CROSS-LINGUISTIC DIFFERENCES BETWEEN ENGLISH AND JAPANESE IN MONOLOGUES: IMPLICATIONS FROM A MICRO-LEVEL BIDIRECTIONAL ANALYSIS OF L1-L2 DESCRIPTIONS IN SPEECH AND GESTURE-IN-TALK[*]

KAZUKI HATA
(Newcastle University)

Abstract

The present study conducts a cross-linguistic bidirectional analysis of gesture-in-talk with English and Japanese speech as both L1 and L2. The discussion is based on a micro-level qualitative analysis without recourse to statistical perspectives, which is rare in recent bidirectional cross-linguistic gesture studies. In this study, 24 speakers participated in a story-narrative task: 6 monolinguals as L1 speakers of each language and 12 bilinguals. As consistently reported in recent work on gesture studies, the results highlight the high degree of semantic synchrony of gestures with accompanying speech units that encode the same type of information simultaneously. These patterns were recognised in all analysed language groups, regardless of the language. The data considered here, however, also demonstrates that the speakers possess different semantic conceptions regarding the meaning of the motion verb *climb* and its Japanese equivalent *noboru* 'climb (up)', making it difficult to conduct an equivalent comparative analysis of speech production in L1 and L2. Thus although the present case study supports the recent findings in the literature regarding the semantic synchrony and uniformity of gestural shapes, it also reveals that the effect of inequivalent semantic variation should be taken into account during cross-linguistic investigation of the speech-gesture relationship in speakers' descriptions of motion events.

1. Introduction

The present study investigates the use of speech and the use of iconic gestures (henceforth, iconics; McNeill 1992) in two target languages, English and Japanese, occurring in story-narrative tasks. The primary research question is whether idiosyncratic variation and preferences affect the use of verbal and gestural expressions by monolingual and bilingual speakers in each target language, and if so, to what extent. Here, it is worth emphasising that the present study considers the expression of manner or path information in speech and gesture, in connection with the effect of semantic variation of specific word forms, via a micro-level analysis with no recourse to statistical perspectives. This micro-level analytic investigation suggests that it is arguably necessary to focus on the semantic concepts in speakers' minds as some words in one language represent semantic information in a different way from their translation equivalents in other languages (Brown & Gullberg 2011: 91). The study design is fully experimental, video-recorded and task-oriented, and therefore, results in the present study may not be comparable to those from natural speech contexts (see Garfinkel 1996: 11, ten Have 1999: 48-49).

---

2. Background

The term 'gesture-in-talk' (i.e. 'gesticulation'; henceforth, 'gesture') is defined as embodied movements with communicative roles, which are systematically intertwined with spoken words by encoding the semantically and/or pragmatically parallel message (Kendon 1980, 1996, 2004, McNeill 1992). This suggests that gestural expressions are neither just co-productions of speech nor simple translations of utterances (McNeill 1992: 105). They can be seen as tightly synchronised symbolic signals co-expressed with speech to relay messages that are not ready-made for spoken forms of language (McNeill 1985: 360, Clark & Krych 2004: 78, Holler and Beattie 2004: 67, Wilcox 2004: 525). Maintaining this, the major difference is that utterances in spoken language generate a meaning by linguistic items as a 'constructed unit', but gesture depicts a complete picture without a strictly obligatory construction process, thus working as a 'presentational unit' (McNeill 1992: 21, Kendon 1994: 49).

Recent findings about gestural functions in L1 and L2 shed light on cross-linguistic variation regarding encoded messages in speech and gesture and patterns of use. In studies based on story-narrative tasks, a key consideration is often the speaker's pattern of encoding information about manner and path in speech and gesture to describe the target motion events in the stimulus. According to Talmy (1985), expressions of manner vary among different languages depending on how path information for motions is encoded. Consider the following examples (1)-(2):

(1) a. I twisted the cork into the bottle.
     b. I twisted the cork out of the bottle.   (Talmy 1985: 72)

(2)   *booru-ga saka-wo korogatte iku*
         ball   hill   roll   go
         'The ball goes rolling on the hill'   (Brown & Gullberg 2008: 228)

English is classified as a satellite-framed language, which usually encodes trajectory in a combination of satellites, that is, 'certain immediate constituents of a verb root' (Talmy 1985: 102, 1991). Thus, the main verb might not be affected by the expression of path information, as in (1a) and (1b). On the other hand, verb-framed languages like Japanese express path information in the main verb, as in (2). In the latter case, manner information could be expressed in either a subordinated verb or adverbial, or it is omitted altogether (Talmy 1985, Slobin 2004a, 2004b).

