5-13-2021

When Gestures Do or Do Not Follow Language-Specific Patterns of Motion Expression in Speech: Evidence from Chinese, English & Turkish

Irmak Su Tutuncu

When Gestures *Do* or *Do Not* Follow Language-Specific Patterns of Motion Expression in Speech: Evidence from Chinese, English & Turkish

by

Irmak Su Tütüncü

Under the Direction of Şeyda Özçalışkan, PhD

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2021

ABSTRACT

Speakers of different languages (e.g., English vs. Turkish) show a binary split in how they package and order components of a motion event in speech and co-speech gesture, but not in silent gesture (Özçalışkan et al., 2016b). In this study, we focused on Chinese that does not follow the binary split in its expression of motion in speech (Slobin, 2004), and asked whether adult Chinese speakers follow the language-specific speech patterns in co-speech but not silent gesture, thus showing a pattern akin to Turkish and English adult speakers in their description of animated motion events. Our results provided evidence for this pattern, with Chinese—as well as English and Turkish— speakers following language-specific patterns in speech and co-speech gesture, but not in silent gesture. Our results provide support for the “thinking-for-speaking hypothesis” (Slobin, 1996), namely that language influences thought only during online, but *not* offline, production of speech.

INDEX WORDS: Motion events, Co-speech gestures, Silent gestures, Speech patterns, Cross-cultural differences

When gestures *do* or *do not* follow language-specific patterns of motion expression in speech:

Evidence from Chinese, English & Turkish

by

Irmak Su Tütüncü

Committee Chair: Şeyda Özçalışkan

Committee: Michael James Beran

Ute Römer-Weyhofen

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

May 2021

ACKNOWLEDGEMENTS

I would like to take this opportunity to show my appreciation to several people who made this project possible and helped me along the way. First off, I want to express my gratitude to my advisor Şeyda Özçalışkan for her guidance in my early years of graduate school, providing me with both academic and personal support in adapting to a new country during the time of a pandemic. I felt her support in every aspect of my academic life within these two years both in person and in countless online meetings. I also thank my collaborators in this project: Jing Paul who helped me with working through the Chinese data as well as collecting it, Samantha Emerson who understood me fully as a graduate student in Cognitive Sciences, and Murat Şengül who collected the Turkish data. I want to give special thanks to Melanie Knezevic, who put a tremendous effort on data coding and for becoming a great friend along the journey. I also thank my committee members Michael James Beran and Ute Römer-Weyhofen for agreeing to be on my committee, and for their wonderful feedback on the thesis. Last but not least, I thank my parents, Özlem and Özkan Tütüncü, who provided me with the best care possible while I wrote drafts of my thesis, allowing its completion in a timely manner.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS .... **IV**

LIST OF TABLES .... **VII**

LIST OF FIGURES .... **VIII**

1 INTRODUCTION **1**

1.1 Packaging and Ordering of Semantic Elements in Speech **1**

1.2 Packaging and Ordering of Semantic Elements in Co-Speech Gesture **5**

1.3 Packaging and Ordering of Semantic Elements in Silent Gesture **8**

1.4 Current Study **10**

2 METHODS **13**

2.1 Participants **13**

2.2 Procedure for Data Collection **13**

2.3 Transcription, Coding and Reliability **15**

2.4 Scoring and Analysis **19**

3 RESULTS **21**

3.1 Packaging of Semantic Elements **21**

3.1.1 Speech ***21***

3.1.2 Co-Speech Gesture ***21***

3.1.3 Silent Gesture ***22***

3.2 Ordering of Semantic Elements **24**

3.2.1 Speech *24*

3.2.2 Co-Speech Gesture *24*

3.2.3 Silent Gesture *25*

4 DISCUSSION **27**

4.1 Packaging of Semantic Elements **27**

4.2 Ordering of Semantic Elements **29**

REFERENCES **34**

APPENDICES **41**

Appendix A **41**

Appendix B **42**

Appendix C **43**

LIST OF TABLES

Table 2.1 List of The Animated Motion Event Scenes 14

Table A.1 Mean distribution of motion elements by semantic packaging and semantic ordering for each scene in speech by language 41

Table B.1 Mean distribution of motion elements by semantic packaging and semantic ordering for each scene in co-speech gesture by language 42

Table C.1 Mean distribution of motion elements by semantic packaging and semantic ordering for each scene in silent gesture by language 43

LIST OF FIGURES

*Figure 2.1 Example stimulus scene of a figure climbing into a treehouse (top) and its description in co-speech gesture (A_1, B_1, C_1,) and silent gesture (A_2, B_2, C_2) by speakers of English (A pictures on left), Chinese (B pictures in the middle) and Turkish (C pictures on right). In **co-speech gesture**, native speakers of English and Chinese preferred to express manner (circling fists) and path (upward trajectory) simultaneously in a single gesture (A_1-B_1), and native speakers of Turkish preferred to express path (leftward trajectory) by itself, omitting manner (C_1). In **silent gesture**, native speakers of English, Chinese or Turkish preferred to express manner and path simultaneously within a single gesture by circling fists upward (A_2, B_2, C_2).* 17

*Figure 2.2 Example stimulus scene of a figure climbing into a treehouse (top) and its description in co-speech gesture (A_1, B_1, C_1,) and silent gesture (A_2, B_2, C_2) by speakers of English (A pictures on left), Chinese (B pictures in the middle) and Turkish (C pictures on right). In **co-speech gesture**, some English speakers preferred to express motion (climb.into) first, followed by ground (treehouse); Chinese and Turkish speakers preferred to express ground (treehouse) first, followed by motion (climb/move into). In **silent gesture**, English, Chinese and Turkish speakers preferred to express ground (treehouse) first, followed by motion (climb.into).* 19

Figure 3.1 Mean number of sentence units with separated (manner-only, path-only, manner-path) or conflated (manner + path) motion elements in speech (A), in gesture with speech (co-speech gesture, B) and in gesture without speech (silent gesture, C); Error bars represent the standard error; the max possible number of sentence units was 16 for silent gesture condition. 23

Figure 3.2 Mean number of sentence units that follow Figure-Ground-MOTION or Figure-MOTION-Ground orders in speech (A), in gesture with speech (co-speech gesture, B) and in gesture without speech (silent gesture, C). Error bars represent standard error; the max possible number of sentence units was 16 for silent gesture condition............... 26

1 INTRODUCTION

Languages differ widely in their expression of different experiential domains, but do these differences influence the way speakers think about such domains? The "thinking-for-speaking hypothesis”—originally proposed by Slobin (1996)—suggests that language might influence cognition during online production of speech, but this effect might not be evident when *not* speaking. In this study, we focused on motion events–an experiential domain that shows systematic variability in its expression across different languages of the world (Talmy, 1985, 2000). We examined three distinct types of languages (Chinese, English, Turkish), which differ in their expression of two key motion components, namely manner (e.g., run, crawl) and path (e.g., exit, ascend). We asked whether the patterns of cross-linguistic variability in speech become evident in gesture when the gestures were produced with speech (i.e., *co-speech gesture*) versus when gestures were produced without speech (i.e., *silent gesture*). If language has an effect on thought only during online production of speech, we would expect to observe similar patterns of cross-linguistic variability in speech and co-speech gesture, but *not in* silent gesture. If, on the other hand, language has an effect on thought both online and beyond online production of speech (i.e., offline), we would expect to find evidence of patterns of cross-linguistic variability in both co-speech and silent gesture.

1.1 Packaging and Ordering of Semantic Elements in Speech

Motion in space constitutes a core human experience (Talmy, 1985, 2000), but the expression of motion varies considerably across different languages (Slobin, 2004). As proposed by Talmy (2000), the world’s languages can be categorized into two types, based on how speakers semantically package the two components of motion: manner (i.e., style of movement) and path (i.e., direction of movement). Speakers of satellite-framed languages (S-languages; e.g., English, German, Polish) use a conflated strategy in speech where they express manner in the main verb (e.g., run) and path outside the verb, typically in a path particle or a preposition (e.g., into), within the bounds of a single clause (e.g., "Adam RUNS INTO the house"). In contrast, speakers of verb-framed languages (V-languages; e.g., Turkish, Spanish, Japanese) use a separated strategy in speech where they opt to express path in the main verb and manner in an additional subordinate clause, (e.g., "Adem eve KOŞARAK GİRER"=Adam house-TO by RUNNING ENTER; Allen et al., 2017; Özçalışkan et al., 2016a, 2016b; Özçalışkan & Slobin, 1999). At the same time, the preference to encode path inside or outside the verb influences the extent to which speakers of each language type express manner or path components of motion. S-language speakers have the option to express manner in the verb and path outside the verb, which allows them to express both manner and path in a single clause frequently in their descriptions of motion (Ibarretxe-Antuñano, 2004; Özçalışkan, 2004, 2009; Özçalışkan & Slobin, 1999, 2000). Contrary to S-languages, V-language speakers use the main verb to express path and use subordinate clauses to express manner of motion. Given the extra processing load such constructions impose on production, V-language speakers are typically more likely to exclude manner from their descriptions, focusing only on path of motion (Cardini, 2010; Choi & Bowerman, 1991; Özçalışkan, 2015; Özçalışkan & Slobin, 1999, 2003; Özçalışkan et al., 2016a, 2016b). Such patterns of expression are pervasive and even influence speakers' expectations for how motion will be described at the neural level (Emerson, Conway, & Özçalışkan, 2020).

