Thinking, Speaking and Gesturing
Grammatical Aspect in Spoken Chinese and English from the Perspective of Multimodal Communication

Ronghua Wang
Thinking, Speaking and Gesturing
Grammatical Aspect in Spoken Chinese and English from the
Perspective of Multimodal Communication

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam en Xiamen University
op gezag van de rectores magnifici
prof. dr. V. Subramaniam en prof. dr. Chongshi Zhu,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geesteswetenschappen
op woensdag 6 September 2017 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Ronghua Wang

geboren te Hengyang, China
promotor: prof. dr. A. J. Cienki
copromotor: prof. dr. G. Jiang
Acknowledgements

This thesis came into being with the generous funding from Xiamen University (XMU), China Scholarship Council (CSC) and Vrije Universiteit Amsterdam (VU). It would also have not existed without the effort and help offered by the relevant administrative staff at XMU and VU.

Prof. dr. Michel ter Hark, thank you for offering me the opportunity to pursue double PhD degrees from VU and XMU, where I can have two great supervisors, Prof. Alan Cienki and Prof. Guiying Jiang. My utmost thanks go out to them from the bottom of my heart. Alan, you are so generous with your time, energy and deep academic thinking to guide me through every chapter of the thesis face-to-face at VU and through Skype meetings abroad. Your enthusiasm for academics and your sharp mind in multimodal communication have been inspiring me and encouraging me to do better. You are a very knowledgeable, kind, modest, responsible, rigorous and considerate teacher. I’m lucky to have you as my supervisor and appreciate you to help me shine my brightest light. Prof. Jiang, you are really a wonderful supervisor. Thank you for introducing me to Cognitive Linguistics and Gesture Studies, and giving me enough freedom to do what I want to explore in academics. You are always there for me at the right time and place, academically and also personally. Your trust in my personal engagement in research and your faith in my capacities have provided powerful energy for me to move forward in research and in life. Without your sharp linguistic insights, rich interdisciplinary research experience, strict requirements and patient guides, this thesis would never have gone so smoothly.

I am also a “lucky dog” for having so many kind and helpful teachers, PhD classmates and research fellows on my side during the four years. I extend my sincere thanks to Prof. Longhai Zhang for your wholehearted support for my Double PhD Degree program. Prof. Xinzhang Yang, you deserve my highest respect. Your profound knowledge of linguistics has nourished my heart and mind. I was also very fortunate to enjoy much academic guidance and support given by Prof. Wenbin Wang from Beijing
Foreign Studies University, Prof. Xu Wen from Southwest University, Prof. Qingping Li from Central South University (CSU), and Prof. Yicheng Wu from Zhejiang University. I have been deeply inspired by your enthusiasm for research, rigorous scholarship and inclusiveness. My sincere thanks go to Mr. Hongde Huang in School of Information Science and Engineering, XMU. Mr. Huang, you are a broad-minded person with great foresight. Thank you for giving me so much help and priceless life inspiration. I am also grateful to Professor Huoxiong Wu from School of Mathematical Sciences, XMU, Ms. Xiaoyan Yang from Overseas Education College, XMU, and Mr Zhifeng Chen from Xiangnan University, who helped me and encouraged me to enjoy the process of my PhD study over the years. I owe my special appreciation to Heng Li, Qiaoyun Chen, Chen Han, Chengchen Li, Fei Ye, Mengyun Zhang, Menglan Wang, Hongyan Lu, Shaowen Sun, Xia Wang, Wanyin Xiao, Min Hu, Jiayin Li, Xie Xi, and other academic sisters and brothers in Prof. Jiang’s research group, and one of my former students, Zhenxing Liu. Thank you all for coding data, doing cross-check, sharing your ideas for my research, and helping me with the administrative affairs when I was not in Xiamen. Suwei Wu, I appreciate all that you have done for me during my stay in Amsterdam. Kasper, I admire your excellent academic capacities and your critical thinking in research. Thank you for critical questions to make me reflect on my research. Yao Tong, your independence and diligence impressed me a lot and are worth learning. Bertie and Nicolina, your optimistic attitude towards life and your persistence in research have inspired me. Sila, I am impressed by your kindness and capacity to listen and share. Xing Xing, Yajuan Su, Lu Jin, Xing Dai, Aizhen Zhang, Lichun Lu, Defa Ren, Yifei Qian, Xiangmei Chen, Juan Hu, Jing Zhang, Gongxing Guo, Jingjing Liu and etc., thank you all for learning, singing, climbing the mountains, eating and drinking out together with me, which are happy memories for my life. The friendship from my good friends living together in Uilenstede 510 Amstelveen have also enriched my life.

I am indebted to Prof. dr. Marianne Gullberg, Prof. dr. Marjolijn Verspoor, Prof. dr. Wenbin Wang, Dr. Geert Brône, and Prof. dr. Mike Hannay. Thank you for serving on the reading committee. Prof. Mike Hannay, I am really moved by your careful
revision, giving me your feedback by scanning the whole manuscript page by page. My sincere thanks also go out to Prof. Mark Turner, Prof. Francis Steen and Prof. Peter Uhrig. Thank you for allowing me to use Red Hen Dataset and CQPweb, as well as providing me with all kinds of technical support. What’s more, Associate Prof. Zheng Zhang, Jianming Kuang, and Juezhi Zhou from CSU, and all the participants from XMU and CSU deserve my deeply-felt thanks. I really owe my gratitude to your selfless contributions to my thesis. I also extend my heartfelt thanks to Prof. dr. Haoran Mao, Prof. dr. Yumei Liu, Prof. Xinzhang Yang, Prof. dr. Zhiying Xin and Dr. Shenai Hu. Your thought-provoking questions, valuable suggestions, and possible solutions have improved the quality of my thesis and also facilitated my scientific research quality.

This thesis could not have been completed without the love and support of my dear family. I’d like to express my utmost appreciation to my husband, Yingshi, thank you for accompanying me throughout the process of chasing our dreams together. You deserve my greatest gratitude for doing the experiments in CSU, being the first reader for each article I wrote, tolerating my complaints about thesis writing, and raising me up when I felt low. Last but not least, I owe my respect to my late father-in-law, who left me with a lot of priceless spiritual fortunes such as self-improvement, self-reliance, and enjoyment of love and being loved. Therefore, I am glad to dedicate this thesis to him.
# Contents

## CHAPTER 1. General Introduction ......................................................... 1
1.1 Research background and context ..................................................... 1
1.2 Research objectives and significance ................................................ 7
1.3 Research questions and methodology ................................................. 8
1.4 Structure of the dissertation .................................................................. 11

## CHAPTER 2. Temporality in Language and Mind: A Multimodal Approach to Grammatical Aspect in Event Conceptualization ................................................. 13
2.1 Linguistic approaches to grammatical aspect ....................................... 14
   2.1.1 Linguistic realizations of grammatical aspect in English and Chinese...... 14
   2.1.2 Cognitive account of the meanings of grammatical aspectual constructions ................................................................. 19
2.2 Psychological approach to grammatical aspect ................................... 31
   2.2.1 Grammatical aspect in language comprehension ................................ 31
   2.2.2 Grammatical aspect in language production .................................... 36
2.3 A multimodal approach to language use .......................................... 39
   2.3.1 A multimodal interaction model .................................................... 39
   2.3.2 A multimodal framework to study grammatical aspect .................... 46

3.1 Introduction .......................................................................................... 50
3.2 Methods ............................................................................................... 51
   3.2.1 Multimodal corpus ......................................................................... 51
5.2 Methods. ................................................................................................................................. 109
  5.2.1 Corpus Description ............................................................................................................ 109
  5.2.2 Data Coding ....................................................................................................................... 109
5.3 Results ..................................................................................................................................... 111
  5.3.1 Dynamicity and the gestures with progressive aspect-marked utterances
                                                                                              ................................................................. 112
  5.3.2 Stativity/Dynamicity and the gestures with durative aspect-marked utterances
                                                                                              ......................................................................................................................... 119
5.4 Discussion and conclusions .................................................................................................. 127
  5.4.1 Pure function of “zài V” construction vs. Multi-functions of “V zhe” ..... 127
  5.4.2 Embodied grammatical meaning ....................................................................................... 129

CHAPTER 6. Polysemy and Multimodality: The Co-Speech Gestures

Accompanying the English Progressive Construction

.................................................................................................................................................. 132

6.1 Introduction ............................................................................................................................ 132
6.2 Methods ................................................................................................................................ 134
  6.2.1 Corpus ............................................................................................................................... 134
  6.2.2 Speech coding .................................................................................................................... 135
  6.2.3 Gesture coding ................................................................................................................... 136
6.3 Results .................................................................................................................................... 136
  6.3.1 Frequency of actual usage and the multimodality of the progressive
                                                                                              construction in spoken English .................................................................................... 136
  6.3.2 The characteristics of gestures accompanying the progressive construction
                                                                                              ................................................................................................................................. 139
6.4 Discussion .............................................................................................................................. 142
  6.4.1 Polysemy of the progressive construction and multimodality ...................... 143
  6.4.2 Progressive aspect and multimodality ............................................................................ 150
6.5 Conclusions............................................................................................................................ 154
CHAPTER 7. FL Learners’ Multimodal Representation of the English
Progressive Construction in Event Construal........ 157

7.1 Introduction .......................................................................................................................... 157
7.2 Background .......................................................................................................................... 160
   7.2.1 Thinking for speaking and gesturing in SLA ............................................................... 160
   7.2.2 The influence of grammatical aspect on event construal in L2 .................................... 161
   7.2.3 The functions of gesture in grammar teaching and learning in SLA ......................... 164
7.3 Methods .............................................................................................................................. 166
   7.3.1 Participants ..................................................................................................................... 166
   7.3.2 Data collection .............................................................................................................. 167
   7.3.3 Data treatment .............................................................................................................. 168
7.4 Results ............................................................................................................................... 172
   7.4.1 Speech results ............................................................................................................... 172
   7.4.2 Gesture results ............................................................................................................. 175
7.5 Discussion ............................................................................................................................ 177
   7.5.1 L1 transfer in L2 progressive aspect use ..................................................................... 178
   7.5.2 The link between learners’ gestural usage and development of conceptual
       understanding ..................................................................................................................... 180
7.6 Conclusions ........................................................................................................................ 184

CHAPTER 8. General Discussion and Conclusions ......................... 187

8.1 Summary of the main findings and general discussion .................................................... 187
8.2 Contributions and limitations ............................................................................................ 191
8.3 Future directions for further studies ................................................................................... 194

BIBLIOGRAPHY .......................................................................................................................... 198

APPENDIX 1: Conversation Prompts .................................................................................... 225
APPENDIX 2: Consent Form ..................................................................................................... 226
APPENDIX 3: Handedness Questionnaire ................................................................................. 228
List of Publications ........................................................................................................... 229
Summary of the Thesis .................................................................................................... 231
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACE</td>
<td>action-sentence compatibility effect</td>
</tr>
<tr>
<td>aIPL</td>
<td>anterior inferior partial lobe</td>
</tr>
<tr>
<td>C</td>
<td>conceptualizer</td>
</tr>
<tr>
<td>CLF</td>
<td>classifier</td>
</tr>
<tr>
<td>COP</td>
<td>copula</td>
</tr>
<tr>
<td>DUR</td>
<td>durative</td>
</tr>
<tr>
<td>EMG</td>
<td>electromyography</td>
</tr>
<tr>
<td>EFL</td>
<td>English as foreign language</td>
</tr>
<tr>
<td>fMRI</td>
<td>functional magnetic resonance imaging</td>
</tr>
<tr>
<td>FL</td>
<td>foreign language</td>
</tr>
<tr>
<td>GA</td>
<td>grammatical aspect</td>
</tr>
<tr>
<td>GEN</td>
<td>genitive</td>
</tr>
<tr>
<td>GSA</td>
<td>Gesture as simulated action</td>
</tr>
<tr>
<td>IS</td>
<td>immediate scope</td>
</tr>
<tr>
<td>LCE</td>
<td>location-sentence compatibility effect</td>
</tr>
<tr>
<td>L1</td>
<td>first language</td>
</tr>
<tr>
<td>L2</td>
<td>second language</td>
</tr>
<tr>
<td>MEG</td>
<td>magnetoencephalograph</td>
</tr>
<tr>
<td>MS</td>
<td>maximal scope</td>
</tr>
<tr>
<td>PROG</td>
<td>progressive</td>
</tr>
<tr>
<td>PRT</td>
<td>particle</td>
</tr>
<tr>
<td>SLA</td>
<td>second language acquisition</td>
</tr>
<tr>
<td>V</td>
<td>verb</td>
</tr>
<tr>
<td>VPs</td>
<td>verb predicates</td>
</tr>
</tbody>
</table>
Coding Gesture with Speech: Typographic Conventions

[ . . . ] Brackets the extent of speech with which a gesture co-occurs.
[[ . . . ]] Nested brackets when there are gestures within a gesture phrase.

**bold**

Boldface stroke phases.

——

Bold underline for pre-stroke hold or post-stroke hold phases

——

Double underline for the hold phase of the last gesture, the bold words indicate an addition of a new gesture stroke of the dominant hand.

//

Two enlarged, boldface italic lines indicate the location of a stroke-hold.

< >

pause, number inside is the duration of pause.
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 ($Vle$) and the experiential aspect ($Vguo$), while the imperfective aspect is made up of the progressive aspect ($zài V$) and the durative aspect ($Vzhe$). 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 speakers\(^1\) 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 a 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* (*Yang Lán Fāng 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 dataset\(^2\). 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](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/](https://tvnews.sscnet.ucla.edu/edge/)), which is the core part of Red Hen dataset ([http://www.redhenlab.org](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 aspecual 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.
Chapter 2. Temporality in Language and Mind: A Multimodal Approach to Grammatical Aspect in Event Conceptualization

Time is fundamental and universal to human cognition. There are many ways to express temporal concepts in language, such as with time adverbials. Temporality is inherent to verbs, which profile a process unfolding with time (Langacker 1987, 1991a, 2008a). Furthermore, relations to time are widely encoded in two grammatical categories, tense and aspect, though some languages only express one or the other category morphologically: some languages are rich in aspect but lack tense forms, such as Chinese, and some languages have both tense and aspect, such as English. Tense is objectively deictic, locating events before, around or after the speaking time, whereas aspect is not related to when the event happens, but only to “different ways of viewing the internal temporal constituency of a situation” (Comrie 1976:3). That is, aspect allows people to construe the same event in different ways. For example, in terms of the past event stick floating, you can present it as a completed event The stick floated down stream, or as an ongoing event The stick was floating down stream. These different linguistic expressions of aspectuality may create different mental images in our mind when we read or hear them. Meanwhile, when a speaker is narrating the event, s/he may also make different gestures, for instance, move her/his right extended index diagonally downward in a straight line from left to right, or move both open hands with palms down to the right in an up and down zigzag line, respectively (Parrill et al. 2013). Therefore, not only speech but also co-speech gestures have the potential to mirror the differences between different grammatical aspects. Grammatical aspect, one of the linguistic means to express temporal information, has been widely studied from the perspectives of linguistics and psychology. This chapter reviews the previous studies on grammatical aspect from such different perspectives, and then proposes a
multimodal interaction model, based on which a multimodal framework to study grammatical aspect is presented to guide the empirical studies in the present dissertation.

2.1 Linguistic approaches to grammatical aspect

Linguistic approaches to grammatical aspect have centered on the identification of grammatical aspect markers, the meanings of different grammatical aspect constructions in different contexts, and the collocation of verb aspects and grammatical aspects.

2.1.1 Linguistic realizations of grammatical aspect\(^3\) in English and Chinese

One traditional linguistic approach has focused on the typological differences between languages in terms of grammatical aspect. This research shows that not all languages have grammatical aspect (Dahl 2001). For instance, Russian, Chinese and English are relatively rich in aspectual distinctions, whereas German, Swedish and Afrikaans are non-aspect languages in terms of morphological expression. The basic aspectual contrast is *perfective aspect* vs. *imperfective aspect* (Comrie 1976). While the former indicates that the speaker takes an “external” viewpoint of an event as an entirety, the latter provides the speaker with an “internal” viewpoint of an event (Comrie 1976; Langacker 1991a: 208; Langacker 2001; Smith 1994).

According to Table 2.1, we can see the differences in aspectual markers between English and Chinese. In English, there is only imperfective aspect, specifically progressive aspect, which is marked by *be ... ing*. It is still under debate on whether there are perfective aspects in English. Some scholars consider the simple past and perfect construction (*have...-ed*) as perfective aspect (e.g., Dillon 1973; Palmer 1987; Trager & Smith 1951), or the perfect construction as perfect aspect (Langacker 1991a:196). However, others argue against it on the ground that the simple past and perfect tense have no indication of the internal temporal structure of an event but have

---

\(^3\) Grammatical aspect (Comrie 1976) is also called viewpoint aspect (Smith 1991), or verb aspect (2002) in the literature. This dissertation will follow Comrie’s practice.
direct relation to the time of utterance, and that the perfect form is compatible with the
progressive or present tense (e.g., Fenn 1987; Michaelis 1994). Therefore, in the present
study, I term them as non-progressive. My focus on English grammatical aspect only
concerns progressive aspect, because I aim at investigating the co-speech gestures that
accompany the grammatical aspect-marked utterances.

In contrast, as a “classic tenseless language” (Binnick 1991:446), Chinese is rich
in aspectual expressions both grammatically and periphrastically. It not only has
imperfective aspect, like English does, but also has an explicit perfective aspect. A set
of grammatical markers are available to focus on different temporal phases of an event:
-le on the realization, -qilai on the inception, -xiaqu on the continuation, zài on the
progressivity, -zhe on the durativity, -guo on the experientiality, and resultative verb
constructions like V-wán on the completion (Xiao & McEnery 2004). However, some
of them have begun the process of grammaticalization as aspect markers, such as –qilái,
-xiàqù and -wán. They are mainly considered as lexical aspectualizers (Liu 2013). Thus,
I would like to review the main grammatical aspects in the present study.

Table 2.1: Grammatical aspect in English and Chinese

<table>
<thead>
<tr>
<th>Languages</th>
<th>Imperfective</th>
<th>Perfective</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Progressive aspect (be ... -ing)</td>
<td>?</td>
</tr>
<tr>
<td>Chinese</td>
<td>Progressive aspect (zài)</td>
<td>Actual aspect (-le)</td>
</tr>
<tr>
<td></td>
<td>Durative aspect (-zhe)</td>
<td>Experiential aspect (-guo)</td>
</tr>
</tbody>
</table>

The main imperfective aspect in Chinese involves the progressive aspect, which
is marked by zài, and the durative aspect, which is marked by -zhe. Zài as a progressive
marker is not so well-established, such that only -zhe, -le, -guo have long been
considered as grammatical aspect markers in Chinese (e.g., Chao1968; Comrie 1976;
Dai 1991, 1997; Shi 1992; Teng 1979; Wang 2015; Zhu 1982). Instead, zài was
considered as a time adverb indicating progressive meaning (e.g., Henne et al. 1977;
Dai 1997). Some treated it as an “aspect concept marker” (Shi 2006). Through corpus
study and the diachronic study on Chinese grammatical aspect, it has been demonstrated that \( zài \) is widely used as a progressive aspect marker signaling progressiveness (Chen 2008; Wang 2011, 2012; Xiao & McEnery 2004; Zuo 2007). This dissertation also treats it as a progressive aspect marker, as it indeed provides a critical means to view the internal temporal constituency of events, and it is also frequently used in spoken Chinese.

The durative aspect \(-\text{zhe}\) has been treated as the only imperfective aspect marker in general, and as the progressive aspect marker in particular, for a long time. It was considered as the “progressive suffix” expressing that a dynamic event is in progress (e.g., Chao 1968: 248; Comrie 1976: 88; Tiee 1986: 93). However, other scholars argued against it on the ground that \(-\text{zhe}\) always indicates the continuation of a stative event or the remaining result of an action (e.g., Yeh 1991; Zhang 1996). In this study, I treat \(-\text{zhe}\) as the durative aspect marker emphasizing the durative feature of an event, either dynamic or stative (e.g., Dai 1997; Smith 1997).

There are some similarities and differences between the progressive aspect \( zài \) and the durative aspect \(-\text{zhe}\). Both of them zoom on the intermediate phase of an event, excluding the initial and end points. Syntactically, the progressive aspect marker \( zài \) occurs before the verb, whereas the durative \(-\text{zhe}\) follows the verb. Semantically, progressive aspect \( zài \) only encodes progressivity, but the durative aspect \(-\text{zhe}\) can not only express progressivity, as the progressive aspect \( zài \) does, but it can also encode temporary or resultative states. Pragmatically, both the progressive aspect \( zài \) and durative \(-\text{zhe}\) have “backgrounding” function. That is, the events described with the progressive aspect or durative aspect can be backgrounded, and other more salient events occur, which are “forgrounded” or profiled.

As Table 2.1 shows, there also exist two main perfective aspect markers in Chinese: the actual aspect, marked by \(-\text{le}\), and the experiential aspect, marked by \(-\text{guo}\). The actual aspect stresses the realization of a bounded event, that is, the event (static or dynamic) has become actual or a reality at the reference time. The actual aspect marker \(-\text{le}\) is mainly attached to verbs located in the middle or the end of an utterance. Sometimes, it can also be used after adjectives to indicate the realization of a change of state, such
as tā liǎn hóng le (Her face flushed). When -le is attached to the verb, it emphasizes that the event mentioned has been actualized. For example, tā qù-le yǐng guó (He has gone to the U.K.) means the event of “his going to the U.K.” is realized as a fact at the default reference time, which coincides with the speaking time here. Whether he has arrived in the U.K. or is still staying there, or has come back is not of concern here. Thus, the actual-le just views the event as a whole without signaling its completion (Chu 1976; Tai 1984; Xiao & McEnery 2004). It is characterized by actuality, a holistic quality, and dynamicity (Dai 1997: 35; Xiao & McEnery 2004: 113).

The experiential aspect is quite similar to the actual aspect, both of which look at the event from the outside, as an entirety. However, the experiential aspect emphasizes the agents’ (physical or quasi-physical) experience of an event, that is, that the experience of such an event denoted by the utterance occurred or existed at least once before the reference point. For instance, tā qù-guo yǐng guó (He has been to the U.K.) indicates that “his going to the U.K.” is not only actualized, but also terminated before the reference time, which implies that he is not in the U.K. now. The experiential -guo indicates the completion of an event (Lü 1980; Shi 1992), and has similar usage to that of the perfect of experience in English to some extent (Xiao & McEnery 2004). It is characterized by experientiality, holism, and dynamicity (Dai 1997: 35; Xiao & McEnery 2004: 113).

Grammatical aspects are abstract and schematic constructions. When the speaker chooses which grammatical aspect to use to view/express an event, they are at least constrained by the intrinsic temporal quality of the verb, i.e., lexical aspect. Lexical aspect is relatively universal across languages. Vendler’s (1967) classification of lexical aspect types is the most influential and the most frequently used in aspect studies. It consists of the four types of States, Activities, Accomplishments, and Achievements according to the three binary temporal features [±DYNAMIC], [±DURATIVE] and [±TELIC] (cf. Shirai 2002: 456; Xiao & McEnery 2004: 39), as Table 2.2 shows. Based

---

4 It is also termed Aktionsart (e.g., Agrell 1908; Comrie 1976; Klein 1994), or situation aspect (Smith 1991, 1997; Xiao & McEnery 2004).
on this, Smith (1997) added a fifth one, Semelfactives.

States are verbs or verb predicates indicating durativity without dynamicity and inherent telicity. Activities are verbs or verb predicates that are dynamic and durative but not inherently telic. Accomplishments are verb predicates which have dynamic actions, temporal extent and logical endpoints. Achievements are verbs or verb predicates that indicate dynamic action with telicity but without inherent duration. Semelfactives are verbs expressing dynamic actions whose initial and final points overlap with each other without duration. Apparently, Vendler’s classification of verb types works at the lexical level, and “what he really did is to propose ontological categories” (Verkuyl 1993: 33).

Table 2.2: The main lexical verb types in English and Chinese

<table>
<thead>
<tr>
<th>Lexical aspect types</th>
<th>[±dyn]</th>
<th>[±dur]</th>
<th>[±telic]</th>
<th>English</th>
<th>Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td>States</td>
<td>—</td>
<td>+</td>
<td>—</td>
<td>love, know</td>
<td>ài, zhī dào</td>
</tr>
<tr>
<td>Activities</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>run, crawl</td>
<td>pǎo, pá</td>
</tr>
<tr>
<td>Accomplishments</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>run for a circle</td>
<td>pǎo yī quān</td>
</tr>
<tr>
<td>Achievements</td>
<td>+</td>
<td>—</td>
<td>+</td>
<td>arrive, win</td>
<td>dào dá, yíng</td>
</tr>
<tr>
<td>Semelfactives</td>
<td>+</td>
<td>—</td>
<td>—</td>
<td>knock, cough</td>
<td>qiāo, kēsōu</td>
</tr>
</tbody>
</table>

According to a perhaps surprising analogy of verbs with nouns in terms of three conceptual dimensions — boundedness, homogeneity and expansibility, Langacker (1982, 1987, 1991a, 2008a) categorized verbs into perfective and imperfective verbs. He pointed out in the footnotes that the perfective verbs subsume Vendler's activities, accomplishments, and achievements, and the imperfective verbs correspond to the states (Langacker 2008a:147). Langacker (2008a) emphasized that such a categorization of verb types is quite flexible. It can be influenced by the event’s participants (i.e., arguments), the nature of the described event, the viewing scope etc. (Langacker 2008a: 149-151). Croft (2012) developed a two-dimensional geometric
model to represent aspectual types at the sentential level. In his model, every event is represented in two dimensions, time (t) and qualitative states (q). He categorized the aspectual types/construal of sentences into STATES, including transitory state (e.g., *The door is open*), acquired permanent state (e.g., *The window is shattered*), inherent permanent state (e.g. *She is French*), and point state (e.g., *The sun is at its zenith*); ACHIEVEMENTS, involving reversible directed achievement (e.g., *The door opened*), irreversible directed achievement (e.g., *The window shattered*), and cyclic achievement (e.g., *The mouse squeaked*); ACTIVITIES, subdivided into directed activities (e.g., *The soup cooled*), and undirected activities (e.g., *The girls chanted*); and ACCOMPLISHMENTS, including incremental accomplishments (e.g., *I ate an apple pancake*), and run-up achievements (i.e., nonincremental accomplishments) (e.g., *Harry repaired the computer*) (Croft 2012: 58-62).

Whatever the classification of aspectual types is, and whether it is at the lexical verb level or at the sentential level, the grammatical aspects are more compatible with certain types. As the only grammatical aspect marker in English, *be ...ing* is so powerful that it is compatible almost with any lexical aspect except the verbs with permanent state such as *be French*. Consequently, the progressive construction in English is polysemous, having various kinds of meaning when co-occurring with different lexical aspects. This will be reviewed in detail in the following section. The Chinese progressive *zài* is dominantly used with activities but incompatible with states. And the durative *-zhe* is more likely to co-occur with activities and then with states. The actual *-le* is able to co-occur with any lexical aspect type, but it prefers achievements and accomplishments. The experiential *-guo* is also compatible with all kinds of lexical aspect type, but it is preferred for accomplishments and activities (Xiao & McEnery 2004).

### 2.1.2 Cognitive account of the meanings of grammatical aspectual constructions

In view of the varied uses of grammatical aspects in English and Chinese, I give a cognitive account of their semantic networks, illustrating them with examples from the
corpora employed in this dissertation.

2.1.2.1 English progressive aspectual meanings

With the core status of the progressive aspect in the English grammatical system, its semantic expansions have drawn great attention from linguists. They have tried to explain the meanings of the progressive within different semantic frameworks. For instance, within the Neo-Davidsonian event semantics framework, some scholars treated the progressive as a predicate operator [HOLD], indicating that the event holds for some time. Meanwhile, they treated non-progressive aspect (i.e., simple past as perfective aspect in their original studies) as a predicate operator [CUL], expressing that the event culminates at some time (e.g., Dowty 1991; Parsons 1990; Szabó 2008). Under the interval-semantic framework, Dowty (1977, 1979) proposed treating the progressive as a modal operator designating “deontic necessity”. In doing so, he aimed at “a uniform semantic analysis of the progressive in contemporary English” in terms of truth conditions (Dowty 1975: 585). However, he came across the famous “imperfective paradox”, that is, when the progressive aspect co-occurs with accomplishment predicates, it cannot entail the attainment of the event’s inherent endpoint. For example, *He was drawing a circle* does not necessarily generate the entailment of *He drew a circle*. A circle may be not attained if the drawer lacked ink or was interrupted to do something else. Thus, the event “he was drawing a circle” can be true only if it continues without any interruption in the “inertia world” (Dowty 1979).

Extending Kratzer’s (1977, 1981, 1991) Framework for Modals to the progressive, Mayerhofer (2014) defended the modal account of the progressive by treating the progressive as a modal operator designating “intentionality”. By doing so, Mayerhofer (2014) effectively resolved “the imperfective paradox” and verified the feasibility of the modal account of the progressive in natural language. Though these studies have deepened our understanding of the English progressive, they have not provided a comprehensively uniform account of the English progressive yet, including both the aspectual-temporal uses and the non-aspectual-temporal uses of the progressive in
modern English. However, the cognitive account has filled the gap, proposing a unified account of the semantics of the English progressive.

The English progressive has been extending from the progressivity construal, to temporary state, habituality and even to the “idiosyncratic” or “special” uses such as futurity and emotion (e.g., Comrie 1976; De Wit & Brisard 2014; Declerck 1991; Declerck et al. 2006; Langacker 1991a, 2008a; Leech et al. 2009; Lyons 1982; Rompaey 2014). Consequently, different scholars proposed different prototypical meanings of the English progressive, such as progressivity (e.g., Comrie 1976), susceptibility to change (William 2002), and epistemic contingency (De Wit & Brisard 2014). Based on the cognitive grammar framework, De Wit & Brisard (2014) effectively established a semantic network for the English present progressive, illustrating how the peripheral meanings derived from and are connected with the prototypical meaning, i.e., epistemic contingency.

Before I review the cognitive account of each progressive meaning derived from the basic meaning as “epistemic contingency”, I would like to introduce the Dynamic Evolutionary Model of Reality as shown in Fig. 2.1, which is adapted from Langacker (2008a). We are living in a world where countless events are happening in every moment. “The history of what has happened in the world” is reality (Langacker 1991a: 276; 2008a: 297), which develops and updates as time flows.

![Fig. 2.1 Dynamic Evolutionary Model of Reality (adapted from Langacker 2008a)](image-url)
What occurs at the moment is the current reality, shown as the leading face of the “growing” cylinder in Figure 2.1 (Langacker 2008a: 301). As a conceptualizer (C), we try to understand the world. What we accept as real and process as “established knowledge” constitutes our conceived reality (Re) (Langacker 2008a: 301), which is inherent in “evolutionary momentum” to drive forward (Langacker 2008a: 306). It grows larger and larger with the development of the conceptualizer’s immediate reality, which is the latest knowledge s/he establishes from moment to moment. The conceptualizer can map some conceived reality onto what has the potential to be real, which constitutes the projected reality. The potential reality is the unknown out of the conceptualizer’s control at the immediate moment. The cognition of the dynamic evolution of reality is also well manifested in the linguistic grounding system, such as tense and modality (Langacker 2008a). For example, the present tense usually describes events occurring in the immediate reality, which is epistemically immediate (i.e. close) to the C at the time of speaking. However, the simple past often designates events occurring in the conceived reality, which is epistemically non-immediate (further away from) to the C. The English progressive construes an event as contingent, that is, epistemically non-immediate to the C. Thus, “its presence or actualization is not seen as necessary and could not have been particularly expected or predicted” at the reference point (De Wit & Brisard 2014: 62). I will review the derived meanings of the English progressive as follows. The schematic semantic structures shown here were based on Langacker’s (2008a) diagrams. The progressivity construal of the English progressive results from either singular events or a series of events.

First, CONTINUOUS ONGOINGNESS is the most entrenched meaning of the progressive construction (De Wit & Brisard 2014; Rompaey 2014). As Fig. 2.2 shows, the progressive construction profiles the internal portion of events as unbounded, homogenous, and state-like in the immediate scope (IS), disregarding the endpoints in the maximal scope (MS) (Langacker 2008a:155; Smith 1997: 73-75). The ongoing events imposed by the progressive construction are parts of the conceptualizer’s immediate reality. Under such circumstances, a singular event is conceptualized as uninterruptedly and continually ongoing at the time of speaking, as in (1a), or other
vantage time, as in (1b) and (1c) (Declerck et al., 2006: 33). The progressive construction can encode events that are “evaluated relative to a larger interval of time, or those expressing the notion of an event viewed as going on at a single point in time” (Bertinetto et al. 2000: 527). The former is referred to as “durative progressives”, as in (1b), and the latter as “focalized progressives”, as in (1c) (Bertinetto, 2000; Bertinetto et al. 2000).

(1) a. I’m using my hands, it’s good to explain. (Ellen Show # 344)
   b. [Yesterday, during my sleep], Ann was playing for two hours all by herself. (from Bertinetto 2000: 571)
   c. When John came, Ann was still working. (from Bertinetto 2000: 564-565)

Fig. 2.2 Continuous ongoingness

Fig. 2.3 Temporary state

Second, when the progressive construction is imposed on state verbs, it leads to the TEMPORARY STATE construal. The incompatibility of the progressive construction with stative verbs in English has long been taken as a defining criterion (e.g., Biber et al., 1999; Croft, 2012; Langacker 2008a; Quirk et al. 1985; Vendler, 1967), because stative verbs denote static (i.e., non-dynamic), internally durative and homogeneous situations. Thus, the use of the progressive on such an occasion would be “redundant” or “superfluous” (Langacker 2011: 57). However, when the speakers choose the progressive aspect to construe such an inherently unbounded situation, they impose implicit temporal boundaries on it in the MS, as shown in Fig. 2.3, and profile its temporariness. Under such circumstances, the progressive has the capacity to
activate an inherently stative verb, making it attain the sense of dynamicity such as in (2a) - (2c) (Michaelis 2004; Ziegeler 1999: 55). In addition, the homogenous temporary state in the progressive results from “the persistence of the same activity” with limited duration (De Wit & Brisard 2014: 78), such as in (2d) and (2e). The conceptualizer’s conceived reality, for instance – the kids not behaving themselves, or sitting as only part of one’s everyday actions – leads the conceptualizer to be aware of the contingency of such events in their immediate reality.

(2) a. I think all three of us are **having** a good time. (Ellen Show # 140)

b. My kids **are being** crazy. (Ellen Show # 26)

c. I **am loving** doing the season. (Ellen Show # 141)

d. I **was wearing** a blonde, curly wig I have named Yolanda. (Ellen Show # 103)

e. But now I **am sitting**. (Ellen Show # 15)

Third, the REPETITION construal of the English progressive arises from the iteration of a series of rather short events in rapid succession (Brinton 1988: 9, 41-42; Comrie 1976: 42; Declerck 1991: 162; Langacker 1991a: 208-9, 2008a: 156). As Fig. 2.4 shows, the short dynamic event represented by dots occurs sequentially in the actual world as part of current reality, but the temporariness and continuation urged the conceptualizer to view it in summary fashion in their mind as a virtual entirety. Consequently, the internal portion of such virtual action in its entirety is profiled in the immediate scope as a part of the immediate reality. A punctual event is bounded by the almost overlapping onset and offset, so it is too brief to be viewed in the normal fashion. Thus, a “semelfactive” or “non-repetitive” punctual event is acceptable with the progressive construction only when it is used as a “frozen action” (Declerck 1991: 162) or in reference to “an after image” (Langacker 2008a: 156). As in (3a) and (3b), *He is blinking/winking/coughing/knocking at the door* can only be appropriate in the context where a single blink can be perceived in a video or a picture. Otherwise, it can only be understood when a series of blinks/winks/coughs/knockings are involved, which are
“construed as constituting a single overall event of bounded duration” (Langacker 2008a: 156). Furthermore, the reoccurrence of an event, such as watching the same news report in (3c), is not generalized enough to be construed as a temporary habit but as repetition, which is strengthened by the adverbial phrase over and over and over again.

(3) a. He is blinking/winking/coughing. (Langacker 2008a: 156)

b. It is, because if you are at home, they may be knocking at your door. (Ellen Show # 214)

c. I have been watching this one news report over and over and over again. (Ellen Show # 119)

Fourth, HABITUALITY is another aspecto-temporal interpretation of the progressive construction. In English, structural habituals are supposed to be marked by the simple present tense, such as in The earth moves round the sun. However, when a series of actions occur repeatedly over an extended period of time, the progressive construction is often used to construe such “temporal habits” (De Wit & Brisard 2014: 81; Declerck et al. 2006: 34). As shown in Fig. 2.5, every single occurrence of the same durative action is an instance of a temporary habit in the actual world as a part of the conceived reality to the conceptualizer, which may be profiled in the immediate scope but not necessarily as being actually performed at the time of speaking in the immediate
Temporality in Language and Mind: A Multimodal Approach to Grammatical Aspect in Event Conceptualization

reality. Instead, it is stored as an entry in the speaker’s structural knowledge, being activated in the process of mental simulation. An example is in (4a), where the listener is obviously not actually flying an airplane while he is sitting there for the interview. Rather, as a pilot, he has been flying at regular times lately. In addition, the present perfect progressive usually invokes the habituality interpretation, as in (4b), where doing boxing and training are also not ongoing at the speech time, but occur repetitively over an extended period in the man’s life. Such habits have not completely fixed, so they are contingent by nature. Thus, the implicit boundaries are also salient in the maximal scope to background the contingency of such habits.

(4) a. You are a good pilot and you do have your license, and you are flying again. And I am sure Callista is thrilled about that (Ellen Show # 959).

b. Ellen: You are getting stronger and stronger and -- what are you doing now?

   Guest: I have been doing a lot of boxing. Terry Norris, the former welterweight champion, has been training me (Ellen Show # 86).

Certain non-aspecto-temporal uses5 have emerged in the English progressive. Such uses no longer bear an explicit relationship with the progressive meaning. Instead, they represent modal meanings (Huddleston & Pullum 2002).

First, the most familiar non-aspectual use of the English progressive lies in its reference to future situations, which are “scheduled” or “already-decided”, to denote futurity (Biber et al. 1999: 471; Huddleston & Pullum 2002; Langacker 2001, 2011; Leech et al. 2009: 131). As Fig. 2.6 demonstrates, the event profiled in the immediate scope is to be actualized in the future rather than the present time. It typically indicates the conceptualizer’s confident PREDICTION of possible occurrence in the projected reality in the potential world from her or his conceived reality. Consequently, the event

5 It is also termed “non-normal uses” (Lyons 1982), “the experiential progressive” (Wright 1995), “modal” (i.e. subjective) (e.g., Kellie 2004; Wright 1994), “not solely aspectual functions” (Smitterberg 2005), or “special uses” (Leech et al. 2009), of the progressive in the literature.
is liable to human fallacy and thus contingent (De Wit & Brisard 2014: 74). It does not refer directly to the actual event. Instead, it recasts a planned future occurrence on a conceptual plan or schedule, either with explicit adverbial phrases, as in (5a), or inferable from the context, as in (5b).

