

Chapter 1. General Introduction

1.1 Research background and context

Our life consists of various kinds of events as time flows. People around the world experience events in the present, recall events in the past, and plan events in the future. However, different languages throughout the world provide their speakers with different linguistic repertoires to characterize how events unfold in time, such as tense and aspect systems. Furthermore, human communication is (variably) multimodal (Cienki 2015, 2016). While narrating events face to face, people use not only speech but also manual gestures to co-construct them and convey meanings. In this light, this dissertation will focus on how English speakers and Chinese speakers conceptualize events via the verbal and gestural “windows”, particularly when they take an explicit aspectual viewpoint on an event in speech. Although, technically, the topic in this dissertation concerns bimodal communication (vocal modality via speech and visual modality via gesture), the conventional term is now “multimodal”, covering even just two modalities of production or reception. Therefore, a multimodal approach is taken to study the role of grammatical aspect in event cognition during the speakers’ face-to-face communication.

Tense is deictic by nature, locating events on a unidirectional timeline with the time of speaking as a reference point (e.g., Comrie 1976, 1985; Klein 1994). An event can precede the time of speaking (i.e., past tense), overlap with the time of speaking (i.e., present tense), or follow the time of speaking (i.e., future tense). The temporal contours of event structures are encoded in aspect. Aspects are “different ways of viewing the internal constituency of a situation” (Comrie 1976: 3). The category of aspect can be realized lexically or grammatically. Lexical aspect is relatively universal semantically, involving the encoding of the inherent temporal information of events in terms of dynamicity, durativity, and telicity (Vendler 1967). Grammatical aspect

conveys, for example, whether an event is ongoing or completed. Grammatical aspectual constructions are learned “form and meaning pairings” (Goldberg, 2006: 3), which consist of grammatical aspect markers and conventional meanings. The basic aspectual contrast can be boiled down to one between perfective vs. imperfective aspect. The perfective aspect views a situation as a complete, single whole with “beginning, middle and end rolled into one”, while the imperfective aspect explicitly refers to “the internal temporal structure of a situation, viewing a situation from within” (Comrie 1976: 24). Aspectual categories can be said to be motivated by the basic human mental scanning capacities: “sequential scanning” and “summary scanning” (Langacker 2008a: 111). People can mentally track an event as it plays out in time, i.e., scanning sequentially the internal phases of an event (Langacker 2008a). This corresponds to the imperfective aspect in language. An alternative viewing mode involves mentally superimposing an event into a single gestalt (Langacker 2008a). This kind of summary scanning is reflected in the perfective aspect in language. The option of conceptualizing an event in either mode of scanning provides the possibility for people to construe the same event in either grammatical aspect. Apparently, grammatical aspect and basic human cognition are highly interrelated.

English and Chinese, as two typologically different languages, have both similarities and differences in expressing temporal information about events in terms of grammatical verb tense and aspect. English has a complete tense system, but it only grammatically marks ongoing events via the progressive aspect (*be V-ing*). Chinese is a tenseless language, but it has a finer-grained aspect system. The basic aspectual distinction is between perfective and imperfective aspect at the macro level. At the micro level, the perfective aspect consists of the actual aspect (*V le*) and the experiential aspect (*V guo*), while the imperfective aspect is made up of the progressive aspect (*zài V*) and the durative aspect (*V zhe*). According to the “thinking for speaking” hypothesis (Slobin 1987, 1996), differences in the grammatical options that different languages or even a single language provide (s) for characterizing events influences thinking in the moment-by-moment processes of speaking. Questions, then, arise including: how might the different grammatical aspectual options available in Chinese affect native Chinese

speakers' thinking for speaking? How might the single grammatical aspect option in English, with a wide range of uses, be structuring native English speakers' thinking for speaking? Moreover, do L1 Chinese L2 English speakers¹ adjust their thinking for speaking while construing events in the L2? Solving such problems can provide a much more comprehensive picture for us to understand the relationship between language and thought, particularly between grammar and cognition in the conceptualization of basic human experiences, i.e., of time and events. However, if we only focus on the written language or even the written transcriptions of spoken language, it is inevitable that our reasoning of language to thought will be limited. However, spontaneous co-speech gestures "offer insights into imagistic and dynamic forms of thinking while speaking and gesturing" (McNeill 2012: 29). In order to reveal what speakers are really thinking while speaking about events, we investigate how they are gesturing while talking about the events in spoken English and Chinese contexts.