Although the literature shows cross-linguistic variation between satellite-framed and verb-framed languages, one source reports that iconics, a type of representational gesture performed in talk, demonstrate a high degree of uniformity with regard to gestural shapes, despite considerable differences between spoken languages (McNeill 1992: 221-222, see the following section for the semiotic features of iconics). This claim can be explained by the Growth Point theory (GP) following Vygotsky (1986), which holds that speakers' minimal idea units can develop from an image into 'a complex structure in which both the gestures and the linguistic structure are integral parts' (McNeill 1992: 29-30). If the GP is a starting point of thought coming from a presupposed background, it should only exhibit slight variation from language to language. However, if different languages encourage different ways of thinking (Slobin 1991: 23), there must be room to suggest that gesture will also vary across languages. In fact, later studies report cross-linguistic variation regarding the use of iconics across different languages (Kita 2000, McNeill & Duncan 2000). Specifically, Kita and Özyürek (2003) found that gestures depicted spatial details that might not be expressed in the accompanying speech of some languages due to inequivalent lexical entries to describe the specific motion event. For

example, there is a lack of linguistic resources to express the swing-across motion in Japanese, which leads to difficulty in articulating arc movement. English uses one clause to express manner and path, whereas Japanese and Turkish tend to use two or more. Given this, Kita and Özyürek propose 'the Interface Hypothesis', based on GP theory, claiming that gestures arise from an interface representation between speaking and spatial thinking, which is information tailored for speech, that is, thinking-for-speaking (McNeill & Duncan 2000). That is to say, gestures are considered to be formed in a different way across languages similarly to the different patterns in linguistic representation, and therefore, linguistic differences influence how spatial information arises in the process of speech production and is expressed by gesture in story narratives.

However, bidirectional cross-linguistic differences between L1 and L2 speakers are arguably still unclear, as there have been relatively few attempts to collect data on Japanese as L2, compared to an accumulation of English samples (see McCafferty 1998, Kita 2000, Kita & Özyürek 2003, Brown & Gullberg 2008, 2011). Given this, there is a research gap in the study of cross-linguistic variation in the expression of manner in speech and gesture between English and Japanese. Consequently, the present study conducts a bidirectional analysis involving the two target languages: both L1 and L2 use of English and Japanese. Considering that gestures might demonstrate idiosyncratic (i.e. inter- or intra-speaker) variation between speakers and/or a single speaker (Kendon 1996, Kipp, Neff & Albrecht 2007), statistical comparison is not a valid and/or reliable procedure. Thus, this study undertakes a micro-level qualitative analysis with no recourse to statistical perspectives in order to investigate the expression of manner and path information in speech and gesture.

3. Methodology
3.1. Participants

The present study recruited 24 people aged 22 to 35 at the time of recording: 12 British English L1 speakers and 12 Japanese L1 speakers. These speakers fell into four different sub-groups depending on their language background: English monolinguals (EM), English bilinguals (EB), Japanese monolinguals (JM) and Japanese bilinguals (JB). In order to standardise their background, the study defines monolingual speakers as people who were born and raised in their home country, and have not had long-term experiences of studying/working abroad and are not acclimatised to other cultures as well as or more than their first culture. This requirement was set to minimise the effect of the speaker's culture on gesture-in-use with the diversity and complexity of meaning (see Birdwhistell 1970, Morris 1979, Richmond & McCroskey 1999). Bilingual speakers in this study, on the other hand, are simply defined as people who use or had used the L2 in their daily activities.

Regarding the categorisation of the participants as monolingual or bilingual, their proficiency level and language experience in L2 were considered, supported by their score in official language proficiency tests for L2 (i.e. IELTS and JLPT; see below) and their experience of studying abroad; all the bilinguals had experience studying or working abroad. The general information on language usage and background, collected by using a pre-test questionnaire, which was based on the language background questionnaire developed by the Max Planck Institute (Gullberg & Indefrey 2003), is summarised in Table 1.

