Amanda Brown* and Jidong Chen

Construal of Manner in speech and gesture in Mandarin, English, and Japanese

Abstract: Two-way typological patterning between satellite- and verb-framed languages in construal of Manner of motion is well attested in speech (e.g. Slobin 2006) and gesture (e.g. McNeill 2001), but contradictory findings exist regarding a third category of equipollently-framed languages (Slobin 2004b). This study examined elicited descriptions of motion from 14 native speakers of Mandarin-Chinese (equipollently-framed), 13 native speakers of English (satellite-framed), and 16 native speakers of Japanese (verb-framed). Results showed that Mandarin and English speakers encoded Manner in speech significantly more than Japanese speakers, and rarely added Manner to gesture when it was absent from speech. However, Mandarin speakers differed from Japanese but also English speakers in gestural highlighting of Manner, where they frequently encoded Manner in speech but not in accompanying gestures, focusing instead on other aspects of the event, specifically Path. These results support a partial three-way typological distinction in construal of motion.

Keywords: gesture, manner, motion event construal, typology, Mandarin Chinese, English, Japanese

*Corresponding author: Amanda Brown: Department of Languages, Literatures, and Linguistics, Syracuse University, Syracuse, NY 13244, USA. E-mail: Abrown08@syr.edu
Jidong Chen: California State University, Fresno, CA, USA

1 Introduction

Since Talmy’s (1991) typological distinction between lexicalization patterns in satellite- and verb-framed languages, much research has focused on cross-linguistic differences in ‘event construal’ (Von Stutterheim and Nüse 2003), i.e. differences in which aspects of an event are considered relevant and therefore verbalized, also referred to as “thinking for speaking” (Slobin 1996a). However, recent work has shown that event construal is not restricted to verbal expression but also includes information expressed in other modalities, e.g. co-speech hand
gestures (e.g. Duncan 2006; Gullberg 2011; Kita and Özyürek 2003; McNeill 2005, inter al.).

This paper focuses on one area of event construal, that of Manner of motion. Numerous studies have found differences between satellite- and verb-framed languages in encoding of Manner in native language speech (e.g. Naigles et al. 1998; Özçalışkan and Slobin 2003; Slobin 1996b, 1997, 2004a, 2004b, 2006; Talmy 1985) and accompanying co-speech gesture (e.g. Brown and Gullberg 2008; Choi and Lantolf 2008; Chui 2009; Duncan 2006; McNeill 2001; McNeill and Duncan 2000; Negueruela et al. 2004; Stam 2006b, 2010). However, few studies have considered construal of Manner in a third category of languages that has recently been proposed by Slobin (2004b), that of equipollently-framed languages, and the findings that exist are contradictory. Here, we examine expression of Manner in three languages to see how speech and gesture contribute to cross-linguistic differences in event construal. Specifically, we systematically compare narrative descriptions of motion from native speakers of Mandarin Chinese, an equipollently-framed language, with those from native speakers of English, a satellite-framed language, and native speakers of Japanese, a verb-framed language, in order to determine how the relationship between spoken and gestural construal of Manner can be characterized in each of the three language types and whether a three-way typological distinction can be observed in this domain.

2 Background

2.1 Construal of Manner of motion in speech

In highly influential work on cross-linguistic expression of motion, Talmy (1991) has claimed that the expression of Manner of motion, i.e. the way in which a protagonist moves (e.g. *jump, roll*), depends on how Path of motion, i.e. the trajectory followed by the protagonist (e.g. *up, down*), is expressed, suggesting that languages can be characterized as either “satellite-framed” or “verb-framed”. In satellite-framed languages like English, the core component of Path of motion is normally lexicalized in a satellite (i.e. verb particle) outside the main verb, leaving the main verb slot free for lexicalization of Manner of motion, as seen in (1). Verb-framed languages like Japanese, on the other hand, typically reserve the main verb slot for Path of motion, leaving Manner of motion to be lexicalized in an adverbial, as in (2) or subordinated verb, as in (3).

(1) *The ball rolls down the hill.*
Typological predictions for lexicalization particularly of Manner, i.e. in the main verb or an accompanying verb-like element, are generally preferences in a language as opposed to grammatical requisites; areas of intra-typological differences exist (for discussion, see e.g. Cadierno 2010), and both satellite- and verb-framed languages usually have alternative options available. English speakers, for example, could lexicalize Manner as an adverbial or a subordinated verb as shown in the translations of (2) and (3) above. Similarly, Japanese speakers could use compound verbs which package Manner and Path in a single lexical item, e.g. koroge-ochiru ‘roll-fall’, as well as a wide variety of ideophones or mimetics, giseigo or gitaigo – “words which imitate sound or shape,” (Weingold 1995: 319), which function as adverbials, e.g. gunguru ‘rotate repeatedly’. However, Slobin (1997, 2004a) has shown that the basic typological differences in lexicalization between verb- and satellite-framed languages affect the size and composition of the lexicon relating to motion, with satellite-framed languages typically possessing a very large and diverse stock of Manner verbs.

Although all languages possess the lexical resources to express Manner, Talmy (1985) hypothesized that independent expression of Manner in a subordinated verb or adverbial “can be stylistically awkward, so that information about Manner . . . is often either established in the surrounding discourse or omitted altogether” (1985: 69). Work by Slobin and colleagues in a range of languages (e.g. Özcализкан and Slobin 2003; Slobin 1996b, 1997, 2004a, 2004b, 2006) has provided substantial empirical support for this proposal (although see e.g. Naigles et al. 1998, and Özcализkan and Slobin 2003, for inclusion of morphosyntactic elements other than the verb, which elevate the frequency of Manner encoding). Such phenomena have motivated frameworks such as “thinking for speaking” (Slobin 1996a), “event conceptualization” (Von Stutterheim et al. 2002), and “event construal” (Von Stutterheim and Nüse 2003) i.e. the notion that speakers will generally select and verbally express aspects of an event that are readily en-

---

1 Abbreviations used in examples are as follows: Acc: accusative case; Clf: classifier; Con: connector; Gen: genitive case; Nom: nominative case; Pfv: perfective aspect; Top: topic marker.
coded by their language, and that, over time, such habitual attention will shape a language-specific rhetorical style.

