difficulties in adjusting themselves to it. It should be pointed out that different event construal patterns only demonstrate different preference patterns (Bylund & Jarvis, 2011; von Stutterheim & Nüse, 2003). They are not absolute principles like grammar rules, which require speakers to obey them obligatorily. So these principles do not block English and Chinese speakers from mentioning the endpoint, as evidenced in this study. In addition, in comparison to L2 Chinese speakers, it seems that L2 English speakers went more towards the target language. This difference might lie in the typological distance between the source and target language.

Chapter 5 Temporal Analysis

Motion events involve a Figure's change in location over time. Hence, besides spatial concepts, temporal concepts are also fundamental in construing motion events. The stimuli used in this study (for more details, see Chapter 3) were designed according to spatial information, i.e., motion events with various kinds of Manner information directed or not directed towards a place at goal, as well as different temporal phases of motion events, i.e., inchoative, intermediate, and final phase. In this chapter, we will analyze the encoding of motion events between L1 and L2 speakers from a temporal perspective.

Languages differ in the linguistic devices used to anchor an event in time. Linguistic devices may be grammaticalized as tense or aspect[72], or may be lexicalized in different languages. For those languages that have no inflectional morphology (e.g., Chinese), tense and aspect do not exist in the grammatical system. This does not mean, however, that such languages cannot express time; rather, they use lexical means, e.g., particles or adverbials, to express an on-going or completed event as well as the past, the present, and the future.

This chapter mainly deals with the following questions:

1 Whether and to what extent have L2 speakers learned to use aspectual marking in L2s?

(1a) Whether and to what extent have L1 German-L2 Chinese speakers learned to use different aspect markers in Chinese?

(1b) Whether and to what extent have L1 German-L2 English speakers learned to use aspectual marking in English?

(1c) Whether and to what extent do L1 German-L2 Chinese speakers differ from L1 German-L2 English speakers in the use of aspectual marking?

2 Chapter 4 has shown that differences in the fixation patterns were found between L1 and L2 speakers. A further question in this regard is: What happens to visual attention over the course of time in relation to aspectual marking or SOTs?

[72] According to Comrie (1976) and Croft (2012), the term 'aspect' can be understood from two levels, that is, the semantic and the formal level. At the semantic level, aspect makes a distinction between perfective and imperfective meaning, regardless of whether this is realized in lexical or grammatical items, while at the formal level, aspect refers to the language-specific forms that encode aspectual information; for example, the aspect marker *zài*, *zhe*, *le*, *guò* in Chinese and the periphrastic expression be-ing in English. In this study, the term 'aspect' refers to the former. That is, it indicates the semantic distinctions between perfective, imperfective, and prospective meaning.

To answer these questions, the linguistic forms encoding temporal concepts in English, Chinese, and German will be introduced first. Following that, the use of aspect markers in L1 and L2 Chinese and the use of progressive marking in L1 and L2 English will be compared. After that, the allocation of visual attention to the moving entity and the endpoint over the course of time in the different languages will be presented.

5.1 Temporal concept encoded in English, German, and Chinese

Tense and aspect are two grammatical categories of verbs[73]. As already discussed in Chapter 2.4, according to Klein (1994, 2009), tense[74] can be characterized as the relation between topic time (TT) and time of utterance (TU), while aspect[75] can be defined as the relation between topic time (TT) and time of situation (Tsit). Furthermore, depending on the presence or absence of TT-contrast, lexical contents can be categorized as 0-state lexical contents, 1-state lexical contents, and 2-state lexical contents (for more details, see Chapter 2). In the following, we will present how these temporal concepts are encoded in English, German, and Chinese, respectively.

5.1.1 English

English has three tense forms: present, past, and future tense. Klein (1994, p. 124) showed the temporal relations between TU and TT for each of these three tense forms as follows:

PRESENT tense: TU INCL TT
PAST tense: TU AFTER TT
FUTURE tense: TU BEFORE TT

[73] There are numerous studies on tense and aspect in different languages (e.g., Comrie, 1976; Croft, 2012; Dahl, 1985; Klein, 1994; Smith, 1997; Vendler, 1967). As Croft (2012) pointed out, the terminology within this domain is confusing. Different scholars differ in their theories on the conceptual properties for aspectual analysis, so the terms they use bring about confusion. The aim of this study is to describe the aspectual distinctions in English, German, and Chinese and the use of the aspectual marking in L1s and L2s. To this end, we find that Klein's (1994, 2009) framework provides a simple and clear way to describe the aspectual systems in different languages. Therefore, in this study, the focus is on Klein's framework, while theories by other researchers will not be discussed.

[74] Comrie (1976) claimed that tense is about the relation between the time of situation and the time of utterance, i.e., the relation between 'before, after and simultaneous'. However, Klein (1994, 2009) argued against this point of view and provided examples to point out the insufficiency of Comrie's definition (for more details, see Klein, 1994, pp. 22-23; Klein, 2009, pp. 42-51). He further proposed that tense is the temporal relation between topic time and time of utterance.

[75] In canonical studies on aspect, aspect is defined as "different ways of viewing the internal temporal constituency of a situation" (Comrie, 1976, p. 3) or as "reference to one of the temporally distinct phases in the evolution of an event through time" (Johnson, 1987, p. 152). However, these definitions are not unproblematic (for more details, see Klein, 1994, 2009).

Here are examples for each of the above-mentioned tense forms:

(1) PRESENT tense: John walks towards a bus stop.
PAST tense: John walked towards a bus stop.
FUTURE tense: John will walk towards a bus stop.

These tense forms can be realized through inflectional endings, e.g., '-s' in *walks* for the present tense, '-ed' in *walked* for the past tense, and periphrastic construction, e.g., *will* in *will walk*, for the future tense.

In addition, in English, aspect can be characterized as having three aspectual relations between TT and Tsit; to be more precise, the relations between TT and the distinguished phase (abbreviated as DP). According to Klein (2000), DP is the only state in the case of a 1-state expression; DP is either the source state or the target state in the case of a 2-state expression. Whether the source state or the target state is selected as DP is language-specific. Klein (2000, pp. 751-54) argued that in the case of 2-state expressions, DP is the source state in English and the target state in Chinese because English is more 'action-oriented' and Chinese more 'result-oriented' (cf. Chu, 1976; Li, 1990; Yong, 1997). These three aspectual relations in English can be represented as follows (Klein, 2000, p. 751):

IMPERFECTIVE: TT IN T-DP
PERFECTIVE: TT OVL T-DP and POSTTIME OF T-DP
PERFECT: TT AFTER T-DP

The imperfective is expressed by the progressive form, the perfective by the simple form, and the perfect by the perfect form, respectively. These different aspectual relations work in situations expressed by lexical contents. Based on Klein's (1994) analysis of situation types, in the case of motion events, such situations can be characterized as 1-state situation or 2-state situation (corresponding to 1-state and 2-state lexical contents). The former refers to a directed motion towards a place (e.g., *a man is walking towards a car*) or an undirected motion (e.g., *a car is driving along a road*), while the latter means that there are two states, i.e., a source state and a target state, which differ in their duration in different expressions: The verb *enter* in the expression *a woman enters a supermarket* describes a transition from a source state to a target state, while in the expression *a dog runs into a house*, the source state has

duration (von Stutterheim et al., 2017). The interaction between grammatical aspect and situations expressed by lexical contents can be illustrated as follows. In the diagrammatic representation, the distinguished phase is represented as +++++, the TT as [ ], and the target state as ----.

1-state situation

(2) IMPERFECTIVE: A man is walking towards a car. +++[++++]+++
PERFECTIVE: A man walked towards a car. +++[++++]
PERFECT: A man has walked towards a car. +++++++ [ ]

2-state situation

(3) IMPERFECTIVE: A dog is running into a house.

```
+++[++++]+++----------
source state      target state
not in house      in house
```

PERFECTIVE: A dog ran into a house.

```
+++++++[+++------]----------
source state      target state
not in house      in house
```

PERFECT: A dog has run into a house.

```
++++++++++---[------]-------
source state      target state
not in house      in house
```

5.1.2 German

Most grammar books list six tense forms in German: future I, future II, present, past, perfect, and past perfect tense. Klein (1994, pp. 126-29) describes these tense forms as follows:

FUTURE I tense: TU BEFORE TT
PRESENT tense: TU INCL TT
PAST tense: TU AFTER TT
PERFECT tense: TU AFTER TT
PAST PERFECT tense: TU AFTER TT and TT AFTER Tsit
FUTURE II tense: TU BEFORE TT and TT AFTER Tsit

Klein (1994, p. 129) argued that tense meaning and aspectual meaning are combined in the past perfect and future II tense in German. In addition, the past tense in German can be replaced by the perfect tense. The past tense is "considered to be the normal form of literary narration" (Klein, 1994, p. 128), while the perfect tense is commonly used in the spoken language. So PERFECT does not necessarily refer to the meaning of perfect. It can also refer to a situation that occurred in the past. Consider the following examples for each of the above-mentioned tense forms:

(4) FUTURE I tense: Ich werde Chinesisch lernen. (I will learn Chinese)
PRESENT tense: Ich lerne Chinesisch. (I learn Chinese)
PAST tense: Ich lernte Chinesisch. (I learned Chinese)
PERFECT tense: Ich habe Chinesisch gelernt. (I have learned Chinese)
PAST PERFECT tense: Ich hatte Chinesisch gelernt. (I had learned Chinese)
FUTURE II tense: Ich werde Chinesisch gelernt haben. (I will have learned Chinese)

Generally speaking, the distinctions between the imperfective and perfective aspect are not systematically grammaticalized in German[76], but they can be encoded in periphrastic constructions like *sein + am/beim + Verb* or *sein + dabei ...zu + Verb* to express an ongoing activity (Ebert, 2000) and the temporal adverbials *schon/bereits* 'already' can be used to express the perfective aspect, as the following examples illustrate:

(5) Ich bin am/beim Lesen. (I am at the reading)
Ich bin dabei, ein Buch zu lesen. (I am in the course, a book to read)
Ich habe das Buch schon gelesen. (I have already read the book)

In our analysis, however, no data was found where the above-mentioned periphrastic constructions were used to express ongoingness in German.

[76] Klein (1994, p. 128) argued that in an appropriate context, the perfect in German can have an aspectual function. To explain this function, the author gave the example *Hans hat schon gegessen*. (lit. Hans hat already eaten) for the explanation why Hans refused to have a meal with his friend. In this case, the perfect in German has the same function as the present perfect in English.

5.1.3 Chinese

Unlike English and German, Chinese does not have inflectional morphology on the verb to mark tense and aspect. Instead, the temporal reference that is encoded in tense in English and German is expressed by virtue of temporal adverbials and aspect is expressed by temporal particles or aspect markers. There are four aspect markers in Chinese: *zài*, *zhe*, *le*, and *guo*[77]. The basic functions of these four aspect markers are described by Li and Thompson (1981, pp. 185-37) as follows[78]:

> - The verbal aspect suffix -*le* expresses perfectivity, that it, it indicates that an event is being viewed in its entirety or as a whole.
> - There are two aspect markers that signal the durative nature of an event: the word *zài* and the suffix -*zhe*. The usage of the durative markers in a sentence depends on the meaning of the verb.
> - The aspect suffix -*guo* means that an event has been experienced with respect to some reference time.

Syntactically, the aspect marker *zài* precedes the verb, while the other three aspect markers follow the verb. Based on time-relational analysis, Klein (2000, p. 754) argued that the aspect markers in Chinese can be represented as follows[79]:

a. *le* TT OVL PRETIME T-DP AND T-DP
b. *guo* TT AFTER T-DP
c. *zai* TT IN T-DP
d. *zhe* TT IN T-DP

As already mentioned in section 5.1.1, in comparison with English, the distinguished phase (DP) in Chinese is the target state. The interaction between aspect and the situations expressed by lexical contents can be better understood through the following examples. In what follows, DP is represented as +++, source state as ----, and TT as [ ].

---

[77] Apart from these four aspect markers, some researchers have considered -*ne* as an imperfective aspect marker that may indicate an ongoing action or the duration of a state. However, its use is restricted to dialogs in colloquial Chinese (e.g., Liu, 1985) or in northern dialects (e.g., Ma, 1987). Also, this aspect marker does not occur in the collected data, so it will not be discussed further.

[78] Similarly, Smith (1997) characterized the functions of the aspect markers in Chinese as follows: "-*le* spans the initial and final endpoints of an event, while the span of -*guo* extends beyond the final endpoint of a situation" (pp. 263-64); "*zài* presents an internal interval of a durative situation, and often has the connotations of activity associated with events" (p. 271); "-*zhe* expresses an imperfective viewpoint that presents a continuous and stable situation without regard to endpoints" (p. 273).

[79] Although both *zài* and *zhe* are imperfective aspect markers, they do not have exactly identical aspectual meanings. *zài* indicates that an activity is in progress, while *zhe* signals the continuation of a situation. *zhe* with this property can mark a situation accompanying another situation (Li & Thompson, 1981).

1-state situation

(6a) zhāngsān kāi le yī-liàng chē. [+++++++++]
zhāngsān drive-LE one-CL car drive car
zhāngsān drove a car.

(6b) zhāngsān kāi guo yī-liàng chē. ++++++++++++ [ ]
zhāngsān drive GUO one-CL car drive car
zhāngsān has driven a car.

(6c) zhāngsān zài kāi yī-liàng chē. ++++[++++++++]++++
zhāngsān ZAI drive one-CL car drive car
zhāngsān is driving a car.

(6d) zhāngsān kāi zhe yī-liàng chē. ++++[++++++++]++++
zhāngsān drive ZHE one-CL car drive car
zhāngsān is driving a car.

2-state situation

As already shown in section 4.3.1.1, the resultative verb compounds[80] in Chinese include two consecutive verb stems, with the first verb stem expressing the source state and the second the target state, e.g., *chi-wan* 'eat-finish'. Hence, resultative verb compounds are transparent to express both the source state and the target state. However, source and target state need not be expressed solely in resultative verb compounds. Instead, they can be expressed simultaneously by a single verb in Chinese, e.g., *dào* 'to arrive', *jìn* 'to enter'. The following examples show 2-state expressions with the four aspect markers in Chinese:

(7a) zhāngsān kāi-dào le hǎidébǎo. ----------[-----++++++++++++]
zhāngsān drive-arrive LE Heidelberg source state target state
zhāngsān drove to Heidelberg. not in Heidelberg in Heidelberg

[80] Resultative verb compounds are a type of serial verb constructions. According to Li and Thompson (1981), in resultative verb compounds, the first element conveys an action or process, while the second element expresses the result caused by the first element.

(7b) zhāngsān kāi-dào guo hǎidébǎo. ----------++++++++++ [ ]
zhāngsān drive-arrive GUO Heidelberg source state target state
zhāngsān has driven to Heidelberg. not in Heidelberg in Heidelberg

(7c) *zhāngsān zài kāi-dào hǎidébǎo. ----------++++[++++]++++
zhāngsān ZAI drive-arrive Heidelberg source state target state
not in Heidelberg in Heidelberg

(7d) *zhāngsān kāi zhe dào hǎidébǎo. ----------++++[++++]++++
zhāngsān drive ZHE arrive Heidelberg source state target state
not in Heidelberg in Heidelberg

Example (7d) should be corrected as shown below:

(7e) zhāngsān kāi zhe chē dào hǎidébǎo.
zhāngsān drive ZHE car arrive Heidelberg
zhāngsān arrived in Heidelberg by driving a car.

In example (7e), by using the durative aspect marker *zhe*, driving a car is presented as the ongoing background to the arrival in Heidelberg. This is because *zhe* can be used "in the first of two clauses to signal that one event provides a durative background for another event" (Li & Thompson, 1981, p. 223).

The imperfective aspect marker *zài* is typically not compatible with 2-state expressions because the distinguished phase in Chinese is the target state and the target state is the result of the action of the source state. It is instantaneous in nature and thus contradicts the imperfective marking, which requires a duration of an event. Thus, TT, which is indicated by *zài*, can neither be related to a target state nor to a source state in Chinese. In English, one can say, *he is entering a room*. This is because the distinguished phase in English is the source state, in this example, not in a room, and the progressive marking can be applied to this source state. By contrast, the same sentence in Chinese * *tā zài jìn yī-ge fángjiān* 'he is entering a room' sounds unnatural, on the grounds that for 2-state expressions in Chinese, the source state is not available for imperfective marking and the target state is not semantically compatible with imperfective marking.

Besides *zài* and *zhe*, there are other imperfective particles, i.e., *zhèngzài* and *zhèng*, which can describe an ongoing activity. *zài* is the short form of *zhèngzài*. Like *zài*, *zhèngzài* also conveys a progressive meaning and cannot occur with 2-state

expressions. However, there is one difference between *zài* and *zhèngzài*. *zài* presents a continuous activity, whereas *zhèngzài* represents "one instance of a dynamic and durative situation" (Jin & Hendriks, 2005, p. 74). Therefore, *zài* can occur with temporal adverbials like *yīzhí* 'all the time' in the sentence *wǒ yīzhí zài kànshū* (I have been reading all the time), whereas *zhèngzài* cannot. * *wǒ yīzhí zhèngzài kànshū* is wrong. In comparison with *zài* and *zhèngzài*, the particle *zhèng* exhibits the punctual characteristic of a situation, which means 'at the moment'. This punctual characteristic of *zhèng* makes it compatible with the instantaneous nature of the target state for 2-state expressions. So the following example is acceptable:

(7f) zhāngsān zhèng jìn yī-ge fángjiān.
zhāngsān ZHENG enter one-CL room
zhāngsān is entering a room.

Apart from the above-discussed aspect markers, i.e., the progressive aspect marker *zài*, the durative aspect marker *zhe*, and the perfective aspect markers *le* and *guo*, there is also the prospective particle *yào*[81] in Chinese, which denotes future events.

1-state situation

(8a) zhāngsān yào qù hǎidébǎo. [ ] ++++++++++++
zhāngsān YAO go Heidelberg go Heidelberg
zhāngsān will go to Heidelberg.

2-state situation

(8b) zhāngsān yào kāi-dào hǎidébǎo le. -------[-----]--++++++++++
zhāngsān YAO drive-arrive Heidelberg CRS[82] source state target state
zhāngsān will arrive in Heidelberg. not in Heidelberg in Heidelberg

[81] *yào* differs in the functions it has in different context. For example, as a modal verb, *yào* means 'to want to, to feel like, to prefer or intend' as in the following sentence:

我要吃面条，不要吃米饭。
wǒ yào chī miàntiáo, bù yào chī mǐfàn.
I want eat noodle, not want eat rice
I want to eat noodles, not rice.

[82] Generally speaking, there are two kinds of *le* in Chinese: verb-final *le* and sentence-final *le*. The former conveys the perfective meaning, while the latter conveys the currently relevant state (CRS). Following Li and Thompson (1981), we adopted the abbreviation CRS for the sentence-final *le*, which means "a state of affairs has special current relevance with respect to some particular situation" (p. 240). In some cases, it is difficult to distinguish the verb-final *le* with perfective meaning from the sentence-final *le* with respect to the currently relevant state (see, e.g., Li, 1990; Rohsenow, 1976, 1978). This study is mainly concerned with the perfective meaning conveyed by the verb-final *le* (i.e., *le* both in the verb-final and sentence-final position).

In comparison with European languages in which tense or aspect is obligatorily inflexed in verbs, the use of aspect markers is optional in Chinese. According to Klein (2000), an utterance without any aspectual marking in Chinese is called "zero marking"[83]. Its meaning can be disambiguated by virtue of other linguistic devices or contextual factors, as shown in the following example:

(9) 一个白色的车子开进一个院子里面。(L1 Ger-L2 Chn P23, C03)

yī-ge báisè-de chēzi kāi-jìn yī-ge yuànzi-lǐmian.
one-CL white-MOD car drive-enter one-CL yard-inside
A white car drives into a yard.

Neither an imperfective nor a perfective aspect marker is found in example (9). However, the target state, that is, *in the yard*, has been incorporated into the lexical content of the resultative verb compound (more precisely, the directional verb compound; for more details, see section 4.3.1.1), i.e., *kāi jìn* 'drive enter'. Also, the use of the postposition *-lǐmian* 'inside' reinforces the meaning that the target state has been reached. Hence, although there is no overt perfective aspect marker *le* in example (9), a perfective reading can be inferred. For sentences with zero marking like example (9), Li and Thompson (1981, p. 206) argues that such a sentence "contains another element that does the job of 'perfectivizing' the verb". In other words, depending on the actual context, there are some other elements in the sentence that can perform the same function as the perfective aspect marker *le* does, i.e., viewing the situation as an unanalyzable whole.

Previous studies have shown that the progressive aspect marker *zài* is closely related with *zài*, which indicates locative meaning (realized as a locative verb or a locative preposition) diachronically and synchronically (e.g., Chao, 1968; Chen, 1978). Given the complications associated with the functions of *zài* in different positions, the following illustrates the functions of the preverbal and the postverbal *zài*-phrase, respectively.

Preverbal *zài*-phrase

Originally, *zài* was a locative verb. Later, it evolved into a locative preposition

[83] Smith (1997) suggested that sentences without any aspectual markings have a neutral aspect, which is the default value of aspect. The meaning can then be interpreted through contextual or world knowledge.

and recently into a progressive aspect marker (Li, 1988, 1993). The preverbal *zài* can occur in the following four patterns:

(10a) zài + location

车在路上。(L1 Ger-L2 Chn P21, B04)

chē zài lù-shàng.
car locate road-top
A car is (located) on the road.

(10b) zài + location + verb[84]

一个人在草地上走。(L1 Chn P08, B03)

yī-ge rén zài cǎodì-shàng zǒu.
one-CL person at grass-top walk
A person is walking on the grass.

(10c) zài + time + verb

他每天(在)早上七点起床。

tā měi-tiān (zài) zǎoshang qī-diǎn qǐchuáng.
he every-day (at) morning seven-o'clock get up
He gets up at seven o'clock every morning.

(10d) zài + verb

女生在走路。(L1 Ger-L2 Chn P12, A07)

nǚshēng zài zǒu-lù.
girl ZAI walk-road
A girl is walking on the road.

[84] "zài + location" can also be located at the beginning of the sentence, as in the construction "zài + location + NP + verb", where NP indicates the subject of the sentence. In this case, this construction has only habitual reading, as the following example shows:

在街上有个女人走过去。(L1 Ger-L2 Chn P11, A01)

zài jiē-shàng yǒu-gè nǚrén zǒu-guò-qù.
at street-top exist-CL woman walk-pass-go
On the street there is a woman who walks over (away from the speaker).

In example (10a), *zài* is a locative verb, while in (10d) it functions as a progressive aspect marker. In example (10c), *zài* precedes a temporal phrase and functions as a preposition, which resembles the English preposition *at* in *at seven o'clock*. In this case, *zài* is generally omitted. It is difficult to determine the function of *zài* in example (10b). There is general agreement on its locative meaning. As shown in the following examples, when the sentence has perfective or habitual reading, then the *zài* phrase functions only as a locative adjunct that specifies the location where a motion takes place. It does not include any temporal information.

(11a) 那个人在草地上走了好几趟。(perfective reading)

nà-ge rén zài cǎodì-shàng zǒu-le hǎo jǐ-tàng.
that-CL person at grass-top walk-LE quite several-CL
That person walked on the grass several times.

(11b) 那个人常常在草地上走。(habitual reading)

nà-ge rén chángcháng zài cǎodì-shàng zǒu.
that-CL person often at grass-top walk
That person often takes a walk on the grass.

Previous studies (Chan, 1980; Chao, 1968; Chen, 1978; Li, 1999, p. 218; Li, 1988, 1990, 1993, 1998; Mathews, 1990, pp. 180-81; Woo, 2013; Xiao & McEnery, 2004, p. 193) have argued that the preverbal *zài* in the construction "zài + location + verb" encodes both locative and temporal information. This point of view, which has diachronic, cross-linguistic, and syntactic justification, is also adopted in this study[85].

The close relation between locative and temporal information is observed in many other languages. Anderson (1973, p. 15) observed "an association between progressive aspect and a locative predication" and Comrie (1976, p. 98) claims that "there is similarity between the formal expression of imperfective aspect, especially progressive aspect, and various locative adverbial phrases" in various languages. In fact,

[85] If there are no other aspect markers (e.g., the perspective aspect marker *le* as in example (11a)) or temporal adverbials (e.g., *chángcháng* 'often' as in example (11b)) in a sentence that can change the progressive reading, then preverbal *zài* in the construction "zài + location + verb" encodes both locative and temporal information (e.g., Chen, 1978; Woo, 2013).

the progressive aspect in English is argued to come from a locative source (e.g., Comrie, 1976; Vlach, 1981). Historical traces can still be found in the English expression *asleep*, which comes from *at sleep*, and in the dialectical and archaic expression *Fred's been a-singing* (Comrie, 1976, p. 99). Although modern English expressions like *he is working* show no traces of being a locative construction, the sentence *he is at work*, which is a locative expression, has the same reading as *he is working*. Other Germanic languages such as German also demonstrate the close relationship between locative and progressive meaning. For example, *er ist am Arbeiten* literally means *he is at the working*. A significant association between spatial location and progressive aspect is also found in Margi, which is spoken in northern Nigeria. In Margi, the locative particle is applied to assign progressivity. The following example is quoted from Anderson (1973, p. 16):

(12) nivar wi.
I-in place-run
I am running.

These diachronic and typological observations justify correlating the locative with progressive meaning in Chinese, as in example (10b). That is, the preverbal particle *zài* not only locates the Figure in space but at a specific point in time, i.e., the topic of time with respect to the time of situation[86]. Just as Comrie (1976, p. 103) pointed out, "it is always the case that progressive meaning is also expressed as a locative, indeed it is usually the case that the same locative construction is used for both meanings".

Concerning the syntactic structure of a sentence that includes the preverbal *zài*

[86] In order to ensure that the progressive meaning is not derived from other constituents in an expression or the null morpheme in Chinese, Woo (2013, pp. 177-79) used *dào* 'arrive/to' and *gěi* 'give/for', which have the same function as *zài* (be located/at), i.e., as a verb or preposition, to examine whether expressions with these two words have progressive reading. As shown in the following, without *zài*, examples i and ii do not have progressive reading, while with *zài*, example c has progressive reading. In other words, the progressive reading is derived from *zài* itself and not from other elements or the null morpheme in Chinese.

i. gōngjiāochē kāi-dào gōngjiāochēzhàn. (L1 Ger-L2 Chn P04, A09)
bus drive-arrive bus station
The bus drove to the bus station.

ii. māma gěi wǒ zuò jiǎozi. (Li & Thompson, 1981, p. 358)
mother for I make dumpling
Mother made dumplings for me.

iii. māma zài gěi wǒ zuò jiǎozi. (Woo, 2013, p. 178)
mother ZAI for I make dumpling
Mother is making dumplings for me.

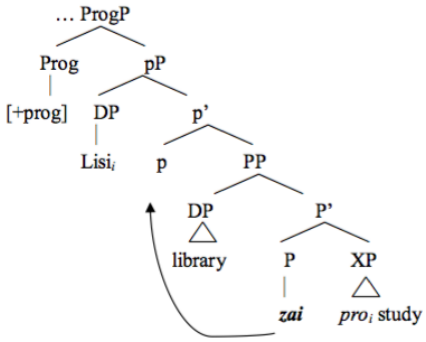
as in example (10b), Woo (2013, p. 185) proposed that "*zài* is a two-place preposition that takes a spatiotemporal element (either a location or a temporal expression) and an EVENT as its arguments". The author further argued that the progressive meaning of *zài* is derived from the spatial location *zài* through several head-movements[87].

In a word, it is diachronically, typologically, and syntactically justified to associate a spatial location with progressivity in Chinese. That is, the preverbal *zài* in the construction '*zài* + location + verb' encodes both locative and temporal meaning.

Postverbal *zài*-phrase

The postverbal *zài* implies the position of an action where a motion takes place (Tai, 1975), termination of an action (Chen, 1978), the location of the subject or direct object as a result of an activity (Li & Thompson, 1981), the location where an activity reaches and a state demonstrates itself (Fan, 1982), and terminative meaning in Northern Mandarin (Chirkova & Lamarre, 2005). Taking an aspectual approach, Liu (2009) argued that depending on the aspectual properties of the verbs with which it co-occurs, the postverbal *zài*-phrase has different functions. When it occurs with a stative verb (e.g., *zhù* 'live in'), the postverbal *zài*-phrase does not offer a boundary for the situation (see example (13a)), whereas when it occurs with a telic dynamic verb (e.g., *diào* 'drop'), the postverbal *zài* phrase explicitly expresses the event boundary inherent in this verb (see example (13b)). When the postverbal *zài* occurs with an atelic dynamic verb (e.g., *zǒu* 'walk', *pǎo* 'run'), it adds a locative boundary to an open-ended event (see example (13c)).

[87] Woo (2013) assumed that in the syntactic structure, there is little *p* and it takes a *zài* phrase as its argument. The function of little *p* is to establish a spatiotemporal relation between Figure (locatee) and Ground (location). The author took the sentence *Lisi zài túshūguǎn kàn shū* (lit. Lisi ZAI library read book) and represented its structure as below (p. 189). In this structural schema, *zài* is base-generated in the head of PP and PP is the complement of pP. *zài* has two arguments: the location *library* and the event *study*. *zài* can be moved to the little *p* and then be attached to it. In addition, the author argued that this sentence *Lisi zài túshūguǎn kàn shū* (lit. Lisi ZAI library read book) has also habitual reading and proposes a syntactic structure for it (for more details, see Woo, 2013, pp. 190-93).



(13a) 我住在海德堡住了三年。

wǒ zhù zài hǎidébǎo zhù-le sānnián.
I live at Heidelberg live-LE three-year
I lived in Heidelberg for three years.

(13b) 书掉在地上。

shū diào zài dì-shàng.
book drop at floor-top
The book dropped to the floor.

(13c) 一个人走在路上。

yī-ge rén zǒu zài lù-shàng.
one-CL person walk at road-top
A person walks on the road.

It is the locative boundary conveyed by the postverbal *zài* phrase as in example (13c) that makes it impossible to add the progressive aspect marker *zài* to the sentence (see example (14b)). In contrast, without the postverbal *zài* phrase, the manner verb *zǒu* 'walk' can occur with the progressive aspect marker *zài* (see example (14a)).

(14a) 一个人在走。

yī-ge rén zài zǒu.
one-CL person ZAI walk
A person is walking.

(14b) *一个人在走在路上。

* yī-ge rén zài zǒu zài lù-shàng.
one-CL person ZAI walk at road-top

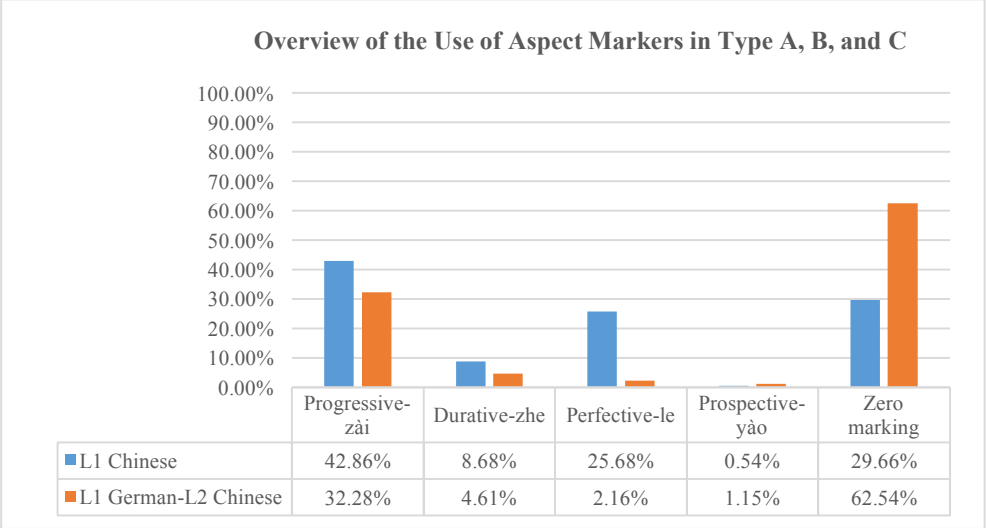
Due to the complicated functions performed by the postverbal *zài* phrase, some L1 German-L2 Chinese speakers have not learned to use it and thus made mistakes, as

in example (14b).

To summarize, according to Klein (1994, 2000), aspectual relations are characterized as the imperfective, perfective, and perfect aspect in English and as the imperfective, perfective, and prospective aspect in Chinese. By contrast, the difference between the imperfective and perfective aspect in German is not grammaticalized systematically. Although there are periphrastic constructions to encode this aspectual difference, they are not frequently used by German speakers. It is thus interesting to see how English learners of German and Chinese learners of German learn to use aspectual marking in English or Chinese when describing motion events. In what follows, we will first examine the use of aspect markers in L1 and L2 Chinese and then compare the use of aspectual marking in L1 and L2 English.