Under the framework of cognitive linguistics, language is considered as an integral part of human cognition, and it consists of symbolic structures. Every symbolic structure is made up of a semantic structure and a phonological structure, which can include not only sounds but also gestures and/or orthographic representations (Langacker 2008a: 15). That is, in face to face communication, conceptualization can be "symbolized by means of sounds and gestures" (Langacker 2008a: 7). Therefore, language has the potential to be a multimodal symbolic structure. By gestures, we mean the unwitting movements of arms and hands that accompany speech and are used as utterances or as part of utterances (Kendon 2004; McNeill 1992, 2005). They are temporally, semantically and pragmatically synchronized with speech, so speech and co-speech gestures are argued to be an integrated system (Kendon 2004; McNeill 1992, 2005, 2012). Manual gestures, reflecting imagistic aspects of the concepts in speakers' thought processes, make certain aspects of meaning visible in our hands. In addition, gesture use also varies in some ways between speakers of different languages,

¹ People whose first language is Chinese, speaking English as a second language.

according to linguistic similarities and differences. In opposition to the generative view that “grammar is autonomous and independent of meaning” (Chomsky 1957: 17), Langacker (1987, 1991a, 2008a) has argued that grammar is inherently meaningful and cognitively motivated. Grammar can not only reflect human cognition but also influence our cognition. Therefore, the basic grammatical aspect contrast in language reflects people’s different ways of cognizing about events. The imperfective aspect allows the speaker and listener an internal view of event descriptions, giving access to details about the event, while the perfective aspect gives them an external view of an event, backgrounding access to event details. Such grammatical differences have also been shown in speakers’ co-speech gestures. While conceptualizing motion events, the gestures accompanying imperfective aspect (e.g., English progressive aspect) utterances were found to be longer and more complex (i.e., encoding more semantic elements of an event) than those with non-imperfective aspect (i.e., English non-progressive aspect, such as the past tense and the perfect tense) (Duncan 2002; Parrill et al. 2013). However, compared with the wide range of studies on linguistic aspect, we still know very little about the imagistic thinking about grammatical aspect in event conceptualization, particularly for the much more complex aspect system in Chinese. To what extent do the finer-grained Chinese grammatical aspect differences in language correlate with different gestural representations of events in natural conversations? Although speech and gestures are said to convey the same underlying ideas, it is unclear whether speakers’ gestures are sensitive to the different meanings with which individual grammatical aspectual forms are used, such as in the polysemous durative aspect in Chinese and the English progressive aspect.

Contrary to the traditional view that language expresses meaning via abstract, amodal and arbitrary words combined by syntactic rules (Chomsky 1957, 1965; Fodor 2000; Jackendoff 2002; Pinker 1994), cognitive linguistics holds that linguistic meaning is embodied, i.e., grounded in our bodily experience (e.g., Barsalou 1999; Lakoff 1987; Lakoff & Johnson 1980; Langacker 1987, 1991a, 2008a). Sensory-motor experiences play a critical role in cognition. We mainly express and understand the non-physical in terms of the physical. Thus, abstract concepts are grounded metaphorically