Table 1. Summary of participants, based on the language background questionnaire

Language background	English monolingual (n=6)	English bilingual (n=6)	Japanese monolingual (n=6)	Japanese bilingual (n=6)
Mean age	36	25	37	27
Mean JLPT / IELTS score	-	N2 (JLPT)	-	6.5 (IELTS)
Mean AoE to L2	0 yrs	5.3 yrs	8 yrs	12.8 yrs
Mean usage of L2 (per day)	-	2.1 hrs	-	5.1 hrs
Mean self-rating in L2 use	-	4.1	-	3.6
Study abroad experience	-	1-3 yrs	-	1-2 yrs

Key: AoE = Age of Exposure; IELTS = International English Language Testing System (Score range: 0.0-9.0 as highest); JLPT = Japanese-Language Proficiency Test (Grade range: N5-N1 as highest)

3.2. Stimulus

The study used an animated colour cartoon *Canary Row* (Freleng 1950) as the stimulus due to its potential not only to evoke a variety of gestures (McNeill & Levy 1982) but also to be used with a variety of speakers, including L2 learners (McNeill 1992: 365). For research purposes, the focus of this study lies on the participants' descriptions of the two climb-up motions shown in the stimulus, as both languages potentially have enough lexical entries to describe these motion events (see Kita & Özyürek 2003 for descriptions of 'swing across' in the same video stimulus). The study used an edited version of the cartoon, which was shortened to approximately two minutes and included the two target motion events, in order to allow it to be easily retold within a single attempt.

3.3. Procedures

Each participant was involved in the experiment for one day in the case of monolinguals or two days in the case of bilinguals; all the bilingual speakers were gathered again and saw the same stimulus for the same task in a different language. In order to counterbalance their L1 and L2 performances, half of the group of participants were randomly required to tell the story in the L1-L2 order, and the other half did so in the L2-L1 order. This procedure was necessary to control the unwelcome effect stemming from their background information on either L1 or L2 descriptions (Gullberg 2003). A second meeting took place one week following the first recordings.

Prior to the recordings, each participant met with one of the participating researchers for a short informal conversation about general topics for two or three minutes. This served as a warm-up task in the language to be used, with two aims: (a) to relax participants and (b) to put them into the 'monolingual mode' (Grosjean 1998, Brown & Gullberg 2008). Following the warm-up section, each participant was given instructions. They were told that they were participating in a story-retelling task and informed that their retellings would be videotaped, and the tapes would be shown to other participants who do not know the story. The latter is a pseudo-instruction, strategically placed to enforce participant output accuracy during production. At that point, no specific information about the stimulus was given to the participants except its length. During the activities, participants were strongly advised to speak not to the video camera but to the researcher in front of them, as some studies showed that

absence of listeners might lead to a reduction in the rate of gesture frequency (Cohen & Harrison 1973, Bavelas, Chovil, Lawrie & Wade 1992, Alibali et al. 2001).

After the participants were given sufficient instruction, each of them signed the written consent form. The focus of this study was not mentioned to them at any point prior to execution of the task, in order to record natural and spontaneous behaviour and to avoid participant attention being directed at the focus (McNeill 1992: 374). After the completion of the experiment, they were informed about the true aim of the study, and were again asked for permission to use their data after the recordings. A post-hoc interview took place immediately following the task. This was recorded and it was confirmed that no participants identified the focus of this study while executing the task.

After the recording phase, a follow-up questionnaire with a semantic judgement task was carried out in order to check their semantic conceptualization of the English word *climb* and its Japanese equivalent *noboru* 'climb (up)'. As initially suggested in Brown and Gullberg (2011: 91), it was uncertain whether Japanese speakers used these words as manner or path verbs. This post-recording interview clarified that bilinguals were seen to have different semantic conceptions of the verb *climb* and its Japanese translation compared to monolinguals.