5.2 Aspectual marking in L1s and L2s

5.2.1 Aspectual marking in L1 and L2 Chinese



	Progressive-zài	Durative-zhe	Perfective-le	Prospective-yào	Zero marking
L1 Chinese	42.86%	8.68%	25.68%	0.54%	29.66%
L1 German-L2 Chinese	32.28%	4.61%	2.16%	1.15%	62.54%

Figure 5.1: Overview of the use of aspect markers in the verbal task of Type A, B, and C videos in total by speakers of L1 Chinese and L1 German-L2 Chinese

Generally speaking, compared to L2 Chinese speakers, L1 Chinese speakers tended to more frequently use the progressive aspect marker *zài* (L1 Chinese 42.86% vs. L2 Chinese 32.28%; $\chi 2$ (1, N = 1247) = 14.786, p < .001), the durative aspect marker *zhe* (L1 Chinese 8.68% vs. L2 Chinese 4.61%; $\chi 2$ (1, N = 1247) = 8.487, p = 0.004), and the perfective aspect marker *le* (L1 Chinese 25.68% vs. L2 Chinese 2.16%; $\chi 2$ (1, N = 1247) = 154.660, p < .001) (see Figure 5.1). By contrast, L2 Chinese speakers

preferred using more zero marking than L1 Chinese speakers (L1 Chinese 29.66% vs. L2 Chinese 62.54%; $\chi 2$ (1, N = 1247) = 133.309, p < .001). In addition, speakers of L1 and L2 Chinese did not differ significantly in the use of the prospective aspect marker *yào* (L1 Chinese 0.54% vs. L2 Chinese 1.15%; $\chi 2$ (1, N = 1247) = 0.706, p = 0.401). The following examples illustrate the use of the various aspect markers.

(15a) Use of the progressive aspect marker *zài*

女孩子在走路，她在右转。(L1 Ger-L2 Chn P01, A01)

nǚháizi zài zǒu-lù, tā zài yǒu zhuǎn.
girl ZAI walk-road, she ZAI right turn
(A) girl is walking. She is turning right.

(15b) Use of the durative aspect marker *zhe*

有一辆车开着。(L1 Ger-L2 Chn P10, B01)

yǒu yī-liàng chē kāi zhe.
exist one-CL car drive ZHE
There is a car driving.

(15c) Use of the perspective aspect marker *le*

这位先生进了教堂的门。(L1 Ger-L2 Chn P01, C08)

zhè-wèi xiānsheng jìn-le jiàotáng-de mén.
this-CL gentleman enter-LE church-MOD gate
This gentleman entered the gate of a church.

(15d) Use of the prospective aspect marker *yào*

一辆白车要进入村子。(L1 Ger-L2 Chn P14, B05)

yī-liàng bái chē yào jìn-rù cūnzi.
one-CL white car will enter-enter village
A white car will enter into the village.

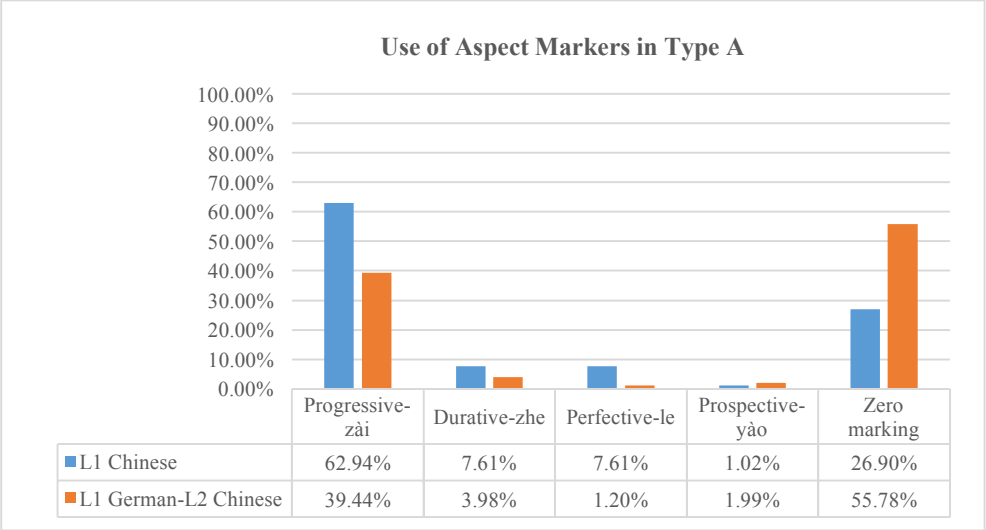
(15e) Use of zero marking

男生进教堂去。(L1 Ger-L2 Chn P17, C08)

nánshēng jìn jiàotáng qù.

boy enter church go

(A) boy enters the church (away from the speaker).



	Progressive-zài	Durative-zhe	Perfective-le	Prospective-yào	Zero marking
L1 Chinese	62.94%	7.61%	7.61%	1.02%	26.90%
L1 German-L2 Chinese	39.44%	3.98%	1.20%	1.99%	55.78%

Figure 5.2: Use of aspect markers in the verbal task of Type A videos by speakers of L1 Chinese and L1 German-L2 Chinese

As shown in Figure 5.2, in the verbal task of Type A videos, L1 Chinese speakers used the progressive aspect marker *zài* more often than L2 Chinese speakers (L1 Chinese 62.94% vs. L2 Chinese 39.44%; $\chi 2$ (1, N = 448) = 24.386, p < .001). That is, L1 Chinese speakers tended to take an intermediate perspective to view the situation in Type A videos, which showed a short trajectory with an obvious endpoint. Speakers of Chinese can use the progressive aspect marker *zài* to describe a motion towards a place or a motion within a place. In contrast, L2 Chinese tended to use more zero marking (L1 Chinese 26.90% vs. L2 Chinese 55.78%; $\chi 2$ (1, N = 448) = 37.524, p < .001). Furthermore, despite the low number of occurrences, L1 Chinese speakers used the perspective aspect marker *le* more often than L2 Chinese speakers (L1 Chinese 7.61% vs. L2 Chinese 1.20%; $\chi 2$ (1, N = 448) = 11.793, p = 0.001). However, speakers of L1 and L2 Chinese did not differ in the use of the durative aspect marker *zhe* (L1 Chinese 7.61% vs. L2 Chinese 3.98%; $\chi 2$ (1, N = 448) = 2.760, p = 0.097) and the prospective aspect marker *yào* (L1 Chinese 1.02% vs. L2 Chinese 1.99%; $\chi 2$ (1, N = 448) = 0.197, p = 0.657).

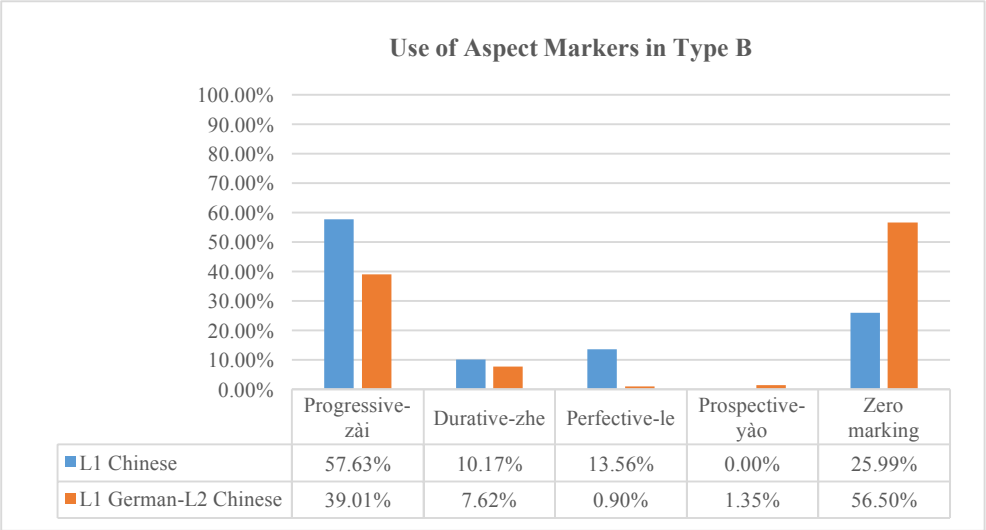
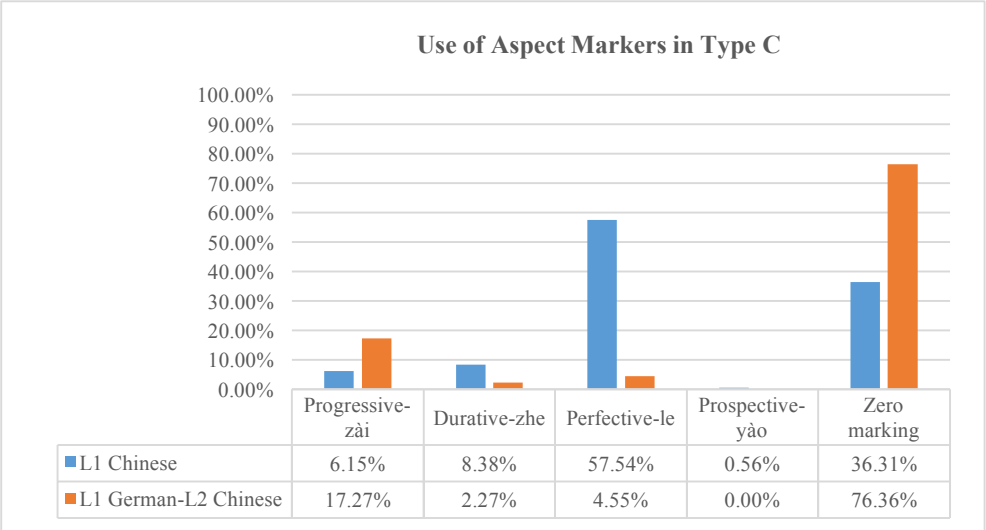


Figure 5.3: Use of aspect markers in the verbal task of Type B videos by speakers of L1 Chinese and L1 German-L2 Chinese

In comparison to the verbal task of Type A videos, no significant effect was found in the use of the progressive aspect marker *zài* by L1 Chinese speakers in the verbal task of Type B videos ($\chi 2$ (1, N = 374) = 1.102, p = 0.294) (see Figure 5.3). The same was true for L2 Chinese speakers when comparing the use of the progressive aspect marker *zài* in the verbal task of Type A and B videos ($\chi 2$ (1, N = 474) = 0.009, p = 0.924)[88]. However, L1 Chinese speakers used the progressive aspect marker *zài* significantly more often than L2 Chinese speakers in the verbal task of Type B videos (L1 Chinese 57.63% vs. L2 Chinese 39.01%; $\chi 2$ (1, N = 400) = 13.717, p < .001). Besides the more frequent use of the progressive aspect marker *zài*, L1 Chinese speakers also used the perfective aspect marker *le*[89] significantly more often (L1 Chinese 13.56% vs. L2 Chinese 0.90%; $\chi 2$ (1, N = 400) = 26.033, p < .001). By sharp contrast, L2 Chinese speakers used more zero marking (L1 Chinese 25.99% vs. L2 Chinese 56.50%; $\chi 2$ (1, N = 400) = 37.485, p < .001) and also used the prospective aspect marker *yào* more often (L1 Chinese 0% vs. L2 Chinese 1.35%) than L1 Chinese speakers. However, L1 and L2 Chinese speakers did not differ in the use of the durative aspect marker *zhe* (L1 Chinese 10.17% vs. L2 Chinese 7.62%; $\chi 2$ (1, N = 400) = 0.801, p = 0.371).

[88] Although no significant effect was found in the use of the progressive aspect marker *zài* between the verbal task of Type A and Type B videos by speakers of L1 and L2 Chinese, its use differed in co-occurrence with spatial concepts encoded in the verbal task of Type A and B videos (for more details, see Chapter 6).

[89] Examples of the use of the perfective aspect marker *le* in the verbal task of Type B videos are expressions like *yī-liàng chē kāi zǒu le*. 'one-CL car drive away LE'. Expressions that are not 2-state expressions nevertheless occurred with the perfective aspect marker *le*; these expressions take the start of an activity as a differentiated point to get the 2-state property. In this example, the change of state from 'not starting to drive' to 'starting to drive' is compatible with the use of the perfective aspect marker *le*.



	Progressive-zài	Durative-zhe	Perfective-le	Prospective-yào	Zero marking
L1 Chinese	6.15%	8.38%	57.54%	0.56%	36.31%
L1 German-L2 Chinese	17.27%	2.27%	4.55%	0.00%	76.36%

Figure 5.4: Use of aspect markers in the verbal task of Type C videos by speakers of L1 Chinese and L1 German-L2 Chinese

A general view regarding the verbal task of Type C videos showed that in comparison to that of Type B, the use of the progressive aspect marker *zài* in both groups decreased, while the use of the perfective aspect marker *le* and of zero marking increased in both groups (see Figure 5.4). However, L1 and L2 Chinese speakers differed in the use of the above-mentioned aspect markers in the verbal tasks of Type C videos. L1 Chinese speakers used the perfective aspect marker *le* significantly more often than L2 Chinese speakers (L1 Chinese 57.54% vs. L2 Chinese 4.55%; $\chi 2$ (1, N = 399) = 136.552, $p < .001$), whereas L2 Chinese speakers used significantly more zero marking than L1 Chinese speakers (L1 Chinese 36.31% vs. L2 Chinese 76.36%; $\chi 2$ (1, N = 399) = 65.164, $p < .001$). Note that Type C videos showed boundary-crossing events. Speakers actually saw a Figure reaching and crossing a boundary characterized by the Ground object. In this case, L1 Chinese speakers preferred using the perfective aspect marker *le* to indicate that the event was completed, while L2 Chinese speakers preferred leaving their utterances unmarked by any aspect markers; i.e., they used zero marking to describe boundary-crossing events. The following are some examples:

(16a) Perfective aspect marker *le*

一个人走进了一个超市。(L Chn P18, C04)

yī-ge rén zǒu-jìn-le yī-ge chāoshì.

one-CL person walk-enter-LE one-CL supermarket

A person walked into a supermarket.

(16b) Zero marking

*一个女人进去超市。(L1 Ger-L2 Chn P20, C04)

* yī-ge nǚrén jìn-qù chāoshì.
one-CL woman enter-go supermarket
A woman enters a supermarket.

Due to the word order problem, example (16b) is grammatically unacceptable. It should be corrected as *yī-ge nǚrén jìn chāoshì qù*. 'lit. one-CL woman enter supermarket go'. L1 Chinese speakers tended to place the perfective aspect marker *le* after the boundary-crossing path verb *jìn* 'enter', while L2 Chinese speakers tended to reserve the same place for the deictic verb *qù* 'go' and leave the utterance unmarked.

In addition, L1 Chinese speakers used the durative aspect marker *zhe* more often than L2 Chinese speakers (L1 Chinese 8.38% vs. L2 Chinese 2.27%; $\chi 2$ (1, N = 399) = 7.731, p = 0.005). Two video clips were responsible for this use in the verbal task of Type C videos: one showing a man riding a bicycle into an entrance and the other one showing a person riding a horse into stable. In these two situations, the durative aspect marker *zhe* occurred together with the manner verb *qí* 'ride' to indicate an accompanying activity, as the following examples illustrate:

(17a) 有一个人骑着自行车进汉学系。(L1 Ger-L2 Chn P14, C09)

yǒu yī-ge rén qí-zhe zìxíngchē jìn hànxuéxì.
exist one-CL person ride-ZHE bicycle enter Chinese Studies department
Someone rides a bicycle into the Department of Chinese Studies.

(17b) 一个人骑着马进了马厩。(L1 Chn P16, C10)

yī-ge rén qí-zhe mǎ jìn-le mǎjiù.
one-CL person ride-ZHE horse enter-LE stable
A person rode a horse into a stable.

Furthermore, in the verbal task of Type C videos, L2 Chinese speakers used the progressive aspect marker *zài* significantly more often than L1 Chinese speakers (L1 Chinese 6.15% vs. L2 Chinese 17.27%; $\chi 2$ (1, N = 399) = 11.344, p = 0.001). This

occurrence of *zài* in L2 Chinese might not be acceptable in Chinese, since some of them occurred with 2-state expressions, which are not allowed in Chinese.

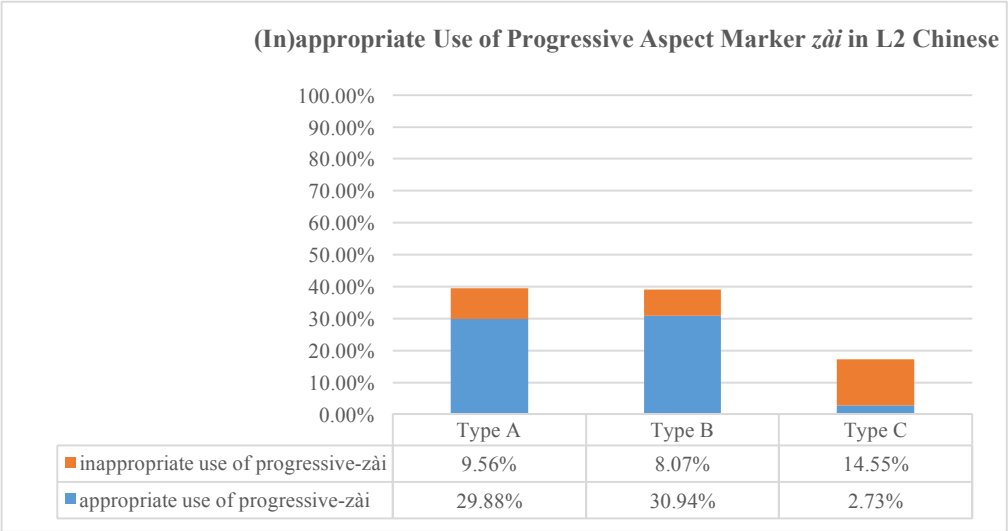


Figure 5.5: Appropriate and inappropriate use of the progressive aspect marker *zài* by L1 German-L2 Chinese speakers in the verbal task of Type A, B, and C videos

As shown in Figure 5.5, L1 German-L2 Chinese speakers used the progressive aspect marker *zài* either appropriately or inappropriately. The percentages of inappropriate use of *zài* was remarkably higher in the verbal task of Type C videos (14.55%) than in that of Type A (9.56%) and Type B videos (8.07%), although the total occurrence of *zài* in the verbal task of Type C videos (17.27%) was lower than in that of Type A (39.44%) and Type B videos (39.01%). The reason why *zài* was used less often in the verbal task of Type C videos lies in the characteristics of this video type, which showed boundary-crossing situations. In Chinese, boundary-crossing situations are expressed through the path verb *jìn* 'enter', which is a 2-state expression and not compatible with the progressive aspect marker *zài*. Some L2 Chinese speakers made this mistake (see example (19)). However, some L1 or L2 Chinese speakers divided the situation in Type C videos into two phases, i.e., an intermediate and a final phase. The progressive marker *zài* can be used in the intermediate phase. It is therefore no wonder that appropriate use of the progressive marker *zài* still occurred in L2 Chinese in the verbal task of Type C videos, as shown in example 18:

(18) 一个人在骑马，骑到房子里。(L1 Ger-L2 Chn P24, C10)

yī-ge rén zài qí mǎ, qí-dào fángzi-li.

one-CL person ZAI ride horse, ride-arrive house-inside
A person is riding a horse. He/she rides it into the house.

In L2 Chinese, the inappropriate use of the progressive aspect marker *zài* can be summarized as co-occurring with 2-state expressions or the preverbal or postverbal *zài*-phrase. As already discussed in section 5.1, the reason for the incompatibility of the progressive aspect marker *zài* with 2-state expressions might lie in the fact that the distinguished phase is the target phase in Chinese. The target state, i.e., the result state, is instantaneous and is not eligible for the progressive aspect marker *zài*, which describes a motion as an ongoing progress (example (19)). The reason for the incompatibility of the progressive aspect marker *zài* with the preverbal *zài*-phrase is that the preverbal *zài*-phrase may have already conveyed the ongoingness. Therefore, the meaning of ongoingness expressed by the progressive marker *zài* and the preverbal *zài*-phrase is duplicated, which leads to unnatural expressions in L2 Chinese (example (20)). The incompatibility of the progressive aspect marker *zài* with the postverbal *zài*-phrase might be attributable to the telic status contributed by the postverbal *zài*-phrase. Depending on which verb is used, the postverbal *zài*-phrase may add a locative boundary to an event, which is incompatible with the meaning of progressivity conveyed by the progressive aspect marker *zài* (example (21)).

Incompatibility of the progressive aspect marker *zài* with 2-state expressions

(19a) ? 一个人在走过去。(L1 Ger-L2 Chn P08, A01)

? yī-ge rén zài zǒu-guò-qù.
one-CL person ZAI walk-pass-go
A person is walking over (away from the speaker).

(19b) ? 一辆汽车在开到小城市。(L1 Ger-L2 Chn P24, B01)

? yī-liàng qìchē zài kāi-dào xiǎo chéngshì.
one-CL car ZAI drive-arrive small city
A car is driving to a small city.

(19c) ? 一辆车在路上开走。(L1 Ger-L2 Chn P03, B08)

? yī-liàng chē zài lù-shàng kāi-zǒu.

one-CL car ZAI road-top drive-leave

A car is driving away on the road.

(19d) * 这位男生在进去教堂。(L1 Ger-L2 Chn P02, C08)

* zhè-wèi nánshēng zài jìn-qù jiàotáng.

this-CL boy ZAI enter-go church

(19e) ? 有一个人在赶上火车。(L1 Ger-L2 Chn P06, C02)

? yǒu yī-ge rén zài gǎn-shàng huǒchē.

exist one-CL person ZAI catch up with-ascend train

There is a person who is catching up with the train.

Incompatibility of the progressive aspect marker *zài* with the preverbal *zài*-phrase

(20) * 有一个白色的车在一条比较大一点的路上在走，然后它越来越远。(L1 Ger-L2 Chn P22, B01)

*yǒu yī-ge báisè-de chē zài yī-tiáo bǐjiào dà yīdiǎn-de lù-shàng zài zǒu, ránhòu tā yuè lái yuè yuǎn.

exist one-CL white-MOD car ZAI one-CL relative big a little-MOD road-top ZAI walk, then it further come further far

There is a white car moving on a relatively big road, and then it moves further and further.

Incompatibility of the progressive aspect marker *zài* with the postverbal *zài*-phrase

(21) * 又是一辆车在开在一个马路上。(L1 Ger-L2 Chn P05, B02)

* yòu shì yī-liàng chē zài kāi zài yī-ge mǎlù-shàng.

again be one-CL car ZAI drive at one-CL road-top

Again this is a car driving on a road.

5.2.2 Aspectual marking in L1 and L2 English

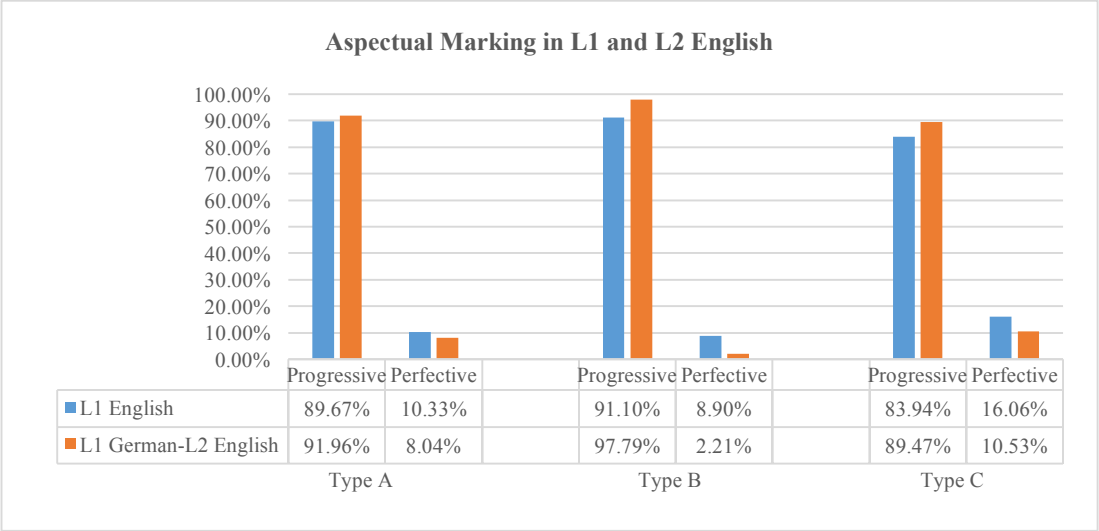


Figure 5.6: Aspectual marking in L1 and L2 English in the verbal task of Type A, B, and C videos

As shown in Figure 5.6, L1 and L2 English speakers used overwhelmingly more imperfective marking in all video types. In the verbal task of Type A and Type C videos, L1 and L2 English speakers did not differ significantly from each other in the use of the imperfective and the perfective aspect (Type A: $\chi 2$ (1, N = 412) = 0.644, p = 0.422; Type C: $\chi 2$ (1, N = 427) = 2.820, p = 0.093). However, in the verbal task of Type B videos, L2 English speakers used more imperfective marking than L1 English speakers ($\chi 2$ (1, N = 372) = 7.810, p = 0.005).

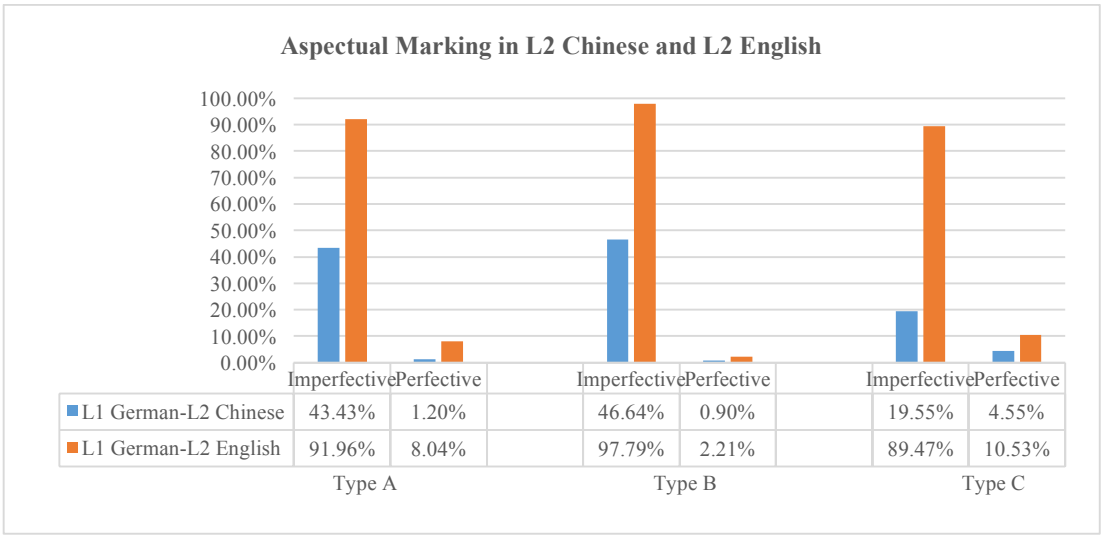


Figure 5.7: Comparison of the use of aspectual marking in L1 German-L2 Chinese and L1 German-L2 English in the verbal task of Type A, B, and C videos

The use of imperfective forms in L2 Chinese includes the occurrence of the progressive marker *zài* and the durative marker *zhe*, while the use of imperfective forms in L2 English means the use of the progressive form. In comparison with L1 German-L2 Chinese speakers, L1 German-L2 English speakers tended to mark their sentences with progressive forms in all video types (Chi-square test: all $p < 0.05$) (see Figure 5.7). The prevalent use of progressive forms in L1 German-L2 English may be attributable to "the transparency of encoding of the new conceptual category" (Schmiedtová et al., 2011, p. 94). That is, in contrast to Chinese, in which the aspect markers are used optionally, the imperfective marking in English is inflexed in the verbs and is perceptually and formally salient. Hence, L2 English speakers are sensitive to imperfective marking, which leads to its prevalent use in L2.

Furthermore, it should be noted that in the verbal task of Type C videos, the English imperfective marking can be applied to the distinguished phase, i.e., the source state of 2-state expressions (e.g., *enter*) in English (e.g., a man is entering a church), whereas the Chinese progressive aspect marker *zài* cannot be applied to the distinguished phase, i.e., the target state of 2-state expressions (e.g., *jìn* 'enter') in Chinese (see example (19d)). In Chinese, a boundary-crossing situation was typically expressed with the 2-state expression *jìn* 'enter' (with or without a manner verb or deictic verb) in the verbal task of Type C videos. Therefore, in comparison with L1 German-L2 English speakers, the use of the progressive marker *zài* occurred remarkably less often in L1 German-L2 Chinese in the verbal task of Type C videos.

Concerning the use of the perfective aspect, L1 German-L2 English speakers used it more often than L1 German-L2 Chinese speakers in the verbal task of Type A and Type C videos, respectively (Chi-square test: both $p < 0.05$). By contrast, L1 German-L2 Chinese speakers preferred using zero marking in all video types (see Figure 5.2, 5.3, and 5.4) rather than the perfective aspect. In comparison to the verbal task of Type A and C videos, despite the low number of occurrences, no significant difference was found in the use of the perfective aspect in the verbal task of Type B videos (Chi-square test: $p > 0.05$).

According to the hypothesis of *seeing for speaking* (von Stutterheim et al., 2012, p. 835), when a language has a particular grammaticalized concept and speakers of this language frequently express it, then they tend to pay attention to the visual features related to this grammaticalized concept. In contrast, when a language does not have the same grammaticalized concept and this concept must be encoded in lexical, phrasal

constituents or in other complex constructions and speakers of this language do not express this concept frequently, they will not, or to a lesser degree, pay attention to the related visual features in that situation. Specifically, the grammaticalized concept in von Stutterheim's study was the temporal-aspectual concept. Regarding the languages at issue, i.e., English, German, and Chinese, English has a grammaticalized imperfective aspect, while German does not have a grammaticalized imperfective or progressive aspect. By contrast, in Mandarin Chinese, the imperfective aspect is encoded in lexical items and is used frequently. In addition, zero marking is also frequently used in Chinese. As already discussed in sections 5.2.1 and 5.2.2, L1 German-L2 Chinese speaker used the imperfective and perfective aspect markers less often, but zero marking more often, than L1 Chinese speakers, while the use of the imperfective aspect is generally comparable for L1 English and L1 German-L2 English. As the next step, it is interesting to examine the visual attention allocated to the moving entity and to the endpoint over the course of time in L1s and L2s.

5.3 Allocation of attention over the course of time in L1s and L2s

Chapter 4 presented the different degrees of attention to the moving entity and to the endpoint in the different language groups. Based on this finding, and to answer the questions proposed at the beginning of this chapter, it will be interesting to find out whether there was any difference in the attention allocated to the moving entity and to the endpoint over the course of time in the different language groups examined in this study. In what follows, we will first compare the relative frequency of the endpoint being mentioned from different aspectual perspectives across the different languages and then illustrate the attention allocated to the moving entity and to the endpoint between L1 and L2 speakers over the course of time.

5.3.1 Allocation of attention in L1 and L2 Chinese

As already discussed in Chapter 4, in the verbal task of Type A videos, 34.52% of the utterances in L1 Chinese mentioned the endpoint. Of these, as shown in Figure 5.8[90], the percentage of those where zero marking was used was 14.72%, followed by

[90] In Figure 5.8, Impf. is the short form for imperfective aspect, Zero stands for zero marking, Perf. for the perfective and Prosp. for the prospective aspect.

13.20% where the imperfective aspect was used, 5.58% where the perfective aspect was used, and 1.02% where the prospective aspect was used. Similarly, 35.46% of the utterances mentioned the endpoint in L2 Chinese in the verbal task of Type A videos. Of these, the percentage of those where zero marking was used was 26.29%, followed by 6.77% where the imperfective aspect was used, 1.99% where the prospective aspect was used, and 0.40% where the perfective aspect was used. The same order was observed for L2 Chinese speakers in the verbal task of Type B videos. That is, 19.28% of the utterances mentioned the endpoint using zero marking, followed by 1.79% where the imperfective aspect was used, 1.35% of utterances where the prospective aspect was used, and 0.90% where the perfective aspect was used. By contrast, only 6.78% of the utterances mentioned the endpoint in L1 Chinese in the verbal task of Type B videos. In 3.39% of these utterances, zero marking was used, while 2.26% used the perfective aspect and 0.56% the imperfective aspect. In the verbal task of Type C videos, 91.62% of the utterances mentioned the endpoint in L1 Chinese, of which 55.31% used the perfective aspect, 32.96% zero marking, 3.35% the imperfective aspect, and 0.56% the prospective aspect. In contrast, there were 82.73% of utterances that mentioned the endpoint in L2 Chinese in the verbal task of Type C videos, of which 67.73% used zero marking, 10.91% the imperfective aspect, and 4.09% the perfective aspect.

