Thinking for Speaking and Gesturing Motion Events: A Typological Perspective

Haizhen Wang
Soochow University, China

Abstract

The thinking-for-speaking hypothesis, formulated by Dan Slobin, has been considered a weak version of or an alternative to linguistic relativity since the 1990s. Both state the influence of language on thought. However, claims regarding thinking-for-speaking are based on verbal evidence, whereas linguistic relativity claims are based on nonverbal evidence. There seems to be no interface between the two. The author argues that thinking for speaking is not an alternative but complementary to linguistic relativity by opening a window into the mind through which the influence of language on cognition is seen with observable behavior, and that the new paradigm of thinking for speaking and gesturing enables a combination of verbal and co-verbal evidence for the effect of language on cognition, as illustrated here by cross-linguistic comparisons of motion event narratives based on Talmy's linguistic typological framework.

Keywords: thinking-for-speaking (TFS) hypothesis, linguistic relativity, motion events, language and cognition

1. Introduction

The relationship between language and cognition has attracted the interest of linguists, psychologists, anthropologists and neuroscientists, who have made assumptions from different perspectives and collected evidence with different tasks. Thinking for speaking refers to thinking that occurs online in the process of speaking. The thinking-for-speaking (henceforth TFS) hypothesis was formulated by Dan Slobin (1987, 1991, 1996a), who viewed that speakers of different languages think differently in the process of mentally preparing content for speech. During the New Whorfian revival in the 1990s, TFS was

wrongly considered an alternative to or a "new" or "weak" version of linguistic relativity (Athanasopoulos & Bylund, 2013; Han & Cadierno, 2010). Both state the influence of language on thought. However, as pointed out by Slobin (1996a), TFS makes no predictions regarding a person's worldview or thinking in general. Instead, it involves attending to and verbalizing those aspects of reality that are readily encodable in the language. The crucial difference between linguistic relativity and TFS is that the former focuses on the effects of linguistic structure on non-verbal behavior, as elicited through tasks that do not involve overt speech production (e.g., perception, classification, sorting, and matching of objects and events), while the latter focuses on the effects of linguistic structure on the cognitive processes involved in speech production, as elicited with narrative tasks (e.g., picture description or film retelling) (Bylund & Athanasopoulos, 2014, 2015). There seems to be no interface between the two.

Criticism has been on either inferring the effect of language on thought by gathering nonverbal behavior with zero involvement of language or observing language effect on merely linguistic performance (not cognition). Later, studies within TFS framework also examine co-verbal behavior, that is, behavior concurrent with speech, such as gesture. The author argues that TFS is not an alternative but complementary to linguistic relativity by opening a window into the mind through which the influence of language on cognition is seen with observable behavior, and that the new paradigm of thinking for speaking and gesturing (Gullberg, 2011; Kita et al., 2017; Pavlenko, 2014) enables a combination of verbal and co-verbal evidence for the study of linguistic relativity, as illustrated in this paper by motion event narratives by speakers of different languages.

2. Linguistic Relativity and Thinking-for-Speaking: Nonverbal or Verbal

The idea of linguistic relativity can be traced back to Johann Gottfried von Herder and Wilhelm von Humboldt in the 18th century, who viewed that languages are linked to the worldviews of their speakers. Franz Boas, in his study of color categorization, commented that the categories of language "impose themselves upon the form of our thoughts" (Boas, [1920] 1966, p. 289). Boas' student Edward Sapir developed the idea further and viewed language as "a guide to social reality", for it "powerfully conditions all our thinking about social problems and processes" (Sapir, [1929] 1949, p. 162). Benjamin Lee Whorf, on the basis of his own research on Hopi, revealed categories unfamiliar to speakers of English, and argued that "users of markedly different grammars are pointed by their grammars towards different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers but must arrive at somewhat different views of the world" (Whorf, [1940] 2012, pp. 282-283).

The basic tenet of the linguistic relativity principle is that cross-linguistic differences in the semantic partitioning of reality give rise to cross-linguistic difference in cognition (Levinson, 2012). Empirical studies on linguistic relativity in the perceptual domains

of color, number, time and space suggest that, while language does not determine our thoughts, and while thinking is very much possible without the aid of language, language nonetheless provides a ready basis of information for the purposes of habitually classifying the world into meaningful categories (Lucy, 1997; Bylund & Athanasopoulos, 2014). Methodologically, researchers within this paradigm follow Lucy (1992a, b), and typically operationalise thought along a range of nonverbal cognitive processes including categorization, sorting, recognition memory, and low-level perception. These processes are nonverbal in the sense that they do not elicit or imply overt speech production or comprehension, but constitute cognitive responses to perceptual stimuli.

As opposed to the studies within the linguistic relativity paradigm, where the focus is on nonverbal behavior, the thinking-for-speaking paradigm places emphasis on verbal behavior. According to Slobin (1987), the expression of experience in linguistic terms constitutes "thinking for speaking"—a special form of thought that is mobilized for communication. Thinking for speaking involves picking those characteristics [of objects and events] that (a) fit some conceptualization of the event, and (b) are readily encodable in the language (Slobin, 1987, p. 435). The selection and organization of information for speech, that is, the conceptualisation stage in Levelt's (1989) terms, is thus central to the TFS paradigm. There is a growing body of evidence showing that speakers of different languages select and structure information differently in speech depending on the lexical and grammatical categories available to them (Berman & Slobin, 1994; Bylund & Athanasopoulos, 2014).

Slobin (1996a) is cautious in pointing out that TFS is not to be confused with linguistic relativity, as it makes no predictions regarding a person's worldview or thought in general. Thus, presenting the TFS hypothesis as a "new" or "weak" version of linguistic relativity (Han & Cadierno, 2010) is at odds with Slobin's original idea. Likewise, claims regarding linguistic relativity based on verbal evidence alone is at odds with the basic methodological premise of linguistic relativity research.