However, the binary classification into S- versus V-languages leaves out a third group of languages (serial-verb languages, such as Mandarin Chinese), which have two main verb slots (one for manner and one for path) both expressed within a single clause, with no explicit marking as to which one is the main verb (Chen & Guo, 2009). For instance, in Mandarin, in order to say "Adam runs into the house" two verbs need to be used, both equal in force: “pǎo” =run (manner verb) and “*jìn*” = enter (path verb; *Yàdāng PǍOJÌN fángzi* = Adam RUN-ENTER house), resulting in difficulties in the classification of Chinese as belonging to either of the two language types (Chen & Guo, 2009). In fact, the classification of Chinese within Talmy’s framework has been a center of debate: some researchers (e.g., Hsueh, 1989; Tai, 2003) argue that the main verb in a serial-verb construction is the path verb, classifying Chinese as a V-language; while others (Chao, 1968; Chang, 2001) argue that the main verb in a serial-verb construction is the manner verb and the path verb serves as a path satellite, placing Chinese within the group of S-languages (see Paul et al., 2021 for further discussion). A more recent approach, however, proposes a third possibility, suggesting that both verbs in a serial verb construction are of equal status and each can be used independently without the other in a clause (Chen & Guo, 2009; Slobin, 2004). According to this theory—originally proposed by Slobin (2004)—Chinese might be neither a V-language nor an S-language, but instead belong to a third category of languages, called "equipollently-framed language (E-language)." Unlike S-framed languages, Chinese allocates comparable weight to both manner and path components of motion, with the use of serial-verb constructions, in which manner and path are expressed in a compound verb (“pǎojìn = run-enter,”), thus designating Chinese as a good fit for the category of E-languages (Brown & Chen, 2013; Chen & Guo, 2009; Xu, 2013).

Languages also differ systematically in their ordering of semantic elements. The possible orders of the basic units of a sentence—the subject (S), verb (V), and object (O)—are not uniformly distributed across the world’s languages. In fact, two orders predominate 90% of the languages spoken around the world (Gell-Mann & Ruhlen, 2011; Gibson et al., 2013; Greenberg, 1963). The two orders include subject-object-verb (SOV; e.g., Turkish, Japanese, Korean, Mongolian) and subject-verb-object (SVO; e.g., English, Russian, French, Italian; Gibson et al., 2013; Shyu, 2001). For example, a description of a scene where a man runs into a house would typically follow the SOV order in Turkish; “Adam eve GİRER” =Adam (S) house-to (O) ENTER (V) and the SVO order in English; “Adam (S) RUNS (V) into house (O)”. Earlier work on speakers’ description of motion events in English (SVO order) and Turkish (SOV order) provided support for these patterns—with the majority of the descriptions in English and Turkish following the language-specific ordering of motion elements (Özçalışkan et al., 2016b, 2018).[1] Chinese, on the other hand, does not rely on a basic word order, mainly because it is a language claimed to be more discourse-oriented. That is, the word order variation is related to variations not only in major syntactic constituents, but also variability in the ordering of modifiers (e.g., marking of time or location) and pragmatic factors (e.g., factors associated with language use; Jin, 2008; Li & Thompson, 1981). For example, in describing simple motion events, Chinese speakers can choose between several options, including an SVO order when describing someone running into a house: “Yàdāng (S) PǍOJÌN (V) fángzi (O)”= Adam RUN-ENTER house (Jin, 2008), and an SOV order when describing someone crawling over a carpet: “Yàdāng (S) ZÀI DÌTǍN shàng (O) páxíng (V)”= Adam at carpet on(top) CRAWL (Goldin-Meadow et al., 2008). Thus, the status of Chinese as an SVO or an SOV language remains inconclusive, with some researchers classifying it as an SVO language (Jin, 2008), while others argue for it to be an SOV language (Shyu, 2001).

Overall, linguistic expression of motion events in different languages show strong but systematic cross-linguistic variation, both in packaging and ordering of semantic elements

[1] This pattern was evident even though Turkish allows for flexibility in word ordering because of its inflectional morphology (i.e., word endings) that marks syntactic information (e.g., tense, subject, direct object)

(Goldin-Meadow et al., 2008; Özçalışkan & Slobin, 1999; Özçalışkan et al., 2016b, 2018; Talmy, 2000). While Turkish (i.e., V-language) speakers rely on separated packaging strategy with SOV ordering; English (i.e., S- language) speakers prefer conflated packaging strategy with SVO ordering. Chinese falls somewhere between these two languages—with researchers classifying it as either an S-language (Talmy, 2000) or an E-language (Slobin, 2004) in its packaging of semantic elements and as following either the SVO (Jin, 2008) or the SOV order (Syhu, 2001) in its ordering of semantic elements.

1.2 Packaging and Ordering of Semantic Elements in Co-Speech Gesture

Speakers use their hands frequently when verbally describing motion events, thus producing co-speech gestures (McNeill, 2000). Earlier research on co-speech gestures accompanying motion event descriptions has shown that co-speech gestures largely exhibit the same cross-linguistic differences observed in speech (see Özçalışkan & Emerson, 2016, for a review). The one exception to this large body of work is an earlier study by McNeill (2000), which showed that Spanish (V-language) speakers used gestures to supplement motion components that they do not convey in speech (i.e., manner gestures accompanying speech expressing path of motion), thus *not* mirroring the patterns found in speech. However, apart from this work, most of the research comparing co-speech gestures in motion event descriptions in S- and V- languages suggests an augmentative role for gestures in which gesture conveys similar information as speech (Kita & Özyürek, 2003; Özçalışkan et al., 2016a, 2018). An earlier study (Kita & Özyürek, 2003) compared the speech and co-speech gestures produced by English (S-language), Turkish and Japanese (both V-languages) adult speakers and found language-specific patterns in both speech and co-speech gesture in each language group. Japanese and Turkish speakers were more likely to use separated co-speech gestures for manner and path of motion (e.g., wiggling fingers to convey running, moving index finger left to right to convey rightward trajectory), while English speakers were more likely to use the conflated strategy, expressing both path and manner simultaneously in their co-speech gestures (e.g., wiggling fingers left to right to convey running in rightward trajectory)—thus mirroring the patterns in their speech about motion.

More recent work extended this earlier work to blind speakers, asking whether the online effect of language-specific speech patterns in co-speech gesture also becomes evident even in the absence of visual access to language-specific gestures. Özçalışkan and colleagues (2016a, 2018) examined the co-speech gestures (along with speech) of blind and sighted adult English and Turkish speakers produced when describing 3-dimensional motion scenes. Their results showed that while Turkish speakers primarily relied on the separated packaging strategy in their co-speech gestures, encoding primarily path of motion, English speakers relied largely on conflated gestures, combining manner and path components into a single gesture—a pattern that remained identical for both sighted and blind speakers. Several other studies that expanded the applicability of packaging strategies to various other S- and V-languages, showed similar patterns as well, with V-language speakers producing more separated gestures (e.g., French; Gullberg et al., 2008; Spanish: Wieselman Schulman, 2004) and S-language speakers using more conflated co-speech gestures (e.g., German; Lewandowski & Özçalışkan, 2018).

In contrast to considerable work on the packaging of semantic elements in co-speech gesture for S- and V-languages, there is very little work that examines gestures produced by speakers of E-languages that show speech patterns characteristic of both S- and V-languages. The few existing studies on the co-speech gestures produced by speakers of E-languages (almost exclusively on Chinese) suggest a predominance of separated gestures that mostly convey path of motion (Brown & Chen, 2013; Chui, 2012). An earlier study (Brown & Chen, 2013) examined the speech and co-speech gestures produced by English (S- language), Japanese (V- language) and Chinese (E- language) speakers based on animated motion scenes. The patterns for English and Japanese motion descriptions replicated earlier work, with Japanese speakers using more separated and English speakers using more conflated speech and co-speech gestures. Chinese speakers, on the other hand, used primarily separated gestures encoding path of motion, even though their speech contained more manner information, particularly compared to Japanese speakers, thus suggesting partial support for the "thinking-for-speaking hypothesis".

Research on ordering of semantic elements in co-speech gesture remains relatively sparse. The few existing studies provide evidence for the "thinking-for-speaking hypothesis" by showing that the ordering of semantic elements in co-speech gesture largely follows the ordering of semantic elements in speech in a given language. An earlier study, in which English (S- language), Turkish and Japanese (V-languages) adults were asked to describe scenes involving simple transitive actions (e.g., woman puts on hat, man swings pail), demonstrated that English speakers follow the Actor-ACTION-Object order, while Turkish and Japanese speakers follow the Actor-Object-ACTION order in their co-speech gestures, thus mirroring the two distinct word order patterns seen in their native speech—SVO vs. SOV—in describing the same scenes (Goldin-Meadow et al., 2008). More recent work by Özçalışkan and colleagues (2016b, 2018) extended this work to motion events, also showing similar alignment between speech and co-speech gesture: English speakers were more likely to use the Figure-MOTION-Ground order (e.g., place index finger in front of body and run it forward to imaginary landmark) which mirrors the SVO order in their speech. In contrast, Turkish speakers were more likely to use Figure-Ground-MOTION order, which corresponds to the SOV order in their speech about motion. However, there is no research examining ordering of semantic elements in co-speech gesture among E-language speakers, such as Chinese.

In summary, packaging and ordering of semantic elements in co-speech gestures produced by S- and V-language speakers show strong cross-linguistic differences, and these differences mirror the differences found in speech, thus providing support for the “thinking-for-speaking hypothesis”. Patterns of co-speech gesture production remains relatively understudied for E-languages, marking a significant gap in the literature.