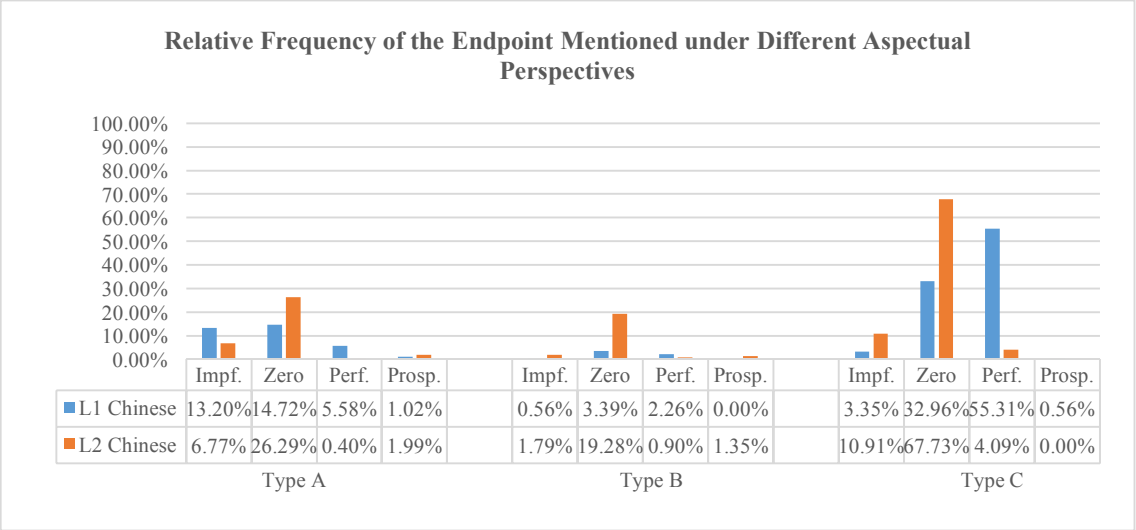


Figure 5.8: Relative frequency of the endpoint being mentioned when using different aspect markers in L1 and L2 Chinese in the verbal task of Type A, B, and C videos

The correlation between the use of aspectual marking and mentioning of the endpoint is summarized in Table 5.1.

Table 5.1: Sequence of use of aspectual marking when mentioning the endpoint in L1 and L2 Chinese in the verbal task of Type A, B, and C videos

	L1 Chinese	L2 Chinese
Type A	Zero Marking>Impf.>Perf.>Prosp.	Zero Marking>Impf.>Prosp.>Perf.
Type B	Zero Marking>Perf.>Impf.>Prosp.	Zero Marking>Impf.>Prosp.>Perf.
Type C	Perf.>Zero Marking>Impf.>Prosp.	Zero Marking>Impf.>Perf.>Prosp.

It appears that when the endpoint was mentioned, zero marking tended to be used by L1 and L2 speakers in all video types. In addition, the imperfective aspect was also used when the endpoint was mentioned. On the other hand, the prospective aspect was used rarely in both L1 and L2 Chinese, and when using the perfective aspect, all speakers paid attention to and referred to the endpoint (see von Stutterheim et al., 2012). Therefore, the attention allocated to the endpoint over the course of time when the perfective or prospective aspect was used will not be examined. Rather, it is more interesting to examine the attention allocated to the moving entity and to the endpoint over the course of time when the imperfective aspect or zero marking were used. Specifically, it is interesting to answer the following questions:

1 Did the attention given to the moving entity and the endpoint differ over the course of time between L1 and L2 Chinese speakers when zero marking or the imperfective aspect was used?

2 If the answer is positive, when did such differences appear after the start of the stimulus?

To answer these two questions, we followed the methods used by Flecken et al. (2015) and Wang (2015), registering the fixations on the moving entity and on the endpoint every 50 ms. Given the naturalistic nature of live-recorded video clips, the participants could fixate on the moving entity (Figure), the endpoint (Ground), neither on the moving entity nor on the endpoint, or no fixations could be registered at all. In this study, we considered the fixations on the moving entity and on the endpoint registered in 50 ms time intervals. Given that we aim at examining the differences during the conceptualization phase, the focus was on the fixations on the moving entity and on the endpoint before SOTs in L1 and L2 Chinese. So a plot was drawn before SOTs for each video type when different aspectual markings were used (before 4000 ms; see Chapter 4 for more details about SOTs). The statistical analysis was conducted on the relative frequency of the fixations on the moving entity and on the endpoint for each participant in consecutive 600 ms time windows (each time window including 12 time intervals).

Type A: Short trajectory with obvious endpoint

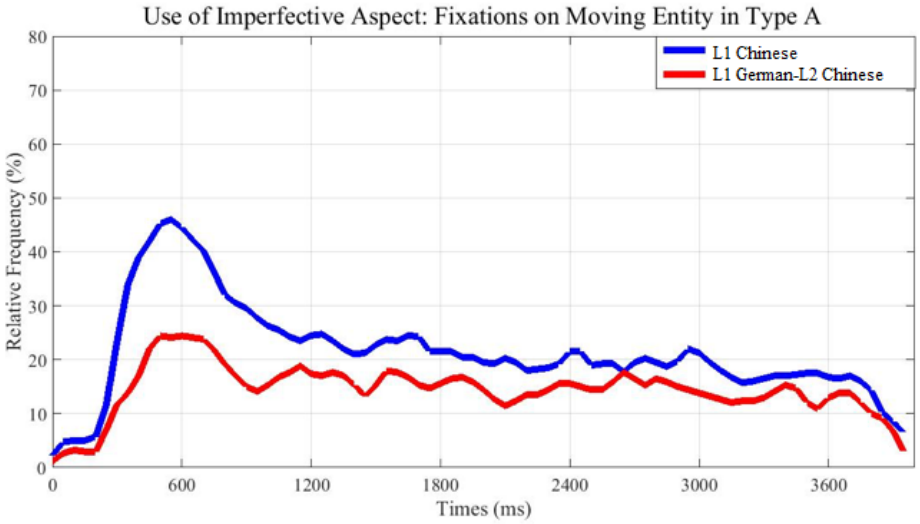


Figure 5.9: Relative frequency of fixations on the moving entity when using the imperfective aspect in L1 and L2 Chinese in Type A videos (time locked to the start of the stimulus)

The fixations on the moving entity when the imperfective aspect was used were analyzed according to language groups (2) and time windows (7 windows, 600 ms for each time window) (see Figure 5.9). Univariate ANOVA showed a significant main effect of the language group ($F (1, 245) = 13.031$, $p < .001$), a significant main effect of the time window ($F (6, 245) = 7.291$, $p < .001$), and a significant language group × time window interaction effect ($F (6, 245) = 3.130$, $p < .05$). In order to shed light on the group difference in the fixations on the moving entity in each time window, one-way ANOVA was carried out for each 600 ms time window. In the 0-600 ms time window, L1 Chinese speakers had significantly more fixations on the moving entity than L2 Chinese speakers ($F (1, 35) = 17.792$, $p < .001$). Similarly, in the 600-1200 ms time window, L1 Chinese speakers fixated more often on the moving entity than L2 Chinese speakers ($F (1, 35) = 8.402$, $p < .05$). By contrast, no significant group difference was found in the subsequent time windows (1200-1800 ms: $F (1, 35) = 1.039$, $p = 0.315$; 1800-2400 ms: $F (1, 35) = 2.488$, $p = 0.124$; 2400-3000 ms: $F (1, 35) = 1.035$, $p = 0.316$; 3000-3600 ms: $F (1, 35) = 0.289$, $p = 0.594$; 3600-4200 ms: $F (1, 35) = 3.349$, $p = 0.076$).

When considering the endpoint fixations under the condition of using the imperfective aspect in L1 and L2 Chinese in the verbal task of Type A videos, as shown in Figure 5.10, univariate ANOVA showed a significant main effect of the language group ($F (1, 238) = 6.429$, $p < 0.05$), a main effect of the time window ($F (6, 238) =$

24.596, p < .001), and a significant language group × time window interaction effect (F (6, 238) = 9.572, p < .001).

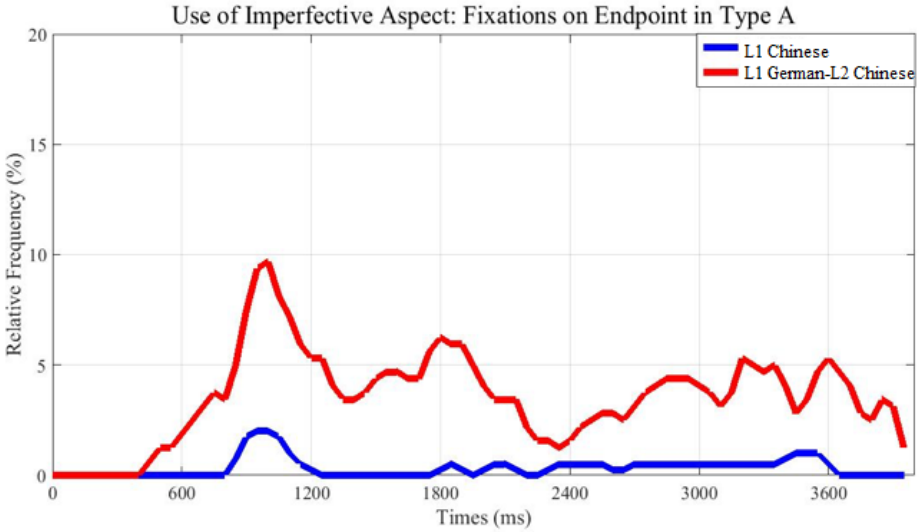


Figure 5.10: Relative frequency of fixations on the endpoint when using the imperfective aspect in L1 and L2 Chinese in Type A videos (time locked to the start of the stimulus)

An in-depth look at the group difference in the fixations on the endpoint in each time window showed that L1 and L2 Chinese speakers did not differ in the 0-600 ms time window (F (1, 34) = 1.259, p = 0.270). In this time window, there were very few fixations on the endpoint. Rather, in the first 600 ms, speakers of both groups fixated more on the moving entity (see Figure 5.9). Despite the low number of fixations in the 0-600 ms time window, L2 Chinese speakers started paying attention to the endpoint earlier than L1 Chinese speakers, with the attention to the endpoint reaching its peak in the 600-1200 ms time window. In this time window, L2 Chinese speakers had significantly more fixations on the endpoint than L1 Chinese speakers (F (1, 34) = 32.267, p < .001). By contrast, no significant difference was found between L1 and L2 Chinese speakers in the subsequent time windows (1200-1800 ms: F (1, 34) = 0.727, p = 0.400; 1800-2400 ms: F (1, 34) = 0.498, p = 0.485; 2400-3000 ms: F (1, 34) = 2.193, p = 0.148; 3000-3600 ms: F (1, 34) = 0.397, p = 0.533; 3600-4200 ms: F (1, 34) = 3.174, p = 0.084).

Like the statistical methods used above, language group (2) × time window (7) univariate ANOVA tests were used to examine the main and interaction effects for the fixations on the moving entity when zero marking was used in L1 and L2 Chinese in the verbal task of Type A videos (see Figure 5.11).

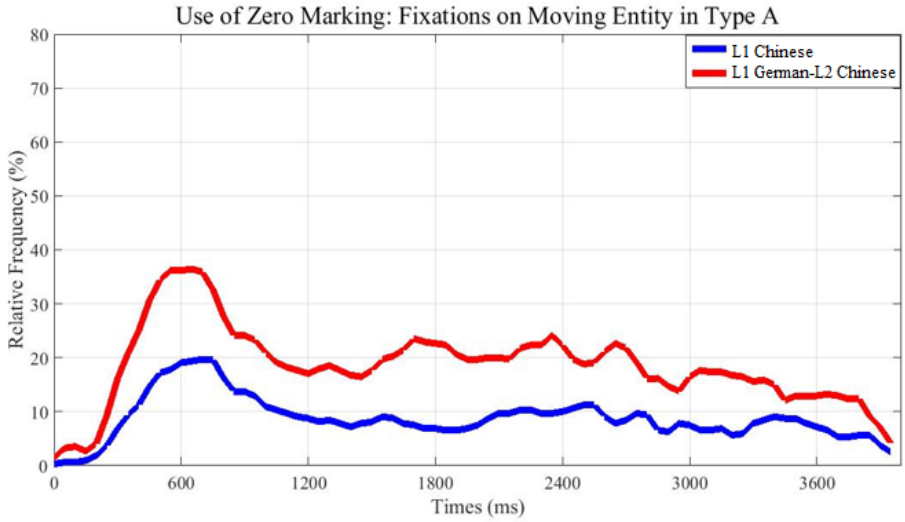


Figure 5.11: Relative frequency of fixations on the moving entity when using zero marking in L1 and L2 Chinese in Type A videos (time locked to the start of the stimulus)

A main effect of the language group ($F(1, 217) = 283.687$, $p < .001$) and the time window ($F(6, 217) = 7.403$, $p < .001$) was found. However, language group × time window interaction showed no significance ($F(6, 217) = 1.398$, $p = 0.217$). A further statistical analysis about the language group difference in the fixations on the moving entity in each time window showed that L2 Chinese speakers had significantly more fixations on the moving entity in each of the subsequent 600 ms time windows than L1 Chinese speakers (0-600 ms: $F(1, 31) = 41.197$, $p < .001$; 600-1200 ms: $F(1, 31) = 39.601$, $p < .001$; 1200-1800 ms: $F(1, 31) = 32.063$, $p < .001$; 1800-2400 ms: $F(1, 31) = 105.580$, $p < .001$; 2400-3000 ms: $F(1, 31) = 38.892$, $p < .001$; 3000-3600 ms: $F(1, 31) = 37.250$, $p < .001$; 3600-4200 ms: $F(1, 31) = 22.472$, $p < .001$).

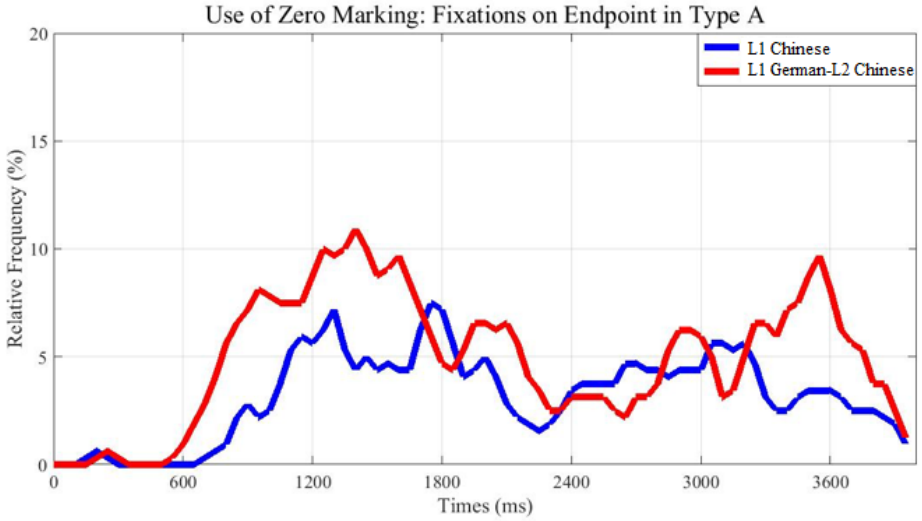


Figure 5.12: Relative frequency of fixations on the endpoint when using zero marking in L1 and L2 Chinese in Type A videos (time locked to the start of the stimulus)

When considering the endpoint fixations under the condition of using zero marking in L1 and L2 Chinese in the verbal task of Type A videos, as shown in Figure 5.12, univariate ANOVA showed a main effect of the language group ($F (1, 210) = 42.625$, $p < .001$), a main effect of the time window ($F (6, 210) = 23.473$, $p < .001$), and a significant language group × time window interaction effect ($F (6, 210) = 7.266$, $p < .001$). A closer look at the impact of the group difference in the fixations on the endpoint in each time window showed that no significant difference was found between L1 and L2 speakers in the 0-600 ms time window ($F (1, 30) = 0.031$, $p = 0.862$). In this time window, very few fixations were registered. By contrast, fixations on the endpoint began to increase between 600 ms and 1200 ms in both language groups. In this time window, L2 Chinese speakers had significantly more fixations on the endpoint than L1 Chinese speakers ($F (1, 30) = 20.759$, $p < .001$). The fixations on the endpoint reached their peak between 1200 ms and 1800 ms for both groups. In comparison with L1 Chinese speakers, L2 Chinese speakers had significantly more fixations on the endpoint in the 1200-1800 ms time window ($F (1, 30) = 16.121$, $p < .001$). The same is true for the 1800-2400 ms time window, where L2 Chinese speakers also had more fixations on the endpoint than L1 Chinese speakers ($F (1, 30) = 6.647$, $p < .05$). In contrast, no significant difference was found in the next two time windows between the two groups, i.e., the 2400-3000 ms time window ($F (1, 30) = 1.409$, $p = 0.244$) and the 3000-3600 ms time window ($F (1, 30) = 2.373$, $p = 0.134$). However, although both groups' fixations on the endpoint decreased between 3600 ms and 4200 ms, L2 Chinese speakers again had more fixations on the endpoint than L1 Chinese speakers ($F (1, 30) = 12.071$, $p < .05$). Note that the average SOTs in L2 Chinese for Type A videos was 3745 ms. L2 Chinese speakers' more frequent fixations on the endpoint shortly before SOTs may reflect their focus on the endpoint, which would help them construe the event with respect to it.

Type B: Long trajectory with possible endpoint

As for Type B videos, the mean value of SOTs in L1 Chinese was 2996 ms and in L2 Chinese 3521 ms, univariate ANOVA tests were used to examine the main effect of the two language groups and the six time windows (each 600 ms interval), as well as the language group (2) × time window (6) interaction effect. That is, in comparison with Type A videos, the statistical analysis was conducted for six time windows (0-3600 ms) instead of 7 time windows (0-4200 ms) for Type B videos, with the aim being to pinpoint early conceptual differences before the production of utterances.

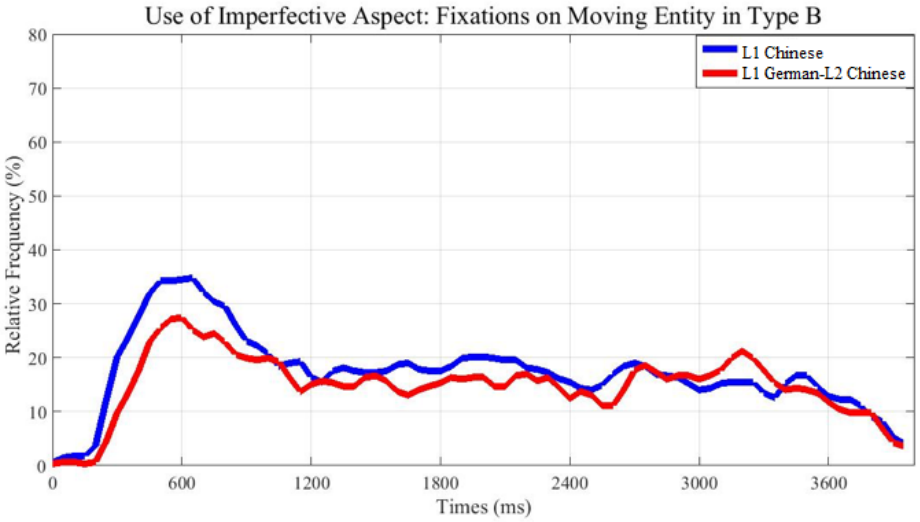


Figure 5.13: Relative frequency of fixations on the moving entity when using the imperfective aspect in L1 and L2 Chinese in Type B videos (time locked to the start of the stimulus)

When considering the fixations on the moving entity under the condition of using the imperfective aspect in L1 and L2 Chinese in the verbal task of Type B videos, a significant main effect of the language group ($F (1, 204) = 8.631$, $p < .001$) and the time window ($F (5, 204) = 5.887$, $p < .001$), respectively, was found (see Figure 5.13). The language group (2) × time window (6) interaction also showed a significant effect ($F (5, 204) = 2.264$, $p < .05$). Given that we want to find out when the conceptual differences in the two language groups occurs, a separate statistical analysis was carried out for each time window. In the first two time windows, i.e., 0-600 ms and 600-1200 ms, the fixations on the moving entity increased and reached their peak around 600 ms in both groups. However, L1 Chinese speakers fixated more on the moving entity in those two time windows than L2 Chinese speakers did (0-600 ms: $F (1, 34) = 12.071$, $p < .05$; 600-1200 ms: $F (1, 34) = 6.137$, $p < .05$). By contrast, no significant difference between the two language groups was found in each of the next four time windows (1200-1800ms: $F (1, 34) = 4.090$, $p = 0.06$; 1800-2400 ms: $F (1, 34) = 0.983$, $p = 0.328$; 2400-3000ms: $F (1, 34) = 0.006$, $p = 0.936$; 3000-3600 ms: $F (1, 34) = 0.905$, $p = 0.348$).

When considering the endpoint fixations under the condition of using the imperfective aspect in the verbal task of Type B videos, as shown in Figure 5.14, for language group (2) × time window (6) interaction, univariate ANOVA revealed a main effect of the time window ($F (5, 174) = 5.080$, $p < .001$) and a significant interaction effect ($F (5, 174) = 3.796$, $p < .05$), but no main effect of the language group ($F (1, 174) = 0.033$, $p = 0.855$).

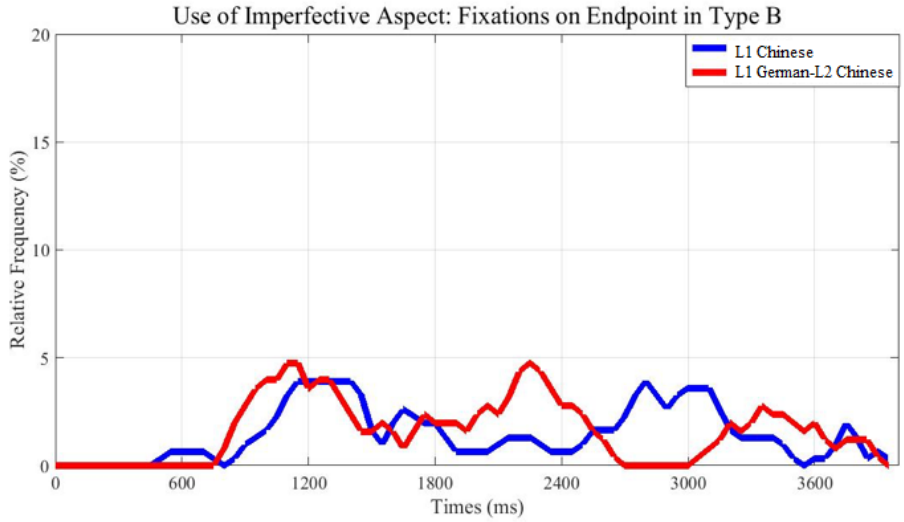


Figure 5.14: Relative frequency of fixations on the endpoint when using the imperfective aspect in L1 and L2 Chinese in Type B videos (time locked to the start of the stimulus)

The statistical analysis of the group difference in each 600 ms time window showed that despite very few fixations on the endpoint in the 0-600 ms time window in both groups, L1 Chinese speakers had more endpoint fixations than L2 Chinese speakers ($F(1, 29) = 7.144$, $p < .05$). Endpoint fixations in both groups increased rapidly and reached their first peak between 600 ms and 1200 ms. In this time window, no significant difference was found between the two groups ($F(1, 29) = 2.711$, $p = 0.110$). In the 1200-1800 ms time window, there was no significant group difference either ($F(1, 29) = 0.005$, $p = 0.944$). By contrast, the second peak of fixations on the endpoint in L2 Chinese occurred between 1800 ms and 2400 ms. In this time window, L2 Chinese speakers had significantly more fixations on the endpoint than L1 Chinese speakers ($F(1, 29) = 4.695$, $p < .05$). In comparison with L2 Chinese speakers, the second peak of fixations on the endpoint in L1 Chinese occurred later, i.e., between 2400 ms and 3000 ms, during which time there were more fixations on the endpoint in L1 Chinese than in L2 Chinese ($F(1, 29) = 10.500$, $p < .05$). The third peak of fixations on the endpoint in L2 Chinese occurred shortly before the mean value of SOTs in Type B videos (3521 ms), i.e., between 3000 ms and 3600 ms. However, in this time window, no significant difference in fixations on the endpoint was found between the two groups ($F(1, 29) = 0.443$, $p = 0.511$).

As shown in Figure 5.15, univariate ANOVA showed a main effect of the language group ($F(1, 156) = 24.020$, $p < .001$) as well as the time window ($F(5, 156) = 3.876$, $p < .05$). For language group (2) × time window (6), a significant interaction effect ($F(5, 156) = 5.162$, $p < .001$) was also revealed. In the 0-600 ms time window,

fixations on the moving entity increased rapidly in both groups and reached their first peak around 600 ms. In this time window, no significant difference was found between the two groups ($F (1, 26) = 0.059$, $p = 0.811$). In the next two time windows, i.e., between 600 ms and 1200 ms and between 1200 ms and 1800 ms, no significant effect was found, either (600-1200 ms: $F (1, 26) = 0.118$, $p = 0.733$; 1200-1800 ms: $F (1, 26) = .000$, $p = 0.996$). However, in the next three times windows, L2 Chinese speakers had significantly more fixations than L1 Chinese speakers (1800-2400 ms: $F (1, 26) = 4.920$, $p < .05$; 2400-3000 ms: $F (1, 26) = 44.234$, $p < .001$ and 3000-3600 ms: $F (1, 26) = 15.843$, $p < .001$).

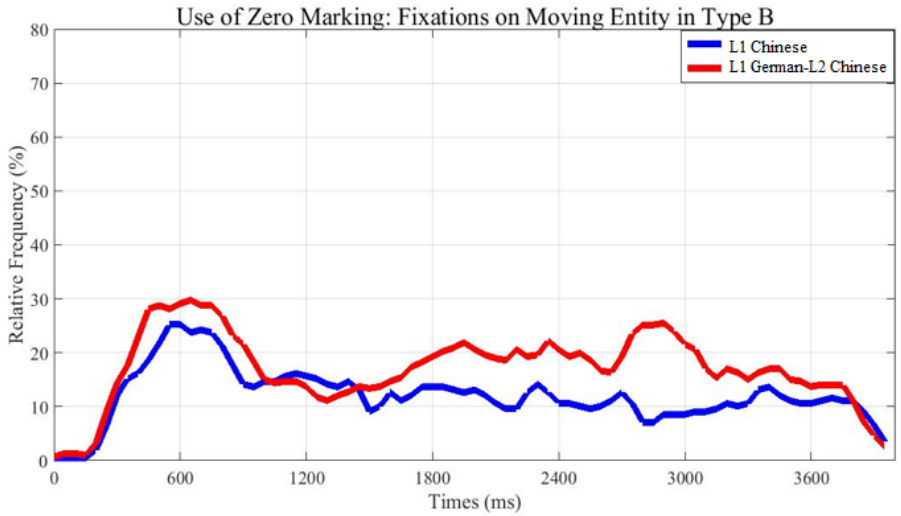


Figure 5.15: Relative frequency of fixations on the moving entity when using zero marking in L1 and L2 Chinese in Type B videos (time locked to the start of the stimulus)

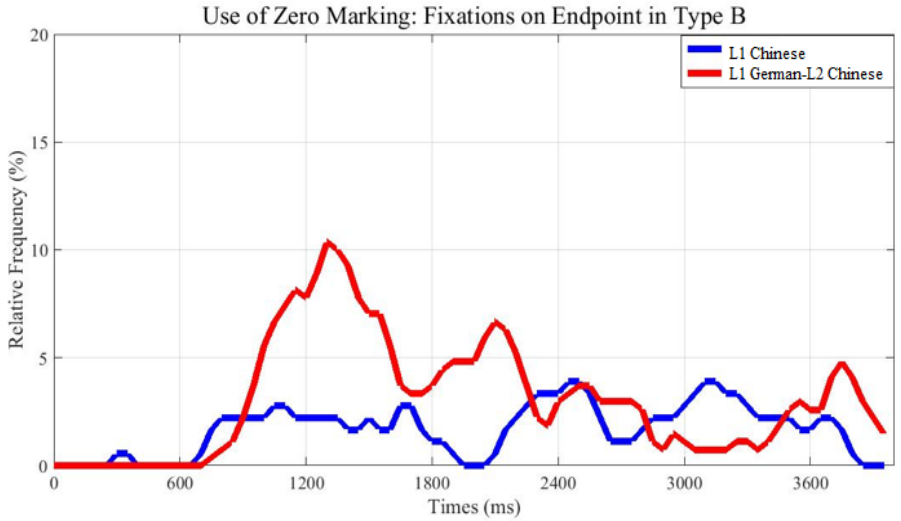


Figure 5.16: Relative frequency of fixations on the endpoint when using zero marking in L1 and L2 Chinese in Type B videos (time locked to the start of the stimulus)

When considering the endpoint fixations under the condition of using zero marking in L1 and L2 Chinese in the verbal task of Type B videos, as shown in Figure 5.16, univariate ANOVA showed a main effect of the language group ($F(1, 138) = 5.988$, $p < .05$) and the time window ($F(5, 138) = 6.819$, $p < .001$). For language group (2) × time window (6), a significant interaction effect ($F(5, 138) = 7.604$, $p < .001$) was found. A separate analysis of the group difference in each time window revealed no significant difference between the two language groups in the 0-600 ms time window ($F(1, 23) = 1.533$, $p = 0.228$) nor in the 600-1200 ms time window ($F(1, 23) = 0.601$, $p = 0.446$). However, fixations on the endpoint increased rapidly between 600 ms and 1200 ms in both groups. L2 Chinese speakers' fixations on the endpoint reached their first peak between 1200 ms and 1800 ms. In this time window, L2 speakers had significantly more fixations than L1 speakers ($F(1, 23) = 7.756$, $p < .05$). In the next time window, i.e., between 1800 ms and 2400 ms, L2 speakers' second endpoint fixations reached the second peak, which was significantly higher than that of L1 speakers ($F(1, 23) = 11.752$, $p < .05$). No significant difference was found between 2400 ms and 3000 ms ($F(1, 23) = 0.065$, $p = 0.801$), while more fixations on the endpoint were found in L1 Chinese between 3000 ms and 3600 ms ($F(1, 23) = 8.414$, $p < .05$).

Type C: Boundary-crossing events

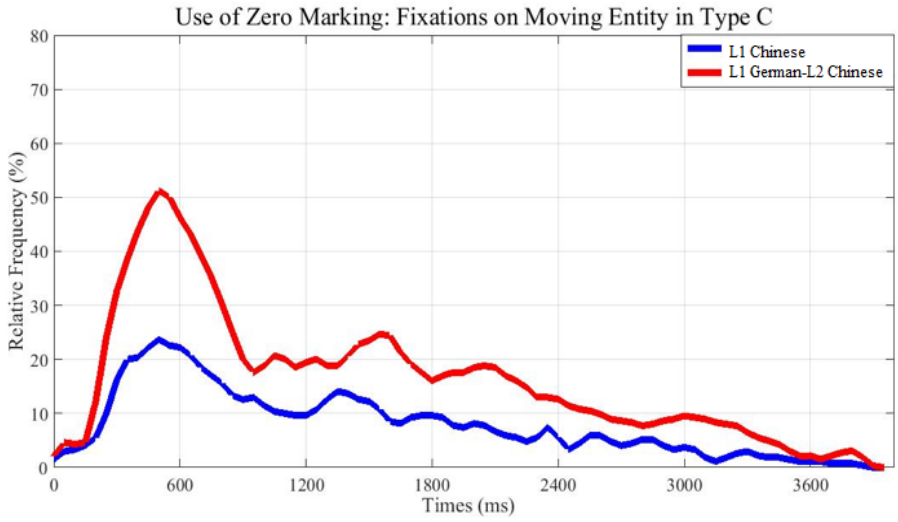


Figure 5.17: Relative frequency of fixations on the moving entity when using zero marking in L1 and L2 Chinese in Type C videos (time locked to the start of the stimulus)

For Type C videos, the fixations on the moving entity and on the endpoint were only compared between L1 and L2 Chinese speakers when zero marking was used. Due

to the rare use of the imperfective aspect markers in L1 and L2 Chinese, fixations on the moving entity and on the endpoint when the imperfective aspect marker was used were not considered. In addition, given that when the perfective aspect marker was used, speakers of all languages paid attention to and referred to the endpoint (see section 4.2.2), no analysis was conducted on fixations on the moving entity or on the endpoint when the perfective aspect marker was used.