The empirical studies cited above have focused on a basic two-way typological distinction between verb- and satellite-framing. However, Slobin (2004b) has proposed a refinement of the typology, arguing for a tripartite classification with equipollently-framed languages as a third category. This category is argued to account for serial-verb languages, bipartite verb languages, and generic verb languages in which Manner and Path are expressed by ‘equipollent’ elements – elements that are equal in formal linguistic terms. An example is given from Mandarin Chinese in (4), in which Manner and Path are each represented by verbs, pao ‘run’ and jin ‘enter’ in the form of a verb compound.

(4) Na ge ren pao-jin le fangzi.
    that Clf person run-enter Pfv house
    ‘That person ran into the house.’
    (Chen 2008: 11)

In Talmey’s (1991) original framework, compound constructions such as pao-jin ‘run-enter’ were analyzed as satellite-framed, with the first verb, a Manner verb, considered a main verb and the second verb, a Path verb, considered a satellite, since such verbs form a relatively small closed class and were claimed not to function independently. However, Slobin (2004b) has argued that these Path verbs can function independently as example (5) illustrates.

(5) Ta jin le fangzi.
    he enter Pfv house
    ‘He entered the house.”
    (Chen 2008:12)

In addition, equipollently-framed languages do not appear to follow general satellite- or verb-framed patterns in domains such as lexical repertoires for Manner and Path or the distributional frequency of Manner verbs and Path verbs. The equipollently-framed classification is supported in a number of recent studies on Mandarin (e.g. Chen 2007; Chen and Guo 2009, 2010; Guo and Chen 2009, although see Huang and Tanangkingsing’s 2005, proposal for a four-way typology of motion, with a distinction between serial-verb languages and macro-event languages, where the motion verb consists of a Manner prefix and Path root).

With respect to empirical work on spoken construal of Manner in the three languages included here, Brown and Gullberg (2008) compared elicited oral narrative descriptions of motion events in English and Japanese and found, as pre-
dicted, that native English speakers predominately used verbs to encode Manner and selected Manner for inclusion in spoken discourse at close to ceiling levels (98% of narratives contained explicit reference to Manner information). Similar results for native English speakers can also be seen in Negueruela et al. (2004) and Stam (2006a), where the focus of study was largely gesture patterns (see below). The native Japanese speakers observed in Brown and Gullberg (2008), in contrast, used a range of devices to encode Manner, including monomorphemic and compound verbs as well as adverbials, but only selected Manner for explicit mention in discourse roughly half of the time (although see Sugiyama 2005, for an investigation of a Japanese literary translation of *The Hobbit*, which contained frequent inclusion of Manner). Duncan’s (2006) comparison of elicited oral narrative descriptions of motion events in English and Mandarin also showed that English speakers frequently mentioned Manner (86% of narratives included explicit mention of Manner). However, Mandarin speakers used Manner verbs but frequently reduplicated them, which yielded what she termed “a super abundance of manner coloration” (2005: 359), although importantly, precise frequencies of mentions of Manner or reduplications were not reported. Similarly, Chui (2009, 2011, 2012) found construal of Manner to be very frequent in Mandarin (e.g. 81.6% of elicited descriptions of motion encoded Manner), most often in the form of monomorphemic Manner verbs as well as in Manner-Path verb compounds. These findings contrast to some extent with those from Chen and Guo (Chen, 2007; Chen and Guo, 2009, 2010; Guo and Chen, 2009), whose analyses of oral and written discourse in Mandarin found a rich and diverse Manner verb lexicon but fewer Manner verb types and tokens than previously found for satellite-framed English, and an equal use of Manner and Path verbs unlike the Manner verb pattern dominant in English (indirect comparisons specifically of the frequencies of Manner construal between Mandarin and English were not reported.)

### 2.2 Construal of Manner of motion in gesture

The basic typological differences described above are not limited to spoken linguistic forms and appear to have repercussions for gesture. Some studies have looked at the relationship between lexicalization and gesture-speech synchrony with respect to gesture placement in native language production (e.g. Kellerman and van Hoof 2003; McNeill 1992; McNeill and Duncan 2000; Stam 2006b, 2010), and others have looked at how speech and gesture combine with respect to the degree of semantic overlap between the two modalities and how this constitutes language-specific patterns of event construal. Particularly in the latter domain, McNeill (2001, 2005) has proposed that gestures can be used to modulate the
salience of Manner in discourse in a way that corresponds to the satellite-framed/verb-framed distinction in speech. In satellite-framed languages, where Manner is almost obligatorily encoded in speech even when it is not particularly salient due to its lexicalization in the main verb, McNeill claims that speakers can downgrade its significance by not encoding it in an accompanying gesture, gesturing instead about other components of the event such as Path, so-called “Manner Modulation” gestures. In contrast, speakers of verb-framed languages, in which explicit mention of Manner in speech may be awkward, can alternatively encode it in gesture. As a Manner gesture produced in the absence of a Manner expression in speech would not have a clear semantic correlate with which to synchronize, McNeill argues that such gestures often “blanket” an entire motion event description, so-called “Manner Fog” gestures.