1.3 Packaging and Ordering of Semantic Elements in Silent Gesture

Gesture mirrors the cross-linguistic differences observed in speech when it is accompanied by speech. Recent research took this one step further asking whether the effect of language on thought (as assessed through gesture) can extend beyond online production of speech and become evident ‘offline’ when describing events without speech, using only gestures (i.e., silent gesture). Earlier studies by Özçalışkan and colleagues (Özçalışkan, 2016; Özçalışkan et al., 2016b, 2018) compared the co-speech and silent gestures produced by adult Turkish and English-speaking participants and showed evidence to the contrary. More specifically, they showed that even though English and Turkish speakers differed in their co-speech gestures about motion, they did *not differ* in their silent gestures. In fact, speakers of both languages almost exclusively used the conflated strategy in their silent gestures, synthesizing manner and path components into a single gesture. This effect has been shown across different groups, including sighted, blind (Özçalışkan et al., 2016b, 2018), as well as monolingual and bilingual participants speaking English or Turkish (Özçalışkan, 2016). These results, thus, suggest that the effect of cross-linguistic variation on the packaging of the semantic elements in gesture appears *only* during online production of language (i.e., co-speech gestures), but not during ‘offline’ production (i.e., silent gesture). Given the high levels of cross-linguistic similarity in the preferred packaging strategy in silent gesture, these studies also raised the possibility for a natural semantic organization that humans might impose on motion events when conveying them nonverbally, in silent gesture (Özçalışkan et al., 2016b). There is no existing research examining patterns of silent gesture production for packaging of semantic elements among E-language speakers.

Research on ordering of semantic elements in silent gestures shows similar patterns to the ones observed for packaging of semantic elements. In an earlier study that involved description of simple transitive action scenes (e.g., captain swings pail), Goldin-Meadow and colleagues (2008) examined how English, Turkish, Spanish and Chinese speakers ordered transitive event components when speaking and subsequently when *not* speaking. Using both gesture and an additional picture sorting task (in which participants were asked to organize pictures of actor, object and action), they found that all four languages used the same Actor-Object-ACTION order on both nonverbal tasks without speech, suggesting that effect of language does not extend beyond online production of speech. More recent studies (2016b, 2018) examining ordering of semantic elements for motion events in silent gestures produced by blind and sighted English speakers (Özçalışkan et al., 2016a, 2018), as well as bilingual English-Turkish speakers (Özçalışkan, 2016), showed that all the participants gestured in the same order when they gestured without speech. More specifically, across the different samples of participants in each study, the English speakers abandoned the Figure-MOTION-Ground (SVO) order that they used in speech and in co-speech gesture and replaced it with Figure-Ground-MOTION (SOV) order in silent gesture, thus mirroring the pattern found in Turkish speakers' silent gestures.

In summary, as evidenced across several studies—mostly focusing on S- and V-languages, the cross-linguistic differences observed in co-speech gesture dissipate when speakers are asked to describe events without speech, only in silent gesture. These results thus suggest that non-verbal representation of events in gesture without speech might elicit a natural order of semantic elements that humans prefer regardless of the language they speak (Özçalışkan et al., 2016a).

1.4 Current Study

Previous research has found that English (i.e., S-language) and Turkish (i.e., V-language) speakers show cross-linguistic differences in their packaging (conflated vs. separated) and ordering (Figure-MOTION-Ground vs. Figure-Ground-MOTION) of semantic elements of an event (Goldin-Meadow et al., 2008; Özçalışkan et al., 2016a, 2016b, 2018). These differences, on the other hand, disappear when the gestures are not accompanied by speech (i.e., silent gesture). Instead, English and Turkish adults demonstrate the same packaging (conflated) and ordering (Figure-Ground-MOTION) patterns when they use silent gesture (Özçalışkan et al., 2016a, 2016b, 2018). These results support the “thinking-for-speaking hypothesis” proposed by Slobin (1996), which argues that language influences thought, but only during the process of online speech production. However, we do not yet know whether in Mandarin Chinese, an E-language, speakers follow the same patterns in their co-speech gestures as in their speech, but differ from speech in their patterns of silent gesture production, thus mirroring the patterns of silent gesture observed among English and Turkish speakers.

In this study, we focused on the patterns of motion event descriptions in speech, co-speech gesture and silent gesture produced by adult Chinese speakers (E-language), comparing their production to the speech and gestures produced by adult English (S-language) and adult Turkish (V-language) speakers.

(1) We first asked whether Chinese, English and Turkish speakers would differ in the way they packaged and ordered motion elements in their *speech about motion*. We expected that our results would replicate earlier work on speech about motion, showing strong cross-linguistic differences between the three languages (Chen & Guo, 2009; Goldin-Meadow et al., 2008; Kita & Özyürek, 2003; Özçalışkan 2016a, 2016b, 2018; Slobin, 2004). More specifically, we expected greater reliance on conflated packaging strategy and Figure-MOTION-Ground order in English, and greater reliance on separated packaging strategy and Figure-Ground-MOTION order in Turkish. We also predicted that Chinese speakers would differ from both English and Turkish speakers in packaging and ordering of semantic elements in their speech about motion. We expected Chinese speakers to rely on the conflated packaging strategy more than Turkish but less than English speakers, a pattern that we expected to be reversed for the separated packaging strategy. We also expected that Chinese speakers would use Figure-MOTION-Ground order more frequently than Turkish but less frequently than English speakers, a pattern that we expected to be reversed for Figure-Ground-MOTION order.

(2) We next asked whether Chinese, English and Turkish speakers would differ in the way they packaged and ordered motion elements in their *co-speech gestures about motion*. We expected co-speech gestures to follow the patterns observed in speech—with greater use of conflated packaging and Figure-MOTION-Ground order in English and separated packaging and Figure-Ground-MOTION order in Turkish, based on earlier work (Goldin-Meadow et al., 2008; Özçalışkan, 2016a, 2016b). We also predicted that Chinese speakers would differ from English and Turkish speakers in packaging and ordering of semantic elements in their co-speech gestures

about motion and mostly follow the patterns observed in their speech (Brown & Chen, 2013; Chui, 2012): Chinese speakers would use the conflated packaging strategy and Figure-MOTION-Ground order less than English but more than Turkish speakers; as a corollary to this, we predicted that they would use the separated packaging strategy and Figure-Ground-MOTION order more than English but less than Turkish speakers.

(3) We last asked whether Chinese, English and Turkish speakers would differ in the way they packaged and ordered motion elements in their *silent gestures about motion*. Based on earlier work showing no cross-linguistic difference in silent gestures of S- and V-language speakers (e.g., Özçalışkan et al., 2016b, 2018), we predicted that speakers of all three languages would rely on the conflated packaging strategy and Figure-Ground-MOTION order in their silent gestures about motion, showing no cross-linguistic differences.

Overall, if the hypotheses were supported, our findings would provide further evidence for the “thinking-for-speaking hypothesis” (Slobin, 1996), which states that language influences thought, but only during online production of speech (i.e., co-speech gesture), but not when thinking ‘offline’ (i.e., silent gesture).

2 METHODS

2.1 Participants

Participants included 60 adult speakers, either with Chinese (n=20, M_{age} = 19.55 [SD = 1.36], range = 18-23, 10 females), English (n=20, M_{age} = 18.95 [SD = 1.10], range = 18-22, 13 females), or Turkish (n=20, M_{age} = 20.8 [SD = 1.76], range = 18-24, 10 females) as their native language. Data from the Chinese, English, and Turkish participants were collected in Jingzhou City Hubei Province (China), Atlanta (USA), and Istanbul (Turkey), respectively by native speakers in each language. The sample size was based on a similar earlier study that showed that n=20 per group was adequate to detect reliable effects at $p < .05$ (Özçalışkan et al., 2016b), along with a power analysis, which indicated that n=20 per group was adequate to detect reliable effects ($p < .05$) with a medium effect size (.25) and power of .99 (Faul et al. 2007). The majority of the participants in each language had some knowledge of another language, but none of the participants was fluent in a second language. The speakers of each language were comparable in education: the majority of participants in each language were either college students or recent college graduates. The participants were compensated by either course credit or small monetary compensation for their participation in the study.

2.2 Procedure for Data Collection

Each participant was interviewed individually by a native speaker in each language, using an animated motion description task, originally developed by Özçalışkan (2016). At the beginning of the interview, participants were introduced to an animated cartoon character named Adam, who performed the motion events in the animations. Each participant was then presented with 16 animated motion events with various manners and paths, one at a time (see Table 2.1 for a list of the 16 animated motion events).

Table 2.1 List of The Animated Motion Event Scenes

Event Description	**Type of PATH**	**Type of MANNER**
Set 1		
Crawl into a house	INTO a bounded space	Crawl
Flip over a beam	OVER a bounded space	Flip
Run out of a house	OUT of a bounded space	Run
Crawls towards a carpet	TOWARD a bounded space	Crawl
Set 2		
Climb into a treehouse	INTO a bounded space	Climb
Jump over a hurdle	OVER a bounded space	Jump
Fly out of a trashcan	OUT of a bounded space	Fly
Climb toward a treehouse	TOWARD a bounded space	Climb
Set 3		
Tumble into a trashcan	INTO a bounded space	Tumble
Crawl over a carpet	OVER a bounded space	Crawl
Crawl out of a house	OUT of a bounded space	Crawl
Flip toward a beam	TOWARD a bounded space	Flip
Set 4		
Run into a house	INTO a bounded space	Run
Jump over a cat	OVER a bounded space	Jump
Tumble out of a treehouse	OUT of a bounded space	Tumble
Crawl toward a house	TOWARD a bounded space	Crawl

They were asked to describe the events in two different ways—once with speech (**co-speech gesture condition**; “You’re going to watch short video clips. Describe each video after each clip. Make sure to use your hands as naturally as possible and also include the landmark in your description”), and once without speech (i.e., **silent gesture condition**; i.e., “Now you’re going to see the same short video clips one more time. This time, I want you to only use your hands while describing the videos without using any speech.”). To avoid influencing the naturalness of co-speech gestures, participants first described the 16 animated clips in the co-speech gesture condition and then in the silent gesture condition. However, the presentation order of the 16 animated motion clips was counterbalanced across participants within both the co-speech gesture and silent gesture condition, with each participant completing the descriptions in one of 4 different set orders in both conditions —following procedures in earlier work (Özçalışkan, 2016; Özçalışkan et al., 2016b). Each participant completed two practice trials prior to describing the scenes in the co-speech gesture and the silent gesture condition to familiarize them with the demands of the task; these trials were not included in the analysis.