When considering the fixations on the moving entity under the condition of using zero marking in L1 and L2 Chinese in the verbal task of Type C videos, as shown in Figure 5.17, univariate ANOVA revealed a main effect of the language group ($F(1, 186) = 122.345$, $p < .001$) and the time window ($F(5, 186) = 6.819$, $p < .001$). However, no significant interaction effect between language group (2) and time window (6) was found ($F(5, 186) = 1.408$, $p = 0.223$). Between 0 ms and 600 ms, fixations on the moving entity increased rapidly and reached their peak in both language groups (see Figure 5.17). However, there were more fixations on the moving entity in L2 Chinese than in L1 Chinese ($F(1, 31) = 16.080$, $p < .001$). After that, fixations on the moving entity gradually decreased over the next five time windows. Still, L2 Chinese speakers fixated significantly more often on the moving entity than L1 Chinese speakers did in each of the next five time windows (600-1200 ms: $F(1, 31) = 20.823$, $p < .001$; 1200-1800 ms: $F(1, 31) = 24.164$, $p < .001$; 1800-2400 ms: $F(1, 31) = 30.827$, $p < .001$; 2400-3000 ms: $F(1, 31) = 12.386$, $p < .05$; 3000-3600 ms: $F(1, 31) = 25.377$, $p < .001$).

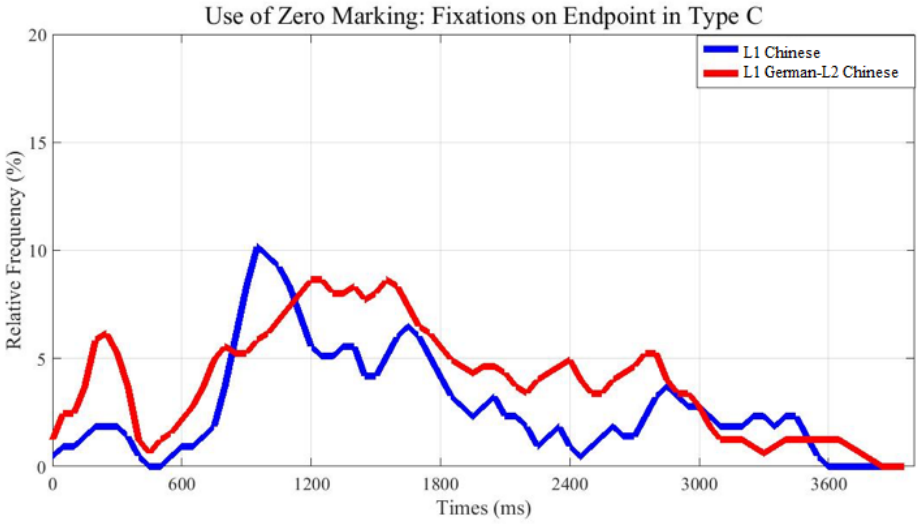


Figure 5.18: Relative frequency of fixations on the endpoint when using zero marking in L1 and L2 Chinese in Type C videos (time locked to the start of the stimulus)

When considering the endpoint fixations under the condition of using zero

marking in L1 and L2 Chinese in the verbal task of Type C videos, as shown in Figure 5.18, univariate ANOVA revealed a main effect of the language group ($F (1, 168) = 29.582$, $p < .001$) and the time window ($F (5, 168) = 7.963$, $p < .001$), as well as a significant language group (2) × time window (6) interaction effect ($F (5, 168) = 2.691$, $p < .05$). In the 0-600 ms time window, L2 Chinese speakers had significantly more fixations on the endpoint than L1 Chinese speakers ($F (1, 28) = 7.561$, $p < .05$). In the next time window, i.e., between 600 ms and 1200 ms, although the fixations on the endpoint reached their peak in L1 Chinese, no significant difference was found between the two groups in this time window ($F (1, 28) = 2.439$, $p = 0.130$). In the time window between 1200 ms and 1800 ms, the peak of the fixations on the endpoint occurred in L2 Chinese; also, in this time window, L2 Chinese speakers had more fixations than L1 Chinese speakers ($F (1, 28) = 7.856$, $p < .05$). This is also true for the next two time windows, i.e., between 1800 ms and 2400 ms ($F (1, 28) = 10.632$, $p < .05$) and between 2400 ms and 3000 ms ($F (1, 28) = 6.155$, $p < .05$), in which L2 Chinese speakers also had more fixations on the endpoint than L1 Chinese speakers. By contrast, no significant group difference in fixations on the endpoint was found between 3000 ms and 3600 ms ($F (1, 28) = 0.188$, $p = 0.668$).

Summary and interpretation

In this section, we analyzed the attention allocated to the moving entity or to the endpoint over the course of time between L1 and L2 Chinese speakers when the imperfective aspect, respectively zero marking was used.

In Type A videos, when the imperfective aspect was used, L1 Chinese speakers tended to fixate on the moving entity, while L2 Chinese speakers did so to a lesser degree and rather preferred fixating on the endpoint. These differences occurred early, i.e., between 0 ms and 1200 ms. In the 0-600 ms time window, fixations on the moving entity reached their peak in both groups; however, L1 Chinese speakers fixated on the endpoint more often than L2 Chinese speakers did. In this time window, very few fixations on the endpoint were registered, and no significant difference in fixations on the endpoint were documented between the two groups. As the time went on, in the 600-1200 ms time window, L1 Chinese speakers continued paying attention to the moving entity, while L2 Chinese speakers fixated more on the endpoint; these fixations on the endpoint reached their peak during this time span. By contrast, when zero marking was used, L2 Chinese speakers had more fixations on the moving entity as well as on the endpoint compared to L1 Chinese speakers. The difference in the

fixations on the moving entity arose as early as between 0 ms and 600 ms, while the difference in the fixations on the endpoint occurred as early as between 600 ms and 1200 ms. In the 0-600 ms time window, fixations on the moving entity reached their peak in both groups; however, L2 Chinese speakers had more fixations than L1 Chinese speakers, which continued until the speech onset of L2 Chinese speakers. In comparison with the fixations on the moving entity in the 0-600 ms time window, very few endpoint fixations were recorded during this time span. In the next time window, i.e., between 600 ms and 1200 ms, however, L2 Chinese speakers began to have more fixations on the endpoint than L1 Chinese speakers. This tendency lasted until 2400 ms and also occurred from 3600 ms to 4200 ms.

In Type B videos, when using imperfective aspect markers, L1 Chinese speakers fixated more often on the moving entity than L2 Chinese speakers did between 0 ms and 1200 ms, whereas speakers of both groups did not differ in the other time windows in this regard. Concerning the attention allocated to the endpoint between the two groups, the first peak of fixations on the endpoint in L2 Chinese (around 1080 ms) was slightly earlier than in L1 Chinese (around 1200 ms). The same is true for the occurrence of the second peak. The second peak of fixations on the endpoint in L2 Chinese occurred around 2240 ms, while the second peak of fixations on the endpoint in L1 Chinese occurred around 2800 ms. The statistical analysis showed that L2 Chinese speakers had more endpoint fixations than L1 Chinese speakers between 1800 ms and 2400 ms, during which time the second peak of fixations on the endpoint arose in L2 Chinese. In turn, L1 Chinese speakers had more fixation on the endpoint in the next time window, i.e., between 2400 ms and 3000 ms, during which time the second peak of fixations on the endpoint occurred in L1 Chinese. When zero marking was used, the difference between the two groups in the fixations on the moving entity occurred between 1800 ms and 3600 ms, during which time L2 Chinese speakers fixated more on the moving entity than L1 Chinese speakers. However, in comparison with the occurrence of the difference in the fixations on the moving entity, the difference between the two groups in the fixations on the endpoint occurred earlier, i.e., between 1200 ms and 1800 ms, during which time L2 Chinese speakers' fixations on the endpoint reached their maximum value. That is, L2 Chinese speakers had significantly more fixations on the endpoint during this time span. In turn, L1 Chinese speakers fixated more on the endpoint than L2 Chinese speakers between 3000 ms and 3600 ms.

In Type C videos, as the imperfective aspect was rarely used in both groups, the

attention allocated to the moving entity or to the endpoint when the imperfective perspective was used was thus not considered. Instead, only the attention allocated to the moving entity and to the endpoint when zero marking was used is considered. When zero marking was used, the difference in the fixations on the moving entity between the two groups occurred in the 0-600 ms time window, during which time the fixations on the moving entity reached their peak in both groups. However, L2 Chinese speakers had more fixations on the moving entity during this time span. The same tendency was found between 1200 ms and 3000 ms, during which time L2 Chinese speakers had more fixations on the moving entity. Likewise, the difference in the fixations on the endpoint between the two groups also occurred between 0 ms and 600 ms. However, the maximum value of the fixations on the endpoint in L1 Chinese speakers (around 1000 ms) occurred slightly earlier than in L2 Chinese speakers (around 1200 ms).

In general, it was found that when the imperfective aspect was used, L1 Chinese speakers tended to have more fixations on the moving entity. This difference between L1 and L2 Chinese speakers occurred between 0 ms and 1200 ms. On the other hand, L2 Chinese speakers tended to have more fixations on the endpoint. This difference arose between 600 ms and 1200 ms in Type A videos and between 1800 ms and 2400 ms in Type B videos. According to von Stutterheim et al. (2012), when speakers of aspect languages, who use the grammaticalized imperfective/progressive aspect frequently, describe goal-oriented motion events, they tend to segment the situation into phases (initial phase, intermediate phase, and terminative phase), while speakers of non-aspect languages, who do not use the grammaticalized imperfective/progressive aspect, tend to take a holistic view, i.e., they tend to pay attention to the endpoint. In line with these findings, we found in our study that L1 Chinese speakers tended to segment the situation into phases and according to the nature of situation, they focused on the different phases. For example, when describing the motion events in Type B videos (long trajectory with possible endpoint), L1 Chinese speakers used more often imperfective, or more specifically, progressive aspect markers (see section 5.2.1). In contrast, L1 German-L2 Chinese speakers used fewer imperfective aspect markers (see section 5.2.1) and although they used the imperfective aspect, traces of L1-mediated conceptualization patterns occurred, i.e., they tended to direct their attention to the endpoint, since the endpoint plays an important role in event construal in L1 German. Furthermore, when zero marking was used, in comparison with L1 Chinese speakers, L2 Chinese speakers tended to pay attention to both the moving entity and the endpoint.

L2 Chinese speakers might have had to "struggle" with which information to select for verbalization. As already discussed in Chapter 4, L1 German speakers tended to structure the path of motion related to the contours of the Ground object or the endpoints, while L1 Chinese speakers tended to either relate the path of motion to the endpoints or conceptualize the situation as "A Figure moving in a certain manner at a certain location". It might be that these two different conceptual patterns – their habitual patterns and the one they have to adapt to – "compete" in L2 Chinese speakers' minds, forcing them to pay attention to both the moving entity and the endpoint.

5.3.2 Allocation of attention in L1 and L2 English

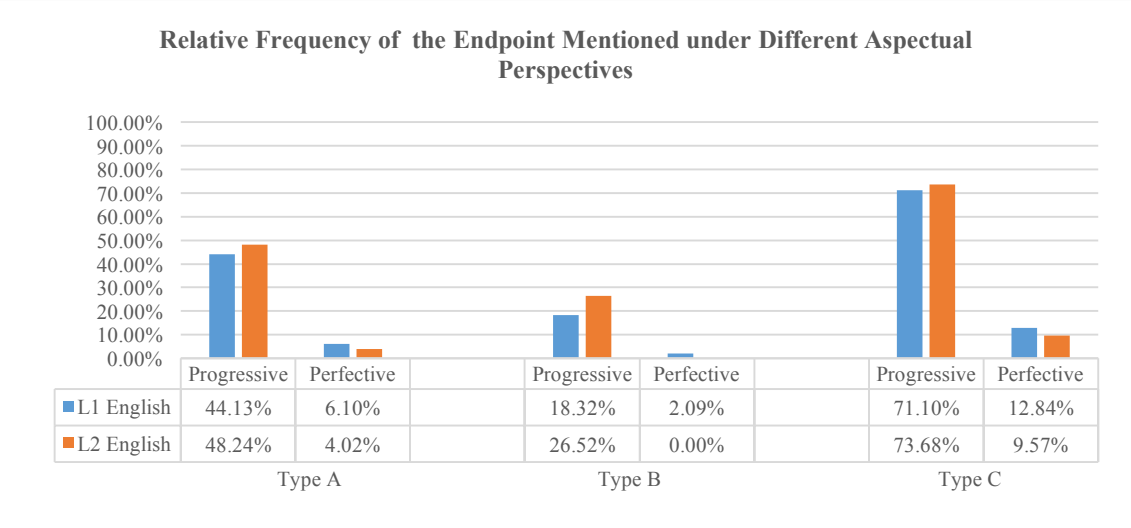


Figure 5.19: Relative frequency of the endpoint being mentioned when using imperfective and aspect marking in L1 and L2 English in the verbal task of Type A, B, and C videos

In Chapter 4, we have shown that speakers of L1, L2 English and L1 German did not differ in the frequency of mentioning the endpoint in the verbal task of Type A, B, and C videos. When considering the use of aspectual markings, 50.23% of the utterances in L1 English mentioned the endpoint in the verbal task of Type A videos. 44.13% of these were marked with the imperfective aspect and 6.10% with the perspective aspect (see Figure 5.19). 52.26% of the utterances in L2 English mentioned the endpoint, of which 48.24% were marked with the imperfective aspect and 4.02% with the perspective aspect. Similarly, in the verbal task of Type B videos, 20.42% of the utterances in L1 English mentioned the endpoint, of which 18.32% were marked with the imperfective aspect and 2.09% with the perspective aspect, while in L2 English, 26.52% of the utterances mentioned the endpoint, all of which were marked with the

imperfective aspect. In the verbal task of Type C videos, a total of 83.94% of the utterances mentioned the endpoint in L1 English, of which 71.10% were marked with the imperfective aspect and 12.84% with the perspective aspect, whereas in L2 English, a total of 83.25% of the utterances mentioned the endpoint, of which 73.68% were marked with the imperfective aspect and 9.57% with the perfective aspect.

From the above-discussed relationship between the mentioning of the endpoint and the use of aspectual marking, especially the imperfective aspect in L1 and L2 English, it was found that when the endpoint was mentioned in an utterance, then this utterance was more likely to be marked with the imperfective aspect. It appears that this observation contradicts the *seeing for speaking* hypothesis, since this hypothesis says that speakers of aspect languages, in which the imperfective/progressive aspect is grammaticalized and used frequently, tend to segment the situation into different phases, and thus pay less attention to the endpoint. However, this hypothesis is actually about the relationship between grammatical forms and visual attention during the conceptualization phase in language production. Quoting von Stutterheim et al.'s own words (2012, p. 835), this hypothesis is about "whether linguistic categories focus speakers' visual attention on certain aspects of a given event while preparing to describe it". In Chapter 4, we showed that speakers of L1, L2 English, and L1 German differed in the fixation on the endpoint before SOTs, i.e., during the conceptualization phase, which validates the *seeing for speaking* hypothesis. The question is then why, when L1 and L2 English speakers mentioned the endpoint, they tended to use the imperfective aspect. The reason might be that on the one hand, the use of the imperfective aspect is by no means incompatible with the mentioning of the endpoint and, on the other hand, as von Stutterheim and Carroll (2006) already explained, speakers of English might at first not pay attention to the endpoint, and after they start to talk, they can also add the endpoint to the sentence that has already been uttered (e.g., A car is driving along the road…to a village). Therefore, it is not that for English speakers, it is not worth mentioning the endpoint; rather, it seems that they tend not to focus on the endpoint in describing goal-oriented motion events during the conceptualization phase.

For this reason, we examined the attention allocated to the moving entity or to the endpoint before SOTs in L1 and L2 English, with the aim of finding out whether there were group differences in the attention allocated to the moving entity or to the endpoint over the course of time before SOTs and, if the answer is positive, when those differences occurred. To achieve this goal, we used the fixations on the moving entity

and on the endpoint before each participant's SOTs in each video clip. Then these fixations were registered in 50 ms time intervals and a plot was drawn before 3000 ms for speakers of L1, L2 English, and L1 German (for more details about SOTs, see Chapter 4). Similar to the analysis for L1 and L2 Chinese, the statistical analysis was conducted on the relative frequency of the fixations on the moving entity and on the endpoint for each participant in consecutive 600 ms time windows (with each time window encompassing 12 time intervals). More specifically, the data point was first analyzed across five consecutive time windows (every 600 ms) among the three language groups and then the group difference in each 600 ms time window was determined.

Type A: Short trajectory with obvious endpoint

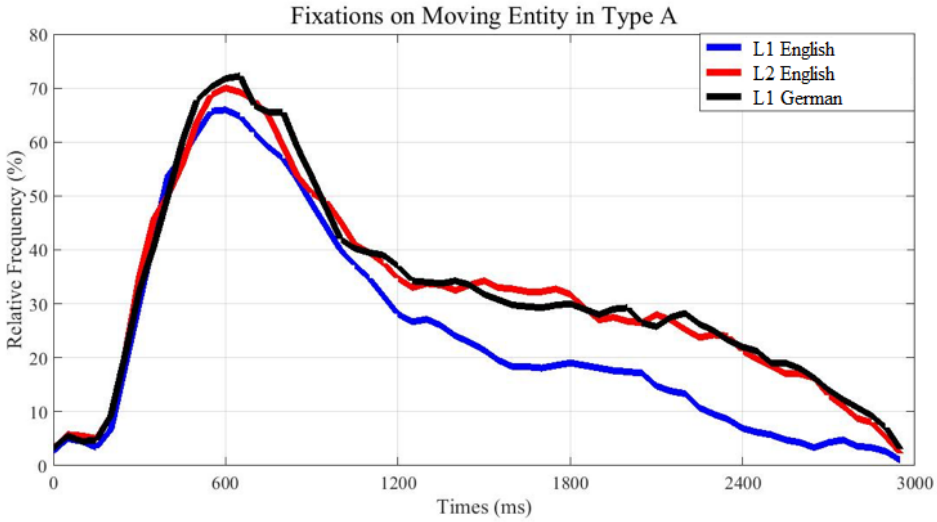


Figure 5.20: Relative frequency of fixations on the moving entity in L1, L2 English, and L1 German before SOTs in Type A videos (time locked to the start of the stimulus)

When considering the fixations on the moving entity before SOTs in L1, L2 English, and L1 German in the verbal task Type A videos, as shown in Figure 5.20, univariate ANOVA showed a main effect of the language group ($F(2, 290) = 19.515$, $p < .001$) and the time window ($F(4, 290) = 65.947$, $p < .001$). However, there was no significant language group (2) × time window (5) interaction effect ($F(8, 290) = 1.309$, $p = 0.238$). An in-depth analysis of the group difference in each 600 ms time window revealed that in the three groups, fixations on the moving entity increased rapidly in the 0-600 ms time window and reached their peak around 600 ms. However, no significant difference was found among the three groups in this time window ($F(2, 58) = 2.037$, $p = 0.140$). Likewise, there was no significant difference in the 600-1200 ms time window

(F (2, 58) = 0.681, p = 0.510). However, a significant group difference was found the in 1200-1800 ms time window (F (2, 58) = 5.304, p < .05) and in the 1800-2400 ms time window (F (2, 58) = 6.732, p < .05), respectively. In these two time windows, Tukey HSD post-hoc tests showed that L1 German speakers had significantly more fixations on the moving entity than L1 English speakers (p < .05), while no significant effect was found between L1 and L2 English speakers nor between L1 German and L2 English speakers (both p > .05). Similarly, a significant effect was also found between 2400 and 3000 ms among speakers of the three groups (F (2, 58) = 10.543, p < .001): Tukey HSD post-hoc tests showed that speakers of L1 German and L2 English fixated on the moving entity more often than L1 English speakers did (both p < .05), while no significant difference was found between speakers of L1 German and L2 English (p > .05).

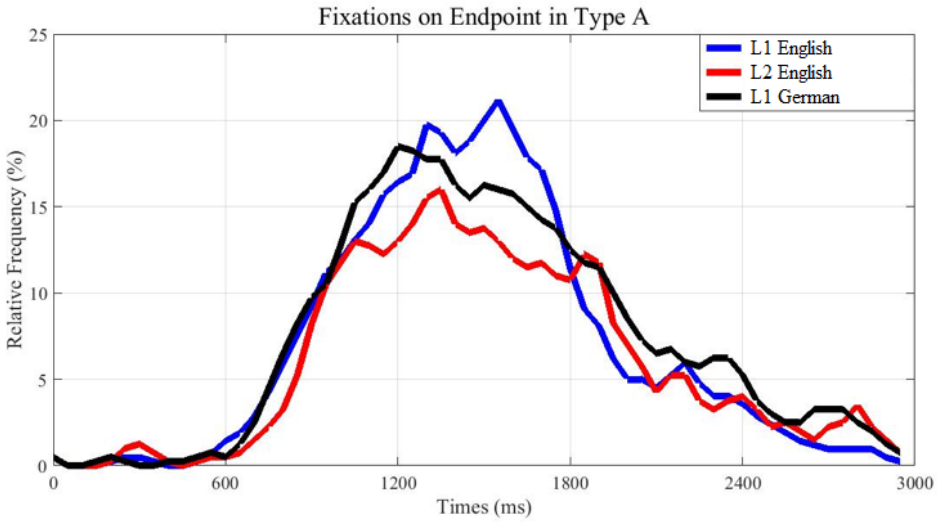


Figure 5.21: Relative frequency of fixations on the endpoint in L1, L2 English, and L1 German before SOTs in Type A videos (time locked to the start of the stimulus)

As shown in Figure 5.21, concerning the fixations on the endpoint in Type A videos in L1, L2 English, and L1 German, univariate ANOVA showed a main effect of the language group (F (2, 290) = 16.639, p < .001) and the time window (F (4, 290) = 129.933, p < .001) as well as a significant language group (2) × time window (5) interaction effect (F (8, 290) = 5.365, p < .001). To investigate the group difference per time window, a separate analysis was conducted for every 600 ms time interval. In the 0-600 ms time window, very few endpoint fixations were registered and no significant difference was found among speakers of the three language groups in this time window (F (2, 58) = 0.976, p = 0.383). By contrast, fixations on the endpoint increased rapidly between 600 ms and 1200 ms in all three groups. L1 German speakers' fixations on the

endpoint reached their maximum value around 1200 ms, while the maximum value of fixations on the endpoint in L1 and L2 English occurred later, namely, between 1200 ms and 1800 ms. In the 600-1200 ms time window, a significant difference was found among speakers of all three groups ($F (2, 58) = 14.981$, $p < .001$). Tukey HSD post-hoc tests showed that L1 German speakers had more fixations than L1 and L2 English speakers (both $p < .05$), while no significant effect was found between L1 and L2 English speakers ($p > .05$). The same holds true for the group difference between 1200 ms and 1800 ms ($F (2, 58) = 10.509$, $p < .001$). Tukey HSD post-hoc tests again showed that more fixations on the endpoint were found in L1 German than in L1 and L2 English (both $p < .05$), whereas no significant difference was found between L1 and L2 English speakers ($p > .05$). In the next two time windows, i.e., between 1800 ms and 2400 ms as well as between 2400 ms and 3000 ms, no group difference was found in these two time windows (1800-2400 ms: $F (2, 58) = 0.441$, $p = 0.645$; 2400-3000 ms: $F (2, 58) = 0.038$, $p = 0.963$).

Type B: Long trajectory with possible endpoint

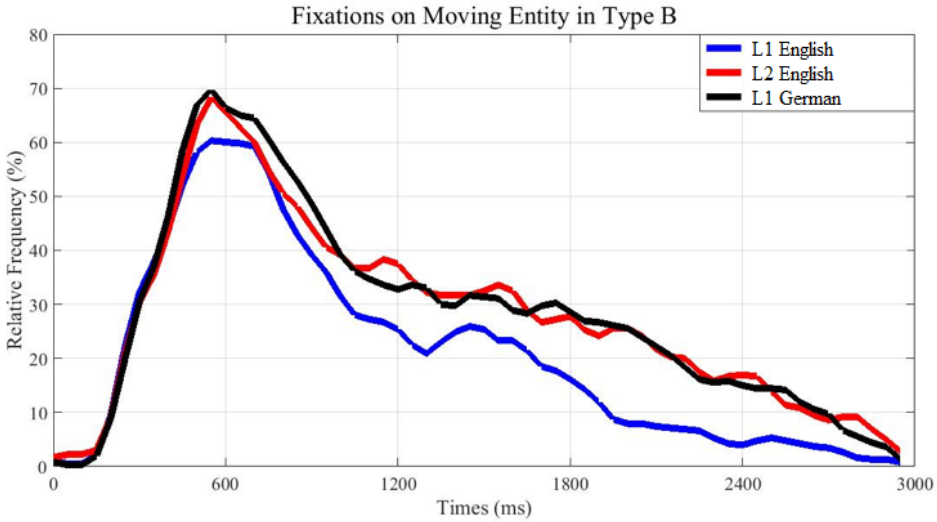


Figure 5.22: Relative frequency of fixations on the moving entity in L1, L2 English, and L1 German before SOTs in Type B videos (time locked to the start of the stimulus)

For the fixations on the moving entity in Type B videos, univariate ANOVA showed a main effect of the language group ($F (2, 290) = 16.088$, $p < .001$) and the time window ($F (4, 290) = 72.912$, $p < .001$), as well as a significant language group (2) × time window (5) interaction effect ($F (8, 290) = 2.697$, $p < .05$) (see Figure 5.22). To explore the group difference in each time window, a detailed analysis was carried out for every 600 ms interval. In the 0-600 ms time window, what was in common in the

three groups was that fixations on the entity increased sharply and reached their peak around 600 ms. The statistical analysis revealed no significant difference in the three groups in this time window ($F (2, 58) = 2.917$, $p = 0.062$). After that, the fixations on the moving entity decreased gradually. No significant group difference was found between 600 ms and 1200 ms and between 1200 ms and 1800 ms, respectively (600-1200 ms: $F (2, 58) = 0.935$, $p = 0.398$; 1200-1800 ms: $F (2, 58) = 2.284$, $p = 0.111$). By contrast, a group difference was found in the last two time windows (1800-2400 ms: $F (2, 58) = 15.815$, $p < .001$; 2400-3000 ms: $F (2, 58) = 10.954$, $p < .001$). Tukey HSD post-hoc tests showed that in these two time windows, speakers of L1 German and L2 English had significantly more fixations on the moving entity than L1 English speakers (both $p < .05$), while no significant difference was found between speakers of L1 German and L2 English ($p > .05$).

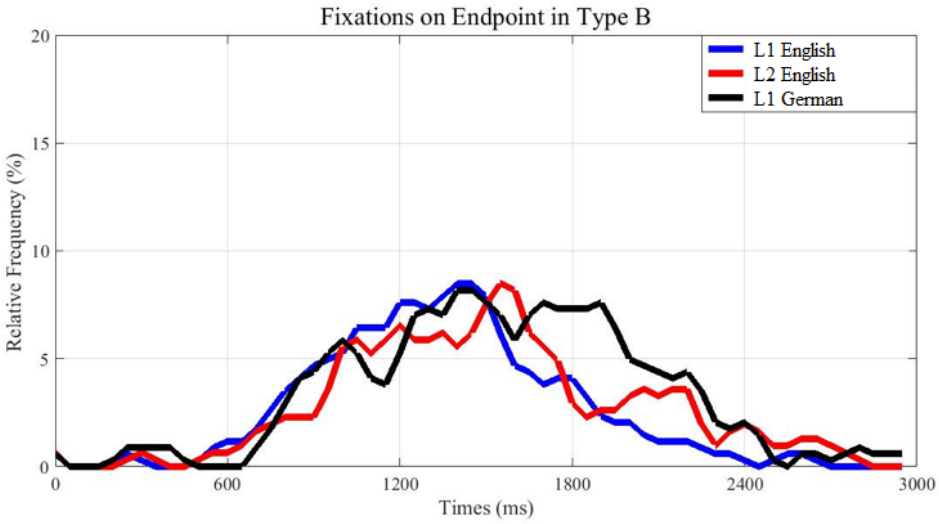


Figure 5.23: Relative frequency of fixations on the endpoint in L1, L2 English, and L1 German before SOTs in Type B videos (time locked to the start of the stimulus)

When we compared the fixations on the endpoint in Type B videos, univariate ANOVA revealed a main effect of the language group ($F (2, 260) = 4.803$, $p < .05$) and the time window ($F (4, 260) = 74.072$, $p < .001$) (see Figure 5.23). However, there was no significant language group (2) × time window (5) interaction effect ($F (8, 260) = 1.571$, $p = 0.134$). A separate analysis for each 600 ms time window showed that very few fixations on the endpoint were registered in the 0-600 ms time window, and no group difference was found in this time window ($F (2, 52) = 0.706$, $p = 0.498$). In all three language groups, fixations on the endpoint increased between 600 ms and 1200 ms and reached their peak between 1200 ms and 1800 ms. In comparison with speakers

of L1 German and L1 English, the maximum value of fixations on the endpoint in L2 English occurred slightly later, between 1200 ms and 1800 ms. However, no significant group difference was found in either of these two time windows (600-1200 ms: $F(2, 52) = 1.108$, $p = 0.338$; 1200-1800 ms: $F(2, 52) = 0.060$, $p = 0.942$). Fixations on the endpoint decreased gradually in the last two time windows in all three groups. Nonetheless, a group difference was found in each of these two time windows, i.e., between 1800 ms and 2400 ms ($F(2, 52) = 4.788$, $p < .05$) and between 2400 ms and 3000 ms ($F(2, 52) = 16.837$, $p < .001$). Tukey post-hoc tests showed that in each of these two time windows, speakers of L1 German and L2 English had more fixations on the endpoint than L1 English speakers (both $p < .05$), while speakers of L1 German and L2 English did not differ in this regard ($p > .05$). That is, both L1 German speakers and L2 English speakers paid more and longer attention to the endpoint than L1 English speakers between 1800 ms and 3000 ms. As shown in Chapter 4, in Type B videos, the mean value of SOTs in L1 English was 1893 ms, while the mean value of SOTs in L2 English was 2577 ms and in German 2596 ms. It appears that speakers of L1 German and L2 English thought longer about how to organize the information, which is presumably related to the endpoint, and then started to talk.

Type C: Boundary-crossing events

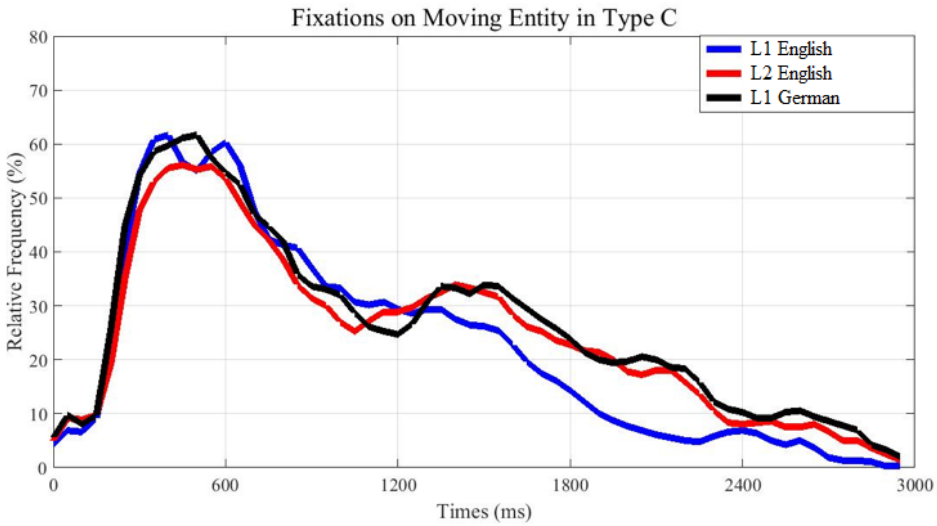


Figure 5.24: Relative frequency of fixations on the moving entity in L1, L2 English, and L1 German before SOTs in Type C videos (time locked to the start of the stimulus)

Regarding the fixations on the moving entity in Type C videos, univariate ANOVA revealed a main effect of the language group ($F(2, 290) = 4.048$, $p < .05$) and the time window ($F(4, 290) = 110.858$, $p < .001$), as well as a significant language

group (2) × time window (5) interaction effect (F (8, 290) = 4.117, p < .001) (see Figure 5.24). An in-depth look at the group difference in each 600 ms time window showed that fixations on the entity increased rapidly and reached their maximum value in all three groups in the 0-600 ms time window, where no significant group difference had been found yet (F (2, 58) = 2.450, p = 0.095). By contrast, there was a group difference between 600 ms and 1200 ms (F (2, 58) = 4.067, p < .05). Tukey HSD post-hoc tests showed that L1 English speakers had significantly more fixations on the entity than L2 English speakers (p < .05), while no significant difference was found between speakers of L1 English and L1 German nor between speakers of L1 German and L2 English (both p > .05). Between 1200 ms and 1800 ms, no group difference was found (F (2, 58) = 0.970, p = 0.385). However, a group difference was found in the 1800-2400 ms time window (F (2, 58) = 9.751, p < .001). Tukey HSD post-hoc tests showed that L1 German speakers fixated more on the moving entity than L1 English speakers did (p < .05), while there was no significant difference between L1 German and L2 English nor between L1 and L2 English. In the last time window, i.e., between 2400 ms and 3000 ms, no group difference was found (F (2, 58) = 2.646, p = 0.08).