McNeill (2001) drew on individual examples from English, a satellite-framed language, and Spanish, a verb-framed language, for his original hypotheses, but subsequent studies of groups of language speakers have provided some additional empirical support. Negueruela et al. (2004) found that a high proportion (73–77%) of Manner gestures among native speakers of verb-framed Spanish were produced in the absence of accompanying Manner speech, i.e. Manner Fog gestures, and that this was much rarer (0–11%) among native speakers of satellite-framed English (see Stam 2006a, for similar results). In contrast, Brown and Gullberg (2008) found that for native speakers of verb-framed Japanese, a much lower proportion of Manner gestures (11%) was produced in the absence of corresponding semantic information in speech, although such gestures were completely absent from the comparison discourse in English. Brown and Gullberg also examined the frequency of ‘Manner Modulation’ gestures, which they showed was a highly robust pattern in English, where monolingual speakers gestured about elements other than Manner while simultaneously encoding Manner in speech roughly half of the time, in stark contrast to monolingual speakers of Japanese, whose speech-gesture production displayed this pattern significantly less frequently (11%). Note that in this study, a distinction between monolinguallism and bilingualism was crucial since bilingual speakers of Japanese and English patterned midway between monolingual speakers of each language.

A few studies have extended McNeill’s (2001) proposal, which originally addressed only satellite- and verb-framed languages, to Slobin’s (2004b) third category of equipollently-framed languages. Duncan (2006) found the existence of Manner Fog gestures in Spanish and focused on how manifestations of Manner in Spanish gestures were intimately coupled with the concept of Ground, yielding movements conflating both concepts, though frequencies of Manner Fog in Spanish, English, or Mandarin were not reported. Moreover, she calculated that
Manner gestures were as frequent in Chinese discourse (88 Manner gestures across 20 participants), where mention of Manner in speech was claimed to be “super abundant”, as in Spanish discourse (88 Manner gestures across 20 participants), where mention of Manner in speech was claimed to be lower, although again comparisons of proportional frequencies of Manner gestures were not computed. Interestingly, in the analysis of raw numbers of gestures, Manner gestures were rather less frequent in English (58 Manner gestures across 20 participants), where mention of Manner in speech was roughly between Mandarin and Spanish. However, Duncan argued that frequencies of Manner gestures in Spanish and Chinese were likely to be accounted for by the prosodic structures of the languages, and that when taking the motion event as the unit of analysis, Manner gesture frequencies were more comparable across the three languages. These results contrast starkly with Chui (2009, 2011, 2012), who reported a very low occurrence of Manner gestures overall in Mandarin. Chui (2009) found only 3.2% of gestures depicting Manner alone in elicited narrative descriptions of a cartoon by native speakers of Mandarin from Taiwan. After expanding the number of participants and adding spontaneous conversational data, Chui (2011) continued to find infrequent use of Manner gestures (4% in elicited narratives and 17.3% in conversations) but very frequent use of Path gestures (63.3% in conversations and 90.3% in narratives). She observed a high incidence of ‘Manner Modulation’ (McNeill 2001) through gesture, where Path gestures occurred with Manner-only expressions, but a complete absence of ‘Manner Fog’ (McNeill 2001), such that Manner gestures never occurred with Path-only expressions in speech.

3 The current study

The basic two-way typological distinction between satellite- and verb-framed languages originally proposed by Talmy (1991) is both well known and highly influential, with empirical support from many languages and repercussions attested in domains beyond that of lexicalization, including the domain of gesture. Thus far, the general consensus from research on construal of Manner is that native speakers of satellite-framed languages frequently select Manner for expression in speech while native speakers of verb-framed languages do so less frequently (Brown and Gullberg 2008; Duncan 2006; Slobin 1996b, 1997, 2004a, 2004b, 2006; Talmy 1991, inter alia.). Furthermore, whereas native speakers of satellite-framed languages may gesture about aspects of the event other than Manner, e.g. Path (Brown and Gullberg 2008; Duncan 2006; McNeill 2001, 2005), native speakers of verb-framed languages may encode Manner information in gesture even when they do not explicitly mention it in speech (Duncan 2006; McNeill 2001, 2005;
Negueruela et al. 2004), although research on satellite-framed languages in this area is currently limited to English and not all verb-framed languages appear to encode Manner in gesture instead of speech to the same extent (cf. Brown and Gullberg 2008). There is much less consensus, however, with respect to patterns of motion event construal in Slobin’s (2004b) proposed third category, that of equipollently-framed languages. Here, the available research findings on frequency of Manner information in speech and gesture, at least in Mandarin Chinese, are somewhat mixed and based on studies with differing methodologies and focus of analyses (Chen 2007; Chen and Guo 2009, 2010; Chui 2009, 2011, 2012; Duncan 2006; Guo and Chen 2009). Therefore, we do not yet have a consensus on the relationship between speech and gesture in event construal in equipollently-framed languages, nor have we determined how such languages pattern in direct comparison to verb- and satellite-framed languages with respect to encoding of Manner (though see e.g. Chui, 2011, 2012, for indirect comparisons between Mandarin and English).

The purpose of this study is to compare construal of Manner in the speech and gesture of three typologically different languages: English, which has long been classified as satellite-framed; Japanese, which is considered verb-framed; and Mandarin, which was originally analyzed as satellite-framed (Talmy 1991) but has recently been classified as equipollently-framed (Slobin 2004b). Three specific areas are addressed. First, we examine how often speakers of each language select Manner for expression in spoken descriptions of motion events. Subsequent analyses focus on the relationship between speech and gesture with respect to expression of Manner. The second analysis identifies how often gestures highlight the Manner information in present speech, referred to here as ‘Manner-highlighting’ gestures. The third analysis examines how often gestures add Manner information otherwise absent from speech, referred to here as ‘Manner-additive’ gestures.