(5) a. Ellen: Are you staying here now?
   Osborn: no, eventually I am leaving California, too bad, at 11:00 a.m., not at 11:00 a.m., at 11:00 p.m. (Ellen Show # 347)
   b. Ellen: So Christmas is coming up. What do you all want for Christmas?
      (Ellen Show # 867)

Fig. 2.6 Futurity

Second, the other well-documented non-aspectual use of the progressive is to express epistemic meanings with emotional connotations, that is, meaning of emotivity (Rompaey 2014). Instead of representing the temporal contour of a situation, the modal use of the progressive incorporates “a high degree of subjective expression of the speaker/writer’s attitude or evaluation of the situation” (Leech et al. 2009: 131). In order to make a statement, wish or question more tentative, speakers tend to make use of the “be...-ing” construction as in example (6a). Furthermore, the emotional overtones of irritation or indignation can be interpreted from the progressive as in example (6b). The
affective or emotive use of the progressive is assumed to frequently co-occur with an adverbial such as *always, just, actually, continually, at all times, forever or constantly*, which denote continuous occurrence/duration (Leech et al. 2009: 134). In addition, the progressive also has the “interpretative” use as in (6c), which can be employed to give “the speaker’s interpretation of some behavior that somebody is engaged in” (Leech et al. 2009: 131). It is used to “ evoke or re-classify a situation that has occurred before or that occurs regularly” (De Wit & Brisard 2014: 85).

(6) a. You **are wondering** who I am. (Ellen Show # 696)

b. There are a lot of kids that are not into exercise and they **are just playing** video games and sitting on the couch and not doing anything. (Ellen Show # 483)

c. I said, your lollipop is on your face. You **are lying** to me. (Ellen Show # 597)

To sum up, the English progressive is a polysemous construction. It reflects the epistemic contingency of an event. The most entrenched aspectual meaning of the English progressive is CONTINUOUS ONGOINGNESS, encoding the current reality around the conceptualizer. When imposing on a static event, the progressive profiles TEMPORARY STATE. It is construed as REPETITION with the iteration of a series of short events. The regular occurrence of an event over time results in the temporal HABITUALITY interpretation. The four readings of the English progressive are mainly based on the conceptualizer’s conceived reality. In addition to the aspect-temporal uses, the English progressive has developed the modal uses. It denotes FUTURITY, encoding the projected reality according to the conceptualizer’s or her/his known schedule. EMOTIVITY arises from the English progressive with the conceptualizer’s subjective evaluation of an event.

2.1.2.2 Chinese grammatical aspectual meanings

The Chinese imperfective aspect also imposes an internal perspective on events. Both
the progressive construction (zài V) and the durative construction (V-zhe) profile the intermediate stage in the immediate scope but exclude the initial and endpoints of an event in the maximal scope as the English progressive, shown in Fig. 2.2. The meaning of the Chinese progressive aspect is also “CONTINUOUS ONGOINGNESS”. It has been widely acknowledged that the progressive zài denotes ongoingness of dynamic actions, but some young scholars stress that the progressive zài expresses ongoingness of dynamic events (e.g., Wang 2012; Yang 2013). An event consists of a series of complex actions, for example, the event of writing a paper includes reading literature, collecting data, writing, editing, etc. (Yang 2013: 58). When any action is carried out, you can say that S/he is writing a paper. This is also consistent with Smith’s idea that the progressive aspect typically represents the “stage property” of a situation with successive stages (Smith 1997: 77). Therefore, the progressive aspect zài is incompatible with state verbs, which lack such successive stages.

There has been a long debate on the meaning of the Chinese durative aspect. Some scholars insisted that the durative aspect -zhe denotes DURATIVITY OF STATES (e.g., Dai 1991, 1997; Li & Thompson 1981; Smith 1991; Teng 1979; Yang 2013), while others proposed that it encodes not only DURATIVITY OF STATES but also ONGOINGNESS OF ACTION (e.g., Xiao & McEnery 2004; Wang 2011, 2012). When it refers to DURATIVITY OF STATES, the durative aspect can not only express TEMPORARY STATE like the English progressive aspect, but also RESULTATIVE STATE. As shown in Fig. 2.3, the former profiles the internal homogeneity resulting from the persistence of an action in the immediate scope with implicit boundaries in the maximal scope, like example (7a). However, the latter profiles the continuation of the resultant state in the immediate scope, as I propose in Fig. 2.7. As in example (7b), the durative existence of a picture on the wall is maintained from the state left by the action of hanging it.

(7) a. tā bāo-zhe nǚér
He hold-DUR daughter.

“He is holding her daughter (in his arms).”
b. qiáng shàng guà-zhe yī-fú huà
  wall on hang-DUR one-CLF picture
  “A picture hung on the wall.”

The Chinese actual aspect -le and the experiential -guo profile an event as a whole in the immediate scope, as shown in Fig. 2.8, encoding the PERFECTIVITY of events. The aspectual meaning of the actual aspect is the ACTUALIZATION of an event (Dai 1997; Xiao & McEnery 2004), which may but not necessarily indicate “completed action” (Chao 1968: 247; Henne et al. 1977; Lü 1980; Tiee 1986; Wang 2015; Zhu 1982). The experiential aspectual construction -guo prototypically represents EXPERIENTIALITY of an event (e.g., Dai 1997; Smith 1997; Xiao & McEnery 2004), which encodes “completed action” (e.g., Dai 1997; Smith 1997; Wang 2015). Thus, there are some overlaps between them and they are interchangeable on some occasions. The semantic differences between them are as follows. First, the experiential aspectual construction represents the existent experience, which has strong relevance to the past. Thus, it can seldom occur with future events. The actual aspect construction expresses the realization of an event, which can be used to refer to events in the past, present and future. Second, an event marked by the experiential aspect is not continuing at present, whereas that marked by the actual aspect may be continuing now, and may continue in the future. Third, the actual aspect construction can indicate some result from the event,
whereas the experiential aspect is not necessary for indicating some result. (Xiao & McEnery 2004).

The different meanings of the perfective vs. imperfective in Chinese and the progressive vs. non-progressive in English have been determined based on linguistic evidence. The question of their psychological reality has stimulated a lot of studies in the field of psychology, which I will review in the following section.

2.2 Psychological approach to grammatical aspect

The existing psychological studies on grammatical aspect mainly center on the effects of the perfective aspect and the imperfective aspect on mental simulation of described events or visual scenes. Mental simulation is the internal engagement of modality-specific brain systems to create or recreate non-present experiences (Barsalou 1999). Most of the studies focus on English and are biased towards the study of language comprehension; few involve other languages, like Chinese, and few consider language production.

2.2.1 Grammatical aspect in language comprehension

The research to date on grammatical aspect and mental simulation has centered on comparing the effects of progressive aspect sentences and non-progressive aspect (simple past tense or perfect) sentences on language comprehension.

Earlier research on aspect in language comprehension has been dedicated to investigating how aspectral cues shaped the construction of mental simulation (i.e., situation models in the original studies). Magliano & Schleich (2000) pioneered the research on the relationship between grammatical aspect (termed verb aspect in the original study) and mental simulation. They found that readers interpreted the imperfective/past-progressive aspect-framed action as ongoing, and the non-progressive (simple past) aspect-framed action as completed when responding to questions about the completion status of the events embedded in stories. Furthermore, they employed probe-word methodologies to investigate the effects of grammatical
aspect on working memory, in which the participants were asked to decide whether the verb phrase occurred in one of the prior progressive or non-progressive sentences s/he read. The results showed that the action information activated in the progressive sentences could be maintained longer than that in the non-progressive sentences in working memory. However, the probe-word methodologies have been questioned in terms of their bias for activation of progressive events, as it is easier for the reader to probe for action and entities within the progressive aspect, which facilitates access to such internal structure, but not for those within the non-progressive aspect, which shuts them off (e.g., Madden & Zwaan 2003). Thus, some psychologists have used the picture presentation method, in which the participants were required to choose which picture representing an event at its intermediate or completed stage, matched the progressive or non-progressive sentences they read. However, such experiments produced some inconsistent results. Madden & Zwaan (2003) found that participants responded to completed-state pictures faster than to ongoing-state pictures after listening to non-progressive (i.e., simple past) sentences, but showed no latency differences after reading past progressive sentences. In order to investigate exactly how grammatical aspects influence the perceptual simulations in language comprehension, Madden and Therriault (2009) employed the replaced-word picture presentation method to test the “use effect” stimulated by the grammatical aspect cues. “Use effect” means that the related instruments in use should be simulated while the reader is reading about an ongoing event, and vice versa. For example, when reading about walking in the rain, the readers are more likely to have the image of an open umbrella rather than a closed one. In the experiments, the participants were asked to read imperfective (past progressive) and non-progressive (past perfect) sentences, word by word, with the critical instrument word replaced by a picture of that instrument in use or not in use. Then the authors tracked the use effect during integration of the picture into the sentences, during the words following the picture, and the sensibility judgment after the sentence. (Madden & Therriault 2009: 1296). The results verified the use effect that participants responded faster, i.e., more quickly processed the pictures of objects in use online (i.e., during the picture integration and following the integration) and offline (i.e.,
sensibility judgment after the sentence) than those not in use, and that the objects in-use facilitation effect only held for the situations in progressive sentences rather than in the non-progressive sentences. The above studies suggest that the progressive aspectual cue was likely to activate the language comprehenders’ mental simulation of the internal structures of events, and possibly the resultant state of events, whereas the non-progressive aspect only stimulated the language comprehenders’ mental simulation of the resultant state of events.

A series of behavioral experiments concerning the effects of grammatical aspect on mental simulation showed that there was an increased action-sentence compatibility effect with progressive sentences, but not with non-progressive (i.e., present perfect) sentences (Bergen & Wheeler 2010). The action-sentence compatibility effect (ACE for short, Glenberg & Kashak 2002) refers to faster responses in sentence judgment tasks when the language comprehender’s hand movement direction is the same as the direction of motion described in a sentence. For instance, it is faster for the participants to move their arm away from the body to push the YES-is-far button when seeing the sentence *John is closing the drawer* – an action that would involve moving one’s hand outward. Bergen & Wheeler (2010) showed that the progressive sentences rather than the “perfect” sentences produced ACE. Furthermore, Liu & Bergen (2016) found an increased Location-sentence compatibility effect (LCE) in understanding concrete progressive aspect-marked sentences rather than the abstract progressive aspect-marked sentences or the non-progressive (i.e., present perfect) sentences. LCE is a phenomenon whereby participants are faster in responding during a sentence judgment task when the location is fixed to make that judgment compatible with the implied meaning of the sentence. For example, the YES-IS-FAR button on the keyboard indicates that the sentence *Kimberly is hanging up the phone* makes sense as a compatible location and sentence combination. Such a LCE leads to a quicker button pressing response from the participants. The authors inferred that the language comprehenders did not only mentally simulate action but also locations of events while reading the concrete progressive aspect-marked sentences. The authors explained that the reason for the lack of compatibility effect of the abstract sentences was that they were harder to process.
than the concrete ones in their experiments. Therefore, the picture-event match experiments and the behavioral experiments implied that the progressive aspect has the potential to profile not only the internal structure but also the final physical state of an event. This challenges the linguistic proposition that the progressive aspect only profiles the internal temporal structure of an event.

Anderson and her colleagues carried out a series of experiments by using computer-mouse tracking (Spivey et al. 2005) to investigate the role that grammatical aspect plays in understanding everyday motion events (Anderson et al. 2008; Anderson et al. 2010; Anderson et al. 2013). In the experiments, the participants were required to place a character in a scene to match the description of motion events in the progressive aspect (past progressive) or non-progressive aspect (simple past) utterances they heard immediately. It was shown that the participants moved the mouse more slowly along the path and dropped the character further away from the end region of the path after listening to a past progressive sentence than to a simple past sentence (Anderson et al. 2008; Anderson et al. 2010). Furthermore, they found that the online processing of motion events was influenced by the interaction effects of grammatical aspect and contextual factors, like settings (difficulty or easy terrain description) where motion events were happening or happened, and temporal distance (recent past or distant past temporal description) when motion events were happening or happened. The contextual factors modulated the motor responses to past progressive more than to simple past: faster movement and finishing the task closer to the goal resulted from an easy terrain description rather than a tough terrain description with past progressive (Anderson et al. 2010); smoother and faster computer mouse movements were found in the matched conditions (past progressive with recent past adverbials, and simple past with distant past adverbials) (Anderson et al. 2013). Thus, these studies provided further evidence that progressive aspect stimulates the comprehenders to focus more on the internal details of motion events, which can be strengthened by the easy environment and recent past temporal description, whereas the non-progressive (simple past) aspect induces an emphasis on the final state of an event, which is less likely to be influenced by the contextual factors.
Recent eye-tracking experiments have also shown that grammatical aspect influences event conceptualization and distribution of attention in processing motion events. Huette et al. (2012) used the blank visual world approach, where participants listened to short descriptions of events in past progressive or non-progressive (i.e. simple past) sentences while looking ahead at a blank screen without visual input. The results showed that participants produced fewer eye movements and longer fixations on the blank screen after listening to events described in past progressive sentences than to those in non-progressive sentences. von Stutterheim et al. (2012) investigated the role of grammatical aspect systems that different languages provide in event construal when talking about motion events via a dynamic video clip stimulus. They found that speakers of languages that do not have grammaticalized aspect of imperfective or progressive aspect with verbs of motion (Czech, Dutch and German) are more likely to take a holistic view, and have more and longer fixations on the endpoint region of motion events than those with imperfective/progressive aspect with motion verbs (English, Arabic, Russian and Spanish). Thus, the grammatical repertoire provided by the native language directly influenced their native speakers’ motion event perception. This is also verified by neuropsychological studies (e.g., Flecken et al. 2015).

The studies mentioned above verified that it is psychologically plausible for the grammatical aspectual contrast (i.e., imperfective/progressive vs. perfective/ non-progressive) to influence the comprehension of events. The psychological evidence is consistent with the linguistic evidence that the progressive aspect facilitates a language comprehender to zoom in on the internal details of an event, whereas the non-progressive aspect leads the language user to take a holistic view and focus on the endpoint of an event. However, some of the studies found that the progressive aspect can stimulate the listeners’ focus not only on the intermediate part but also on the resulting state of an event, and that such contextual factors as abstract language, temporal distance and the environmental conditions in which the events occurred exhibited greater influence on the perceptual simulation of events in the progressive aspect sentences than in the non-progressive aspect sentences.
2.2.2 Grammatical aspect in language production

Compared with the studies on grammatical aspect in language comprehension, there has been less research on the role of grammatical aspect in producing everyday language.

Matlock and her colleagues carried out a series of experimental studies on how grammatical aspectual cues influence people’s conceptualization and description of events in language production. Matlock (2011) found that participants produced more multiple-action main clauses when completing sentences with progressive aspect adverbial frames (e.g., when John was walking to school, he felt sick and went home) than those with simple past adverbial frames (e.g. when John walked to school, he got a hamburger). Through asking the participants to produce descriptions about the videotaped car accidents in different aspectual-framed cues (what happened? vs. what was happening?), Matlock et al. (2012) further revealed that the progressive aspect cue facilitated the speakers to produce more motion verbs and even more iconic gestures than the non-progressive cue did. The results suggest that the imperfective/progressive framing led people to focus more on action details in reporting the past events.

Being one of the most important grammatical devices, grammatical aspect has been gaining more and more attention in research on co-speech gestures, which, at least according to the GSA framework, are regarded as reflecting mentally simulated actions (Hostetter & Alibali 2008).

Duncan (2002), as a pioneer in the study of aspect and gesture, qualitatively compared the gestural behaviors accompanying perfective and imperfective-aspect marked utterances in Chinese and English. She found that the gestures occurring in imperfective aspect-marked speech were longer temporally and more complex in semantic content encodings than those in the perfective aspect speech contexts in both languages, and that gestural agitation, such as complex motion strokes, iteration of gesture movements, etc., characterized the imperfective aspect-marked utterances representing either physical motion events or stative/mental events.

McNeill (2003) also qualitatively investigated the relationship between aspect in
speech and actual action in manipulation. He found that the aspectual perspective (perfectivity vs. imperfectivity) was effectively monitoring the temporal alignment between speech, actual action and lexical aspect (logical action types, in his words). In addition, the action the speakers produced accompanying the imperfective aspect also took longer than those in the perfective aspect context.

Recently, Parrill et al. (2013) carried out a strictly controlled experimental study to investigate the effects of English grammatical aspect on the gestural conceptualization of linguistically presented events. They further confirmed that gestures accompanying progressive utterances were longer-lasting and more complex than those with the non-progressive verb forms. However, such a situation only occurs when the described events were also originally presented in the progressive aspect.

Furthermore, the grammatical aspectual framing is also reported to influence people’s reasoning and decision-making in everyday practice. In a series of experiments carried out by Matlock and her colleagues, the participants were asked to read sentences framed with progressive aspect (e.g., Bob was planting pine trees along his driveway last week/John was painting houses last summer) or those framed with non-progressive aspect (i.e., simple past) (e.g., Bob planted pine trees along his driveway last week/John painted houses last summer), after which they had to estimate the length of the driveway or the number of houses that had been painted. The results showed that their estimates were much higher with the progressive sentences than those with the simple past (Matlock 2011; Matlock et al. 2012). Similarly, when reading election messages about a political candidate's past bad actions framed with past progressive (e.g., Last year, Mark was having an affair with his assistant and was taking money from a prominent constituent) or with simple past (e.g., Last year, Mark had an affair with his assistant and took money from a prominent constituent), the voters had higher estimates about the amount of hush money and were more likely to refuse to vote for the candidate whose political message was framed with past progressive (Fausey & Matlock 2010). Recent studies have also shown that the grammatical aspect even influences people’s solution rates for problems. For problems that depended on the action of the sentence, solution rates were higher when the insight problem was manipulated in the
imperfective aspect, and vice versa (Salomon et al. 2013). In addition, one’s legal judgment was also shown to be indirectly affected by grammatical aspect. The murder degree (first vs. second-degree murder) was more likely to be judged higher when the provocation and murder events were manipulated in the imperfective/progressive aspect than in the perfective aspect (i.e. simple past) (Sherrill et al. 2015).

To sum up, studies on the effects of grammatical aspect on language production or reasoning have made it clear that progressive aspectual framing results in more actions in the speaker’s mental simulation of events than non-progressive aspect framing does.

The psychological studies on the role of grammatical aspect in language comprehension and production have provided ample evidence for the psychological reality of the basic aspectual contrast (imperfective/progressive vs. perfective/non-progressive aspect) in language users’ mental simulation of events. The major contributions that these studies have made are that they effectively demonstrate that grammatical aspectual structures are meaningful, and that aspeo{4}ical meanings can modulate how people conceptualize events in processing and producing language, and eventually can influence people’s reasoning and decision-making related to everyday actions. However, the studies only focused on the basic meanings of progressive aspect and non-progressive aspect (CONTINUOUS ONGOINGNESS vs. COMPLETION). They are biased towards English and seldom mention non-European languages like Mandarin Chinese, which is much richer in its grammatical aspect system.

Many questions remain: To what extent can the results from the laboratory experiments be generalized to more complex natural conversations in human communication? What specific elements of events are simulated in speakers’ minds when they adopt different grammatical aspects? How do the other meanings of the grammatical aspects modulate speakers’ event conceptualizations? How does the L2 grammatical aspect system modulate the event construal of L2 speakers, in whose mind exist two different grammatical aspect systems? In order to explore these questions, I propose a multimodal approach to investigate what speakers are speaking and gesturing about events, using different aspectual frames within and across languages.
2.3 A multimodal approach to language use

When people speak with each other face to face, they do not only use speech, but also various kinds of bodily movements such as manual gestures, facial expression, eye gaze, etc. simultaneously to convey meaning and intentions. Such visible actions are a necessary part of our everyday communicative processes. As the core of human interaction, face-to-face communication is by nature multimodal (Cienki 2015, 2016; Müller et al. 2013: 1; Vigliocco et al. 2014). With the recognition of great significance of gestures in communication and the development of video-recording technologies, more and more scholars from different fields have adopted a multimodal approach to investigate language use. Co-speech gestures, which are produced spontaneously while talking, have been argued to constitute an integrated system with speech. Co-speech gestures are produced in close temporal, semantic and pragmatic alignment with speech (Kendon 2004; McNeill 1992, 2005). They share the burden of communication (Goldin-Meadow 2006). Thus, co-speech gestures constitute “an integral part of language” with speech (McNeill 1992, 2005; Perniss et al. 2015). Based on the previous studies, the present dissertation lays the foundation for studying multimodal construal of events via different grammatical aspect forms and different uses of gesture in English, Chinese, and English as FL.

2.3.1 A multimodal interaction model

As Fig. 2.9 shows, the MEANING of human interaction is what is at the heart of human communication. Instead of mapping onto things in the external world, meaning is dynamic in the minds of the speakers who are involved in the immediate communication (Langacker 2008a; Gärdenfors 2014). As with meaning, online thinking and established thought are also invisible in the speakers’ minds. The ways that people convey meanings and make minds meet are through the speech they can hear and the gestures they can see. While speaking and gesturing, meanings emerge dynamically in the interactions between the speakers during the process of thinking.
In the multimodal interaction model, speakers communicate their thoughts via two separate modalities, i.e., a linguistic component in the vocal modality, and an imagistic component in the visual modality (McNeill 2005, 2012; Marstaller & Burianová 2014). Such linguistic (i.e., language) and imagistic (i.e., gesture) forms present information as a closely integrated unit, but they are completely different in the way they express ideas by virtue of being produced in separate channels. Language (in terms of sound as speech or written symbolic form) is a “linear” code that is digital in its ontology, while gesture, with its multi-dimensional forms, is analog in nature (McNeill 1992:11). Thus, the former presents ideas in a discrete and segmental manner, whereas the latter packages them up in a global way. Speech and gesture are material carriers (Vygotsky 1986), which refer to the embodiment of meaning in a concrete enactment or material experience (McNeill 2005: 98). Meaning as conceptualization does not occur in a vacuum but is situated in a certain environment and is embodied by the visible manual gestures (Cienki 2013). As Streeck states, hand gestures link the abstract meanings with the manual experiences or interaction with the material world:
The meanings manufactured by human hand gestures are more than externalizations of mental structures and processes: they follow from the ways in which human hands in practical action make and experience the world (Streeck 2013: 74).

Iconicity plays a critical role in embodiment. The GSA framework (Hostetter & Alibali 2008, 2010) argues that representational gestures are outward manifestations of the action components involved in mental simulations of spatial and motor events (Hostetter 2014: 1469). In gesturing, the hands are acting out the meaning in the speaker’s thoughts or online thinking and are iconic representations of the objects, events, temporal and spatial relations conceptualized in everyday specific situations. While speaking, the motor images (performing or observing actions in the previous events) or visual image (perceiving the objects involved in the previous events) may be activated in the speaker’s mind. When the images are so strongly simulated that they surpass a gesture threshold, or a barrier to gesture, they will be realized as a gesture, making the embodied meaning visible in our hands.

Iconicity is also “one of the most powerful means of constructing gestural meaning” (Müller 1998: 321). A lot of gesture researchers have pointed out the ways that gestures portray meanings (e.g., Kendon 1997, 2004; Müller 1998, 2009, 2014). Gestures are iconically motivated by four basic modes of representation: acting, molding, drawing, and representing (Müller 1998, 2009, 2014). In the acting mode, the hands reenact the everyday instrumental actions. For instance, alternately moving the fists back and forth acts out what one might do when running. In the tracing mode, the hands draw out the shape of an object or the path of a motion. For example, moving the two extended index fingers upward, outward and downward in the opposite directions traces the shape of heart. In the molding mode, the hands mold the shape and/or size of an object. For instance, holding the hands with the curled index fingers and thumbs apart and other fingers curled into the palms in front of the speaker molds the round shape and size of a cake. In the representing mode, the hand itself represents an object. For example, a flat palm-up hand represents a piece of paper. These techniques are not mutually
exclusive, because sometimes a gesture incorporates more than one of them. For example, the gesture that moves the right extended index finger touching on the thumb in a zigzag way from left to right integrates both the technique of acting and that of representing. Whichever mode is used to depict meaning, any gesture is to some extent just schematic, extracting the core elements of objects, actions, events or spatial relations to metaphorically and metonymically map out meanings in a gestalt way.

In face-to-face communication, gestures have been found to be conducive for both the listener and the speaker. A lot of studies showed that visibility of the listener is one of the important contextual factors that influence the speaker’s gesture production. Generally speaking, speakers indeed make more gestures when the listeners are present than when they are absent (e.g., Alibali et al. 2001; Bavelas et al. 2008; Pine et al. 2010). However, a recent study revealed that speakers become less sensitive to the visibility of the audience as the descriptions of objects or events are more about bodily actions (Hostetter 2014). Thus, visibility studies can only show that gestures are partly to help the listeners understand the speaker’s meanings. There is no doubt that gestures can establish common ground, like pointing, disambiguating meanings and even evoking different image schemas than speech alone (e.g., Cienki 2005; Kendon 2004; McNeill 2005). Maybe gestures can facilitate the listener’s processing not only of the immediate information but also help them predict the following utterances. As Wilson (2006: 211) said, “the body schema acts as a prediction device”. When the listener perceives the initial stage of a gesture with which they are familiar, their brains generate imitative motor programs (Streeck 2013: 82). Neuroscientists also proposed that there exist putative mirror neurons in humans; groups of neurons in the primate brain fire not only when an action is performed, but also when it is perceived (Iacoboni et al. 2005; Kohler et al. 2002; Rizzolatti & Craighero 2004). Thus, producing and observing gestures help conduct one’s thoughts more directly and quickly, as perception and motor action form an integral or coupled system or feedback-loop (Merleau-Ponty 1962; Streeck 2013: 88).

When congenitally blind people communicate with blind or sighted people, they produced gestures which are as frequent and qualitatively similar as in communication
between ordinary hearing-seeing people (Iverson & Goldin-Meadow 1998). Such evidence, together with gesturing on the phone, illustrates that gestures are also cognitively motivated. They are of great significance to speakers in the following aspects. First, gestures can help the speaker to package spatial and motor information into units that fit verbal expression (Alibali et al. 2000; Kita 2000; Kita & Özyürek 2003; McNeill 1992). Second, gesturing can help speakers quickly retrieve lexical items from their mental lexicons (Butterworth & Hadar 1989; Krauss 1998; Krauss et al. 1996). Third, gestures can help the speaker decrease the cognitive load, sparing more resources to make communication go more smoothly or to do dual tasks (Goldin-Meadow et al. 2001; Kita & Davies 2009; Ping & Goldin-Meadow 2010). This dissertation will center around the cognitive functions of gestures, with some mention of their communicative functions. I am going to investigate how the speakers express their conceptualization of events in co-speech gestures.

Language and thought are inseparable. The relations between them have been an ongoing topic of research (see Lucy 2016 for a recent review). Relating language to worldview or habitual thought, the Linguistic Relativity Principle, also known as the Sapir-Whorf Hypothesis, states that the language we speak influences our thought or worldview (Whorf 1956). This idea has been criticized by some scholars, like Eric Lenneberg, who object that Whorf’s studies did not demonstrate the causality between a linguistic phenomenon and a mental phenomenon (Lenneberg 1953). Brown and Lenneberg (1954) moved the research on linguistic relativity ahead through work on color perception. They found that speakers who have more color categories at their disposal in their language are more capable of recognizing and remembering different colors than those in command of fewer color terms, and that the different color categorization in different languages indeed resulted in different memory patterns. Later, universalists such as Chomsky and Pinker, who believe that linguistic structures are mostly innate and that all languages share the same underlying structure, also criticized Whorf for his unclear clarification of how language influences thought, arguing against what they found to be cyclic reasoning (Chomsky 1957, 2006; Pinker 1994). Pinker (1994) even argued that thought is independent of language. However,
this idea is apparently too radical. In contrast, the relativists such as Lucy and the cognitive linguists such as Lakoff supported Whorf’s linguistic relativity theory. Lucy conducted a series of non-linguistic studies on the relations of language structure and cognition in English and Mayan Yucatec. The results showed that different languages lead to different ways of viewing reality (e.g., Lucy 1992a, b). Lakoff (1987) found that the different conceptual metaphors and linguistic categories employed by the speakers of different languages revealed how languages to some extent influence their speakers in conceptualizing the world. Language and thought are both multi-dimensional. Both of them are grounded in our experience. Thus, as Casasanto (2016) recently put forward, there is a continuum between different facts of experiential relativity, i.e., linguistic relativity, cultural relativity, and bodily relativity. He demonstrated that language, non-linguistic cultural practices and direct bodily experiences can all shape the way we think, feel and reason via some shared mechanisms.

Instead of relating the static entities, language and thought, Slobin turned to focus on the dynamic entities, thinking and speaking. Slobin (1987, 1996) proposed the thinking for speaking hypothesis, which gave rise to a lot of productive studies on linguistic relativity in the past two decades. It has become more operational and more direct to see the linguistic influence on cognition, as this “special kind of thinking that is intimately tied to language” is “carried out, on-line, in the process of speaking” (Slobin 1996: 75). While constructing verbal expressions on-line, the speakers’ thinking for speaking mechanism automatically guides their attention to the features of objects and events that are encodable in the linguistic repertoires. Thinking for speaking is “a special form of thought that is mobilized for communication” (Slobin 1987: 436; Slobin 1996: 76). However, the hypothesis has a risk of circular reasoning if we infer the different thinking-for-speaking patterns of languages via investigating speech alone. By studying the spontaneous gestures accompanying speech, gesture researchers have reformulated the hypothesis to cover thinking for speaking and gesturing (McNeill & Duncan 2000; Cienki & Müller 2008b). They propose that the differences in the grammatical options available across (and perhaps also within) languages influence the speakers’ thinking in the moment-by-moment processes of talking and gesturing. This
hypothesis has been widely confirmed by the studies in motion event conceptualization in terms of both lexicalization patterns for manner and path (e.g., Brown 2015; Brown & Gullberg 2008; Choi & Lanolf 2008; Slobin 1996; Slobin et al. 2014), and grammatical aspect (e.g., Athanasopoulos & Bylund 2013a; Bylund et al. 2013; Flecken 2011a; von Stutterheim et al. 2012).

In recent years, the investigation of thinking for speaking (and gesturing) has been extended to the study of bilingualism and second language acquisition (SLA). Slobin (1996) proposed that the thinking for speaking patterns developed in the process of acquiring a native language are “exceptionally resistant to restructuring in adult second language acquisition” (p. 89). However, many studies on whether learners can learn the thinking for speaking (and gesturing) patterns of the target language have shown that bilinguals or second language learners have the potential to adjust to the L2 thinking for speaking (and gesturing) patterns as long as their language proficiency is well developed and the length of exposure to the target language is substantial (e.g., Athanasopolous et al. 2015; Athanasopoulos & Albright 2016; Brown 2015; Brown & Gullberg 2008, 2011, 2012; Cadierno 2008, 2013; Choi & Lantolf 2008; Flecken et al. 2015; Stam 2010, 2015).

Most of the above studies were limited to the lexicalization patterns in event cognition across typologically different languages. There were few studies that focused on the grammatical structural differences with a focus on aspect marking across languages (Athanasopoulos & Bylund 2013; Athanasopolous et al. 2015; Athanasopoulos & Albright 2016; Flecken et al. 2015). Up to now, we have known little about how the grammatical options provided within a single language influence the speakers’ conceptualization of events in general, and the multimodal construal of events in particular. Based on the multimodal interaction model, this dissertation is dedicated to exploring the effects of the grammatical aspect options available in Chinese on native speakers’ thinking for speaking and gesturing. In addition, the present dissertation is interested in whether the native American English speakers’ gesturing can be correlated with the different meanings of English progressive aspect in the process of speaking, whether and to what extent L2 English learners of Chinese
can adjust to the L2 thinking for speaking and gesturing patterns.

2.3.2 A multimodal framework for studying grammatical aspect

With the increasing awareness of, and interest in, the multimodality of language use in human communication, more and more work on gesture has begun to incorporate grammar into multimodal interaction analyses. These studies are mainly devoted to revealing how grammar and gesture coalesce in multimodal interactions. Some scholars propose accommodating gesture into existing grammatical models. Steen & Turner (2013) proposed the idea of “multimodal construction grammar” by integrating gestures (in the broad sense) into the Construction Grammar Framework based on a multimodal corpus. With more of a focus on co-speech gestures, Kok and his colleagues assessed the degree to which gestures could have a place in Hengeveld & Mackenzie’s (2008) Functional Discourse Grammar Framework and Langacker’s (1987, 1991a) Cognitive Grammar Framework in view of the multifunctionality of gestures in interactions (Kok 2016; Kok & Cienki 2016; Kok et al. 2016). Their theoretical frameworks have systematically elaborated the feasibility and plausibility of using those two grammatical models to analyze gestures, and also pointed out the challenges in practice.

Other scholars mainly dealt with certain grammatical structures in a multimodal perspective to investigate how the speakers employ multimodal resources to realize grammatical conceptualizations. There have been two ways to deal with grammar and gesture in previous studies: One was starting from certain gestures, such as the cyclic gesture, the away-gesture family, and then the grammatical structures that they co-occurred with were explored (e.g., Bressem & Müller 2014; Kendon 2002, 2004; Ladewig 2011, 2014). The other was starting from certain grammatical constructions, and then exploring the features of gestures accompanying them. Such studies concerned negation (e.g., Andrén 2014; Harrison 2009, 2010, 2014), quoted speech (e.g., Park 2009; Parrill et al. 2016; Stec et al. 2015), references (e.g., Meyer & Gelman 2016), markedness (Ferré 2014), epistemic stance (Harrison 2009; Roseano et al. 2016), grammatical aspect (Duncan 1996, 2002; Harrison 2009; McNeill 2003; Parrill et al. 2016).
2013), and so on. The present dissertation will follow the latter route as the following framework presents (see Fig. 2.10).

According to Fig. 2.10 below, people experience events in reality via perception or action. As Langacker (2008a: 296) stated, “by their very nature, events do not endure — rather they occur, and typically their occurrence is quite brief” as compared with the long existence of objects. Thus, the experienced events are processed, stored, and left traces, or are even filtered out of our brains. When talking about the events in our mind, we verbalize them according to the constraints our linguistic systems impose on us and according to the communicative purposes we intend to achieve. When we describe the events we have in mind, we may use only speech. In this case, the mono-modal representation of events is formed. However, when we communicate with others in natural conversations, we usually cannot help making gestures in addition to speech. On such occasions that the multimodal representations of events arise.

![Fig. 2.10 A multimodal framework to study grammatical aspect](image-url)
The focus of this dissertation will be on the multimodal representations of events that involve grammatical aspect-markers in English and Chinese utterances and are also accompanied by gestures. To know how grammatical aspects and gesture integrate, we systematically examine the form and/or functions of the gestures accompanying the respective grammatical aspect-marked utterances. In terms of gesture forms, we will investigate the gesture handedness (right hand, left hand or both hands) and gesture qualities like iteration (repeated or single) and stroke features (dynamic or static). The first two are reported to reflect the differences between perfective and imperfective aspects in motion events (Duncan 2002; Parrill et al. 2013). The third one is expected to reflect the differences between the dynamicity of the ongoing events encoded by the progressive aspect in English and Chinese and the stativity of the states encoded by the durative aspect in Chinese or the progressive aspect in English.

In terms of the functions of gestures, I am concerned with whether the gesture is iconic or not. Iconic gestures are supposed to be more likely to accompany the imperfective aspect-marked utterances, which focus on the internal part of events. Iconic gestures can reflect the specific elements of events that are salient in the speaker’s mind while speaking. Specifically, an iconic gesture may be iconic to entities that participate in an event, action as the nucleus of an event, the process of an abstract event, the path of a motion event, or a combination of some of these elements. A canonical event, as the most typical kind of occurrence, is “a bounded, forceful event in which an agent acts upon a patient to induce a change of state” (Langacker 2008a: 357). As I examine all kinds of events in conversations, various kinds of participants beyond the agent and patient are involved in the events, such as mover in motion events, experiencer in mental events, instrument, etc. Thus, entities subsume all kinds of participants in events. Actions, including motions of concrete and metaphoric events as well as the schematic processes of inherently verbalized events, are considered as the most salient elements. The path of a motion event can be gestured alone, or integrated with manner in a single gesture. As Kok et al. (2016) argue, gestures tend to be multifunctional. A gesture can depict several semantic elements. Therefore, I can hypothesize that gestures encoding combinations of elements reflect that the speaker is
paying more attention to the details of events. I can also hypothesize that they are more likely to co-occur with imperfective aspect-marked utterances.

Such a multimodal approach to grammatical aspect use in conversation can visually show the speakers’ online thinking to some extent while speaking. What the salient elements of events are in their mind can be “read” easily, instead of being inferred from the linguistic or psychological data. If grammar is inherently meaningful and an integral part of our cognition, co-speech gestures accompanying the grammatical aspect-marked utterances should also reflect the aspectual distinctions and be influenced by different meanings encoded by even the same grammatical aspect construction. The multimodal framework of grammatical aspect is used here to guide the empirical studies on exploring speakers’ thinking for speaking and gesturing patterns in event conceptualization within a language and across languages in the following chapters.

3.1 Introduction

In face-to-face conversations, people subconsciously make gestures to express how an event unfolds through time, to depict what an agent or patient of an action looked like, and to model what states an action result in, etc. Such gestures bring the thought about an event into existence on a concrete plane (McNeill 2005: 99). The co-occurring speech and gesture are considered as an integrated “ensemble” to convey a “single thought” (Kendon 2004; McNeill, 1992, 2005). As a window to thought, gestures provide access to the speaker’s mental imagery or simulation of a particular experienced event, which can be vividly brought back into the immediate communicative context. However, to what extent the speaker goes into the details of an event is dependent on the temporal perspectives from which s/he chooses to construe it. Thus, aspect plays an important role in modulating the conceptualization of events (e.g., Madden & Zwaan 2003; Bergen & Wheeler 2010; Matlock et al. 2012; Duncan 2002; McNeill 2003; Parrill et al. 2013).

Comrie (1976) defines aspect as “different ways of viewing the internal temporal constituency of a situation” (p.3), which are realized grammatically and/or lexically across languages. In spite of being a tenseless language, Mandarin Chinese is rich in grammatical aspect, which has a series of particles to indicate the temporal contouring of events, such as -le to express the actualization of an event, -guò to express someone’s having experienced the event, and zài on the event’s progressive nature, -zhe on its durativeness. The former two represent perfective aspect, viewing the event as an indecomposable whole from an external perspective, whereas the latter two express
kinds of imperfective aspect, viewing the extension of an event from an internal perspective. That is, the conceptual information encoded by the perfective aspect takes an external viewpoint on an event and focusing its boundaries, while that encoded by the imperfective aspect takes an internal viewpoint on an event and defocuses its boundaries. Previous studies on grammatical aspect and gesture (or even instrumental, actual goal-directed actions in McNeill’s experiment) show that gestures (actions) accompanying imperfective aspect-marked utterances last longer temporally and display more complexity in form (Duncan 2002; McNeill 2003; Parrill et al. 2013).

Most previous studies of aspect and gesture focused on English aspects, used elicited narrative data, and narrowed their lens on physical motion events and concrete iconic gestures. Although Duncan (2002) took Mandarin Chinese into consideration, she conflated the different types of Chinese aspect-marked event construal. Based on these studies, I systematically explore how Mandarin Chinese speakers conceptualize events as expressed in their co-speech gestures accompanying grammatical aspect-marked utterances in natural conversations, which concern not only concrete but also abstract events. The following specific questions are addressed.