2.3 Transcription, Coding and Reliability

All speech produced in the co-speech gesture condition was transcribed and segmented into sentence-units by native speakers in each language. A sentence-unit was defined as a segment of speech that contains at least one verb and its associated arguments and subordinate clauses, following earlier work (Özçalışkan et al., 2016a; e.g., “Adam RUNS into a house”; “Adem eve KOŞARAK GIRER” = Adam house-to by RUNNING ENTERS; “Yàdāng PAO JÌN fángzi” = Adam RUN-ENTER house). All gestures that accompanied each sentence-unit in the co-speech gesture condition and that were produced on their own in the silent gesture condition were also coded. Gesture was defined as a communicative hand or body movement that was communicative and conveyed meanings; only gestures that indicated or characterized motion, figure or the ground associated with the animated stimulus scenes were coded, following earlier work (Özçalışkan 2016).

Each sentence-unit was further coded for packaging of semantic elements and for the ordering of semantic elements, separately for speech, co-speech gesture, and silent gesture, following earlier work (Özçalışkan et al. 2016b, 2018). For the **packaging of semantic elements**, each sentence-unit was coded as either *conflated* (i.e., manner and path are both described within a single clause or within a single gesture) or *separated* (manner and path are described in separate clauses or in separate gestures). A sentence-unit was classified as

separated if it contained manner-only (e.g., “he RUNS”, “Adam KOŞAR= he RUNS”, “Tā PǍO = he RUNS”), path-only (e.g., “he ENTERS the house”, “Adam eve GİRER = he house-TO ENTER”, “Tā JÌN ru fangzi= he ENTERS house”), or manner and path, conveyed in two separate clauses (e.g., “he enters the house by running”, “Adam eve KOŞARAK GİRER=Adam house-to by RUNNING ENTERS”, “Ta tong guo PǍOJÌN ru fangzi= he by RUNNING-ENTERS house to”). A gesture was classified as **separated** if it contained manner-only (e.g., wiggling fingers rapidly as if running), path-only (e.g., moving index finger left to right as if conveying rightward trajectory) or manner and path conveyed in two separate gestures (e.g., wiggling fingers in place as if running then moving index finger left to right as if conveying rightward trajectory). A sentence-unit or gesture was classified as **conflated** if it synthesized manner and path into a single clause (e.g., "Adam RUNS INTO the house”; “Adam eve KOŞAR” =Adam house-TO RUN; “Yàdāng PǍOJÌN fangzi” = Adam RUN-ENTER house) or a single gesture (e.g., wiggling fingers left to right as if running left to right (see Figure 2.1).

*Figure 2.1 Example stimulus scene of a figure climbing into a treehouse (top) and its description in co-speech gesture (A_1, B_1, C_1,) and silent gesture (A_2, B_2, C_2) by speakers of English (A pictures on left), Chinese (B pictures in the middle) and Turkish (C pictures on right). In **co-speech gesture**, native speakers of English and Chinese preferred to express manner (circling fists) and path (upward trajectory) simultaneously in a single gesture (A_1-B_1), and native speakers of Turkish preferred to express path (leftward trajectory) by itself, omitting manner (C_1). In **silent gesture**, native speakers of English, Chinese or Turkish preferred to express manner and path simultaneously within a single gesture by circling fists upward (A_2, B_2, C_2).*

For the **ordering of semantic elements** each sentence-unit was coded as either Figure-Ground-MOTION or Figure-MOTION-Ground, according to the placement of the primary motion element. The primary motion element was the main verb of the sentence unit in speech and the gestural element conveying motion in gesture (manner only, path only, manner-path sequential, manner+path conflated gesture) within a single sentence unit (Özçalışkan et al., 2018). For example, if a participant placed his left palm on the left (i.e., Ground) and right finger on the right (i.e., Figure), and then moved his right finger towards the left palm (i.e., MOTION), the sentence-unit was coded as Figure-Ground-MOTION. If, however, the participant placed her finger on the right (i.e., Figure), wiggled it left to right (i.e., MOTION), then placed left cupped hand on the left (i.e., Ground), the sentence-unit was coded as following the Figure-MOTION-Ground order (see Figure 2.2).

*Figure 2.2 Example stimulus scene of a figure climbing into a treehouse (top) and its description in co-speech gesture (A_1, B_1, C_1,) and silent gesture (A_2, B_2, C_2) by speakers of English (A pictures on left), Chinese (B pictures in the middle) and Turkish (C pictures on right). In **co-speech gesture**, some English speakers preferred to express motion (climb.into) first, followed by ground (treehouse); Chinese and Turkish speakers preferred to express ground (treehouse) first, followed by motion (climb/move into). In **silent gesture**, English, Chinese and Turkish speakers preferred to express ground (treehouse) first, followed by motion (climb.into).*

2.4 Scoring and Analysis

We computed the total number of sentence-units with separated versus conflated motion packaging strategy and the total number of sentence-units with Figure-MOTION-Ground versus Figure-Ground-MOTION order that each speaker produced in speech, co-speech gesture and silent gesture, separately in each language. We analyzed differences, using two-way mixed ANOVAs, with language as a between subjects factor (English, Turkish, Chinese) and either packaging type (separated, conflated) or ordering type (Figure-MOTION-Ground, Figure-Ground-MOTION) as within subject factors, separately for speech, co-speech gesture and silent gesture. In two of analyses (ordering in co-speech and silent gesture) where the assumptions of the ANOVA were violated, we first tested differences with ANOVAs to observe possible interaction effects, but we also followed up these analyses with non-parametric tests (i.e., Kruskal-Wallis) to determine whether the patterns of similarities and differences remained the same in both tests.

All gestures and speech produced by speakers in each language were coded by native speakers of the language, who were trained for speech and gesture coding. Inter-coder reliability was established by additional coders in each language—also native speakers— who coded a randomly selected 20% of the speech, co-speech and silent gestures in each language. Reliability was assessed with independent coders; agreement between coders was 97% for identifying gestures, 99% for describing gesture form, and 97% and 98% for coding motion elements in speech and gesture, respectively.

3 RESULTS

3.1 Packaging of Semantic Elements

3.1.1 Speech

Speakers of Chinese, English and Turkish differed in their packaging of motion components in speech (Language x Packaging interaction: ($F(2, 57) = 10.59$, $p < .001$, η^2_p = 0.64). English and Chinese speakers produced more conflated than separated responses (Bonferroni, *p's* < .001): They expressed manner and path in the same clause, using predominantly manner verbs with path satellites in English (e.g., *He RUNS INTO house*; 72.38%), and manner verb+path verb serial verb constructions (e.g., *Tā PǍOJÌN fángzi = Adam RUN-ENTER house*; 64.18%) or manner verbs with path satellites (e.g., *Tā WǍNG fángzi LǏ PǍO = He TOWARDS house INSIDE RUN*; 7.45%) in Chinese. Turkish speakers showed the opposite pattern, using more separated than conflated packaging strategies (Bonferroni, $p <$ .001): they described the same scenes by expressing only path (e.g., *Adam eve GIRER = Adam house-TO ENTERS*; 36.73%), only manner (e.g., *Adam KOŞUYOR = Adam is RUNNING*; 33.16%) or path in one clause and manner in a subordinate clause (e.g., *Adam eve KOŞARAK GİRER"=Adam house-TO by RUNNING ENTERS*; 21.58%)[2]; see Figure 3.1.A.

3.1.2 Co-Speech Gesture

The co-speech gestures of the speakers of the three languages mirrored the patterns observed in speech, showing a significant Language × Packaging interaction; $F(2, 57) = 23.14$, *p*

---

[2] We also found a main effect of packaging ($F(1, 57) = 14.54$, $p < .001$, $\eta^2_p = .20$)—with greater production of conflated than separated responses ($M = 11.86$, $SE = .43$ vs. $M = 9.43$, $SE = .38$, $p < .001$), and a main effect of language ($F(2, 57) = 16.81$, $p < .001$, $\eta^2_p = .37$)—with greater production of speech responses by Chinese speakers ($M = 12.72$, $SE = .45$) than both English ($M = 9.32$, $SE = .45$, $p < .001$) and Turkish ($M = 9.88$, $SE = .45$, $p < .001$).

$< .001$, $\eta^2_p = .45$). English and Chinese speakers both produced more conflated than separated responses (Bonferroni, $p = .001$, $p < .001$, respectively) by synthesizing manner and path into the same gesture (e.g., wiggling fingers left to right as if running into; English: 67.80%, Chinese: 77.65%). Turkish speakers showed the opposite pattern, producing more separated than conflated responses in gesture (Bonferroni, $p = .001$): they used gestures to express only path (e.g., moving both hands left to right as if entering; 7.55%), only manner (e.g., circling fists rapidly next to body as if running; 47.79%), or separate gestures for manner and path but within the bounds of a sentence-unit (e.g., first moving both arms up and down with fisted hands next to body as if running then moving both hands left to right as if entering; 8.81 %; Figure 3.1.B) [3].