Previous analyses of part of the current data showed that English speakers encoded Manner more often than Japanese speakers in speech, that English speakers frequently gestured only about Path when Manner information was present in speech and never encoded Manner in gesture when the information was absent from speech, and that Japanese speakers occasionally gestured only about Path when Manner information was present in speech and occasionally encoded Manner in gesture when it was absent from speech (Brown and Gullberg 2008). The current paper contributes new analyses of English and Japanese as well as the inclusion of an entirely new dataset in Mandarin. Given that previous research has shown construal of Manner to be higher in Mandarin than in verb-framed languages in speech (e.g. Chen and Guo, 2010; Guo and Chen 2009), we predicted that Mandarin would not pattern with Japanese in any of the measures.
However, with existing comparisons between satellite- and equipollently-framed languages somewhat incomplete and yielding mixed results, it was not clear where Mandarin speakers would fall in comparison to English speakers with respect to construal of Manner in speech and gesture.

4 Methodology

4.1 Participants

A total of forty-three adults aged between 18 and 48 participated in this study, distributed across three groups: native Japanese speakers residing in Kobe, Japan (n = 16), native English speakers residing in Boston, USA (n = 13), and native Mandarin-Chinese speakers residing in Shanghai, China (n = 14). As previous research has shown an influence of bilingualism in this area (Brown and Gullberg 2008), it was important to control for second language knowledge, and biographical information and information on language usage was gathered using a detailed questionnaire (Gullberg and Indefrey 2003), translated where appropriate. The questionnaire revealed that the study participants had all had exposure to another language at some point in their lives. Although the English speakers had never learned Mandarin or Japanese, the Mandarin and Japanese speakers had all been exposed to English, as is generally the case with such a global language. However, as all participants had had relatively minimal exposure to their second language, were not engaged in active study of a second language, and did not use a second language in their everyday lives, they were considered ‘functionally monolingual’.

4.2 Stimuli

As in previous studies in this area (e.g. Slobin 2004b inter alia), data were obtained through a narrative retelling task. Oral descriptions of the six-minute, animated, Sylvester and Tweety Bird cartoon, Canary Row (Freleng 1950) were elicited from participants. This cartoon, which is commonly used in gesture research on motion events (e.g. Kita and Özyürek 2003; Kita et al. 2007; McNeill 2005; Stam 2006b), was divided into scenes following McNeill (1992), and shown scene-by-scene with a blank screen between each scene followed by participant descriptions, in order to increase the likelihood of mention of individual motion events. Two scene sequences were constructed, which maintained the first and last scenes in first and last position. Participants were shown one of the two se-
quences to control for any order effects and were not informed of the focus on gesture until after the data had been collected. From the stimulus material, four motion events were selected for coding and analysis, which yielded the following Manner-Path combinations: climb through, roll down, clamber up, and swing across. These events were targeted since some had been analyzed in previous studies (e.g. Brown and Gullberg 2008; Duncan 2006; McNeill 2001), they were salient enough in the story line to warrant description by the majority of participants, they illustrated voluntary as opposed to caused motion, they were not at the beginning or end of the experimental task, and they contained types of Manner and Path that could largely be reliably and unambiguously coded in speech and gesture.2

4.3 Procedure

All participants participated individually in the study accompanied by an experimenter, who was a native speaker of the language of the experiment. The participant and experimenter first engaged in a brief warm-up, consisting of small talk in the target language, in order to relax participants and increase the likelihood of gesturing. Next, the experimenter told participants that they would be watching a series of animated scenes from a cartoon on a computer screen and should retell what they had seen in as much detail as they could remember. The experimenter was trained to appear fully engaged in listening to the participants’ narratives, but to avoid asking questions, crucially those related to motion.

4.4 Speech segmentation, coding, and analyses

All narratives were transcribed from digital video by a native speaker of the relevant language. ELAN (Wittenburg et al. 2006), a digital video tagging software program that enables a frame-by-frame analysis (at 40 ms intervals) of movement as well as sound, was used to code the data.

---

2 Regarding the selection of events for coding and analysis, most previous research using this stimulus has considered at most two motion events, generally the climbing and rolling events (e.g. Duncan 2006; McNeill 2001), though Brown and Gullberg (2008) analyzed all four events. In addition, interpretability of particular motions both in speech and gesture is critical. A motion event involving the verb “fall” or its equivalent across languages may be considered Path if the trajectory is prioritized, Manner if the lack of volitional control is prioritized, or Manner and Path if both are considered. Furthermore, the difficulty of distinguishing Manner from Path in gestures depicting a falling event would affect the reliability of coding.
Narratives were first divided into clauses, defined as “any unit that contains a unified predicate ... (expressing) a single situation (activity, event, state),” following procedures laid out in Berman and Slobin (1994: 660). Next, clauses describing the four target motion events were identified and coded. Analyses of speech were based on presence or absence of Manner and were conducted at two levels, event description and clause, in order to address broad and narrow patterns of event construal. Target event clauses also formed the basis for subsequent gesture analyses, which focused on the extent of overlapping Manner information in speech and gesture.

Since calculations of the degree of Manner construal depend on the exact morphosyntactic resources examined, i.e. whether analyses are restricted to verbs or include elements other than verbs (Naigles et al. 1998), a fairly inclusive system was followed by which any element encoding information about the way in which the protagonist underwent its translocational motion was coded as Manner, including adverbials (mimetics and comparison phrases). Examples of speech segmentation and coding of Manner for descriptions of the swing across event in each of the three languages appear in (6) to (8) below with clause boundaries marked by brackets and Manner expressions underlined.