3.4. Annotation process

The present study identified that all gestures which accompanied speech referring to the climb-up motion event are iconics. Iconics, a semiotic type of gesture-in-talk, are seen as more semantically tied to the accompanying speech than other types of gestures, showing a close synchronicity between speech and gesture (McNeill 1992, see also Beattie & Shovelton 2002, Holler & Beattie 2002, 2004). This type of gesture represents the same feature of the action or event as its verbal description and is thus semantically parallel to the verbal expression (McNeill 1985, 1992). Iconics are also used to encode manner or path information which is not directly expressed in accompanying speech (Brown & Gullberg 2008). As they encode manner or path information, iconics can represent the viewpoint of the narrator: the character's viewpoint (C-VPT) and the observer's viewpoint (O-VPT) (McNeill 1992: 123-133). For example, a speaker depicts the event from the C-VPT when moving his/her arms up and down, demonstrating the character's motion going up the ladder from the character's viewpoint. Alternatively, he/she can refer to the same event from the O-VPT when raising up his/her hand, depicting the path information of the character rising upward (McNeill 1992: 118-119).

The study used transcriptions and annotations of video data created through ELAN annotation software, a free multi-modal annotation tool for audio and video data originally developed at the Max Plank Institute for Psycholinguistics (Wittenburg et al. 2006). The software was utilised for spatio-temporal synchronisation between the different types of data with separate annotation tiers for two different types of communicational signals, speech and gesture. All spoken data were represented in orthographic transcription and synchronised with the gesture annotations and the corresponding audio and video data in the software. As the study investigates non-linguistic communicative signals (i.e. gesture) and non-English language in use, transcription and its translation, as well as gesture annotations, were double-checked by three annotators through discussion, in order to establish 'inter-rater reliability'; this is arguably a mandatory process to avoid too much selectivity arising from a single annotator in any transcription/annotation processes (McNeill 1992: 375, Chawla & Krauss 1994: 588, Kendon 1996: 5). The study recruited three Japanese speaking L2 English speakers to translate Japanese utterances into their English equivalents, and an English-speaking annotator with experience in annotating gestures. In each case, a consensus was reached through discussions among the annotators. The list of motion verbs uttered by different groups of participants is shown in Table 2.

Within the annotation process, gestures were annotated utilising McNeill's semiotic gesture categorisation. The present study analysed only the stroke phase of each gesture, as this segment is considered obligatory and thus the most significant part of a gestural movement (McNeill 1992, Kita, van Gijn & van der Hulst 1998, Kita 2000). Each stroke phase was identified through annotation of the gesture's internal temporal structure, which identifies the start and end point of the stroke or hold phase (Kipp et al. 2007: 330). In the present study, each gesture annotation is described with co-expressed accompanying speech by stating the stroke phase in boldface and bracketing the entire gesture phrase, followed by the description of the actual movement, as in example (3).

(3)   The cat is [**climbing** on the wall]
        Manner: wiggling movement of hands depicting the manner information for the climb-up motion of the character

Table 2. List of motion verbs linguistically expressed in monologues

	English monolingual	English bilingual	Japanese monolingual	Japanese bilingual
English	climb (+up) crawl get (+into / up) go (+up / through)	climb (+up) get (+into/up)		climb (+up) approach (+to) enter go (+up)
Japanese		*noboru* (climb) *agaru* (go up) *iku* (go)	*noboru* (climb up) *iku* (go) *chikazuku* (approach) *haiagaru* (climb up)	*noboru* (climb up) *kuru* (come) *iku* (go) *chikazuku* (approach) *haiagaru* (climb up)

Key: 1. (+up) = the word 'climb' was in some cases followed by the adverbial 'up', but not always.
2. Semantically, '*noboru*', '*agaru*' and '*haiagaru*' are always only associated with the description of upwards motion; however, it should be noted that this semantic value was violated by English-Japanese bilinguals.

4. Analysis of speech-gesture synchrony

Remarkably, all the manner-encoding gestures were strictly tied with the motion verbs referring to the climb-up event, regardless of the speaker's language background. The Japanese speaking participants who used manner-encoding gestures always synchronised the stroke phase with the motion verb rather than adverbials or noun phrases, thus directly expressing the climb-up motion. Consider the following examples (4)-(6).