3.1.3 Silent Gesture

As expected, the three languages showed similarities in the packaging of motion elements in silent gesture, with a main effect of packaging type ($F(1, 57) = 240.71$, $p < .001$, $\eta^2_p = .81$), a main effect of language ($F(2, 57) = 7.79$, $p = .001$, $\eta^2_p = .22$), but no interaction between language and packaging ($F(2, 57) = 2.21$, $p = .12$). Speakers—regardless of language—produced a greater number of conflated than separated responses ($M_{conflated\ all}$= 12.38 [SE= .38] vs. $M_{separated\ all}$= 2.30 [SE= .31]) (see Figure 3.1.C).

[3] There was no main effect of language $F(2, 57) = 2.57$, $p = .08$, but a main effect of packaging in co-speech gesture, $F(1, 57) = 12.06$, $p = .001$, $\eta^2_p = .18$—with overall greater use of conflated than separated packaging strategy ($M = 8.53$, $SE = .50$ vs. $M = 5.87$, $SE = .41$).

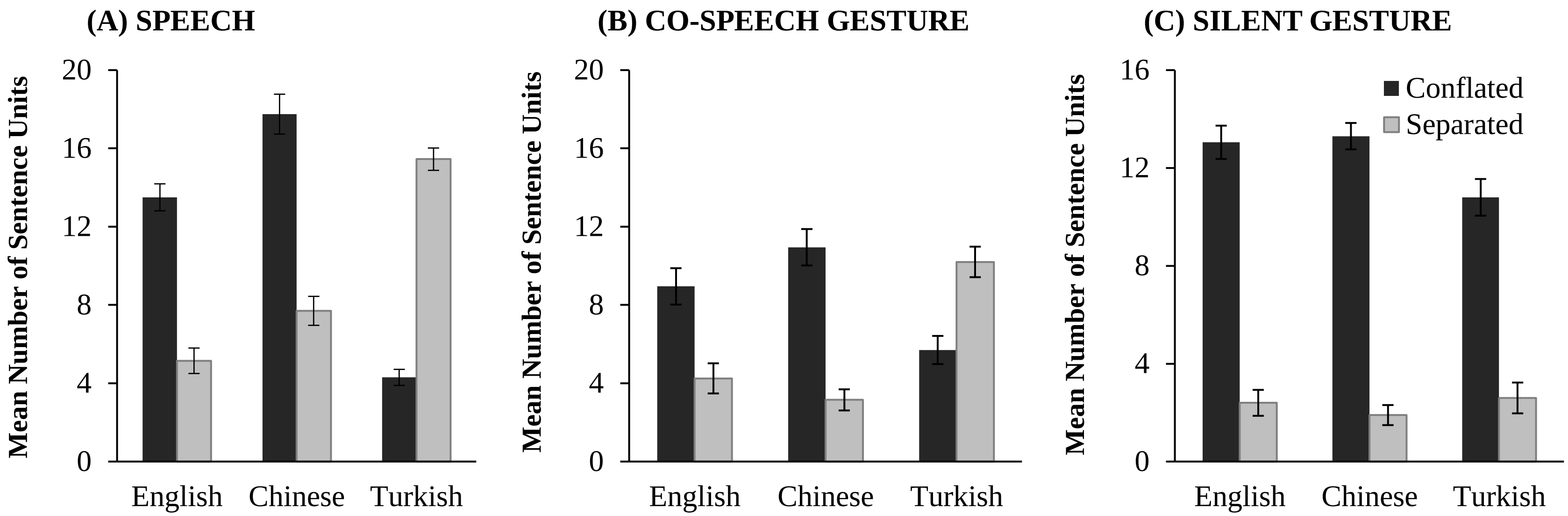

Figure 3.1 Mean number of sentence units with separated (manner-only, path-only, manner-path) or conflated (manner + path) motion elements in speech (A), in gesture with speech (co-speech gesture, B) and in gesture without speech (silent gesture, C); Error bars represent the standard error; the max possible number of sentence units was 16 for silent gesture condition.

3.2 Ordering of Semantic Elements

3.2.1 Speech

Speakers of Chinese, English and Turkish differed in their ordering of semantic elements in speech (Language × Order interaction, $F(2,57)=324.48$, $p < .001$, $\eta^2_p = 0.92$). As expected, English speakers produced the Figure-MOTION-Ground order more often than Figure-Ground-MOTION order (Bonferroni, $p < .001$). The pattern was reversed for Turkish speakers: they produced Figure-Ground-MOTION order more often Figure-MOTION-Ground order (Bonferroni, $p < .001$). Chinese speakers, on the other hand, did not differ in their production of the two order types, producing each at comparable rates (Bonferroni, $p = .15$[4]; see Figure 3.2.A).

3.2.2 Co-Speech Gesture

Speakers of Chinese, English and Turkish also differed in their ordering of semantic elements in co-speech gesture (Language × Order interaction, $F(2,57) = 324.48$, $p = .006$, $\eta^2_p = 0.16$). Chinese speakers produced significantly more Figure-Ground-MOTION than Figure-MOTION-Ground order (Bonferroni, $p < .001$)—a pattern that was also observed but not reliable in Turkish (Bonferroni, $p = .14$). English speakers tended to produce the opposite pattern, with slightly more Figure-MOTION-Ground than Figure-Ground-MOTION order responses in their co-speech gestures, however this tendency was not significant either (Bonferroni, $p = .60$; see Figure 3.2.B)[5].

---

[4] We found no main effect of order ($F(1,57)=0.009$, $p = .93$) or language ($F(2,57)=2.58$, $p = .08$) in speech.

[5] We found no effect of language ($F(2, 57) = 1.64$, $p = .20$) but a main effect of order ($F(1,57) = 8.91$, $p = .004$, $\eta^2_p = .14$) with greater production of Figure-Ground-MOTION than Figure-MOTION-Ground order ($M = 2.00$, $SE = .36$ vs. $M = 0.85$, $SE = 0.15$) in co-speech gesture. Participants infrequently combined gestures into strings, resulting in non-normal distribution of the data. We therefore followed up the ANOVAs, with Kruskal-Wallis tests and found reliable differences only for the Figure-Ground-MOTION order ($H(2)= 13.77$, $p = .001$)--with English

3.2.3 Silent Gesture

Turning next to silent gesture, again we found that Chinese-, Turkish- and English-speakers did not display the differences found in their speech and co-speech gesture. Even though there was a cross-linguistic difference in their ordering of semantic elements in silent gesture (Language × Order interaction, $F(2,57) = 14.98$, $p < .001$, $\eta^2_p = 0.34$), the difference was an outcome of more pronounced preference for Figure-Ground-MOTION ordering in Chinese than in English or Turkish (Bonferroni, $p's < .001$). Speakers showed a main effect of ordering, with speakers of all three languages producing Figure-Ground-MOTION order significantly more than Figure-MOTION-Ground order in their silent gestures ($M_{Figure\text{-}Ground\text{-}MOTION}$=7.51 [SE= .56] vs. $M_{Figure\text{-}MOTION\text{-}Ground}$ =.72 [SE= .20], $F(1,57) = 115.12$, $p < .001$, $\eta^2_p = 0.68$)[6]; see Figure 3.2.C.

---

speakers producing significantly fewer responses with Figure-Ground-MOTION order than both Chinese and Turkish speakers ($p < .05$).

[6] We also found a main effect of language ($F(2,57) = 10.98$, $p < .001$, η^2_p =0.28) with greater production of silent gesture by Chinese speakers ($M = 5.85$, $SE = .48$) than both English ($M = 3.78$, $SE = .45$, $p < .001$) and Turkish ($M = 2.72$, $SE = .48$, $p < .001$) speakers. Participants showed individual variability in their likelihood of combining gestures into strings in silent gesture, resulting in non-normal distribution of the data. We therefore followed up the analysis with Kruskal-Wallis tests. We found differences both for the Figure-Ground-MOTION order ($H(2)= 19.70$, $p < .001$) and Figure-MOTION-Ground order ($H(2)= 11.68$, $p = .003$) --with Chinese speakers producing significantly more responses with Figure-Ground-MOTION order and less Figure-MOTION-Ground order compared to English and Turkish speaking participants (*ps* < .05). English and Turkish did not show any differences for either order (*ps* > .05)