in embodied and situated knowledge (Gibbs 1994; Lakoff & Johnson 1980). Temporality, realized by grammatical aspect, is abstract by nature. Are grammatical aspects embodied in our experience? How are they represented in speakers' co-speech gestures? According to the gestures as simulated action (GSA) framework, gestures derive from "the perceptual and motor simulations that underlie embodied language and mental imagery" (Hostetter & Alibali 2008: 502). The progressivity encoded by the progressive aspect has been reported to be represented in some cases by cyclic gestures (i.e., a continuous circling motion of the loose hand) (Harrison 2009; Ladewig 2011). The cycle schema of temporality results from our perception of time in everyday life, such as the cyclic change of day and night, the four seasons in a year, etc. In addition, the durative characteristics of English progressive have also been found to be mapped sometimes onto the metaphoric container gesture (i.e., holding both hands in parallel apart in front of the speakers' body at chest height with palms facing each other) (Matsumoto & Dobs 2016). Such a gesture is doubtlessly embodied in our everyday activity, such as counting the time spent in finishing an event of running 100 meters measured by a line segment on the playground. Such gestural features were only based on the qualitative observation of inherently abstract events, i.e., highly schematic expressions of events, (e.g., *there's something going on in the city*). What are the gestures like in terms of the role of grammatical aspect in abstract events, not only the inherently abstract but also metaphorically abstract events (e.g., *The price of oil is climbing*)? According to psychological studies on the role of grammatical aspect in language comprehension and production, in contrast with the perfective aspect (non-progressive aspect in English), the imperfective aspect (English progressive aspect) has been found to activate more vivid mental simulation of actions and locations (Bergen & Wheeler 2010; Liu & Bergen 2016), allocate more attention to the middle part of an event (Magliano & Schleich 2000; Madden & Zwaan 2003; Athanasopoulos & Bylund 2013a; Huette et al. 2014), and cause people to infer more actions in a given period (Matlock 2010, 2011; Matlock et al. 2012). All of these previous studies focused on concrete (motion) events, except Liu and Bergen (2016), who took abstract events into consideration. They found that concrete language facilitates the mental simulation of

location of an event in the progressive aspect, whereas abstract language inhibits it, regardless of the use of the progressive aspect or the non-progressive aspect. Such findings may challenge the embodied view of language. Thus, we will study whether gestures accompanying concrete and abstract events can provide new evidence to deepen our understanding about the embodiment of grammatical aspect in abstract events.

Tense and aspect has invariably been one of the most difficult points to acquire in second language acquisition (SLA). This is particularly the case for native Chinese speakers learning English in the foreign language context (e.g., Yang et al. 2012). Although the progressive aspect markers in both English and Chinese have similar spatial sources for their development, they have evolved into different grammatical structures (Ruan & Wang 2015: 75). English progressive aspect (*be ... -ing*) does not only conflate tense with aspect, but also encodes a wide range of meanings, such as progressive, habitual, repetitive, temporary state, future and epistemic situations. In contrast, the two Chinese imperfective aspect constructions — progressive aspect (*zài V*) and durative aspect (*V zhe*) — encode only progressive, temporary or resultative state situations. The existing studies on morphological development of tense/aspect marking in SLA have focused on the L2 learners' association of grammatical tense and aspect marking with particular lexical aspects to verify the Aspect Hypothesis (e.g., Andersen & Shirai 1996; Bardovi-Harlig 1994, 1999, 2000; Collins 2002; Lee & Kim 2007). The Aspect Hypothesis claimed that progressive marking begins with activities, and then extends to accomplishments and achievements; progressive markings are not incorrectly overextended to statives (Bardovi-Harlig 2000: 227). There is an extreme research lag in the study of L2 learners' or foreign language (FL) learners' semantic expansion of the progressive use in English, and their multimodal competence of using progressive constructions in event representations. Slobin proposed that the thinking for speaking patterns developed in L1 acquisition are “exceptionally resistant to restructuring in adult second language acquisition” (Slobin 1996: 89). This proposal stimulated a lot of studies on whether the thinking for speaking and gesturing patterns are changeable in SLA. However, the research centered on motion lexicalizations and

produced some inconsistent results. Some found that the bilinguals maintained the L1 thinking for speaking and gesturing patterns (e.g., Kellerman & van Hoof 2003; Negueruella et al. 2004), and others found that the L2 learners are able to restructure their thinking for speaking and gesturing patterns in the L2, particularly when their language proficiency is very high (e.g., Brown 2015, Brown & Gullberg 2008, 2011, 2012; Lewis 2012; Stam 2008, 2010, 2015). This gives rise to the following question. Is it possible for FL learners to change their thinking for speaking and gesturing about events in natural conversations when acquiring grammatical categories, such as the English progressive aspect? This kind of investigation is conducive for us to make clear FL learners' development of mental representations of events in terms of their multimodal competence and to provide cues for teaching and learning the grammatical categories in a multimodal manner. Generally speaking, looking at not only the speech but also the co-speech gestures can help us know more about "the thought processes that co-occur with speech production during language use and acquisition" (Athanasopoulos & Bylund 2013b: 91).