The distinction between verbal and nonverbal evidence plays a crucial role in making the TFS hypothesis a complement to rather than a substitute for linguistic relativity. Verbal evidence is behavioral data concerning overt speech production or comprehension, with narrative tasks (e.g., picture descriptions or film retellings) constituting the most commonly used elicitation technique. Nonverbal evidence, in contrast, is elicited through tasks that do not involve overt production or comprehension of speech, such as perception, classification, sorting, and matching of objects and events. While the author does not deny that the use of nonverbal tasks is necessary in order to "avoid the circularity that arises when inferences about nonverbal behavior are made on the basis of verbal evidence alone" (Bylund & Athanasopoulos, 2014, p. 952), she also argues that nonverbal behavioral data with the nonuse or zero residue of language is neither legitimate nor sufficient as evidence for the effects of language on thought. The "non-linguistic" status of the non-verbal tasks (e.g., eye-tracking task, similarity judgment task, recognition task) on performance only leads to questionable results that cross-linguistic differences

on non-verbal tasks are interpreted as evidence of either prior or subsequent linguistic encoding. The rationale for focusing on both verbal and co-verbal behaviour (i.e. behavior concurrent with speech) is that human beings are, most of the time, engaged in preparing, producing, or interpreting verbal messages, and therefore research into language and thought is incomplete without attention to the mental processes that are concurrent with speech production (Bylund & Athanasopoulos, 2015).

3. From Thinking-for-Speaking to Thinking-for-Speaking-and-Gesturing

3.1 Thinking for speaking

Vygotsky (1986) points out that the only way to study an internal process is to externalize it experimentally. In 1987, Dan Slobin turned a familiar pairing of terms—thought and language—into a more dynamic expression, replacing the abstract nouns with gerunds, i.e. thinking for speaking. He purposely turns our attention from static entities of "linguistic structures", "mental representation" or "objective reality" to the dynamic activity of thinking employed in the activity of speaking, for "a particular utterance is never a direct reflection of 'objective' or perceived reality or of an inevitable and universal mental representation of a situation", and "each language provides a limited set of options for the grammatical encoding of characteristics of objects and events" (p. 435). The language or languages that we learn in childhood are "not neutral coding systems of an objective reality". Rather, each one is "a subjective orientation to the world of human experience" (Slobin, 1991, p. 23), and this orientation affects the ways in which we think while we are speaking. This cross-linguistically variable cognitive process of thinking-for-speaking enables us to study those systems that "cannot be experienced directly in our perceptual, sensorimotor and practical dealings with the world" (ibid.). Different languages inherently present different thinking for speaking patterns, targeting different "contents of experience that are capable of expression in linguistic terms" (ibid.).

The hypothesis that languages not only provide speakers with a framework for the expression of experiences, events, and thoughts but also restrict how experiences, events, and thoughts are expressed at the time of speaking, was testified by Slobin's (Berman & Slobin, 1994; Slobin, 1987, 1991) early cross-linguistic studies of L1 narrative development in several languages (English, German, Hebrew, Icelandic, Japanese, Mandarin, Russian, Spanish, and Turkish) using the story *Frog, Where Are You?* (Mayer, 1969). The participants included children at various ages and adults. The findings showed that speakers of different languages have different patterns of thinking for speaking in how semantic domains such as aspect and motion are indicated lexically and syntactically.

3.2 Speech and gesture

The research on gesture dates from Roman times with emphasis on rhetorical gestures—

the mannered performances of orators, with the hands and body comprising more or less deliberate gestured embellishments on spoken public performances. Then, it went unnoticed under the influence of structuralism. Commencing with Kendon in 1972 and continuing with ever increasing vigor to the present day, gestures are now regarded as parts of language itself—not as embellishments or elaborations, but as integral parts of the processes of language and its use (McNeill, 2000).

Kendon (2000, p. 49) defines gestures as a "range of visible bodily actions that are, more or less, generally regarded as part of a person's willing expression". There is little disagreement that gestures play an important role in communication, as speech and gesture jointly express the speaker's message in a coordinated way. That is to say, gestures that spontaneously accompany speech convey information coordinated with the concurrent speech. Thus, gesture production is partly motivated by the speaker's desire to enhance communication. However, a growing body of evidence shows that gesture production also affects the gesturer's own cognitive processes and representations (Kita et al., 2017). Gestures have not only communicative functions but also self-oriented cognitive functions. Examining both participants' speech and gesture gives researchers an enhanced window into the mind through which mental representations and processes can be observed. These gestures are external manifestations of a speaker's online thinking for speaking (McNeill & Duncan, 2000). Stam (2006) further proposes that looking not only at learners' speech but also at their accompanying gestures gives us a clearer and more complete picture of their progress in learning another language than looking at speech alone.

3.3 Thinking for speaking and gesturing

People spontaneously produce gestures during speaking and thinking. "To make a gesture...is to bring thought into existence on a concrete plane, just as writing out a word can have a similar effect" (McNeill, 2005, p. 99). It is through the partnership between gesture and speech that utterance meaning is achieved, which can be illustrated with a simple example (McNeill, 2000, p. 142). A speaker raises her hand upward to mean that a character in a story is climbing up. The hand and its movement present thought in action. The rising hand expresses upwardness, and so does the speech, "[and he climbs **up the pipe**]". "Just as binocular vision brings out a new dimension of seeing, gesture reveals a new dimension of the mind"—the imagery of language which has laid hidden (McNeill, 1992, p. 1). The co-expressiveness of speech and gesture suggests that the speaker was thinking in terms of a combination of imagery and linguistic categorical content: thinking happened in the form of a holistic image of the character rising upward coordinated with the analytic, linguistically categorized meanings of "up" and "the pipe". Speech and gesture arise from an interaction of mental operations of opposite character—imagistic and linguistic. Speech (linear and segmental) and gestures (global and synthetic) co-occur and are co-expressive in the act of speaking.

Gestures exhibit images that cannot always be expressed in speech, as well as the

images that the speaker thinks are concealed. With these kinds of gestures speakers unwittingly display their inner thoughts and ways of understanding events of the world. These gestures are "the person's memories and thoughts rendered visible" (McNeill, 1992, p. 12). Studies of motion expression in speech and gesture have demonstrated that cross-linguistic differences in motion encoding guide speakers of different languages to focus on different aspects of motion for verbalization and gesturing and result in systematic differences in habitual lexicalization patterns, the ease of lexical access, and interpretation of novel verbs (Pavlenko, 2014). These findings extend the notion of "thinking for speaking" to "thinking for speaking and gesturing". The following section provides more evidence from a cross-linguistic comparison of motion events that supports the interplay between speech and gesture production.