(6) English: E9
[he’s going to swing on a rope]
[so then Sylvester swings over]

(7) Japanese: J4
[jitaku-no mado kara roopu-wo kou yurashite]
home-Gen window from rope-Acc like swing.Con
[ikouto shitandesukedo]
try.to.go did.but
’(He) kind of swung on a rope from his window, and tried to go’

(8) Mandarin: C25
[zhenghao dang guo qu]
just swing cross go
[xiang zai guo qu]
want again cross go
’(He) just swung over; (he) wanted to cross over again’

The event description from a native English speaker in (6) consisted of two clauses, both of which encoded Manner in the verb *swing*. Therefore, in both the event description-level and clause-level analyses, this example was coded as including mention of Manner. The event description from a Japanese speaker in
Amanda Brown and Jidong Chen

(7) also consisted of two clauses. However, this time, only one clause encoded Manner in the verb yurasu ‘swing’. Therefore, while this description was coded as including mention of Manner at the event level, in the clause analysis, only one of the clauses was coded for inclusion of Manner. A similar situation held for the Mandarin example in (8), where the first clause contained expression of Manner in dang ‘swing’ but the second clause encoded only Path.

4.5 Gesture segmentation, coding, and analyses

Co-speech gestures were also observed as part of the study since their inclusion offers a more complete picture of motion event construal, particularly in a cross-linguistic context (e.g. Brown and Gullberg 2008; Duncan 2006; Kita and Özyürek 2003; McNeill 2000, 2001; McNeill and Duncan 2000; Negueruela et al. 2004). Iconic gestures were observed, specifically the ‘stroke’ component of the gesture, which is the most effortful part and contains semantic information (see Kendon 1972; Kita et al. 1998; Seyfeddinipur 2006 for discussion of gesture strokes). Iconic gesture strokes that depicted target motions and co-occurred with clauses containing target motion event speech were coded for presence or absence of Manner.

In line with Duncan (2006), gestures were assigned a Manner code if the movement was “agitated” (2006: 357) in a way that could be interpreted as depicting the Manner component of the following target motion events: CLIMB THROUGH, ROLL DOWN, and CLAMBER UP. This generally involved movement of fingers or rotation of the wrist, e.g. wiggling of fingers, twirling of the arm, or the portrayal of a climbing action, with or without a concurrent trajectory from one point to another (i.e. gestures conflating Manner and Path or those expressing Manner alone). For the swing across event, gestures depicting the existence of an arc-shaped trajectory as opposed to a straight-line trajectory were taken to reflect the Manner component (swinging action) of this event.

Analyses focused on the degree of semantic overlap between speech and gesture with respect to expression of Manner. We first examined the highlighting of Manner information through gesture, where speech included mention of Manner and associated motion gestures also depicted Manner. We next examined the adding of Manner information through gesture, where Manner was absent from a spoken clause but included in the corresponding gesture.

Representative examples of gesture coding processes as applied to clauses describing the clamber up event in each of the three languages are shown in (9) to (11) with individual gesture strokes in boldface and numbered, Manner expressions underlined, and clauses marked by brackets.
(9) English: E₁₁

\[ \text{you see} \ [\text{Sylvester}^{\text{1st}} \ (\text{pause})^{\text{2nd}} \text{slithering up the pipe}^{\text{4th}}] \]

1\text{st} Path gesture: hand moving in upward trajectory with no concurrent finger movement or wrist rotation.
2\text{nd} Path gesture: hand moving in upward trajectory with no concurrent finger movement or wrist rotation.
3\text{rd} Path gesture: hand moving in upward trajectory with no concurrent finger movement or wrist rotation.
4\text{th} Manner-Path gesture: hand moving in upward trajectory with concurrent finger movement representing climbing.

(10) Japanese: J₂

\[ \text{ue-no hou made itte}^{\text{1st}} \]

up-Gen side to go.Con ‘(He) goes to the top’

1\text{st} Manner gesture: rotation of wrist and movement of fingers, hand depicting a climbing action in place (i.e. without upwards trajectory).

(11) Mandarin: J₁₆

\[ \text{wang shang pa}^{\text{1st}} \]

toward top climb ‘(He) climbs up’

1\text{st} Path gesture: hand moving in upward trajectory with no concurrent finger movement or wrist rotation.

The fourth gesture in example (9) illustrates a Manner-highlighting gesture, where Manner is expressed in speech and also in gesture. Gestures one to three in example (9) and the single gesture in example (11), from English and Mandarin speakers respectively, where clauses with spoken encoding of Manner are accompanied by Path gestures, illustrate a lack of Manner highlighting through gesture. Example (10) from Japanese, where a clause with no spoken encoding of Manner is accompanied by a gesture depicting Manner, illustrates a ‘Manner-Additive’ gesture.

5 Results

Three main quantitative analyses were conducted: expression of Manner in speech, expression of Manner in speech and associated gesture (Manner-highlighting gesture), and expression of Manner in gesture with associated speech expressing only Path (Manner-additive gesture). Due to the sample sizes,
all statistical analyses employed non-parametric tests, specifically Kruskal-Wallis for preliminary multiple group analyses and Mann-Whitney for follow-up pairwise analyses.

5.1 Construal of Manner in speech

Two analyses of construal of Manner in speech were conducted, the first at the level of the event description and the second at the level of the clause. Figure 1 shows the mean proportion of event descriptions containing Manner information in speech. For this analysis, the total number of target motion event descriptions in each of the three languages was considered, yielding 161 descriptions (52 event descriptions from Mandarin, 49 from English, and 60 from Japanese).3

The main analysis revealed a significant difference among the groups in inclusion of Manner in spoken event descriptions ($\chi^2(2, N = 43) = 33.363, p < .001$). Follow-up tests showed that monolingual speakers of Japanese ($M = 46\%$) en-

![Fig. 1: Mean proportion of target event descriptions containing Manner information in Mandarin, English, and Japanese.](image)

---

3 As participants were free to describe the stimulus in any way they wanted, not all target motion events were described by all participants. In this case, three English-speaking, three Japanese-speaking, and three Mandarin-speaking participants did not describe the CLAMBER UP event, one Japanese-speaking participant did not describe the SWING ACROSS event, and one Mandarin-speaking participant did not describe the ROLL DOWN event.
coded Manner significantly less often than both monolingual speakers of English ($M = 98\%; z = -5.113, p < .001$) and monolingual speakers of Mandarin ($M = 95\%; z = -4.728, p < .001$), who did not significantly differ from each other ($z = .465, p = .642$).