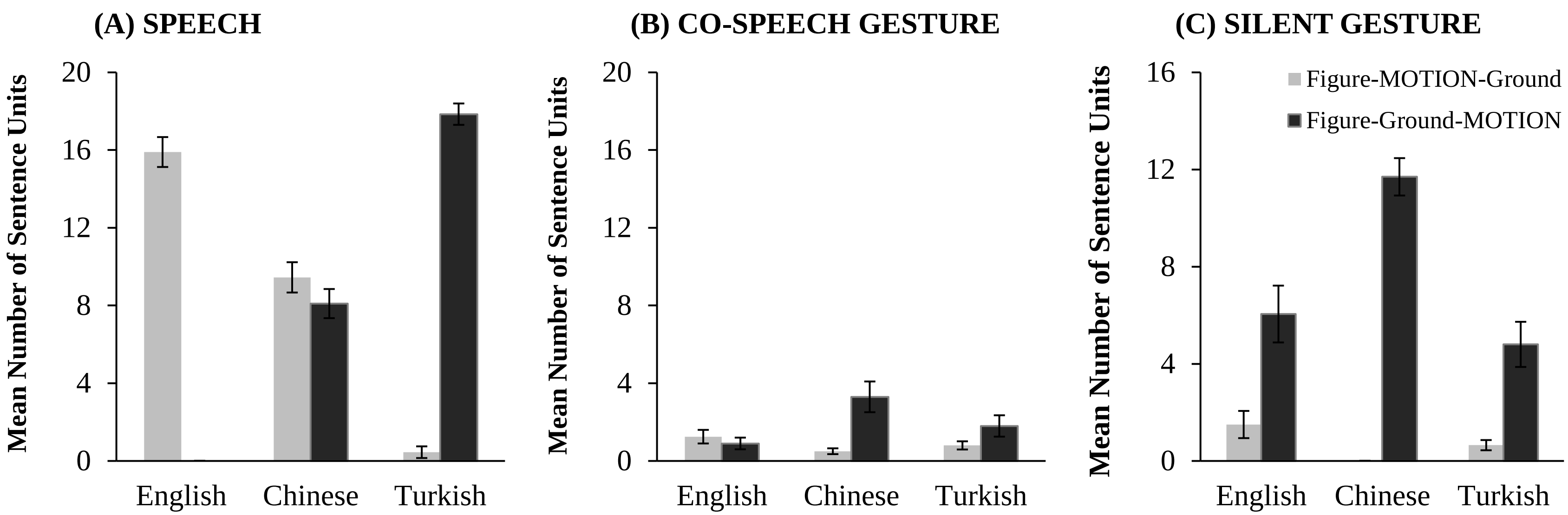

Figure 3.2 Mean number of sentence units that follow Figure-Ground-MOTION or Figure-MOTION-Ground orders in speech (A), in gesture with speech (co-speech gesture, B) and in gesture without speech (silent gesture, C). Error bars represent standard error; the max possible number of sentence units was 16 for silent gesture condition.

4 DISCUSSION

Previous research focusing on S- and V- languages demonstrated that languages differ in the packaging (conflated vs. separated) and ordering (Figure-MOTION-Ground vs. Figure-Ground-MOTION) of semantic elements of a motion event in speech and co-speech gesture but not in silent gesture (e.g., Özçalışkan et al., 2016a, 2016b, 2018). In the current study, we took these findings one step further and asked whether patterns of packaging and ordering of motion events found in speech, co-speech gesture and silent gesture extend to Chinese, an E- language. We found that English, Turkish and Chinese speakers showed cross-linguistic differences in the way they packaged and ordered semantic elements of a motion event in their speech and co-speech gestures. However, these cross-linguistic differences dissipated in silent gesture: all languages including Chinese demonstrated the same packaging (conflated) and ordering (Figure-Ground-MOTION) patterns. Our results thus provide further support for the "thinking-for-speaking hypothesis" proposed by Slobin (1996), which argues that language influences thought only during the process of online speech production (i.e., speech and co-speech gesture), but not in offline production (i.e., silent gesture).

4.1 Packaging of Semantic Elements

Our study follows earlier work in patterns of speech production, with English speakers showing a preference for conflated and Turkish speakers showing a preference for separated packaging strategies (Kita & Özyurek, 2003; Özçalışkan et al., 2016a) and Chinese following a pattern akin to English (Chen & Guo, 2003; Paul et al., 2021) also relying on conflated packaging strategy in their speech about motion. In addition to replicating earlier work on patterns of speech production, our study also extended these findings to gesture, showing that the co-speech gestures—including those produced by Chinese speakers— mirrored the patterns found in speech. These results thus provide further evidence that speech and co-speech gesture form a tightly integrated system (Kita & Özyürek, 2003; McNeill, 1992).

One of our predictions was that Chinese participants would show patterns characteristic of both English and Turkish participants in terms of their packaging of motion in both speech and co-speech gesture. Our results, however, showed that Chinese speakers were more closely aligned with English speakers, also showing greater preference for the conflated strategy in both speech and co-speech gesture. One likely explanation could be the use of serial-verb constructions in Chinese ("*pǎo jìn* = run-enter"), which allows for easy encoding of manner and path within the bounds of a single clause (Brown & Chen, 2013). This pattern also follows recent work (Paul et al., 2021), which showed that Mandarin Chinese speakers relied heavily on the conflated strategy, using primarily serial verbs (70%) in describing motion events.

Importantly, the differences that we observed in co-speech gesture were not evident when speakers described each scene only in silent gesture, without using language—a pattern that has been shown for both English and Turkish in an earlier study (Özçalışkan et al., 2016b, 2018), but one that has been shown for the first time in Chinese. Speakers of all three languages including Chinese resorted to the conflated packaging strategy when describing events in silent gesture—even in Turkish, where the speakers had to abandon the separated packaging strategy which they relied on almost exclusively in their speech and co-speech gestures. Furthermore, even English and Chinese speakers boosted the level of conflated gestures they used compared to their co-speech gestures when describing the same scenes in silent gesture. What might underlie this preference for conflated packaging that expresses both manner and path together in a single clause or gesture? One possible explanation—as also suggested by Özçalışkan and colleagues (2016b)—could be that the preference might be driven by the need to convey maximal information with limited effort. Turkish speakers need an adjunct or a subordinate clause to express manner in descriptions that use path verbs, thus resulting in greater syntactic complexity and greater omission of manner information in speech about motion (Özçalışkan & Slobin, 1999). Gesture, on the other hand, provides an opportunity to express both the manner and the path information simultaneously, thus creating a relatively easy tool to express both components of motion.

4.2 Ordering of Semantic Elements

The ordering of motion elements showed a similar pattern to packaging where cross-linguistic differences observed for ordering in speech and co-speech disappeared in silent gesture. Chinese speakers used both Figure-MOTION-Ground and Figure-Ground-MOTION orders equally in their speech thus replicating earlier work on speech ordering in Chinese (Goldin-Meadow et. al., 2008; Jin, 2008). Turkish and English differed strongly, however, with a preference for a Figure-MOTION-Ground order in English and Figure-Ground-MOTION order in Turkish, also replicating earlier work (Özçalışkan et al., 2016b, 2018).

We also expected to see a cross-linguistic difference in co-speech gestures mirroring the order patterns found in speech for each language. However, similar to previous research on ordering of semantic elements in co-speech gesture (Özçalışkan et al., 2016b), we found that none of the languages including Chinese produced many gesture strings when speaking. Even though English speakers produced slightly more Figure-MOTION-Ground order and Turkish speakers produced slightly more Figure-Ground-MOTION order— mimicking the patterns of their speech in ordering— none of these patterns were statistically reliable largely due to the limited number of gesture strings available in co-speech gesture. Chinese speakers, who also produced relatively fewer gesture strings than speakers of both other languages, showed a preference for Figure-Ground-MOTION order—an order that differed from their speech, which showed evidence of both orders. However, the relatively small numbers of gesture strings that all three groups of speakers produced in co-speech gesture (Figure-MOTION-Ground order: English 6.88%, Chinese 1.46%, Turkish 3.76%; Figure-Ground-MOTION order; English 4.76%, Chinese 9.64%, Turkish 8.71%) renders it impossible to draw broader conclusions from these patterns.

Importantly, however, speakers showed a lack of a cross-linguistic difference when describing the same scenes only with their hands, in silent gesture, with an overwhelming preference for Figure-Ground-MOTION order, thus replicating previous work on motion events (Özçalışkan et al., 2016a, 2016b, 2018) as well as other event types (i.e., transitive events; Goldin-Meadow et. al., 2008). However, this time it was English and Chinese speakers who abandoned the Figure-MOTION-Ground order that they used in their speech. What might explain this preference for Figure-Ground-MOTION ordering? One possible explanation—also suggested in earlier work (Özçalışkan et al., 2016b) —could be that Figure-Ground- MOTION order might be cognitively easier (Gentner, 1982; Goldin-Meadow et. al., 2008; Özçalışkan et al., 2016b). More specifically, describing the figure and the ground before the motion might ease the processing load and it might be specifically effective in silent gestures as gesture does not allow for grammatical marking of who did what to whom (Goldin-Meadow et. al., 2008; Özçalışkan et al., 2016b). Moreover, the Figure-Ground-MOTION order has been shown to appear in emerging sign languages (Goldin-Meadow & Feldman, 1977; Goldin-Meadow & Mylander, 1998). For instance, a study by Goldin-Meadow and Mylander (1998) showed that American and Chinese deaf children whose hearing parents did not expose them to a conventional sign language displayed consistent OV (Ground-MOTION) order in their gestures which may indicate a default way that humans conceptualize the order of motion events across languages.

In our study, we focused on standard dialects of each of the three languages. However, we know from previous research that the expression of motion in speech also shows significant dialect-based variability (e.g., Paul et al., 2021). As such, future work that examines how dialectical variability within a language system could influence patterns of co-speech and silent gesture production could shed further light on the strength of similarities between speech and co-speech gesture within a dialect and similarities in silent gesture between different dialects of a language.

Our study also focused on adults speaking each of the three languages. Earlier work focusing on children's speech across different languages showed evidence of early-emerging differences in children's speech about motion events (e.g., Gullberg et al., 2008, Hickmann et al., 2009, Özçalışkan, 2009; Özçalışkan & Slobin, 1999). Relatively less is known about patterns in co-speech gesture in children; and the existing research suggests inconclusive results: some studies show early attunement (ages 3-4; Özçalışkan, Gentner & Goldin-Meadow, 2014; Özçalışkan & Goldin-Meadow, 2011) and others show later attunement (Özyürek et al., 2008) to language-specific patterns in co-speech gesture. Research on the developmental trajectory of silent gestures is even sparser, with only one study suggesting early emerging similarities in silent gesture in both packaging and ordering of motion elements in both S- and V-languages (Özçalışkan, Lucero & Goldin-Meadow, 2021). As such, future work that examines cross-linguistic similarities and differences in co-speech and silent gestures across a broader set of languages over developmental time can provide further insight into the etiology of cross-linguistic variability (or its lack) in gesture when speaking and when not speaking.