(4)   *isshou-kenmei [**nobotte**   iku]*
        with his best   climb up   go
        '(he) goes **climbing up** with his best'
        Manner: wiggling movement of hands depicting the manner information for the climb-up motion of the character

(5)   *haisuikan-wo   [**nobori-masu**]*
        the drainpipe   climb up
        '(he) **climbs up** the drainpipe'
        Manner: wiggling movement of hands depicting the manner information for the climb-up motion of the character

(6)   he tries to [**climb up** on the wall]
        Manner: wiggling movement of hands depicting the manner information for the climb-up motion of the character

        Example (4) shows that the stroke of the manner-encoding gesture starts by saying *nobotte* 'climbing (up)'. The sentence pattern seen in (4), packaging a manner verb and a path-verb within the same unit (Ohara 2002, Sugiyama 2005), is called a quasi-lexical unit (Slobin 2006: 69). Interestingly, this synchrony is also seen in descriptions by the English bilinguals in their L2 narratives, such as in (5); although a quasi-lexical unit was not seen in this example. There is inter-speaker variation in the length of the stroke; some of the speakers hold their hand movement right after *climb*, but the others continue their moving gestures until they complete the sentence. Nevertheless, the starting point of the stroke is similar and always tied with accompanying spoken words encoding the same type of information (i.e. manner). Interestingly, this pattern is found not only in the Japanese-speaking narratives but also in the English-speaking attempts, as in (6). Whatever the circumstances, the speakers always executed the stroke phase when uttering the main verbs as a reference to the climb-up motion, when they encoded the manner information in gestures.
        Significantly, such a strict speech-gesture synchrony, regardless of language selection, is also seen in path-encoding gestures. For all speakers, the stroke of path-encoding gestures is tied to the adverbials but not to the main verbs, regardless of the language spoken; compare (7)-(9).

(7)   *amadoi-wo   [**totte**   chikazuku]*
        the drainpipe   through   come close
        '(he) comes close (by going) **through** the drainpipe'
        Path: upward movement of one hand depicting the trajectory information of the climb-up motion of the character

(8)   *nakanihai-tte [**ue-made**   itte]*
        going inside   into up   go
        '(he) gets into the inside (of the pipe) then goes **up**'
        Path: upward movement of one hand depicting the trajectory information of the climb-up motion of the character

(9)   he climbs up [**through** the drainpipe]
        Path: a hand movement rising up depicting the trajectory information of the climb-up motion of the character

The results show that the stroke of the path-encoding gestures is synchronised with an expression of path information in speech. Some speakers hold gestural shapes until they complete the sentence in speech, but others simply tie path-encoding gestures to the accompanying adverbials.

Note that these findings do not generate new perspectives, but rather support the claim of speech-gesture synchrony in the literature (e.g. McNeill 1985, 1992, Kendon 1997, Gullberg 2003), by claiming that gestures tend to be performed at the same time as semantically parallel linguistic items. Thus, gestures co-expressed with speech typically encode the same information derived from the same 'idea unit'; the meanings conveyed by the speech and the gesture complement each other in depicting this identical information (McNeill 1992: 27). In fact, the data in this study show that each dynamic stroke with a core message is seen as semantically synchronised with the content of the accompanying speech unit. The data also demonstrate that the stroke never crosses clause boundaries in speech although gestures are held until the sentence is completed, similarly to McNeill's discussion of the optional hold phase (1985: 361).

Another notable point is that semantic synchrony is seen in all samples regardless of the speaker's language background and the language to be used in the narratives, which casts doubt on the idea of cross-linguistic variation in gesture-in-use. As McNeill claims, the synchrony rule will not always be clear since some iconics are so vague that it is difficult to show that the gestures clearly encode the same idea unit as the speech does (1992: 27). Nevertheless, the study provides evidence of the semantic synchrony between gestures and accompanying speech units, which holds for all the speakers recruited in the present study. Participants in the present study demonstrated uniformity of gestural shape usage. Regarding this, the inter-speaker (or intra-speaker) variation of gestures is arguably still in debate; it has been on the one hand reported that there is consistency of gestural shapes between different language speakers unless they describe the same event (McNeill 1992), but on the other hand it is claimed that there might be significant variation (Kendon 1996, Kipp et al. 2007). Maintaining that different language speakers would demonstrate different ways of thinking (Slobin 1991), it is expected that gestural expressions have significant inter-speaker variation. On the other hand, this paper will demonstrate that all the speakers, regardless of their language background, exercised manner- or path-encoding gestures with similar shapes to describe the same motion event.