1) Are the gestures accompanying speech with imperfective aspect, including the progressive and durative aspect, more complex in form than those with perfective aspect, including the actualized and experiential aspect?

2) Is speech with imperfective aspect, including the progressive and durative aspect, accompanied by more iconic gestures than that with the perfective aspect, including the actualized and experiential aspect?

3.2 Methods

3.2.1 Multimodal corpus

The materials in this study are from Yang Lan One on One. Twenty shows consisting of deep interviews with native Chinese speakers were randomly selected; they were recorded between July 2014 and October 2015. They cover a wide range of topics
(about work, life experiences etc.), which provide relatively rich and natural Chinese conversation data. Each video, without the background introduction and advertisements, lasts for about 20 minutes. This multimodal corpus is about 8.5 hours long in total.

3.2.2 Participants

Thirty people (10 females) were involved in the interview corpus. Their mean age is 49.53 (SD = 10.82), ranging from 24 to 69. Most of them are actors or actresses. All of them are native Chinese speakers.

3. 2. 3 Coding procedures

3.2.3.1 Speech

There were 342 multimodal utterances marked by aspect markers, which were all affirmative and marked by just a single aspect particle. The negative utterances were all excluded. I also excluded utterances marked by the combination of two aspect markers such as chī guò le (eat-EXP-ACTL), as the categorical data can only be designated to one category for \( \text{Chi-square} \) analysis later. One coder coded the utterance aspect for the whole dataset, and a second coded the grammatical aspect in 50% of the dataset. Grammatical aspect was coded at the macro level for perfective and imperfective, and at the micro level for ACTL-le, EXP-guò, PROG zài and DUR-zhe; the inter-rater agreement was 98%.

The events described by each multimodal utterance were also coded as concrete events or abstract events (with metaphoric and abstract use of verbs). Concrete events refer to those that depict what physically happens in our daily life, whereas abstract events depict what happens without concrete action. The abstract event involves the metaphoric use of concrete verbs in construing abstract concepts (such as life attitude, beliefs, changes in quantity etc.), and inherently abstract verbs (such as the inner psychological state predicate xiāng (think), and the superordinate verb zuò (do)). Example (1a) is a concrete event, in which holding a machine is physically done and is
visible, while (1b) is an abstract event with the metaphoric use of a concrete verb "bàoyòu" (hold), in which the life attitude of indifference and calmness is abstract; “holding it” is an invisible mental state. Finally, (1c) is an abstract event with the use of an inherently abstract predicate "biànhuà" (change).

(1) a. Yìmòu yōngyuǎn bǎo zhe jīqi.
   Yimou forever hold DUR machine
   “Yimou was always holding the machine.”

b. tā yōngyuǎn bǎo yǒu zhe yǐ-zhǒng huā kāi huā luò
ever hold have DUR a-CLF flower open flower fall
dé tānrán hé cóngróng.
DE unperturbed and calm.
   “He always holds the life attitude of being unperturbed and calm, like flowers blossom and fall naturally.”

c. shìchǎng zài biànhuà.
   market PROG change.
   “The market is changing.”

### 3.2.3.2 Gesture

Previous studies reported that gesture forms such as handedness and movement quality (i.e. iteration) were related with aspects (Duncan 2002; Parrill et al. 2013). The handedness of gestures accompanying durative aspect tended to be more complex than those with perfective aspect and progressive aspect. The gestures accompanying the imperfective aspect were more likely to be iterated (or agitated) than those with the perfective aspect. Thus, those forms of gestures accompanying the aspect-marked utterances were coded. In terms of handedness, I considered whether the speaker used a single hand or both hands. In the case of both hands, they were coded as both hands mirroring each other in shape and position, or as both hands contrasting with each other in either one of these respects. Movement qualities here were coded as single when the
movement of a gesture stroke occurred only once, or iteration when the same movement pattern of a gesture stroke was repeated twice or more.

All the aspectual speech synchronous gestures were also coded according to gesture types adapted from McNeill (1992, 2005) and Kendon (2004). They were categorized into *concrete iconic* (hands represent the shape, size and/or action of the concrete events under description), *metaphoric iconic* (hands represent the shape, size and/or action of the abstract events), *deictic* (hands point at some physically present or imagined entities in the immediate context), and *pragmatic* (hands contribute to the pragmatics of the utterance meaning; they may function as performative, modal, or parsing gestures). The first two are iconic gestures, and the other two non-iconic. The iconic gestures focus more on the internal structure of events, such as being iconic for the properties of entities including agents or patients of events, iconic for the manner of action or path left by the action, or iconic for the ongoing or durative process of the events being described. However, the non-iconic gestures are more related to attention alignment via deictic gestures or discourse management via pragmatic gestures.

One coder coded the entire dataset, and a second coded 25% of the dataset. The inter-rater agreement was good: 100% for handedness, 97% for iteration, and 96% for gesture type.

### 3.2. 4 Data analysis

The data in this study are all categorical, so the Chi-square test was used to investigate the significance of differences between two variants. When the interaction effects were calculated, the *Loglinear Model* was used. The alpha level was set at 0.05.

### 3.3 Results

I first report the results of gesture forms of the aspect-marked utterances. Then the usage of gesture types accompanying different grammatical aspects will be thoroughly examined.
3.3.1 Gesture handedness

Generally speaking, a single-handed gesture is less salient than a 2-handed gesture. Furthermore, a 2-handed mirror gesture is also less complex than a gesture where the two hands have contrasting forms or motions. Now, I am going to see whether such complexities were embodied in gestures accompanying different aspect-marked utterances.

3.3.1.1 At the macro-level

As displayed in Table 3.1, the difference in handedness of gestures accompanying speech marked for perfective and imperfective aspectual distinctions is statistically non-significant ($\chi^2_{(2)} = 5.124, p > 0.05$). Both the perfective and imperfective aspect-marked utterances were dominantly accompanied by single-handed gestures, less frequently with gestures where the hands mirrored each other, and the least when the hands were making contrasting gestures, as presented in Fig. 3.1.

<table>
<thead>
<tr>
<th>Macro GA</th>
<th>Gesture handedness</th>
<th>Single hand</th>
<th>Hands mirror</th>
<th>Hands contrast</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Perfective</td>
<td></td>
<td>121</td>
<td>52.4</td>
<td>62</td>
<td>26.8</td>
</tr>
<tr>
<td>Imperfective</td>
<td></td>
<td>47</td>
<td>42.3</td>
<td>43</td>
<td>38.7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>168</td>
<td>49.1</td>
<td>105</td>
<td>30.7</td>
</tr>
</tbody>
</table>

Table 3.1: Frequencies of gesture handedness across macro grammatical aspect contexts
3.3.1.2 At the micro-level

According to Table 3.2, there was no significant difference of gesture handedness among the micro grammatical aspect-marked utterances ($\chi^2(6) = 6.567, p > 0.05$). The actualized aspect, the experiential aspect, the progressive aspect and the durative aspect marked speeches were all mainly accompanied by single-handed gestures, then by the hands as mirror gestures, and the least by hands as contrast gestures, as displayed in Fig. 3.2. The gestures in durative aspect contexts were even less often two-handed as contrasted both with those in the progressive aspect contexts and with those in the perfective (the actualized and experiential aspects) contexts.

**Table 3.2**: Frequencies of gesture handedness across micro grammatical aspects

<table>
<thead>
<tr>
<th>Micro GA</th>
<th>Gesture handedness</th>
<th>Single hand</th>
<th>Hands mirror</th>
<th>Hands contrast</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td></td>
<td></td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
</tr>
<tr>
<td>ACTL-le</td>
<td></td>
<td>108</td>
<td>54</td>
<td>46</td>
<td>208</td>
</tr>
<tr>
<td></td>
<td></td>
<td>51.9</td>
<td>26.0</td>
<td>22.1</td>
<td>60.8</td>
</tr>
<tr>
<td>EXP-guo</td>
<td></td>
<td>11</td>
<td>8</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47.8</td>
<td>34.8</td>
<td>17.4</td>
<td>6.7</td>
</tr>
<tr>
<td>PROG zai</td>
<td></td>
<td>26</td>
<td>21</td>
<td>12</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td></td>
<td>44.1</td>
<td>35.6</td>
<td>20.3</td>
<td>17.8</td>
</tr>
<tr>
<td>DUR-zhe</td>
<td></td>
<td>22</td>
<td>22</td>
<td>8</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42.3</td>
<td>42.3</td>
<td>15.4</td>
<td>15.2</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>167</td>
<td>105</td>
<td>70</td>
<td>342</td>
</tr>
<tr>
<td></td>
<td></td>
<td>48.8</td>
<td>30.7</td>
<td>20.5</td>
<td>100</td>
</tr>
</tbody>
</table>

The handedness of gestures produced together with the macro- and micro-
grammatical aspect utterances in this study did not show the preferences revealed by Duncan’s study: the perfective and progressive aspect contexts preferred 2-hand/mirror gestures, whereas the durative aspect contexts favored 2-hand/contrast gestures (Duncan 2002: 197). This may result from the different types of data investigated. Duncan only considered motion events and iconic gestures, but I examined natural conversation which included not only motion events but also other events like causative events and states, and different dimensions of gestures, like iconic, metaphoric, deictic and pragmatic.

![Fig. 3.2 Frequencies of gesture handedness across micro grammatical aspects](image)

3.3.2 Movement quality

3.3.2.1 At the macro-level

According to Table 3.3 and Fig. 3.3, although both the perfective and imperfective marked speeches predominantly co-occurred with the single movement gestures, the imperfective aspect-marked utterances were accompanied by twice as many more gestures with iteration than the perfective aspect-marked utterances were. The differences were reliably significant ($\chi^2_{(1)}=13.159$, $p<0.05$).
Table 3.3: Frequencies of movement quality of gestures across macro-grammatical aspects

<table>
<thead>
<tr>
<th>Macro GA</th>
<th>Single</th>
<th>Iteration</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Perfective</td>
<td>197</td>
<td>85.3</td>
<td>34</td>
</tr>
<tr>
<td>Imperfective</td>
<td>76</td>
<td>68.5</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>273</td>
<td>79.8</td>
<td>69</td>
</tr>
</tbody>
</table>

Fig. 3.3 Movement quality of gestures across macro-grammatical aspects

3.3.2.2 At the micro-level

As Table 3.4 shows, the differences of gestural iteration among the micro grammatical aspect-marked utterances were highly significant ($\chi^2(3) = 25.563$, $p < 0.05$). Like the actualized aspect- and experiential aspect-marked speech, the durative aspect-marked speech was mainly accompanied by gestures with a single occurrence, as displayed in Fig. 3.4. Thus, the differences mainly resulted from the gestural behavior of progressive aspect-marked speech, which was far more iterated than those in the other three aspect contexts.

To sum up, there was a significant difference between the frequencies for perfective and imperfective utterances in terms of gestural iteration, but almost half of the progressive aspect-marked utterances co-occurred with iterated gestures in contrast with the situation with the durative aspect-marked utterances.
Table 3.4: Frequencies of movement quality of gestures across micro grammatical aspect

<table>
<thead>
<tr>
<th>Micro GA</th>
<th>Movement Quality</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single</td>
<td>Iteration</td>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTL-le</td>
<td>178</td>
<td>30</td>
<td>14.4</td>
<td>208</td>
<td>60.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXP-guo</td>
<td>19</td>
<td>4</td>
<td>17.4</td>
<td>23</td>
<td>6.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PROG zai</td>
<td>33</td>
<td>26</td>
<td>44.1</td>
<td>59</td>
<td>17.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DUR-zhe</td>
<td>43</td>
<td>9</td>
<td>17.3</td>
<td>52</td>
<td>15.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>273</td>
<td>69</td>
<td>20.2</td>
<td>342</td>
<td>100</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 3.4 Movement quality of gestures across micro grammatical aspect

3.3.3 Gesture types/functions across grammatical aspects

3.3.3.1 At the macro-level

Table 3.5 shows that the multimodal utterances marked by the perfective aspect were twice as frequent as those with the imperfective aspect (231 vs. 111). This result is consistent with the case in English, in which the imperfective aspect (the progressive forms) is used less frequently than the perfective aspect (the non-progressive forms like the simple past tense) (e.g., Biber & Reppen 2002). This makes sense for natural conversational discourse, in which the speakers mostly narrate past events in a snapshot way.
According to Table 3.5 and Fig. 3.5, we can see that the utterances marked by the imperfective aspect were accompanied by significantly more iconic gestures than those in the perfective aspect contexts (69.4% vs. 51.1%). In the imperfective aspect speech contexts, twice as many iconic gestures were produced than non-iconic ones (69.4% vs. 30.6%), whereas in the perfective aspect-marked speech contexts, there was almost equal use of iconic and non-iconic gestures (51.1% vs. 48.9%). The Chi-square test of significance showed a reliable effect ($\chi^2(1) = 10.231$, p < 0.05).

Table 3.5: Frequencies of iconic gesture across macro grammatical aspect

<table>
<thead>
<tr>
<th>Macro GA</th>
<th>Gesture type</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Iconic</td>
<td>N</td>
<td>%</td>
<td>Non-iconic</td>
<td>N</td>
<td>%</td>
<td>Total</td>
<td>N</td>
</tr>
<tr>
<td>Perfective</td>
<td>Iconic</td>
<td>118</td>
<td>51.1</td>
<td>Non-iconic</td>
<td>113</td>
<td>48.9</td>
<td></td>
<td>231</td>
</tr>
<tr>
<td></td>
<td>Imperfective</td>
<td>77</td>
<td>69.4</td>
<td></td>
<td>34</td>
<td>30.6</td>
<td></td>
<td>111</td>
</tr>
<tr>
<td>Total</td>
<td>Iconic</td>
<td>195</td>
<td>57.0</td>
<td>Non-iconic</td>
<td>147</td>
<td>43.0</td>
<td></td>
<td>342</td>
</tr>
</tbody>
</table>

Fig.3.5 Frequency of iconic gesture across macro grammatical aspect

The results suggest that the imperfective event descriptions were accompanied by more complex gestures, which is consistent with the results of the previous studies (Duncan 2002; Parrill et al. 2013). The visible gestural evidence further proves that the internal structures of situations in linguistic imperfective contexts are more likely to be mentally stimulated.
### 3.3.3.2 At the micro-level

As Table 3.6 shows, for the multimodal events, the actualized aspect was dominantly employed by the Mandarin Chinese speakers, whereas the experiential aspect-marked multimodal utterances were the least frequent in natural conversations. The progressive aspect and durative aspect were almost equally used, both of which were twice more frequent than the experiential aspect and almost four times less frequent than the actualized aspect.

**Table 3.6: Frequencies of micro gesture type across micro grammatical aspect**

<table>
<thead>
<tr>
<th>Micro Grammatical aspect</th>
<th>Deictic N</th>
<th>Deictic %</th>
<th>ConIconic N</th>
<th>ConIconic %</th>
<th>MetaIconic N</th>
<th>MetaIconic %</th>
<th>Pragmatic N</th>
<th>Pragmatic %</th>
<th>Total N</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTL-le</td>
<td>30</td>
<td>14.4</td>
<td>30</td>
<td>14.4</td>
<td>80</td>
<td>38.5</td>
<td>68</td>
<td>32.7</td>
<td>208</td>
<td>60.8</td>
</tr>
<tr>
<td>EXP-guo</td>
<td>4</td>
<td>17.4</td>
<td>1</td>
<td>4.3</td>
<td>7</td>
<td>30.4</td>
<td>11</td>
<td>47.8</td>
<td>23</td>
<td>6.7</td>
</tr>
<tr>
<td>PROG zai</td>
<td>3</td>
<td>5.1</td>
<td>8</td>
<td>13.6</td>
<td>35</td>
<td>59.3</td>
<td>13</td>
<td>22.0</td>
<td>59</td>
<td>17.3</td>
</tr>
<tr>
<td>DUR-zhe</td>
<td>11</td>
<td>21.2</td>
<td>23</td>
<td>44.2</td>
<td>11</td>
<td>21.2</td>
<td>7</td>
<td>13.5</td>
<td>52</td>
<td>15.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>48</td>
<td>14.0</td>
<td>62</td>
<td>18.1</td>
<td>133</td>
<td>38.9</td>
<td>99</td>
<td>28.9</td>
<td>342</td>
<td>100</td>
</tr>
</tbody>
</table>

Notes: ConIconic=Concrete Iconic, MetaIconic=Metaphoric Iconic

![Fig. 3.6 Frequencies of micro gesture type across micro grammatical aspect](image)

As Fig. 3.6. and Table 3.6. show, the actualized aspect utterances were mainly accompanied by metaphoric iconic gestures and pragmatic gestures, the experiential aspect utterances were predominantly accompanied by pragmatic gestures, then by
metaphoric iconic gestures, and the least by concrete iconic gesture. Although the actualized aspect and experiential aspect all construe the events perfectly as an entirety, the gestural behaviors with them display some differences. If I conflate the concrete iconic and metaphoric iconic gestures into iconic gestures, and the deictic and pragmatic gestures into non-iconic gestures, you can see that, like the progressive and durative aspect-marked utterances, the actualized aspect-marked utterances were accompanied slightly more by iconic gestures than by non-iconic gestures (52.9% vs 47.1%), whereas the experiential aspect-marked utterances were accompanied more by non-iconic gestures than by iconic gestures (65.2% vs 34.7%). This result suggests that the speakers were also more prone to simulate the internal structure of events while using the actualized aspect. When the experiential aspect was used, pragmatic gestures were produced most frequently. The progressive aspect-marked utterances predominantly co-occur with metaphoric gestures, at a level three times that with durative aspect-marked utterances and twice as much as with those of actualized and experiential aspects-marked utterances (59.3% vs. 21.2% vs. 38.5% vs. 30.4%). However, the durative aspect-marked utterances were mainly accompanied by concrete iconic gestures, almost ten times more than utterances with the experiential aspect, and four times more than those with actualized aspect and progressive aspect (44.2% vs. 4.3% vs. 13.6% vs. 14.4%). The differences between them are statistically significant as the Chi-square test shows ($\chi^2 = 49.753$, p < 0.05). The results reveal an interesting phenomenon, namely, that the progressive aspect may be preferentially chosen to construe abstract events, but the progressive aspect may be more frequently employed to conceptualize concrete events. The next section will examine this phenomenon in detail.

To our knowledge, these results have been discovered here for the first time, which is of great significance for us to understand the micro differences of profiled focus in conceptualizing events between different aspect-framed contexts.
3.3.4 Aspect and gesture in concrete vs. abstract events

3.3.4.1 At the macro-level

As is shown in Table 3.7, there is no significant difference between the perfective and imperfective aspects ($\chi^2(1) = 1.497, p > 0.05$) when it comes to their use in characterizing concrete and abstract events. Both of them were employed more frequently with the abstract events than with the concrete events in conversational discourse, as indicated in Fig. 3.7. Therefore, when studying aspect and gesture use by speakers of Mandarin Chinese, we need to pay attention not only to concrete motion events as previous studies did, but also to abstract events, which are more frequently framed by the aspect particles.

<table>
<thead>
<tr>
<th>Macro GA</th>
<th>Event types</th>
<th>Concrete</th>
<th>Abstract</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Perfective</td>
<td>74</td>
<td>32.0</td>
<td>157</td>
<td>68.0</td>
</tr>
<tr>
<td>Imperfective</td>
<td>43</td>
<td>38.7</td>
<td>68</td>
<td>61.3</td>
</tr>
<tr>
<td>Total</td>
<td>117</td>
<td>34.2</td>
<td>225</td>
<td>65.8</td>
</tr>
</tbody>
</table>

Fig. 3.7 Macro grammatical aspect in concrete and abstract events
In order to know the relationship among the macro grammatical aspect, the event type, and the macro gesture types in multimodal contexts, I carried out a LogLinear Selection Analysis. The results showed that there was no interaction effect between macro grammatical aspect, event type and macro gesture type ($\chi^2_{(1)} = 1.199$, $p > 0.05$). But the interaction between macro grammatical aspect and macro gesture type was significant ($\chi^2_{(1)} = 8.578$, $p < 0.05$). It revealed a main effect of event type ($\chi^2_{(1)} = 34.696$, $p < 0.05$). Such results imply that event type influences not only the use of grammatical aspect, but also the native Chinese speakers’ gestural behaviors when they take aspectual views on events.

3.3.4.2 At the micro-level

According to Table 3.8, the actualized aspect, experiential aspect, and progressive aspect in particular were used more frequently with abstract events than with concrete events, whereas the durative aspect was employed more frequently with concrete events than to characterize abstract events in conversational discourse, which is clearly displayed in Fig. 3.8. The differences are also shown to be significant via the Chi-square test ($\chi^2_{(3)} = 12.978$, $p < 0.05$).

**Table 3.8: Distribution of micro grammatical aspects across event types**

<table>
<thead>
<tr>
<th>Event Type</th>
<th>ACTL-\textit{le}</th>
<th>EXP-\textit{guo}</th>
<th>PROG \textit{zai}</th>
<th>DUR-\textit{zhe}</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Concrete</td>
<td>67</td>
<td>32.2</td>
<td>8</td>
<td>34.8</td>
<td>15</td>
</tr>
<tr>
<td>Abstract</td>
<td>141</td>
<td>67.8</td>
<td>15</td>
<td>65.2</td>
<td>44</td>
</tr>
<tr>
<td>Total</td>
<td>208</td>
<td>60.8</td>
<td>23</td>
<td>6.7</td>
<td>59</td>
</tr>
</tbody>
</table>
In order to know the interaction effects concerning the micro grammatical aspect, the event type, and the micro gesture types in multimodal contexts, I also carried out a LogLinear Selection Analysis. The results showed that the interaction effect of micro grammatical aspect, event type and micro gesture type was not significant ($\chi^2(9) = 8.677$, $p > 0.05$). However, the interaction between micro grammatical aspect and micro gesture type was significant ($\chi^2(9) = 34.450$, $p < 0.05$), as it was between event type and micro gesture type ($\chi^2(3) = 111.308$, $p < 0.05$). Such results suggest that both micro grammatical aspect and event type influence the speakers’ co-speech gestural functions. Thus, it furthered our awareness of the influence of event type on the grammatical aspect usage and the speakers’ gestural performance.

3.4 Discussion

3.4.1 Grammatical aspect and gesture forms

In terms of the physiological form of gestures accompanying aspect-marked utterances, the study revealed that there is no significant relation between aspect and gesture handedness, but aspects at both macro and micro- levels and gesture iteration were found to be interrelated.
When Mandarin Chinese speakers construe events using grammatical aspects in speech, they did not differ in the use of handedness between the imperfective and perfective speech contexts, and among micro-grammatical aspect contexts. The use of one-handed or two-handed (symmetrical or asymmetrical) gestures may not be framed by the linguistic aspects but the contexts of their use. Maybe due to the principle of least effort (Zipf 1949), single-handed gestures dominated the utterances with different grammatical aspect markers. Two-handed mirror (i.e. symmetrical) gestures were mainly used in the contexts where (1) the original action was done by two hands simultaneously, as in example (2a) below; (2) where two events with the same patient occur at the same time, as in (2b), in which the first event was mostly marked by the durative aspect; (3) where a metaphorical event with implied two-handed action was involved, as in (2c).

In (2a), a famous actress was narrating how she practiced throwing paper fans simultaneously before performing this scene in a film. As the left picture from the original film segment in Fig. 3.9 (2a). shows, the actress grabbed two paper fans downward and threw them upward above her head, and grabbed them again when they fell down at the same time. Simulating the original action in the film, the actress made a grabbing downward gesture in the preparation phrase while saying nà-gè (that), then rotated and threw her two hands upward while uttering the topic shànzi (paper fan). The symmetrical hands as mirror gesture (an upward thrust) that accompanied the spoken reference to the shànzi (paper fan) was expressively the same as the verb liàn (practice), so it was held after the predicate liàn (practice) appeared.

\[ (2) \text{a. [nà-gè shànzi wǒ liàn le wǔ-gè yuè].} \]

That-CLF fan I practice PERF five-CLF month.

“I practiced (throwing) those paper fans simultaneously for five months.”
b. [wǒ ná zhe máojīn níng]
I hold DUR towel wring.
“I held the towel and wrung it.”

c. [[zuìhòu tā yǎngài le] [zhēnshì de shichāng de xiānxìăng]]
finally it cover ACTL real GEN market GEN phenomenon.
“Finally, it covered the real phenomenon of markets.”

In (2b), the actor made two fists as if grabbing a towel and twisted his two hands simultaneously and repeatedly when speaking ná zhe máojīn níng (hold the towel and wring it). The event of holding the towel and the event of twisting the towel were integrated in the symmetrical hands as a mirror gesture, as Fig. 3.9 (2b) shows.

In (2c) and Fig. 3.9 (2c), the speaker held his two palm-down hands up to the center in front of his neck as a preparation, then his two similar hands moved apart from the center to either side while uttering yǎngài le (covered), and then repeated it in a
diminishing way when speaking (zhēn)shì de (real); it was held through when the abstract patient (object of the verb) shichāng de xiànxì (the phenomenon of market) was mentioned. In this situation, we can see the metaphor REALITY OR TRUTH IS OBJECT. Covering it was acted out as if covering a table or something like that. Such symmetrical hands as mirror gestures occurred very frequently when metaphorical events were depicted.

The two-handed contrast (i.e., asymmetrical) gestures mainly occurred in the contexts where (1) the agent and goal of a motion event were profiled as in (3a), and (2) two different events with different patients happened simultaneously, one as the backgroundering event with durative aspect marker, and the other as a foregrounding one like (3b), and (3) when a reference point of the place where an event happened was profiled as in (3c).

In (3a), there was a long pre-stroke hold during the speaker’s self-correction as the left part in Fig. 3.10 (3a) shows: the left palm-towards-center open hand standing for the “gate of hell” was held statively on the left side as the goal of the motion event, and the right palm-down open hand represented the actors, which moved laterally to the left hand while uttering the aborted dì (earth). Then the fingers of the right hand were inserted between the fingers of the left hand one by one when speaking jinrù le (entered) as displayed in the right part in Fig. 3.10 (3a), expressing the path of the motion in speech; however, the gesture did not only reveal this, but also a very special manner of entering. Finally, the fingers of crossed hands curled down intensely associated with diyu zhī (hell’s) and were held to mén (gate), and this gesture complemented the speech by adding the information of “the immediate closure of the gate of hell after you enter it”. The semantic components of such a metaphorical motion event were integrated vividly and comprehensively into this asymmetrical two-hands as contrast gestures.

(3) a. [[ni jù hui dào jinrù* hàoxiàng di*] [jinrù le d] [iyù zhī mén]].
you then can reach enter seem earth enter ACTL hell GEN gate

“It seemed that you entered the gate of hell.”
In (3b) and Fig. 3.10 (3b), two simultaneous events with different patients were involved. When introducing the new topic by uttering *ránhòu wǒmen jiù* (then we), the left unbounded cupped hand as if holding a cup moved away from his body to the center, and the right palm-down clenched hand as if holding a round object moved downward
towards the left hand. Such an abrupt gesture preceded all the gesture-linked speech. He repeated and enlarged this gesture when the relevant information occurred: the thumb and the index came into contact, like the shape of the round mouth of a cup, and the other fingers stayed slightly curled when uttering ná zhe gè (held a) and were held until the zhǐ bēi (paper cup) appeared, which was accompanied by an additional gesture of the right palm-down clenched hand moving rapidly and directly towards the left hand. Then when the predicate of dào le hénduō xiāngbīn (poured a lot of champagne) was spoken, as a reference point, the posture of the left hand was maintained, and the right hand moved cyclically towards the left hand. These asymmetrical hands, contrasting in the gesture, clearly demonstrated that the two hands were involved in different actions at the same time in an event. The movement pattern distinguished pouring a lot of cups of champagne rather than pouring much champagne into the same cup.

In (3c) and Fig. 3.10 (3c), the actress narrated her experience of performing a scene for a film on a sea by driving a boat with the other antagonist. When they started, the sea was normal. However, when they reached the center of the sea, a sudden rising tide occurred. This was located by her left palm down hand as a reference point, where nearby they turned around abruptly as displayed in Fig. 3.10 (3c). The right palm-down open hand representing the boat moved laterally in a smooth way to the center, parallel to her neck, with the speech chābūdūō dōu kuài (almost) and held through the words dào zhè’ér de shíhòu (when reached here), then moved semi-cyclically and very fast to her left side, co-occurring with the speech jiù guăiwăn le (turned around).

The results also showed that the gestures accompanying imperfective aspect speech were more likely to be iterated than those in perfective aspect speech, but actually only the progressive aspect-marked utterances were significantly more frequently accompanied by gestures with iteration. This further proves that when taking the progressive aspect perspective to describe an event, the speakers were more prone to focus on the “ongoingness” and event-internal details of that event. As the previous behavioral studies showed, the progressive aspect resulted in more actions and longer duration of actions in a given period of time than the perfective aspect did (Fausey & Matlock 2011; Madden & Zwann 2003; Matlock 2011). The uses of progressive aspect
were more likely to lead to more gestural action with language production here than the perfective aspect and even the durative aspect. Events described with the progressive aspect were homogeneous, with each portion having the same features and occurring successively (Langacker 2008a). The above conceptual features can be argued to have motivated the repeated gestures with the progressive-aspect speech, such as in (4a) and (4b).

In (4a), the speaker narrated that his son worked very hard until 2 or 3 a.m. every night writing and drawing. You can see in Fig. 3.11 (4a) that his index finger touching upon the thumb of his right hand like the head of a pen moved horizontally from the center to the right as if writing three times while uttering dōu zài xiē a huà a (was writing and drawing).

In example (4b), when the speaker mentioned zhè-gè liàng-gè shī’ér (those two things), which referred to the industrial structure transformation and the aging problem, he held up his two palm-down open hands with extended fingers to the center from his legs, then he moved the two hands, one forward, and the other backward alternatively three times while uttering zài pāobù (are running), as Fig. 3.11 (4b) presents. The nucleus of such a metaphorical event was simulated as the concrete event of running.

(4) a. [tā dōu zài <...0.5s> dōu zài xiē a huà a].
   he all PROG all PROG write PRT draw PRT
   “He was writing or drawing all the time.”

b. [zhè-gè liàng-gè shī’ér zài pāobù]
   this-CLF two-CLF things PROG run.
   “These two things are running.”
The durative aspect also focused on the intermediate phase of an event, emphasizing the durativity of an ongoing event. Therefore, in most cases, where the durative aspect-marker co-occurs with a verb to describe an event, the speakers usually hold the gesture stroke rather than repeat it, as in (5a). The gesture accompanying this durative event preceded the speech. During the 0.5s pause before the speech, the gesture of “holding a rope” (the left palm-up fist in Fig. 3.12 (5a) came into being, and it was held until the ending of this utterance. Even when another subsequent event (looking down) was added, the “holding a rope” gesture was maintained rather than repeated.

He made another new gesture by pointing downward at the ground under the desk before him with a downward gaze upon uttering kàn zhe (was looking at). The deictic gesture was also held to the ending of this utterance. However, the gestures accompanying two simultaneous events in durative aspect speech contexts were always repeated as in the progressive aspect speech contexts like example (2b). On closer examination, we can find that the gesture for ná zhe máojīn (was holding the towel) (the two grabbing hands) held through the utterance after it came into existence, and the iteration for three times of the gesture resulted from the forgrounding event níng (máojīn) (wring the towel). This is also the case in example (3b). Furthermore, the repeated gestures accompanying the durative aspect utterances may also result from the implied meaning in the discourse as (5b) and Fig. 3.12 (5b). He was describing the situation where people think about whether it is better to deposit money or invest the money to buy a house. When speaking of ná zhe yí-gè fāngzì (are having a house), he
moved his two palm-down open hands up and down alternatively and repeatedly three times, as if weighing things on a balance scale.

(5) a. \[/[//...0.5s> wòmen sān-gè zài shàng-tou a ná zhe shéng]\]

we three-CLF at above PRT hold DUR rope

[kàn zhe dìxià nà-gē].

look DUR below that.

"We three stood at the top of hill, holding ropes and looking down at that below."

b. [ná zhe yì-gè fángzǐ].

hold DUR one-CLF house.

“(When they) are having a house.”

Fig. 3.12 (5a) Gesture with “hold-DUR rope”

Fig. 3.12 (5b) Gesture with “hold-DUR house”

The first use of durative aspect is typical, while the last two are not so common. Thus, durative aspect speech is more likely to co-occur with single movement gestures.

The perfective aspect takes a holistic view towards the event, so co-speech gestures with it are not likely to be repeated. However, gestures with iteration in the actualized and experiential aspect speech contexts were motivated by the repetitive or multiple readings of utterances, as in example (6a) and (6b).

(6) a. [zǐjī hái shǒu yòng fán tí]  chāo le]

yourself also hand use traditional Chinese characters copy ACTL

"You also copied the whole book with hand in traditional Chinese characters."

b. [nǐ jīn nián yàn guó de yī xíliè diàn yìng dǎngzhōng]
Among the series of films you played in recent years.

As shown in (6a) and Fig. 3.13 (6a), the speaker’s thumb touched the index finger on her right hand was molded like a pen head, and it moved laterally from the middle to her right side twice. The mimetic gesture of “writing action” expressed by chāo (copy) here preceded the speech and synchronized with the thematic elements zìjǐ hái shōu yòng fǎntīzì (you also (copied it) in the traditional Chinese characters by hand). As copying a script involves a lot of writing action, the repetitive construal implied here is demonstrated by the repeated gesture. In example (6b), the speaker moved her left open hand cyclically twice while uttering jīn nián yǎn guò de (you played in recent years), as Fig. 3.13 (6b) shows. The gestural repetitions were caused by the repetitive readings of the objects yī xīliè (a series of) mentioned later.

3.4.2 Grammatical aspect and gesture complexity

In addition to probing the hypothesis about the relationship between grammatical aspect and gesture forms, this chapter also examines whether macro grammatical aspects and micro aspectual categories differ in gesture complexity during event conceptualization. The findings provide a mixed answer to this question. Whereas imperfective aspect speech contexts were accompanied by more iconic gestures than perfective aspect contexts, actualized aspect-marked utterances also more frequently co-occurred with
Multimodal Events in Chinese Conversations: The Co-speech Gestural Representation of Grammatical Aspects

Iconic gestures. This result thus suggests that when the speakers chose the actualized aspect, a subtype of perfective aspect, to construe events in speech, they still tended to focus on the internal features of them in conceptualization, which were realized via the co-speech gestures. However, the perfective aspect was less likely to occur with representation of the internal structure of an event (perhaps because it prevented conceptual access to it), but rather with highlighting of the resulting static endpoint (e.g., Anderson et al. 2008; Anderson et al. 2013; Comrie 1976; Dowty 1977; Huette et al. 2014). With the other subtype of perfective aspect – the experiential aspect – speech contexts were mainly accompanied by pragmatic and deictic gestures, which encoded little information about the internal features of events. The experiential aspect emphasizes the completion of an event and the effect the past event had on the entities, so the gestures here were more pragmatically related to the immediate discourse and deictically related to the agent or location where an event took place. In this study, the actualized aspect was found to behave similarly to the imperfective aspect rather than to the experiential aspect in terms of gesture complexity about the events. It shows that the actualized aspect did not block access to the mental simulation of the internal features of a described event in language production. This may be interpreted as “aspect may not always indicate mental construal of event structure” to some extent (Parrill et al. 2013: 150). But there are many other possibilities for this result. One possibility is that the types of verb aspect, the inherent aspect of predicates (Vendler 1967), in the actualized and imperfective (the progressive and durative aspect) aspect speech contexts were more similar to each other than to those in the experiential aspect speech contexts. The interactions between the grammatical aspect and the verb aspect may affect event construal not only in speech but also in gesture (Becker et al. 2011). Another possibility is that although the gestures accompanying the actualized, progressive, and durative aspect speech contexts were all more likely to be iconic, they may be iconic in relation to different elements of an event, like the entities (agents, patients), action, path or settings. These points will be examined thoroughly in the following chapters.

In addition, the results also showed that the iconic gestures accompanying the
actualized, experiential and durative aspect speech contexts, which appeared more often in abstract events, were more likely to be metaphoric, while the iconic gestures with the durative aspect utterances, more frequently compatible with concrete events, were more concretely iconic. Why are there such preferences concerning grammatical aspect and event type? This may be related to our cognitive preferences. People tend to have more access to the concrete events which keep the action or maintain the resultant states of that action for a period of time. We may experience the durativity of concrete events like holding something or looking at how others hold something now and then in everyday life. Therefore, the concrete iconic gestures accompanying the durative aspect embody our recurrent durative experiences.

3.5 Conclusions

This study adds to our understanding of how grammatical aspect relates to gestural production in event construal. Gestures accompanying the progressive aspect were more likely to be iterative than those with the other grammatical aspects. These results suggest that the progressive aspect activates the dynamic intermediate phase of an event with more action. In addition, the gestures co-occurring with the imperfective aspect, including both the progressive and the durative aspect, and the actual aspect of the perfective aspect, were more complex with iconicity for the internal features of events than those with the experiential aspect of the perfective aspect. Such results have some implications for aspect studies. First, conflating the sub-aspect categories into the basic aspectual distinctions between perfective and imperfective aspects will probably cover some important differences in their gestural behaviors. Thus, it is of great significance to investigate the grammatical aspect at macro- as well as micro-levels. Second, the effects of grammatical aspectual distinctions on their co-speech gestures are gradient rather than categorical in language production. While constructing situation models, speakers may mentally simulate past events, which they acted in or perceived, for language production. The prototypical features of the events’ internal structure are
represented in speech via the imperfective aspect and co-occurring co-speech gestures, but they are also depicted in the gestures of perfective aspect, particularly the actualized aspect, to some extent. Third, the gestural construal of events was likely to be affected not only by the grammatical aspect, but also by other elements like verb aspect, arguments, and the event type itself. The following chapter will examine the relationship between grammatical aspect and gestures in different event types in terms of concreteness.
4.1 Introduction

In everyday conversations, we often talk about events we can neither witness nor experience. Whether exclaiming about the passage of time, describing changes of emotional states, or discussing changes in the stock market, we have to conceptualize and describe such events that we cannot directly perceive or perform. Then how can we communicate about those events to which we have no direct access without physical experience? The embodiment approach to language and cognition provides an answer to this question. That is, people understand and experience abstract concepts in terms of richer and more experience-based concrete concepts (Boroditsky & Ramscar 2002; Gibbs 1996; Lakoff & Johnson 1980). The way people talk about concrete and abstract events can provide some evidence for this proposal. The same verb can be used literally and metaphorically to describe both concrete and abstract domains. Action verbs in Mandarin Chinese provide a case in point. In their literal uses, they describe the actual action of entities, as in “tā zài pāo” (She is running), “zhàdàn zhà le” (The bomb exploded), and “shuǐ miàn zài zhǎng” (The water level is increasing). In their metaphorical uses, which are rich in everyday speech, action verbs are descriptive of time, emotions, change, and other abstract domains, as in “shí jiān pāo dé hǎo kuài” (Time is running very quickly), “tā yǐxià zhà le” (She exploded at once), and “gūpiào zài zhǎng” (The stock is increasing). Obviously, these abstract concepts are grounded metaphorically in embodied and situated knowledge. More and more cognitive linguists and psychologists are providing evidence that metaphor is not only a matter of language but also a matter of thought and action (Boroditsky & Ramscar 2002; Gibbs 2006; Lakoff & Johnson 1980). Consequently, metaphor is realized not only by language but
also by gestures (Cienki 1998a, b; Cienki & Müller 2008a, b). Furthermore, the same event can be construed in different ways and an event’s objective properties are insufficient to predict its construal (Langacker 1991a, 2008a). In viewing and describing a situation, people may choose either the ongoing intermediate phase or the completion of the entire event to focus on, which can be realized in relevant grammatical aspect forms in language (Comrie 1976; Langacker 1987, 2008a).