4. Evidence from Motion Events

4.1 Motion events and language typology

The cognitive domain of motion has been selected as the central area for cross-linguistic comparisons, for motion is fundamental and universally central to human experience across cultures, and different languages exhibit striking differences in how they select and structure information about motion. To testify the thinking-for-speaking hypothesis, cross-linguistic research has been conducted in the domain of motion events, i.e. movements of entities through space, in a number of different languages (Danish, Dutch, English, German, Hebrew, Icelandic, Korean, Japanese, Mandarin, Russian, Spanish and Turkish). Motion events include the following components (Talmy, 1985, 1991, 2001): (a) *Figure*, the moving object, (b) *Ground*, the reference-point object with respect to which the Figure moves, (c) *Path*, the direction or trajectory of the motion, and (d) *Motion*, the movement. A typical example of an expression of a motion event would be: "The girl ran out of the house". In this sentence, *the girl* expresses the Figure, *the house* is the Ground, and the Path is expressed by the particle *out of*. The verb root (*run*) itself conflates Manner and Motion.

Studies of L1 acquisition and of gestures of deaf children suggest that all children begin at a default starting point paying equal amount of attention to manner and path, and then the modulation by linguistic input results in language-specific lexicalization patterns (Pavlenko, 2014). On the basis of where a language encodes path, Talmy (1985, 1991) classifies languages into two types: verb-framed languages (V-languages) that encode path in the main verb, leaving the marking of manner optional, and satellite-framed languages (S-languages) that encode manner of motion in the main verb and path in the satellites, such as prefixes or particles. English, along with all branches of the Indo-European language family except the Romance languages, is satellite framed. S-languages include (Germanic) Dutch, English, German, Icelandic, Swedish; (Slavic) Polish, Russian; (Finno-Ugric) Finnish; and (Sino-Tibetan) Chinese. Verb-framed languages include not only Spanish, but all Romance languages, as well as families such as the Semitic

and Polynesian languages. They are (Romance) French, Italian, Portuguese, Spanish; (Semitic) Arabic, Hebrew; (Turkic) Turkish; (Sign) American Sign Language, and Greek, Japanese, and Korean.

The contrast in the satellite-framed/verb-framed dichotomy is illustrated in the following examples in English and Spanish respectively (Sharpen, 2016).

(1) a. *She ran out of the house.*
b. *Salió de la casa corriendo (*alternatively, *Salió corriendo de la casa).*
'She left the house [by] running'

In English, the verb "to run" encodes motion with manner, while the path of the motion event is expressed with satellite elements, i.e. the adverb "out" plus the preposition "of". In accord with the verb-framed encoding pattern, motion and path are encoded in Spanish in the verb *salir*, 'to leave', while manner is encoded in the verb *correr*, 'to run', in the gerund form. Note that the gerundive constituent, *corriendo* in (1b), can sit either at the end of the phrase or directly after the verb of motion.

Linguistic differences interact with universal perceptual processes, increasing our sensitivity to particular aspects of events and decreasing the salience of others. That is to say, experiences are "filtered" through language for speaking purpose. Slobin's (1991, 1996a) manner salience hypothesis proposes that language-specific motion encoding serves a "filtering" function: S-languages, which encode manner in the main verb and rely on finite, high-frequency verbs, draw attention to manner, while V-languages, which encode path in the main verb and manner in non-finite verbs and low-frequency lexical items, phrases or clauses, reduce the salience of manner and draw attention to path. The hypothesis is supported by studies of lexicalization patterns through narratives elicited by picture books of motion events, showing that children and adults speaking S-languages use more frequently a wider range of manner verbs than speakers of V-languages (Slobin, 1996a, b).

The typological difference between languages are important. When acquiring a native language, the child learns particular ways of thinking for speaking and gesturing. Each native language has trained its speakers to pay different kinds of attention to events and experiences when talking about them. This training is carried out in childhood and is exceptionally resistant to restructuring in adult second-language acquisition. If different patterns of thinking for speaking and gesturing exist in the L1 and the L2, the learners must learn another pattern of thinking for speaking and gesturing in order to be proficient speakers in their L2 (Stam, 2010).

4.2 Inter-typological variation in thinking for speaking and gesturing motion events

To understand better the typological differences between languages, two S-languages (English, Chinese) and two V-languages (Spanish, Japanese) are selected here to make inter-typological comparisons of thinking for speaking and gesturing motion events in

both first (L1) and second (L2) languages. Representative studies are used respectively to make the illustration (Table 1).

Table 1. Cross-linguistic inter-typological research on motion events

		S-languages	
		English	Chinese
V-languages	Spanish	McNeill & Duncan (2000); Stam (2006, 2010, 2014, 2015)	Duncan (2006)
	Japanese	Kita & Özyürek (2003); Brown & Gullberg (2008, 2011)	Brown (2015)

4.2.1 Thinking for speaking and gesturing motion events in L1

Spanish speakers and English speakers are found to have different patterns of thinking for speaking and gesturing about path of motion (McNeill & Duncan, 2000; Stam, 2006, 2010, 2014, 2015). Spanish speakers express path linguistically on verbs, their path gestures tend to occur with path verbs, and their manner gestures may occur without manner in speech, whereas English speakers express path linguistically on satellites (adverbs or particles), their path gestures tend to occur with satellite units, and their manner gestures rarely occur without manner in speech. Examples (2) and (3) (adapted from Stam, 2006) illustrate the patterns. These findings support McNeill's idea (McNeill, 1992; McNeill & Duncan, 2000) that spontaneous gestures are synchronous with speech, and that co-produced speech and gesture express the same underlying idea unit but do not necessarily express identical aspects of it. Sometimes speech and gesture represent the same entities, and sometimes they complement each other, where the gestures indicate an aspect of the speaker's thought that is present but not expressed through speech (Stam, 2010).