There are two patterns of gestures used to depict either the manner or path information of the climb-up motion. Figure 1 represents a typical example of manner-encoding gestures showing the climb-up motion event resulting from the stimulus. The speakers wiggled both hands one after the other, which depicted the motion of the cat from the C-VPT. This form of the gesture resembles the climb-up motion of the cat wiggling both paws when climbing up the drainpipe in the stimulus. Regardless of the use of words or phrases spoken, the accompanying manner-encoding gestures of different speakers are similar in terms of their shapes. Another example of path-encoding gestures, with one hand moving up, depicted the path of the character from the O-VPT as shown in Figure 2.

Figure 1. The screenshots at the point when speakers wiggled both hands one after the other to depict manner information of the character

Figure 2. The screenshots at the point when speakers moved the right hand with right finger raised up to depict the path of the character

Considering the high degree of similarity between gestural shapes among all the speakers, the effect of specific language preferences is questioned. Based on the results, it is reasonable to suppose that gestures are derived from the GP as a starting point of thought coming from the information in the stimulus rather than from the effect of linguistic preferences regarding information encoding (McNeill 1992, McNeill & Duncan 2000). As speech is expressed in completely or slightly different ways in different languages (Talmy 1985, Slobin 1991), it is natural to expect that, if linguistic preferences affect the means of encoding information in gestures, the gestures used by speakers with different language backgrounds would demonstrate significant variation according to their L1. However, this assumption does not explain the high degree of uniformity of gesture shapes with semantic synchrony across all the speakers observed in this study. Whereas it is acknowledged that there is a slight difference in terms of gesture space, in which kinesic forms and movements are performed (see McNeill 1992, Kendon 1996), similar shapes are seen in the gestures used in the present study. However, note that the result is tentative with only six speakers for each group, and thus further investigation is required to generalise this point.

5. Discussion: an implication of the impact of translation equivalents in L1-L2

As described so far, no Japanese L1 speakers in L1 narratives depicted the event with manner expressions in their speech, whereas the English L1 speakers in L1 narratives tended to use manner information encoded in their speech. This pattern is actually also seen in their gestural expressions. Namely, a greater number of English speakers used manner-encoding gestures than Japanese L1 speakers. This point is supported by recent findings claiming that languages

stimulate different forms of thinking even when speakers express the same motion events (Kita 2000, McNeill & Duncan 2000, Kita & Özyürek 2003). More specifically, this cross-linguistic difference in speech and gesture patterns between Japanese and English conforms to the suggestion by Brown & Gullberg (2008) that more manner expressions in gestures are found for English speakers than Japanese speakers. This result suggests that linguistic preference has an effect on gestural preferences in encoding information.

Nevertheless, it should also be acknowledged that not every EM, EB, and JB speaker when speaking English used manner-encoding gestures. Rather, even speakers of English as L1 decided to encode the path information rather than the manner information in more than half of the total cases; this does not seem so significant when compared to the data from the Japanese L1 speakers. Arguably, this tendency contradicts the idea that due to linguistic preference the speakers paid more attention to manner information of the motion and hence used more manner-encoding gestures. Unfortunately, the study could not conduct a statistical analysis due to the limited number of samples, even though it has offered a micro-analytic perspective, and thus leaves this question for future studies.

Rather, what the micro-level analysis highlights is that very few studies to date have considered the difference in the semantic information contained in specific words between the original word and its translation equivalent. For example, the Japanese word *noboru* 'climb (up)' is claimed to be a path-encoding verb as it comprises only the meaning of the up direction, although the word is in most cases translated as *climb* in English, a manner-encoding verb, which can be co-expressed with adverbials encoding the meaning of both 'up' and 'down' directions (Brown & Gullberg 2011: 91). As for speech descriptions, the emphasis should be firstly on the classification of verbs as encoding either manner or path information. The participants recruited in the present study used a variety of verbs encoding the climb-up motion event (refer back to Table 2). Here, the novel finding from the post-recording questionnaire is that bilingual speakers judged semantic concepts of motion verbs differently from monolingual speakers when speaking their L2. Interestingly, the study reveals that JB speakers treated *climb* in English as a path verb since they associated the word with only the meaning of the up direction, saying that *climb down* is semantically unacceptable. This wrong semantic judgement is also seen in EB speakers, as they treated *noboru* 'climb (up)' as a manner verb because they thought that the Japanese verb could be used to refer to either the up or the down direction, although Japanese L1 speakers do not use the verb in the latter case.