Previous studies demonstrated that the basic aspectual distinctions between the imperfective aspect and the perfective aspect modulate how people conceptualize events. Events described in imperfective aspect (realized by be... -ing in English; the progressive aspect particle zài and durative aspect particle zhe in Mandarin) stimulate the comprehenders to focus more on the middle portion of a path or dynamic nucleus of motion (Anderson et al. 2008; Anderson et al. 2010; Athanasopoulos & Bylund 2013a; Madden & Zwaan 2003), to yield greater facilitation of the manual ACE or LCE (Bergen & Wheeler 2010; Liu & Bergen 2016), to infer more action in a given period (Fausey & Matlock 2011; Matlock 2011), and to produce more complex and longer co-speech gestures (Duncan 2002; Parrill et al. 2013) than the perfective aspect (realized by the simple past or perfect in English; by the actualized aspect marker -le and the experiential aspect marker -guo in Mandarin). The results can be interpreted as more mental simulation of the internal structure of events involved in the imperfective aspect. All of the above studies compared the concrete (motion) events in the imperfective aspect description and the perfective aspect in artificial tasks. It is unclear how well the findings could be generalized to real-world natural conversation situations, which are replete with metaphorical, abstract, as well as concrete events. Whether, and to what extent, abstract concepts, including grammar constructions like grammatical aspect, are also grounded in action and perception is a critical, but relatively less explored question in embodied cognition theories, particularly from the perspective of language production. Co-speech gestures as visible components of utterances provide the most direct window for us to see how people mentally represent the concrete, metaphorical, and inherently abstract events. According to the GSA framework (Hostetter & Alibali 2008, 2010), gestures arise when speakers activate representations during speech
production that evoke mental simulations of spatial and motor events. The hypothesis
was based on physical manual actions without taking metaphorical and inherently
abstract actions into consideration. Therefore, this study aims at investigating how
Mandarin Chinese speakers conceptualize the concrete, metaphoric and abstract events
in speech with grammatical aspect markers and in co-speech gestures in natural
conversations.

4.2 Mental simulation in literal, metaphoric and abstract language
processing

Studies of the activation of the motor system in literal, metaphoric and abstract action
language comprehension using neurophysiological techniques have yielded somewhat
inconsistent results due to different techniques and tasks. For instance, Aziz-Zadeh et
al. (2006) used functional magnetic resonance imaging (fMRI) to test the activation of
areas in the human premotor cortex that is responsible for executing and observing
actions when people watch actions and read (literal or metaphorical) phrases relating to
three effectors – foot, hand, or mouth actions. They found effector-specific activations
in the premotor cortex of the left hemisphere for literal action (e.g., biting the peach,
grasping the scissors, pressing the car brake), but not for metaphorical action (e.g.,
chewing over the details, grasping the idea, kicking off the year). Similarly, through a
set of fMRI studies, Raposo et al. (2009) found significant activation in motor and
premotor regions when participants listened to action verbs in isolation (e.g., kick) and
to a lesser extent in literal sentences (e.g. kick the ball), but not in idiomatic sentences
(e.g. kick the bucket). In contrast, Boulenger et al. (2012) used Magnetoencephalograph
(MEG) and found somatotopic activation (the activation of motor-related areas in the
brain) both in silent reading of literal sentences (Pablo kicked the ball) and idiomatic
sentences (Pablo kicked the bucket). But none of the above studies considered
inherently abstract events. Desai et al. (2011) filled this gap. They used fMRI studies
to compare the neural responses to literal action (e.g. The daughter grasped the flowers),
metaphoric action (e.g. *The public grasped the idea*), and abstract action (e.g. *The public understand the idea*) sentences of varying familiarity. The results showed that the left anterior inferior partial lobe (aIPL) was activated both in literal and metaphoric contexts, that the left superior temporal areas were activated in both metaphoric and abstract contexts, and that the right aIPL regions associated with abstract language was also activated with the metaphoric sentences. Importantly, the more familiar and conventionalized the literal and metaphoric sentences are, the less the primary motor and biological motion perception areas were activated. Thus, understanding metaphoric language involves not only the process of understanding literal language but also that of abstract language. Relatively few studies explored the involvement of the motor system during language production. Morsella & Krauss (2005) employed Electromyography (EMG) to measure electro-muscular activity in participants’ hands and arms when they retrieved difficult lexical items. The results showed that retrieving concrete words led to more activity than retrieving abstract words. They interpreted this as evidence that motor activation is involved in language production.

Some behavioral studies found that mental simulation happens only in response to literal language but not in metaphoric and abstract language understanding. For example, Bergen et al. (2007) used a visual object categorization task to investigate whether language for concrete, metaphorical and abstract upward or downward motion (e.g. “The mule climbed”, “Oil prices climbed above $ 51 per barrel”, “The price multiplied”) interferes with visual processing in the same area of the visual part on the computer screen. They found that both the up- or down-associated nouns and verbs in concrete motion events, rather than in metaphoric and abstract motion events, activated the mental simulation of visual imagery.

Both the neuropsychological studies and behavioral studies reviewed above demonstrated that mental simulation is affected not simply by words but also by the sentential contexts where the concrete, metaphoric and abstract language are embedded. Grammar, as the other end of a lexico-grammar continuum (Langacker 1987; Talmy 2000), must also have an influence on mental simulation. As one of the most important grammatical devices in the world’s languages, grammatical aspect has recently gained
great attention in this field. For example, Liu & Bergen (2016) carried out a location-sentence compatibility experiment to investigate in what conditions (progressive aspect vs. perfect aspect (i.e., non-progressive in English); concrete events vs. abstract events) language comprehenders mentally simulate the ending location of a motion event. They found that the progressive aspect in both the concrete and abstract events facilitates mental simulation not simply of the central action but also of the ending location of the event, whereas the perfect aspect description of events did not elicit simulation of locations. Furthermore, the concreteness of events led to different simulation effects of locations even within the progressive aspect context: the concrete events showed a Location-sentence compatibility effect, whereas the abstract events showed an incompatibility effect. Thus, they conclude that mental simulation of locations must be different between concrete vs. abstract events.

Based on the findings surveyed above, this study intends to explore the following two questions.

1) Is the internal structure of an event more highlighted by the gestures accompanying imperfective aspect utterances than perfective aspect utterances in reference to concrete, metaphoric, and inherently abstract events? If the metaphoric and inherently abstract description of events simulates visual and motor images in the same way as the description of concrete events does, then imperfective aspect utterances should be accompanied by more iconic gestures than perfective aspect utterances in reference to metaphoric and inherently abstract events, as it does in reference to concrete events.

2) Are gestures accompanying imperfective and perfective aspect speech in concrete, metaphoric and inherently abstract event descriptions iconic for similar or different event semantic elements? If metaphoric and abstract events yield mental simulation similar to that performed in producing concrete events, then this would suggest that grammatical elements such as imperfective aspect and perfective aspect have the same effect on the three sorts of language.
4.3 Methods

4.3.1 Data

The multimodal data used for this study were taken from 20 videos from Yang Lan One on One. Each video lasts for about 20 minutes, and 8.5 hours long in total. Our corpus consists of 30 native Chinese speakers (10 females). There were 342 grammatical aspect marked utterances which were accompanied by gestures.

4.3.2 Annotation

I annotated the linguistic features and the gestural features of the 342 multimodal utterances with aspect markers.

4.3.2.1 Speech

In terms of speech, I encoded the utterance as either perfective aspect or imperfective aspect. In Mandarin, the perfective aspect is mainly realized by the particles le and guo. The particles -le attached to verbs usually in the middle, sometimes also in the end of an utterance, is used to express the actuality of an event, whereas -guo attached to verbs indicates the experientiality of an event. The imperfective aspect is mainly realized by the particles zài and zhe. The particle zài in front of verbs encodes an ongoing event, and -zhe attached to verbs focuses on the durativity of an event.

Then each aspect marked utterance was categorized into three types of events: concrete, metaphoric or abstract. A concrete event used a concrete action verb (e.g., dā kāi chuāng hù “open a window”, bào jīqì “hold a machine”) to depict a physical action. A metaphoric event used a concrete action verb in a figurative manner (e.g. dā kāi shìyē “broaden the vision”, bào zhe xī wàng “hold the hope”), so that no physical action was involved. An abstract event used an inherently abstract verb (e.g. xiǎng “think”, dān xīn “worry”) to describe the inner psychological state or a highly schematic verb (e.g. zuò “do”, gōngzuò “work”) to over-generalize physical actions.
One coder coded the utterance aspect for the whole dataset, and a second coded 50% of the dataset. The inter-rater agreement for the perfective and imperfective aspect coding was 98%, and for the concrete, metaphoric, and abstract events coding was 96%.

4.3.2.2 Gestures

Gestures were coded as iconic and non-iconic gestures. Iconic gestures referred to hand/arm movements that enacted the semantic content of the accompanying speech, including concrete iconic (i.e. depicting the shape, size and action of concrete events) and metaphoric iconic gestures (i.e. depicting the shape, size and action of abstract events). Non-iconic gestures were gestures that represented no semantic information of the speech itself, including the deictic gestures (i.e. pointing to concrete or abstract entities, locations etc. in the gesture space), and pragmatic gestures (i.e. hand movements contribute to the pragmatics of the utterance meaning, they may function as performative, modal, or parsing gestures).

The iconic gestures were further coded in terms of their iconicity for relevant semantic elements as follows. (1) Iconic for entities: gestures depicted the shape, size or other properties of the agent, patient or settings of an event. (2) Iconic for action: gestures reenacted out the movements of an event. (3) Iconic for path: gestures traced the path left by a motion event. (4) Iconic for process itself: gestures depicted the schematic processing of an event. For example, while uttering “she’s sitting there worrying”, the speaker rotated her two hands around one another alternatingly to indicate the dynamic duration of this abstract event (Duncan 2002: 202). However, these semantic elements in gestures were not mutually exclusive. Thus, I added the item (5) iconic for combination. For example, when uttering “He was writing”, the speaker mimicked the action of writing by moving the right hand horizontally from the left to right, and the handshape in which the right index finger touched upon the thumb and other fingers curled in looked like the head of a pen. Thus, this gesture is iconic for the action of “writing” and the entity “pen” at the same time.

The author of this thesis coded the entire gesture dataset, and a second coded 25%
of the dataset. The inter-rater agreement was good: 96% for gesture type; 94% for gestural iconicity relating to semantic elements of speech.

4.4 Results

4.4.1 Concrete events

As Table 4.1 shows, the utterances with imperfective aspect were accompanied by more iconic gestures than those in the perfective speech contexts (72.7% vs. 51.6%). When imperfective aspect was used, it co-occurred far more frequently iconic gestures than non-iconic gestures (72.7% vs. 27.3%). The differences were statistically significant enough ($\chi^2_{(1)} = 4.796$, $p < 0.05$). The results suggest that the speakers taking the imperfective aspect to construe events were more prone to mentally simulate the internal structures of concrete events than those taking the perfective aspect.

<table>
<thead>
<tr>
<th>Grammatical aspect</th>
<th>Gesture type</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Iconic</td>
<td></td>
<td>Non-iconic</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Perfective</td>
<td>32</td>
<td>51.6</td>
<td>30</td>
<td>48.4</td>
</tr>
<tr>
<td>Imperfective</td>
<td>32</td>
<td>72.7</td>
<td>12</td>
<td>27.3</td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td>60.4</td>
<td>42</td>
<td>39.6</td>
</tr>
</tbody>
</table>

In terms of the iconic gestures accompanying the aspect-marked utterances in the concrete events, I examined what elements they were iconic for in detail. As Fig. 4.1 reveals, the gestures accompanying the imperfective aspect were predominantly iconic for the combination of semantic elements of events (e.g., the combination of entity and action) that the speech denoted, and twice more so than gestures with the perfective aspect (51.5% vs. 22.9%). However, those accompanying the perfective aspect were more iconic for entity or action than those in the imperfective aspect contexts (31.4% vs. 15.2%, 25.7% vs. 15.2% respectively). These results suggest that the gestures accompanying imperfective aspect speech were more complex than those in the
perfective aspect contexts for concrete events, which is consistent with previous studies on concrete motion events (Duncan 2002; Parrill et al. 2013).

In the perfective aspect speech contexts, the speakers usually depicted the entities involved in the concrete events. As is the case in (1a). When the actress narrated her experience of searching for jobs, she mentioned the effect of the outstanding résumé she designed. Once initiating the topic related to the résumé, she extended her left palm-up open hand, which was iconic for the résumé. The gesture preceded the lexical affiliate jiǎnlì (résumé) and was synchronous with dōu shì kàn wǒ de (all just look at my). It held to the ending of this turn, though there were new gestures added to it in the following utterances. In the second actual aspect-marked utterance, as Fig. 4.2 (1a) depicts, the speaker lifted her left hand a little and moved her gaze from the addressee to her left hand when the predicate kàn (look) was said. In the third negative utterance, she shook her right hand several times. In the fourth actual aspect-marked utterance, her extended right index finger pointed down at the left-hand-as- résumé, which was metonymic for herself. The gesture was synchronous with the adverb zhījì jiù (directly then), and held until its lexical affiliate dìng le wǒ (decide to hire me) appeared. The series of gestures reflected that the speaker simulated the situation where the directors made the decision to hire her due to her excellent résumé. In fact, she didn’t witness the situation, so all of the situation was recreated by her via imagination.

Some gestures accompanying the perfective aspect speech also combined the
semantic elements of concrete events on some occasions. For example, in (1b) and Fig. 4.2 (1b), the position of the thumb and index of both hands was iconic for holding a camera, and the movements of the speaker’s index fingers as if clicking the shutter button was iconic for the action of taking a picture. Such a gesture reflects that he used a camera rather than a mobile phone to take a picture at that time.

(1) a. [[[tāmen dōu shì kàn wǒ de jiānlǐ.] [kàn le jiānlǐ yǐhòu],
   they all be look I DE résumé, look ACTL résumé after,
   [bù yòng jiàn wǒ rén], [zhǐjiē jiù dìng le wǒ.]]
   no use meet I people, directly then hire ACTL me.
   “All of them just had a look at my résumé. After looking at it, they directly decided to hire me without needing to give me an interview.”

b. [ránhòu wǒ pāi le yī-gè zhàopiān].
then I take ACTL a-CLF picture.
   “Then I took a picture.”

c. [wǒmen jiù ná zhe gē zhī bèi ]
we then hold DUR CLF paper cup.
   “Then we held a paper cup.”

d. [[tā hái zài pá], [cái qī-gè duō yuè], [tā zài pá]]
he still PROG crawl, still seven-CLF more month, he PROG crawl.
   “He was still in the phase of crawling, as he was only more than seven months old, and he was crawling.”
Gestural iconicity for the combination of semantic elements of concrete events was dominant in the imperfective aspect speech contexts. However, it occurred mainly in the durative aspect speech contexts, and the progressive aspect-marked utterances were usually accompanied by gestures iconic for action only. For example, in (1c) and Fig. 4.2 (1c), the speaker’s left thumb and index finger touched each other to be iconic for the rim of the paper cup, the other fingers curled around as if holding it. He simulated the entity and action carried out by the actress beside him in the past with details. The progressive aspect speech such as in (1d) went together with the speakers’ focus on the action. As Fig. 4.2 (1d) depicts, the actress’s two fists moved forward and backward in alternate ways three times, each movement being synchronous with the predicate of each utterance. Although the second utterance did not use the predicate zài pá (was crawling), the lexical affiliate of the gestures, she also made the gesture to complement her explanation. Such a series of gestures demonstrates the characteristics of a crawling phase, as the speech described the state of the child as a whole.

4.4.2 Metaphoric events

It is widely acknowledged that people understand abstract concepts metaphorically in terms of concrete notions (Lakoff 1987; Lakoff & Johnson 1980). Metaphorical experiences of abstract concepts are expressed not only in language but also in gestures.
We saw in the previous section that speech encoding the literal meaning of concrete verbs to construe events was more likely to be accompanied by iconic gestures in imperfective aspect than the perfective aspect contexts. Will this effect be extended to the utterances that include verbs denoting metaphorical meanings? As Table 4.2 reveals, there was no significant difference in the gesture use between the two grammatical aspect situations with metaphorical events ($\chi^2(1) = 0.325, p > 0.05$). Different from the pattern in the concrete event contexts, the gestures co-occurring with not only imperfective but also perfective speech contexts were highly iconic (78.3% vs. 72.1%), and with both aspects they occurred three times more frequently with iconic than with non-iconic gestures. This result suggests that metaphoricity is a much more important predictor for recalling the details of events when they are metaphoric events, without regard to aspectual distinctions.

<table>
<thead>
<tr>
<th>Grammatical aspect</th>
<th>Gesture type</th>
<th>Iconic</th>
<th>Non-iconic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Perfective</td>
<td>44</td>
<td>72.1</td>
<td>17</td>
<td>27.9</td>
</tr>
<tr>
<td>Imperfective</td>
<td>18</td>
<td>78.3</td>
<td>5</td>
<td>21.7</td>
</tr>
<tr>
<td>Total</td>
<td>62</td>
<td>73.8</td>
<td>42</td>
<td>39.6</td>
</tr>
</tbody>
</table>

Taking into consideration the semantic elements of the iconic gestures referred to, both the imperfective and perfective aspect-marked utterances were accompanied by gestures which were also dominantly iconic for the action they depicted (50.0% and 42.9% respectively) (see Fig. 4.3). The gestures co-occurring with the perfective aspect were more iconic for combinations (e.g. entity and action) with the path than those with the imperfective aspect, which were more iconic for process and entity.
Unlike in the concrete events, not so many gestures were iconic for the entities of the metaphoric events, as in (2a). The speaker’s palm-up open right hand was iconic for the track of the playground, and this stroke hold was synchronous with *xiǎo hái shénme* (the child what). When uttering the topic *xiǎo hái shénme* (the child what), his left index finger drew a straight line on his right palm, and he held it through the comment of the whole sentence *shū zài le qǐpāo xiàn* (lose at the starting line), as shown in Fig. 4.4 (2a). The concrete domain RUNNIng MATCH was mapped onto the abstract domain LIFE in speech, consisting in the metaphor “LIFE IS A RUNNIng MATCH”. The competitors, the track, the starting point and the winning or losing results were all projected onto the abstract concept LIFE. His gestures highlighted the track and starting line.

There were also a number of motion events in metaphoric contexts. Their co-speech gestures were more likely to be iconic for the path in the perfective aspect speech contexts than in the imperfective speech contexts. For example, as in (2b), the speaker was narrating that he liked digging holes (i.e., to direct some TV series or films whose topics were very disputable and dangerous) as he progressed smoothly and successfully in his career. His right index finger traced the path of walking around the round hole, as presented in Fig. 4.4 (2b). We can infer from the gesture that he mentally simulated
the abstract event of overcoming difficulties in work or life as the concrete event of walking around a dangerous hole.

(2) 

a. jiù shì [tā //pà zhègè xiǎo hái shénme shū zài le qǐ pào xiàn.] then be she fear this-CLF little child what lose at ACTL starting line.

“Then she feared that the little child lost at the starting point (in life).”

b. dānshì [[[hòulái shì kěng de biān,][biān shàng zhuàn le yī quān], but later be at hole DE side side on walk ACTL a round, [méi diào xiàqù.]]

test fall down

“But later I walked around the hole and did not fall into it.”

c. [dǎkāi le yì-gè xīn de zhègè rénmen yán jiū]

open ACTL a-CLF new DE this people study

zhè-gè shì jiè de yì-gè] fāngshì huòzhě shuō dàolù.

this-CLF world DE a-CLF way or say road.

“(It) opened a new way or road for people to study the world.”

d. [[yǒu hěnduō yě yǒu.. zhèngfǔ…(0.6s)] [zài tuídòng nèixiē diàn]zì shāngwù.

have many also have government PROG push those electronic commerce.

“There were many..also governments which were pushing those e-commerces.”
In metaphorical events, gestures accompanying both perfective and imperfective aspect speech were more likely to be iconic for the action embodied in the concrete events the speakers experience here and there. Take (2c) for instance; the speaker mapped the concrete action of opening doors, windows, cupboards etc. in everyday life onto the abstract action of opening a new way or road to study the world. As Fig. 4.4 (2c) depicts, the speaker moved his two palms towards the center, open hands moving forward and upward, separating from each other when uttering dākāi le (opened), and he held them this way for a while. Similarly, in example (2d) and Fig. 4.4 (2d), the speaker moved his palm downwards open right hand up and held at the height of his right shoulder before the stroke when uttering hěnduō yě yǒu (many also have), and then moved it downwards and outwards quickly when uttering the topic of this utterance zhèngfǔ (government). When the progressive aspect-marked comment zài tuīdòng (is pushing) occurred, the same gesture was repeated twice in a diminishing manner. Thus, the speaker simulated the abstract concept of facilitating e-commerce development via the concrete action of pushing a button or something in everyday life. In most metaphorical events, the agents or patients are abstract, but the actions have an embodied basis. Therefore, the actions were more likely to be activated and represented by gestures.
4.4.3 Inherently abstract events

Abstract utterances, like the metaphoric utterances in the last section, represented abstract events but did so using verbs that did not have a concrete meaning denoting the physical action either (verbs such as biàn “change” and xiǎng “think”). The strong version of the embodied approach to language proposes that representation and comprehension of all concepts, be they concrete or abstract, are all embodied in our experience in, and interaction with, the world. Therefore, it is of great importance to investigate whether the gestures accompanying them will behave similarly to those in the concrete contexts when the inherently abstract events are framed by grammatical aspects.

As Table 4.3 shows, the gestures accompanying the imperfective aspect speech were more likely to be iconic than those in the perfective aspect speech contexts (61.4% vs. 38.9%). This was also the case with the concrete events. The imperfective aspect-marked utterances were accompanied by more iconic gestures than non-iconic gestures (61.4% vs. 38.6%), and the situation was the opposite in the perfective speech contexts (38.9% vs. 61.1%). The differences between them were significant via the Chi-square test ($\chi^2(1) = 6.371, p < 0.05$). The results suggest that the speakers choosing the imperfective aspect to describe an event with inherently abstract verbs were more prone to mentally simulate the internal structure of the event than they were when choosing the perfective aspect.

<table>
<thead>
<tr>
<th>Grammatical aspect</th>
<th>Gesture type</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Iconic</td>
<td>Non-iconic</td>
<td>Total</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Perfective</td>
<td>42</td>
<td>38.9</td>
<td>66</td>
</tr>
<tr>
<td>Imperfective</td>
<td>27</td>
<td>61.4</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>69</td>
<td>45.4</td>
<td>83</td>
</tr>
</tbody>
</table>

As Fig. 4.5 shows, the iconic gestures accompanying the imperfective aspect were dominantly iconic for the actual process itself of an event, to a much greater extent than
that in the perfective aspect context (59.3% vs. 33.3%). Furthermore, the gestures accompanying the imperfective aspect speech were two times more likely to be iconic for the combination (mainly combination of metaphoric entity and process) than those in the perfective aspect speech contexts (11.1% vs. 4.8%). The iconic gestures with the perfective aspect were more iconic for the entity and action referred to in an inherent abstract event than those with the imperfective aspect (38.1% vs. 14.8%, and 21.4% vs. 14.8% respectively).

![Fig. 4.5 Gestural iconicity for different semantic elements in abstract verb contexts](image)

Previous studies reported that the “cyclic” gesture (i.e. a continuous circling motion of the loose hand) (Ladewig 2011) often accompanied the English progressive construction (*be* *v-ing*) to express progressivity (Harrison 2009). Taking a close look at the examples given by Harrison (2009), I found that the cyclic gestures encoding progressivity all occurred in inherently abstract events, such as “it’s turning out to be quite expensive”, “there’s something going on in the city that...”, “Because I never know what’s gonna change”. I also found some, but not so many, examples in our corpus, such as in example (3a). The actress made a two-handed (mirroring) cyclic gesture with the whole predicate by moving the hands downward-outward-inward vertically and then ended it with a palm-up open hand gesture with a hold for 10 seconds before the next utterance, as shown in Fig. 4.6 (1a). Such cyclic gestures “directly symbolize duration and dynamism” of the processing of an inherently abstract event.
(3) a. [wǒ juéde shì zài jinxìng yì-gè xǐlì.] I think be PROG doing one-CLF baptism.

“I think we were going through the baptism (of a tough life)”

b. [shichāng lăobăixìng yě zài tiáozhēng] market ordinary people too PROG adjust

“The ordinary people are adjusting themselves to markets too”

Fig. 4.6 (3a) Cyclic gesture for PROG
Fig. 4.6 (3b) Back and forth gesture for PROG

However, the prototypical gesture accompanying the imperfective aspect in Mandarin Chinese conversations was the “back and forth” gesture rather than the cyclic gesture. The prototypical “back and forth” gesture was made with an intense or loose hand(s) alternating away and towards the speaker’s body. Take (3b) for instance; in Fig. 4.6 (3b), the speaker makes a “back and forth” gesture by alternating the tense open right hand away and towards his body while uttering the imperfective aspect-marked utterance lăobăixìng yě zài tiáozhēng (ordinary people are also adjusting).

Due to the affordances of gestures and influence of context, inherently abstract events are usually accompanied by gestures iconic for the implicit agents, patients, actions or circumstances located in the events. Take (4a) as an example; before this utterance, the speaker recalled the preparation his son was making for directing a film. He located the center gesture space as the preparation period, the left gesture space as
the ending period and the middle period of film making by making a metaphorical
gesture like Fig. 4.7 (4a). His two spread flat palms facing towards center consisted of
a bounded container to represent the abstract concept TIME. When he said his son spent
much more time in the preparation period than in the postproduction phase or even the
shooting period, he explained that because “he was thinking and doing” during it. As
verbs like think and do are abstract, their actions are not available for realizing as
gestures. Therefore, the speaker referred back to the conduit metaphor gesture (McNeill
1992) in the center. This implies that the abstract events of “his thinking and doing”
occedurred in the preparation period. Similarly, in example (4b), the actress made a
gesture in which her right index and thumb horizontally paralleled, and the other fingers
curled onto the palm while uttering wǒ cèngjīng (I once), it held statively when the
abstract mental predicate hèn guo (hated) was mentioned, then she moved the tensed
index-thumb upward quickly in front of her right eye when uttering yī xiǎo xiàzǐ (for a
little while), as Fig.4.7 (4b) presents. Therefore, I can infer that the tensed index-thumb
was metaphorically iconic for the length of TIME yī xiǎo xiàzǐ that her hate lasted for.
However, when we look at the following utterance, we see that the tensed index-thumb
gesture was simulated as “hate” at the same time. When she said zìjǐ yòu bù jiàn le
(itself disappeared), the tensed index finger and thumb slowly came together to touch
each other and meanwhile moved from the center to the extreme right, which
metaphorically mimicked the action of bù jiàn le (disappeared). It is not hard to see that
what the gestures represent was very flexible and dynamic, strongly depending on the
context.

(4) a. [[yīnwèi tā zài xiǎng, tā zài zuò.]
   Because he PROG think, he PROG do.

   “Because he was thinking and doing.”

b. [[[dànshí wǒ cèngjīng hèn guo] [//yī xiǎo xiàzǐ,]
   but I once hate EXP a small CLF,
   [dàn hòulái e zìjǐ yòu bù jiàn le.]]
   but later PRT itself again not see ACTL.
“But I once hated (my ex-boyfriend) for a while, but later it disappeared again.”

c. [qíshí wǒmen zuò guò yī-gè diào yī-gè diàoyán.]  
actually we make EXP a-CLF invest.. a-CLF investigation  
“Actually, we made an investigation.”

d. [nǐ kàn zhè fāngdìchǎn liǎkè jiù.. jiù nánshòu le.]  
you see the real estate immediately then unwell ACTL.  
“You see that the real estate immediately felt unwell.”

In the perfective speech contexts, the speakers’ gestures were prone to focus on the entities in the inherently abstract events, such as in example (4c). In speech, the speaker used a super-ordinate verb zuò (make or do) and chose the experiential aspect
to construe the experience of making an investigation of the excellent resources like education, healthcare, science and technology, the job opportunities, and environment, which the ordinary people were concerned about while choosing a city to live in. As Fig. 4.7 (4c) depicts, in the co-speech gesture, his palm-up open left hand and the stretched forearm modeled the questionnaire sheet (using the embodying mode of representation, introduced in Chapter 2), and his open right hand with the palm towards the body moved from the top to the elbow of the left hand, modeling the arrangement of the excellent resources items on the questionnaire sheet. The gesture was iconic for the entity “questionnaire sheet”, which was not explicit in speech but complementary with it to clarify how they made the investigation.

In some inherently abstract events, the speakers sometimes used more schematic verbs to describe an event, but the co-speech gestures represented the action beyond words. Although the implied verbs were seldom employed explicitly before or after what the multimodal utterance said, they could be inferred from the context. In example (4d), in speech, the speaker used the mental state verb nánshòu (unwell, uncomfortable) to depict the situation of market once the broad money (i.e. M2 in economy) was adjusted. Its implied meaning was “The housing prices became unstable, fluctuating up and down”, which was elaborated by the co-speech gesture. The speaker moved his right open hand up and down twice while uttering lìkè jiù... jiù nánshòu le (felt unwell immediately) as in Fig. 4.7 (4d).

Like in the metaphoric events, there were few gestures iconic for the combination of entity and action or process in the inherently abstract events. For instance, in example (5a) the speaker was narrating how his colleague constructed a data model to predict the Academy Award in 2014. He moved his two hands to the center from the armchair in rest position in the preparation phase, then slightly shook his two palms towards center with open hands while uttering yīzhī zài biàn (was changing all the time) as Fig. 4.8 (5a) displays. In the following utterance, he clarified what the deictic zhègè (this) in the last utterance referred to was móxing (model); thus we can infer that his two palm-towards-center open hands were iconic for the entity, i.e., the “data model”. Furthermore, he moved the “data model” from the center to the right side with a slight
shake at the same time while elaborating “you see that his model was changing all the time”. The slight shake was to be iconic for the ongoing process of the abstract event (CHANGE IS MOTION [Grady 1997]), and the movements of change in SPACE can be interpreted as metaphorically iconic for change in QUANTITY of the data model. Similarly, in example (5b) and Fig. 4.8 (5b), the speaker’s right index finger and thumb were bent, and the other fingers curled onto the palm. Such a tensed index-thumb arrangement, as if modeling the length of three or four days, moved laterally from the center to the extreme right quickly while starting the topic tā yǐzhī zài (he was all the time). After a short pause, he completed the utterance while moving the tensed index-thumb back parallel to his right shoulder. In speech, the abstract verb nèijiù (guilty) was used to depict his son’s mental activity after seeing the quarrel between him and his wife. The holding and movement of the tensed index-thumb from the beginning to the end of the utterance can be seen as metaphorically iconic for the combination of entity and process to represent the duration and progressivity of the abstract mental action.

(5) a. [[tā zhègè yǐzhī zài biàn],
   he this all the time PROG change,
   [ni kàn tā de móxing yǐzhī zài biàn.]]
   you see he DE model all the time PROG change.
   “This was changing all the time. You see that his model was changing all the time.”

b. [[tā yǐzhī zài (.0.2s)] [zhè sān sì tiān tā zài nèijiù]].
   he all the time PROG this three four day he PROG guilty.
   “He was guilty all the time during the three or four days.”
To investigate the interaction effects between grammatical aspects, verb types and gesture type, I carried out the LogLinear Model Selection process. The results showed that there was no interaction effect between grammatical aspects, event type and gesture type ($\chi^2_{(2)} = 0.739, p > 0.05$). However, there were significant interaction effects between grammatical aspect and event type ($\chi^2_{(2)} = 5.490, p < 0.05$), between grammatical aspects and gesture type ($\chi^2_{(1)} = 8.419, p < 0.05$), and between the event type and gesture type ($\chi^2_{(2)} = 20.326, p < 0.05$). The results suggest that the functions of the co-speech gestures accompanying the grammatical aspectual utterances are influenced by the grammatical aspect and event type, respectively.

4.5 Discussion

The study reported in this chapter investigated the gestures accompanying the grammatical aspect-marked utterances in concrete, metaphorical and inherently abstract events in Mandarin Chinese conversations. In this section, I will first summarize the results so that I can answer the two research questions, and then I will interpret them respectively.

I examined the frequencies of iconic gestures co-occurring with grammatical aspect speech to answer the first question I put forward: is the event internal structure more highlighted by the gestures accompanying imperfective aspect utterances than
perfective aspect utterances in concrete, metaphoric, and abstract events? The results showed that gestures accompanying imperfective aspect speech were more prone to be (concrete or metaphoric) iconic than gestures accompanying perfective aspect speech in the concrete and inherently abstract events, whereas the gestures that speakers produced in both imperfective and perfective aspect-marked speech contexts were equally frequently iconic in the metaphoric events. The more iconic gestures represent the events, the more internal structures were highlighted. Thus, the results suggest that the event-internal structures were more highlighted by the gestures accompanying the imperfective aspect-marked utterances than the perfective aspect-marked utterances in the concrete and inherently abstract events, and were equally highlighted in the metaphoric events.

According to GSA framework, representational gestures emerge from the perceptual and motor simulations that underlie embodied language and motor imagery (Hostetter & Alibali 2008). The speakers who produced more iconic gestures in event descriptions should activate more perceptual and motor information about the internal structures of the events. Our results further suggest that describing metaphoric and inherently abstract events often stimulates the motor system as when describing the physically experienced or perceived concrete events, as evidenced by the motoric movements of gesturing. They were consistent with the results from some neurophysiological studies reviewed in section 4.2 that understanding literal sentences (i.e., describing concrete events), metaphoric sentences (i.e., describing metaphoric events) and abstract sentences (i.e., describing inherently abstract events) could produce somatotopic activation (Boulenger et al. 2012; Desai et al. 2011). The embodied cognition approach proposes that many abstract concepts are understood and experienced via analogies to sensation and action (Bergen et al. 2007; Gibbs 2006; Lakoff & Johnson 1980, 1999). This may be useful for explaining why the metaphoric events were more likely to be accompanied by iconic gestures to highlight their internal structures regardless of the temporal perspectives the speakers took to construe the events. However, in the concrete and inherently abstract events, the imperfective aspect viewpoint appears to correlate with the speakers’ focusing more on the internal
structures of events in contrast with the perfective aspect via their co-speech gesture use.

Metaphoric and inherently abstract events had the same potential to be represented iconically as the concrete events did in grammatical aspect speech contexts. But metaphorical and inherently abstract events have a different nature and can activate different regions in the brain in contrast with concrete events (Desai et al. 2011). Let us now turn to the second question: Are gestures accompanying the imperfective and perfective aspect speech in concrete, metaphorical and inherently abstract events descriptions iconic for similar or different event semantic elements? In this regard, I found that co-speech gestures have “preferences” in terms of iconicity for different semantic elements with different types of events. Gestures accompanying the imperfective aspect-marked utterances were more iconic for the combination of event semantic elements, whereas those co-occurring with the perfective aspect speech were more iconic for entity and for action of the concrete events. For the metaphorical events, gestures accompanying the imperfective and perfective aspect speech were all predominantly iconic for action. Gestures accompanying the imperfective aspect speech were more iconic for the process, whereas those accompanying the perfective aspect were more iconic for entity in the inherently abstract events. Our results extended the previous studies (Duncan 2002; Parrill et al. 2013), as gestures accompanying imperfective-aspect marked speech were shown to be more complex (representing more semantic elements) than those with the perfective-aspect speech not only in concrete motion events but also in various kinds of concrete events. However, this effect did not arise in the metaphorical and inherent abstract events. Behavioral and electrophysiological studies also showed that progressive aspect descriptions of concrete events activate more richly detailed information, such as details about location and the entities of an event (Carreiras et al. 1997; Ferretti et al. 2007). Furthermore, concrete events can provide more detailed information in mental imagery, and the imperfective aspect facilitated greater fine-grained simulation in concrete imagery. Thus, the entities, locations, actions etc. were more available to be combined in gestural representations of the events. However, in the metaphorical and inherently abstract events,
either the entities or actions involved were tangible without direct access being possible to their mental images of the (abstract) target referents.

Previous studies showed that use of the imperfective aspect focused on the dynamic ongoing actions of an event, whereas the perfective aspect emphasized the completed end-state or affected entities of an event. As our study showed, this only occurred in the concrete and inherently abstract events rather than in the metaphorical events. In the metaphorical event descriptions, the verbs were concrete, embodied in the physical actions we carry out in everyday life. No matter what grammatical aspect the speaker employed to construe the events, motion or action may be simulated first to bridge the mapping between the concrete domain and the abstract domain.

Grammatical aspect is closely related to the process of events and cognitively motivated by our ability to simulate action (Barsalou 2009; Langacker 2008a). With abstract events, the verbs were inherently abstract without direct access to a physical action we can perceive or do. As the semantic poles of their symbolic structures were both highly schematic, their relevant cognitive iconicity depended on a correspondingly schematic phonological structure (Wilcox 2004). The qualitative results about inherently abstract events show that the entities represented by the co-speech gestures in perfective aspect were mainly schematic as conduit metaphors of time to describe the temporal locations of the abstract events. The movements of the gestures did not have some clear and concrete action forms like the concrete and metaphorical use of verbs, but rather the schematic “back and forth” or “cyclic” forms, which were the schematic phonological form of the imperfective aspect. Our results were inconsistent with some psychological or behavioral studies which reported that understanding abstract events did not activate motor systems of the brain or stimulate mental simulation of visual imagery (Aziz-Zadeh et al. 2006; Bergen et al. 2007; Raposo et al. 2009). This may result from the differing contextual scopes provided by psychological experimental materials and by natural conversations. The psychological studies all used one-sentence stimuli, whereas the natural conversations consist of bigger discourse units. The discourse contexts influence the growth points of gestures accompanying speech (McNeill 2005). Thus, the metaphorical entities mentioned in the perfective-aspect
utterances, and the implicit action implied by the grammatical aspect-marked utterances, were also represented in gestures to keep the conversations coherent.

4.6 Conclusions

A comparison of gestures accompanying grammatical aspect-marked utterances in concrete, metaphoric and abstract events revealed that grammatical aspect modulated the conceptualization of concrete and inherently abstract events rather than that of metaphoric events. Gestures co-occurring with imperfective-marked speech were more likely to be iconic than those with perfective-marked speech for the concrete and inherently abstract events, whereas both types of speech were equally accompanied by gestures which were highly iconic for the metaphoric events. Furthermore, the gestures in the imperfective speech contexts were more iconic for a combination of semantic elements such as entity and action for concrete events, and for process for inherently abstract events, whereas those in the perfective speech contexts were more iconic for the (concrete or metaphoric) entities in the two kinds of events. However, the gestures accompanying the imperfective and perfective aspect speech were not different, and for both of them were mainly iconic for actions in metaphoric events. The speakers made more complex gestures when producing imperfective aspect utterances than when producing perfective aspect ones only in the concrete events.