(2) *Native Spanish - Sylvester and bowling ball*

a. *al momento de que sale*-PATH

at-the moment of that he-exits

'at the moment that he exits'

b. ***sale*-PATH** *el gato pero con la bola adentro*

exits the cat but with the ball inside

'the cat exits but with the ball inside'

gesture: both hands move down along body <ball going into and down Sylvester >

(3) *Native English - Sylvester and bowling ball*

a. [*Sylvester* ***falls*-MANNER *back*-PATH**][*down*-PATH

gesture: right hand in front of head moves straight down <Sylvester and bowling ball fall down pipe>PATH

b. *the drain**pipe rolls***-MANNER ***down*-PATH *th***][***e street into*-PATH *the bow***]*ling alley* 1a 1b

1a gesture: right hand moves from left to right across body changing direction (angling up on diagonal) with *the street* <Sylvester and bowling ball move across/down street>PATH

1b gesture: (continuation of previous PATH gesture) right hand moves up on a diagonal and holds <Sylvester and bowling ball going into bowling alley >PATH

In Example (2), when narrating Sylvester [the cat] and the bowling ball going down the drainpipe, the Spanish speaker produced two clauses with the verb *sale* 'exit' indicating path. One path gesture that occurred in (2b) accompanied the verb *sale* 'exit'. However, the co-produced speech and gesture do not express the same underlying meaning. The gesture here serves as a "meta-level comment on the discourse" (Stam, 2006, p. 159). The speech with the verb *sale* 'exit' indicates that the cat is exiting [the drainpipe], but the gesture indicates path (the ball going into and down Sylvester) and tells the interlocutor that the ball had gone into Sylvester, an aspect of the speaker's thought that is not expressed through speech. In such case, speech and gesture complement each other.

In Example (3), the native English speaker narrated the motion event in two clauses, using two satellites and one preposition of motion to indicate path that were added one after the other: *down*, *down*, and *into*, and two manner verbs *fall* and *roll*, which indicate both movement and manner. During the speech production, the speaker had three path gestures, either co-occurring with a satellite or a verb + satellite or with verbs and ground noun phrases. The first, (3a), occurred on a verb + satellite *falls back*; the second, (3b)1a, on a ground noun phrase, the following verb and satellite *pipe roll down*; and the third, (3b)1b, on a ground noun phrase, the next satellite and the ground noun phrase following that satellite *street into the* with a hold on part of *bowling*. In other words, English speakers tend to have their path gestures co-occurring with satellite units.

Duncan (2006) examined over 100 videotaped, unrehearsed, cartoon story narrations collected from adult Spanish, English, and Mandarin Chinese speakers for their speech and speech-synchronous gestures related to manner of motion. First, the three languages are found to form a continuum with respect to the tendency to use manner-expressive verbs: Spanish speakers' preference of path verbs at one end, and Chinese speakers' heavy use of manner verbs at the opposite end, with English in between. This finding further illustrates the cline of manner salience. Second, speakers of all three languages are found to produce manner-expressive gestures similar in type and frequency. This finding contrasts McNeill (2001), who proposed a typological variation in gesture: S-language manner modulation gestures versus V-language manner fog gestures. In S-languages like English and Mandarin, where manner is almost obligatorily encoded in speech due to its lexicalization in the main verb, speakers can modulate its significance by not encoding it in an accompanying gesture, gesturing about other components instead of the event such as path. In V-languages like Spanish, where explicit mention of manner in speech may be awkward, gestures may be used as an alternative carrier of

manner information. Thus, Duncan suggested that motion event description may in fact involve conceptualization of manner to roughly the same extent in all three languages. Moreover, gestures' role is not only limited to a compensation for content speech lacks, and as Duncan proposed, gesture and speech "jointly highlight" shared or congruent semantic content.

Kita and Özyürek (2003) elicited narratives in Turkish, Japanese, and English using an animated cartoon as the stimulus and the informational coordination between speech and gesture was compared across languages. It was found that gestural expression of the same motion event (Swing Event) shows both systematic cross-linguistic variation and similarity. The gestural content varies cross-linguistically in a manner that parallels to how the three languages package information about the Swing Event in speech. In S-language English, where there is a readily accessible linguistic means to package the change of location and the arc-shaped trajectory, speakers' gestures represent change of location with an arc-shaped trajectory. By contrast, in V-language Turkish and Japanese, where readily accessible linguistic means cannot encode the arc trajectory, the majority of the speakers produce a change of location gesture without the arc-shaped trajectory. These findings demonstrate a language effect on both speech and gestural representations and a synchrony of speech and gesture. However, regardless of the trajectory shapes and the language types, the Swing-Event gestures regularly encode the directional information in the Swing Event that is never verbalized, i.e. gestures encode some aspects of the stimulus scene that are not expressed in the accompanying speech. Therefore, it can be concluded that a gesture is simultaneously shaped both by readily accessible, concise linguistic packaging of relevant information and by the spatio-motoric properties of the referent that are never verbalized.

4.2.2 Thinking for speaking and gesturing motion events in L2

In a longitudinal study of a Spanish-English learner's thinking for speaking patterns about motion over a fourteen-year period, Stam (2010, 2014, 2015) found that with increased exposure to the target language, the learner's gestural expression of path changed in both her L1 (Spanish) and L2 (English), and her gestural expression of manner changed in her L2, and that the learner's thinking for speaking about path in English became more native-like, but her thinking for speaking about manner did not. This change suggests that manner, a pattern acquired in childhood, may not be resistant to restructuring in L2 acquisition after all. L2 thinking for speaking is not static. It can change over time, but not all aspects of thinking for speaking change equally. The learner's increased L2 proficiency and use of L2 on a daily basis may affect her pattern of thinking for speaking and gesturing.