With this classification in mind, it is observed that there were no cases of manner expressions in speech by JM and JB speakers in L1 narratives. This is consistent with the suggestion that Japanese, a verb-framed language, tends to encode manner in adverbials rather than verbs (Talmy 1985, Sugiyama 2005, Slobin 2006). This is not the case for the EM speakers and EB speakers when speaking Japanese as their L2. Regarding the latter case, their Japanese descriptions include the use of '*noboru*' as a manner verb. Although it is true that some EM and EB speakers encode the path information verbally with path verbs, such as 'go', all Japanese L1 speakers encoded only the path information. This supports the claim that speakers of a verb-framed language do not encode manner as much as those of a satellite-framed language (Hohenstein, Eisenberg & Naigles 2006, Brown & Gullberg 2008). In the present study, L1 speakers of Japanese used various verbs to depict the motion event. However, all of them are semantically classified as path verbs encoding the up direction. From this perspective, it is indeed arguable that no cases of manner-encoding verbs in the descriptions by Japanese speakers could be simply explained by limited linguistic entries in Japanese; since EB speakers also encoded no manner information in Japanese speech when strictly considering *noboru* 'climb (up)' as a verb containing the meaning of 'only' an upward direction in a strict sense.

Given these points, it should be acknowledged that linguistic resources for depiction of a climb-up motion are not equivalent between the two languages due to the variety of semantic

conceptions of words and limited lexical entries in Japanese compared to English. Although the data samples evidence L1 speakers of English focusing on manner, such language qualities arguably do not allow an equivalent comparison of different patterns in speech between these two languages. Therefore, the study does not take into account the different patterns in speech depicting manner or path information of the motion event.

6. Conclusion

The present study shows the high degree of semantic synchrony between gestures and their accompanying speech units. Furthermore, it is notable that there is a uniformity of gesture-in-use not only in the gestural phrase but also in the actual shapes encoding information. These patterns are highly recognisable in both languages in this study, regardless of speakers' linguistic background. Although these findings support the findings of previous empirical gesture studies, this study goes further in that it highlights significant challenges present when investigating cross-linguistic differences in terms of monolingual and bilingual speakers. This micro-level analysis of speech and gesture production by monolinguals and bilinguals has implications regarding the inequivalent semantic concepts of *climb* and its Japanese translation. Nonetheless, the findings presented here require qualification, as there were only six speakers recruited for each group.

References

Alibali, M. W., Heath, D. C. & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language* 44, 169-188.

Bavelas, J. B., Chovil, N., Lawrie, D. A. & Wade, A. (1992). Interactive gestures. *Discourse processes* 15(4), 469-489.

Beattie, G. & Shovelton, H. (2002). An experimental investigation of some properties of individual iconics that mediate their communicative power. *British Journal of Psychology* 93(2), 179-192.

Birdwhistell, R. L. (1970). *Kinesics and context: Essays on body motion communication*. Philadelphia: University of Pennsylvania Press.

Brown, A. & Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture: A study of Japanese speakers of English. *Studies in Second Language Acquisition* 30(2), 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.

Chawla, P. & Krauss, R. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology* 30(6), 580-601.

Clark, H. H. & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language* 50(1), 62-81.

Cohen, A. A. & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of Personality and Social Psychology* 28(2), 276-279.

Freleng, I. (Director). (1950). *Canary Row* [Animated film]. New York: Time Warner.

Garfinkel, H. (1996). Ethnomethodology's program. *Social Psychology Quarterly* 59 (1), 5-21.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition* 1(2), 131-149.

Gullberg, M. (2003). Gestures, referents, and anaphoric linkage in learner varieties. In Dimroth, C. & Starren, M. (eds.), *Information structure and the dynamics of language acquisition*, 311-328. Amsterdam: John Benjamins.

Gullberg, M. & Indefrey, P. (2003). *Language history questionnaire*. http://www.mpi.nl/departments/other-research/research-projects/the-dynamics-of-multilingual-processing/tools/ [Retrieved on 06 September, 2014]

Hohenstein, J., Eisenberg, A. & Naigles, L. (2006). Is he floating across or crossing afloat? Cross-influence of L1 and L2 in Spanish-English bilingual adults. *Bilingualism: Language and Cognition* 9(3), 249-261.