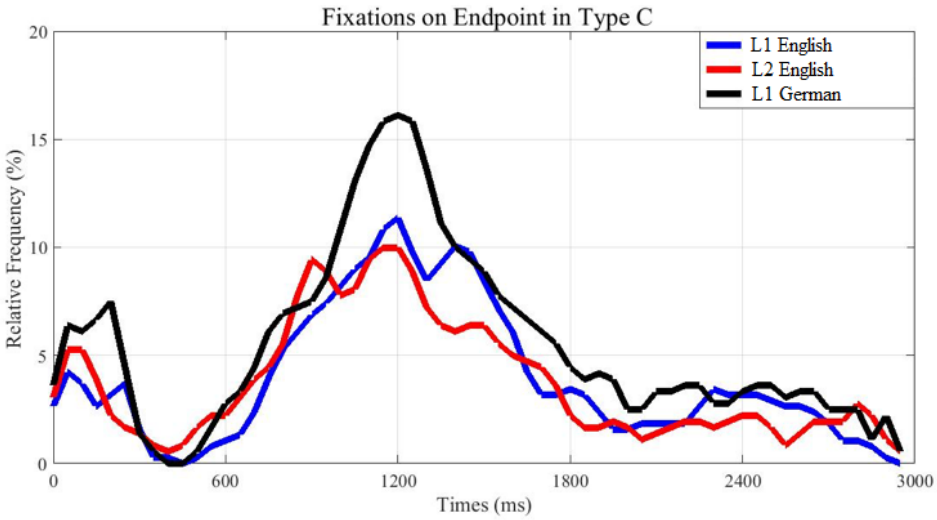


Figure 5.25: Relative frequency of fixations on the endpoint in L1, L2 English, and L1 German before SOTs in Type C videos (time locked to the start of the stimulus)

When considering the fixations on the endpoint in Type C videos, univariate ANOVA was carried out again to test the main and interaction effect of language group and time window. The results showed a main effect of the language group (F (2, 290) = 6.693, p < .05) and the time window (F (4, 290) = 44.575, p < .001), as well as a significant language group (2) × time window (5) interaction effect (F (8, 290) = 3.054,

$p < .05$) (see Figure 5.25). A closer look at the group difference in each 600 ms time window revealed that in the 0-600 ms time window, there were several fixations on the endpoint at the onset of the stimulus in all three groups. After that, these fixations decreased before 600 ms and began to rise again just shortly before 600 ms. In the 0-600 ms time window, a group difference was found ($F (2, 58) = 8.131$, $p < .05$). Tukey HSD post-hoc tests revealed that L1 German speakers fixated more on the endpoint than L1 English speakers did ($p < .05$), while no significant difference was found between L1 German and L2 English speakers nor between L1 and L2 English speakers (both $p > .05$). Around 1200 ms, the fixations on the endpoint reached their maximum value in all three groups. In the 600-1200 ms time window, a group difference is found ($F (2, 58) = 8.752$, $p < .001$). Tukey HSD post-hoc tests showed that both L1 German speakers and L2 English speakers had more fixations than L1 English speakers (both $p < .05$), while L1 German and L2 English speaker did not differ in this regard ($p > .05$). From 1200 ms on, the fixations on the endpoint showed a gradually decreasing tendency in all three groups. In the last three time windows, no group difference was found in each of these time windows (1200-1800 ms: $F (2, 58) = 1.538$, $p = 0.224$; 1800-2400 ms: $F (2, 58) = 0.444$, $p = 0.644$; 2400-3000 ms: $F (2, 58) = 0.572$, $p = 0.567$)

Summary and interpretation

The statistical analysis of the attention allocated to the moving entity showed that between 0 and 600 ms, speakers of all three languages tended to first fixate on the moving entity in the different video types. This is because the moving entity, as its name indicates, was the only moving object in the video clip, while other objects like house or mountain were stationary. It may be easier for movable objects to capture the speakers' attention. Therefore, in the 0-600 ms time window, there was a rapid increase in the number of fixations on the moving entity in all three groups. Flecken et al. (2015) assumed that this time span includes the phase for global event apprehension. That is, speakers of all languages tend to look around in order to get a general view of the video. According to Flecken et al. (2015), this global event apprehension phase lasts about 300 ms for the utterance production task (cf. Griffin & Bock, 2000), during which time very few fixations are registered. But this phase serves as the basis for the further utterance planning processes. What was found to be in common in the different video types was that the fixations on the moving entity reached their peak around 600 ms in all three language groups for all three video types.

Furthermore, speakers of the three groups differed in the attention they paid to

the endpoint over the course of time before 3000 ms in Type A, B, and C videos, respectively. In Type A videos, L1 German speakers had more endpoint fixations than L1 and L2 English speakers from 600 ms to 1800 ms, while L1 and L2 English speakers did not differ in this regard. In addition, the maximum value of fixations on the endpoint in L1 German occurred earlier (at around 1200 ms) than in the other two groups for Type A videos, while the maximum value in L2 English (around 1350 ms) was slightly earlier than that of L1 English speakers (around 1575 ms). That is to say, both speakers of L1 German and L2 English paid attention to the endpoint in Type A videos slightly earlier. In Type B videos, speakers of L1 German and L2 English had more fixations on the endpoint than L1 English speakers between 1800 ms and 3000 ms, while no group difference was found between L1 German and L2 English speakers in this regard. That is, from 1800 ms on, L1 English speakers paid less attention to the endpoint. In contrast, although the fixations on the endpoint also decreased in L1 German and L2 English from 1800 ms on, in comparison with L1 English speakers, speakers of L1 German and L2 English continued to pay attention to the endpoint, which reflects their tendency to focus on the endpoint. In Type C videos, a group difference in the fixations on the endpoint was found between 0 ms and 600 ms as well as between 600 ms and 1200 ms: in these two time windows, L1 German speakers had more fixations on the endpoint than L1 English speakers, while L2 English speakers did not differ from L1 German speakers in this regard. In addition, L2 English speakers did not differ from L1 English speakers between 0 ms and 600 ms. Nevertheless, they had more fixations on the endpoint than L1 English speakers between 600 ms and 1200 ms.

The study of Flecken et al. (2015) showed that L1 French-L2 German speakers tended to allocate their attention to the moving entity, which follows their L1 French's conceptualization patterns, according to which the spatial concepts for structuring the path of motion are derived from the moving entity (e.g., alignment with and distance from the endpoint). Similarly, our study also revealed traces of L1-mediated conceptualization patterns in L1 German-L2 English, since L2 speakers more or less tended to allocate their attention to the endpoint, which is a typical conceptualization pattern in L1 German (see also, e.g., von Stutterheim et al., 2012).

5.4 Summary

In Chapter 4, we compared the lexicalized concept encoded in different

languages and showed the L1 effect on the lexicalized concept. In this chapter, we extended the research on the lexicalized concept to the grammatical concept, i.e., the temporal-aspectual concept, in the domain of motion events.

Languages have different linguistic devices to express time. To encode the temporal concept, generally speaking, German speakers typically use different tense forms, adverbials, and periphrastic constructions, while English speakers typically use different tense and aspect forms and Chinese speakers typically use temporal particles or aspect markers and adverbials. Speakers of all three groups can also use other means to encode the temporal concept, such as discourse principles.

As the occurrence of aspectual marking in L1 German was found to be 0%, the occurrence of aspectual marking in L1 German will not be discussed further. An overview of the use of aspect markers in L1 and L2 Chinese revealed that L1 Chinese speakers used the progressive aspect marker *zài*, the durative aspect marker *zhe*, and the perfective aspect marker *le* more often than L2 Chinese speakers, while L2 Chinese speakers tended to use zero marking in different video types; speakers of both groups did not differ in the use of the prospective aspect marker *yào*. The characteristics of the three different video types had an impact upon the use of different aspect markers in L1 and L2 Chinese to some degree. In the verbal task of Type A videos, L1 Chinese speakers tended to use the progressive aspect marker *zài* and the perfective aspect marker *le*, while L2 Chinese speakers tended to use zero marking. In addition, speakers of both groups did not differ in the use of the durative aspect marker *zhe*. In the verbal task of Type B videos, the prospective aspect marker *yào* was found more often in L2 Chinese than in L1 Chinese. The use of the other aspect markers in the verbal task of Type B videos was the same as for Type A videos: The progressive aspect marker *zài* and the perfective aspect marker *le* were found more often in L1 Chinese, while zero marking was found more frequently in L2 Chinese. There was no difference in the use of the durative aspect marker *zhe* between the two groups. By contrast, in the verbal task of Type C videos, L1 Chinese speakers preferred using the perfective aspect marker *le* to indicate that the event was completed, while L2 Chinese speakers tended to use the perfective aspect less often, but zero marking more often (Chapter 6.3 will provide an explanation for the lesser use of the perfective aspect in L2 Chinese). Furthermore, L2 Chinese speakers used the progressive aspect marker *zài* more often than L1 Chinese speakers in the verbal task of Type C videos, either appropriately or inappropriately. L2 speakers used it appropriately when they divided the whole

situation into two phases, an intermediate and a final phase. *zài* can only be applied for the intermediate phase, but not for the final phase in Chinese. L2 speakers used the progressive aspect marker *zài* inappropriately when they combined *zài* with 2-state expressions or with the postverbal *zài*-phrase.

Both L1 English and L2 English speakers overwhelmingly used imperfective marking in their verbal task. In comparison with L1 German-L2 Chinese speakers, L1 German-L2 English speakers used more imperfective marking. There might be three reasons for this. First, ongoingness is expressed in lexical items in Chinese, but these are not used on an obligatory scale. Second, progressivity is inflexed in verbs in English and is more salient and transparent for L2 speakers to learn. Therefore, L2 English speakers might be more sensitive to progressivity in English than L2 Chinese speakers are to the progressive aspect markers in Chinese. Third, there are restrictions on the use of the progressive aspect marker *zài* in Chinese: *zài* cannot co-occur with 2-state expressions. In contrast, imperfective marking English can co-occur with 2-state expressions. That is, it can be applied to the distinguished phase in English, i.e., the source state of 2-state expressions (e.g., *enter*). For these three reasons, imperfective marking was found to prevail in L2 English.

To find out under which aspectual perspectives L2 Chinese speakers operate to verbalize the endpoint, the utterances in which the endpoint was mentioned were sorted according to the aspectual marking used. In comparison with L1 Chinese speakers, L2 Chinese speakers tended to mention the endpoint when they used zero marking or the imperfective aspect. Another step in this research was to find out whether there was a difference in the attention allocated to the moving entity and to the endpoint over the course of time in L1 and L2 Chinese when zero marking or the imperfective aspect was used. The results showed that L1 and L2 Chinese speakers differed in the attention they allocated to the moving entity and to the endpoint as time went on. When the imperfective aspect was used, L1 Chinese speakers tended to pay attention to the moving entity. This difference occurred between 0 ms and 1200 ms. On the other hand, L2 Chinese speakers tended to pay attention to the endpoint. This difference occurred between 600 ms and 1200 ms in Type A videos and between 1800 ms and 2400 ms in Type B videos. It appears that the endpoint played an important role in event construal for L2 Chinese speakers. When zero marking was used, L2 Chinese speakers tended to pay attention to both the moving entity and the endpoint compared to L1 Chinese speakers, which might reflect their “struggling” with the selection and organization of

information in the source and target language. That is, L1 Chinese speakers tended to conceptualize the path of motion as "A Figure moving in a certain manner at a certain location". This conceptual pattern was more typically found in Type B videos. In contrast, L1 German speakers tended to structure the path of motion related to the contours of the Ground object or to the endpoint. These two different conceptual patterns might have led L2 Chinese speakers to pay attention to both the moving entity and the endpoint.

The relationship between endpoint encoding and aspectual contrasts in L1 and L2 English, i.e., the imperfective and the perfective aspect, shows that when L2 English speakers mentioned the endpoint, they tended to use the imperfective aspect. As von Stutterheim and Carroll (2006) noted, speakers of English tended to segement the situation into different phases and defocused on the endpoint in the conceptualization phase. However, after they started to talk, they can also pay attention to the endpoint and add the endpoint to a sentence that has already been uttered (e.g., A car is driving on the road…to a village). Therefore, it is more significant to investigate the attention allocated to the moving entity or to the endpoint over the course of time before SOTs, i.e., during the conceptualization phase, in L1 and L2 English, rather under the aspectual contrasts of imperfective and perfective. The fixations on the moving entity reached their peak around 600 ms by speakers of all three groups, while there was a great difference among speakers of the three groups with regard to the attention paid to the endpoint. In Type A videos, L1 German speakers fixated more on the endpoint than L1 and L2 English speakers between 600 ms and 1800 ms, while L1 and L2 English speakers did not differ in this regard. In Type B videos, both L1 German and L2 English speakers had more endpoint fixations than L1 English speakers between 1800 ms and 3000 ms, while L1 German speakers and L2 English speakers did not differ in this regard. In Type C videos, L1 German speakers fixated more on the endpoint than L1 English speakers as early as between 0 ms and 600 ms, while L1 and L2 English speakers did not differ during this time span. However, between 600 ms and 1200 ms, both speakers of L2 English and L1 German had more endpoint fixations than L1 English speakers. In contrast, speakers of L2 English and L1 German did not differ in endpoint fixaitons between 600 ms and 1200 ms. That is to say, L2 English speakers differed more or less from L1 English speakers in their early attention to the endpoint, which plays an essential role in motion event construal in L2 English.

Chapter 6 From Space to Time: The Interaction between Spatial and Temporal Concepts

In this study, we showed the encoding of spatial concepts in Chapter 4 and the use of aspectual marking in Chapter 5 regarding motion event descriptions in L1s and L2s. The next question is then how these two categories interact with each other among speakers of different languages. Von Stutterheim et al. (2017) demonstrated the importance of relating spatial concepts with temporal concepts in describing goal-oriented motion events. Their study on two Arabic varieties showed that although Tunisian Arabic and Modern Standard Arabic are closely related, they differ in the linguistic repertoire they have to express spatial and temporal concepts. When describing motion events, Modern Standard Arabic has various linguistic forms to encode spatial concepts, while Tunisian Arabic has few such forms. But it has rich aspectual forms. The authors found that when describing goal-oriented motion events, speakers of Tunisian Arabic tended to use aspectual forms, i.e., the progressive aspect for directed motion to convey information that is encoded in expressions for spatial concepts in Modern Standard Arabic. In order to explore the various options of motion event encoding systematically, it is important to examine the interaction between spatial and temporal concepts in L1s and L2s. In what follows, we will identify the correlation between spatial concepts and aspectual marking in L1 and L2 Chinese as well as in L1 and L2 English, respectively.

6.1 Spatial concepts with aspectual marking in L1 and L2 Chinese

The data collected in the verbal task in L1 and L2 Chinese may include spatial concepts without Path specification (i.e., manner verbs occurring alone or manner verbs occurring together with locative information), or spatial concepts with Path specification (i.e., manner verbs occurring together with path verbs or adjuncts or path verbs occurring alone). Therefore, depending on the presence and absence of the Path information encoded in the spatial concepts, the spatial concepts with aspectual marking can be categorized as spatial concepts with Path specification with aspectual marking and spatial concepts without Path specification with aspectual marking. The

percentages of each construction with aspectual marking were calculated as the frequency of occurrence of this type of construction with different aspectual markings used over the total number of utterances. For example, there were a total of 251 utterances in L2 Chinese in the verbal task of Type A videos, of which thirteen utterances were made up of bare manner verbs (i.e., manner verbs occurring alone without Path specification) marked with the progressive aspect marker *zài*. One utterance consisted of bare manner verbs with the durative aspect marker *zhe* and five utterances contained bare manner verbs with zero marking. So in the verbal task of Type A videos, the percentage of bare manner verbs used together with the progressive aspect marker *zài* was 5.18% (13/251), that of bare manner verbs used with the durative aspect marker was 0.40% (1/251), and that of zero marking was 1.99% (5/251).

All data was analyzed on a qualitative basis. A quantitative analysis was only conducted if there was an appropriate number of utterances for such a statistical analysis (a chi-squared test requires that there are at least five data points for each level of a factor). So whenever there was a low number of utterances, no statistical analysis was conducted for the interaction between different factors.

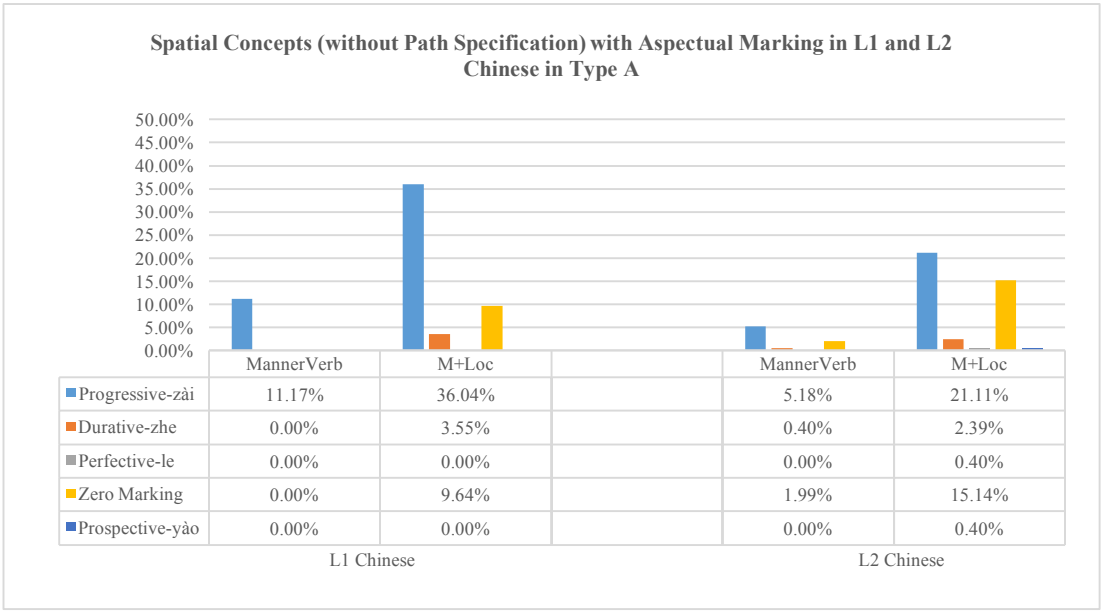
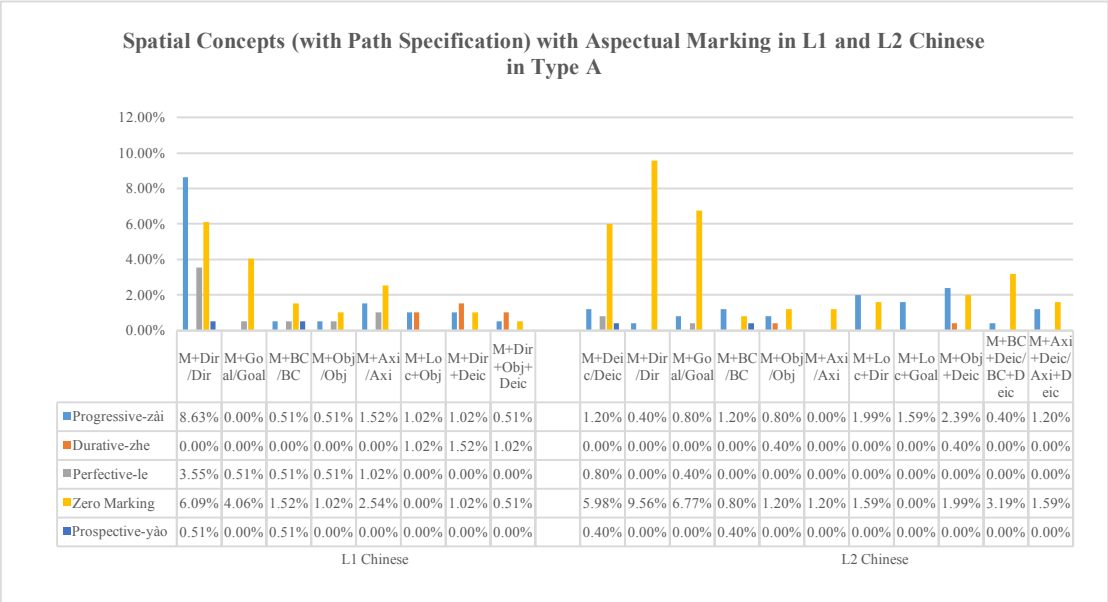


Figure 6.1: Aspectual marking for spatial concepts without Path specification in L1 and L2 Chinese in the verbal task of Type A videos

As shown in Figure 6.1, in the verbal task of Type A videos, when L1 Chinese speakers used bare manner verbs (i.e., manner verbs occurring alone), they tended to mark them with the progressive aspect marker *zài* (11.17%), while L2 Chinese speakers

also used the progressive aspect marker *zài* to mark utterances with a bare manner verb (5.18%), but to a lesser degree. Besides the use of the progressive aspect marker, L2 speakers also used zero marking for utterances with bare manner verbs (1.99%). Concerning the use of locative spatial concepts, L1 Chinese speakers tended to mark these with the progressive aspect marker *zài* (36.04%), followed by zero marking (9.64%) and the durative aspect marker *zhe* (3.55%). In contrast, L2 Chinese speakers made less use of the progressive aspect marker *zài* (21.11%) and the durative aspect marker *zhe* (2.39%), but used zero marking more often (15.14%).



	M+Dir /Dir	M+Goal/Goal	M+BC /BC	M+Obj /Obj	M+Axi /Axi	M+Loc+Obj	M+Dir +Deic	M+Dir +Obj+ Deic	M+Deic/Deic	M+Dir /Dir	M+Goal/Goal	M+BC /BC	M+Obj /Obj	M+Axi /Axi	M+Loc+Dir	M+Loc+Goal	M+Obj +Deic	M+BC +Deic/ BC+Deic	M+Axi +Deic/ Axi+Deic
Progressive-zài	8.63%	0.00%	0.51%	0.51%	1.52%	1.02%	1.02%	0.51%	1.20%	0.40%	0.80%	1.20%	0.80%	0.00%	1.99%	1.59%	2.39%	0.40%	1.20%
Durative-zhe	0.00%	0.00%	0.00%	0.00%	0.00%	1.02%	1.52%	1.02%	0.00%	0.00%	0.00%	0.00%	0.40%	0.00%	0.00%	0.00%	0.40%	0.00%	0.00%
Perfective-le	3.55%	0.51%	0.51%	0.51%	1.02%	0.00%	0.00%	0.00%	0.80%	0.00%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Zero Marking	6.09%	4.06%	1.52%	1.02%	2.54%	0.00%	1.02%	0.51%	5.98%	9.56%	6.77%	0.80%	1.20%	1.20%	1.59%	0.00%	1.99%	3.19%	1.59%
Prospective-yào	0.51%	0.00%	0.51%	0.00%	0.00%	0.00%	0.00%	0.00%	0.40%	0.00%	0.00%	0.40%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Figure 6.2: Aspectual marking for spatial concepts with Path specification in L1 and L2 Chinese in the verbal task of Type A videos

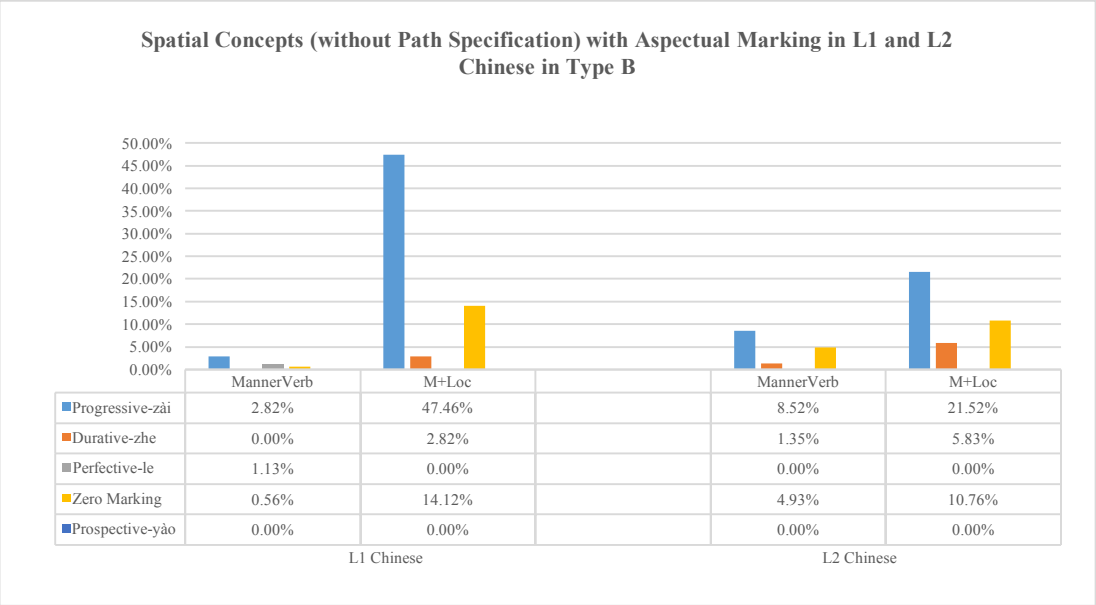
In the data of this study, there are various constructions with Path specification that are compatible with different aspectual markings. Given the very low occurrence of some of these constructions, they will not be discussed further here. So Figure 6.2 only shows a part of those constructions that occurred with different aspect markers in Chinese[91]. The frequency of occurrence of the combination of the spatial concept "manner+ deictic" amounted to 8.38% in L2 Chinese, with 5.38% being used with zero marking and 1.20% with the progressive aspect marker *zài*. The use of directional spatial concepts amounted to 18.78% in L1 Chinese, with 8.63% being marked with the progressive aspect marker *zài*, 6.09% with zero marking, and 3.55% with the

[91] In this bar chart, Deic is the short form for deictic concept, Dir for direction, BC for boundary-crossing, Obj for object-related concept, and Axi for axial concept. M+Dir/Dir means the occurrence of the spatial concepts Manner + Direction or the occurrence of the spatial concept Direction.

perfective aspect marker *le*, while the use of directional spatial concepts amounted to 9.96% in L2 Chinese, most of which consisted of zero marking (9.56%); very few were marked with the progressive aspect marker *zài* (0.40%). Concerning goal-oriented concepts, their use amounted to 4.57% in L1 Chinese and 7.97% in L2 Chinese. In L1 Chinese, 4.06% of the respective utterances used zero marking, whereas in L2 Chinese, 6.77% used zero marking. There was rather low occurrence of boundary-crossing concepts in both L1 and L2 Chinese. In L1 Chinese, 1.52% were used with zero marking. By contrast, in L2 Chinese, 1.20% were marked with the progressive aspect marker *zài*. The co-occurrence of boundary-crossing concepts and the progressive aspect marker *zài* may have led to unacceptable utterances in L2 Chinese, since the progressive aspect *zài* is not compatible with 2-state expressions like *jìn* 'enter' in Chinese. There was also low occurrence of object-related concepts in L1 and L2 Chinese. 1.02% were used with zero marking in L1 Chinese, while 1.20% were used with zero marking in L2 Chinese. Regarding the use of axial concepts (with or without a manner verb), 2.54% were used with zero marking, 1.52% with the progressive aspect *zài*, and 1.02% with the perfective aspect marker *le* in L1 Chinese, while 1.20% of the utterances used zero marking in L2 Chinese.

In general, the occurrence of a manner verb with two or more path adjuncts or verbs was low in L1 and L2 Chinese. Regarding the construction "manner + location + direction", 1.99% of the utterances were marked with the progressive aspect and 1.59% with zero marking in L2 Chinese. In L1 Chinese, 1.02% were marked with the imperfective aspect for the construction "manner + location + object". Regarding the construction "(manner) + direction + deictic", 1.02% were marked with the progressive aspect, 1.52% with the durative aspect, and 1.02% with zero marking. In comparison with L1 Chinese, the combination of different spatial concepts was found to be more varied in L2 Chinese: 1.59% of the utterances were marked with the progressive aspect for the construction "manner + location + goal", and 1.20% were marked with zero marking for the construction "manner + direction + goal". Regarding aspectual marking for combinations of deictic concepts with other spatial concepts, it was found that when deictic concepts occurred with spatial concepts like object, boundary-crossing, or axial, they tended to be unmarked by any aspect markers: 2.39% were marked with the progressive aspect and 1.99% had zero marking for the construction "manner + object + deictic"; 3.19% had zero marking for the construction "manner + boundary-crossing + deictic"; 1.59% had zero marking and 1.20% were marked with the progressive aspect

for the construction “manner + axial + deictic”. Note that some of the above-listed constructions with the progressive aspect in L2 Chinese are not acceptable or sound unnatural. These constructions include manner + object + deictic, manner + boundary-crossing + deictic, manner + axial + deictic, and manner + location + object + deictic, which are 2-state expressions, which makes them incompatible with the progressive aspect marker *zài*.

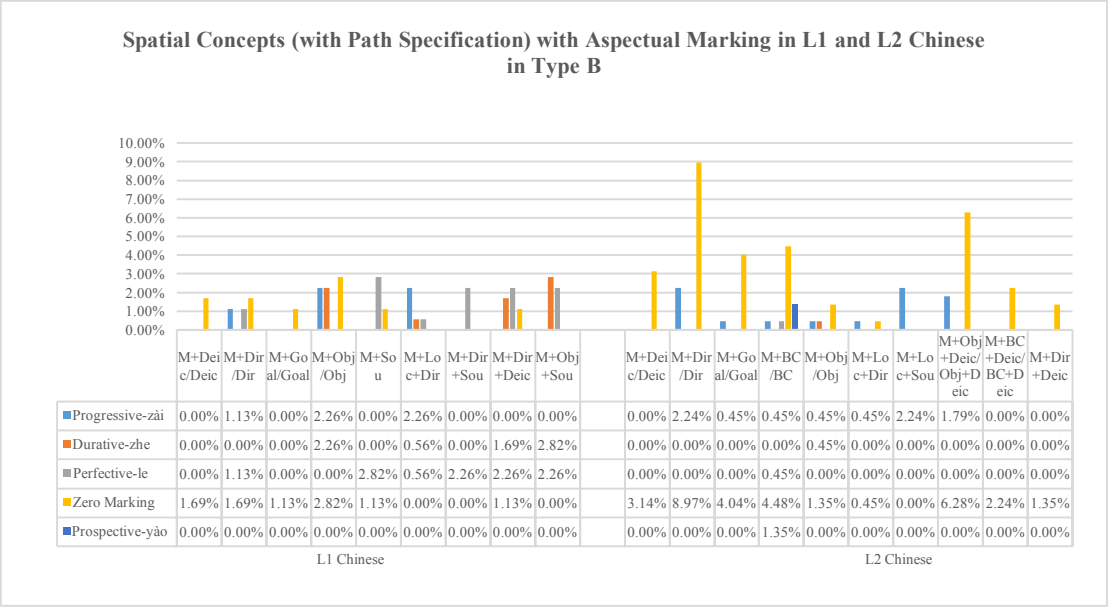


	MannerVerb	M+Loc		MannerVerb	M+Loc
Progressive-zài	2.82%	47.46%		8.52%	21.52%
Durative-zhe	0.00%	2.82%		1.35%	5.83%
Perfective-le	1.13%	0.00%		0.00%	0.00%
Zero Marking	0.56%	14.12%		4.93%	10.76%
Prospective-yào	0.00%	0.00%		0.00%	0.00%

Figure 6.3: Aspectual marking for spatial concepts without Path specification in L1 and L2 Chinese in the verbal task of Type B videos

As shown in Figure 6.3, in the verbal task of Type B videos, the frequency of occurrence of bare manner verbs amounted to 4.51% in L1 Chinese and to 14.8% in L2 Chinese. respectively. In L1 Chinese, 2.82% of bare manner verbs were marked with the progressive aspect, 1.13% with the perfective aspect, and 0.56% had zero marking, while in L2 Chinese, 8.52% of bare manner verbs were marked with the progressive aspect, 4.93% with zero marking, and 1.35% with the durative aspect. When considering aspectual marking with locative concepts, it was found that the frequency of occurrence of “manner + location” amounted to 64.40% in L1 Chinese, of which 47.46% occurrences were marked with the progressive aspect, 14.12% with zero marking, and 2.82% with the durative aspect. In sharp contrast, the frequency of occurrence of the same construction “manner + location” amounted to 38.11% in L2 Chinese, which is significantly lower than that in L1 Chinese. In these constructions, 21.52% were marked with the progressive aspect, 10.76% with zero marking, and 5.83%

with the durative aspect in L2 Chinese.