Brown and Gullberg's (2008, 2011) analysis of speech and gestures of Japanese-English speakers reveals that in the encoding of manner the participants pattern with L1 Japanese speakers in both L1 and L2, displaying L1 influence on L2 English. Yet they also pattern with L1 English speakers by encoding manner in speech but not in the

accompanying gestures, a pattern interpreted as L2 influence on L1 performance. The bidirectional influence is also found in Brown (2015), who investigates further bilingual versus monolingual construal of manner of motion in speech and gesture across three languages—Mandarin, Japanese, and English. Narratives descriptions of motion are elicited in the L1 and L2 from bilingual Mandarin-English and Japanese-English speakers and from monolingual speakers of the three languages. Representative examples of speech and gesture coding for descriptions of the CLAMBER-UP event are illustrated in (4)-(10) (adapted from Brown, 2015, pp. 71-73).

(4) Monolingual English

[*you see*] [***Sylvester***[1st] **(pause)**[2nd 3rd] ***slithering*** ***up the pipe*** [4th]

1st-3rd path gestures: hand moving in upward trajectory with no concurrent finger movement or wrist rotation

4th manner-path gesture: hand moving in upward trajectory with concurrent finger movement representing climbing

In Example (4), a monolingual English speaker expressed manner in speech in the verb *slither*, and produced the first three gestures indicating path only, which illustrate a lack of manner-highlighting. The fourth gesture, in contrast, encoded both manner and path, indicating a manner-highlighting gesture.

(5) Monolingual Japanese

[***Ue-no hou made itte***[1st]]

Up-GEN side to go.CON

'(He) goes to the top.'

1st manner gesture: rotation of wrist and movement of fingers, hand depicting a climbing action in place

In Example (5), a monolingual Japanese speaker did not mention manner in her entire description of the motion event, indicating an absence of manner in speech. However, the accompanied gesture depicted manner and thus illustrates a manner-additive gesture.

(6) Monolingual Mandarin

[*wang shang* ***pa***[1st]]

Toward top climb

'(He) climbs up.'

1st path gesture: hand moving in upward trajectory with no concurrent finger movement or wrist rotation

In Example (6), a monolingual Mandarin speaker expressed manner in speech in the verb *pa* 'climb', while the accompanying gesture expressed only path, hence illustrating a

lack of manner-highlighting.

(7) Bilingual Mandarin-English in L1

[*Xia shui guan xia shui guan*[1st], yan zhe nei ge[2nd] *xia shui guan wang shang pa*]

drainpipe drainpipe along that drainpipe toward up climb

'The drainpipe, it is climbing up along that drainpipe.'

1st manner-path gesture: rotation of wrist and movement of fingers, hand depicting a climbing action, conflated with upward trajectory

2nd path gesture: hand moving in upward trajectory with no concurrent finger movement or wrist rotation

(8) Bilingual Mandarin-English in L2

[*The cat*[1st] *climbed*[2nd] *up*[3rd] *again*[4th]]

1st-4th path gestures: hand moving in upward trajectory with no concurrent finger movement or wrist rotation

In Examples (7) and (8), the same bilingual Mandarin-English speaker expressed manner in speech in the verb *pa* 'climb' in Mandarin (L1) and *climb* in English (L2). In her L1, the speaker produced one gesture conflating manner and path, which was coded as a manner-highlighting gesture. However, all other gestures produced in her L1 and L2 depicted path only and were thus coded as an absence of manner highlighting.

(9) Bilingual Japanese-English in L1

[*Sono suki ni shirubesutaa-wa haisuikan paipu kara nobotte kite*[1st]]

that chance at Sylvester-TOP drainpipe pipe from climb.ascend.CON come.CON

'With that chance, Sylvester is climbing up from the drainpipe.'

1st path gesture: hand moving in upward trajectory with no concurrent finger movement or wrist rotation

(10) Bilingual Japanese-English in L2

[He **climb up**[1st] the climbed **up to**[2nd] **the**[3rd] Tweety's window]

1st-3rd path gestures: hand moving in upward trajectory with no concurrent finger movement or wrist rotation

In Examples (9) and (10), the bilingual Japanese-English speaker expressed in her L1 neither manner in speech nor manner in gesture, indicating an absence of manner in speech and an absence of manner-additive gesture. In her L2, the speaker started the utterance with a false start accompanied by a path gesture, and then she encoded explicitly manner in speech in the verb *climb*, but did not gesture about manner, illustrating a lack of manner highlighting through gesture.

Results of the above within- and between-participant comparisons revealed that both

groups of bilinguals in L1 speech patterned with their monolingual counterparts and in line with typological predictions of manner encoding, i.e. Mandarin-English speakers encode manner more frequently than Japanese-English speakers, and that encoding of manner in L2 speech is characterized by universal features of development, while construal of manner in gesture is characterized by bidirectional interactions between properties of the source and target languages involved, yielding a convergence between the L1 and L2. In addition, the rare occurrence of manner-additive gestures in both sets of L2 data despite the frequency of L2 gesturing overall supports the claim that gestures are not necessarily used as a compensatory device in L2 interaction.

4.3 Intra-typological variation in thinking for speaking and gesturing motion events

We must not assume that languages within the same typological group would be completely identical in terms of motion event lexicalization. The existence of intra-typological variation challenges Talmy's dichotomy between verb-framed and satellite-framed languages. Hijazo-Gascón and Ibarretxe-Antuñano (2013) elicited data from speakers of three verb-framed languages (French, Italian, and Spanish) and found significant differences in the way speakers provide motion descriptions. Italian speakers, for example, describe Path in more detail than speakers of French or Spanish, prefer Plus-ground verb constructions and mention more than one Path element per verb, and French speakers tend to omit Manner information more systematically.

Here, three satellite-framed languages are selected to illustrate intra-typological differences in thinking for speaking and gesturing motion events: English, Chinese and Russian. For S-languages, path is not encoded in the verb, but in a satellite—an adjunct to the verb, and manner is encoded in the main verb.

Both Chinese and English have a large lexicon of manner verbs. Duncan (2006) reported that in speech production of the Climbing-Event, English speakers often use the manner-expressive verb *climb* in one or more phrases to describe the cat's (Sylvester) ascent, and the phrase often combines in sequence with one or more additional phrases incorporating a non-manner-expressive verb *get*, as in (11), or *go* and *come*. In contrast, Chinese speakers often repetitively use manner verbs within utterances in a way not attested in the English data, as shown in (12). That is to say, Chinese speakers use manner verbs more heavily than English speakers do.