The second analysis considered expression of Manner in the clause. Figure 2 shows the mean proportion of clauses containing Manner information in speech. For this analysis, the total number of clauses describing target motion events in each of the three languages was considered, yielding 334 clauses (153 clauses from Mandarin, 74 from English, and 107 from Japanese).

The analysis revealed a significant difference among the groups in inclusion of Manner in clauses describing target events ($\chi^2(2, N = 43) = 26.581, p < .001$). Follow-up tests showed again that monolingual Japanese speakers ($M = 32\%$) included Manner significantly less often in clauses than both monolingual English speakers ($M = 80\%; z = -4.641, p < .001$) and monolingual Mandarin speakers ($M = 76\%; z = -4.120, p < .001$), who did not significantly differ from each other ($z = .584, p = .559$).

Overall, analyses of speech showed that native speakers of Mandarin, an equipollently-framed language, appear to include explicit mention of Manner in speech at both the event description and clausal level to the same high degree as native speakers of English, a satellite-framed language, which contrasts with the picture from native Japanese, a verb-framed language, where explicit mention of Manner is much less frequent.

![Fig. 2: Mean proportion of target event clauses containing Manner information in Mandarin, English, and Japanese.](image-url)
5.2 Highlighting of Manner information through gesture

For the analysis of Manner highlighting in gesture, we first isolated target event clauses expressing Manner and containing a gesture, which yielded 187 gestures (91 gestures from Mandarin, 62 from English, and 34 from Japanese). We then calculated the mean proportion of motion gestures encoding Manner when Manner also appeared in the accompanying clause. Figure 3 shows the frequency of Manner-highlighting gestures as a proportion of all gestures accompanying clauses with spoken encoding of Manner in descriptions of target motion events.

The analysis of highlighting of Manner through gesture revealed a significant difference among the three groups ($\chi^2(2, N = 34) = 14.081, p = .001$). Post-hoc tests found that monolingual Mandarin speakers ($M = 16\%$) highlighted Manner through gesture significantly less often than both monolingual English speakers ($M = 39\%$; $z = −2.003, p = .045$) and monolingual Japanese speakers ($M = 67\%$; $z = −3.745, p < .001$), who approached but did not reach a significant difference ($z = −1.705, p = .088$).

Fig. 3: Mean proportion of gestures highlighting Manner out of gestures accompanying target motion event clauses containing Manner in Mandarin, English, and Japanese.

---

As not all participants in the study gestured and of those that did, not all mentioned Manner, this analysis represents 12 Mandarin, 11 English, and 11 Japanese speakers.
5.3 Adding of Manner information through gesture

For the final analysis of Manner-additive gestures, target event clauses that contained a gesture but did not express Manner in speech were isolated, yielding 80 gestures (25 gestures from Mandarin, 5 from English, and 50 from Japanese). We then calculated the mean proportion of motion gestures depicting Manner when no expression of Manner appeared in the accompanying spoken clause. Figure 4 shows the frequency of gestures adding semantic information about Manner to the discourse as a proportion of gestures accompanying clauses with no spoken mention of Manner in descriptions of target motion events.

Descriptively, native speakers of Japanese appeared to add Manner to the discourse through gesture as opposed to speech the most frequently, followed by native speakers of Mandarin, and then by native speakers of English, who never did this, at least in this dataset. However, the statistical analysis of foregrounding of Manner through gesture revealed a difference among the groups (Mandarin $M = 8\%$, English $M = 0\%$, Japanese $M = 24\%$) that approached but did not reach statistical significance ($\chi^2 (2, N = 25) = 5.158, p = .075$).

![Fig. 4: Mean proportion of gestures adding Manner out of gestures accompanying target motion event clauses without Manner in Mandarin, English, and Japanese.](image)

5 As not all participants in the study gestured, and of those that did, many explicitly mentioned Manner in speech, this analysis represents 9 Mandarin, 4 English, and 12 Japanese speakers.
6 Discussion

This paper sought to compare spoken and gestural construal of Manner of motion in three typologically different languages: English, a satellite-framed language, Japanese, a verb-framed language, and Mandarin, an equipollently-framed language. In the first analysis of speech, the results found in analyses of construal of Manner both at the event description level and the clausal level revealed a bipartite difference between Mandarin and English on the one hand and Japanese on the other hand. More specifically, speakers of Mandarin and English included explicit mention of Manner at close to ceiling levels in narratives and to a very high degree in individual clauses describing motion events. These results contrast with those from Chen and Guo (Chen, 2007; Chen and Guo, 2009; Guo and Chen, 2009), who proposed differences between the Mandarin data in their studies and English data in prior studies, at least in various analyses of type and token frequencies. However, as their analysis only considered Manner verbs, the higher mention of Manner reported in Mandarin here may simply be a reflection of our inclusion of adverbials expressing Manner. Our findings also contrast to some extent with Duncan’s (2006) claims that Mandarin is higher on a continuum of spoken Manner expression than English. However, this claim was made on the basis of reduplication of Manner expressions in individual descriptions and not on the quantitative comparisons of presence or absence of Manner information at the narrative discourse or clausal levels undertaken here. Since Mandarin was classified as a satellite-framed language in Talmy’s (1991) original framework and Mandarin does express Manner in verbs, it seems reasonable to expect that speakers of Mandarin would pattern similarly to speakers of other satellite-framed languages such as English with respect to encoding of Manner in spoken discourse. However, clearly more research is needed with replications of current analyses in other equipollently-framed languages before fully claiming that no quantitative distributional distinctions exist between equipollently- and satellite-framed languages with respect to construal of Manner in speech.