	M+Deic/Deic	M+Dir/Dir	M+Goal/Goal	M+Obj/Obj	M+Sou	M+Loc+Dir	M+Dir+Sou	M+Dir+Deic	M+Obj+Sou	M+Deic/Deic	M+Dir/Dir	M+Goal/Goal	M+BC/BC	M+Obj/Obj	M+Loc+Dir	M+Loc+Sou	M+Obj+Deic/Obj+Deic	M+BC+Deic/BC+Deic	M+Dir+Deic
Progressive-zài	0.00%	1.13%	0.00%	2.26%	0.00%	2.26%	0.00%	0.00%	0.00%	0.00%	2.24%	0.45%	0.45%	0.45%	0.45%	2.24%	1.79%	0.00%	0.00%
Durative-zhe	0.00%	0.00%	0.00%	2.26%	0.00%	0.56%	0.00%	1.69%	2.82%	0.00%	0.00%	0.00%	0.00%	0.45%	0.00%	0.00%	0.00%	0.00%	0.00%
Perfective-le	0.00%	1.13%	0.00%	0.00%	2.82%	0.56%	2.26%	2.26%	2.26%	0.00%	0.00%	0.00%	0.45%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Zero Marking	1.69%	1.69%	1.13%	2.82%	1.13%	0.00%	0.00%	1.13%	0.00%	3.14%	8.97%	4.04%	4.48%	1.35%	0.45%	0.00%	6.28%	2.24%	1.35%
Prospective-yào	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.35%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

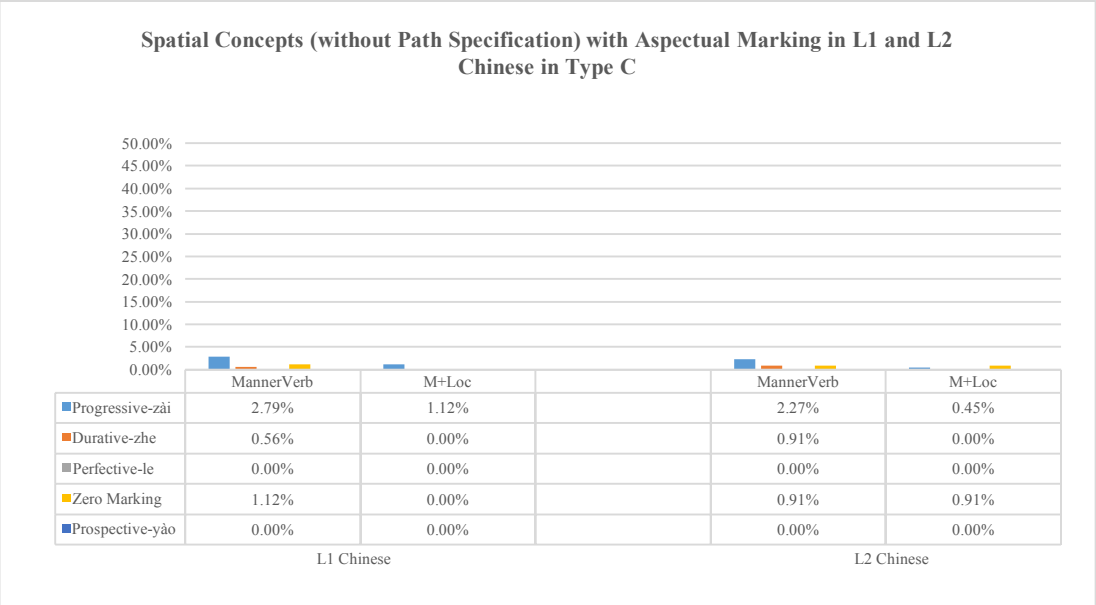
Figure 6.4: Aspectual marking for spatial concepts with Path specification in L1 and L2 Chinese in the verbal task of Type B videos

As shown in Figure 6.4, deictic concepts were encoded with zero marking in both groups: 1.69% in L1 Chinese and 3.14% in L2 Chinese were used with zero marking[92]. The frequency of directional concept encodings was higher in L2 Chinese (11.21%) than in L1 Chinese (3.95%). In L1 Chinese, 1.69% were used with zero marking, 1.13% with the progressive and the perfective aspect, respectively, while in L2 Chinese, 8.97% were used with zero marking and 2.24% with the progressive aspect. The same is true for the encoding of goal-oriented concepts, which was found more often in L2 Chinese (4.49%) than in L1 Chinese (1.13%): all of them were used with zero marking in L1 Chinese (1.13%), while 4.04% were used with zero marking in L2 Chinese. Regarding the encoding of object-related concepts, 2.82% were used with zero marking in L1 Chinese and 2.26% with the progressive and the durative aspect, respectively, while in L2 Chinese, 1.35% were used with zero marking. In comparison with L1 Chinese speakers, L2 Chinese speakers also encoded boundary-crossing concepts (6.73%), of which 4.48% were used with zero marking and 1.35% with the prospective aspect.

Just as in the verbal task of Type A videos, the occurrence of a manner verb with two path adjuncts or verbs was low in L1 and L2 Chinese in the verbal task of

[92] Sou is the short form of source in Figure 6.4.

Type B videos. L2 Chinese speakers tended to combine deictic concepts with other spatial concepts and use zero marking for those combinations: Regarding the construction "manner + direction + deictic", 2.26% were marked with the perfective aspect, 1.69% with the durative aspect, and 1.13% with zero marking in L1 Chinese, while 1.35% were used with zero marking in L2 Chinese; regarding the construction "manner + object + deictic", 6.28% were used with zero marking and 1.79% with the progressive aspect in L2 Chinese; regarding the construction "manner + boundary-crossing + deictic", 2.24% were used with zero marking in L2 Chinese. That is, when deictic concepts occurred with other spatial concepts, L2 Chinese speakers tended to leave the constructions unmarked by any aspect markers.



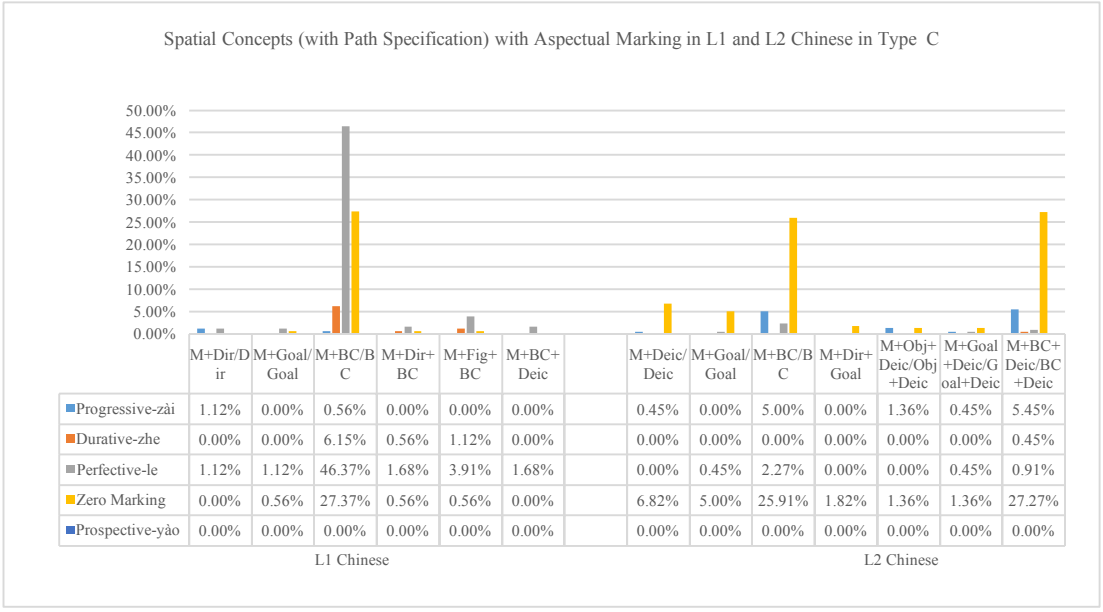
	L1 Chinese			L2 Chinese	
	MannerVerb	M+Loc		MannerVerb	M+Loc
Progressive-zài	2.79%	1.12%		2.27%	0.45%
Durative-zhe	0.56%	0.00%		0.91%	0.00%
Perfective-le	0.00%	0.00%		0.00%	0.00%
Zero Marking	1.12%	0.00%		0.91%	0.91%
Prospective-yào	0.00%	0.00%		0.00%	0.00%

Figure 6.5: Aspectual marking for spatial concepts without Path specification in L1 and L2 Chinese in the verbal task of Type C videos

As shown in Figure 6.5, in the verbal task of Type C videos, L1 and L2 Chinese speakers did not differ in marking their utterances with bare manner verbs with different aspect markers: 2.79% were marked with the progressive aspect and 1.12% with zero marking in L1 Chinese, while 2.27% were marked with the progressive aspect, and 0.91% with the durative aspect and zero marking, respectively, in L2 Chinese. There were only low numbers of occurrences of the construction "manner + Loc" in both groups in the verbal task of Type C videos.

Type C videos showed boundary-crossing events, so speakers of both groups tended to encode boundary-crossing concepts rather than other spatial concepts.

However, as shown in Figure 6.6, L1 and L2 Chinese speakers differed significantly in terms of the correlation of boundary-crossing concepts with aspectual marking ($\chi 2$ (4, N = 217) = 73.737, p < .001): L1 Chinese speakers preferred marking them with the perfective aspect marker *le* (46.37%), while L2 Chinese speakers used the perfective aspect marker significantly less often (2.27%) and preferred using zero marking instead (25.91%). Apart from using the perfective aspect marker, L1 Chinese speakers also used zero marking (27.37%), the durative aspect marker (6.15%), and the progressive aspect marker *zhèng*[93] (0.56%) to mark boundary-crossing concepts, whereas L2 Chinese speakers used the progressive aspect marker *zài* with boundary-crossing path verbs, which led to unacceptable utterances in Chinese. This is because *zài* is not compatible with 2-state expressions, while *zhèng* "at the moment" is punctual in nature and thus can occur together with the boundary-crossing path verb *jìn* 'enter' in Chinese. Deictic concepts tended to be used with zero marking (6.82%) in L2 Chinese. This was true for the goal-oriented concepts in L2 Chinese (5.00% were used with zero marking).



	M+Dir/Dir	M+Goal/Goal	M+BC/BC	M+Dir+BC	M+Fig+BC	M+BC+Deic	M+Deic/Deic	M+Goal/Goal	M+BC/BC	M+Dir+Goal	M+Obj+Deic/Obj+Deic	M+Goal+Deic/Goal+Deic	M+BC+Deic/BC+Deic
Progressive-zài	1.12%	0.00%	0.56%	0.00%	0.00%	0.00%	0.45%	0.00%	5.00%	0.00%	1.36%	0.45%	5.45%
Durative-zhe	0.00%	0.00%	6.15%	0.56%	1.12%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.45%
Perfective-le	1.12%	1.12%	46.37%	1.68%	3.91%	1.68%	0.00%	0.45%	2.27%	0.00%	0.00%	0.45%	0.91%
Zero Marking	0.00%	0.56%	27.37%	0.56%	0.56%	0.00%	6.82%	5.00%	25.91%	1.82%	1.36%	1.36%	27.27%
Prospective-yào	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Figure 6.6: Aspectual marking for spatial concepts with Path specification in L1 and L2 Chinese in the verbal task of Type C videos

Just as in the verbal task of Type A and B videos, the occurrence of a manner verb with two or more path adjuncts or verbs with aspectual marking was low in L1 and L2 Chinese, except for the construction "manner + boundary-crossing + deictic" in

[93] As already discussed in Chapter 5, the progressive aspect marker *zhèng* meaning "at the moment" can occur together with the boundary-crossing path verb *jìn* 'enter'.

L2 Chinese. This construction tended to be used with zero marking (27.27%), followed by the use of the progressive aspect marker (5.45%) and the perfective aspect marker (0.91%). Note that its co-occurrence with the progressive aspect marker may have led to unacceptable utterances in L2 Chinese. In contrast, L1 Chinese speakers used this construction significantly less often and marked it with the perfective aspect marker (1.68%). When considering the combination of deictic concepts with other spatial concepts, it was found that L2 Chinese speakers tended to leave those combinations unmarked by any aspect markers: Regarding the construction "manner + object + deictic", L2 Chinese speakers used zero marking (1.36%) and the progressive aspect to mark it (1.36%); regarding the construction "manner + goal + deictic", L2 Chinese speakers used zero marking (1.36%) to mark it.

To summarize, L1 Chinese speakers preferred using the progressive aspect marker to mark the construction "manner + Location", especially in the verbal task of Type B videos. In comparison with L1 Chinese speakers, L2 Chinese speakers used the progressive aspect marker less often to make this construction. Instead, for the various spatial concepts, they preferred zero marking. This was extremely remarkable in the verbal task of Type C videos, where L2 Chinese speakers preferred zero marking for boundary-crossing concepts, whereas L1 Chinese speakers preferred using the perfective aspect to mark these concepts.

6.2 Spatial concepts with aspectual marking in L1 and L2 English

Just as in L1 and L2 Chinese, the linguistic data collected in the verbal task in L1 and L2 English may include spatial concepts that do not specify Path information (i.e., manner verbs occurring alone or manner verbs occurring together with locative adjuncts) or spatial concepts that specify Path information (i.e., manner verbs occurring together with path adjuncts or path verbs occurring alone). In the following, we will present the interaction between spatial concepts and aspectual marking from these two perspectives. The calculation was the same as in section 6.1. Given the low number of occurrence of some combinations of spatial concepts, these will not be listed and discussed further.

As shown in Figure 6.7, in the verbal task of Type A videos, different

combinations of spatial concepts tended to be marked with the imperfective aspect by L1 and L2 English speakers. Bare manner verbs occurring alone without further specification of path information were found in 4.23% of the utterances in L1 English and in 1.51% in L2 English, all of which were marked with the imperfective aspect. Regarding the construction "manner verb + locative adjunct", 9.86% were marked with the imperfective aspect in L1 English and 10.05% in L2 English.

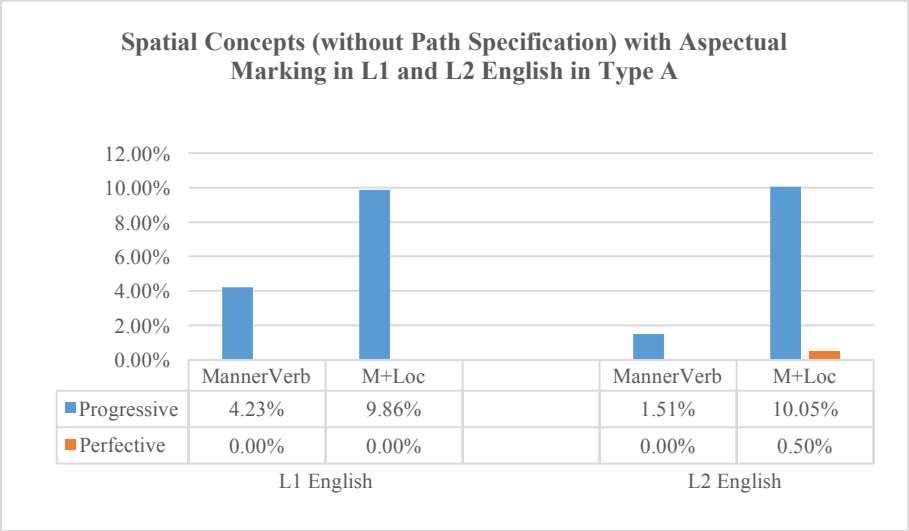


Figure 6.7: Aspectual marking for spatial concepts without Path specification in L1 and L2 English in the verbal task of Type A videos

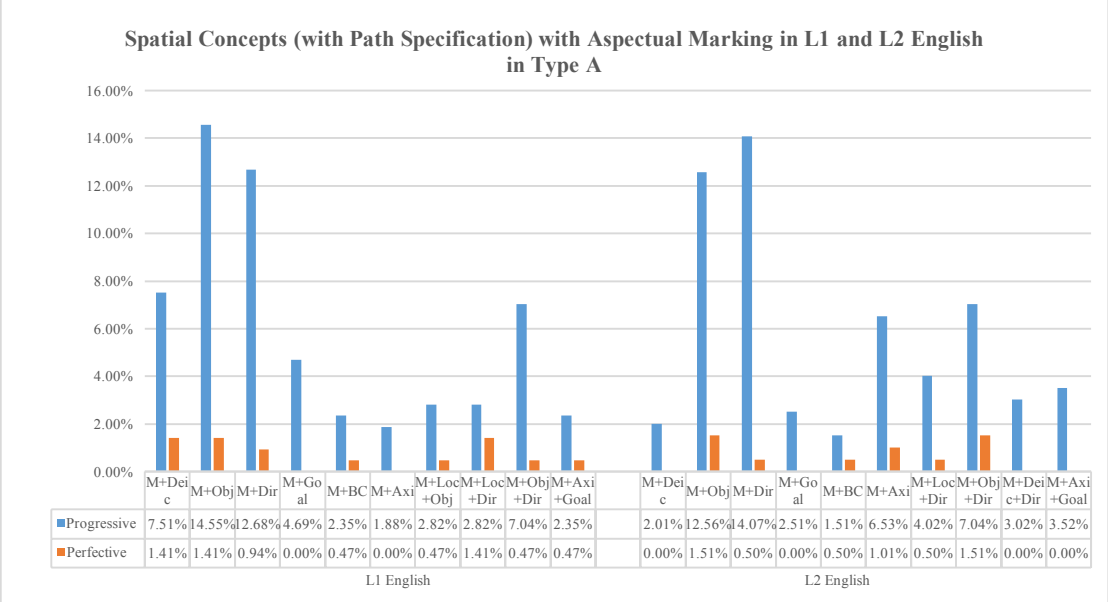


Figure 6.8: Aspectual marking for spatial concepts with Path specification in L1 and L2 English in the verbal task of Type A videos

Note that different combinations of spatial concepts occurred in L1 and L2

English (see Figure 6.8). Given the low numbers of occurrence of some combinations, these will not be discussed here. Concerning the construction "manner verb + deictic adjunct", L1 English speakers used the imperfective aspect (7.51%) and the perfective aspect (1.41%) to mark it, while L2 English speakers only used the imperfective aspect (2.01%) for it. Speakers of both groups used both the imperfective aspect (L1 English 14.55% vs. L2 English 12.56%) and the perfective aspect (L1 English 1.41% vs. L2 English 1.51%) to mark the construction "manner verb + object-related adjunct". The same was true for the construction "manner verb + directional adjunct": 12.68% were marked with the imperfective aspect in L1 English, while 14.07% were marked with the imperfective aspect in L2 English. In addition, both groups chose to use the imperfective aspect to mark the construction "manner verb + goal-oriented adjunct" (L1 English 4.69% vs. L2 English 2.51%), while for the construction "manner verb + boundary-crossing adjunct", L1 English speakers used the imperfective (2.35%) as well as the perfective (0.47%) aspect. L2 English speakers also used the imperfective (1.51%) and the perfective aspect (0.50%) to mark it. Regarding the construction "manner verb + axial adjunct", the perfective aspect (1.01%) was only found in L2 English, while both groups tended to mark it with the imperfective aspect (L1 English 1.88% vs. L2 English 6.53%).

In addition, Figure 6.8 also shows the co-occurrence of manner verbs with two or more path adjuncts, which generally occurred rarely in L1 and L2 English in the verbal task of Type A videos. Regarding the construction "manner verb + locative adjunct + object-related adjunct", L1 English speakers used the imperfective aspect (2.82%) to mark it. As for the constructions "manner verb + locative adjunct + directional adjunct" and "manner verb + object adjunct + directional adjunct", both groups used the imperfective aspect (for the former: L1 English 2.82% vs. L2 English 4.02%; for the latter: L1 English 7.04% vs. L2 English 7.04%) and the perfective aspect (for the former: L1 English 1.41% vs. L2 English 0.50%; for the latter: L1 English 0.47% vs. L2 English 1.51%). Concerning the construction "manner verb + axial adjunct + goal-oriented adjunct", both groups tended to use the imperfective aspect (L1 English 2.35% vs. L2 English 3.52%).

As in the verbal task of Type A videos, the imperfective aspect was also mostly used in both groups in the verbal task of Type B videos (see Figure 6.9). L1 and L2 English speakers tended to mark bare manner verbs with the imperfective aspect (L1 English 2.09% vs. L2 English 2.21%). Regarding the construction "manner verb +

locative adjunct", L1 and L2 English speakers tended to mark it with the imperfective aspect (L1 English 13.09% vs. L2 English 12.71%). In addition, L2 English speakers also marked it with the perfective aspect (0.55%). But these occurrences were rather low.

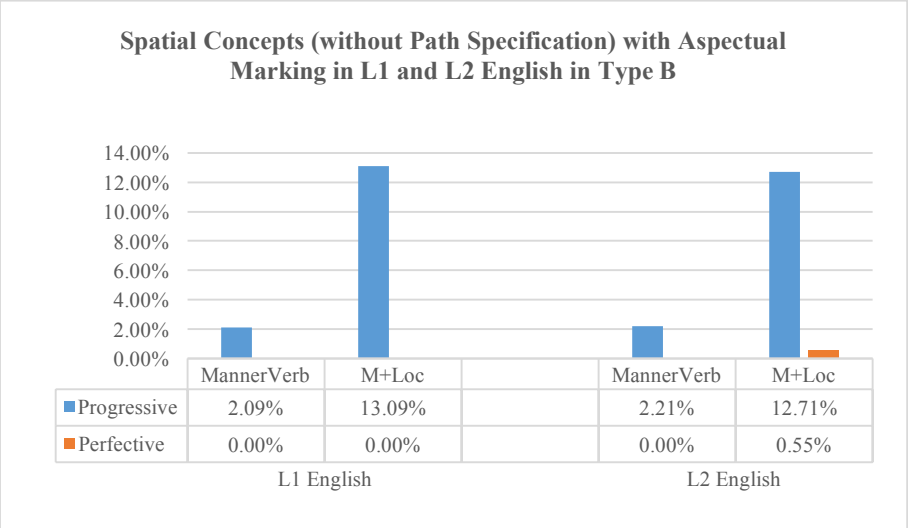


Figure 6.9: Aspectual marking for spatial concepts without Path specification in L1 and L2 English in the verbal task of Type B videos

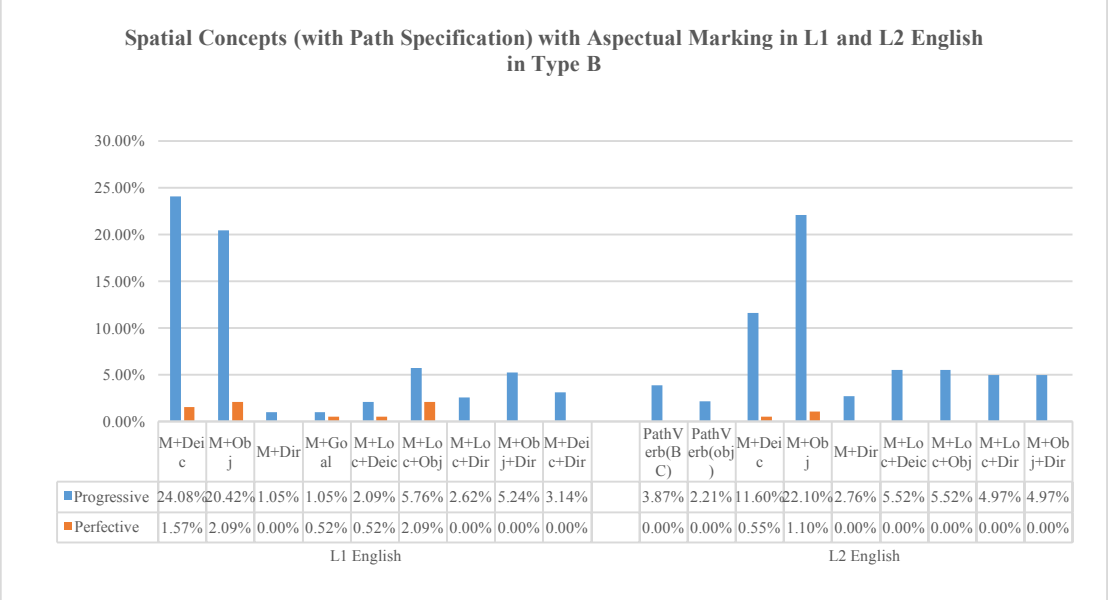


Figure 6.10: Aspectual marking for spatial concepts with Path specification in L1 and L2 English in the verbal task of Type B videos

Apart from the use of bare manner verbs, L2 English speakers also used the imperfective aspect to mark bare boundary-crossing path verbs (3.87%) and object-related path verbs (2.21%) (see Figure 6.10). Concerning the constructions "manner

verb + deictic adjunct" and "manner verb + object-related adjunct", speakers of both groups marked them with both the imperfective aspect (for the former: L1 English 24.08% vs. L2 English 11.60%; for the latter: L1 English 20.42% vs. L2 English 22.10%) and the perfective aspect (for the former: L1 English 1.57% vs. L2 English 0.55%; for the latter: L1 English 2.09% vs. L2 English 1.10%). By contrast, L1 and L2 English speakers marked the construction "manner verb + directional adjunct" only with the imperfective aspect (for the former: L1 English 1.05% vs. L2 English 2.76%; for the latter: L1 English 0.52% vs. L2 English 1.10%).

When considering the combination of manner verb with two path adjuncts in L1 and L2 English in the verbal task of Type B videos, it was found that almost all of the constructions listed in Figure 6.10 were marked with the imperfective aspect, except for the constructions "manner verb + locative adjunct + deictic adjunct" and "manner verb + locative adjunct + object adjunct" in L1 English, in which L1 English speakers used the imperfective aspect (for the former: L1 English 2.09%; for the latter: L1 English 5.76%) as well as the perfective aspect (for the former: L1 English 5.76%; for the latter: L1 English 2.09%), while L2 English speakers used the imperfective aspect to mark these two constructions (for the former: L2 English 5.52%; for the latter: L2 English 5.52%). In contrast, L1 and L2 English speakers used the imperfective aspect to mark the constructions "manner verb + locative adjunct + directional adjunct" (L1 English 2.62% vs. L2 English 4.97%), and "manner verb + object adjunct + directional adjunct" (L1 English 5.24% vs. L2 English 4.97%).

In the verbal task of Type C videos, the occurrence of bare manner verbs was low in both groups (see Figure 6.11). In L2 English, 5.26% were marked with the imperfective aspect. Type C videos showed boundary-crossing events. Therefore, the use of the construction "manner verb + boundary-crossing adjuncts" was found more often than other constructions (see Figure 6.12). However, Pearson chi-square tests showed that L1 and L2 did not differ in the use of aspectual marking for this construction ($\chi 2$ (1, N = 193) = 0.259, p = 0.611). In addition, L2 English also used bare boundary-crossing path verbs: 15.31% were marked with the imperfective aspect and 2.39% with the perfective aspect, while in L1 English, 5.50% were marked with the imperfective and 0.92% with the perfective aspect. As for the construction "manner verb + object adjunct", L1 English speakers used both the imperfective (4.13%) and the perfective aspect (2.29%) to mark it, respectively, while L2 English speakers only used the imperfective aspect to mark it (4.78%). Regarding the constructions "manner verb

+ deictic adjunct" and "manner verb + directional adjunct", speakers of both groups marked them with the imperfective aspect (for the former: L1 English 4.13%; for the latter: L1 English 3.21% vs. L2 English 1.44%).

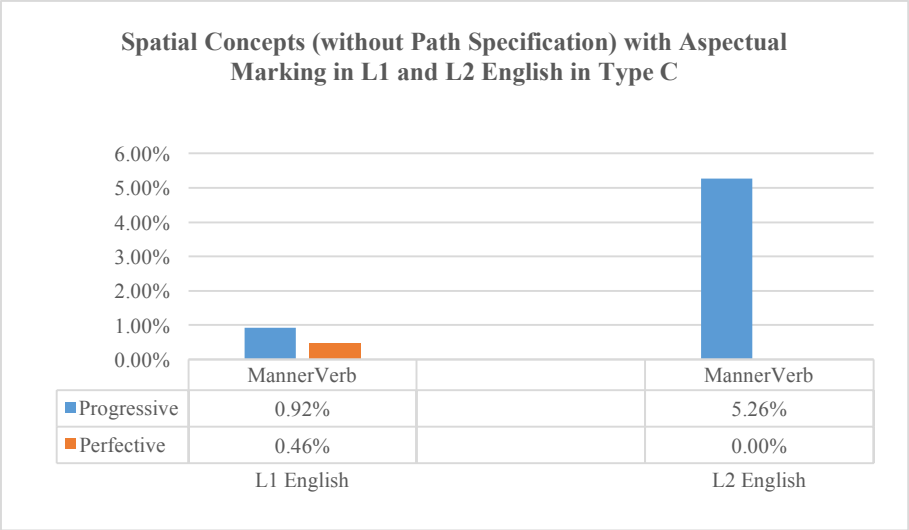


Figure 6.11: Aspectual marking for spatial concepts without Path specification in L1 and L2 English in the verbal task of Type C videos

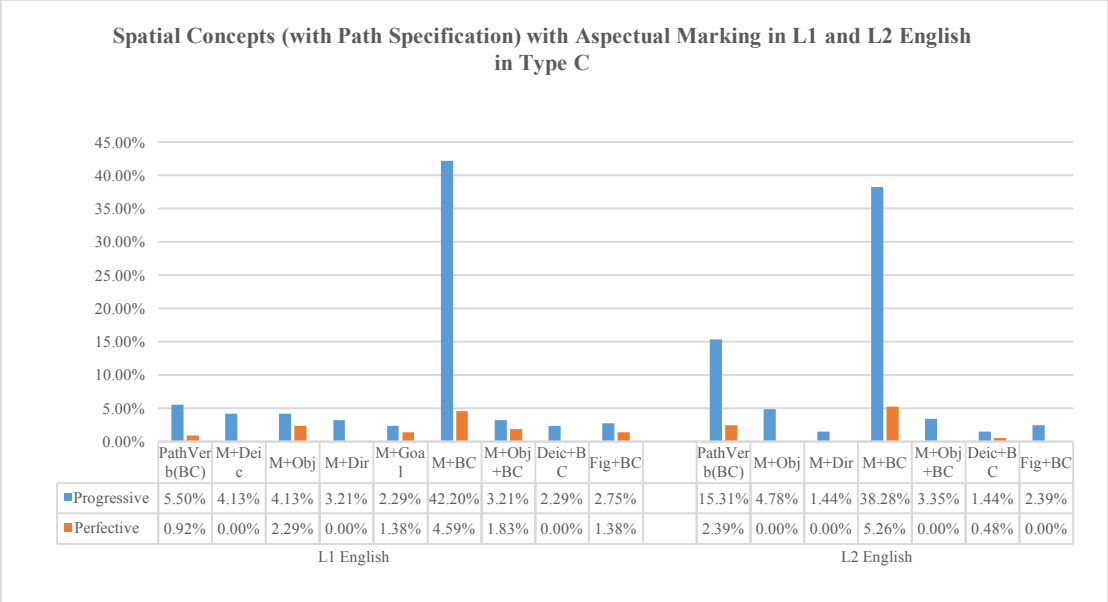


Figure 6.12: Aspectual marking for spatial concepts with Path specification in L1 and L2 English in the verbal task of Type C videos

As shown in Figure 6.12, regarding the construction "manner verb + object adjunct + boundary-crossing adjunct", in addition to its co-occurrence with the imperfective aspect in both language groups (L1 English: 3.21%; L2 English: 3.35%), L1 English speakers also used the perfective aspect (1.83%) to mark it. Concerning the

construction “deictic path verb + boundary crossing adjunct”, L1 English speakers marked it only with the imperfective aspect (2.29%), while L2 English speakers marked it both with the imperfective (1.44%) and the perfective aspect (0.48%). Regarding the construction “figure-related path verb + boundary-crossing adjunct”, L1 English speakers marked it with the imperfective (2.75%) and the perfective aspect (1.38%), while L2 English speakers only used the imperfective aspect (2.39%).