(11) *Sylvester climbs up the drainpipe gets to the top*

(12) *mao kai-shi cong shui-guan pa pa-pa-pa pa-shang-qu*
cat begin via drainpipe climb climb-climb-climb climb-up-go
'the cat starts to climb the drainpipe he climbs and climbs he climbs up'

As regards motion event gestures, the English and Chinese descriptions provide further illustrations of the gesture-speech semantic synchrony. However, the Chinese-

specific pattern is a gesture that occurs earlier in the temporal sequence of speech than the timing of English, as shown in (13) (McNeill & Duncan, 2000, p. 153).

(13) *lao tai-tai* [*na -ge da bang hao*]-*xiang gei ta da-xia*
old lady hold CLASSIFIER big stick seem CAUSE him hit-down
'The old lady apparently knocked him down with a big stick'

As the speaker said *da bang* ("big stick"), she performed a downward-blow gesture. Her hand then promptly relaxed and went to the rest position well before the verb phrase emerged in speech. It is as if the gesture shifts forward in the surface stream, in the direction of the utterance-initial position in Chinese speech. McNeill and Duncan (2000) interpreted this phenomenon as evidence that links the typological distinction of topic- vs. subject-prominence with different patterns of thinking-for-speaking, which indicates that Chinese speakers are able to create a topicalizing frame with conjoined gestures and speech, whereas the packaging of motion information in English is encased within the grammatical structure of the sentential predicate to a degree not found in Chinese.

Pavlenko and Volynsky (2015) examined lexicalization of motion in narratives elicited with the use of a picture book *Frog, Where Are You?* (Mayer, 1969) from L1 speakers of Russian, L1 speakers of English, and Russian-English bilinguals who were subdivided into early, childhood, and late bilinguals. Quantitative and qualitative analyses of motion verb corpora revealed that L1 Russian speakers segment motion events in a more fine-grained way and encode the manner, directionality, and spatiotemporal contours of motion events significantly more frequently than speakers of L1 English. Bilinguals follow language-specific lexicalization patterns in both languages but late bilinguals display reduced lexical diversity in L2 English. These findings are linked to differences in obligatoriness, boundedness, and complexity of encoding of motion components in the two languages. For example, in some motion events, L1 Russian speakers used manner verbs one hundred percent of the time, while L1 English speakers also appealed to non-manner verbs. The fact that English does not require obligatory manner encoding raises the question about the use of English as a canonical contrastive language. Hence, Pavlenko (2014, 2015) suggests a replacement of obligatory encoding with frequent encoding as the source of language effects.

5. Discussion and Conclusion

The studies of motion events reviewed here present compelling evidence that cross-linguistic variation in motion encoding affects all aspects of the speech planning and verbalization process, suggesting that linguistic differences are not superficial—while they do not determine cognition, they interact with universal perceptual processes, increasing our sensitivity to particular aspects of events and decreasing the salience of others. The synchrony of speech and gestures during verbalization demonstrates an

interaction of mental operations in both linguistic and imagistic forms. Co-produced speech and gestures are two simultaneous aspects of a single idea unit (McNeill, 1992). Therefore, it is reasonable to extend Slobin's notion of thinking for speaking to thinking for speaking and gesturing. The cross-linguistic comparisons of thinking for speaking and gesturing motion events demonstrate the existence of a continuum of manner/path salience, the cognitive functions of gestures, and the complexity of reconceptualization in L2 production.

5.1 A continuum of manner/path salience

The expression of motion events is restricted linguistically and gesturally by the structure of the language. For example, the expression of path linguistically and gesturally in Spanish and English is restricted by the structure of the two languages. Path is encoded on the verb in Spanish and the satellite in English. The number of English speakers' gestures that co-occur with verbs is found significantly fewer than the Spanish speakers' and is reflective of the difference between verb-framed and satellite-framed languages (Stam, 2006).

It is also reported that speakers of V-languages encode information about manner of motion around 20% of the time when describing motion, whereas speakers of S-languages do so 70% of the time; however, when judging motion similarity, speakers of S-languages rely on manner of motion 50% of the time, whereas V-language speakers do so 30% of the time (Bylund & Athanasopoulos, 2014). These results suggest that V- and S-languages are not in a bipolar position but fall into a continuum of manner salience. English is more salient in encoding manner of motion in the main verb, and hence categorized as a satellite-framed language. Yet compared with Russian, another S-language, English does not encode manner in an obligatory fashion and English-speakers commonly rely on high-frequency non-manner verbs *come*, *go* or *get*. In contrast, in Russian, manner encoding is near obligatory, and the Russian lexicon has only a few non-manner verbs (Pavlenko, 2014). The high intra-typological variation shows that the manner salience continuum happens not only across typologically different languages but also between languages of the same typology. There also exists a continuum of path salience. For example, Italian is a high-path salient V-language whereas Spanish and French are low-path salient V-languages (Hijazo-Gascón & Ibarretxe-Antuñano, 2013).

5.2 Cognitive functions of gestures

Gestures have not only communicative functions, but also cognitive functions. Several hypotheses have been proposed regarding types and functions of gestures. McNeill (2001) proposed that gestural depiction of motion exhibits cross-linguistic differences following typological patterning in speech, and thus a distinction was made between manner modulation gestures in S-languages and manner fog gestures in V-languages. For example, studies by Kita and Özyürek (2003) and Brown and Gullberg (2008) revealed

systematic cross-linguistic differences in gestures accompanying motion speech. Speakers of the S-language English focused more on manner and tended to conflate manner and path in a single verbal clause, and a single gesture, showing that speakers of S-languages tend to modulate the significance of manner by gesturing about path. In contrast, speakers of V-languages French, Japanese, and Turkish focused more on path and when they did encode manner it often appeared in separate clauses, accompanied by manner-only and path-only gestures. On the other hand, speakers of the V-language Spanish produced manner gestures in the absence of spoken manner encoding. For V-language speakers, gestures may be used as an alternative carrier of manner information.