These findings have a number of clear implications. First, metaphoric events and inherently abstract events can strongly activate mental simulation of events as the concrete events did when they were marked by grammatical aspects in speech production. Gestures as simulated action have been argued to be visible manifestations of the speakers' mental simulation of visual or motor images related to events. While (re)constructing situation models, the speakers made more iconic gestures than non-iconic gestures in the concrete and abstract events rather than in the metaphoric events when they are marked by imperfective aspect rather than perfective aspect.

Second, what elements of events the speakers focus on in event conceptualization
is modulated by the grammatical aspectual categories, which rely on the event types with which they occur. The imperfective aspect facilitated the speakers’ focus on the dynamic nucleus, whether it was concrete or schematic, in contrast with the perfective aspect, where the focus was on the entities, in the concrete and abstract events. Metaphoricity based on physical actions overrides the grammatical aspect to mainly simulate the action in both imperfective and perfective aspects speech contexts.

There are some limitations of this study. First, the corpus is relatively small, particularly the samples of metaphoric events. Second, the advantages of natural conversations, which can provide various kinds of events, larger discourse contexts, and show what people actually say and do in event construal, also entail disadvantages at the same time. There are too many factors influencing the gestural behaviors. I cannot definitely say what features are uniquely inherent to the grammatical aspect in different event types. Therefore, the results of this study need to be verified in future studies by strictly controlling other factors such as the nature of agents, patients, and use of the same stimuli to elicit speech and gestures.
Chapter 5. Dynamicity and Stativity of the Mandarin Chinese Imperfective Aspect: A Multimodal Perspective

5.1 Introduction

Aspect is concerned with “the way the grammar marks the duration of type of temporal activity denoted by the verb” (Crystal 1991: 27). It relates to “the internal temporary constituency” (Comrie 1976: 5), or “the internal temporal structure” (Trask 1993: 21), of a situation. Grammatical aspect allows for varying perspectives on the very same events (Givón 2001; Langacker 2001; Smith 1994). There are two basic grammatical aspect distinctions found in many languages, that is, perfective aspect vs. imperfective aspect. The perfective aspect views an event as a whole from an external perspective, whereas the imperfective aspect is characterized by “explicit reference to the internal temporal structure of a situation, viewing a situation from within” (Comrie 1976: 24). Consequently, the imperfective aspect registers finer-grained details and increases imagined image accuracy (Tversky & Paul 1998). Furthermore, the imperfective aspect is divided into two categories: habitual aspect and continuous aspect. The latter is subdivided into progressive aspect and non-progressive aspect (Comrie 1976: 28).

The imperfective aspect can be realized in one or several grammatical forms in different languages. For example, English only has one imperfective aspect marker – *be*...*ing* to encode events as ongoing. Langacker (2001: 259) treats the imperfective aspect in English (i.e. progressive aspect *be*...*ing*) as a device of “zooming in” and “taking an internal perspective” of an event. However, Mandarin Chinese has two main separate imperfective aspect markers – *zài* and *zhe*. Historically, *zhe* has long been considered as the single progressive aspect marker in Chinese. With the grammaticalization of *zài*, it is gradually taking the place of *zhe* as the commonly
acknowledged progressive aspect marker. By contrast, *zhe* is seen as the durative aspect marker. Syntactically, *zài* is used before the verb, whereas *-zhe* follows the verb. The former is autonomous, while the latter is dependent on other elements to form a sentence. For example, *tā zài chī fàn* (*He is having dinner*) is grammatical, but *tā chī zhe fàn* (*He is having dinner*) is ungrammatical. In order to be acceptable, it needs some other elements to be added like *ne* (a particle). Semantically, there has been a long debate about the differences between *zài* and *zhe*, and particularly, about *zhe* among the linguists. Some researchers proposed that the dualism of dynamicity vs. stativity is the basic semantic distinction between *zài* and *zhe* (e.g., Comrie 1976; Li & Thompson 1981; Smith 1991; Teng 1979). They concluded that *zài* dominantly co-occurs with active verb predicates, representing dynamicity. *Zhe* has been taken to usually be attached to stative verb predicates to express stativity (e.g., Dai 1991, 1997; Li & Thompson 1981; Smith 1991; Teng 1979). Smith (1991) even termed it the “stative imperfective *zhe*”. Some scholars argue that the core grammatical meaning of *zhe* is “durativity”, only focusing on the continuation of events or states (Dai 1991, 1997; Xiao & McEnery 2004). Yang (2013) suggests that the construction meaning of “*zài + V*” is “event is being on”, and that of “*V + zhe*” is “constant state”. However, others argued that there were two kinds of *zhe* in Mandarin Chinese: *zhe*₁ is the durative aspect marker, and *zhe*₂ is the continuous or progressive aspect marker, but they experienced different grammaticalization pathways (Liang 2010). While *zhe*₁ encodes the duration of static events, *zhe*₂ can also focus on the ongoingness of dynamic events, as *zài* does when the verb is an action verb (e.g., Lü 1980; Lu 1999; Wang 2012; Zhu 1982). Some hold that there is only one *zhe* in Chinese, and that it has two grammatical meanings, that is, the ongoingness of dynamic action and durativity of states (Lü 1980; Zhu 1982). Both Jin (1991) and Qian (2000) claimed that *zhe* cannot denote “ongoingness/progressive”, and “the basic grammatical meaning of *zhe* is ‘state’, signaling that the event is in a state and in a relatively stable and homogenous developmental stage” (Jin 1991: 26). Jin (1991) further argued that the differences of stativity and dynamicity resulted from the related verbs themselves rather than the grammatical meanings of *zhe*. Wang (2012) explains that the semantic contrast of dynamicity/stativity can be only applicable to *zài*,
which co-occurs with dynamic and durative verb predicates. She argues that the “homogeneity” of verb predicates (VPs) plays a central role in the semantic interpretation of *zhe*, which only co-occurs with homogeneous VPs. The overlapping use of *zài* and *zhe* lies in encoding the ongoingness of “dynamic and homogeneous events” (Wang 2012: 243). The debate on the aspectual construal of progressive and durative aspects in Mandarin Chinese is still ongoing among Chinese linguists. However, focusing only on the linguistic evidence, as the previous studies have done, easily leads to cyclic argumentation in reasoning about linguistic forms and their meaning. Co-speech gestures have been discussed as a possible means of indicating aspectual distinctions in event representations. Gestures with the imperfective aspect categories in both English and Chinese are longer and more complex than those with the perfective aspect (Duncan 2002; McNeill 2003, Matlock et al. 2012; Parrill et al. 2013). Duncan (2002: 197) has also noted that gestures with durative aspect are more complex in form and represent more semantic content than those with progressive aspect. It is still unclear whether and how the widely discussed dynamic-static distinction of progressive aspect and durative aspect in linguistics can also be represented in gestures.

Cognitive Linguistics claims that “(grammatical) meaning is identified with conceptualization” (Langacker 2008a: 30). Meaning as conceptualization has the potential to be realized not only in language, but also in co-speech gestures as well (Langacker 2008a, b; McNeill 1992). As simulated action (Hostetter & Alibali 2008), gesture is a window to “part of the speaker’s current cognitive being, her very mental existence, at the moment it occurs” (McNeill 2005:92). According to the thinking for speaking and gesturing hypothesis (Cienki 2008; Slobin 1987, 1996), differences in the grammatical options provided in a language or across languages can influence the speaker’s thinking in the moment-by-moment processes of talking and gesturing. Therefore, it is of great significance to study whether and how the two Mandarin imperfective aspect categories relate to Chinese speakers’ speaking and gesturing. This chapter intends to investigate whether the insight into the conceptualization of events provided by gestural data in natural conversations supports a particular semantic model.
of aspect for the imperfective in Chinese: that of the imperfective as one coherent whole category; that of zài and zhe as each expressing their own internally consistent semantic category; or that of zhe as comprising two subcategories, one of which functions more like zài. This will be investigated by finding the answers to two empirical questions. First, do gestures accompanying imperfective utterances represent the semantic differences of dynamic vs. static distinctions between the progressive aspect and the durative aspect? Second, in what contexts do gestures with dynamic or static qualities accompany the imperfective constructions?

5.2 Methods

5.2.1 Corpus description

Our data were extracted from 32 videos from Yang Lan One on One. Each video lasts for about 20 minutes (excluding the advertisements and inserted video clips), so the multimodal corpus is 640 minutes in total. I went through each video carefully, and sampled out all instances of the two kinds of imperfective aspect-marked utterances with iconic (concrete or metaphoric) gestures, which are described below. In total, there were 48 progressive zài and 40 durative zhe marked-utterances with such explicit concrete or metaphoric iconic gestures. The 88 imperfective aspect-marked utterances with iconic gestures constitute a sub-corpus of multimodal data marked with the imperfective aspect.

5.2.2 Data coding

All the data were coded in ELAN, video annotation software developed by the Max Planck Institute for Psycholinguistics (see Lausberg & Sloetjes (2009) and http://tla.mpi.nl/tools/tla-tools/elan/ for details).
5.2.2.1 Speech

All the utterances marked by the imperfective aspect markers were annotated as the progressive aspect or the durative aspect, but only the utterances marked by progressive aspect marker *zài* and durative aspect marker *zhe* with iconic gestures were further analyzed. Two coders coded all the 88 imperfective aspect-marked utterances. The agreement was 98%.

5.2.2.2 Gesture

Every gesture phrase, i.e. a whole gestural movement, consists of several possible “phases”: preparation, stroke, stroke hold, hold, and retraction (Kendon 2004; McNeill 2005; Kita 1990). The preparation phase refers to the phase when the hand rises from the rest position to the position for starting the stroke. The stroke is essential in terms of the movement structure of a gesture, involving the peak of the effort exerted (McNeill 1992). It “displays semantic content related to that of the speech with which the gesture co-occurs” (Duncan 2002: 193). It may be synchronous with co-expressive speech or slightly precede the speech with which it links semantically. Stroke holds are “independent holds” and relatively static, which stand on their own. The hold is the phase in which the hands maintain the stroke’s final position and form motionlessly. However, sometimes beats rhythmically oscillate the hold with the prosody of the speech that was co-expressive with the gesture stroke or stroke hold. Retraction is the phase in which the hands return to rest at the place where they started or some other relaxed position. Only the stroke or stroke hold phase is obligatory to constitute a gesture, and the other phases are all optional.

The iconic (concrete or metaphoric) gestures accompanying the progressive *zài* or the durative *zhe* marked utterances were annotated in this study. In concrete iconic gestures, the physical entities, actions, and spatial properties being mentioned in speech are schematically mapped out by the shape and/or motion of the hands with some degree of iconicity. For example, while uttering *She walked along the path*, the speaker may move her right extended finger from left to right in front as if drawing out the path.
she actually took. A metaphoric iconic gesture represents an abstract concept (i.e., the abstract entities, processes, and properties) by iconically depicting some physical referent in terms of which the abstract concept can be understood (Cienki 1998a, b, 2008a, b). For example, while talking about honesty or honest people with a verbal metaphoric expression (e.g., she's straight as an arrow) or even without it, the speaker may make a gesture with a tense, open, vertically flat (straight) hand, thumb at the top to iconically depict the abstract concept HONEST as STRAIGHT (Cienki 1998a, b).

Such iconic gestures were annotated in detail in terms of gesture phases. Then we coded the property of the gesture stroke or stroke hold as dynamic or static. When the gesture stroke was in continuous motion, representing the action (concrete or abstract) involved in an event, it was coded as “dynamic”. When the gesture stroke or stroke hold was sustained, and even when there were biphasic beats added, it was coded as “static”. This was done because the beats do not represent action-related meaning in speech, but they are only important for the timing of vocalizations (Bernard et al. 2015). Two coders coded all 88 gestures. The agreement for gesture feature (dynamic versus static) was 97%.

5.3 Results

According to Table 5.1 below, you can see that almost all of the strokes of gestures accompanying the progressive zài are dynamic, which is far more frequently dynamic than those with the durative zhe utterances (93.75% vs. 55%). The strokes of gestures co-occurring with the durative zhe tend to be a little more dynamic than static (55% vs. 45%). The differences are significant ($\chi^2_{(1)} = 18.032$, $p < 0.05$). The results suggest that gestures accompanying the progressive aspect do represent its aspeсtual meaning, that is, dynamicity, and those co-occurring with the durative aspect encode either dynamicity or stativity. The gestural data are consistent with the linguistic evidence that the progressive aspect zài encodes “progressivity” or “ongoingness” (e.g., Comrie 1976), and “dynamicity” of dynamic events (e.g., Smith 1991; Xiao & McEnery 2004;
Dai 1997), and that the durative aspect zhe can encode not only “stativity” of static events but also “dynamicity” of dynamic events (e.g., Lu 1999; Wang 2012). However, the gestural result also shows that there is no predominant use of durative zhe to encode the durative states rather than the dynamic events, which the previous linguistic studies emphasized.

**Table 5.1:** The frequencies of gesture feature with the imperfective aspect-marked utterances

<table>
<thead>
<tr>
<th>Imperfective aspect</th>
<th>Gesture stroke feature</th>
<th>Dynamic</th>
<th>Static</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Progressive zài</td>
<td>Dynamic</td>
<td>45</td>
<td>93.75</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Static</td>
<td>22</td>
<td>55</td>
<td>18</td>
</tr>
<tr>
<td>Durative -zhe</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dynamic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Static</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>67</td>
<td>76.14</td>
<td>21</td>
</tr>
</tbody>
</table>

Keeping the quantitative results into mind, I will now analyze the qualitative data to see the detailed contexts in which the dynamic and static gestures occurred for two imperfective aspect-marked utterances in the following sections. In addition, I also discuss some revealing examples about how co-speech gestures contribute to sentence meaning.

**5.3.1 Dynamicity and the gestures with progressive aspect-marked utterances**

The strokes of gestures accompanying progressive zài-framed utterances were prototypically dynamic, which explicitly represented the ongoingness of the relevant progressive aspect-framed actions. Although the handshape, orientation, movement quality etc. were varied, the gesture strokes shared the schematic feature of dynamicity with energetic motion for either concrete or abstract events.

The iconic gestures are argued to be constructed via different modes of representation or techniques of representation (Kendon 2004; Müller 1998, 2014, see Chapter 2 for detailed review). For concrete events, there are a lot of progressive aspect-
framed utterances, allowing for layered aspectual expression in natural conversations. That is, while the verb predicate is used with progressive aspect, other elements, adverbials in particular, coerce the whole utterance to be imperfective aspect, but not progressive aspect any more. In such cases, the most salient features of actions encoded by the verb predicates were still profiled by the co-speech gestures, such as in example (1a) and (1b). In (1a), the speaker recounted that his son worked very hard until 2 a.m. or 3 a.m. every night, writing and drawing. In speech, the verb predicate zài xiě a huà a (was writing and drawing) expresses the ongoingness of the dynamic writing and drawing events. However, the adverbial měi tiān wǎn shàng (every night) indicates that the whole utterance denotes habitual aspect. It means that the event of his son writing and drawing until 2 a.m. or 3 a.m. occurred repeatedly during the film making. Such dynamics and iterativity have been mapped onto the gesture. As presented in Fig. 5.1a, his index finger touching upon the thumb of his right hand moved horizontally from the left to the right in a zigzag way as if holding a pen or pencil and reenacting writing or drawing with it three times while uttering dōu zài xiě a huà a (he was even writing or drawing).

(1) a. [[[wǒ měi tiān wǎn shàng kàn dào][tā liǎng][sān diǎn zhōng]
   tā] [dōu zài <0.5s> dōu zài xiě a huà a]]
   “I saw that he was even writing and drawing every night at 2 a.m. or 3 a.m.”

   b. [[[tā hái zài pá, cái][qǐ-gè duō yuè], [tā zài pá]]
   “He was still in the crawling phase. Because he was only more than seven months old, he was crawling.”
Before the utterance in (1b), the speaker was narrating the situation where her son was playing near a large pond in their garden. His son was too little to walk. It was dangerous that he crawled to the side of the pond. The verb predicate of the first sentence zài pá (be crawling) in (1b) encodes progressive aspect, but the adverbial hái (still) shows that the aspectual type of the whole sentence is imperfective, i.e., state (non-progressive) instead of progressive. As Fig. 5.1b shows, the speaker alternately moved her two fists back and forth once while uttering tā hái zài pá (he was still in the phase of crawling), and then held for a moment; then she made the same gesture again while uttering the second utterance even without a lexical affiliate for pá (crawl), and she repeated the gesture again for the third progressive utterance tā zài pá (he was crawling). Hence, the gesture illustrates the crawling behavior that is typical of this phase of a seven-month-old child, but not the aspectual type of being in this phase, which is static. The actual action of writing or drawing on an imagined piece of paper with a pen, and the characteristics of crawling on the ground, were presumably simulated in the speakers’ mind and was reenacted vividly by the hand gestures. The dynamic and iterated gesture motion visually demonstrated not only that such actions were ongoing, but also how the actions were unfolding with time. They can be seen not only as a reflection of the mental image of the speakers’ immediate thinking, but also as a stimulus for activating the listeners’ mental image of such actions.

A concrete gestural image can provide the substrate for metaphorical depiction of an abstract event (Cienki 1998a, b; Cienki & Müller 2008a, b). Thus, for abstract events
framed by the progressive aspect, co-speech gestures occurring with them were iconic for physical actions being used to describe metaphoric events, as in example (2a), or iconic for the schematic process of ongoing inherently abstract events, as in example (2b).

(2) a. [zhè yàng de háng yě zài zhú jiàn de kāifāng].
   such GEN field also PROG gradual PRT open
   “Such fields are also gradually opening to the public.”

b. [tā jiù gēn wǒ yīzhí zài.. yīzhí zài gā gā gā gā]
   he then with I all the time PROG all the time PROG (onomatopoeia)
   “Then he was talking to me all the time.”

In (2a), the speaker moved her two hands with palm towards center from the center to the two sides as if opening something concrete (see Fig. 5.2a), which was synchronous with the progressive aspect construction zài zhújiàn de kāifāng (are gradually opening). The opening of the previous monopoly industries that she was discussing was simulated as if embodying something opening outward to the sides, like curtains or sliding doors. It demonstrates that the actions of abstract entities can be rooted in everyday actions.

In (2b), the speaker was recounting that the director was persuading her to do something she was not willing to do. She enacted out the embarrassing situation as presented in Fig. 5.2b. She thrusts her right hand with palm towards body between the addressee and herself alternately four times through the whole utterance. She used the onomatopoeia
gā (the sound of the duck quacking to depict talking quickly and a lot) in speech, but her co-speech gestures metaphorically represented the process of talking: the opening and closing of her fingers can be interpreted as acting out the opening and closing of a mouth while talking. The schematic duration of the process of persuasion is embodied in the iterated back and forth movement of her hand. The metaphor is expressed in her gesture: COMMUNICATION IS LIKE TRANSFERING SOMETHING IN THE AIR (the CONDUIT metaphor of communication, known from Reddy [1979/1993]). The reduplication use of gā in speech was represented by the dynamic, cyclic and iterated gestures, co-expressing the continuous and ongoing events of persuasion. In addition, her closed eyes and lowered head implied that she refused to communicate with the director.

The progressive aspect-marked utterances with generic verbs such as zuò (do), jinxing (proceed), etc. refer to abstract conceptual actions that do not themselves physically exist in the world. Consequently, they have no concrete images in the speakers’ mind. Instead, they were prototypically accompanied by the back and forth gestures and occasionally by cyclic gestures (Ladewig 2011, 2014), which are known to characterize the ongoingness and duration of the inherently abstract events, as example (3a) and (3b) demonstrate, respectively.

(3) a. [shìchāng  lǎobāixìng  yě  zài  tiáozhèng]
    market  ordinary people  too  PROG  adjust
    “The ordinary people are adjusting themselves to markets too”

b. [wǒ  jué  de  shì  zài  jinxíng  yí-gē  xǐ].
   I  think  PRT  COP  PROG  doing  one-CLF  baptism.
   “I think we were going through the baptism (of tough life)”
In (3a), the speaker used the generic verb tıáozhěng (adjust) to express the abstract action people took towards the change of market. As Fig. 5.3a shows, he made a “back and forth” gesture by alternating the intense open right hand away and towards his body while uttering láobāixíng yě zài tıáozhěng (the ordinary people also are adjusting). The “back and forth” gesture may stem from the BALANCE schema. The oscillation of hand(s) embodies the imbalance of an ongoing process with a changing state. In (3b), the actress employed the generic verb jìnxíng (proceed). She made a two-handed mirroring cyclic gesture with the whole predicate by moving the hands downward-outward-inward vertically and then ending with an open hand palm up gesture with a hold for 10 seconds before the next utterance (see Fig. 5.3b). The highly schematic gestural form can be seen as abstracted from the perception of dynamic events in our lives that are cyclic in nature, which form the CYCLE schema in our minds.

In addition, there were also four gestures iconic for abstract entities (TIME) mentioned in progressive aspect utterances, as in example (4a) and (4b). Three were static and one was dynamic. In (4a), the speaker held her two hands, palms towards center in front, as if molding the length of something, as Fig. 5.4a presents. The static stroke was synchronous with the time adverb xiànzài (now), and it was held with energetic beats for emphasis when the progressive construction zài zuò (be doing) occurred. Such a gesture is very popular in the corpus when time is referred to. Thus, I consider it as a metaphoric gesture (TIME IS A CONTAINER or A LIMITED AMOUNT OF TIME IS A BOUNDED SPACE), which profiled the duration of the
general event “doing this thing” to some extent. Such a gesture has also been found to be used to teach one of the characteristics of English progressive aspect, i.e., duration, in second language classroom (Matsumoto & Dobs 2016).

(4) a. [wò xiànzài zài zuò zhègè shì].
   I now PROG do this thing
   “I’m doing this thing now.”

b. [tā yīzhí zài <0.2s> zài zhè sān sì tiān tā zài nèijiù]
   he all the time in in this three four day he PROG guilty
   “He felt guilty throughout the three or four days.”

Fig. 5.4a Gesture with “now PROG do”

Fig. 5.4b Gesture of “PROG feel guilty”

Furthermore, the mental activity verb “nèijiù a” (feel guilty) is also not affordable for gestural representation. As shown in (4b) and Fig. 5.4b, the speaker moved his right hand to the upper center with the extended thumb and index finger and the other fingers curled into the palm as preparation during the ending of the last utterance and starting of this utterance, then he moved it quickly from the center to the right periphery as he uttered tā yīzhí zài (he all the time), and then the gesture held with beats for the lexical affiliates sān sì tiān (three or four days) but kept static for the progressive construction zài nèijiù (feel guilty) until the end of this utterance. Such a vertically paralleling thumb and extended index finger representing a scale can also metaphorically stand for the abstract concept PERIOD OF TIME (three or four days), as the diametrically opposed and separated pads of the thumb and index finger are often used to indicate an extent
of something (Hassemer 2015). Although this gesture stroke was also dynamic, it is
different from the other dynamic gestures accompanying the progressive aspect
utterances discussed above. This dynamic gesture may imply the continuous flowing of
time or maybe it just helped bring the idea forward to the addressee. In a word, the
schematic verbs used in speech were correlated with the speakers’ way of construing
events in gestures; the metaphoric gestures for time can be seen as reifying the abstract
actions or just the period of time the actions take.

According to the above qualitative analyses, I can conclude that the prototypical
gestures accompanying the progressive zài-framed utterances dynamically reenact out
the physical actions of concrete or metaphorical events, even when the layered aspectual
type is imperfective like habitual or state rather than progressive aspect any longer. The
“back and forth” gestures and cyclic gestures represent the schematic process of
abstract events. Moreover, the peripheral gestures accompanying the progressive zài-
framed utterances statically mold the temporal period for which the abstract event lasted.
Thus, the ongoingness and duration of the progressive aspect-framed events are
embodied by their dynamic co-speech gestures.

5.3.2 Stativity/Dynamicity and the gestures with durative aspect-marked
utterances

Now I will turn to the features of gesture strokes of the durative aspect-marked
utterances. The qualitative analyses will present a panoramic picture of the linguistic
contexts in which the static and the dynamic gestures are accompanying the durative
aspect-marked utterances.

5.3.2.1 Static durativity

Among the 40 instances, there were 18 cases with a static stroke hold. The verbs
inherently involve long duration: cáng (hide), chéngzài (carry), ná (pick up, hold, carry),
bào (hold or carry), tí (carry), zhà (open), chān (support), kàn (watch), and two
adjectives indicated the acquired states: pàng (fat) and zhòng (swollen). The manner
verbs have ambiguous uses. On the one hand, they can denote movements leading to a certain configuration. On the other hand, they may denote just the resulting state. For example, the verb *ná* either denotes a dynamic action in *tā ná qí yī běn shū* (She picks up a book), or denotes the resulting state *tā ná zhe yī běn shū* (She is holding a book in her hand). When such manner verbs co-occur with the durative *zhe*, the co-speech gestures are characterized by static stroke hold with long duration.

Most of the static stroke holds enacted out the maintenance of one portion of the actual actions as in examples (5a) and (5b), some added complementary action as in example (5c), and those with adjectives demonstrated the resultant state as in example (5d). When the speaker was recounting that his daughter was crying at the beginning of a live TV show, he needed to hold her in his arms at that time in (5a). As shown in Fig. 5.5a, he kept his two crossed hands in front of his chest as if holding something in his arms, which was synchronous with the durative construction *bào zhe tā* (holding her).

\[(5)\text{ a. } [qíshí wǒ jiù //bào zhe tā]\
\text{"Actually, I was holding her in my arms at that time."}\
\]

\[(5)\text{ b. } [Yīmóu tā //yòngyuǎn shì bào zhe jǐqi]\]
\text{"Yimou was holding the machine forever."}\

**Fig. 5.5a** Gesture with “hold DUR her”

**Fig. 5.5b** Gesture with “hold DUR machine”
In (5b) and Fig. 5.5b, the speaker recounted that the director Yìmóu was afraid of breaking the recording machine, so he was always holding it by himself even when climbing mountains. The speaker held each hand with palm lateral towards center at his side, which is the apex of the holding action and synchronous with yǒngyuàn shì bào zhe jīqì (was holding machine forever). The objects for holding in (5a) and (5b) were different, so the locations, handshapes of gestures differed from each other. This is also the case for other frequently used verbs like nà (hold, carry). However, the common characteristics for them are that the stroke hold was relatively static (except on some occasions when there were beats added to it for emphasis) and often maintained for the whole verb predicates, several utterances following or even until the ending of the whole turn.

(5) c. [wǒ zhè //yī biǎn kàn zhe diànhì, yǐbìān gěi tā hǎn zhe wù sì]
I this one side watch DUR TV, one side help her shout DUR five four]

“While I was watching TV, I was counting down ‘five, four’ with her.”

d. [zhēng tiān dà yǎnpāo zhōng zhe]
whole day big upper eyelid swell DUR

“My upper eyelids were swollen in the whole day.”

In (5c), the speaker was narrating the situation when he was waiting for his wife to give birth to his son in the hospital. He moved his right fist from the rest position over his shoulder, which lacked a lexical affiliate in speech. As Fig. 5.5c displays, this
stroke hold was actually iconic for holding a TV remote control in his hand. His eye
gaze moved away from the addressee upwards and stared at an imagined TV like one
may find overhead in public places, which was synchronous with the hand gesture and
also with the *yī biān kàn zhe diànnǐ* (was watching TV). Then he lowered his head and
counted down “five, four” when speaking the next utterance with durative aspect
construction *yībiān gēi tā hàn zhe wù sì* (was counting down: five, four…). During the
whole process, the right hand gesture stayed in place, encoding the static duration of
the action of holding TV remote control. The hand gesture, eye gaze and body
movement encoded things like the visual and spatial perspective/orientation of the
speaker (e.g., looking upward) and added a lot of information, which could not be
interpreted from the verbal information alone. The total meaning of such a multimodal
utterance is of great potential significance for communication. The static gesture
starting before the durative aspect-marked utterances and being held through the whole
situation instantiates the long maintenance and duration of the event.

In (5d), the speaker was narrating the results of his weight loss by drinking a lot
of water without eating something else. The resultative state is demonstrated by the
gesture in Fig. 5.5d. He put his open right hand with palm towards body over his right
eye to mold (as if touching) the big and swollen eye, which was synchronous with the
whole durative aspect-marked speech. Such a static stroke hold does not represent the
nucleus of an action as those in (5a) to (5c), but profiles the duration of the resultant
state. This is consistent with previous linguistic results indicating that the durative
aspect *zhe* has a resultative state reading (e.g., Smith 1991).

When taking a closer look at the static gestures with the durative aspect-marked
utterances, I found that the stroke hold usually preceded the lexical affiliate and
occurred with the emergence of the topic, and that stroke holds tended to last over at
least the clause the lexical affiliate was embedded in and even over several clauses until
the turn ended. The long static duration of gestures with spoken durative aspect would
seem to characterize the focus on the nucleus of an action in an event, perhaps “to
facilitate the greater complexity of depiction” of some simultaneous events (Duncan
5.3.2.2 Dynamic durativity

As Table 1 shows, 55% of the gestures accompanying durative aspect speech exhibited the kind of dynamicity found with the progressive aspect. The dynamic gestures accompanying imperfective durative aspect speech mainly occurred in the following three typical linguistic contexts.

First, they occurred with verbs that inherently express strong dynamicity. When the verbs in speech were highly dynamic, the gestures with the durative aspect were always as dynamic as those with the progressive aspect in speech. Consider (6a) and (6b), for instance. The same verb tuīdòng (push) was used with the two different aspect categories in metaphoric events with similar semantic purpose. Both of their co-speech gestures dynamically reenacted the action of pushing something as if in actual situations. In (6a) and Fig. 5.6a, the speaker in the same interview also moved his left hand with palm towards body away and towards his body alternately three times to reenact pushing the abstract “e-commerce” as pushing some concrete things. Similarly, in (6b) and Fig. 5.6b, the other guest moved his palm-down right hand downward with diminishing repetitions to act out pushing the abstract “e-commerce” as if pushing some physical object. Therefore, the gestures revealed that the durative aspect-marked event of pushing encoded the “ongoingness” and “dynamicity” as the progressive aspect-marked event of pushing did. In addition, the interlocutors demonstrated their “interactive alignment” both lexically (with the verb tuīdòng) and gesturally (with acting mode of repeated gestures) (Oben & Brône 2016: 32).

(6) a. [zhèyàng de shìqíng shì tuīdòng zhe zhōngguó de zhègè diànzǐ shāngwù.] such GEN thing COP push DUR China GEN this electronic commerce
   “Such things were pushing the Chinese e-commerce.”

b. [[yǒu hěnduō yě yǒu zhèngfū...(0.6s)] [zài tuīdòng nèixī diàn]]zǐ shāngwù.
   have many also have government PROG push those electronic commerce.
   “There were many..also governments which were pushing those e-commerces.”
Second, when the monosyllabic action verbs such as bào (hold), chāo (copy), chī (eat), hē (drink) are followed by zhe and reduplicated as “V zhe V zhe”, the construction conveys that while the event denoted by the verb was ongoing, another new event encoded by the following clause or utterance occurs. In example (7a), the speaker was narrating the situation where he needed to throw a baby onto the ground for a film scene. Of course, he used a doll instead, but he was still very nervous. At one point, he thought he felt the doll move, as if it had been a real baby. Thus, he made a gesture in which his two hands with palm towards body shook in front of him, as Fig. 5.7a presents. The dynamic gesture was synchronous with the whole utterance with reduplicated durative aspect and the following clause. In speech, the durative aspect-framed event bào zhe (holding the doll in his arms) is static, and the second event dòng (move) is dynamic. The dynamicity of the new event was integrated into the gesture accompanying the durative utterance here. Hence, the co-speech gesture combined the duration of the holding action and the moving action, profiling the simultaneous occurrence of the two actions.

(7) a. [wǒ bào zhe bào zhe, jiǎ de jiù jué de dòng]
I hold DUR hold DUR, fake PRT then feel PRT move.
“When I was holding the doll, I felt it moved.”


(7) b. [wǒ ná zhe máojīn níng.]

I hold DUR towel twist.

“While holding the towel, I twisted it.”

c. [wǒmen jiù ná zhe gě zhī bēi]

we then hold DUR CLF paper cup
dào le hěnduō xiāngbīn, bā bā bā bā]
pour ACTL much Champagne, [onomatopoeia].

“Then while holding a paper cup, we poured much Champagne.”

Third, the nà zhe (NP)V (hold and do something) construction was also usually accompanied by dynamic gestures to profile the duration of the second verb action while the action of “holding” was backgrounded. When two events occur
simultaneously, the durative aspect is likely to occur, because it functions to background events in discourse (e.g., Duncan 2002). Co-speech gestures with such utterances also represented this characteristic, as in example (7c) and (7d). As example (7c) and Fig. 5.7c displays, the speaker rotated his two fists simultaneously three times as if performing the action of holding something in his hands and twisting it, which was synchronous with the whole durative aspectual utterance. The salient motion in the gesture foregrounded the twisting event, and the fists backgrounded the event of holding a towel horizontally. In contrast, in example (7d), the speaker made a very complex gesture in contrasting handshape, space and movement quality for representing the two simultaneous events. The gesture involves a combination of stativity in one hand and dynamicity in the other hand. As Fig. 5.7d shows, his left hand with the index finger touching the thumb and the other fingers slightly curled represented the “paper cup” and its static maintenance backgrounded the long duration of the events of nà zhe gè zhī bēi (was holding a paper cup). In comparison, his right fist, which enacted the holding of a round wine bottle, moved cyclically while he uttered dào le hěnduō xiāngbīn (poured much Champagne) and changed to iterated back and forth motion while he uttered bā bā bā bā (using the onomatopoeia to simulate the sound of pouring wine). The dynamic right hand gesture foregrounded the event of pouring Champagne, which revealed that the speaker poured much Champagne into many different paper cups for a lot of people there, rather than poured a lot of Champagne into a single paper cup. Apparently, the speaker provided gestural repair of a performance error in the speech. The speech and gesture form a composite (compare Enfield 2009 on composite utterances) to express the total meaning of the events, which go far beyond the speech alone.

To recapitulate: when a durative aspect-framed utterance expressing long duration or resultant state is framed by the durative aspect, its co-speech gesture tends to keep static and be held in place to represent stativity of a durative event; when the durative aspect frames such constructions as V zhe NP (verbs with high dynamicity), V zhe V zhe (V is reduplicated) , and ná zhe V (hold/carry zhe V), the co-speech gestures are prototypically dynamic, like those in the progressive aspect-framed constructions.
5.4 Discussion and conclusions

When describing events with progressive aspect, almost all native Chinese speakers in the present dataset produced iconic gestures that were highly dynamic. They represented the finer-grained details of the actions in concrete and metaphoric events, or the schematic progress and temporal duration of the actions in inherently abstract events. When describing events with durative aspect, the Chinese speakers made static gestures with long duration to profile the maintenance of certain core features of actions in certain situations or the resultant state of certain actions. However, they tended to make dynamic gestures accompanying such durative aspect constructions as \( V \text{ zhe } NP \) (verbs with high dynamicity), \( V \text{ zhe } V \text{ zhe} \) (\( V \) is reduplicated), and \( nà \text{ zhe } V \) (hold/carry \( zhe V \)). Most of the gestures accompanying either the progressive aspect or the durative aspect speech focused on the internal core structures of events. The multimodal results provide new evidence for prototypical meanings of the two different imperfective aspect markers: the progressive aspect \( zài \) focuses on the internal stages of non-static situations and has a dynamic conceptual meaning, whereas the durative aspect \( zhe \) can focus not only on the internal stages of non-static situations but also on the external stages, like the resultant stages of static situations. It thus has dynamic as well as static conceptual meaning potentials, depending on the linguistic contexts in which it occurs. The findings are worth discussing in terms of the following two points.

5.4.1 Pure function of “\( zài V \)” construction vs. Multi-functions of “\( V \text{ zhe} \)”

A lot of grammar book and other researchers have pointed out that the functions of two morphological forms –\( zài \) and \( zhe \)– are the same when it comes to expressing progressive aspect. Both of them denote that actions are ongoing (e.g. Chao 1979; Wang 2011, 2012; Zhu 1982). However, our multimodal data reveal that there are clear divisions of labor between \( zài \) and \( zhe \), though they also have some overlapping functions. Both the linguistic and gestural evidence showed that the progressive aspect construction \( zài V \) has a single function of denoting an ongoing event. It profiles the
dynamic internal structure of an event. Such high dynamicity in conceptualization requires the use of activity verbs expressing strong action in speech and motivates the continuously dynamic movement in the co-speech gestures. In contrast, the $V$ zhe construction has multiple functions, resulting in the complexity of its linguistic and gestural behaviors.

Firstly, when it co-occurs with the action verbs with strong dynamicity, the function of $V$ zhe overlaps with that of $zài$ $V$, focusing on the dynamic internal structure of an event and denoting its on-goingness without interruption. Accordingly, co-speech gesture produced with it also displays high dynamicity.

Secondly, the $V$ zhe construction has the function of “zooming in” and focusing on the “frozen part” of the internal structure of an event, such as bào zhe (keep holding/carrying something). Meanwhile, the static gestures accompanying the zhe-framed utterances in such circumstances were usually sustained for a very long time, even extending over several following utterances. The stativity of such gestures may provide a possible explanation for the reason why Chinese linguists termed zhe a durative aspect particle, though both the $zài$- and zhe-framed utterances encode durative situations.

Thirdly, the $V$ zhe construction has the function of focusing on the external stage of an event – the resultant state beyond the completion of an event, such as the final state of being fat or swollen. This is consistent with the recent empirical results of research on mental simulation in language comprehension that English progressive aspect not only highlights the internal structure of an event, but also the final physical state (Liu & Bergen 2016: 193). Such static gestures co-occurring with the durative zhe aspect in such situations may be motivated by the mental simulation of resultant states acquired from a dynamic event. This would be consistent with the attention phenomenon of “endpoint focus”, known from the early work in cognitive semantics (e.g., Lakoff 1988).

Fourthly, the $V$ zhe construction has a backgrounding function, so the maintenance of action is usually embedded in the simultaneous ongoing event. Accordingly, the gestures accompanying zhe aspect-marked utterances in such cases had the dynamic
features found also in those with zài aspect-marked utterances. However, they had some different motivations behind them. In most cases, the non-stativity (i.e. dynamicity) and continuousness of the second event rather than the zhe-framed event in the serial verb constructions as ná zhe V motivated the dynamic expression of such complex events, realized in the gestures. The durativity of the zhe-framed event was backgrounded in the face of dynamicity of the following event. This revealed that when two events occurred simultaneously, the function of V zhe, like ná zhe V, was to provide the background information to facilitate foregrounding of the second event (Duncan 2002; Li 2014).

Smith (1991: 363) proposed that “zhe is changing from a stative imperfective to a general imperfective, gradually taking over the function of zài”. We do not agree with this proposition. It seems to be more reasonable to infer that the expansion of zhe’s functions promoted the occurrence of zài as a progressive aspect marker. Diachronically, only zhe has long been grammatically used to represent the events that were in progress (Zuo 2007); gradually, its semantic function extended to focus on the resultant state, and even the modal domain such as the use in imperative contexts (Chao 1979). Thus, the functional indeterminacy of the V zhe construction may have triggered the development of the zài V construction to focus on the pure progressivity of an event.