6.3 Summary

In this chapter, we examined the correlation between spatial concepts and aspectual marking in L1s and L2s. In general, we found in our study that L1 Chinese speakers tended to encode locative and directional concepts in the verbal task of Type A videos, locative concepts in the verbal task of Type B videos, and boundary-crossing concepts in the verbal task of Type C videos. Among these spatial concepts, L1 Chinese speakers tended to use the progressive aspect to mark locative and directional concepts and the perfective aspect to mark boundary-crossing concepts. L2 Chinese speakers also tended to encode locative and directional concepts in the verbal task of Type A videos: Locative concepts tended to be marked with the progressive aspect and directional concepts tended to have no aspectual markings, i.e., zero marking. In the verbal task of Type B videos, apart from marking the locative concepts with the progressive aspect, L2 Chinese speakers also encoded more directional, goal-oriented, and boundary-crossing concepts, all of which tended to have no aspectual markings. In addition, boundary-crossing concepts were also marked with the prospective aspect in L2 Chinese in the verbal task of Type B videos. The prospective aspect marks an event that will happen in the future, so the co-occurrence of boundary-crossing concepts and the prospective aspect is acceptable for describing the situations in Type B videos, although it is not the typical way the motion events in Type B videos were described by L1 Chinese speakers. Unlike the boundary-crossing concepts marked with the perfective aspect in L1 Chinese in the verbal task of Type C videos, these concepts tended to have no aspectual markings (zero marking) in L2 Chinese. In addition, boundary-crossing concepts also occurred with the progressive aspect (5.00%) and with the perfective aspect (2.27%) in the verbal task of Type C videos in L2 Chinese. However, the co-occurrence of the progressive aspect marker *zài* and boundary-crossing concepts results in unacceptable utterances in Chinese since boundary-

crossing concepts are realized as 2-state expressions in Chinese. The distinguished phase of 2-state expressions in Chinese is the target state, which is stative in nature. It is thus not compatible with the progressive aspect marker *zài*, which describes a motion as ongoing. In addition, it was found that L2 Chinese speakers tended to combine deictic concepts with other spatial concepts, especially with boundary-crossing concepts in the verbal task of Type C videos; these combinations tended to have no aspectual markings. That is, in the verbal task of Type C videos, L2 Chinese preferred using zero marking for the construction "manner verb + boundary-crossing path verb + deictic verb". By contrast, in the verbal task of Type C videos, L1 Chinese speakers preferred using the construction "manner verb + boundary-crossing path verb" and marked it with the perfective aspect marker *le*, which can be represented as "manner verb + boundary-crossing path verb + perfective aspect marker *le*". In comparison with L1 Chinese, the deictic verbs that immediately followed the boundary-crossing path verbs in the serial verb constructions in L2 Chinese took the position of the perfective aspect marker *le*. The question is why L1 German-L2 Chinese speakers did not use the perfective aspect marker *le* but rather used deictic verbs to describe boundary-crossing situationd. L1 German speakers typically used the construction "manner verb + boundary-crossing adjunct" to describe the boundary-crossing situations in Type C videos; for example, *eine Frau geht in das Haus* 'a woman walks into the house'. In this kind of expression, L1 German speakers did not formally differentiate between saying whether this woman is actually in the house or on her way to the house, as both of these situations are expressed as *in das Haus* 'into the house'. If German speakers want to focus on 'into the house', then verb particles like *hinein* 'thither-in' is used in addition. The corresponding description in Chinese is *yī-wèi nǚshì zǒu-jìn zhe-ge fángzi*. 'lit. one-CL woman walk-enter this-CL house'. In this Chinese description, the focus is on the transition from the outside to the inside of the house, leaving the direction to the house unspecified. In other words, the concept of the resultant state is important in Chinese for describing a boundary-crossing situation, while this is not the case in German. Rather, the directionality expressed by the German verb *gehen* 'go by walking' in this example plays a prominent role in construing the boundary-crossing situation in German. Triggered by the preference for encoding directionality in L1 German, L1 German-L2 Chinese speakers preferred using the deictic verb *qù* rather than the perfective aspect marker *le*. It seems that it was difficult for the L1 German-L2 Chinese speakers to learn the use of the perfective aspect marker *le*. As Li and Thompson (1981,

p. 216) pointed out, "learning to control -*le* is one of the most difficult tasks facing a European-language speakers attempting to master Mandarin, partly because European languages have no feature quite like it." This observation can be confirmed for the L1 German-L2 Chinese learners in our study. Typically, L1 German speakers used spatial expressions, as reflected in the construction "manner verb + boundary-crossing adjunct", to describe the boundary-crossing situations in Type C videos. L2 Chinese speakers tended to follow this pattern. They preferred using spatial means, i.e., deictic verbs, and obviously had not gotten used to employing the perfective aspect marker *le* to describe boundary-crossing events as L1 Chinese speakers did.

Furthermore, as already discussed in Chapter 5, *zài* in the preverbal *zài*-phrase has two functions: It encodes both locative and temporal meaning[94]. Locative concepts tended to be marked with the progressive aspect in L1 and L2 Chinese, especially in L1 Chinese in the verbal task of Type B videos. L1 Chinese speakers preferred describing the situations in Type B videos as "a Figure moving in a certain manner at a certain location". When doing this, no Path information is provided. (It should be pointed out that in comparison with L1 Chinese speakers, this type of description occurred less often in L2 Chinese.) The question is whether this way of description is sufficient for describing the motion events shown in Type B videos. In line with the findings in von Stutterheim et al. (2017), it was interesting for us to find that in our study, speakers of Chinese were quite similar to speakers of Tunisian Arabic in the way they described directed motion events. Due to the lack of directed path verbs, speakers of Tunisian Arabic tended to encode locative information and used a neutral motion verb with the progressive aspect to describe directed motion events. Von Stutterheim et al. (2017) argued that in Tunisian Arabic, the absence of path information in the descriptions of motion events "'is compensated for' by temporal aspect in the form of progression" (p. 245). The meaning of progression per se has a spatial origin and can be extended to the temporal category, which indicates "moving ahead on the one-dimensional time line" (p. 245). In this sense, motion events with the progressive aspect can be understood as "specific and directed towards some point of change on the way" (p. 245). The Chinese language data in our study provides further evidence for this interpretation. As mentioned above, L1 Chinese speakers tended to use manner verbs with the preverbal

[94] When a sentence has a perfective or habitual reading, i.e., if it has the perfective aspect marker *le* or a temporal adverbial like *chángcháng* 'often', then *zài* in the preverbal *zài*-phrase does not convey progressivity (see examples (11a) and (11b) in Chapter 5.1.3).

zài-phrase, which provides information on location and progressivity, to describe a directed motion, especially in the verbal task of Type B videos. Although there was only locative information without further path information about the trajectory or the endpoint (which contrasts with the typical description in L1 German; for more details, see Chapter 4), the progressive aspect enables us to understand that a motion is progressing along the time axis.

The analysis of the interaction between spatial concepts and aspectual marking in L1 and L2 English showed that the imperfective aspect was frequently used with various spatial concepts. There were rather low numbers of occurrence of the perfective aspect with different spatial concepts in L1 and L2 English. In comparison with Chinese, in which there are restrictions on the use of the imperfective aspect (incompatibility with 2-state expressions), the imperfective aspect can co-occur with all kinds of spatial categories in English. Hence, unlike L2 Chinese speakers, L2 English speakers did not make any mistakes in the use of aspectual marking. When we took a closer look at the use of the perfective aspect in L1 and L2 English, it was found that the perfective aspect was applied for the boundary-crossing situations in the verbal task of Type C videos. Regarding the construction "manner verb + boundary-crossing adjunct", 4.59% of the utterances in L1 English and 5.26% of those in L2 English were marked with the perfective aspect. That is, although they did so rarely, some of the L1 and L2 speakers viewed the boundary-crossing events from a holistic perspective.

Given the different aspectual systems in Chinese and English, the interaction between spatial concepts and aspectual marking differed between L1 German-L2 Chinese and L1 German-L2 English. In comparison with Chinese, the imperfective aspect is compatible with various spatial categories in English. L1 German-L2 English speakers preferred using the imperfective aspect to mark different spatial concepts. By contrast, the use of the imperfective aspect was much less frequent in L1 German-L2 Chinese than in L1 German-L2 English, as it cannot occur with all spatial categories (for example, 2-state expressions) in Chinese. In L1 German-L2 Chinese, the imperfective aspect tended to occur with locative concepts. As already explained in Chapter 5, the more frequent use of the imperfective aspect in L1 German-L2 English is related to the fact that it is inflexed in the verbs (*-ing* form), which is salient for L2 speakers to understand and use, while aspect markers in Chinese are used optionally and there are restrictions on their use. Some L1 German-L2 Chinese speakers made mistakes in the use of the progressive aspect, while L1 German-L2 English speakers

did not make any mistakes in this regard. In addition, it was found that the perfective aspect was applied to describe the boundary-crossing situations in Type C videos by both learner groups. However, its co-occurrence was rather low in both groups. As mentioned above, L1 German-L2 Chinese speakers used the perfective aspect significantly less frequently, while using more deictic verbs than L1 Chinese speakers. This reflects the preference of L1 German-L2 Chinese speakers to encode directionality.

Chapter 7 Discussion

In this study we compared the expression of motion events in L2s with respect to L1s and found interesting relevant results. The event framing in terms of the spatial domain shows clear traces of the L1. This finding is in line with the research results in Carroll et al. (2012), Hendriks et al. (2008, 2015), Schmiedtová et al. (2011), Schmiedtová and Sahonenko (2008), von Stutterheim (2003), and von Stutterheim and Carroll (2006). Our main results show that the less transparently conceptual categories are marked in a language, the more difficulty L2 speakers have in acquiring them. Whether or not to choose encoding endpoints is not determined by grammatical rules. L2 speakers have to detect and infer this from the input (L1 German speakers preferred encoding endpoints, while L1 Chinese and L1 English speakers tended not to do so). As for aspectual encoding, ongoingness is grammaticalized in English (via the *-ing* form), whereas it is realized in the form of temporal particles in Chinese. The use of temporal particles in Chinese is used optionally and there are restrictions regarding their co-occurrence with spatial concepts. Therefore, learning to use temporal particles poses challenges for L1 German-L2 Chinese speakers. According to our findings, unlike L1 German-L2 English speakers, L1 German-L2 Chinese speakers have not mastered the use of aspect markers in typical contexts in Chinese. This chapter provides a summary of our major findings in two respects: the language and eye tracking data.

Language data: encoding of spatial concepts

When comparing how the concept of motion is encoded in different languages, we found that speakers of L1 Chinese and L1 German-L2 Chinese preferred to encode Manner and Path information in verbs, while L1 German speakers typically encoded Manner information in manner verbs and Path information in adjuncts. This corresponds to the typological features of E-languages and S-languages, respectively. Chinese is a representative of E-languages in which Manner and Path information is encoded in verbs. German is a representative of S-languages in which Manner is encoded in verbs and Path in adjuncts. When taking a closer look at the use of verb types among speakers of the different groups, we found that L1 Chinese speakers used manner verbs in SVCs, while L1 German-L2 Chinese speakers preferred the use of manner verbs alone. As for the path verbs, both L1 and L2 Chinese speakers preferred to use them in SVCs. In addition, the use of bare path verbs was found more often in

L1 German-L2 Chinese than in L1 Chinese. The choice regarding the use of path verbs was found to be related to the characteristics of the different video types (A, B, and C). In Type A and Type B, L2 Chinese speakers used more deictic, goal-oriented, and boundary-crossing path verbs than L1 Chinese speakers. The more frequent use of deictic path verbs in L2 Chinese is related to the phonological and morphological status of modern Mandarin Chinese, which tends to be disyllable or bipartite, depending on the directional meaning it indicates, and the holistic perspectives L2 speakers tended to take. The more frequent use of goal-oriented and boundary-crossing path verbs in Type A and Type B reflects their L1s' conceptual patterns: Previous empirical studies have shown that L1 German speakers typically took a holistic perspective and focused on the endpoints in describing motion events (e.g., Carroll & von Stutterheim, 2003; von Stutterheim, 2003; von Stutterheim & Carroll, 2006; von Stutterheim et al., 2012). When goal-oriented and boundary-crossing path verbs are used by L2 Chinese speakers, especially in Type B which shows a figure moving along a long trajectory, possibly towards an endpoint, then the endpoint has to be encoded with those concepts. The use of boundary-crossing path verbs in Type B reflects L2 Chinese speakers' focus on the endpoints. This is because in a boundary-crossing situation the endpoint has a high profile, so it has to be encoded. This contrasts sharply with L1 Chinese speakers who typically conceptualized the situation in Type B videos as "Figure moving in a certain manner in a certain location". That is, L1 Chinese speakers typically did not encode directional or goal-related information in Type B, but rather locative information. We interpret this as meaning that L1 German speakers' habitual conceptual patterns play a role when speakers of L1 German-L2 Chinese describe the motion events. In Type C videos, L1 and L2 Chinese speakers did not differ in the use of boundary-crossing path verbs. This is because Type C showed a boundary-crossing situation, so all speakers tended to express it explicitly. Manner verbs and path verbs are combined to form SVCs in Chinese. In general, L1 Chinese speakers generally selected the combination motion verbs with manner verbs, or manner verbs with path verbs, while L2 Chinese speakers preferred to combine path verbs with deictic verbs, or manner verbs with path verbs and deictic verbs. That is, L2 Chinese speakers preferred to combine deictic verbs with other path verbs to form the construction "path verb + deictic verb" or "manner verb + path verb + deictic verb". Current data showed that L1 German speakers typically used the construction "manner verb + path adjunct" to describe motion scenes. As mentioned above, deictic verbs in Chinese are directional in nature. The more frequent use of

deictic verbs in the SVCs in L2 Chinese might be triggered by the directional motion expressed in the German verb *gehen* ('go by walking') or the directionality encoded in adjuncts in L1 German. Concerning the use of adjuncts among speakers of the three groups in our study, L1 German speakers used significantly more adjuncts than L1 and L2 Chinese speakers. The use of specific adjuncts was found to be related to the features of the different video types. In Type A, speakers of L1 German and L2 Chinese used more goal-oriented adjuncts than L1 Chinese speakers. In addition, L1 German speakers used more boundary-crossing adjuncts than L1 and L2 Chinese speakers. This corresponds to the more frequent use of boundary-crossing path verbs in L2 Chinese in Type A. In Type B, L1 Chinese speakers preferred the use of locative adjuncts, while L2 Chinese speakers encoded less locative adjuncts and used more goal-oriented and boundary-crossing adjuncts. In Type C, L1 German speakers typically encoded boundary-crossing concepts in adjuncts. In contrast, speakers of Chinese generally encoded boundary-crossing concepts in verbs. For the same concept, different language speakers have different means to encode it.

Like German, English is also an S-language. In our study, speakers of L1 and L2 English typically encoded Manner information in manner verbs and Path information in adjuncts, which conforms with the typological features of S-languages. Speakers of L1, L2 English, and L1 German did not differ in the use of manner verbs. Despite the low occurrence of path verbs, L2 English speakers used more boundary-crossing path verbs than speakers of L1 English and L1 German in Type B and Type C videos, respectively. The more frequent use of boundary-crossing path verbs in L2 English in Type B may reflect their focus on the endpoints, since in a boundary-crossing situation the endpoint has a high profile and cannot be neglected. This shows the traces of their L1 conceptual patterns (i.e., focus on the endpoints) in L2 English speakers. Generally speaking, L1 and L2 English speakers did not differ in the use of various adjuncts. However, in comparison with L1 German speakers, L2 English speakers used less goal-oriented adjuncts in Type A, but more object-related adjuncts in Type B, while L1 and L2 English speakers did not differ in this regard. The less frequent use of object-related adjuncts in L1 German was related to the fact that L1 German speakers also encoded this concept in verb particles (for more details, see Chapter 4).

When comparing the use of verb types and adjuncts between L1 German-L2 Chinese and L1 German-L2 English speakers, we found that L1 German-L2 English speakers preferred to use more manner verbs than L1 German-L2 Chinese speakers. In

contrast, L1 German-L2 Chinese speakers used more path verbs than L1 German-L2 English speakers. Speakers of both groups did not differ with regard to the use of motion verbs and other verbs. Furthermore, L1 German-L2 English speakers used more adjuncts than L1 German-L2 Chinese speakers. Taking a closer look at the spatial concepts encoded (both in verbs and adjuncts), we found that speakers of the two learner groups did not differ in the use of locative adjuncts. However, L1 German-L2 Chinese speakers preferred to encode deictic and goal-oriented concepts, which were mostly encoded in path verbs. In contrast, L1 German-L2 English speakers used object-related and directional concepts, which were more likely to be encoded in adjuncts. Both L2 groups did not differ in the total encoding of boundary-crossing concepts, which tended to be encoded in path verbs by L2 Chinese speakers and in adjuncts by L2 English speakers.

Furthermore, typological contrasts between S-languages and E-languages play a role in the formation of event units, i.e. whether one or two utterances are produced when describing motion events. What was found as a common feature for speakers of different groups is that speakers of L1s and L2s all tended to produce one utterance in describing motion events. In comparison with the occurrence of one utterance, the occurrence of two utterances was found to be lower in all language groups. However, L1 and L2 Chinese speakers differed in the use of two utterances: L2 Chinese speakers produced two utterances more often than L1 Chinese speakers, while L1 and L2 English speakers did not differ in this regard. Taking a closer look at the occurrence of one utterance, we found that German and Chinese differed in their capacity to combine different spatial concepts in one utterance, while German and English did not differ in this regard. The difference in the capacity to combine different spatial concepts between German and Chinese lies in the verb type chosen as predicate which determines the event type. In German, the event type is characteristically framed by manner verbs in which various path segments can be stacked after manner verbs (e.g., Hickmann & Hendriks, 2010; Slobin, 1996). In Chinese, the event type is typically framed by manner verbs or a combination of manner verbs and path verbs. When the event type in Chinese is determined by a combination of manner verbs and path verbs, due to the syntactic and semantic features given by the path verbs, there are restrictions on the combination of different path segments in one utterance. This form of event unit formation posed challenges for L1 German-L2 Chinese speakers who either produced unacceptable utterances with different spatial concepts in one utterance (although there were few

occurrences of this) or solved this problem by producing two utterances. That is to say, in order to accommodate the syntactic and semantic restrictions imposed by the path verbs, L1 German-L2 Chinese speakers chose to produce two utterances. Our findings have shown that two utterances are more likely to be produced when the figure's orientation changes in a situation, which is in line with the findings of Gerwien and von Stutterheim (2018). When producing two utterances, L1 German-L2 Chinese speakers preferred to encode the path of the motion with directional information in the first utterance and mentioned the endpoint in the second utterance. In contrast, L1 Chinese speakers preferred to use manner verbs, with or without locative adjuncts in the first utterance and expressed directional information, or selected a second utterance to encode the endpoint. This procedure by L1 Chinese speakers in producing two utterances is on par with the findings of Carroll et al. (2012) and von Stutterheim et al. (2017). It showed that when speakers of V-languages (e.g., French, Arabic) produced two utterances in describing motion events, they preferred to express the manner of the motion at a certain location in the first utterance and used path verbs to encode the endpoint in the second utterance. In contrast, in the present study, L1 German-L2 Chinese speakers preferred to express directional or dynamic information rather than locative information in the first utterance and then mentioned the endpoint in the second utterance. In comparison with L1 German-L2 English speakers, L1 German-L2 Chinese speakers produced more responses with two utterances in Type A and B videos, while the L2 learner groups did not differ in the production of two utterances in Type C videos. The difference between the two learner groups resides in the lexicalization differences and the availability of morphosyntactic constructions in each language (Bohnemeyer et al., 2007). As in German, English allows different path segments, i.e., path adjuncts, to be stacked after manner verbs, while syntactic and semantic restrictions are imposed by the path verbs in combinations of different spatial concepts in Chinese. When L1 German-L2 Chinese speakers wanted to express the same amount of information as L1 German-L2 English speakers they had to divide the event into sub-events and expressed these in separate utterances.

Language data: encoding of aspectual concepts

In our study, we extended the analysis of lexicalized concepts to grammaticalized concepts, i.e., temporal-aspectual concepts, in the domain of motion events (see, e.g., Bylund & Jarvis, 2011; Schmiedtová & Sahonenko, 2008; von Stutterheim, 2003; von Stutterheim & Carroll, 2006; von Stutterheim et al., 2012, 2017).

This study has shown that progressive aspect in English can occur with different verb types, while this was not the case in Chinese, where the features of different types of video were found to have an impact upon the use of a specific temporal category. To be more specific, in Type A videos, L1 Chinese speakers preferred the use of the progressive aspect, while L1 German-L2 Chinese speakers preferred the use of zero marking. The same holds true for Type B videos. The use of the progressive aspect was found more frequently in L1 Chinese speakers, while the use of zero marking was found more frequently in L2 Chinese speakers. In addition, L2 Chinese speakers also used the prospective aspect, whereas no data was found for L1 Chinese speakers regarding the use of prospective aspect. The use of prospective aspect in L2 Chinese speakers makes it possible for the learners to mention the endpoints in Type B. This can be attributed to the fact that although in Type B videos the actual phase of reaching an endpoint is not shown, prospective aspect, which indicates a situation that will happen in the future, is nevertheless suitable in verbalizing the endpoints in Type B. In Type C videos, L1 Chinese speakers used perfective aspect more often to indicate the completion of boundary-crossing motion events and used durative aspect more often to specify the accompanying situation compared to L1 German-L2 Chinese speakers, respectively. In contrast to L1 Chinese speakers, L1 German-L2 Chinese speakers tended to use zero marking and progressive aspect instead of perfective aspect to mark a boundary-crossing situation. Progressive aspect might be used inappropriately by L2 Chinese speakers in Type C: When L2 speakers combined progressive aspect with a postverbal *zài*-phrase or two-state expressions, they used it inappropriately. This is because the postverbal *zài*-phrase adds a locative boundary to the situation and the distinguished phase of two-state expressions is stative in nature; both are not compatible with the progressive aspect marker *zài*[95]. In comparison to L1 German-L2 Chinese speakers, L1 German-L2 English speakers were found to use imperfective aspect more frequently across different video types. This is most likely attributable to the fact that imperfective aspect in English can be combined with any type of verb. It is inflexed on verbs and thus salient and transparent for L2 English learners, while progressive aspect marker in Chinese is a lexical item and used optionally. The possibility of using aspect markers optionally in Chinese makes their usage both flexible and constraining: Speakers of

[95] When the aspect marker *zhèng* with its meaning 'at the moment' occurs with a postverbal *zài*-phrase or two-state expressions, the sentence is acceptable. However, there were only two utterances in total with *zhèng* by L2 Chinese speakers in this study: One occurred in Type B and the other in Type C.

Chinese can choose whether to use aspect markers or not. However, this flexibility is confined by the condition that the progressive aspect marker *zài* cannot occur with two-state expressions.

Language data: interaction between spatial and aspectual concepts

There is a dependency between the use of spatial and aspectual concepts and the type of motion events in different languages. In the verbal task of Type A videos in our study, L1 Chinese speakers typically used the locative and directional concepts with the progressive aspect, while L1 German-L2 Chinese speakers preferred marking locative concepts with the progressive aspect and directional concepts with zero marking. In the verbal task of Type B videos, the majority of L1 Chinese speakers preferred expressing locative concepts by marking them with the progressive aspect, while L1 German-L2 Chinese speakers did so to a lesser degree and additionally expressed more directional, goal-oriented, and boundary-crossing concepts using zero marking. Apart from using zero marking for boundary-crossing concepts, L1 German-L2 Chinese speakers also marked them with the prospective aspect in the verbal task of Type B videos. Since the prospective aspect indicates a situation that will happen in the future, the co-occurrence of boundary-crossing concepts with the prospective aspect in the verbal task of Type B videos was appropriate in L2 Chinese. In the verbal task of Type C videos, L1 Chinese speakers preferred marking boundary-crossing concepts with the perfective aspect, while these concepts tended not to have aspectual markings in L2 Chinese. In addition, boundary-crossing concepts were also marked with the progressive aspect in L2 Chinese, which led to unacceptable expressions. Furthermore, L1 German-L2 Chinese speakers preferred combining boundary-crossing concepts with deictic concepts and using zero marking for this combination in the verbal task of Type C videos. On par with the findings in von Stutterheim et al. (2017), about 60% of L1 Chinese speakers and 30% of L1 German-L2 Chinese speakers used the construction "locative adjunct + manner verb" and marked it with the progressive aspect in the verbal task of Type B videos. Motion events with progression can be understood as "specific and directed towards some point of change on the way" (von Stutterheim et al., 2017, p. 245). Therefore, although L1 and L2 Chinese speakers did not specify the path of motion, in their use of the construction "locative adjunct + manner verb" the progressive aspect lets us understand that the motion is moving ahead along the timeline. In this sense, the integration of an aspectual category into the description of a motion

event can furnish the spatial concepts that are encoded in verbs or other linguistic means in some languages.

L1 and L2 English speakers preferred using the imperfective aspect to mark various spatial concepts. This is because unlike Chinese, the imperfective aspect in English can combine with different spatial categories. In addition, L1 and L2 English speakers also used the perfective aspect to mark spatial concepts, especially boundary-crossing concepts in the verbal task of Type C videos. However, the occurrence of the perfective aspect was rather low in L1 and L2 English. It seems that L1 and L2 English speakers also took a holistic perspective to describe the boundary-crossing situations in Type C videos. It should be pointed out that the habitual event construal patterns in different languages only demonstrate different preference patterns (Bylund & Jarvis, 2011; von Stutterheim & Nüse, 2003). They are not absolute principles, unlike grammar rules, which speakers must obey. Therefore, although speakers of English preferred segmenting a motion scene into different phases, this did not prevent them from taking a holistic perspective on viewing the situation, as L1 German speakers typically did. So it is not surprising that some L1 and L2 English speakers also used the perfective aspect, i.e., took a holistic perspective when viewing boundary-crossing situations.

Eye-tracking data: speech onset times

Speech onset times (SOTs) imply the time it takes speakers to process relevant visual and linguistic information. When comparing the SOTs between L1s and L2s, it was found that L2 speakers had longer SOTs than L1 speakers. We cannot exclude the L2 effect for the longer SOTs in L2s. However, the L2 speakers had been selected carefully according to their reported language proficiency and time spent learning foreign languages, so they were qualified as advanced learners (see Chapter 3). So in addition to the L2 effect, there should be other reasons for the longer SOTs in both L2 groups. Further analysis of the SOTs in the different video types showed that the order of the SOTs from longest to shortest in L1 German-L2 Chinese was Type A > Type B > Type C, while the order of the SOTs in L1 German-L2 English was Type A > Type C > Type B. That is to say, the two L2 groups differed in the SOTs for Type B videos. L1 German-L2 Chinese speakers had longer SOTs than L1 German-L2 English speakers for Type B videos. According to the spatial analysis, we observed that in the verbal task of Type B videos, L1 German-L2 Chinese speakers encoded deictic, goal-oriented, and boundary-crossing concepts more often than L1 Chinese speakers, while L1 German-L2 English speakers encoded boundary-crossing concepts more often than L1 English

speakers. (It should be pointed out that the occurrence of boundary-crossing concepts was low in L1 and L2 English). To encode these concepts, L2 speakers had to wait until the endpoint became recognizable and identifiable as the motion scene unfolded, and then began to speak. So in comparison with L1 speakers, the two learner groups had longer SOTs. This also reflects L1 German's habitual conceptual patterns in both learner groups. In addition, L1 German-L2 Chinese speakers had longer SOTs than L1 German-L2 English speakers in the different video types. This might be due to language-specific features in Chinese, i.e., the syntactic complexity of serial verb constructions (see Chapter 4), the compatibility of different path segments in one utterance (see Chapter 4), the choice of using aspect markers or not, and the compatibility of aspect markers with spatial concepts (see Chapters 5 and 6). All these language-specific features required L1 German-L2 Chinese speakers to think longer about how to select and organize the information. So in comparison with L1 German-L2 English speakers, they had longer SOTs.

Eye-tracking data: endpoint mentioning and fixation patterns

As for the mentioning of the endpoint in their utterances, L1 and L2 Chinese speakers exhibited a remarkable difference in the verbal task of Type B videos. Speakers of L1 German and L1 German-L2 Chinese mentioned the endpoint significantly more often than L1 Chinese speakers, while L1 German speakers and L1 German-L2 Chinese speakers did not differ in this regard. This was in line with the eye-tracking data, which showed that L1 German speakers and L1 German-L2 Chinese speakers had more fixations on the endpoint both in total and before SOTs than L1 Chinese speakers in Type B videos, while L1 German speakers and L1 German-L2 Chinese speakers did not differ in this regard. As for the duration of the fixations on the endpoint, L1 German speakers fixated on the endpoint longer, both in total and before SOTs, than L1 Chinese speakers, while L1 German-L2 Chinese speakers did not differ from L1 German speakers in these regards. That is to say, like L1 German speakers, L1 German-L2 Chinese speakers preferred paying attention to the endpoint during the conceptualization phase in the verbal task of Type B videos, which provided evidence for our explanation for the late SOTs in L1 German-L2 Chinese in Type B videos and is in accordance with the results for the mentioning of the endpoint among speakers of the three language groups.

L1 and L2 English speakers did not differ in the mentioning of the endpoint in the different video types. However, the eye-tracking data showed interesting results for

Type B videos: Regarding the total number of fixations on the endpoint, L1 English speakers fixated more often on the endpoint than L2 English speakers did, while L2 English speakers did not differ from L1 German speakers in this regard; neither did L1 English speakers differ from L1 German speakers. In contrast, regarding the fixations on the endpoint before SOTs, L1 German speakers fixated more often on the endpoint than L1 English speakers, while L2 English speakers did not differ from L1 German speakers, nor did they differ from L1 English speakers in this regard. In other words, if we used a scale and placed these three groups of speakers on it in terms of their attention to the endpoint before SOTs, L1 German speakers, who preferred paying attention to the endpoint during the conceptualization phase, could be placed on one end of this scale, and L1 English speakers, who paid less attention to the endpoint, could be put on the other end of this scale. L2 English speakers, who differed neither from L1 German speakers nor from L1 English speakers, could be placed in the middle of this scale. It appears that L2 English speakers kept their L1 German speakers' habitual conceptual patterns on the one hand and tried to adjust them in order to follow L1 English speakers' conceptual patterns on the other hand.

Furthermore, the two learner groups did not differ in the mentioning of the endpoint in the different video types. The eye-tracking data further showed that these two learner groups did not differ in the fixations on the endpoint in total and before SOTs in Type B and C videos. What is interesting is that in Type A videos, L1 German-L2 Chinese speakers fixated more often and longer on both the endpoint and the moving entity than L1 German-L2 English speakers did before SOTs. In other words, during the conceptualization phase in the verbal task of Type A videos, L1 German-L2 Chinese speakers preferred directing their attention to both the endpoint and the moving entity. They may have needed to think about which information to select and how to structure this information in Chinese in relation to the endpoint and to the moving entity. Both learner groups could conceptualize the situation in Type A videos as "A Figure moving in a certain manner", "A Figure moving in a certain manner at a certain location", or "A Figure moving in a certain manner towards or to a place at goal". In English, all these different conceptual components can co-occur in one utterance, like "A Figure moving in a certain manner towards or to a place at a certain location". L2 English speakers can add different components to a sentence they already uttered. This is not the case in Chinese, which, by contrast, has restrictions on the combination of path

segments. Therefore, L2 Chinese speakers have to determine which information to select for verbalization during the conceptualization phase.

Eye-tracking data: early allocation of attention

As an additional window on conceptual processing, the allocation of visual attention over the course of time unveils when differences in the fixation patterns illustrated above occur. In this study, we further examined the attention allocated to the endpoint and to the moving entity over the course of time in L1s and L2s. When considering under which aspectual perspectives the endpoint tended to be mentioned in the verbal task in L1 and L2 Chinese, it was found that L2 Chinese speakers tended to mention the endpoint when they used zero marking and the imperfective aspect. The results showed that no matter which aspect L1 and L2 Chinese speakers used, the fixations of the moving entity rapidly increased in both groups between 0 ms and 600 ms and reached their peak around 600 ms. The reason for this might lie in the fact that as the video clip began to play, the speakers may have first needed to scan the whole situation in order to get the gist of it and an overall understanding of it, since they did not know which parts of the background information would be relevant for answering the question (what is happening?). Then they directed their attention to the moving object along the trajectory in order to see what would happen. As the scene unfolded over time, the speakers' conceptual representation evolved. They may have needed to direct their attention to the object at the endpoint in order to retrieve relevant information for naming it. When they used the imperfective aspect, L1 Chinese speakers tended to pay attention to the moving entity; this difference occurred between 0 ms and 1200 ms in the different video types. On the other hand, L2 Chinese speakers tended to pay attention to the endpoint; this difference occurred as early as between 600 ms and 1200 ms in Type A videos and between 1800 ms and 2400 ms in Type B videos. L1 Chinese speakers' behavior confirmed the hypothesis of *seeing for speaking*, which says that speakers of aspect languages, who use the imperfective/progressive aspect frequently, tend to segment the situation into different phases, while speakers of non-aspect languages, who do not use grammaticalized aspects frequently, tend to take a holistic perspective and encode the endpoint. Although L1 German-L2 Chinese speakers used the imperfective aspect, they also paid more attention to the endpoint. This fact does not contradict the hypothesis of *seeing for speaking*; rather, it reflects that their L1's conceptual patterns had an effect in their L2. Furthermore, when zero marking was used, in contrast to L1 Chinese speakers, L2 Chinese speakers tended to

direct their attention to both the moving entity and the endpoint as time went on. This was because L1 Chinese speakers and L1 German speakers differed in their habitual conceptual patterns for event construal: L1 Chinese speakers typically referred to the location of the Figure, which was atypical in German. In contrast, L1 German speakers tended to describe the situation related to the endpoint or the trajectory traversed by the Figure in motion. L1 Chinese speakers' conceptual patterns typically required speakers to focus on the Figure, whereas L1 German speakers' conceptual patterns required speakers to additionally pay attention to the endpoint. This might have been the reason why L1 German-L2 Chinese speakers paid more attention to both the moving entity and to the endpoint in the different video types.