In addition to traditional analyses of speech, complementary analyses of co-speech hand gestures were conducted since research has found patterns in gesture that correspond to typological distinctions in speech, which reveal a broader picture of motion event construal (e.g. Gullberg 2011; Kita and Özyürek 2003; McNeill 2001, 2005; McNeill and Duncan 2000). Thus, this paper investigated the extent to which gestures were used to highlight information about Manner already present in speech or add information about Manner otherwise absent from spoken descriptions.

In the analysis of Manner highlighting, we focused only on those gestures accompanying spoken clauses containing Manner in each of the three languages.
Since we already know that Manner is differentially expressed in speech cross-linguistically, this analysis leveled the playing field across languages somewhat as we removed from the calculation any gestures accompanying clauses expressing only Path, which were quite common in Japanese (50 out of 84 gestures). Even with such a leveled starting point, however, there were clear cross-linguistic differences, with equipollently-framed Mandarin patterning differently from the other two languages types. Native speakers of Mandarin simultaneously encoded Manner in speech and gesture quite rarely, and significantly less often than native speakers of English and Japanese. These results contrast with Duncan (2006), who found considerable construal of Manner in gesture in Mandarin.

Taking these results together with the results from the analyses of construal of Manner in speech, we see that Japanese speakers frequently omit Manner from their narratives and clauses leaving fewer clauses containing Manner and available for highlighting of that information through gesture. As Talmy (1985, 1991) has argued, Manner may be expressed by speakers of verb-framed languages only if it is salient; therefore, production of clauses containing Manner by Japanese speakers would likely express salient Manners, and such salient Manners would likely be highlighted or ‘double marked’ additionally in gesture. That being said, it is rather striking that one third of the gestures accompanying Manner information in spoken clauses in Japanese did not highlight spoken Manner. This suggests that multiple levels of Manner salience exits (see below), that the presence of Manner in speech alone may not be a completely reliable indicator of the level of salience of Manner for speakers of verb-framed languages, and that gesture may be used to modulate the significance of Manner information even in verb-framed discourse.

With respect to Mandarin and English, previous findings for English (Brown and Gullberg 2008) have supported McNeill’s (2001) proposal that speakers of satellite-framed languages, who generally talk about Manner whether it is salient or not, may use gesture as a mechanism for downplaying the significance of the Manner information in speech. Contrary to Duncan (2006), the results presented here suggest that speakers of equipollently-framed Mandarin may exploit the same gestural mechanism for modulation of Manner information in speech, a finding that supports data in Chui (2009, 2011, 2012). However, the difference between Mandarin and English in frequency of Manner-highlighting gestures is rather striking given the lack of difference between the languages in construal of Manner in speech.

There are various possible explanations for the gestural difference between Mandarin and English, one being a potential difference in overall gesture frequency across the languages. If Chinese culture is characterized by comparatively lower gesture frequencies, as has been suggested (Nicoladis et al. 2011; So 2010)
and Mandarin speakers gesture less frequently than English speakers, while both cultures may by default depict Path in gesture as the defining component of a motion event, any additional gestures produced by English speakers may include Manner information, making gestural highlighting of Manner information more frequent in English. This potential pattern is illustrated in examples (9) and (11), where speakers of both English and Mandarin, respectively, include Manner and Path in spoken descriptions and highlight only Path in several accompanying gestures, but it is the English speaker who adds a further gesture depicting Manner. At face value, an explanation based on differences in gestural frequency seems rather unlikely for this dataset given that 116 gestures were produced by the 12 gesturing speakers of Mandarin versus 67 gestures produced by 11 gesturing speakers of English. However, in addition to the fact that the Mandarin-speaking and gesturing group was slightly larger than the English-speaking and gesturing group, gesturing and non-gesturing speakers of Mandarin also produced many more clauses overall (153) than speakers of English (74), yielding more opportunities for speakers to gesture. Thus, the cross-linguistic question of gesture frequency has not been adequately addressed here and remains one for future research, which would need to determine the most appropriate unit of analysis, i.e. syntactic versus prosodic, not an uncomplicated matter in and of itself (cf. Nicolaidis et al. 2011). Other possible explanations include cultural differences in perceptions of the salience of Manner information in the stimulus, particularly as the stimulus used here (a Sylvester and Tweety cartoon) is arguably biased towards the culture of the U.S., or potential differences in the participants themselves, for example, differing socio-economic factors associated with functional monolingualism in each of the two countries. However, an alternative explanation may come from the nature of spoken expression of Manner itself. As noted previously, Duncan (2006) claimed that Mandarin is characterized by “a super abundance of manner coloration” (2006: 359) due to the existence of reduplication of spoken Manner expressions, e.g. gun gun gun ‘roll roll roll’. When Manner information is so abundant in speech, such as in cases of reduplicated Manner items, it may be less necessary for Mandarin speakers to highlight the information in gesture. If so, this pattern may even be evident in Japanese, where reduplication is prominent in mimetics, e.g. gorogoro gorogoro “roll roll”. An interesting line for future research, then, would be to examine motion event descriptions with reduplicated Manner expressions in speech to determine whether they are less likely to be accompanied by Manner-highlighting gestures.

The final analysis of additive Manner gestures found that speakers of Mandarin rarely encoded Manner in gesture when it was absent from speech and patterned like speakers of English, who never exhibited the phenomenon. This can be explained by the fact that speakers of Mandarin and English generally
included Manner in spoken clauses to a very high degree, thereby reducing the number of clauses containing only Path where Manner could be added through gesture. Although speakers of Japanese appeared to add Manner through gesture more often than speakers of Mandarin and English, the results were not statistically significant, possibly due to statistical power, and overall frequencies were quite low and highly variable in Japanese, lower and more variable than the frequencies associated with Manner highlighting through gesture.