5.4.2 Embodied grammatical meaning

As an “aspect-prominent” language, Mandarin Chinese has grammaticalized aspektual distinctions to mark dynamic durative events and static durative events. This study revealed that the grammatical meanings of the imperfective aspectual constructions are also externally embodied in co-speech gestures. The gestures accompanying the progressive aspect-marked utterances dynamically enacted out the durative actions of the concrete and metaphoric events. In addition, the “back and forth” or cyclic gestures accompanying the progressive aspect-marked utterances also dynamically enacted out the schematic feature of an inherently abstract event in progress. The static durative events were grammatically marked by the imperfective aspect marker zhe and
accompanied by the static gestures that were maintained for a long time. When *zhe* encoded dynamic durative events, the accompanying gestures also displayed high dynamicity. The strong gestural sensitivity to the different aspectual categories and their different uses should result from the underlying mechanisms by which such grammatical categories affect our thinking for speaking and gesturing (Cienki & Müller 2008b; Slobin 1996). The schematic meanings of dynamicity and stativity manufactured by the gestures are an externalization or visualization of the speakers’ conceptualizations (i.e. mental structures and processes) of events (McNeill 1992). Moreover, dynamicity and stativity represented by the gestures derive from the ways in which hands make practical actions while experiencing concrete events in the world, as Streeck (2009) has claimed. The proposal here is that the schematic meaning (dynamicity or stativity) residing in the construal of events affects what grammatical aspectual perspective the speakers take and how the co-speech gestures are produced, though what the gestures act out is determined by the specific conceptual content inherent in the content verbs. This is also the case in the mental simulation in language comprehension: the lexicon determines what to mentally simulate, whereas grammatical categories, such as aspect, modulate how one simulates events that were described verbally (Bergen & Wheeler 2010).

Our findings provide empirical evidence from the perspective of grammatical conceptualization for the proposal that speech and gesture production are coordinated, forming an integrated system (e.g., McNeill 1992, 2005; Kendon 2004). The aspect-marked linguistic forms and their co-speech gestures were co-equal and co-expressive of the same aspectual information in different events. Temporally, all the strokes or stroke holds of the co-speech gestures overlapped with the imperfective aspect constructions, though some of them occurred before them. Semantically, the gestures accompanying the imperfective aspect markers *zài* also presented the same or closely related aspectual meanings in speech. For the same morphological item *zhe*, when it occurred in different linguistic contexts expressing different aspectual meanings, the gestural behavior adjusted with it accordingly. The same underlying conceptual potentials of the imperfective constructions (i.e. dynamicity or stativity) can be argued
to have directed the speakers’ speaking and gesturing. The semantic synchronization between the two forms of imperfective aspect-marked speech and their co-occurring gestures can be seen as orchestrated by the single integrated speech-gesture system.

6.1 Introduction

Polysemy refers to “the association of two or more related senses with a single linguistic form” (Taylor 1995: 99). As a ubiquitous phenomenon in natural language and “the pivot of semantic analysis” (Ullmann 1951: 117), polysemy has gained wide attention in Cognitive Linguistics. The prototype approach to word meaning in Cognitive Linguistics proposed that the senses of a polysemous word center around a prototype and are systematically interrelated by a family resemblance structure, and that they are organized as a schematic network via links of schematization and extension (e.g., Lakoff 1987; Lakoff & Johnson 1980; Langacker 1987, 1991a, 1991b). Many studies found that polysemy involves the extension of a central meaning to other non-central meanings by means of basic cognitive mechanisms such as metaphor, metonymy, image schema, proposition, generalization, and specialization etc. (e.g., Brugman & Lakoff 1988; Johnson 1987; Lakoff 1987). According to Cognitive Linguistics, meaning is equated with conceptualization, and grammar is meaningful by nature (Langacker 1987, 1991a, 1991b, 2008a). Like lexical items, grammatical items are frequently polysemous. However, their meanings are often abstract and reside in our conceptualization (Langacker 2008b: 16). As the only explicitly marked aspecual form in English, the progressive construction (be V-ing) has developed as a polysemous construction, resulting in a rather complex semantic network (Bybee et al. 1994; Dahl 1985; De Wit & Brisard 2014; Langacker 1987, 2001, 2008a).

The use of the English progressive construction has temporal and modal motivations. While the aspecto-temporal uses of the progressive construction have the
semantic potentials of expressing current ongoingness, habituality, iteration, and temporary state, its modal (i.e., non-aspecto-temporal, subjective) uses give rise to the futurate reading or the epistemic readings of surprise, tentativeness and intensification (see section 2.2.1 in Chapter 2 for details). As the prototypical use of the progressive construction, the progressive aspect is a grammatical device for “zooming in” and taking an “internal perspective” on an event, as if the conceptualizer were watching it unfold sequentially rather than viewing it holistically as a whole (Comrie 1976; Langacker 1991; Langacker 2001; Smith 1997). Grammatical aspect categories not only depict grammatical features but also mirror conceptual structures, and hence have psychological reality. Psychological studies have shown that the English progressive aspect plays a central role in mental simulation. It facilitates language comprehenders’ mental simulation of actions and locations, producing a large Action-sentence Compatibility Effect and Location-sentence Compatibility Effect (Bergen & Wheeler 2010; Glenberg & Kaschak 2002; Liu & Bergen 2016). The English progressive aspect also activates language comprehenders to conceptualize more action in a given time (Huette et al. 2012; Madden & Zwaan 2003; Matlock 2010, 2011), and on the part of speakers, its use correlates with the production of longer and more complex gestures (Duncan 2002; Parrill et al. 2013). The previous studies only center on the “continuous ongoingness” of progressive aspect uses in motion events. It is unclear whether and how the other senses of the progressive construction affect speakers’ mental simulation in natural conversations.

Speech and co-speech gestures have been claimed to be an integrated system (e.g., Kendon 2004; McNeill 1992, 2005). If so, one might expect that the gestures accompanying utterances with different meanings of the progressive construction should be distinct from each other. The GSA framework proposes that gestures derive from internal motor or visual simulation of action (Hostetter & Alibali 2008). Representational gestures – representing concrete or abstract objects, actions, and ideas – are regarded as “outward manifestations of the action components involved in mental simulations of spatial and motor events” (Hostetter 2014: 1469). The meaning of the speaker’s message is “externally embodied in representational gestures” (Marghetis &
Bergen 2014: 2000). Therefore, this chapter is interested in investigating to what extent and how the different meanings of the progressive construction are externally embodied in gestures. The specific questions are as follows.

1) To what degree is the polysemous progressive construction in modern spoken English multimodal?

2) What are the prototypical gestures accompanying the different meanings of the progressive construction?

3) What are the gestural characteristics of iconic gestures accompanying the English progressive aspect?

Answering these questions will indicate whether the different functions of the progressive aspect in English have different multimodal manifestations or not. If so, gesture can provide insight into how the mental simulations behind the different functions differ – an important point if we take the cognitive linguistic approach to meaning as conceptualization.

6.2 Methods

6.2.1 Corpus

The data used in this research come from “The Ellen DeGeneres Show”. I had access to the data via the UCLA Library Broadcast NewsScape via the Distributed Little Red Hen Database, co-directed by Francis Steen and Mark Turner. In order to extract the utterances with a “be V-ing” construction only from The Ellen DeGeneres Show, I created a subcorpus—ellen degeneres show— in CQPweb, a web-based corpus analysis system combining power, flexibility and usability (Hardie 2012). Then I carried out a standard query in the form of “{be/V} (_RB)? _VBG”, developed by Peter Uhrig. Since this cannot extract out the abbreviated forms like “s’/re V-ing”, I obtained 77,545 matches in 1,638 different texts. Then I exported them into Excel and manually eliminated the instances which have no progressive “be V-ing” constructions as in (1a), the equative constructions containing main verb BE and gerund as in (1b), and the -ing
forms used as predicative adjectives as in (1c). I have also left out all the *be going to* constructions as in (1d), because it has been grammaticalized into the expression of future tense.

(1) a. His new movie is “The Hateful Eight”. (Ellen Show #913)
   b. And one of my favorite things is waiting for the show night to see what flowers Chris will send her. (Ellen Show #138).
   c. The end result is something that is very captivating (Ellen Show # 294)
   d. While you look at how cute it is, I am going to take a nap. (Ellen Show # 1)

In addition, I also eliminated those constructions with progressive forms from commercial advertisements, song scripts, the breaking news report, the game unit and when the speakers were invisible, because they all involved different kinds of demonstrations, showing other images, or real actions without equal opportunities for gestural expression.

As gesture coding is laborious and time-consuming work, I just examined the first 200 progressive utterances made by the guests and the hostess respectively in the interview part, leaving out all the data from monologue and game parts. Thus, the corpus consists of 400 progressive utterances in total. As Ellen has a different role in the interaction from the guests, participating in each episode, I separated her data from that of guests for analyses to avoid individual bias. The 400 progressive utterances are from the interview between Ellen and guests of 35 episodes featured from November 10, 2015 to January 19, 2016. Each interview lasts for about 20 minutes, so the whole corpus consists of about 700 minutes of material.

### 6.2.2 Speech coding

In terms of the meanings of the progressive construction (see section 2.1.2.1 in Chapter 2 for details), each utterance was categorized into “continuous ongoingness”, “habituality”, “repetition”, “temporary state”, “futurity” or “epistemicity”. Two coders...
annotated all 400 utterances. The agreement for the meaning annotation was 92% (N = 400).

6.2.3 Gesture coding

Each progressive utterance was coded for whether it occurred with gesture or without gesture. Based on the previous studies on gesture types (Kendon 2004; McNeill 1992, 2005), each gesture was annotated as deictic, concrete iconic, metaphoric iconic or pragmatic. Twenty percent of the progressive utterances with gestures (Total N = 239) were coded by the second rater. Then the iconic gestures (concrete and metaphoric) were flagged out, and further annotated (N = 60). In terms of the semantic elements it was iconic to, each iconic or metaphoric gesture was coded for iconicity to entity, action, path of motion, or abstract process involved in an event. There were five utterances that were accompanied by multiple gestures; I analyzed only the gesture that was synchronous with the progressive construction. All of the iconic gestures were coded by two raters according to the coding manual. Inter-rater reliability of this judgment across two independent observers was 93% for gesture type, and 97% for the semantic elements the iconic gestures are iconic to.

6.3 Results

I report the results of the guests as a community and the hostess ELLEN as an individual in separate ways as follows. To find out the potential multimodality of different meanings of the progressive construction, I compared the frequency distributions of the gestures accompanying them. Then to investigate the role the gestures play, I compared the frequency distributions of their functions.

6.3.1 Frequency of actual usage and the multimodality of the progressive construction in spoken English

Table 6.1 shows the distribution of gestures in the different usage types of the
progressive construction in our dataset. Overall, the aspecto-temporal use dominates in the conversation context, while the non-aspectual use is relatively not frequent (79.5% vs. 20.5%). More than one third of the utterances with the progressive construction triggered the “continuous ongoingness” interpretation (37.0%), which is the prototypical meaning of this construction. This result is consistent with previous studies of the progressive construction uses in an oral corpus or a mixed corpus of both spoken and written data (De Wit & Brisard 2014; Rompaey 2014). “Temporary state” use occurred even more frequently than the “habituality” reading (25.5% vs. 17.0%). There was no use of “repetition” with be V-ing, a use which was also found to be marginal in other studies (De Wit & Brisard 2014; Rompaey 2014). The non-aspectual use of be V-ing includes almost twice as many instances of “emotivity” (13.0%) as “futurity” (7.5%).

### Table 6.1: Gesture use for the meanings of the progressive construction by GUESTS

<table>
<thead>
<tr>
<th>Meanings</th>
<th>With gesture</th>
<th>Without gesture</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Ongoingness</td>
<td>51</td>
<td>25.5</td>
<td>23</td>
</tr>
<tr>
<td>Habituality</td>
<td>20</td>
<td>10.0</td>
<td>14</td>
</tr>
<tr>
<td>Repetition</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Temporary State</td>
<td>27</td>
<td>13.5</td>
<td>24</td>
</tr>
<tr>
<td>Futurity</td>
<td>10</td>
<td>5.0</td>
<td>5</td>
</tr>
<tr>
<td>Emotivity</td>
<td>15</td>
<td>7.5</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>123</td>
<td>61.5</td>
<td>77</td>
</tr>
</tbody>
</table>

Generally speaking, the progressive construction shows much greater potential to be multimodal than mono-modal (61.5% with gesture vs. 38.5% without gesture). No matter what the meaning of the progressive construction is, the utterances accompanied by gestures are more frequent than those without gestures. The multimodality of the utterances denoting “continuous ongoingness” was the highest (25.5%), and that of the “futurity” was the lowest (5.0%). However, the Chi-square result showed no significant differences of gesture use among the progressive construction with different meanings ($\chi^2(4) = 3.729, p > 0.05$) (the repetition category was excluded).

In Table 6.2, you can see that Ellen also made far more aspecto-temporal use than
non-aspectual use of the progressive construction (64% vs. 36%). However, there is almost equal use of the progressive construction with the meaning of “continuous ongoingness” (21.0%), “temporary state” (23.0%), and “emotivity” (21.5%). There are also very few progressive utterances denoting “repetition”. In contrast with the results of the guests (see Table 6.1), Ellen made far greater use of progressive utterances with non-aspectual meanings than the guests did (36% vs. 20.5%). There is less “continuous ongoingness” use of the progressive construction (21.0% vs. 37.0%). “Temporary state” use is equally high between them (23.0% for Ellen and 25.5% for the guests), and “repetition” use is scarce (1.0% and 0, respectively).

<table>
<thead>
<tr>
<th>Meanings</th>
<th>With gesture</th>
<th>Without gesture</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Ongoingness</td>
<td>27</td>
<td>13.5</td>
<td>15</td>
</tr>
<tr>
<td>Habituality</td>
<td>24</td>
<td>12.0</td>
<td>14</td>
</tr>
<tr>
<td>Repetition</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>Temporary State</td>
<td>25</td>
<td>12.5</td>
<td>21</td>
</tr>
<tr>
<td>Futurity</td>
<td>15</td>
<td>7.5</td>
<td>14</td>
</tr>
<tr>
<td>Emotivity</td>
<td>24</td>
<td>12.0</td>
<td>19</td>
</tr>
<tr>
<td>Total</td>
<td>116</td>
<td>58.0</td>
<td>84</td>
</tr>
</tbody>
</table>

In terms of gestural tendency, the progressive utterances denoting “continuous ongoingness”, “habituality” and “emotivity” are more likely to be accompanied by gesture, that is, they have greater potential to be multimodal. However, there is almost 50% chance of the progressive utterances with “temporary state” and “futurity” construal being accompanied by gestures. Overall, however, the differences between meanings and gesture rate are not significant ($\chi^2_{(4)} = 1.954, p > 0.05$).

To summarize, the progressive construction as a whole has greater potential to be multimodal than monomodal in our data set. “Continuous ongoingness” as the prototypical meaning of the progressive aspect is dominantly used most often and the most likely to be multimodal. The English progressive construction has expanded to encode “temporary state” in modern English conversations, but only half of these cases...
are accompanied by gestures. The hostess Ellen used far more progressive constructions to encode “futurity” and “emotivity” than the guests did, almost half of which were multimodal representations.

6.3.2 The characteristics of gestures accompanying the progressive construction

The characteristics of gestures accompanying the progressive construction are revealed through the general functions and specific functions (i.e., semantic elements) in the following subsections.

6.3.2.1 Frequencies of gestures accompanying the progressive construction with different meanings

Table 6.3 presents the distribution of the four gesture types in the five kinds of meanings denoted by the progressive construction. Overall, the pragmatic gesture accompanying the progressive utterances dominated in the corpus (44.7%), which is twice as much with the deictic gesture (21.1%). There are almost equal amounts of concrete iconic gesture (17.9%) and metaphoric iconic gesture (16.3%), but they are both far less frequent than pragmatic gesture. Utterances with different meanings of the progressive construction are most frequently accompanied by pragmatic gestures. However, there are almost six times and three times more concrete iconic gestures accompanying the utterances with “continuous ongoingness” construal than with “temporary state” and with “futurity” respectively (14.6% vs. 2.4% vs. 4.5%). Progressive utterances with the construal of “temporary state” are more often accompanied by metaphoric iconic gestures than those of “current ongoingness”, “emotivity” and “habituality”. The deictic gestures are relatively evenly distributed among the progressive utterances with different meanings. There are no concrete iconic gestures co-occurring with the progressive utterances denoting “habituality” and “emotivity”. There is no metaphoric iconic gesture co-occurring with the “futurity” interpretation of the progressive utterances. This result suggests that the non-aspectual use of the progressive construction (i.e., “futurity” and “emotivity”) is not likely to be accompanied by iconic
Chapter 6

gestures.

Table 6.3: Gesture types for the meanings of the progressive construction by GUESTS

<table>
<thead>
<tr>
<th>Meanings</th>
<th>Deictic</th>
<th>Iconic</th>
<th>Metaphoric</th>
<th>Pragmatic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Ongoingness</td>
<td>5</td>
<td>4.1</td>
<td>18</td>
<td>14.6</td>
<td>7</td>
</tr>
<tr>
<td>Habituality</td>
<td>4</td>
<td>3.3</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Tem. state</td>
<td>7</td>
<td>5.7</td>
<td>3</td>
<td>2.4</td>
<td>8</td>
</tr>
<tr>
<td>Futurity</td>
<td>4</td>
<td>3.3</td>
<td>1</td>
<td>4.5</td>
<td>0</td>
</tr>
<tr>
<td>Emotivity</td>
<td>6</td>
<td>4.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>26</td>
<td>21.1</td>
<td>22</td>
<td>17.9</td>
<td>20</td>
</tr>
</tbody>
</table>

* Notes: Tem is short for Temporary.

Table 6.4: Gesture types for the meanings of the progressive construction by ELLEN

<table>
<thead>
<tr>
<th>Meanings</th>
<th>Deictic</th>
<th>Iconic</th>
<th>Metaphoric</th>
<th>Pragmatic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Ongoingness</td>
<td>12</td>
<td>10.3</td>
<td>7</td>
<td>6.0</td>
<td>1</td>
</tr>
<tr>
<td>Habituality</td>
<td>8</td>
<td>6.9</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Repetition</td>
<td>1</td>
<td>0.9</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tem. state</td>
<td>8</td>
<td>6.9</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Futurity</td>
<td>8</td>
<td>6.9</td>
<td>1</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>Emotivity</td>
<td>7</td>
<td>6.0</td>
<td>1</td>
<td>0.9</td>
<td>3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>44</td>
<td>37.9</td>
<td>9</td>
<td>7.8</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 6.4 shows that pragmatic and deictic gestures (84.5% in total) dominate in the gesture use for Ellen, whereas the concrete iconic and metaphor iconic gestures relatively rarely (15.6% as a whole) co-occurred with her progressive utterances. The prototypical gesture accompanying the progressive utterances denoting “continuous ongoingness” and “futurity” is the deictic gesture, whereas the typical gesture with those denoting “habituality”, “temporary state” and “emotivity” is the pragmatic gesture.

Compared with the results of GUESTS, you find that there are twice as many progressive utterances made by the GUESTS co-occurring with concrete iconic and metaphor iconic gestures as there are made by ELLEN (34.2% vs. 15.6%). Similarly, only the progressive utterances with “continuous ongoingness” were more likely to be
accompanied by the concrete iconic gestures for both the GUESTS and ELLEN.

The findings suggest that only when the progressive construction has the “continuous ongoingness” construal can it be the most likely to co-occur with iconic gestures; otherwise, the co-speech gestures mainly have a pragmatic function related to the discourse. Thus, the progressive construction in the interview context did not effectively facilitate the native English speakers individually or as a community to focus on the event-internal structure. This is inconsistent with what previous linguists have claimed and as psychologists operating in labs have found.

6.3.2.2 Semantic elements for which the gestures with progressive utterances are iconic

As the sample of iconic (concrete and metaphoric) gestures is small (N = 60), I combined the totals from the GUESTS and ELLEN for analysis in this section. In addition, due to the rather scarce co-occurrence of iconic (concrete and metaphoric) gestures with the progressive construction denoting non-aspectual meanings by both the GUESTS and ELLEN, I left them out (N = 7) and took only the iconic and metaphoric gestures accompanying the progressive construction with temporal-aspectual use into consideration (N = 53). Consequently, there were 33 iconic (concrete and metaphoric) gestures accompanying progressive utterances with a “current ongoingness” construal, 6 accompanying those with a “habituality” construal, and 14 accompanying those with a “temporary state” construal. Fig. 6.1a presents the distributions of the semantic elements that they are iconic for. The gestures co-occurring with progressive construction with “continuous ongoingness” construal are the most iconic for the actions of events (39.4%), and then next most frequently for the path of events (24.2%). In addition, it is just the progressive construction with “continuous ongoingness” construal that is accompanied by gestures showing a combination of semantic elements (most are combinations of action and entity) involved in events (15.2%). Both the progressive utterances denoting “habituality” and “temporary state” are predominantly accompanied by gestures that are iconic for
entities participating in the events (50% for the former and 35.7% for the latter), and second most often by those that are iconic for the process of abstract events (33.3% for the former and 28.6% for the latter).

![Graph showing semantic elements of iconic gestures with progressive construction.](image)

**Fig. 6.1a** Semantic elements of the iconic gestures with progressive construction

### 6.4 Discussion

In order to examine whether, and if so, how, the polysemy of a grammatical construction correlates with multimodality, I investigated the relationship between the co-speech gestures and the progressive construction with rich meanings in natural conversation.

There are three main findings from our study. First, the utterances with a progressive (*be V-ing*) construction had greater potential to be multimodal than to be monomodal, both for the general population and for the individual. This pattern appears to stem from the fact that the progressive utterances with “continuous ongoingness” were far more likely to be accompanied by gestures.

Second, the prototypical gesture type for the progressive construction in general was pragmatic gestures. For the general population, this is the case regardless of its meanings, whereas for Ellen as an individual with a special role in the conversations,
the prototypical gesture type with the progressive construction with the construal of “habituality”, “temporary state” and “emotivity” was pragmatic gestures, and with the construal of “continuous ongoingness”, “repetition” and “futurity” it was deictic gestures.

In terms of our third finding, two pieces of evidence appear to further clarify the view that it is the use of the English progressive aspect to express “continuous ongoingness”, rather than the progressive construction in general, that facilitates the speakers’ mental simulation of the internal structure of events. First, only when the progressive utterances denote “continuous ongoingness” did the speakers produce predominantly (concrete and metaphoric) iconic gestures. Second, only the iconic gestures accompanying the progressive utterances with a “continuous ongoingness” construal was more likely to be iconic for the action, the nucleus of an event, and then to be iconic for more than two event elements.

Therefore, I will discuss the results in detail in terms of the multimodality of the polysemous progressive construction, in general, and of the progressive aspect, in particular.

6.4.1 Polysemy of the progressive construction and multimodality

The dominant use of co-speech gestures accompanying utterances with the progressive construction provides further evidence for the general claim that grammar is (at least sometimes) multimodal in nature (Fricke 2008; Harrison 2009). Grammatical meanings are abstract and schematic; accordingly, there is no particular fixed gestural form for any particular meaning denoted by the progressive construction. Thus, the grammatical construction just correlates with gesture rather than with a particular kind of gesture.

The prototypical gestures for the utterances with the polysemous progressive construction were pragmatic gestures, and deictic gestures were also frequently involved. I believe such results derive from the inherent nature of the interview as involving a particular kind of spoken language usage event (Cienki 2015), that is, ones with a high level of interaction and involvement. Interviews, as face-to-face dialogue,
have particular social as well as semantic and syntactic features (Bavelas 1990; Bavelas et al. 1992). Pragmatic gestures refer directly to the interlocutor rather than to the topic, which serves to include the interactant (Bavelas et al. 1992). The native speakers of American English in the data make systemic use of pragmatic gestures to signal and maintain interaction and involvement not only with the interlocutor in conversations but also with the audience in the studio or the audience watching the show on TV.

The forms of the pragmatic gestures are relatively simple, predominantly with open hand gestures. As Fig. 6.1b shows, more than half of the pragmatic gestures were palm-up open hand (PUOH) gestures (58.2% for the guests and 64.8% for Ellen). Almost 15% of them were palm-down open hand gestures (PDOH). The guests used three times as many palm-lateral open hand gestures (PLOH) as Ellen did. The other forms, like extended vertical index finger, palm vertical open hand gestures, were 12.8% for the guests, and 14.8% for Ellen. However, the overall differences between Ellen and the guests were not significant ($\chi^2 (3) = 2.465, p > 0.05$).

The PUOH gestures are also called “palm presenting” gestures (Kendon 2004: 273), which are assumed to be derived from the everyday action of presenting something, offering or giving something, or showing readiness to receive something via the extended open hand with palms up (Kendon 2004; Müller 2004). In our corpus, the native speakers of American English frequently made use of them to present to the
addressee or the audience what they were doing or are doing, the temporary state they were in, and attitudes or comments they had towards something as if on the empty palms of their open hands. In (2a), they were talking about the performance Gwen Stefani made in a video clip that had just been played on the screen. While explaining that it was not to be the video as Ellen said but only a live show, Stefani presented the fact of what was ongoing — “we were doing the live show” — on her open hands with palm up towards the addressee, Ellen (see Fig. 6.2a). In (2b), after seeing a picture in which the guest, Ricky Gervais, was running, Ellen used the progressive utterance with epistemic construal i.e., inferring the purpose of his running. As Fig. 6.2b shows, she extended her open hands with palms up towards the addressee, Ricky, while uttering “you are actually”, and the gesture held through “trying to get in shape”. Thus, she presents her inference via her PUOH gesture for Ricky’s agreement, which was signaled further by the next utterance “right?”. Unfortunately, Ricky denied her inference by saying “well, no.” and clarified that he worked out nearly every day to keep healthy instead of running only over a short period to keep in shape for participating in The Golden Globe Award show. Such PUOH gestures towards the addressee directly do not represent any information about the speech content but function to include the addressee to an extreme extent in interaction.

(2) a. Ellen: Because you were so raw in the video, too. You can just see how you are feeling throughout the whole—it’s just one take.

Gwen: I’m going to cry when I look at that. The thing is, that was never going be the video. That was just going to be -- [We were doing the show live], we were doing the song live for the first time in New York.

b. Ellen: [You are actually trying to get in shape for the] (..). It seems like you joke about it, but you are running, getting in shape for the globes, right?

Ricky: Well, no. Believe it or not, I work out nearly every day.
In contrast, the PUOH gestures towards the audience rather than the speaker in Fig. 6.2c and Fig. 6.2d tended to get the audience involved in their interaction. In (2c), Ellen and the guest, Charlie Puth, were talking about the process of getting in touch with Charlie and inviting him to appear on the Show. Charlie did not believe that he was able to be invited, so he took them as crank calls and hung up the phone every time. Consequently, the Ellen Show team had to look for him via his high school, etc. After Charlie’s narration of this, Ellen reidentified what they did with the progressive utterance “we were tracking you down”. Then Charlie confirmed her opinion and further reiterated it by saying “you trackers like were tracking my town down”. Meanwhile, he made a PUOH gesture towards the audience and turned his eye gaze from Ellen to the audience too (see Fig. 6.2c). Apparently, he presented the idea on his open hands with palms up to the audience to seek their agreement. Accordingly, the audience replied with laughter in return.

(2) c. Charlie: ...I’m like, “Cool, Bye”. And then you called my high school, and you called my grammar school, like, everybody.

Ellen: We were tracking you down.

Charlie: You really did. [You trackers like were tracking my town down].

d. Ellen: I’m so inspired, I’m going to do that flagpole. I’m going to do it by the end
of the season.

Kevin: [laughter] you are going to do flagpole?

Ellen: All right. [And **now** you are laughing at me. I’m going to do it.]

In (2d), Kevin was talking about getting stronger and stronger by doing flagpole exercises (i.e., a male fitness item for developing strong muscles) and tried to inspire more people to engage in this sport. Ellen responded that she was going to do it too, which stimulated Kevin Hart’s laughter and doubt about it. At the same time, the audience were clapping and laughing too. Therefore, Ellen made a progressive utterance “you are laughing at me” with the epistemic construal of re-identification of what Kevin and the audience were doing. Such a progressive utterance was accompanied by a PUOH gesture and also an eye gaze adjustment from Kevin to the audience in reply to the audience’s non-verbal involvement in the interaction too (See Fig. 6.2d).

The deictic gesture was the second most frequent gesture accompanying the progressive utterance with various kinds of meanings except those denoting “current ongoingness”. Deictic gestures are generally taken to be employed by speakers to draw other people’s attention or to establish a joint focus of attention (Bavelas et al. 2014). In the corpus, the deictic gestures are usually in the form of extended index finger, the thumb and the palm up open hand pointed at the speaker, the addressee, the audience, the non-present person or object involved in the reported events, or those in the pictures or videos on the screen. As one of the most important characteristics of face-to-face
dialogue, the pronouns, you and I in particular, were mainly the subjects of utterances. Thus, there are many progressive utterances that are accompanied by the deictic gestures of pointing at the speaker himself/herself or at the addressee to emphasize or clarify the agent of an event. The speaker also usually pointed at a certain object which was the shared common ground with the audience, as in (3a), when Ellen was questioning how the guest, Matt Lauer, could walk in the high-heeled shoes he had on. Matt explained it was not the first time he had done so and said to Ellen “you made a resolution and you came on my show”. Upon hearing this, the audience were laughed loud. Thus, Ellen reidentified the usage event that “They (the audience) are still laughing at your feet”, and she pointed at Matt’s shoe with her extended left index finger while uttering “they are still laughing at” (see Fig. 6.3a). The special female high-heeled shoes the male guest, Matt, was wearing caught the hostess’ and the audience’s attention the moment he stepped on the stage. The deictic gesture here just highlighted such shared common ground (cf. Holler & Wilkin 2009). In addition, the speaker can also make deictic gestures to point at abstract objects, such as in (3b), where Ellen asked what Leonardo DiCaprio did for Christmas. Leonardo told about the party he had with his family. In the previous gestures accompanying “they have loads of kids. Their kids had kids”, he set up the “family tree” model of family lineage in vertical tiers (successive generations below previous ones), using his open hands with vertical palm away from his body. While uttering the progressive utterance “their kids’ kids are having kids” with the construal of temporary state, he emphasized his point by pointing to the imaginary individuals and their kids on the previous “family tree” model he constructed in some downward sequences (see Fig. 6.3b). In this example, his pointing to the kids may not only function to draw the listeners’ attention, but also to organize his own thoughts of how the generations flow.

(3) a. Matt: You made a resolution and you came on my show (..)

Ellen: I think [they’re still laughing at your] feet.

b. Leonardo: I have lots of brothers and sisters. And they have loads of kids. Their kids had kids, [and now their kids’ kids are having kids.]
Sometimes, the speakers profiled the setting of an ongoing event with deictic gestures. In (3c), Hilary Clinton, pointed at the backstage with her two separated index fingers to indicate to the audience the place where the event of her watching took place (see Fig. 6.3c). Furthermore, the deictic gestures can be used to clarify the information source which the speech did not provide, as in (3d), in which the girl, a member of the audience, explained why she had come to the show. In Fig. 6.3d, you can see that the girl sitting on the arm of the sofa pointed backwards at the screen with her right thumb when uttering “were saying like before” and held it through “you were single”. The boys said “they are single” in the video clips that played before the girl came to the stage. Therefore, the epistemic use of a progressive utterance was embodied in such a deictic gesture, citing the source of information.

(3) c. Ellen: You look rested and fantastic. I don’t know how that’s possible, but you do.

Hilary: You are so nice to say that. [I was watching backstage]and I saw you talking to Ashlyn and Brett.

d. Ellen: Why did you really come here?

Audience: Well, I was (..) Okay, well, I’m a massive fan, Legit, but you guys [were saying like before, you were single], and I’m single.
6.4.2 Progressive aspect and multimodality

Previous linguistic and psychological comprehension studies stressed that the English progressive aspect should facilitate the focus on the internal structures of an event and stimulate the mental simulation of the ongoing process of an event. Although the prototypical gesture accompanying the utterances with progressive aspect is the pragmatic gesture, the progressive aspect-marked utterances in the data set with “continuous ongoingness” are most frequently accompanied by iconic (concrete and metaphoric) gestures. This result implies that for the American speakers in this real face-to-face interaction context, the English progressive construction was not being used with salient mental imagery that is specific to it. This could be because conceptual processing is highly flexible (e.g., Binder 2016; Dove 2016; Zwaan 2016), and the activation of grounded features varies across contexts. Relevant salient mental imagery was not strongly stimulated enough in the progressive aspect situations to give rise to gestures or iconic gestures in the following contexts. First, when there were pictures or videos showing the past events in the immediate environment, the speaker tended to make no gesture or deictic gestures rather than iconic gestures, as in (4a) and (4b). As (4a) and Fig. 6.4a show, when the video on the screen is demonstrating the event that the little girl is acting out the animals, no manual gesture occurred. However, the eye gaze of the little girl, Brielle, her mother, and that of Ellen, converged on the video on the screen in front. In (4b), Ellen asked Charlie to explain his big kiss with Mechan at
the end of his concert which was displayed in the video. When he said Mechan was his best friend and tried to explain something else, Ellen interrupted him, as she thought none of the audience was listening to him, because they were looking at the picture on the screen. Ellen pointed backwards at the picture on the screen with her right thumb when uttering “They are looking”, as in Fig. 6.4b. Meanwhile, the audience also started laughing.

(4) a. Ellen: You played “Head’s UP” with the animal noises.
    Brielle: Yes.
    Ellen: What animals did you (..), oh, you are acting it out.

b. Charlie: Anyway, so I made this record (..) Machan was (..)
    Ellen: No one is listening to you. [They are looking] at that.

![Fig. 6.4a No gesture with “are acting it out”](image1)

![Fig. 6.4b Gesture with “are looking at”](image2)

On the other hand, it was only the progressive utterances denoting “continuous ongoingness” that were the most likely to be accompanied by (concrete and metaphoric) iconic gestures. This gestural result provides further evidence that taking the progressive aspectual view reflects the speaker and enables the listener to view the event as ongoing dynamically (Comrie 1976; Madden & Zwaan 2003). The ongoingness and dynamicity of events apparently affords strong activation of the mental imagery of them. Consequently, the perceptual and motor simulations that underlie the speakers’ thinking and speaking are stimulated, which results in the representational gestures (Hostetter 2014; Hostetter & Alibali 2008). Most iconic
gestures mimicked the action that the agent of an event executed or the movement that the speaker may have imagined in fictive events, as in (5a). When talking about the plastic dots around her mouth on a helmet that she wore during the film shooting, the speaker explained that the function of them was to capture her expressions. In fact, the dots as technical devices were attached to the helmet, recording her expression. However, the speaker alternately moved her two vertical hands with palms away from her body up and down twice while uttering “are capturing” and held it during “my expression” (see Fig. 6.5a). Thus, I can say her gestures reflected her mental simulation of the working mechanism of the plastic dots.

(5) a. Ellen: What do the dots do? What are they doing?

Nyong’o: [They are capturing my expression], so I had four cameras that were hooked to like a helmet on my head.

b. Ellen: But you are in crazy costumes and then behind the camera directing.

Eva: [laughter] I (...) we had a beach scene, so [I was running back and forth] in this outfit.

In addition, unlike the results of the previous studies on the effect of aspect on the gestural representations of motion events, the concrete iconic gestures with the progressive utterances were more complex than those with the non-progressive utterances, in that the former were more likely to encode the manner of the motion events, and the latter to encode the path of the motion events (Duncan 2002; Parrill et
al. 2013). Our study showed that the iconic (concrete or metaphorical) gestures co-occurring with the progressive aspect marked utterances of motion events primarily only represented the path of concrete or fictive motion events, as in (5b) and (5c). In the concrete motion event of running back and forth in (5b), the guest, Eva Longoria, alternately moved her left extended index from right to left several times to represent the path “back and forth” of her running (see Fig. 6.5b). In the metaphorical events like (5c), the hostess, Ellen, commented on what the guest, Krasinski, did in a film by saying “you are running towards danger”. Meanwhile, she made a gesture that was also iconic to the fictive path of the metaphorical motion of “running towards”. As Fig. 6.5c demonstrates, Ellen moved her right hand with lateral palm towards her body quickly from right to left to depict the path. The reasons why (American) English speakers are more likely to profile the path rather than the manner of motion events in natural conversations may be as follows. First, as English is a satellite-framed language (Talmy 2000), path is the focus of motion events. According to the phenomenon of “manner modulation” (McNeill 2005: 201), the speakers of satellite-framed languages often omit manner in gesture and profile the direction of motion by gesturing about its path. Co-speech gestures with progressive aspect-marked utterances highlight the path to encode the duration and ongoingness of a motion event. Second, there are many metaphorical events in natural conversations, the actions of which are less capable of being simulated directly (Bergen 2007, 2010). However, paths are salient for sequential mental scanning. Third, the acting out of a path demands less effort on the part of the hands than expression of the manner of the motion events. Thus, it fits more with the principle of the least effort in communication (Zipf 1949).

(5) c. Krasinski: I don’t know we would have the courage to do the same.

    Ellen: And you don't have to do that. [You are running towards danger] instead of away from it.

d. Bryan: So when we got to [the scene where we are shooting in the jail],

    we thought, this is going to be important to see a man completely, physically stripped of everything.
There were also a few utterances with progressive aspect accompanied by gestures which were iconic for entities involved in the events. When taking a close look at them, I found that most of them were iconic for the entities (concrete or metaphoric) of the setting in which the event occurred, as in (5d). As Fig. 6.5d shows, Bryan Cranston held his two hands with palms towards center in front to mold the scene or the jail, that is, the setting or location where the event of film shooting took place. The maintenance of the gesture may imply the long duration of the film shooting there. It implies that speakers “mentally simulated locations” and then profiled them in the gestures while producing progressive aspect-marked utterances. This complements Liu & Bergen’s (2016) results that language comprehenders “did indeed simulate locations” (p. 181) in listening to progressive aspect-marked utterances, and this is also the case in language production. However, entities were the least profiled by the iconic gestures accompanying the progressive aspect-marked utterances denoting “continuous ongoingness” in conversations. This may have been caused by the rich use of pronouns (you or I) as the subject of utterances and the priority of profiling the action or the path left by motion in the progressive aspect-marked speech context.

### 6.5 Conclusions

This study investigated the multimodal representations of polysemous progressive
construction utterances. The results showed that the gesture rate and the prototypical gesture did not significantly differ in the polysemous progressive utterances, but the rate and characteristics of (concrete and metaphorical) iconic gestures accompanying them showed differences. Utterances with the progressive aspect have greater potential to be multimodal, particularly when they denote “continuous ongoingness”. The prototypical gesture accompanying the utterances with progressive construction was pragmatic gesture, the second most frequent being deictic gesture, no matter what meaning it denoted. However, only when the progressive construction represented the “continuous ongoingness” of an event was the co-speech gesture more likely to be iconic, and predominantly iconic to the action or path of motion.

There are two clear implications of the findings. First, context plays the central role in determining the multimodality of the polysemous progressive construction in general. The context of talk show or conversation requires high interaction and involvement across the speaker, the addressees, and the immediate and potential audience. Those characterizations of context result in the prototypical pragmatic gestures accompanying the progressive utterances, regardless of the meanings they denoted. Reference to the speaker, the addressee, the audience, and the pictures or video clips being talked about, in the immediate environment via deictic gestures, are also a critical reflection of the high interaction and involvement in the talk show. Without doubt, this needs further experimental study to compare the gestures produced in monologues like retelling stories and in dialogues, like retelling stories, and in dialogues, like semi-spontaneous conversations.