The findings of our study have several broader implications. One important implication of this work is that learning a particular language involves mastery of not only speech—but also gesture patterns. There is, in fact earlier work, that suggests that advanced second language (L2) learners of English—with Turkish as first language (L1)—showed L1 effects on their co-speech gestures while showing L2 speech patterns when describing events in their L2 English (Özçalışkan, 2016). Recent work also shows that bilingual speakers face greater challenges when speaking an L2 with a more detailed system of expression than their L1 (Lewandowski & Özçalışkan, 2021). However, we do not yet know whether a similar challenge also becomes evident in the co-speech gestures that accompany the L2 descriptions in speech. As such, our study raises the possibility that mastery of a language—L1 or L2—should be studied in terms of not only speech but also co-speech gesture production to provide a more complete assessment of language learning.

Another important implication of this work is that speakers do *not* show the cross-linguistic differences in the way they represent events in their gesture when *not* speaking, suggesting that silent gesture might be a useful instructional device in conveying key concepts that are not expressed in a particular language. For example, L2 learners of English could benefit from instruction with silent gestures that express different types of manner to understand the nuances in manner encoding in English—particularly for L2 learners with L1s that do not encode such fine-grained manner distinctions. Future studies that examine the effect of observing gesture in L2 acquisition contexts can shed further light on the efficacy of instruction with gesture in attaining L2 proficiency both in speech and in gesture.

In conclusion, our study showed that speakers of all three types of languages (S-, V- and E-languages) display cross-linguistic differences in their speech and co-speech gestures about motion; but they do not rely on language-specific patterns in silent gesture, and instead show cross-linguistic similarities. These findings, thus, give further support to the “thinking-for-speaking hypothesis” proposed by Slobin (1996), which argues that language influences thought only during the process of online speech production (i.e., speech and co-speech gesture), but not during offline production (i.e., silent gesture).

REFERENCES

Allen, S., Özyürek, A., Kita, S., Brown, A., Furman, R., Ishizuka, T., & Fujii, M. (2007). Language-specific and universal influences in children's syntactic packaging of Manner and Path: A comparison of English, Japanese and Turkish. *Cognition*, *102*, 16–48.

Brown, A., & Chen, J. (2013). Construal of manner in speech and gesture in Mandarin, English, and Japanese. *Cognitive Linguistics*, *24*(4), 605–631.

Cardini, F. E. (2010). Evidence against Whorfian effects in motion conceptualisation. *Journal of Pragmatics*, *42*(5), 1442-1459.

Chang, J., (2001). The syntax of event structure in Chinese. Ph.D. Dissertation. University of Hawaii.

Chao, Y., (1968). *A grammar of spoken Chinese*. University of California Press.

Chen, L., & Guo, J. (2009). Motion events in Chinese novels: Evidence for an equipollently-framed language. *Journal of Pragmatics*, *41*(9), 1749-1766.

Chen, L., & Guo, J. (2010). From language structures to language use: A case from Mandarin motion expression classification. *Chinese Language and Discourse*, *1*(1), 31-65.

Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, *41*(1-3), 83-121.

Chui, K. (2012). Crosslinguistic comparison of representations of motion in language and gesture. *Gesture*, 12(1), 40–61.

Emerson, S. N., Conway, C. M., & Özçalışkan, Ş. (2020). Semantic P600—but not N400—effects index crosslinguistic variability in speakers' expectancies for expression of motion. *Neuropsychologia*, *149*(107638), 1–21.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.

Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences, USA*, *108*(42), 17290–17295.

Gibson, E., Piantadosi, S. T., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy channel account of crosslinguistic word order variation. *Psychological Science, 24*(7), 1079–1088.

Goldin-Meadow, S., & Feldman, H. (1977). The development of language-like communication without a language model. *Science*, *197*(4301), 401-403.

Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, *391*(6664), 279-281.

Goldin-Meadow, S., So, W. C., Özyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences*, *105*(27), 9163-9168.

Greenberg, J.H. (Ed.). (1963). *Universals of language*. M.I.T. Press.

Gullberg, M., Hendriks, H., & Hickmann, M. (2008). Learning to talk and gesture about motion in French. *First Language, 28*(2), 200–236.

Hsueh, Feng-Sheng, (1989). The structure meaning of BA and BEI constructions in Mandarin Chinese. In J. Tai, & F.-H. Hsueh (Eds.), *Functionalism and Chinese grammar* (pp. 95–125). Chinese Language Teachers Association.

Ibarretxe-Antuñano, I. (2004). Language typologies in our language use: The case of Basque motion events in adult oral narratives. *Cognitive Linguistics*, *15*(3), 317-349.

Jin, L. (2008, April). Markedness and second language acquisition of word order in Mandarin Chinese. In *Proceedings of the 20th North American Conference on Chinese Linguistics* (pp. 297-308). East Asian Studies Center.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16-32.

Lewandowski, W., & Özçalışkan, Ş. (2018). How event perspective influences speech and co-speech gestures about motion. *Journal of Pragmatics*, *128*, 22-29.

Lewandowski, W. & Özçalışkan, Ş., (2021). The specificity of event expression in first language influences expression of object placement events in second language. *Studies in Second Language Acquisition*, 1-32.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar* (Vol. 3). University of California Press.

McNeill, D. (2000). Analogic/analytic representations and crosslinguistic differences in thinking for speaking. *Cognitive Linguistics*, *11*(1–2), 43–60.

Mei, K. (1980). Is modern Chinese really a SOV language? *Cahiers de Linguistique-Asie Orientale*, *7*(1), 23-45.

Özçalışkan, Ş. (2004). Typological variation in encoding the manner, path, and ground components of a metaphorical motion event. *Annual Review of Cognitive Linguistics*, *2*(1), 73-102.

Özçalışkan, Ş. (2009). Learning to talk about spatial motion in language-specific ways. In J. Guo, E. Lieven, S. Ervin-Tripp, N. Budwig, K. Nakamura, & Ş. Özçalışkan (Eds.), *Cross-linguistic approaches to the psychology of language: Research in the tradition of Dan Isaac Slobin* (pp. 263–276). Psychology Press.

Özçalışkan, Ş. (2015). Ways of crossing a spatial boundary in typologically distinct languages. *Applied Psycholinguistics*, *36*(2), 485-508.

Özçalışkan, Ş. (2016). Do gestures follow speech in bilinguals' description of motion? *Bilingualism: Language and Cognition, 19*(3), 644-653.

Özçalışkan, Ş., & Emerson, S. (2016). Learning to talk, think and gesture about motion in language-specific ways: Insights from Turkish. In N. Ketrez & B. Haznedar (Eds.), *Trends in Language Acquisition Research: The acquisition of Turkish in childhood*. (pp. 177-191). John Benjamins.

Özçalışkan, Ş., & Slobin, D. I. (1999). Learning 'how to search for the frog': Expression of manner of motion in English, Spanish and Turkish. In A. Greenhill, H. Littlefield, & C. Tano (Eds.), *Proceedings of the 23rd Boston University conference on language development* (pp. 541–552). Cascadilla Press.

Özçalışkan, Ş., & Slobin, D. I. (2000). Climb up vs. ascend climbing: Lexicalization choices in expressing motion events with manner and path components. In *Proceedings of the 24th Annual Boston University Conference on Language Development* (Vol. 2, pp. 558-570). Somerville, MA, Cascadilla Press.

Özçalışkan, Ş., & Slobin, D. I. (2003). Codability effects on the expression of manner of motion in Turkish and English. In *Studies in Turkish Linguistics. Istanbul: Boğaziçi*.

Özçalışkan, Ş., Gentner, D. & Goldin-Meadow, S. (2014). Do iconic gestures pave the way for children's early verbs? *Applied Psycholinguistics, 35*, 1143–1162.

Özçalışkan, Ş., Lucero, C. & Goldin-Meadow, S. (2021). How early does speaking shape the native language of gesture? [Manuscript submitted for publication]. Department of Psychology, Georgia State University.

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016a). Is seeing gesture necessary to gesture like a native speaker? *Psychological Science, 27*(5), 737-747.

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2016b). Does language shape silent gesture? *Cognition, 148*, 10-18.

Özçalışkan, Ş., Lucero, C., & Goldin-Meadow, S. (2018). Blind speakers show language-specific patterns in co-speech gesture but *not* silent gesture. *Cognitive Science. 42*(3), 1001-1014.

Özçalışkan, Ş., Stites, L. J., & Emerson, S. (2017). Crossing the road or crossing the mind: How differently do we move across physical and metaphorical spaces in speech or in gesture. In I. Ibarretxe-Antuñano (Ed.), *Motion and space across languages and applications* (pp. 257-277). John Benjamins.

Özyürek, A., Kita, S., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2008). Development of cross-linguistic variation in speech and gesture: Motion events in English and Turkish. *Developmental Psychology*, *44*(4), 1040-1054.

Paul,. J, Emerson, S., & Özçalışkan, Ş. (2021). Does dialect matter in the expression of motion? Babao vs. Standard Mandarin [Manuscript submitted for publication]. Asian Studies Department, Agnes Scott College.

Shyu, S. I. (2001). Remarks on object movement in Mandarin SOV order. *Language and Linguistics, 2*(1), 93-124.

Simmons, R. V. (1999). Chinese dialect classification: A comparative approach to Harngjou, Old Jintarn, and Common Northern Wu. In E. F. K. Koemer (Ed.), *Current issues in linguistic theory, Vol. 188*. John Benjamins Publishing Company.