In addition, when comparing under which aspectual perspectives the endpoint tended to be mentioned in the verbal task in L1 and L2 English, it was found that both L1 and L2 speakers tended to mention the endpoint when they used the imperfective aspect. Speakers of English tended to decompose the situation into different phases and may not pay attention to the endpoint in the conceptualization phase; after they started to talk, they can add the endpoint to an already uttered sentence (see von Stutterheim, 2003; von Stutterheim & Carroll, 2006). So it was no surprise that L1 and L2 English speakers were able to use the imperfective aspect and verbalized the endpoint. Therefore, it was more significant to examine the attention allocated to the moving entity and to the endpoint before SOTs in L1 and L2 English, i.e., during the conceptualization phase, rather than when the imperfective or perfective aspect was used. In Type A videos, L1 German speakers fixated on the endpoint more often than L1 and L2 English speakers as early as between 600 ms and 1800 ms, whereas L1 and L2 English speakers did not differ during this time span. In Type B videos, both speakers of L1 German and L1 German-L2 English had more endpoint fixations than L1 English speakers between 1800 ms and 3000 ms, while no difference in the fixations on the endpoint was found in L1 German and L1 German-L2 English during this time span. Note that for Type B videos, the average SOT in L1 English was 1893 ms, while the SOT in L1 German-L2 English was 2577 ms and in L1 German 2596 ms. It appeared that during the conceptualization phase, speakers of L1 German-L2 English and L1 German tended to fixate longer on the endpoint in Type B videos than L1 English speakers. In Type C videos, L1 German speakers fixated on the endpoint more often than L1 and L2 English speakers as early as during the first 600 ms. Furthermore, between 600 ms and 1200 ms, both speakers of L1 German and L1 German-L2 English

had more fixations on the endpoint than L1 English speakers, while no difference in the fixations on the endpoint was found between speakers of L1 German and L1 German-L2 English during this time span. Overall, L1 German-L2 English speakers differed more or less from L1 English speakers in the early allocation of attention to the endpoint. It therefore appears that the endpoint plays an important role in the early attention patterns in motion events construal in L1 German-L2 English.

To summarize, we have shown that although both L1 German-L2 Chinese and L1 German-L2 English speakers had generally learned to use the linguistic forms of the L2 and their appropriate functions, they are apparently unable, to some degree, to map these forms onto the principles required for event construal in their L2s (Chinese and English, respectively). In comparison with L1 German-L2 Chinese speakers, L1 German-L2 English speakers went further towards the target language. It seems that the typological distance between source and target language plays a role in motion events construal in adult second language acquisition. In addition, we found that it was difficult even for advanced L2 learners to adjust their L1 language-specific *thinking and seeing for speaking* to their L2.

The question is then which mechanism drives the impact of the verb structure and the grammatical aspect upon motion events cognition. According to Slobin (1996, 2004, 2006), von Stutterheim (2003), von Stutterheim and Carroll (2006), and von Stutterheim et al. (2012), when a specific linguistic construction is used frequently, it will become entrenched and salient in the speakers' minds. Following Langacker (2008), the frequency of a given schema being activated determines the degree to which this schema becomes the cognitive routine of the speakers' minds. Casasanto (2008) also argued that the habitual use of language structures (i.e., lexical or grammatical means) will activate the relevant conceptual categories. The process of using lexical or grammatical means reinforces the association between language structures and the relevant conceptual categories. In the case of grammatical aspects, different temporal viewing frames (i.e., taking a holistic view) are schematized in the minds of speakers of different languages to different degrees. Frequent use of certain lexical or grammatical aspects would make such aspectual features salient in speakers' minds. Therefore, speakers of aspect languages, which have grammaticalized aspects, such as English, prefer segmenting the situation into different phases. In contrast, speakers of non-aspect languages, which do not have grammatical aspects, such as German, prefer taking a holistic view and focusing on the endpoint. It is possible for the speakers of

non-aspect languages to turn to lexical or peripheral constructions to describe ongoingness. Therefore, the differences between speakers of aspect and non-aspect languages lie in "the degree of entrenchment of particular time schemas" (Bylund et al., 2013, p. 945). In our study, the difficulty that the two learner groups have to confront is that they must learn the L2's principles of information organization, i.e., perspective taking. That is, L1 German-L2 Chinese and L1 German-L2 English speakers have to switch from their habitual L1 perspective, i.e., a holistic endpoint-oriented perspective (e.g., von Stutterheim, 2003; von Stutterheim & Carroll, 2006), to the perspective of phasal decomposition in Chinese and English, respectively. However, those underlying principles might be subtle and difficult to change, so even for advanced L2 learners, their L1 habitual and highly automated routines might have an effect when they speak an L2.

This study has various implications for second language teaching. First, Stam (2015) and Flecken et al. (2015) argued that if teachers could explicitly give instructions or pay attention to the semantic components that make up language-specific spatial concepts, then they could help learners to master them. However, even if foreign language teachers may be aware of this insufficiency, for those L2 language productions that result from the conceptual transfer from learners' L1 the teachers may fail to point out the problems that lie in the language-specific conceptualization patterns from a typological perspective. Therefore, foreign language teachers should get training in this respect (see Flecken et al., 2015). Second, it is important to teach the learners in a native-like discourse environment and pay less attention to practicing grammatical knowledge at the sentence level. The strategy of providing a native-like discourse environment is adopted in first language acquisition, but it is not fully used in the SLA classroom. Our study has shown that although the advanced L1 German-L2 English learners were able to produce utterances without grammatical mistakes, they nevertheless showed traces of L1 conceptual patterns. They need to be exposed more to a native-like environment in order to fit the L2 *thinking and seeing for speaking*. Such exposure to a native-like environment could be achieved by using multimodal pedagogical tools, such as audiovisual media. The research results in Bylund and Athanasopoulos (2015b) showed that L1 Swedish-L2 English speakers' conceptual patterns could be restructured through exposure to English audiovisual media. In addition, given that conceptual categories are represented multimodally (e.g., in images, schemas, or visual impressions; see Bylund & Athanasopoulos, 2015b), foreign

language teachers should not limit their teaching to textbooks, i.e., to the description of grammar rules. They should design activities to get the learners to process the new concepts in different modalities. For example, teachers could first present some semantic features and then ask L2 learners to distinguish the manner of motion according to these semantic features (see Jarvis & Pavlenko, 2008); or they could organize attention exercises for L2 learners in the foreign language classroom (see Pavlenko & Volynsky, 2015) in order to train learners in paying attention to those components in motion events that are typically encoded in L2.

References

Anderson, J. (1973). *An essay concerning aspect*. Berlin: De Gruyter Mouton.

Athanasopoulos, P. (2011). Cognitive restructuring in bilingualism. In A. Pavlenko (Ed.), *Thinking and speaking in two languages* (pp. 29-65). Bristol: Multilingual Matters.

Athanasopoulos, P., & Bylund, E. (2013). Does grammatical aspect affect motion event cognition? A cross-linguistic comparison of English and Swedish speakers. *Cognitive Science, 37*, 286–309.

Becker, A. (1994). *Lokalisierungsausdrücke im Sprachvergleich. Eine lexikalisch-semantische Analyse von Lokalisierungsausdrücke im Deutschen, Englischen, Französischen and Türkischen*. Tübingen: Niemeyer.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum.

Beyer, Y.-H. (2016). *Raum-zeitliche Konzeptualisierung bei der Darstellung von Ereignissen. Ein Vergleich zwischen chinesischen, englischen und deutschen Sprechern.* (Doctoral thesis), University of Heidelberg.

Bohnemeyer, J., Enfield, N. J., Essegbey, J., Ibarretxe-Antuñano, I., Kita, S., Lüpke, F., & Ameka, F. K. (2007). Principles of event segmentation in language: The case of motion events. *Language*, 495-532.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43*(1), 1-22. doi:https://doi.org/10.1006/cogp.2001.0748

Brown, A., & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in second language acquisition, 30*(2), 225-251.

Bylund, E. (2009). Effects of age of L2 acquisition on L1 event conceptualization patterns. *Bilingualism: Language and Cognition, 12*, 305–322.

Bylund, E., & Athanasopoulos, P. (2014). Linguistic relativity in SLA: Toward a new research program. *Language learning, 64*(4), 952-985.

Bylund, E., & Athanasopoulos, P. (2015a). Introduction: Cognition, motion events, and SLA. *The Modern Language Journal, 99*, 1-13.

Bylund, E., & Athanasopoulos, P. (2015b). Televised Whorf: Cognitive restructuring

in advanced foreign language learners as a function of audiovisual media exposure. *The Modern Language Journal, 99*, 123-137.

Bylund, E., Athanasopoulos, P., & Oostendorp, M. (2013). Motion event cognition and grammatical aspect: Evidence from Afrikaans. *Linguistics, 51*(5), 929–955.

Bylund, E., & Jarvis, S. (2011). L2 effects on L1 event conceptualization. *Bilingualism: Language and Cognition, 14*(1), 47-59. doi:10.1017/S1366728910000180

Cadierno, T. (2004). Expressing motion events in a second language: A cognitive typological perspective. In M. Achard & S. Neimeier (Eds.), *Cognitive linguistics, second language acquisition and foreign language pedagogy*. Berlin: De Gruyter Mouton.

Cadierno, T. (2010). Motion in Danish as a second language: Does the Learner's L1 make a difference? In Z. Han & T. Cadierno (Eds.), *Linguistic relativity in SLA: Thinking for speaking* (pp. 1-33). Bristol: Multilingual Matters.

Cadierno, T. (2012). Thinking for speaking in second language acquisition. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Hoboken, NJ: Wiley-Blackwell.

Cadierno, T., & Ruiz, L. (2006). Motion events in Spanish L2 acquisition. *Annual Review of Cognitive Linguistics, 4*, 183-216.

Carroll, M. (1993). Deictic and intrinsic orientation in spatial descriptions: A comparison between English and German. In J. Altarriba (Ed.), *Advances in psychology* (Vol. 103, pp. 23-44). Netherlands: North-Holland.

Carroll, M. (2000). Representing path in language production in English and German: Alternative perspectives on figure and ground. In C. Habel & C. von Stutterheim (Eds.), *Räumliche Konzepte und sprachliche Strukturen* (pp. 97–118). Tübingen: Max Niemeyer Verlag.

Carroll, M., & Lambert, M. (2006). Reorganizing principles of information structure in advanced L2s: French and German learners of English. In H. Byrnes, D. H. Weger-Guntharp, & K. Sprang (Eds.), *Educating for advanced foreign language capacities* (pp. 54-73). Georgetown: Georgetown University Press.

Carroll, M., & von Stutterheim, C. (2003). Typology and information organisation: Perspective taking and language-specific effects in the construal of events. In A. G. Ramat (Ed.), *Typology and second language acquisition* (pp. 365-402). Berlin: De Gruyter Mouton.

Carroll, M., von Stutterheim, C., & Nüse, R. (2004). The language and thought debate:

A psycholinguistic approach. In P. Thomas & H. Christopher (Eds.), *Approaches to language production* (pp. 183-218). Berlin: Mouton.

Carroll, M., Weimar, K., Flecken, M., Lambert, M., & von Stutterheim, C. (2012). Tracing trajectories: Motion event construal by advanced L2 French-English and L2 French-German speakers. *Language, Interaction and Acquisition, 3*(2), 202-230.

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. *Language learning, 58*, 63-79.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition, 106*(2), 579-593.

Chan, M. (1980). Temporal reference in Mandarin Chinese: An analytical-semantic approach to the study of the morphemes le, zai, zhe, and ne. *Journal of the Chinese Language Teachers Association, 15*(3), 33-79.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, Los Angeles: University of California Press.

Chappell, H., & Peyraube, A. (2008). Chinese localizers: Diachrony and some typological considerations. In D. Xu (Ed.), *Space in languages of China: Cross-linguistic, synchronic and diachronic perspectives* (pp. 15-37). Dordrecht: Springer.

Chen, C.-y. (1978). Aspectual features of the verb and the relative positions of the locatives. *Journal of Chinese Linguistics, 6*(1), 76-103.

Chen, L. (2007). *The Acquisition and Use of Motion Event Expressions in Chinese*. Munich: Lincom Europa.

Chen, L., & Guo, J. (2009). Motion events in Chinese novels: Evidence for an equipollently-framed language. *Journal of Pragmatics, 41*(9), 1749-1766. doi:https://doi.org/10.1016/j.pragma.2008.10.015

Cheng, Y. (1988). *Deutsche und chinesische Bewegungsverben: Ein sprachdidaktischer Vergleich ihrer Semantik und Valenz*. Berlin: De Gruyter Mouton.

Chirkova, K., & Lamarre, C. (2005). The paradox of the construction [V zai NPloc] and its meanings in the Beijing dialect of Mandarin. *Cahiers de Linguistique - Asie Orientale, 34*(2), 169-220.

Chu, C. C. (1976). Some semantic aspects of action verbs. *Lingua, 40*(1), 43-54.

Chuang, S.-y. (2010). *Sprache und Sprachverwendung am Beispiel der Raumreferenz: Eine vergleichende Studie zwischen dem Chinesischen und dem Deutschen.*

(Doctoral thesis), University of Heidelberg.

Comrie, B. (1976). *An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Croft, W. (2012). *Verbs: Aspect and clausal structure*. Oxford: Oxford University Press.

Dahl, Ö. (1985). *Tense and aspect systems*. Oxford: Blackwell.

Daller, M. H., Treffers-Daller, J., & Furman, R. (2011). Transfer of conceptualization patterns in bilinguals: The construal of motion events in Turkish and German. *Bilingualism: Language and Cognition, 14*(1), 95-119.

Ebert, K. H. (2000). Progressive markers in Germanic languages. In Ö. Dahl (Ed.), *Tense and aspect in the languages of Europe* (pp. 605-654). Berlin: De Gruyter Mouton.

Fan, J. (1982). Lùn jiècí duǎnyǔ zài+chùsuǒ [On the adpositional phrase zài+ location]. *yuyan yanjiu [Language Research], 2*, 71-86.

Flecken, M., Carroll, M., Weimar, K., & Von Stutterheim, C. (2015). Driving along the road or heading for the village? Conceptual differences underlying motion event encoding in French, German, and French–German L2 users. *The Modern Language Journal, 99*, 100-122.

Flecken, M., Gerwien, J., Carroll, M., & Von Stutterheim, C. (2015). Analyzing gaze allocation during language planning: a cross-linguistic study on dynamic events 1. *Language and Cognition, 7*(1), 138-166.

Flecken, M., von Stutterheim, C., & Carroll, M. (2014). Grammatical aspect influences motion event perception: Findings from a cross-linguistic non-verbal recognition task. *Language and Cognition, 6*(1), 45-78.

Gennari, S. P., Sloman, S. A., Malt, B. C., & Fitch, W. T. (2002). Motion events in language and cognition. *Cognition, 83*(1), 49-79.

Gentner, D., & Goldin-Meadow, S. (2003). *Language in mind: Advances in the study of language and thought*. Cambridge, MA, US: The MIT Press.

Gerwien, J., & von Stutterheim, C. (2018). Event segmentation: Cross-linguistic differences in verbal and non-verbal tasks. *Cognition, 180*, 225-237. doi:https://doi.org/10.1016/j.cognition.2018.07.008

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science, 11*(4), 274-279.

Griffin, Z. M., & Spieler, D. H. (2006). Observing the what and when of language production for different age groups by monitoring speakers' eye movements.

Brain and Language, 99(3), 272-288.

Gumperz, J. J., & Levinson, S. C. (1996). *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.

Habel, C., & Tappe, H. (1999). Processes of segmentation and linearization in describing events. In C. von Stutterheim & R. Meyer-Klabunde (Eds.), *Processes in language production* (pp. 117-152). Opladen: Westdeutscher Verlag.

Hasko, V. (2010). Semantic composition of motion verbs in Russian and English: The case of intra- typological variability. In V. Hasko & R. Perelmutter (Eds.), *New approaches to Slavic verbs of motion* (pp. 197-224). Amsterdam: John Benjamins Publishing Company.

Hendriks, H., & Hickmann, M. (2011). Space in second language acquisition. In V. Cook & B. Bassetti (Eds.), *Language and bilingual cognition* (pp. 315-339). Hove, UK: Psychology Press.

Hendriks, H., & Hickmann, M. (2015). Finding one's path into another language: On the expression of boundary crossing by English learners of French. *The Modern Language Journal, 99*, 14-31.

Hendriks, H., Hickmann, M., & Demagny, A.-C. (2008). How adult English learners of French express caused motion: A comparison with English and French natives. *Acquisition et interaction en langue étrangère*(27), 15-41.

Hickmann, M., & Hendriks, H. (2010). Typological constraints on the acquisition of spatial language in French and English. *Cognitive Linguistics, 21*(2), 189-215.

Hohenstein, J., Eisenberg, A. N. N., & Naigles, L. (2006). Is he floating across or crossing afloat? Cross-influence of L1 and L2 in Spanish–English bilingual adults. *Bilingualism: Language and Cognition, 9*(3), 249-261. doi:10.1017/S1366728906002616

Hsueh, F. (1989). The structure meaning of bă and bèi constructions in Mandarin Chinese. *Functionalism and Chinese grammar*, 95-125.

Huang, C. T. J. (1988). Wǒ pǎo de kuài and Chinese phrase structure. *Language, 64*(2), 274-311. doi:10.2307/415435

Jackendoff, R. (1990). What would a theory of language evolution have to look like? *Behavioral and Brain Sciences, 13*(4), 737-738.

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. New York: Routledge.

Jin, L., & Hendriks, H. (2005). The development of aspect marking in L1 and L2 Chinese. *Working Papers in English and Applied Linguistics, 9*, 69-99.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

Kang, J. (2001). Perfective aspect particles or telic Aktionsart markers? Studies of the directional verb compounds. *Journal of Chinese Linguistics, 29*(2), 281-339.

Kellerman, E., & van Hoof, A.-M. (2003). Manual accents. *International Review of Applied Linguistics in Language Teaching, 41*(3), 251-269.

Kersten, A. W., Meissner, C. A., Lechuga, J., Schwartz, B. L., Albrechtsen, J. S., & Iglesias, A. (2010). English speakers attend more strongly than Spanish speakers to manner of motion when classifying novel objects and events. *Journal of Experimental Psychology: General, 139*(4), 638-653. doi:10.1037/a0020507

Klein, W. (1994). *Time in language*. London: Routledge.

Klein, W., & Li, P. (2009). *The expression of time*. Berlin, New York: Mouton de Gruyter.

Klein, W., Li, P., & Hendriks, H. (2000). Aspect and assertion in Mandarin Chinese. *Natural Language & Linguistic Theory, 18*(4), 723-770.

Lamarre, C. (2008). The linguistic categorization of deictic direction in Chinese — with reference to Japanese —. In D. Xu (Ed.), *Space in languages of China: Cross-linguistic, synchronic and diachronic perspectives* (pp. 69-97). Dordrecht: Springer.

Langacker, R. W. (2008). *Cognitive grammar: A basic introduction*. New York: Oxford University Press.

Levelt, W. J. (1982). Linearization in describing spatial networks. In S. Peters & E. Saarinen (Eds.), *Processes, beliefs, and questions* (pp. 199-220). Dordrecht: Springer.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, Massachusetts: The MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1-75. doi:10.1017/S0140525X99001776

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Li, M. (1999). *Negation in Chinese*. Manchester: University of Manchester.

Li, P. (1988). Acquisition of spatial reference in Chinese. In P. Jordens & J. Lalleman (Eds.), *Language development*. Dordrecht: Foris.

Li, P. (1990). *Aspect and Aktionsart in child Mandarin.* (Doctoral thesis), University of Leiden, Netherlands.

Li, P. (1993). The acquisition of the zai and ba constructions in Mandarin Chinese. In J. C. P. Liang & R. P. E. Sybesma (Eds.), *From Classical Fü to 'Three Inches High': Studies on Chinese in Honor of Erik Zürcher* (pp. 103-120). Leuven/Apeldoorn: Garant Publishers.

Li, P., & Bowerman, M. (1998). The acquisition of lexical and grammatical aspect in Chinese. *First Language, 18*(54), 311-350.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition, 83*(3), 265-294. doi:https://doi.org/10.1016/S0010-0277(02)00009-4

Li, Z. (2018). *Morphosyntaktische Eigenschaften der Chinesischen Serial-Verb-Konstruktion*. Master's thesis. University of Heidelberg.

Liu, D. (2002). Circumpositions in Chinese. *Contemporary Linguistics, 4*, 241-253.

Liu, D. (2008). Syntax of space across Chinese dialects: Conspiring and competing principles and factors. In D. Xu (Ed.), *Space in languages of China: Cross-linguistic, synchronic and diachronic perspectives* (pp. 39-67). Dordrecht: Springer.

Liu, F.-h. (2009). Aspect and the post-verbal zài phrase in Mandarin Chinese. In J. z. Xing (Ed.), *Studies of Chinese linguistics: Functional approaches* (pp. 103-129). Hong Kong: Hong Kong University Press.

Liu, N. (1985). Lùn –zhe jíqí xiāngguān-de liǎnggè dòngtài fànchóu (On -zhe and its two related verbal categories). *yuyan yanjiu (Language Research), 9*, 117-127.

Liu, Y. (1998). *Qūxiàng bǔyǔ tōngshì (On directional complements)*. Beijing: Beijing Language and Culture University Press.

Ma, X.-w. (1987). Běijīng fāngyánlǐ-de -zhe (-zhe in the Beijing dialect). *Fangyan (Dialectology), 1*, 17-22.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences, 8*(3), 108-114. doi:https://doi.org/10.1016/j.tics.2004.01.003

Matthews, S. J. (1990). *A cognitive approach to the typology of verbal aspect.* (Doctoral

thesis), University of Southern California.

Mayer, M. (1969). *Frog, where are you?* New York: Dial.

Müller, S., & Lipenkova, J. (2009). Serial verb constructions in Chinese: A HPSG account. In S. Müller (Ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar* (pp. 234-254). Stanford, CA: CSLI Publications.

Navarro, S., & Nicoladis, E. (2005). *Describing motion events in adult L2 Spanish narratives.* Paper presented at the Selected proceedings of the 6th conference on the acquisition of Spanish and Portuguese as first and second languages.

Negueruela, E., Lantolf, J. P., Jordan, S. R., & Gelabert, J. (2004). The "private function" of gesture in second language speaking activity: A study of motion verbs and gesturing in English and Spanish. *International Journal of Applied Linguistics, 14*(1), 113-147. doi:10.1111/j.1473-4192.2004.00056.x

Olsen, S. (1996). Partikelverben im deutsch-englischen Vergleich. In E. Lang & G. Zifonum (Eds.), *Deutsch- typologisch* (pp. 261-288). Berlin, Boston: de Gruyter.

Packard, J. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.

Papafragou, A., Hulbert, J., & Trueswell, J. (2008). Does language guide event perception? Evidence from eye movements. *Cognition, 108*, 155-184.

Papafragou, A., & Selimis, S. (2010). Event categorization and language: A cross-linguistic study of motion. *Language and cognitive processes, 25*(2), 224-260.

Pavlenko, A. (2005). Bilingualism and thought. In K. F. Judith & D. G. M. B. Annette (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 433-453). Oxford: Oxford University Press.

Pavlenko, A., & Volynsky, M. (2015). Motion encoding in Russian and English: Moving beyond Talmy's typology. *The Modern Language Journal, 99*(S1), 32-48.

Qi, H. (1998). *Xiàndài hànyǔ kōngjiān wèntí yánjiū (Research on space in Chinese)*. Shanghai: Xuelin Press.

Qian, X. (1997). Rìběn liúxuéshēng hànyǔ qūxiàng bǔyǔ-de xídé shùnxù (The sequence of acquiring Chinese directional complements by Japanese learners). *World Chinese Teaching, 1*, 94-101.

Rohsenow, J. (1976). A unified treatment of lexical, verbal, and sentential aspect in Mandarin Chinese. In S. S. Mufwene, C. A. Walker, & S. B. Steever (Eds.),

Papers from the 12th regional meeting of the Chicago linguistic society. Chicago: Chicago Linguistic Society.

Rohsenow, J. (1978). *Syntax and Semantics of the Perfect in Mandarin Chinese.* (Doctoral thesis), University of Michigan, Ann Arbor.

Schmiedtová, B. (2013). Traces of L1 patterns in the event construal of Czech advanced speakers of L2 English and L2 German. *International Review of Applied Linguistics in Language Teaching, 51*(2), 87-116.

Schmiedtová, B., & Flecken, M. (2008). Aspectual concepts across languages: Some considerations for second language learning. In S. De Knop & T. De Rycker (Eds.), *Cognitive approaches to pedagogical grammar* (pp. 357-384). Berlin: Mouton de Gruyter.

Schmiedtová, B., & Sahonenko, N. (2008). Die Rolle des grammatischen Aspekts in Ereignis-Enkodierung: Ein Vergleich zwischen tschechischen und russischen Lernern des Deutschen. In M. Walter & P. Grommes (Eds.), *Fortgeschrittene Lernervarietäten: Korpuslinguistik und Zweitspracherwerbforschung* (pp. 45-71). Berlin, Boston: De Gruyter.

Schmiedtová, B., Von Stutterheim, C., & Carroll, M. (2011). Language-specific patterns in event construal of advanced second language speakers. In A. Pavlenko (Ed.), *Thinking and speaking in two languages* (pp. 66-107). Bristol: Multilingual Matters.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. Gumperz & S. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70--96). Cambridge: Cambridge University Press.

Slobin, D. I. (2000). Verbalized events: A dynamic approach to linguistic relativity and determinism. In S. Niemeier & R. Dirven (Eds.), *Evidence for linguistic relativity* (pp. 107-138). Amsterdam: John Benjamins Publishing Company.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought.* (pp. 157-192). Cambridge, MA: The MIT Press.

Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömqvist & L. Verhoeven (Eds.), *Relating events in narrative: Vol. 2. Typological and contextual perspectives* (pp. 219-257). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Slobin, D. I. (2006). What makes manner of motion salient? Explorations in linguistic typology, discourse, and cognition. In M. Hickmann & S. Robert (Eds.), *Space in Languages: Linguistic systems and cognitive categories* (pp. 59-81). Amsterdam: John Benjamins Publishing Company.

Slobin, D. I. (2008). Relations between paths of motion and paths of vision: A crosslinguistic and developmental exploration. In V. C. M. Gathercole (Ed.), *Routes to language: Studies in honor of Melissa Bowerman* (pp. 197-221). Mahwah, NJ: Lawrence Erlbaum Associates.

Slobin, D. I., & Hoiting, N. (1994). Reference to movement in spoken and signed Languages: Typological considerations *Proceedings of the Twentieth Annual Meeting of the Berkeley Linguistics Society: General Session Dedicated to the Contributions of Charles J. Fillmore* (pp. 487-505).

Smith, C. S. (1997). *The parameter of aspect*. Dordrecht: Kluwer.

Stam, G. (2001). *Gesture and second language acquisition*. Paper presented at the TESOL Convention, St Louis, Missouri.

Stam, G. (2015). Changes in thinking for speaking: A longitudinal case study. *The Modern Language Journal, 99*, 83-99.

Stiebels, B. (2015). *Lexikalische Argumente und Adjunkte: zum semantischen Beitrag von verbalen Präfixen und Partikeln*. Berlin: Akademie Verlag.

Tai, J. H. (1973). *On the center of predication in Chinese verb-complement construction.* Paper presented at the Annual meeting of linguistics society of America.

Tai, J. H. (1975). On two functions of place adverbials in Mandarin Chinese. *Journal of Chinese Linguistics, 3*, 154-179.

Tai, J. H. (2003). Cognitive relativism: Resultative construction in Chinese. *Language and Linguistics, 4*(2), 301-316.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description: Vol. 3. Grammatical categories and the lexicon* (pp. 225-282). Cambridge: Cambridge University Press.

Talmy, L. (1988). The relation of grammar to cognition. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics*. Amsterdam: John Benjamins Publishing Company.

Talmy, L. (2000). *Toward a cognitive semantics: Vol. ll. Typology and process in*

concept structuring. Cambridge: The MIT Press.

Talmy, L. (2009). Main verb properties and equipollent framing. In J. Guo, E. Lieven, N. Budwig, S. Ervin-Tripp, K. Nakamura, & S. Ozcaliskan (Eds.), *Crosslinguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 389-402). New York: Psychology Press.

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J.-R. (2009). Unconscious effects of language-specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences, 106*(11), 4567. doi:10.1073/pnas.0811155106

Vendler, Z. (1967). *Linguistics in philosophy*. Ithaca, NY: Cornell University Press.

Vlach, F. (1981). The semantics of the progressive. In P. Tedeschi & A. Zaenen (Eds.), *Syntax and semantics (Vol.14: Tense and aspect)* (pp. 271-292). New York: Academic Press.

von Humboldt, W. (1963). *Schriften zur Sprachphilosophie*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Von Stutterheim, C. (2003). Linguistic structure and information organisation: The case of very advanced learners. In S. H. Foster-Cohen & S. P. Doehler (Eds.), *EuroSLA yearbook* (Vol. 3, pp. 183-206). Amsterdam: John Benjamins Publishing Company.

von Stutterheim, C., Andermann, M., Carroll, M., Flecken, M., & Schmiedtová, B. (2012). How grammaticized concepts shape event conceptualization in language production: Insights from linguistic analysis, eye tracking data, and memory performance. *Linguistics, 50*(4), 833-867.

von Stutterheim, C., Bouhaous, A., & Carroll, M. (2017). From time to space: The impact of aspectual categories on the construal of motion events: The case of Tunisian Arabic and Modern Standard Arabic. *Linguistics, 55*(1), 207-249.

von Stutterheim, C., & Carroll, M. (2006). The impact of grammaticalized temporal categories on ultimate attainment in advanced L2 acquisition. In H. Byrnes, D. H. Weger-Guntharp, & K. Sprang (Eds.), *Educating for advanced foreign language capacities. constructs, curriculum, instruction, assessment* (pp. 40-53). Georgetown: Georgetown University Press.

von Stutterheim, C., Carroll, M., & Klein, W. (2009). New perspectives in analyzing aspectual distinctions across languages. In W. Klein & P. Li (Eds.), *The expression of time* (pp. 195-216). Berlin: De Gruyter Mouton.

von Stutterheim, C., Flecken, M., & Carroll, M. (2013). Introduction: Conceptualizing in a second language. *International Review of Applied Linguistics in Language Teaching, 51*(2), 77-85.

von Stutterheim, C., & Lambert, M. (2005). Crosslinguistic analysis of temporal perspectives in text production. In H. Hendriks (Ed.), *The structure of learner varieties*. Berlin: Mouton de Gruyter.

Von Stutterheim, C., & Nüse, R. (2003). Processes of conceptualization in language production: Language-specific perspectives and event construal. *Linguistics, 41*(5), 851-881.

Wang, X. (2015). *Der Einfluss der sprachlichen Struktur auf die Konzeptualisierung von Bewegungsereignissen.* (Master's thesis), University of Heidelberg.

Whorf, B. L. (2012). Languages and logic (1941). In J. B. Carroll, S. C. Levinson, & P. Lee (Eds.), *Language, thought, and reality* (pp. 299-314). Cambridge, London: The MIT Press.

Whorf, B. L. (2012). Language, mind, and reality (1941). In J. B. Carroll, S. C. Levinson, & P. Lee (Eds.), *Language, thought, and reality* (pp. 315-344). Cambridge, London: The MIT Press.

Woo, I.-h. (2013). *The syntax of the aspectual particles in Mandarin Chinese.* (Doctoral thesis), Boston University.

Wu, F. (2008). Origin and evolution of the locative term hòu 'BACK' in Chinese. In D. Xu (Ed.), *Space in languages of China: Cross-linguistic, synchronic and diachronic perspectives* (pp. 229-247). Dordrecht: Springer.

Wu, S. L. (2011). Learning to express motion events in an L2: The case of Chinese directional complements. *Language learning, 61*(2), 414-454.

Wu, S. L. (2014). Influence of L1 thinking for speaking on use of an L2: The case of path expressions by English-speaking learners of Chinese. In Z. Han (Ed.), *Studies in second language acquisition of Chinese*. Bristol: Multilingual Matters.

Xiao, R., & McEnery, T. (2004). *Aspect in Mandarin Chinese: A corpus-based study*. Amsterdam: John Benjamins Publishing Company.

Yang, D. (2003). Sequence of acquiring the directional complement by English speaking learners of Chinese. *Chinese Teaching in the World, 64*(2), 52-65.

Yong, S. (1997). The grammatical functions of verb complements in Mandarin Chinese. *Linguistics, 35*(1), 1-24.

Zhu, D. (1982). *Yǔfǎ jiǎngyì (Notes on Grammar)*. Beijing: Shangwu Press.

Zlatev, J., & Yangklang, P. (2004). A third way to travel: The place of Thai in motion event typology. In S. Strömqvist & L. Verhoeven (Eds.), *Relating events in narrative (Vol. 2: Typological and contextual perspectives)* (pp. 159-190). Mahwah, NJ: Lawrence Erlbaum Associates.