Second, it is the meaning rather than the form of a grammatical structure that determines the variation of its co-speech gestures. As a polysemous grammatical construction, the progressive utterances denoting different meanings significantly influenced the iconic gesture representations. Only when the progressive constructions denoted “continuous ongoingness” were they more likely to be accompanied by iconic (either concrete or metaphorical) gestures. According to the GSA framework (Hostetter 2014; Hostetter & Alibali 2008, 2010), representational gestures derive from the strong simulation of imagistic-motor information of an event. Thus, the “continuous
ongoingness” denoted by the progressive construction doubtlessly facilitated the speakers’ mental simulation of the spatial or motor information of the events’ intermediate stage, the actions or the path of motion in particular. The other meanings denoted by the same grammatical structure may have made the internal stages of an event less accessible, or provided no spatial or motor information to be mentally simulated at all. Thus, the speakers’ gestural behaviors adapted to the meanings of a polysemous grammatical construction in an appropriate way. The strong sensitivity to grammatical meanings in speech and co-speech gestures supports the position not only that is grammar meaningful, but also that the meanings of a grammatical structure can relate to different kinds of cognitive processing by speakers, which can be seen in embodied simulations, in gesture.
Chapter 7. FL Learners’ Multimodal Representation of the English Progressive Construction in Event Construal

7.1 Introduction

According to the Linguistic Relativity Hypothesis, the language we speak shapes the way we perceive the world and construe the reality around us, so language and thought/worldview are closely intertwined (Whorf 1956). Consequently, it is widely acknowledged that language provides a window onto how people construe the world, and that learning a new language is acquiring a new thinking pattern. As thought or worldview were labels that he found too static, Slobin (1987, 1996) proposed the dynamic categories “thinking and speaking” in place of the static entities “language and thought”. He put forward the “thinking for speaking” hypothesis, which states that the obligatory grammatical options provided by a language affect the speaker’s thinking in the moment-by-moment processes of speaking. It was further extended as the “thinking for speaking and gesturing” hypothesis (Cienki & Müller 2008b; McNeill & Duncan 2000). In recent years, studies on language and thought in SLA have gained increasing popularity. This research has investigated the cognitive restructuring, and the conceptual representations in particular, in the bilingual or L2 learner’s mind6 (e.g., Athansopoulos et al. 2015; Athansopoulos & Albright 2016; Bylund & Jarvis 2011; Stam 2010, 2015). Researchers focused on the cross-linguistic differences in the

---

6In the SLA literature, the term second language (L2) learners was originally used to refer to those learning a second language other than their mother tongue in the context of the target language, such as Chinese speakers learning English in the U.S. Now, it has been extended to anyone who learns another language rather than their native language in any context. Foreign language (FL) learners refers to those who learn another language other than their mother tongue via classroom instruction in the foreign language context, such as Chinese speakers learning English in China.
endpoint salience as a function of the presence or absence of progressive aspect, and in packaging the manner and path information of motion events across languages. The results showed that the extent of changes in bilingual or L2 learners’ thinking for speaking patterns depended on their language proficiency in, and exposure to, the L2 language. Generally speaking, the more proficient the bilingual or L2 learners are, and the longer exposure they have to the L2, the more likely the L2 learners will adopt the L2 thinking for speaking patterns to construe events. Thus, contrary to Slobin’s proposal that language-specific thinking patterns developed in L1 acquisition are “exceptionally resistant to restructuring in adult second language acquisition” (Slobin 1996: 89), these studies have preliminarily revealed that L2 learners are able to restructure their conceptual representation. However, it is unclear whether and to what extent FL learners, who have fewer opportunities for L2 exposure, can also restructure their thinking for speaking patterns. In rethinking for speaking in L2 (Robinson & Ellis 2008), the cross-linguistic influence, the influence of a speaker’s knowledge of one language on his/her knowledge or use of another language (Jarvis & Pavlenko 2008; Kellerman & Sharwood-Smith 1986; Odlin 2005), may originate from the conceptual transfer, i.e., the transfer of mental concepts and patterns of conceptualization in one language to another (Jarvis 2007, 2011, 2016).

Conceptual representation is multimodal by nature, which is realized not only in verbal (e.g., speech) but also in nonverbal experiential elements (e.g., co-speech gestures) (Pavlenko 1999; Pulvermüller 2003; Langacker 2008a). Gestures are argued to act as windows into the speaker’s “thinking for speaking” in their L2 (Gullberg 2008a, 2008b; Kellerman & van Hoof 2003; Gullberg 2013; Stam 2015). Like language, gestures can be transferred in the process of L2 learning. For instance, gesture rates, iconic gesture rates in particular, can be transferred from a high frequency gesture language/culture to a low frequency gesture language/culture7 (Efron 1941; Pika et al.

---

7 The classification of high or low gesture frequency culture is relative. English belongs to the high-gesture languages, as opposed to Chinese as a low-gesture language; it is also a low-gesture language when compared with Italian, Spanish and French as high-gesture languages.
2006; Scheflen 1972; So 2010). Bilingual learners were found to make more gestures than monolingual speakers, particularly more iconic gestures than speakers from a low-gesture language. The present study pursues a multimodal approach to investigate the FL learners’ changes in thinking for not only speaking but also gesturing patterns in event construal.

As the previous studies focused on a completely different cross-linguistic phenomenon, it is unclear what the situation is like when there are cross-language similarities and differences in certain aspects, for instance, progressive aspect in English and Chinese. Although the progressive aspect markers in both English and Chinese have similar spatial source for their development, they have evolved into different grammatical structures (Ruan & Wang 2015: 75). As noted in Chapter 2, Both Chinese and English provide aspectual means to encode ongoingness of events, but Chinese progressive aspect differs from English progressive aspect in at least three respects. First, Chinese has two progressive aspect markers, zài and zhe. The former focuses on the progressivity of an event, whereas the latter concentrates on the durativity of an event. However, English only uses the verb suffix -ing to mark the event as ongoing and specific. Second, the conceptual meanings are different between them. Chinese progressive aspect marker zài only represents ongoingness and zhe encodes either ongoingness or temporary/resultative states. But English progressive aspect not only has such temporal-aspectual uses as the ongoingness, habituality, temporary state and repetition, but has also extended into the epistemic domain, including categories such as futurity and emotivity. Third, English progressive aspect integrates with tense, which is absent in Chinese. Such structural differences may have been caused by the characterized ways of thinking, that is, the particularity of spatiality in Chinese, and the predominant traits of temporality in English (Ruan & Wang 2015; Wang 2013). Given these cross-linguistic similarities and differences, this study intends to investigate whether and to what extent Mandarin speakers of L2 English change their thinking for speaking and gesturing patterns while taking the progressive aspect to construe events.
7.2 Background

7.2.1 Thinking for speaking and gesturing in SLA

With the increasing awareness of the multimodal nature of language learning and teaching, more and more studies have been done on whether L2 learners can acquire the thinking for speaking and gesturing patterns of the target language. These studies mainly focused on the motion event domain. Talmy (1991, 2000) categorized the languages that lexicalize path of motion in a satellite or verb particle, and that encode manner of motion in the main verb, as satellite-framed languages (such as English, Dutch), and those that encode paths in a main verb and manner in a subordinate verb or adverbial as verb-framed languages (such as Spanish, Japanese). Slobin (2004) added a certain category of equipollently framed languages (such as Chinese), which package manner and path of motion in equipollent elements, that is, “both manner and path of the motion receive equal weight in serial verbs” (Wu 2011: 419). Based on Talmy’s semantic framework, a lot of cross-linguistic research has shown that the typologically different patterns in motion conceptualization in speech are also mapped onto their gestural depiction of motion events. In satellite-framed languages like English, the co-speech gestures usually encode other components of the event, path in particular, or conflate manner and path to highlight manner (e.g., Brown 2015; Brown & Gullberg 2008; Choi & Lantolf 2008; Duncan 2002; Negueruela et al. 2004). In verb-framed languages such as Japanese and Spanish, the co-speech gestures tend to encode manner and path in two gestures, or add manner information which is lacking in speech (Brown 2015; Brown & Gullberg 2008). The research results in equipollently-framed Chinese are rather inconsistent. Some found that the co-speech gestures frequently represent path but seldom manner or manner and path (Brown & Chen 2013; Chui 2011, 2012), whereas others found that manner of motion was frequently encoded in both speech and co-speech gestures (Duncan 2006). In fact, the gestural encoding of motion events is also affected by other grammatical categories, such as grammatical aspect. The co-speech gestures accompanying progressive aspect-marked utterances tend to more
frequently conflate manner and path, whereas those with perfective aspect-marked utterances are more likely to only express path, in both English and Chinese (Duncan 2002; Parrill et al. 2013).

Correspondingly, previous studies in SLA mainly focused on what changes have taken place in L2 learners’ thinking for speaking and gesturing patterns about the path or manner of events while learning typologically different languages (e.g., Brown 2015; Brown & Gullberg 2008, 2013; Kellerman & van Hoof 2003; Levis 2012; Negueruela et al. 2004; Stam 2008, 2010, 2015). Three main findings could be concluded from such speech and co-speech gesture studies. First, the L2 (e.g. English) speakers still maintained their L1 (e.g. Spanish) thinking patterns for gesturing while talking about path in L2 (e.g., Kellerman & van Hoof 2003; Negueruela et al. 2004). Second, the L2 speakers adopted L2 thinking for speaking and gesturing patterns for path or manner in their L2, particularly when their L2 proficiency increased enough (e.g., Lewis 2012; Stam 2015). Third, the L2 speakers developed mixed thinking for gesturing patterns for path or manner, resulting from the bidirectional interaction between L1 and L2 (e.g., Brown 2015; Brown & Gullberg 2008, 2013; Stam 2008, 2010). Generally speaking, it is easier for L2 speakers to adapt their thinking for speaking patterns to those of L1 speakers, whereas the gestural patterns, in particular for manner, are relatively more resistant to change.

7.2.2 The influence of grammatical aspect on event construal in L2

Cross-linguistic studies found that grammatical aspect has a great influence on motion event cognition. Speakers of languages without marked grammatical aspect (e.g., German, Swedish, Afrikaans) prefer to take a holistic view, mentioning and paying attention to endpoints when describing motion events. However, speakers of aspect languages (e.g., English, Chinese, Russian) tend to express the ongoingness and direct attention to the intermediate part of motion events (Athanasopoulos & Bylund 2013; Bylund et al. 2013; Flecken 2011a, 2011b; Flecken et al. 2015; Slobin 1987; von Stutterheim & Nüse 2003; von Stutterheim et al. 2012). Therefore, the way that a
language is configured in terms of grammatical aspect equips its native speakers with certain preferred patterns of event construal. However, there is still debate about whether such differences in aspect languages and non-aspect languages influence the bilinguals’ or L2 learners’ event conceptualization.

Bylund & Jarvis (2011) found that there was L2 influence (Swedish, non-aspect language) on L1 (Spanish, aspect language) conceptualization of motion events. They asked L1 Spanish-L2 Swedish bilinguals to provide oral descriptions in Spanish of the video clips involving motion events with different degrees of endpoint orientation. A grammaticality judgment test concerning aspectual contrasts was also carried out. The results showed that the bilinguals mentioned the endpoints of motion events more frequently than the monolingual Spanish speakers did, and that there were negative correlations between the bilinguals’ discrimination of aspectual errors in the grammaticality judgment test and the possibility of encoding the endpoints in motion events. The study reveals that these bilinguals tend to adapt their thinking for speaking about motion events from L1 to L2.

Flecken (2011b) found that there were mixed effects of Dutch-German on event construal, leading to a bilingual-specific event conceptualization pattern. She employed a verbal description task and non-verbal data via eye-tracking to study the effects of grammatical aspect on motion event cognition in early bilingual speakers of Dutch (an aspect language, according to her criteria) and German (a non-aspect language). The results revealed that the early bilinguals not only frequently applied the progressive aspect, but they also tended to combine the progressive aspect with endpoints, which is not prototypical for native Dutch. The eye-tracking data showed that like the monolingual Dutch speakers, the bilingual Dutch speakers allocated more attention to the ongoing action rather than to the agents performing the action, which positively correlates with the high frequency of use of the progressive aspect.

Hilberink-Schulpen, Nederstigt & Starren (2014) investigated the grammatical preferences in aspect marking in L1s and L2s which differed in the availability of progressive form to encode ongoing events. They asked native Dutch, English and German speakers, on the one hand, and L1 Dutch L2 English and L2 German speakers,
on the other hand, to rate the likelihood of using a particular utterance (utterances in the
simple form, progressive form, simple form+object, or progressive form+object) on a
7-point Likert scale. The results showed that the grammatical repertoires of languages
(having a grammatical marker for the progressive or not) influenced their speakers’
event perception. Whereas the native Dutch speakers had no overall preference for any
kind of sentence, the native English speakers preferred to use progressive utterances to
describe the non-locomotion events, regardless of the presence or absence of objects in
view. However, the native German speakers preferred a simple form with an object
utterance. Moreover, both the L1 Dutch L2 English speakers and the L1 Dutch L2
German speakers demonstrated similar patterns as the respective target native speakers
did in event descriptions. Dutch learners of English preferred a progressive sentence,
just as the native speaker of English did, while Dutch learners of German preferred a
simple sentence, like the native speaker of German did, to describe motion events with
a clearly visible object, but preferred none of the sentences to describe those without a
visible object. Thus, the bilingual learners were more likely to adopt the L2 thinking
for speaking patterns.

In contrast, von Stutterheim (2003) found that the patterns of event
conceptualization developed through L1 were not easily susceptible to change. She
investigated how the English and German natives who were also advanced L2 learners
of German or English construe motion events with clear, reachable endpoints or without
any endpoints. The participants orally described the events immediately when they
recognized what was happening in the film clip in both the L1 and the L2 respectively.
The results showed that the English speakers preferred not to mention endpoints in
either English or German, whereas the German speakers tended to refer to endpoints in
both German and English. Thus, the L2 learners maintained their patterns of event
construal in L1, rather than changing to those of the target languages.

The cross-linguistic influences of grammatical aspect on L2 learners’ event
construal documented by the above studies reflect that there was reverse transfer i.e.,
the effects of the L2 on the L1 (Bylund & Jarvis 2011; Hilberink-Schulpen et al. 2014),
forward transfer, i.e., the effects of L1 on the L2 (von Stutterheim 2003), and mixed
effects of both L1 and L2 (Flecken 2011b).

7.2.3 The functions of gesture in grammar teaching and learning in SLA

There has been a serious written language bias in SLA, overlooking the multimodality and embodied nature of human communication (Block 2014). Fortunately, many studies have shown that gestures could improve language comprehension (Sueyosmohi & Hardison 2005), enhance L2 vocabulary memorization (Kelly et al. 2009; Tellier, 2008), organize spatial thoughts (Kita 2000), and reveal L2 learners’ thinking pattern (e.g., Stam 2010, 2015). Whether it is also of great help to teach and learn grammatical categories awaits further and systematic research. As grammatical concepts are rather abstract, the speakers and learners cannot directly perceive and experience them in daily life. Thus, it also poses great difficulties for teachers to teach and learners to learn. In recent years, there has been more and more research into teaching a variety of grammatical categories in a multimodal way, such as locative prepositions (Nakatsukasa 2016), tense and aspect (Kimura & Kazik 2017; Matsumoto & Dobs 2016). These studies showed that the teacher’s gestures could facilitate students’ learning of grammar, enhance their interaction in the classroom, and help learners organize their thinking to finish linguistic encoding online in the activity of speaking. In addition, the students appropriate the teachers’ gestures to externalize their developing understanding of a certain grammatical concept.

For example, Kimura & Kazik (2017) analyzed the microgenetic development of an English as second language learner’s understanding of the progressive aspect. They found that the teacher made the cyclic gesture (rotating both hands in an alternating manner) as an “in-progress catchment” to demonstrate the characteristics of English progressive aspect (p.18). Their Arabic subject — Mada’s — microgenetic development of the progressive aspect via gesture showed that she integrated her initial incomplete conceptual understanding and the gestural information provided by the teacher, making a gesture by rotating both hands outward alternately while moving them to the left and back to the right. Her gesture merged the “in progress” of progressive aspect with the
lateral representation of temporal duration. Matsumoto & Dobs (2016) carried out qualitative analyses to investigate the functions of gesture in teaching and learning tense and aspect in grammar classes of English as a second language. The teachers repeatedly and consistently employed abstract deictic gestures by pointing backward or left for the past tense, forward or right for the future tense, and downward in front of body for the present tense. The metaphoric gestures (for time as space) were used to teach and learn grammatical aspect in class. While explaining one of the characteristics of present progressive — extended period, the teacher used a metaphoric container gesture, i.e., holding both hands in parallel apart in front of their body at chest height with palms facing each other. Moreover, the metaphoric circular-motion gestures (i.e. repeatedly moving the hand(s) in a circular manner) were used to teach the habitual aspect. Students also effectively employed such gestures to demonstrate their understanding of the English tense and aspect system and interact with the teachers in classroom. The above two case studies reported two different instructional uses of gesture for the English progressive aspect in the SLA classroom: rotating (both) hands alternately (i.e., the cyclic gesture) and holding hands in parallel apart. What they have in common is that gesture has been shown to be one of the critical pedagogical means to teach and learn such abstract temporal concepts in the second language classroom. However, it is unclear whether L2 learners represent their conceptual understanding of English progressive aspect via the cyclic gesture, the gesture of holding hands in parallel apart, or something else, in natural conversational contexts.

Based on the findings reviewed in this section, I investigated what FL learners actually speak and gesture in semi-spontaneous conversations when using progressive aspect to construe events. The speech and gesture produced by the native Chinese speakers and native English speakers in this respect were used as baselines to see whether and to what extent FL learners change their thinking for speaking and gesturing, as evidenced by their verbal and co-verbal behavior. The present study addressed two primary questions.

1) Given the similarities and differences between the ways that Chinese and English speakers adopt to construe events with progressive aspect, what are the
potential problems for FL learners in using the English progressive aspect?

Specifically: (a) What formal errors do FL learners make in using the progressive aspect? (b) In light of the multi-functions of the progressive aspect in the target language, English, can the FL learners expand the semantic scope of the English progressive?

2) In what ways are FL learners’ gestural patterns with progressive utterances similar or different from those of L1 English speakers or L1 Chinese speakers?

Specifically: (a) Do FL learners gesture more frequently than monolingual speakers, the low-gesture L1 Chinese speakers in particular? (b) Are the ways in which FL learners use iconic gestures for reference similar to, or different from, the patterns used by L1 English speakers?

7.3 Methods

7.3.1 Participants

There were three groups of participants in this study: FL learners, native Chinese speakers, and native English speakers. The FL learners were 20 L1 Chinese L2 English speakers. They all had experience of attending the national speech or interpretation contests in China and won the first or second class prize. Their self-rated reading and writing proficiency was more than 8 on average, but the average spoken English was about 7 in a 10-point scale in the Language Experience and Proficiency Questionnaire (LEAP-Q) (Marian et al. 2007). It revealed that they are advanced FL learners in terms of reading and writing, but high-intermediate FL learners in terms of speaking. According to the results of a handedness questionnaire, all of the participants are right-handed people. As the present study focused on the progressive aspect, three learners who made no use of progressive aspect from three different pairs were excluded. Thus, data produced by 17 participants were analyzed (mean age: 22.65, SD = 2.344; age ranges 19-25 years; 11 females and 6 males). They have the experience of learning English on average for 14 years. Seven of them had been abroad as a traveler or
exchange student from 1 week to 1 year.

There were 17 native Chinese speakers in this study (average age: 47.41, SD = 10.666; age ranges 29-68 years; 5 females, 12 males), as well as seventeen native English speakers (mean age: 43.18, SD = 12.386; age ranges 24-69; 8 females, 9 males). All of them are celebrities from the entertainment, sports, economic or political fields in China, in the U.S. or internationally.

7.3.2 Data collection

Data of the FL learners were collected using the research design from Becker et al. (2011). The semi-spontaneous data were elicited via a semi-structured, semi-spontaneous format in which participants talked about personal experiences with their friends. This is in part to overcome the limitation of previous work investigating the role of grammatical aspect in event conceptualization which used video clips or written narratives as prompts, resulting in a bias towards physical motion events (Duncan 2002; Parrill et al. 2013). The conversation prompts used in the present study are appropriate to obtain real usage data from the speaker’s own experience without a bias towards any event type. The data consist of three topics of telling about: 1) your favorite place in the world (city, region, or place you have visited), as a warm-up question; 2) a difficult situation you have experienced, to elicit talk on events of a longer duration; and 3) an unusual situation you have witnessed, to focus more on quick or sudden events. A list of relatively everyday situations was available to choose from under items 2) and 3). The details about the conversation prompts can be found in Appendix 1. Each pair of participants as friends or classmates was asked to interview each other about the three topics. They could choose the same item or different item under 2) and 3) to tell the stories they had experienced in detail rather than briefly comment on them. They were told that the aim of the study was to investigate their communicative competence and I mentioned nothing about gestures in the experimental instructions. Before the recording, the participants were asked to read and sign the consent forms (see Appendix 2) if they agreed to take part in the experiment. Each pair of participants sat on chairs without
arms next to each other at about 45 degrees in a classroom or in an office. The video
camera was positioned in front of them, with a wide-angle view from the torso up. After
the recording, they were asked to sign another consent form, where they could choose
how their data would be used, and possibly published, for research purposes, and to fill
in the LEAP-Q and a handedness questionnaire, adapted from the Edinburgh
handedness inventory (see Appendix 3).

Ten pairs of participants were recorded. Each pair of participants discussed the
three topics for as long as they liked. Their conversation ranges from 17 to 32 minutes
with an average of 22.94 minutes, and the whole FL corpus consists of approximately
3.8 hours. 136 progressive utterances in total were produced for further analysis in the
present study.

The data of the native Chinese and English speakers were randomly sampled from
the popular TV talk shows — “Yang Lan One on One” in China, and The Ellen
DeGeneres Show in the U.S. They could be used as the baseline data for the FL learners
for the following reasons. First, they were also semi-structured spontaneous interactions.
Second, the interviews also centered on the difficulties or unusual/unforgettable events
the interviewees were involved in or witnessed in their personal life. As the hostesses
have a special role in the interviews, their data were not taken into consideration in both
Chinese and English. Consequently, 17 native Chinese speakers produced 146
progressive utterances in 13 videos, each of which lasted for approximately 20 minutes,
so approximately 4.3 hours in total. 17 native English speakers produced 96 progressive
utterances in 18 videos, each of which lasted for approximately 10 minutes, so
approximately 3 hours in total.

7.3.3 Data treatment

Each pair of FL participants was asked to help transcribe their dialogue without any
revision of errors. I checked all of the data for the use of progressive utterances. Then
all of the data in this study were annotated in ELAN and coded, as described below.
7.3.3.1 Speech coding

All of the narratives from the three groups were divided into clauses. Following Berman & Slobin (1994), a clause refers to any unit involving a finite or nonfinite verb that encodes a single activity, event, or state. Those with aspectual and modal verbs counted with main verbs as one clause, for instance, clauses with begin (to sleep), go (shopping), and try (to surprise her). The self-referential and paranarrative clauses, such as, I think, you know, were not counted as a clause. However, in this study, only the clauses with progressive aspect, in the form of be V-ing or present participle in English or zài V or V zhe in Chinese, were counted in data analyses.

For the FL learners’ data, I sampled out all the clauses with the progressive aspect or potential progressive use. Then grammatical errors related to progressive constructions were coded. They were categorized into the following error types, as illustrated in the examples numbered below: common tense error (1a), absence of copula verb in a main clause (1b), no aspectual inflection (1c), combination of tense and aspect (TA) errors (1d), and incorrect use of the progressive construction (i.e., using the progressive construction to construe a completed event) in (1e). However, when there were such self-corrections as in (1f), it was considered grammatical-error free. Two coders judged all of the sampled progressive sentences, and the agreement between them was 98% (N = 136).

(1) a. She’s standing in the line, and people are pushing her. [Participant 11, for past events]
   b. Many sad stories taking place in dormitory on the campus. [Participant 18]
   c. They were listen to me. [Participant 12]
   d. I was so frustrated, because everyone in the dorm is sleep. [Participant 1]
   e. the phone is dropping into your bag or something. [Participant 12, a completed event]
   f. We’re driving... we were driving ... around the city. [Participant 1]
The meaning expressed by each progressive clause was coded. While the English progressive utterances can encode ongoingness, temporary state, habituality, repetition, futurity, and emotivity, the Chinese progressive utterances represent either continuous ongoingness or temporary state (See section 2.1.2 in Chapter 2 for details). Thus, it is of great significance to see whether and to what extent the FL learners expanded their semantic system of the progressive aspect in English. When the progressive construction depicts that an event is in progress, it means “continuous ongoingness” as in example (2a) in English, and (2b) in Chinese. When the progressive construction is used with stative verbs or adjectives, it is construed as “temporary state” like in example (2c) in English, and (2d) in Chinese. When the same action repeatedly occurs at a regular interval, the English progressive construction is read as “habituality” as in example (2e), whereas when the same action is continually repeated in a series, it is construed as “repetition” as in example (2f). In addition, the English progressive construction is often used to depict the event happening in the future as in example (2g), and to express the speaker’s emotivity as in example (2h).

(2) a. Everybody was pushing up against me.
   b. tā zài pá.
       he PROG crawl
       “He was crawling.”
   c. So Chewbacca was sitting there.
   d. tā zuò zhe.
       he sit DUR
       “He was sitting.”
   e. Terry Norris, the former welterweight champion, has been training me.
   f. They are knocking at the door.
   g. So I suppose you are returning to your hometown after your graduation.
   h. You must be kidding.
Two coders independently coded the meaning of each progressive utterance according to the contexts. The agreement for the Chinese, FL learners, and English progressive meanings was 98% (N = 136), 93% (N = 146) and 95% (N = 96), respectively. The cases of disagreement were further discussed with an expert, then the agreed upon interpretation was used as the final one.

7.3.3.2 Gesture coding

Only gestures whose gesture stroke was overlapping with the verb marked by the progressive aspect were considered throughout this study. We first decided whether the progressive utterances were accompanied by gesture or not. The agreement between two coders was 100% for all data corpora. When the gesture bore direct resemblance to the semantic content of the speaker’s speech, the gesture was categorized as iconic (concrete and metaphoric); otherwise it was coded as non-iconic (McNeill 1992, 2005; Kendon 2004). Iconic gestures are important in this study, as they can visually depict what elements of events are profiled in the speakers’ mind. We can see whether the progressive aspect in Chinese, English and L1 Chinese L2 English have similar or different influences on their speakers’ thinking for speaking. The iconic gestures were further coded according to how they represented the referents (Müller 1998, 2014, see section 2.3.1 in Chapter 2 about the modes of representations for detail.). The iconic gestures can represent the entities, actions, process, path or combination of them of an event (see section 2.3.2 in Chapter 2 about this distinction for detail). When the gesture traced the outline or shape, molded the size of an entity, or represented something directly, it was coded as iconic to entity. When the gesture acted out the real or imagined action of concrete events, it was coded as iconic to action. When the gesture acted out the duration of an abstract event, it was coded as iconic to process. When the gesture traced the path of a motion event, it was coded as iconic to path. When a gesture simultaneously represented the elements of an event such as entity and action, it was coded as “combination”. The agreement between two coders for gesture types of the three groups was 100%, while that for the elements the iconic gestures refer to of the
native Chinese speakers, FL learners, and native English speakers was 97% (N = 52), 94% (N = 64) and 93% (N = 22) respectively.

7.4 Results

The results are presented in two sections. First, the form features of the progressive construction employed by the FL learners are reported and the relative meanings of progressive construction across languages are compared. As not all FL learners made grammatical errors, only a subset was included in the analysis. Second, the gestural behaviors of FL learners, native Chinese speakers and native English speakers are compared. I used nonparametric tests, Kruskal-Wallis for preliminary multiple group analyses and Mann-Whitney for further between-group analyses. In all statistical analyses, the alpha level is set at 0.05; a p value below that is significant.

7.4.1 Speech results

7.4.1.1 Form features of the progressive construction employed by the FL learners

Of the 146 progressive constructions produced by the FL learners, 134 involved be V-ing (91.78%) and 12 involved V-ing (8.22%) to construe events. Our analysis of the errors was taken into account. The grammatical errors made by the FL learners were almost a third of the total use of progressive construction in the present study (47 out of 146). As shown in Fig. 7.1, you can see that 57.45% of the grammatical errors lay in the wrong use of tense, 29.79% resulted from the absence of copula in the main clauses, and that there were few errors in non-inflection of aspect, combination of TA error, and incorrect use of progressive aspect. The results demonstrated that there was apparent and serious L1 negative transfer to the FL learning system in speech. The L1, Chinese, as a tenseless language, is characterized by no use of a copula in progressive constructions and non-inflection of the main verb in a clause. In contrast, the L2, English, as a tense language, conflates tense and aspect. This can be taken as the reason why progressive aspect is one of the most difficult grammatical categories for L1
Chinese L2 English learners to grasp. Thus, the conceptualization transfer easily occurs; that is, the L1 Chinese L2 English learners tend to transfer their L1 thinking for speaking patterns of the ongoing events while construing them in L2. The results also imply that the L2 learners have mastered the concept of the progressive construction, that is, the progressive construction is used to encode the ongoing rather than the completed events, which is similar in both the L1 and L2 systems. Thus, the positive concept transfer of progressive construction may have existed in the L2 learners’ interlanguage system.

Fig. 7.1 Grammatical errors of progressive construction in L2

7.4.1.2 Frequencies of meanings of progressive constructions across groups

No matter whether it was in the native Chinese, FL learners or native English corpus, there was no progressive construction that solely encoded “repetition”. This may have resulted from the serious constraints of the verb: only punctual verbs are compatible with the progressive construction to construe a series of repetitive events.

Consequently, Table 7.1 shows the frequencies of semantic distribution of the progressive utterances made by the native Chinese speakers, FL learners, and native English speakers. From this, you can see that “continuous ongoingness” as the prototypical meaning of the progressive construction was encoded the most in all of the three groups. FL learners used far more progressive constructions to construe ongoing events than did the native English speakers (61.64% vs. 43.75%), but less than the
native Chinese speakers did (61.64% vs. 72.06%). Furthermore, temporary state was the second most mapped meaning to the progressive construction across the three groups, and among the FL learners in particular. They even overextended the use of the progressive construction to construe completed events, such as *One is missing/The sun... the sun was already setting on the mountain*. In contrast to the native English speakers’ much wider use of the progressive construction to encode “habituality” and “emotivity”, the FL learners were more used to employing it to encode future events. Thus, generally speaking, while the L1 Chinese L2 English FL learners started to expand the use of the progressive construction, becoming more like native English speakers, they were still more likely to maintain the semantic pattern of the native Chinese in their interlanguage system.

**Table 7.1:** Frequencies of meaning distributions of progressive construction across groups

<table>
<thead>
<tr>
<th></th>
<th>Ongoingness</th>
<th>State</th>
<th>Habituality</th>
<th>Futurity</th>
<th>Emotivity</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Native Chinese</td>
<td>98</td>
<td>72.06</td>
<td>38</td>
<td>27.94</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FL Learners</td>
<td>90</td>
<td>61.64</td>
<td>44</td>
<td>30.14</td>
<td>1</td>
<td>0.68</td>
</tr>
<tr>
<td>Native English</td>
<td>42</td>
<td>43.75</td>
<td>18</td>
<td>18.75</td>
<td>16</td>
<td>16.67</td>
</tr>
</tbody>
</table>

The speech results suggest that the FL learners mostly keep their L1 thinking for speaking pattern in FL conversations. They negatively transferred the formal features of the progressive aspect from L1 to L2, and also maintained the prototypical use of progressive aspect with some preliminary expansions.
7.4.2 Gesture results

In the analyses of multimodal potentials of the progressive constructions in native Chinese, FL learners, and native English, I tested not only the gesture potentials in general but also in relation to the iconic gestures in particular.

7.4.2.1 Gesture potentials of the progressive constructions across groups

Fig. 7.2a shows the mean proportion of gestures accompanying the progressive constructions out of the total number of progressive constructions in speech. There was no significant difference showing the gestures accompanying the progressive constructions between the native Chinese, FL learners, and native English speakers ($\chi^2(2) = 2.116$, $p > 0.05$).

![Fig. 7.2a Mean proportion of gestures accompanying the progressive construction](image)

7.4.2.2 Iconic gesture potentials of the progressive construction across groups

Fig. 7.2b shows the mean proportion of iconic gestures out of the total number of gestures accompanying the progressive constructions. There was a significant difference between the groups in the iconic gesture potentials of the progressive constructions ($\chi^2(2) = 17.725$, $p < 0.05$). Specifically, the native Chinese group employed more iconic gestures accompanying the progressive constructions than both the native English group ($z = -4.009$, $p < 0.05$), and the FL learner group ($z = -2.556$, $p < 0.05$).
However, the FL learner group used significantly more iconic gestures accompanying the progressive constructions than the native English group ($z = -2.029$, $p < 0.05$).

![Fig. 7.2b Mean proportion of iconic gestures accompanying the progressive construction](image)

### 7.4.2.3 Distribution of the referents of iconic gestures accompanying the progressive constructions across groups

As the progressive constructions employed by the native English speakers were accompanied by too few iconic gestures, I could not carry out any statistical tests here. Thus, Table 7.2 only shows the distribution of the semantic meaning indicating which referent (i.e., protagonist/object, action or schematic process etc.) was represented in the gesture. From the table, you can see that both native Chinese and FL learners encoded action the most in gestures, whereas native English speakers represented process the most. The three groups did not differ in representing more than two referents simultaneously in one gesture (i.e., the combination category). The FL learners represented action less than the native Chinese did (31.25% vs. 51.92%), both of whom did it far more often than the native English speakers did (18.19%). As the action or process constitutes the nucleus of an event, if we add them up, we see that the iconic gestures of native Chinese speakers refer to the nucleus the most (73.07%), far more than those of the FL learners (54.68%) and the native English speakers (50.01%). The FL learners’ iconic gestures expressed the path of motion almost the same amount as the native English did (21.88% vs. 22.73%), which was much more than that of the
native Chinese (3.85%). The entity was less frequently represented by the FL learners’ gestures than by the gestures of the native English and the native Chinese speakers. Generally speaking, L2 learners neither completely maintained the native Chinese pattern nor followed the native English pattern. It seemed that their gestural semantic pattern was influenced both by the L1, Chinese, and by the L2, English.

Table 7.2: Frequencies of the referents of iconic gestures accompanying the progressive constructions across groups

<table>
<thead>
<tr>
<th></th>
<th>Native Chinese</th>
<th>FL learners</th>
<th>Native English</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
</tr>
<tr>
<td>Entity</td>
<td>5</td>
<td>9.62</td>
<td>5</td>
</tr>
<tr>
<td>Action</td>
<td>27</td>
<td>51.92</td>
<td>20</td>
</tr>
<tr>
<td>Process</td>
<td>11</td>
<td>21.15</td>
<td>15</td>
</tr>
<tr>
<td>Path</td>
<td>2</td>
<td>3.85</td>
<td>14</td>
</tr>
<tr>
<td>Combination</td>
<td>7</td>
<td>13.46</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>52</td>
<td>100</td>
<td>64</td>
</tr>
</tbody>
</table>

7.5 Discussion

I investigated the FL learners’ thinking for speaking and gesturing about the events framed by progressive aspects by comparing their progressive speech and the co-speech gestures with those of the native Chinese speakers and native English speakers. Our study reported two main findings. First, the FL learners maintained the native Chinese thinking for speaking pattern of employing the progressive aspect in form and meaning. Second, while the FL learners gestured as frequently as the native Chinese and native English, the number of iconic gestures accompanying the progressive utterances made by the FL learners was greater than the number made by the L1 English speakers but smaller than the number made by the L1 Chinese speakers. They displayed their own
gestural patterns of representing the referents in the progressive utterances. The results will be discussed below.

7.5.1 L1 transfer in L2 progressive aspect use

According to the grammatical error analyses of the FL learners’ use of progressive aspect construction, I found that in speech, the FL learners usually tended to use the present progressive to describe what was happening in the past, and ignored the addition of the copula to form the progressive. Furthermore, they overlooked the inflection of the verb to the -ing form to constitute the progressive aspect, and overgeneralized the use of progressive aspect to encode the resultant state of an event on rather few occasions. Such error patterns supposedly originated from the negative L1 transfer to L2 speech and the high cognitive demand of speaking in L2. Cross-language similarities are likely to be positively transferred to L2 learning, whereas cross-language differences tend to be negatively transferred to L2 use (MacWhinney 2005; Tolentino & Tokowicz 2014). Chinese is a tenseless language, which lacks syntactic realization of tense and lacks copula use in progressive constructions. Thus, simplification of tense is very popular in the tenseless FL learners’ language production, both in writing and speaking. In FL learners’ interlanguage, the absence of the obligatory copula in the progressive aspect construction in finite clauses has something to do with the learner’s native language. While the copula is obligatory to support the morphological realization of tense in English progressive constructions, there is no need for a copula in a progressive sentence in Chinese. As Chinese has syntactic realization of progressive aspect, like English, FL learners seldom made errors in aspectual inflection and use. However, why did FL learners focus on using the present tense to express past events? This may result from the native Chinese extended present time viewpoint of the flow of event time (Chen et al. 2013). According to Chen et al.’s study, speakers of Chinese adhered less closely to the implicit tri-phasic temporal framing in verbal descriptions of imminent, ongoing, and completed events compared with speakers of English. Furthermore, Chinese speakers produced a wider, extended time window than English
speakers did in non-linguistic tasks like marking the actual duration of imminent, ongoing and completed actions. The authors concluded that “the absence of tense in Chinese leads speakers to focus by default on temporal continuity as opposed to temporal segmentation” (Chen et al. 2013: 90). Consequently, Chinese speakers tend to adopt a temporal viewpoint of the extended present as default, so they consider an event as current related to the speech time, even though it may have happened or will happen (Alloway & Corley 2004; Chen et al. 2013). Such a linguistic difference (i.e., having tense or not) resulted in different thinking for speaking patterns in native speakers (Slobin 1996). The entrenched L1 thinking for speaking pattern is also apparently maintained even in highly proficient FL learners’ minds.

As the previous studies observed, the elements of grammar were significantly more accurate in written narrative tasks than in oral narrative tasks (e.g., Dykstra-Pruim 2003). By comparing the interaction of lexical aspect and simple past marking in spoken narratives and their corresponding written ones by L1 Chinese L2 English FL learners, Yang et al. (2012) found that the use of simple past tense was more likely to be correct in the written narratives than that in the spoken context. In addition, the L1 Chinese L2 English FL learners also provided less past tense marking in oral production than in written production for obligatory contexts (Yang & Lyster 2010). This is also reflected in the present study, though I did not compare the development of progressive aspect in oral and written language use. However, I asked the FL participants to transcribe their conversations. Interestingly, although the instruction emphasized that they were required to write down what they originally said without grammatical revision, the FL learners sometimes could not help correcting the wrong use of tense and added the aspect marker -ing. Such behaviors may imply that the FL learners are aware of the appropriate form of English progressive aspect, but speaking in an L2 involves intensive online processing, which is time limited and cognitively costly. Therefore, the fluency of speaking may be at the cost of the accuracy of grammatical elements like progressive aspect.