As Path is argued to be the defining core of a motion event (Talmy 1991), its salience can be assumed and its subsequent construal in spoken and gestured descriptions expected. However, determining whether or not Manner will be perceived as salient and thus expressed by a given speaker for a given event is much more problematic, given the differences in possibilities for lexicalization across languages. Taking all the results above together, we propose that a speaker’s perception of the status of the Manner component in a motion event may be estimated, but only by observing both the speech and gesture they use to describe an event. We further suggest that three levels of Manner salience may potentially be communicated through speech and gesture. At the first level, Manner is considered a non-salient component of a motion event and is not selected for expression in either speech or gesture, leaving specification of only the trajectory. In the second level, Manner is considered fairly salient and is either selected for expression in speech but not highlighted in gesture, particularly for speakers of satellite- and equipollently-framed languages but also for speakers of verb-framed languages to some extent, or not selected for expression in speech but added to gesture, particularly for speakers of verb-framed languages but potentially also for speakers of satellite- and equipollently-framed languages. For this level-two Manner salience, speakers may choose between spoken or gestural construal of Manner depending on the codability of Manner, i.e. whether it is more efficiently encoded in a syntactically simple lexical construction or in a gesture. In the third level, Manner is considered highly salient and is encoded in speech and highlighted in gesture. This proposal is illustrated in Figure 5.

The framework outlined in Figure 5 would predict, all things being equal, first that speakers of equipollently-, satellite-, and verb-framed languages would pattern similarly with respect to non-salient and highly salient Manners, and second that speakers of verb-framed languages would construe in gesture the very same Manners construed in speech by speakers of satellite- and equipollently-framed languages. Although the initial results from this study do not seem to support such a hypothesis given the differing frequencies of Manner-highlighting and Manner-additive gestures and the differences between Mandarin and English in Manner-highlighting gestures, we cannot assume that ‘all things are equal’ in this domain, in particular that all languages and cultures visually perceive, categorize
and recall Manner in a similar way (see e.g. Gennari et al. 2002; Papafragou et al. 2002, for discussion of the potential for cross-linguistic differences in perception and recall of Manner information). Further research should test the validity of a three-tiered perceptual cline of Manner salience and examine whether and how such a cline might relate to the two classes of Manner expressions that have been proposed elsewhere, i.e. first-tier Manner expressions lexicalized by all languages versus second-tier Manner expressions lexicalized by satellite-framed languages (Slobin 1997).

Finally, further research in this area might also consider an additional stimulus that enables analyses of a greater variety of motion events than those included here. The task and stimulus employed in this study facilitated comparisons with prior research as well as the examination of motion event construal in at least a semi-naturalistic setting, which has a greater potential to reveal the rhetorical speech and gesture patterns ‘characteristic’ of particular languages than would the short, targeted descriptions of motion elicited from stimuli where Manner and Path are designed to be equally salient (see e.g. Allen et al. 2007, where previously attested cross-linguistic differences in construal of Manner disappear among adult speakers of English, Japanese, and Turkish). However, the cost of such semi-naturalistic data with this particular stimulus was the restricted number of motion events that were analyzable due to participant freedom in which aspects of the stimulus they chose to describe and which motion depictions could be reliably interpreted in speech and gesture. This restriction may have affected statistical power in some of the analyses, though possibly not in the analyses of Manner-additive gestures since the number of relevant clauses in

---

**Fig. 5:** Levels of Manner salience and associated spoken and gestural encoding of Manner across language types.
English and Mandarin is always likely to be low given that both languages encode Manner in speech at close to ceiling levels. Thus, the ideal stimulus would be one that elicits descriptions long enough for participants to engage in the rhetorical style most characteristic of their language but with a variety of motions that are salient enough to be mentioned and can be expressed in speech and gesture in clear and codable ways.

7 Conclusion

In conclusion, from the results presented here, we argue that equipollently-framed languages pattern generally like satellite-framed languages with respect to construal of Manner in speech, both at the level of the event description and at the level of the clause, and in part with respect to construal of Manner in gesture, specifically, the extent to which speakers use gesture to add information about Manner when it is absent from speech. However, speakers of equipollently-framed languages, specifically Mandarin, may have a lower propensity to highlight spoken Manner information through gesture, which may be related to the characteristics of Manner information within an utterance, e.g. extent of reduplication. Like McNeill (e.g. 1985, 1992, 2005) and others (e.g. Gullberg 2011; Negueruela et al. 2004; Stam 2006b), we view the role of representational gestures as integral to communication in general and construal of motion events in particular. These gestures are tightly coupled with speech and may function, in part, as a complementary and in some cases regulatory mechanism though which speakers of all languages can convey their perceptions of the salience of events in the world.

**Acknowledgments:** Eriko Higashida, Liinlin Wang, Xianwei Wu, Riko Yasunaga, and Star Zhou provided assistance with transcriptions. Two anonymous reviewers offered helpful comments and suggestions on a previous version of this article. All of these contributions are acknowledged with grateful thanks.

This research received technical and financial support from Syracuse University, California State University at Fresno, and the Max Planck Institute for Psycholinguistics.
References


Cadierno, Teresa. 2010. Motion in Danish as a second language: Does the learner’s L1 make a difference? In Zhaohong Han & Teresa Cadierno (eds.), Linguistic relatvity in SLA 1–33. Bristol, UK: Multilingual Matters.


Kita, Sotaro, Aslı Özyürek, Shanley Allen, Amanda Brown, Reyhan Furman, & Tomoko Ishizuka. 2007. Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. Language and Cognitive Processes 22(8), 1212–1236.


Slobin, Dan I. 1996a. From “thought and language” to “thinking for speaking”. In John J. Gumperz & Steven C. Levinson (eds.), Rethinking linguistic relativity 70–96. Cambridge: Cambridge University Press.