Studies that examined both speech and gestures showed that the two modalities form a tightly integrated system where information may sometimes be encoded in one modality and not the other (Pavlenko, 2014). Brown and Chen (2013) compared manner-highlighting gestures, gestures that highlight information about manner already present in speech, and manner-additive gestures, gestures that add manner information otherwise absent from speech. Results revealed that monolingual speakers of satellite-framed English tend to highlight manner in speech with accompanying manner gestures less often than monolingual speakers of verb-framed Japanese, and that manner-additive gestures are much less frequent overall, with no significant differences across languages. The above findings suggest different functions of gestures, i.e. gestures and speech may jointly highlight shared information, or gestures may compensate for what speech lacks.

Kita and colleagues (2017) recently reviewed four existing hypotheses regarding the cognitive functions of gestures and proposed a new one—the gesture-for-conceptualization hypothesis. The four established hypotheses focus on how gestures facilitate speaking. First, the lexical retrieval hypothesis holds that speakers' gestures help increase activation on items in their mental lexicons, therefore facilitating lexical access. Second, the image activation hypothesis holds that gestures maintain visuospatial imagery, providing better quality information for the speech production. Third, the information packaging hypothesis holds that gestures help speakers package or chunk spatio-motoric information into units appropriate for verbal encoding. Fourth, the cognitive load reduction hypothesis holds that gestures reduce the amount of cognitive resources needed for formulating speech, thus lightening the cognitive load of speaking.

Further, Kita and colleagues (2017) proposed that gesture shapes the way people conceptualize information through four functions: gesture activates, manipulates, packages, and explores spatio-motoric information for the purposes of speaking and thinking. These four functions are shaped by gesture's ability to schematize information, that is, to focus on a small subset of available information that is potentially relevant to the task at hand. By schematizing spatio-motoric information for these four functions, gesture facilitates cognitive processing, generates novel ideas, strategies and solutions, and makes representations more flexible and open to change. That is, schematization

via gesture focuses on spatio-motoric information, "stripping away other types of information (activation), makes it possible to efficiently modify representations (manipulation), focuses on small chunks of spatio-motoric information appropriate for speaking and thinking (packaging), and creates a unique landscape in which information can be explored (exploration)" (p. 258). Furthermore, according to the gesture-for-conceptualization hypothesis, speakers' schematization of information in gesture influences not only speakers' but also listeners' thinking. Gesture plays a central role in human cognition.

5.3 Complexity in L2 learners' reconceptualization

The interface between language and cognition remains complex. Speakers of more than one language raise further challenges, challenges to understand how experience is processed through the filter provided by one language and what happens to that filter when an alternative filter provides different options. The accumulated native speech-gesture evidence indicates considerable cross-linguistic differences in how motion is linguistically conceptualized. These differences have consequences for L2 speakers who must reconceptualize motion, that is, shift attention to different types of information, potentially expressed in different linguistic categories and structures, to be target-like (Gullberg, 2011). The studies reviewed here suggest that multiple perspectives are possible and that L2 reconceptualization is gradual and dynamic. Some studies reveal persistent conceptual transfer from the L1 with L2 speakers continuing to select L1-typical information and express it in L1-typical categories (e.g. Stam, 2006). Some studies show evidence of shifts toward L2-typical conceptualization, that is, of reconceptualization to match the target language (e.g., Stam, 2010, 2014, 2015). There is also evidence of bidirectional influence with Japanese-English speakers whose encoding of manner in gesture is characterized by bidirectional interactions between properties of the source and target languages (Brown & Gullberg, 2008, 2011; Brown, 2015). Moreover, linguistic conceptualization or reconceptualization is not impervious to change, as evidenced by Stam's longitudinal studies (2010, 2014, 2015).

However difficult it is for L2 speakers to reconceptualize motion regardless of the direction of typological transition, a general observation is that all studies suggest that restructuring is possible, even if few studies indicate a fully target- or monolingual-like conceptualization. Thinking for speaking and gesturing data reveal that reconceptualization is gradual and not necessarily an all-or-nothing process even within a single domain. L2 learners or bilinguals rarely behave as "two monolinguals in one". Instead, L2 learners experience a complex cognitive restructuring process, involving stages of L1 influence on the L2, destabilization, internalization of new categories, co-existence or maintenance of respective patterns, convergence, L2 influence on L1, and attrition (Pavlenko, 2014). In the process, several factors have been identified which affect cognitive restructuring or reconceptualization, including age of arrival in the L2 environment, the context of L2 acquisition, the length of exposure to target language,

frequency of language use, language dominance and language proficiency. The interactive effects of these factors are complex. For example, Bylund and Athanasopoulos (2015) reported that those Swedish speakers of FL English who often watched television in English approximated the ongoing preference of the English native speakers, which suggest that event cognition patterns may be restructured through exposure to FL audiovisual media. Brown and Gullberg (2008, 2011) found that Japanese-English bilinguals with intermediate proficiency levels displayed evidence of L2 influence on L1 in speech and gesture, regardless of the context of acquisition and the context of language use. It is time we started seeing (re)conceptualizations as dynamic and changing, and future studies need to adopt longitudinal designs to trace the complexity of L2 learners' linguistic and cognitive systems, taking into consideration interactive factors of reconceptualization.

All in all, by looking at the speech and the gesture jointly, we are able to infer what may not be obvious from the speech alone. Analyses of speech and gesture, combined in verbal and co-verbal productions, constitute new lens through which the effect of language on thought can be observed. Gestures provide rich imagistic and multidimensional data, activating, manipulating, packaging, and exploring spatio-motoric information for the purposes of speaking and thinking. As evidenced by studies of motion events narrated by speakers of different languages, speech and gesture illuminate different aspects of the same underlying conceptualization, and shed new light on the complex dynamic process of L2 reconceptualizations.

References

Athanasopoulos, P., & Bylund, E. (2013). The "thinking" in thinking-for-speaking: Where is it? *Language, Interaction, and Acquisition*, *4*, 91-100.

Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Boas, F. (1920, reprinted in 1966). The methods of ethnology. In F. Boas (Ed.), *Race, language, and culture* (pp. 281-289). New York: The Free Press/Macmillan.