The FL learners’ patterns of encoding the meanings of the progressive construction were also more similar to those of the L1 Chinese speakers than to those of the L1
English speakers. This result can be interpreted in two ways. First, the prototypical meanings of both Chinese and English progressive aspect, that is, ongoingness and/or temporary state, have been entrenched and transferred to the FL learning of the English progressive aspect. We can note the preferred One-to-One Principle (Andersen 1984), which claims that FL learners generally tend to associate one meaning/function with one grammatical form, and they prefer to adopt one-to-one mapping as a learning strategy to eventually acquire complex form-meaning pairs. Therefore, the expansion from the prototypical meaning to the peripheral meanings of the English progressive aspect seems to be not easy even for the high-intermediate FL learners in this study. Second, the English textbooks, grammar books and classroom instructions all emphasize that the progressive aspect means that the event is ongoing at the moment or in a period of time, and that it can express future events, an emphasis which neglects the other meanings of the progressive construction. Consequently, the focus on particular meanings and scarce input of other meanings may have impeded the FL learners’ semantic expansions of the progressive construction.

7.5.2 The link between learners’ gestural usage and development of conceptual understanding

Gesture frequency is said to be transferred from the relatively high-gesture culture to the relatively low-gesture culture, such as from Spanish culture to (Canadian) English culture (Pika et al. 2006), and from American culture to Chinese culture (So 2010). However, the results of this study showed that the native English speakers did not gesture more frequently than the native Chinese speakers while using progressive aspect constructions in general. On the contrary, the native Chinese speakers produced iconic gestures three times as many as the native English speakers did. Even the FL learners employed iconic gestures nearly twice as often as the native English speakers. The results show that iconic gesture frequency was transferred from the so-called relatively low gesture culture, L1 Chinese, to the relatively high gesture culture, English L2. Such results that are inconsistent with previous research can be interpreted in two
ways. First, the classification of gesture frequency culture and the direction of gesture frequency transfer may be affected by the linguistic items under investigation. So’s (2010) study focused on bilinguals’ gestures of motion events in terms of lexicalization patterns. However, our study centered on FL learners’ gestures accompanying a grammatical construction, i.e., grammatical aspect. In addition, the previous studies counted the gesture rate in terms of gestures per clause or per word, whereas our study calculated only the gestures overlapping with the grammatical aspect-marked verbs per clause. Second, Chinese and English differ in the extent of the progressive aspect facilitating mental simulation of the event details. According to the GSA framework (Hostetter & Alibali 2008), representational gestures (i.e. iconic gestures in this study) derive from simulations of the action or spatial information in the mind. Therefore, while the Chinese progressive aspect facilitated the speakers’ access to the details of the event, the English progressive construction may not be available with salient mental imagery that is specific to it in the context of interviews.

The results also showed that there were bidirectional effects on the referents that the iconic gestures represent with FL learners: the iconicity to action from L1 was maintained, and iconicity to path from L2 was incorporated into the FL learners’ interlanguage gesture system. The Chinese progressive aspect led the native Chinese to pay more attention to the nucleus of events, i.e., the action itself. Such an entrenched thinking for speaking and gesturing pattern may be more easily transferred in FL learning. Native English speakers “have a strong preference to encode manner in speech by default” (Brown & Gullberg 2008: 245), considering manner information secondary to path information (Talmy 1991). Thus, their gestures tend to represent the path information. While speaking in an L2, FL learners in the present study also preferred to complement the manner information in speech with path information in gestures. They even extended the potential path information in a stative event in their gestures as in example (3b) and Fig. 7.3b, presented and explained below. The downward path was added to the gesture in the description of the stative sitting position.

The qualities of FL learners’ gestures can be interpreted as revealing their emerging thinking for speaking patterns, which are not detectable in fluent and correct
L2 speech, particularly in the following two respects. First, although the FL learners encoded temporary states with progressive aspect in speech, most of their gestures reflected that they conceptualized the state in the form of dynamic ongoingness, as in Fig. 7.3a and Fig. 7.3b.

(3) a. Because others were do uhmm [just putting some make-ups] [and wearing dresses]

b. and I saw my father [sitting on the er... couch]

In example (3a), the speaker (the girl in black in Fig. 7.3a recounted that she felt embarrassed when she took part in a party without any preparation, because others prepared well. She depicted the temporary states of the other participants with the progressive constructions in speech, as Fig. 7.3a shows; meanwhile, she alternately moved her vertical hands with palms facing herself towards and away from her body as if she was acting out the other participants’ schematic action of putting some cosmetics on their face. In example (3b), the speaker described that when she opened the door, she saw her father was sitting on the couch with her mother. In her speech, she employed the progressive construction to depict her father’s sitting position rather than the process of sitting down, but she moved her open hands with palm downwards from the level of her chest to the level of her thighs, as Fig. 7.3b shows. Such gestures accompanying temporary states did not occur in the native Chinese and native English data corpora. The gestures suggest that in the FL learners’ mind, they mentally
simulated the stative events as ongoing events. Therefore, the gestures rather than the speech alone betrayed that the FL learners may have not completely mastered the “temporary state” use of English progressive aspect in conceptualization.

Second, while encoding ongoing events with abstract verbs in speech, the co-speech gestures produced by the native Chinese were mainly “back and forth” gestures, away from and towards the body, but those by the native English were more likely to be “cyclic gestures”. The “cyclic gesture” (rotating hands alternately in a clockwise or counter-clockwise way) has been reported as accompanying the progressive aspect in natural conversation (Ladewig 2011, 2014; Harrison 2009) and in the L2 instructional context (Kimura & Kazik 2017). The “back and forth” gesture with the progressive aspect seems to be particular to native Chinese speakers. However, there was convergence between source and target language in the L2 gestures in the present study. The “cyclic gestures” and “back and forth” gestures occurred almost equally in the FL learners’ gestures that were iconic to schematic processes (i.e., 6 for the former, and 5 for the latter; the total was 11).

(4) a. [It totally get my brain’s working]  
   b. [as we talking]

![Fig. 7.4a. Gesture with “my brain’s working”](image) ![Fig. 7.4b. Gesture with “as we talking”](image)

In example (4a), the speakers were talking about insomnia at night. The girl in yellow at the right side in Fig. 7.4a explained that talking about some exciting things with her friends before going to bed made it difficult for her to fall asleep. While
uttering the sentence *It totally get my brain’s working*, her extended left index moved cyclically around the left side of her brain, as shown in Fig. 7.4a. Such a cyclic gesture acted out the invisible progressivity and duration of her brain working. However, in some cases, the FL learners maintained their L1 “back and forth” gestures to indicate the dynamicity and duration of ongoing events such as in example (4b), while uttering *as we talking*, the speaker alternately moved her open hands with palm towards center back and forth as in Fig. 7.4b. Thus, such bidirectional effects of source language and target language on the FL learners’ gestural patterns could be interpreted as illustrating that they are on the way to conceptual development of the English progressive aspect.

### 7.6 Conclusions

In summary, taking the multimodal approach to study the FL learners’ representation of progressive aspect in speech and gesture, I have shown that FL learners do follow the L1 Chinese thinking for speaking patterns of using the progressive aspect in form and meaning in their L2 English speech. The use of present progressive to construe what was happening, the absence of the copula necessary for progressive aspect in the main clause etc., were transferred from the tenseless L1 to the L2 in speech. The FL learners transferred their conceptualization of L1 progressive aspect encoding “ongoingness” and “temporary state”, which limited their semantic expansions to the more varied meanings of the English progressive construction.

I have provided a number of insights into the FL learners’ gestural patterns in taking an internal viewpoint of events with the use of the progressive construction, namely, they did not differ in gestural frequency from both the L1 Chinese and L1 English speakers, but the iconic gesture frequency was transferred from L1 to the FL learners’ interlanguage gesture system. The FL learners integrated the most entrenched L1 Chinese speakers’ use of iconic gestures for actions and the very salient L1 English use of iconic gestures for path accompanying the progressive utterances. Furthermore, the emergent dynamic gestures with the progressive aspect encoding a temporary state
revealed that the FL learners may not have understood the conceptual meaning of the progressive aspect in construing states.

The results have some significant implications for FL teaching and learning of grammatical aspect with Chinese learners of English. First, FL teachers or peers should improve the awareness of the appropriate English progressive aspect form in oral English training. Immediate corrective feedback should be provided to FL learners when they make grammatical mistakes on the progressive constructions, as it is important to promote the FL learners’ attentional focus on critical aspects of different grammatical features between L1 and L2 (Robinson 1995; Schmidt 1990). Second, more input of the peripheral meanings, in addition to the prototypical meanings, of the English progressive aspect should be incorporated into the teaching syllabus. According to the usage-based approach to SLA, FL learners extract and build the form-meaning pairings from the exemplars with saliency they encounter. Thus, for FL learners, enough original data is the best key to jump over the hurdle of semantic expansions of the English progressive construction. Third, gestures should be integrated into FL grammar teaching and learning. This is one of the promising ways to train and cultivate students’ thinking ability in FL, as the thinking-based FL education calls for (Wen 2013). Co-speech gestures make the grammatical meaning and the characteristics of the abstract grammatical categories visible. Iconic gestures, representing the action and/or path of motion in particular, could be conducive to helping FL learners understand the nature of progressive aspect, i.e., focusing on the dynamic and durative ongoingness of events. Furthermore, cyclic gestures with the progressive aspect construction can also be of great importance to help FL learners understand the schematic duration and ongoingness of the abstract events. What is more, focusing on the FL learners’ gestures can help the teachers know more about the development of their conceptual understanding. For example, using dynamic gestures with the progressive aspect encoding “temporary state” revealed that the L2 learners conceptualize the stative event as a dynamic ongoing one. Under such circumstances, the teacher had better illustrate the “temporary state” use of progressive aspect immediately. They can use the stative gestures like holding the hand as if molding the
resultant position or state that the actions caused to demonstrate the meaning of “temporary state”. In addition, it may be necessary to inform FL learners that they need not use as many iconic gestures accompanying progressive utterances in dialogue as native English speakers do. Doubtlessly, when, whether and how to integrate gestures into grammar teaching and learning is an issue waiting to be systematically explored in future research.
Chapter 8. General Discussion and Conclusions

This thesis has focused on Chinese and English speakers’ multimodal representation of the temporal contours of events (i.e., grammatical aspect) while thinking for speaking and gesturing. It has provided a comprehensive account of multimodal grammatical aspect-marking constructions in Chinese and English by employing qualitative and quantitative corpus-analytical methodologies. In this final chapter, the main findings of the dissertation are summarized and discussed, the contributions and limitations of the dissertation are addressed, and directions for future research on the multimodality of aspect are also proposed.

8.1 Summary of the main findings and general discussion

Before a series of empirical studies were carried out, the theoretical framework of the thesis was presented in Chapter 2. I reviewed the previous studies on grammatical aspect in English and Chinese from the perspectives of linguistics and psychology. The main linguistic realizations of grammatical aspect in English and Chinese were presented and the cognitive account of their meanings was provided as the main objectives for the empirical studies that followed. Then I reviewed the psychological studies on grammatical aspect in language comprehension and production. Based on them, a multimodal approach to grammatical aspect was proposed to accommodate the multimodal realities of human interaction.

This thesis consists of five empirical studies from Chapter 3 to Chapter 7, with the first (see Chapter 3) investigating and comparing the gestural forms and gestural functions of the Chinese grammatical aspects at the macro as well as micro levels. It was revealed that at the macro level (perfective vs. imperfective aspect), there were no significant differences in gesture handedness, but the gestures accompanying the imperfective aspect were much more likely to be repeated in form and had greater potential to be iconic in function than those with perfective utterances. At the micro-
level, there were no significant differences of gesture handedness either, but the gestures accompanying the progressive aspect showed a greater tendency to be repeated than those with the other aspect forms. The gestures accompanying the progressive aspect had the greatest potential to be iconic (metaphorically in particular), followed by the durative aspect (concretely iconic in particular). Gestures with the actual aspect and the experiential aspect were less likely to be iconic. However, the gestural behaviors accompanying actual aspect are much more similar to those with imperfective aspect rather the experiential aspect of the perfective aspect. In addition, the perfective aspect and the imperfective aspect were both used more for abstract events than for concrete events, except that the durative aspect was more frequently used for concrete events.

Chapter 4 furthered the study in Chapter 3 by investigating gestures accompanying grammatical aspect-marked utterances (imperfective vs. perfective aspect) in three different event types — concrete events, metaphoric events and inherently abstract events — in natural Mandarin Chinese conversations. The results showed that the gestures accompanying imperfective aspect speech were more prone to be iconic than those with perfective speech for concrete and inherently abstract events. The gestures were mainly iconic for the combination of semantic elements and for the schematic process in imperfective aspect speech contexts, whereas they were predominantly iconic for entities in perfective aspect speech contexts in the concrete and inherently abstract events. The gestures accompanying the imperfective and perfective aspect-marked utterances in the metaphoric events demonstrated the same patterns in both the frequencies to be (highly) iconic and the semantic elements (i.e., action) that they were iconic for. The findings extended the research on multimodal event construal, suggesting that the aspectual modulation of event conceptualization increases with concrete and highly abstract events, and diminishes for metaphoric events.

In Chapter 5, I used evidence from speakers’ gestures to augment semantic debates on the Chinese progressive aspect versus the durative aspect, the different meanings of the durative aspect in particular. I investigated the dynamic or static properties of the gestures accompanying utterances with the progressive zài and the durative zhe. I found that almost all of the gestures co-occurring with the zài-framed utterances were
dynamic, iconically encoding the action or process of the events, but that gestures accompanying the zhe-framed utterances were either static or dynamic, depending on the situation. The multimodal evidence suggests that dynamicity is the prototypical meaning of the progressive, and that the durative has the semantic potential for either dynamicity or stativity, depending on the linguistic contexts that instances are embedded in.

Chapter 6 extended the research on multimodality of Chinese imperfective constructions to the polysemous or multi-functional aspect constructions — the progressive (be V-ing) construction in English. It compared the potentiality to be multimodal and the gestural functions (i.e. types) for the different meanings of the progressive construction in the community as well as in an individual. In addition, it was made clear what elements of the events were referred to by iconic gestures. The results showed that the progressive construction as a whole was more likely to be multimodal, regardless of the level of the community or the individual. In addition, the prototypical function of gestures accompanying the progressive construction was pragmatic. Only when the progressive denotes “continuous ongoingness” were the gestures more likely to be iconic, representing the action of an event.

Some preliminary differences between co-speech gestures accompanying progressive utterances among the native Chinese speakers and the native English speakers were observed in Chapter 5 and Chapter 6. Based on them, Chapter 7 furthered the investigation to whether, and to what extent, FL learners (L1 Chinese L2 English) changed their thinking for speaking and gesturing patterns from L1 to L2. I examined FL learners’ verbal and gestural features when taking the progressive aspect to construe events, using those data from native Chinese and English speakers as baselines. The results showed that FL learners maintained their L1 thinking for speaking patterns by making the negative transfer of Chinese progressive aspect in form and following the semantic uses of the Chinese progressive aspect. In addition, FL learners transferred the iconic gesture frequency from the L1 Chinese thinking for gesturing pattern. However, they demonstrated a mixed gesturing pattern in terms of the referents that their gestures were iconic for.
The findings further attest the close interrelation between grammar, gesture and cognition. While conceptualizing events in different aspectual views, the speakers may have simulated the events in different ways, represented by the gestural indices. Some gesture forms can directly predict the aspectual differences (e.g., iteration and dynamicity of gestural movements), but others cannot (e.g., gesture handedness). Gestural repetition characterizes the progressive aspect in Chinese. The ongoingness of the progressive aspect-marked events is embodied in the dynamic gesture stroke, while the durativity of the durative aspect-marked events is encoded by the stative gesture stroke hold. The single hand dominates with gestures accompanying speech, regardless of the aspectual view the speaker takes. In addition, the gestural functions reflected that the gestures co-occurring with imperfective aspect utterances were more likely to iconically represent the elements of events, and the nucleus of events, i.e., action or process, in particular. Despite the similar function of the actual aspect and the experiential aspect (for expressing perfectivity), and the progressive aspect and the durative aspect (for expressing imperfectivity), their co-speech gestures revealed that it appears not to be the case that mental simulation and access to the details of events are in fact totally comparable. Thus, types of thinking-for-speaking apparently differ not only across the basic aspectual contrast but also within the same aspectual category.

In addition, the findings also highlight that grammatical aspect plays different roles in the expression of concrete and abstract events, so it suggests that speakers cognitively process concrete and abstract events in different manners. For concrete and metaphoric events, grammatical aspect does not correlate with what is gestured (i.e., gesture content is determined by the lexical aspect), but with how the speaker gestured (i.e., in the iterated and more complex gesture forms), while for inherently abstract events, the grammatical aspect used probably correlates with what to gesture in certain image schemas. English and Chinese prefer different image schemas to demonstrate the ongoingness and durativity of the progressive constructions. Native English speakers prefer the CYCLE schema, whereas the Chinese speakers tend to use the BALANCE schema.

Furthermore, the dissertation found that the degree of progressive aspect use (that
presumably facilitates access to mental imagery of events) is different across languages according to the different cross-linguistic gestural behaviors. The gestures accompanying Chinese progressive utterances were more likely to be iconic, representing multiple semantic elements of an event, and action or process in particular. However, the gestures accompanying English progressive utterances were predominantly pragmatic, which were closely related with the discourse management in the conversations. This pattern is apparently even transferred to the L1 Chinese L2 English FL learners’ interlanguage system. Therefore, Chinese progressive constructions are more likely to activate the native Chinese speakers’ mental simulation of the details of events than the English progressive construction does for the native English speakers.

8.2 Contributions and limitations

This dissertation has refreshed the study of grammatical aspect with a new multimodal perspective by investigating its living use in spoken language (i.e., spoken Chinese and English). Previous studies in this area have typically focused on the use of grammatical aspects in written language, and were often based on the researchers’ intuitions about usage. Such a multimodal approach to grammatical aspect has presented a more comprehensive picture of speakers’ online thinking for speaking and gesturing about events when using different grammatical aspects. The gestures accompanying grammatical aspect-marked utterances have provided new insight into the visual image of what event details are saliently activated while thinking for speaking. The multimodal evidence partly confirms the linguistic and psychological claims about grammatical aspects and also provides new perspectives on them. The more iterated, more complex (encoding more than one event element in a gesture) and/or more action or process focusing gestures accompanying the imperfective utterances (c.f., Chapter 3 and Chapter 4) confirmed the linguistic claim that the Chinese imperfective aspect takes an internal view on the events, focusing on the internal details of them. However, this
is only restrained to concrete and inherently abstract events. In metaphoric events, the perfective aspect-marked and the imperfective aspect-marked utterances have equally high potential to correlate with iconic gestures, metaphorically representing the action as the nucleus of the events (cf. Chapter 4). Such significant similarities between the two basic aspectual contrasts might have been blurred by linguistic evidence alone and not noticed in the psychological studies. To our knowledge, this is the first study that brought the importance of the nature of events to the investigation of the basic aspectual distinction. The multimodal approach to grammatical aspect doubtlessly deepened our understanding of the role of grammatical aspect in mental simulation of various events and complemented what cannot be found in psychological studies about abstract events.

In addition, through its converging evidence focusing on the study of grammatical aspect, this dissertation to some extent has also empirically verified the basic tenets of cognitive linguistics that grammar is inherently meaningful in everyday language and meaning as conceptualization is embodied in our experience in or interaction with the world. Cross-linguistic data (Chinese and English) and diversified kinds of subjects (native Chinese speakers, native English speakers, and FL learners) were taken into consideration. In terms of the polysemous grammatical aspect constructions, both the Chinese durative aspect construction and the English progressive aspect construction, coordination has been found of specific meanings with specific gestural behaviors (c.f. Chapter 5 and Chapter 6). For instance, when the Chinese durative aspect expresses the stative duration of an event, the co-speech gesture tends to be a static stroke hold, whereas when the same durative aspect expresses the dynamic ongoingness of an event, the co-speech gesture correspondingly exhibits a dynamic stroke (c.f. Chapter 5). Such gestural evidence is not only conducive to settling the semantic dispute on the durative aspect in Chinese, but also illustrates which meaning the durative aspect encodes in what specific contexts. The results can also be used to teach the imperfective aspect for L1 English L2 Chinese FL learners. The durative and ongoing nature of the imperfective (or English progressive) aspect is embodied by iterated gestures or the conduit metaphor gesture (holding hands with palm towards center apart in front) (c.f. Chapter 3 and Chapter 5), mostly by the back and forth gesture in native Chinese
speakers, and mostly by the cyclic gestures in native English speakers, with equal use of the last two in the FL learners (c.f. Chapter 3, 6 and 7). The abstract grammatical meanings have been visibly materialized in the speakers’ hands.

This dissertation has also contributed to our understanding of the relations between language and thought, or thinking and speaking, from the perspective of multimodal communication. It touched upon the thinking for speaking patterns within a single language (c.f. Chapter 3 and Chapter 4), across languages (c.f. Chapter 6 and Chapter 7), and from the L2 developmental view (c.f. Chapter 7). Within language, the fine-grained grammatical aspectual options do influence the native Chinese speakers’ speaking and gesturing about events. The multimodal evidence revealed that the internal or external view on the events between perfective and imperfective aspect is not a black or white question but a matter of degree in the process of speaking. In terms of access to the details of events while thinking for speaking and gesturing, the continuum appears to be as follows: (Totally internal view) the progressive aspect > the durative aspect> the actual aspect >the experiential aspect (Mostly external view). Across languages, though the progressive aspects in both Chinese and English were claimed to take an internal view on events, focusing on their intermediate part or phase, the co-speech gestural data suggest that native Chinese speakers and native English speakers think rather differently for speaking and gesturing. It appears that the progressive aspect (including the durative aspect) in Chinese motivated its native speakers to pay much more attention to the internal structures of events via rich iconic gestural representations. However, the progressive aspect in English did not show this tendency, and it appears to have motivated the native English speakers in my study to focus more on the discourse of the conversation itself via rich pragmatic gestures. From the developmental view of thinking for speaking and gesturing in FL learning, FL learners have been found to transfer their L1 thinking for speaking about events in the progressive aspect, and adopt mixed thinking for gesturing patterns from the most salient L1 and L2 gesture features while taking an internal viewpoints on events. Such interesting results provide a lot of new ideas about the influence of grammatical aspect on speakers’ thought or thinking about events. They encourage more studies on language and thought, or thinking for
speaking, in future multimodal research.

Beyond the contributions of the dissertation, some limitations should also be mentioned. First, the aspectual meaning was only based on the grammatical aspect constructions without considering the influence of the lexical verbs, arguments, temporal adverbials, and so on at the sentential level. Langacker (1987, 1991a, 2008a), Talmy (2000) and Croft (2012) valued the importance of the verbs, arguments etc. in determining aspectual meanings on the sentential level, taking a uni-dimensional approach to aspect study with a consideration of English grammatical aspect markers as a stativizer. Thus, we should express caution on some occasions when interpreting the results in the dissertation. Second, some of the data samples are a little small, holding the risk of some results having occurred by chance. The progressive aspects in both Chinese and English are not so frequently used, so the samples of them were relatively small, particularly when the use with iconic co-speech gestures was concerned. Consequently, some results relative to the iconic gestures were not able to be subjected to statistical analyses in some sections of the dissertation. Third, the proficiency of FL participants’ spoken English is not advanced enough to be comparable to that of near-native English speakers. Previous studies showed that the level of language proficiency and the amount of exposure to the target language influence the adaptation to L2 thinking for speaking and gesturing patterns (e.g., Athanasopolous et al. 2015; Athanasopoulos & Albright 2016; Brown 2015; Brown & Gullberg 2008, 2013; Choi & Lantolf 2008; Flecken et al. 2015; Stam 2008, 2010, 2015). Thus, highly advanced FL learners would be needed to investigate the change of thinking for speaking and gesturing without difficulties in L2 linguistic expression.

8.3 Future directions for further studies

Although research on aspect has a very long history dating back to Aristotle, it has become a flourishing field since the 1960s. While in earlier decades the focus of research was on the logical action type (i.e., lexical aspect), the focus has shifted to the
constraints on the integration of lexical aspects into the grammatical aspect. With increasing interest and developments in research on grammar and cognition, the role of grammatical aspect in event cognition and mental simulation has attracted attention from psycholinguists, cognitive linguists and gesture researchers. The present dissertation has attempted to add to their work with a systematic study on grammatical aspect from a multimodal perspective. It has provided a number of interesting findings and solved some critical problems that are impossible to deal with from the study of linguistic (verbal) data alone. Indeed, aspect is a fundamental field of research which still calls for much research in the future within the framework of multimodal communication.

First, the interaction effects of grammatical aspect and lexical aspect on gestures warrant further explorations. Lexical aspect (i.e., verbs or verb phrases) represents the inherent qualities of the event. When different lexical aspects co-occur with the same grammatical aspect (i.e., temporal perspective on the event), different aspectual meanings may arise, which we would also expect to be represented in co-speech gestures. In addition, most of the events in spoken English and Chinese are not marked by grammatical aspect, i.e., zero aspect (Xiao & McEnery 2004: 236). When speakers take such a “neutral viewpoint” on events (Smith 1997), it is of great significance to examine how they present such an in-between perfective and imperfective view in their gestures. It would also be interesting to compare the gestures accompanying the same lexical aspect category with the perfective aspect, imperfective aspect and zero aspect, in order to provide a more exact idea about how grammatical aspect relates to gesture use in event cognition.

Second, studies should investigate the thinking for speaking and gesturing differences between aspect languages and non-aspect languages. Although there have been more and more studies on the behavioral differences such as attention allocation to the different parts of motion events between aspect language speakers and non-aspect language speakers, no studies, to my knowledge, have touched upon the gestural differences between them. Grammatical aspects are viewpoints highlighted by how language is used to present events, but it may be universal for people to perceive how
events unfold with time in reality. The ongoingness or completion of events is not grammatically marked in non-aspect language, but it is possible for such information to be complementary in co-speech gestures. Thus, the nature of the relations between language and thought or thinking and speaking awaits to be comprehensively understood via more studies on the grammatical options from the viewpoint of multimodality.

Third, proficiency is considered as an important predictor of adaptation to L2 thinking for speaking and gesturing, so a study with different proficiency levels should be designed in the future. The high-intermediate L1 Chinese L2 English FL learners have still negatively transferred the L1 thinking for speaking patterns, and tended to adopt some L2 thinking for gesturing patterns when construing events in the progressive aspect in the present study. Thus, it is of great importance to see whether beginners completely follow the L1 thinking for speaking and gesturing about events, and whether highly advanced FL learners mainly adopt the L2 thinking for speaking and gesturing about events. It is worth noting that the conversation prompts used to elicit data can only lead to very limited use of the progressive aspect, even with three among twenty speakers making no progressive utterance at all. Therefore, researchers should carefully develop some other better or more useful methods to collect data. In addition, according to a dynamic usage-based perspective on language learning and use, L2 development is a dynamic process (e.g., Larsen-Freeman 2006; Verspoor et al. 2011). As Chan et al. (2015) found, even identical twins have different dynamic development patterns in speaking versus writing in terms of syntactic complexity. It is of great importance for us in the future to investigate the development patterns of syntactic complexity (like grammatical aspects) and gestural complexity based on longitudinal studies.

Finally, it is also promising to apply the results in this dissertation to FL or SL teaching about grammatical aspects. Grammatical aspect is one of the most difficult teaching and learning points in SLA. The integration of tense with aspect and the wide range of semantic functions of the English progressive aspect is a great challenge for L1 Chinese L2 English FL speakers to learn. The absence of a formally expressed
perfective aspect in the English grammatical aspect system and the finer-grained aspectual distinction within the imperfective aspect in Chinese impose great pressure on L1 English L2 Chinese FL learners. Our studies have shed light on the visible differences in the gestural representations of different grammatical aspects underlying the conceptualization of events. Grammar is potentially multimodal and embodied. However, the field of SLA has a lingual bias, overlooking the multimodality and embodied nature of human communication (Block 2014). Therefore, it is time for FL teaching and learning to take a multimodal turn to make grammar teaching and learning more effective and more enjoyable.
Bibliography


Andrén, M. (2014). Multimodal constructions in children: Is the headshake part of


the Gap (pp.189-204). Stanford: CSLI Publications.


Dahl, Ö. (2001). Languages without tense and aspect. In Ebert, K. H., & Zúñiga (Eds.),
Aktionsart and Aspectotemporality in Non-European Languages (pp. 159-173). Zürich: Semina für Allgemeine Sprachwissenschaft.


Jin, F. M. (1991). Zhùci “zhe”de jīběn yǔfā yìyi (The basic grammatical meaning of


Discourse Grammar perspective. *Functions of Language*, 23 (2), 149-178.


Perceptual and Socio-Cultural Factors (pp.133-147). Amsterdam: John Benjamins.


Cognitive Linguistics, 24 (1), 135-158.


Neuropsychologia, 47 (2), 388-396.


Yang, Y. F., Gong, W., & Liang, L. (2012). Qíngzhùáng tǐ yù yì bān guò qǔ shí xī dé de kǒu bǐ yú dì bǐ yán jiū (A study of lexical aspect and the acquisition of simple past based on spoken and written narratives). Xiàn dàì wàiyǔ (Modern Foreign Languages), 35 (2), 158-166.


Appendix 1: Conversation Prompts

I. Tell about your favorite place in the world (city, region, or place you have visited).

II. Pick one from either of these two categories:

1. Tell me about a time when you had difficulty doing one of these things:
   a. Dealing with bureaucracy
   b. You ordered dinner in a restaurant in a foreign language.
   c. You were trying to get to sleep.

2. Tell me about a time when you did something you were proud of:
   a. You took up a personal challenge
   b. You organized a surprise party for a beloved one
   c. A close friend told you how much you had helped him/her.

III. Pick one from this category

1. Tell me about a time when you witnessed one of these situations:
   a. An accident
   b. Winning a sports event
   c. A random act of kindness
   d. A bizarre event
   e. A confrontation
   f. Something surprising
Appendix 2: Consent Form

I freely give my consent to participate in an experiment under the direction of Wang Ronghua. I understand that for the purpose of this study I will be asked to discuss several topics with a fellow student and then to fill out some surveys. My participation will not take more than 30-40 minutes.

I also understand that our discussion will be video recorded. The recording will be used for research on communication. At the end I will have the opportunity to specify how the videorecorded material can be used (e.g., who will be allowed to see it). I am guaranteed that any personal information collected on the surveys will not be made public and that my name will not appear in any publications or reports that may be produced based on this study.

At the end of the experiment the person conducting it will explain to me the purposes of the research. I am aware that at any point I may end my participation in the study without any consequences. In that case, the video recording of me will be erased.

Signature: 

Place and date:

Name written clearly:
Participant #_____
Your agreement for use of the recording

With your agreement a video recording was just made of you for the purpose of research on communication. You have the right to view the recording if you wish.

Please indicate below how the recording may be used. The recording will only be used in the manner(s) you agree to. Please indicate all conditions below where we have your consent.

1. **Anonymized information may be used from the recording in research publications and presentations (e.g., transcripts without my name, an artist’s rendering of my gestures).**

                      ........................................

[please sign if you agree]

2. **My image and voice from the recording may be used in research publications and presentations (e.g., for academic journal articles, teaching purposes, talks at academic conferences).**

                      ........................................

[please sign if you agree]

3. **Portions of the video and audio recording with my image and voice may be posted on academic websites (e.g., as a demonstration of a research project, as a demonstration of an example in a research article).**

                      ........................................

[please sign if you agree]

I give my permission for the recording to be used as indicated above.

Date:

Name (written clearly):
Appendix 3: Handedness Questionnaire


**Instructions**

For each of the activities below, please indicate:

Which hand you prefer for that activity?  Do you ever use the other hand for the activity?

<table>
<thead>
<tr>
<th>Which hand do you prefer to use when:</th>
<th>No pref</th>
<th>Do you ever use the other hand?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drawing:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Throwing:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using Scissors:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using a Toothbrush:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using a Knife (without a fork):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using a Spoon:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using a broom (upper hand):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Striking a Match:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Opening a Box (holding the lid):</td>
<td></td>
<td></td>
</tr>
<tr>
<td>items below are not on the standard inventory:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holding a Computer Mouse:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Using a Key to Unlock a Door:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holding a Hammer:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holding a Brush or Comb:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Holding a Cup while Drinking</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
List of Publications

Articles included in this dissertation


**Ronghua, Wang.** Thinking, speaking and gesturing: Foreign language learners’ multimodal representation of the progressive aspect. *Submitted.* [Chapter 7].

**Ronghua, Wang.** Grammatical aspect and co-speech gesture in different types of events. *Manuscript in preparation.* [Chapter 4]


Other published articles


12 (3), 84-90.


Summary of the Thesis

As one of the most important linguistic devices to represent events in Chinese and English, grammatical aspect reflects and influences the speakers’ cognition of events. This thesis addresses the multimodality of grammatical aspect in event representation to investigate the interrelations between grammar, gesture and cognition. The main questions we explore here are 1) what differences there are, if any, in gestural expressions of the basic Chinese grammatical aspectual contrast (perfective vs. imperfective aspect); 2) how gestures correlate with different meanings of the imperfective aspect in Chinese and English respectively; and 3) whether there is evidence that 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. Adopting qualitative and quantitative corpus-analytical methodologies, we systematically examined native Chinese speakers’, native English speakers’ and highly-intermediate L1 Chinese L2 English FL learners’ multimodal representations of aspectual distinctions while thinking for speaking and gesturing about events.

Chapter 2 reviews previous studies on grammatical aspect from the perspectives of linguistics and psychology, and then proposes a multimodal approach to grammatical aspect to accommodate the realities of multimodal human interaction. The linguistic studies on grammatical aspect point out that the most acknowledged grammatical aspect markers in Chinese are zài indicating progressivity, zhe indicating durativity, le indicating actualization, and guo indicating experientiality. In contrast, the progressive aspect realized by be...ing is the only grammatical aspect marker beyond dispute in English. The meanings of the grammatical aspects were provided with a cognitive account, particularly those of the polysemous English progressive aspect. The psychological studies on language comprehension and production showed that it is psychologically plausible for the imperfective aspect (i.e., progressive aspect) to focus
more on the internal structures of events than the non-progressive aspect in English. Most of these studies only focused on the English progressive aspect encoding “continuous ongoingness” and in concrete motion events. We constructed a multimodal interaction model to reveal the relation between dynamic process of thinking, speaking and gesturing, and the relation between static products of thought, speech and gestures. Based on it, we proposed a multimodal approach to grammatical aspect in natural or semi-spontaneous conversations. By investigating the forms and functions of gestures accompanying the grammatical aspect-marked utterances, we carried out a series of empirical studies to explore the thinking for speaking and gesturing patterns in native Chinese speakers, native English speakers and FL leaners in the following chapters.

Using the multimodal data from one of the most popular Chinese talk shows, Chapters 3 and 4 present studies that investigate the formal and functional differences of gestures accompanying Chinese perfective aspect utterances (including the actual aspect .beh and the experiential aspect.guo). vs. imperfective aspect utterances (including the progressive aspect.zài and the durative aspect.zhe). The results in Chapter 3 showed that the gestures accompanying grammatical aspect at the macro- and micro- levels were not different in the handedness, but in iteration of movement and functions. At the macro-level, the gestures with imperfective aspect were more likely to be iterated and to be iconic in form than those with perfective aspect. At the micro-level, the gestures with progressive aspect.zài were more likely to be iterative than those with the actual aspect .le, the experiential aspect.guo and the durative aspect.zhe. In addition, the gestures co-occurring with progressive aspect.zài and durative aspect.zhe was more likely to be iconic, but the former tended to be more metaphoric iconic, and the latter were mainly concrete iconic. Those with experiential aspect.guo was the most likely to be pragmatic, and they also tended to be metaphoric iconic as with the actual aspect .le. We further found that both the imperfective aspect and the perfective aspect was more likely to occur in abstract event descriptions. However, specifically, most of the actual aspect, the experiential aspect and the progressive aspect were used with abstract events, but the durative aspect was employed more with concrete events. The results suggest that gestures with imperfective aspect have the potential to be more complex than those
with perfective aspect in Chinese, and that the gestural behaviors of the micro-grammatical aspect in Chinese have their own salient characteristics, and should therefore be treated separately to reveal finer-grained qualities of event cognition.

Based on the findings that the uses of grammatical aspects are influenced by event types, Chapter 4 investigated co-speech gestures with grammatical aspects in different types of events in natural Chinese conversations. Imperfective utterances were more likely to be accompanied by iconic gestures representing combinations of event elements than perfective utterances in concrete events. The gestures with both perfective and imperfective utterances were equally highly iconic for action in metaphoric events. The gestures with imperfective aspect were more likely to be iconic for schematic processes than were those with perfective aspect for inherently abstract events. The results suggest that imperfective aspect has the greater potential to facilitate the speakers to focus on the internal structures or details of concrete or inherently abstract events than perfective aspect does. The findings also support the view that metaphor is a matter of thought, so that the speakers tend to experience the metaphoric abstract events in terms of the concrete events, regardless of the temporal perspectives they take on them (i.e., grammatical aspect they use in speech).

We narrowed down our “lens” on the gestural representations of the meanings of imperfective aspect in Chinese and English (progressive aspect) in Chapters 5 and 6. In Chapter 5, by examining whether the gesture stroke accompanying the two Chinese imperfective aspect constructions (i.e., the progressive construction and durative construction), we found that almost all of the gestures co-occurring with progressive construction were dynamic, reenacting out the physical actions of concrete or metaphoric events, or the schematic process of abstract events. In addition, a little more than half of the gestures with durative construction were dynamic in the linguistic contexts such as $Vzhe\ NP$ (verbs with high dynamicity), $Vzhe\ Vzhe$ ($V$ is reduplicated), and $ná\ zhe\ V$ (hold/carry $zhe\ V$), and the other half were static, enacting out the maintenance of one portion of the actual actions or demonstrating the resultant states. The findings suggest that the prototypical meaning of the progressive aspect is dynamicity, and the durative aspect has the potential to encode dynamicity and stativity,
depending on the contexts.

By investigating the gestures accompanying the English progressive construction from one of the most popular American talk shows, we found that the “continuous ongoingness” and “futurity” interpretations of the English progressive aspect were found to have greater potentials to be multimodal. The prototypical gestures accompanying the English progressive constructions were predominantly pragmatic, regardless of the meaning it denotes. Furthermore, only when the progressive construction encoding “continuous ongoingness”, the co-speech gestures were more likely to be iconic, representing actions involved in the events. The findings in Chapter 6 suggest that the English progressive construction does not in and of itself facilitate speakers’ focus on the internal structures that much as a whole, but can do so only when it has the “continuous ongoingness” construal in the interview context.

The last empirical study examines the changes of L1 Chinese L2 English FL learners’ thinking for speaking and gesturing patterns when they construe events with progressive aspect. The results showed that FL learners negatively transferred their L1 thinking for speaking about events to the L2 construal of events with wrong use of tense, absence of copula etc. They positively transferred the frequency of iconic gestures from L1 Chinese to L2 English, but have developed a mixed thinking for gesturing patterns in their interlanguage. The findings revealed that it is not only possible for FL learners to restructure their thinking for speaking and gesturing patterns, but it is promising for them to develop their own patterns with integration of the most salient characteristics of L1 and L2.

All of the findings in the empirical studies support the view that grammatical aspects in Chinese and English are embodied in their speakers’ physical experiences, and that their meanings as conceptualization modulate the correlation of speech with their co-speech gestures. In the final chapter, the findings of the thesis are summarized and generally discussed; the contributions, limitations as well as future directions are presented.