Holler, J. & Beattie, G. (2002). A micro-analytic investigation of how iconics and speech represent core semantic features in talk. *Semiotica* 142, 31-70.

Holler, J. & Beattie, G. (2004). The interaction of iconic gesture and speech in talk. In *Gesture-Based Communication in Human-Computer Interaction*, 63-69. Berlin: Springer.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In Key, M. R. (ed.), *The relationship of verbal and nonverbal communication*, 207-228. New York: Mouton Publishers.

Kendon, A. (1994). Human gesture. In Gibson, K. R. & Ingold, T. (eds.), *Tools, language and cognition in human evolution*, 43-62. Cambridge: Cambridge University Press.

Kendon, A. (1996). An agenda for gesture studies. *Semiotic review of books* 7(3), 8-12.

Kendon, A. (1997). Gesture. *Annual Review of Anthropology* 26, 109-128.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kita, S. (2000). How representational gestures help speaking. In McNeill D. (ed.), *Language and gesture,* 162-185. Cambridge: Cambridge University Press.

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

Kita, S., van Gijn, I. & van der Hulst, H. (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In Wachsmuth, I. & Fröhlich, M. (eds.), *Gesture and sign language in human-computer interaction*, 23-35. Berlin: Springer.

Kipp, M., Neff, M. & Albrecht, I. (2007). An annotation scheme for conversational gestures: how to economically capture timing and form. *Language Resources and Evaluation* 41(3), 325-339.

McCafferty, S. G. (1998). Nonverbal expression and L2 private speech. *Applied Linguistics* 19(1), 73-96.

McNeill, D. (1985). So you think gestures are nonverbal?. *Psychological Review* 92(3), 350-371.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. London: University of Chicago Press.

McNeill, D., & Duncan, S. D. (2000). Growth points in thinking-for-speaking. In McNeill, D. (ed.), *Language and gesture,* 141-161. Cambridge: Cambridge University Press.

McNeill, D. & Levy, E. T. (1982). Conceptual representations in language activity and gesture. In Jarvella, R. J. & Klein, W. (eds.), *Speech, place, and action: Studies of deixis and related topics,* 271-295. New York: Wiley.

Morris, D. (1979). *Gestures, their origins and distribution*. London: Jonathan Cape.

Ohara, K. H. (2002). Linguistic encodings of motion events in Japanese and English: A preliminary look. *Hiyoshi Review of English Studies* 41, 122-153.

Richmond, V. P. & McCroskey, J. C. (1999). *Nonverbal behavior in interpersonal relations*. 4th ed, Boston: Allyn and Bacon.

Slobin, D. I. (1991). Learning to think for speaking: Native language, cognition, and rhetorical style. *Pragmatics* 1(1), 7-25.

Slobin, D. I. (2004a). How people move: Discourse effects of linguistic typology. In Moder, C. L. & Martinovic-Zic, A. (eds.), *Discourse across languages and cultures,* 195-210. Amsterdam: John Benjamins.

Slobin, D. I. (2004b). The many ways to search for frog: linguistic typology and the expression of motion events. In Stromqvist, S. & Verhoeven, L. (eds.), *Relating events in narrative, Volume 2: Typological and contextual perspectives,* 219-257. Mahwah: Lawrence Erlbaum Associates.

Slobin, D. I. (2006). What makes manner of motion salient?. In Hickmann, M. & Stephane, R. (eds.), *Space in languages: Linguistic systems and cognitive categories,* 59-81. Amsterdam: John Benjamins.

Sugiyama, Y. (2005). Not all verb-framed languages are created equal: The case of Japanese. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 31(1), 299-310.

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

ten Have, P. (1999). *Doing conversation analysis: A practical guide*. London: Sage.

Vygotsky, L. S. (1986). *Thought and language* (A. Kozulin, Trans.). Cambridge: The Massachusetts Institute of Technology Press.

Wilcox, S. (2004). Language from gesture. *Behavioral & Brain Sciences* 27(4), 525-526.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A. & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of International Conference on Language Resources and Evaluation 2006*, 1556-1559.

Kazuki Hata
School of Education, Communication and Language Sciences
King George Building VI
NE1 7RU
United Kingdom

k.hata@newcastle.ac.uk