Brown, A. (2015). Universal development and L1-L2 convergence in bilingual construal of manner in speech and gesture in Mandarin, Japanese, and English. *The Modern Language Journal*, *99* (Supplement), 66-82.

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

Brown, A., & Gullberg, M. (2008). Bidirectional cross-linguistic influence in L1-L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in Second Language Acquisition*, *30*, 225-251.

Brown, A., & Gullberg, M. (2011) Bidirectional cross-linguistic influence in event conceptualization? Expressions of path among Japanese learners of English. *Bilingualism: Language and Cognition*, *14*(1), 79-94.

Bylund, E., & Athanasopoulos, P. (2014). Linguistic relativity in SLA: Toward a new research

program. *Language Learning, 64*(4), 952-985.

Bylund, E., & Athanasopoulos, P. (2015). Introduction: Cognition, motion events, and SLA. *The Modern Language Journal, 99* (Supplement), 1-13.

Duncan, S. D. (2006). Co-expressivity of speech and gesture: Manner of motion in Spanish, English, and Chinese. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 27,* 353-370.

Gullberg, M. (2011). Thinking, speaking and gesturing about motion in more than one language. In A. Pavlenko (Ed.), *Thinking and speaking in two languages* (pp. 143-169). Bristol: Multilingual Matters.

Han, Z., & Cadierno, T. (Eds.). (2010). *Linguistic relativity in L2 acquisition: Thinking for speaking*. Bristol, UK: Multilingual Matters.

Hijazo-Gascón, A., & Ibarretxe-Antuñano, I. (2013). Same family, different paths: Intratypological differences in three Romance languages. In J. Goschler & A. Stefanowitsch (Eds), *Variation and change in the encoding of motion events* (pp. 39-54). Philadelphia: John Benjamins B. V.

Kendon, A. (2000). Language and gesture: Unity or duality. In D. McNeill (Ed.), *Language and gesture* (pp. 47-63). Cambridge: Cambridge University Press.

Kita, S., Alibali, M. W., & Chu, M. (2017). How do gestures influence thinking and speaking? The gesture-for-conceptualization hypothesis. *Psychological Review, 124*(3), 245-266.

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*, 16-32.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

Levinson, S. (2012). Foreword. In S. Levinson & P. Lee (Eds.), *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (pp. vii-xv). Cambridge, MA: The MIT Press.

Lucy, J. (1992a). *Language diversity and thought: A reformulation of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lucy, J. (1992b). *Grammatical categories and cognition. A case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press.

Lucy, J. (1997). Linguistic relativity. *Annual Review of Anthropology*, *26*, 291-312.

Mayer, M. (1969). *Frog, Where Are You?* New York: Dial.

McNeill, D. (1992). *Hand and mind*. Chicago: The University of Chicago Press.

McNeill, D. (2000). *Language and gesture*. Cambridge: Cambridge University Press.

McNeill, D. (2001). Imagery in motion event descriptions: Gestures as part of thinking-for-speaking in three languages. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 23,* 255-267.

McNeill, D. (2005). *Gesture and thought*. Chicago: The University of Chicago Press.

McNeill, D., & Duncan, S. D. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp.141-161). Chicago: The University of Chicago Press.

Pavlenko, A. (2014). *The bilingual mind and what it tells us about language and thought*. Cambridge: Cambridge University Press.

Pavlenko, A., & Volynsky, M. (2015). Motion encoding in Russian and English: Moving beyond

Talmy's typology. *The Modern Language Journal, 99* (Supplement), 32-48.

Sapir, E. (1929). The status of linguistics as a science. *Language*, *5*, 207-214. Reprinted in Sapir, E. (1949). *Selected writings of Edward Sapir in language, culture, and personality* (pp. 160-166). Berkeley/Los Angeles, CA: University of California Press.

Sharpen, R. (2016). L1 conceptual transfer in the acquisition of L2 motion events in Spanish and English: The thinking-for-speaking hypothesis. *Open Linguistics, 2*, 235-252.

Slobin, D. I. (1987). Thinking for Speaking. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 13*, 435-445.

Slobin, D. I. (1991). Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics, 1*, 7-26.

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

Slobin, D. I. (1996b). Two ways to travel: Verbs of motion in English and Spanish. In M. Shibatani & S. Thompson (Eds.), *Grammatical constructions: Their form and meaning* (pp. 195-219). Oxford: Clarendon Press.

Stam, G. (2006). Thinking for speaking about motion: L1 and L2 speech and gesture. *International Review of Applied Linguistics, 44*(2), 145-171.

Stam, G. (2010). Can an L2 speaker's patterns of thinking for speaking change? In Z. Han & T. Cadierno (Eds.), *Linguistic relativity in SLA: Thinking for speaking* (pp. 59-83). Buffalo, NY: Multilingual Matters.

Stam, G. (2014). Further changes in L2 Thinking for Speaking? In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body-language-communication: An international handbook on multimodality in human interaction* (pp.1875-1886). Berlin/Boston: De Gruyten Mouton.

Stam, G. (2015). Changes in thinking for speaking: A longitudinal case study. *The Modern Language Journal*, *99* (Supplement), 83-99.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shopen (Ed.), *Language typology and syntactic description: Vol 3. Grammatical categories and the lexicon* (pp. 57-149). Cambridge: Cambridge University Press.

Talmy, L. (1991). Path to realization: A typology of event conflation. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society, 17,* 480-519.

Talmy, L. (2001). *Towards a cognitive semantics: Vol 2. Typology and process in concept structuring*. Cambridge, MA: The MIT Press.

Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: The MIT Press.

Whorf, B. (1940). Linguistics as an exact science. *Technology Review*, *43*, 61-63/80-83. Reprinted in Whorf, B. (2012). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf* (2nd ed.) (pp. 281-297). Cambridge, MA: The MIT Press.

(Copy editing: Alexander Brandt)

About the author

Haizhen Wang (uuwanghz@sina.com) is Associate Professor of Linguistics at the School of Foreign Languages, Soochow University, Suzhou, China. She obtained her PhD in applied linguistics from Nanjing University, Nanjing, China. Her research interests include SLA, language testing and assessment, and English language education.