1.2 Research objectives and significance

The overall objective of the dissertation is to reveal the interplay of grammar and gesture in human cognition and communication to whatever degree it is present. I focus on how English and Chinese people express their conceptualization of events via grammatical aspect-marked speech and co-speech gestures in multimodal communication. Given this scope, four more specific aims are set. The first aim is to investigate to what degree the grammatical aspectual distinctions are mapped to the formal and semantic/functional differences in the speakers' co-speech gestures. The second aim is to gain some insights into the effects of grammatical aspectual distinctions on the expression of abstract events. The third aim is to explore the ways in which the different meanings of imperfective aspect in Chinese or progressive aspect in English play out in gestures. The fourth aim is to see whether FL learners change

their thinking for speaking and gesturing in their L2 when taking an internal view (i.e., using progressive aspect) on events.

This dissertation aims for both theoretical significance and important practical relevance. Theoretically, it refreshes our understanding of grammar, grammatical aspect in particular, from a multimodal viewpoint. By systematically looking at the gestures accompanying grammatical aspect-marked utterances, we can understand how grammar takes part in speakers' online conceptualization of events in natural human communication. The co-speech gestural evidence can visually present what properties or elements of the events are prominently simulated in the speakers' mind when they use different grammatical aspects. Different gestural patterns going with different meanings of the same grammatical constructions, such as the Chinese durative aspect construction and the English progressive constructions, would empirically demonstrate that grammar is inherently meaningful and embodied in our physical experience. Moreover, it can reveal how meaning as conceptualization modulates speakers' linguistic as well as imagistic thinking for speaking. On the practical level, this dissertation can lay a foundation for a multimodal approach to language teaching and learning. Native Chinese speakers or native English speakers have different gestural patterns when construing events, but L2 speakers may still gesture in their L1 manner when communicating with the native speakers. As a result, misunderstanding or even irritation may arise in cross-cultural and cross-linguistic communication. Ultimately, this could hinder communication in audio-video conferencing, for example, and more broadly, could have negative effects in international negotiations in political or economic contexts. Thus, it will be argued that it is urgent and fundamental to integrate a multimodal view of spoken language into the L2 or FL teaching curriculum.

1.3 Research questions and methodology

To achieve the objectives above, the following research questions are addressed to guide the studies in the present dissertation:

I . What differences are there, if any, in gestural expressions of the basic Chinese grammatical aspectual contrast (perfective vs. imperfective aspect)?

1. What gestures are prone to accompany the Chinese perfective vs. imperfective aspect-marked utterances?

2. How do the event types affect the gestures accompanying the Chinese perfective vs. imperfective aspect-marked utterances?

II. How do gestures correlate with different meanings of imperfective (progressive) aspects in Chinese and English?

1. What are the characteristics of gestures accompanying imperfective utterances in Chinese marked by two different grammatical aspect particles *zài* and *zhe*?

2. What are the characteristics of gestures accompanying the English progressive construction with different meanings?

III. Do L1 Chinese L2 English FL learners change their thinking for speaking and gesturing patterns to construe events when taking an internal viewpoint (i.e. progressive aspect) in L2?

Following a usage-based approach to language (e.g., Langacker 1987; Barlow & Kemmer 2000), I use qualitative and quantitative corpus-analytical methods in this dissertation. Escaping the bias towards written language in linguistics (Linell 2005), I focus on actual spontaneous spoken language. As stated below, actual usage of language is necessary and essential for the usage-based approach to language:

“You cannot have a usage-based linguistics unless you study actual usage — as it appears in corpora in the form of spontaneous, non-elicited language data or as it appears in an online and elicited from in experimental settings” (Geeraerts & Cuyckens 2007:17).

Spontaneous conversations in a multimodal corpus reflect the way people construe the world around them and interact with it. The results are directly relevant to demonstrating how cognitive motivations manifest themselves in real-world behaviors. Thus, it is of great significance to take “a phenomenological perspective” to understand

speakers' cognition of events and the meaning of grammatical aspects (Zlatev 2016). I have developed three multimodal corpora in this thesis, as follows.