Slobin, D. I., (2004). The many ways to search for a frog: linguistic typology and the expression of motion events. In S. Strömqvist & L. Verhoeven, L. (Eds.), *Relating events in narrative: Typological and contextual perspectives* (pp. 219–257). Lawrence Erlbaum Associates.

Slobin, D. I. (1996). From ‘thought’ and ‘language’ to ‘thinking for speaking. In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge University Press.

Slobin, D. I. (2003). Language and thought online: Cognitive consequences of linguistic relativity. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 157-192). The MIT Press.

Tai, James, (2003). Cognitive relativism: Resultative construction in Chinese. *Language and Linguistics, 4*(2), 301–316.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and semantic description. Vol. III: Grammatical categories and the lexicon* (pp. 36–149). Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics*. MIT Press.

Wieselman Schulman, B. (2004). *A crosslinguistic investigation of the speech-gesture relationship in motion event descriptions*. Ph.D. Dissertation, University of Chicago.

Xu, Z. (2013). An empirical study of motion expressions in Mandarin Chinese. *English Language and Literature Studies, 3*(4), 53-67.

Yuan, J., (2001). *Outline of Chinese dialects* (2nd edition). Language and Culture Press.

APPENDICES

Appendix A

Table A.1 Mean distribution of motion elements by semantic packaging and semantic ordering for each scene in speech by language

	ENGLISH		CHINESE		TURKISH		ENGLISH		CHINESE		TURKISH	
	CONF	SEP	CONF	SEP	CONF	SEP	F-M-G	F-G-M	F-M-G	F-G-M	F-M-G	F-G-M
Set 1												
Crawl into a house	0.95	0.20	1.05	0.25	0.05	1.25	1.05	0.0	0.75	0.10	0.0	1.15
Flip over a beam	0.95	0.10	1.20	0.50	0.70	0.35	0.85	0.0	0.75	0.50	0.10	0.90
Run out of a house	1.15	0.0	1.20	0.05	0.0	1.05	1.15	0.0	0.15	0.85	0.0	1.05
Crawls towards a carpet	0.80	0.30	1.30	0.25	0.05	1.40	1.10	0.0	0.85	0.35	0.10	1.35
Crawl into a house												
Set 2												
Climb into a treehouse	0.90	0.30	0.95	0.40	0.50	0.95	1.00	0.0	0.55	0.20	0.10	1.30
Jump over a hurdle	0.85	0.25	0.85	0.35	0.90	0.10	0.90	0.0	0.75	0.25	0.05	0.95
Fly out of a trashcan	0.85	0.35	1.20	0.20	0.10	1.00	0.85	0.0	0.10	0.75	0.0	0.85
Climb toward a treehouse	0.80	0.45	1.30	0.60	0.55	1.00	1.10	0.0	1.00	0.45	0.05	1.25
Set 3												
Tumble into a trashcan	0.70	0.30	0.60	0.90	0.0	1.05	0.90	0.0	0.45	0.40	0.0	1.00
Crawl over a carpet	1.00	0.20	1.15	0.85	0.25	1.00	1.05	0.0	0.70	0.35	0.0	1.20
Crawl out of a house	1.00	0.0	1.00	0.05	0.0	1.00	0.95	0.0	0.05	0.90	0.0	1.00
Flip toward a beam	0.30	1.10	0.50	1.25	0.15	1.10	0.60	0.0	0.40	0.45	0.0	0.90
Set 4												
Run into a house	0.90	0.20	1.45	0.35	0.0	1.30	1.00	0.0	0.70	0.45	0.0	1.30
Jump over a cat	0.95	0.25	2.05	0.80	0.75	0.70	1.30	0.0	1.60	0.70	0.0	1.35
Tumble out of a treehouse	0.55	0.60	0.65	0.40	0.10	0.95	0.85	0.0	0.0	0.85	0.05	0.90
Crawl toward a house	0.85	0.55	1.30	0.50	0.20	1.25	1.25	0.0	0.65	0.55	0.0	1.40

SEP: separated. CON: conflated. F-G-M: Figure-Ground-Motion. F-M-G: Figure-Motion-Ground

Appendix B

Table B.1 Mean distribution of motion elements by semantic packaging and semantic ordering for each scene in co-speech gesture by language

	ENGLISH		CHINESE		TURKISH		ENGLISH		CHINESE		TURKISH	
	CONF	SEP	CONF	SEP	CONF	SEP	F-M-G	F-G-M	F-M-G	F-G-M	F-M-G	F-G-M
Set 1												
Crawl into a house	0.45	0.35	0.40	0.20	0.30	0.80	0.10	0.00	0.0	0.15	0.0	0.0
Flip over a beam	0.50	0.35	1.00	0.10	0.25	0.65	0.15	0.10	0.0	0.40	0.05	0.30
Run out of a house	0.55	0.30	0.55	0.30	0.10	0.90	0.00	0.05	0.0	0.10	0.0	0.0
Crawls towards a carpet	0.55	0.25	0.60	0.25	0.25	0.75	0.10	0.05	0.0	0.20	0.10	0.0
Set 2												
Climb into a treehouse	0.55	0.30	0.45	0.20	0.45	0.70	0.05	0.05	0.05	0.15	0.05	0.05
Jump over a hurdle	0.80	0.0	0.65	0.05	0.65	0.20	0.10	0.05	0.05	0.15	0.05	0.40
Fly out of a trashcan	0.40	0.15	0.80	0.15	0.45	0.60	0.05	0.00	0.0	0.55	0.0	0.40
Climb toward a treehouse	0.55	0.35	0.45	0.35	0.50	0.60	0.10	0.05	0.15	0.25	0.10	0.10
Set 3												
Tumble into a trashcan	0.65	0.25	0.90	0.25	0.25	0.60	0.15	0.10	0.10	0.10	0.15	0.0
Crawl over a carpet	0.45	0.30	0.75	0.20	0.50	0.40	0.10	0.05	0.0	0.15	0.0	0.0
Crawl out of a house	0.60	0.25	0.65	0.25	0.20	0.80	0.00	0.05	0.0	0.25	0.0	0.0
Flip toward a beam	0.50	0.50	0.90	0.10	0.30	0.65	0.20	0.00	0.0	0.25	0.05	0.05
Set 4												
Run into a house	0.35	0.45	0.55	0.35	0.25	0.80	0.00	0.10	0.10	0.15	0.0	0.05
Jump over a cat	0.85	0.05	1.20	0.00	0.55	0.45	0.15	0.10	0.05	0.10	0.05	0.20
Tumble out of a treehouse	0.70	0.05	0.60	0.05	0.45	0.45	0.00	0.10	0.0	0.20	0.05	0.25
Crawl toward a house	0.50	0.35	0.50	0.35	0.25	0.85	0.00	0.05	0.0	0.15	0.15	0.0

SEP: separated. CON: conflated. F-G-M: Figure-Ground-Motion. F-M-G: Figure-Motion-Ground

Appendix C

Table C.1 Mean distribution of motion elements by semantic packaging and semantic ordering for each scene in silent gesture by language

	ENGLISH		CHINESE		TURKISH		ENGLISH		CHINESE		TURKISH	
	CONF	SEP	CONF	SEP	CONF	SEP	F-M-G	F-G-M	F-M-G	F-G-M	F-M-G	F-G-M
Set 1												
Crawl into a house	0.80	0.15	0.95	0.0	0.60	0.10	0.15	0.35	0.0	0.85	0.05	0.15
Flip over a beam	0.90	0.05	0.80	0.15	0.70	0.20	0.0	0.35	0.0	0.80	0.05	0.45
Run out of a house	0.70	0.25	0.95	0.05	0.55	0.25	0.05	0.55	0.0	0.75	0.0	0.30
Crawls towards a carpet	0.80	0.15	0.90	0.05	0.80	0.05	0.10	0.40	0.0	0.70	0.05	0.10
Set 2												
Climb into a treehouse	0.90	0.10	0.75	0.05	0.80	0.10	0.10	0.25	0.0	0.50	0.10	0.35
Jump over a hurdle	0.95	0.0	0.80	0.15	0.85	0.0	0.0	0.50	0.0	0.90	0.0	0.55
Fly out of a trashcan	0.65	0.25	0.55	0.35	0.55	0.35	0.0	0.35	0.0	0.55	0.0	0.50
Climb toward a treehouse	0.75	0.20	0.95	0.05	0.85	0.15	0.15	0.20	0.0	0.65	0.20	0.15
Set 3												
Tumble into a trashcan	0.85	0.15	0.75	0.15	0.45	0.35	0.20	0.40	0.0	0.75	0.05	0.35
Crawl over a carpet	0.85	0.10	0.95	0.05	0.75	0.10	0.20	0.35	0.0	0.90	0.0	0.30
Crawl out of a house	0.90	0.10	0.90	0.1	0.70	0.05	0.0	0.65	0.0	0.70	0.0	0.50
Flip toward a beam	0.80	0.15	0.75	0.25	0.80	0.15	0.10	0.30	0.0	0.80	0.15	0.20
Set 4												
Run into a house	0.60	0.40	0.90	0.1	0.35	0.25	0.25	0.35	0.0	0.75	0.0	0.15
Jump over a cat	0.95	0.05	0.85	0.05	0.60	0.15	0.00	0.25	0.0	0.70	0.0	0.10
Tumble out of a treehouse	0.80	0.15	0.60	0.35	0.65	0.25	0.10	0.40	0.0	0.75	0.0	0.55
Crawl toward a house	0.85	0.15	0.95	0.0	0.80	0.10	0.10	0.40	0.0	0.65	0.0	0.10

SEP: separated. CON: conflated. F-G-M: Figure-Ground-Motion. F-M-G: Figure-Motion-Ground