The first one is a Chinese corpus of native Chinese speakers, consisting of original video clips from *Yan Lang One on One* (*Yáng Lán Fǎng Tán Lù*). It has been one of the most popular talk shows in China since 1998. This program is broadcast once a week via Beijing TV. The hostess, *Yang Lan*, interviews one or two celebrities in the fields of entertainment, politics, business, or education either in Chinese or English (with foreign guests like William (Bill) Clinton, Hillary Rodham Clinton, Michael Jordan, etc.). However, I only randomly selected the interviews with the native Chinese speakers. Each episode lasts for about 20 minutes, excluding the advertisements.

The second one is an English corpus of native English speakers, consisting of original video clips from the *Ellen DeGeneres Show*, retrieved from CQPweb via the Red Hen dataset². It is one of the most popular American television talk shows. Each episode consists of four units: the hostess Ellen shares some stories or encyclopedic knowledge with the audience in her monologue; she interviews celebrities from the fields of entertainment, sports, politics, etc.; she plays games with the guests or members of the audience; and there is a breaking news report. However, I only use the interviews from each episode, each of which lasts for about 20 minutes. Further details about the corpus are provided in Chapter 6.

The third one is an FL corpus of L1 Chinese L2 English speakers. It was elicited via a semi-structured, semi-spontaneous format in which participants interviewed each other about their personal experiences with their friends. Ten pairs of participants were recorded. Each pair of participants talked about three events as long as they liked. Their conversations range from 17 to 32 minutes with an average of 22.94 minutes, and the whole FL corpus consists of approximately 3.8 hours of recording. Further details about

² CQPweb (<https://corpora.linguistik.uni-erlangen.de/newsscape/newsscape>) was created by Andrew Hardie from Lancaster University. It is a web-based corpus analysis system, combining power, flexibility and usability to a very high level. It provides the links to the videos we used in this thesis at UCLA Communication Studies Archive (<https://tvnews.sscnet.ucla.edu/edge/>), which is the core part of Red Hen dataset (<http://www.redhenlab.org>).

the corpus are provided in Chapter 7.

The grammatical aspect-marked utterances are the focus in the dissertation. When analyzing the data for different purposes from Chapter 3 to Chapter 7, I present the quantitative results first, and then provide the qualitative analyses to discuss some interesting examples behind the numbers.

1.4 Structure of the dissertation

The dissertation is organized as follows. **Chapter 2** lays the theoretical foundations. It first reviews the previous studies on grammatical aspect in Chinese and English from the fields of linguistics and psychology, and then proposes a multimodal framework to guide the empirical studies carried out in the following chapters. **Chapter 3** investigates the co-speech gestural representation of events marked by different grammatical aspects at the macro- and the micro- levels in Chinese conversations. It reveals to what extent the finer-grained grammatical aspectual distinctions in Chinese are mapped to the formal and functional features of the co-speech gestures in various events. **Chapter 4** addresses the similarities and differences of the speakers' gestures accompanying grammatical aspect-marked utterances (imperfective vs. perfective aspect) of three different event types — concrete events, metaphoric events and inherently abstract events expressed non-metaphorically in natural Mandarin Chinese conversations. It extends the study of Chapter 3 to see how native Chinese speakers visually embody (metaphorically and inherently) abstract events in relation to the concrete events. In **Chapter 5**, I investigate the dynamic or static properties of the gestures accompanying the imperfective utterances with the progressive *zài* and the durative *zhe* to augment the semantic debate on them in Chinese linguistic studies. **Chapter 6** is concerned with the multimodality of the polysemous English progressive construction. Based on the view of meaning as conceptualization, it investigates whether the different meanings or functions of the progressive aspect in English have different multimodal manifestations or not. In view of the results in Chapter 5 and Chapter 6 about the multimodality of

Chinese and English progressive aspect, **Chapter 7** takes a step further to analyze whether, and to what extent, there is evidence that L1 Chinese and L2 English FL learners change their thinking for speaking and gesturing patterns in their L2. **Chapter 8** summarizes the main findings, discusses the possible relationship between grammar, gesture and cognition, points out the limitations of the studies in the present dissertation, and suggests some directions for future research.