

IMAGERY FOR SPEAKING

David McNeill

University of Chicago

INTRODUCTION

The Whorfian hypothesis has alternately attracted and annoyed linguists and psycholinguists for generations. The polar reactions tend to come in waves. We currently seem to be entering a phase of attraction, due in no small part to Dan Slobin's innovative extension of the Whorfian hypothesis to encompass thinking-for-speaking. The classic Whorfian hypothesis is fundamentally static. It presumes the synchronic view of language that has dominated linguistics ever since Saussure's famous *Course* (Saussure, 1966, original compiled posthumously by his students from lectures and published around 1915). As usually understood, the Whorfian hypothesis (Whorf, 1956) is the doctrine that holds that language influences 'habitual thought'—the very term a synchronic reference: thought abstracted from realtime dynamics to form a system of relationships viewed *in toto*, visible at a single theoretical instant. Lucy's (1992a, 1992b) elucidation of the Whorfian hypothesis confirms this crystalline structure, in the form of projected analogies between language and thought that by their nature are grasped synchronically. It is to Dan Slobin in his Berkeley Linguistics Society paper, "Thinking for Speaking" (Slobin, 1987), that we turn to get the first sight of a truly *dynamic* version of the Whorfian hypothesis—thinking generated, as Slobin says, because of the requirements of a linguistic code: "'Thinking for speaking' involves picking those characteristics that (a) fit some conceptualization of the event, and (b) are readily encodable in the language"¹ (p. 435). That languages differ in their thinking-for-speaking affordances is a version of the relativity hypothesis, now realized on the realtime dimension of speech and its unfolding.

My contribution to this approach is to bring in gestures. The imagery embodied in gestures also differs across languages. Duncan and I wrote on this theme in relation to thinking-for-speaking in McNeill and Duncan (2000). The current contribution is an updating of our joint paper, drawing on the growth point hypothesis we presented there and the considerable further development of the hypothesis in McNeill (2005). I will demonstrate gestures at work in thinking-

¹ The expression, 'thinking-*for*-speaking' suggests to some readers a temporal sequence: thinking first, speaking second. We posit instead an extended process of thinking-*while*-speaking, but keep the thinking-*for*-speaking formulation to maintain continuity with Slobin and his writings, and to capture the sense of an adaptive function also conveyed by *for*, with the caveat that we do not mean by this a thinking→speaking temporal sequence.

for-speaking in four languages. To provide a theoretical framework, I will first define “gesture”, and then present a theoretical dynamic model—the growth point or GP—to explicate the role performed by gestures in thinking-for-speaking. The overall approach is presented at length in McNeill (2005).

Duncan and I observed in our paper that a skeptical view of thinking-for-speaking could maintain that it operates only at the level of linguistic expression; there are indeed differences across languages in how data and experience are expressed, but to infer also from these differences in thinking risks circularity. To counter such a view, some way is needed to externalize cognition in addition to language. We thus considered speech and gesture jointly as an enhanced ‘window’ onto thinking and showed how the co-occurrences of speech and gesture in different languages enabled us to infer thinking-for-speaking in Slobin’s sense (McNeill and Duncan, 2000).

WHAT IS ‘GESTURE’?

Kendon (1980) distinguished five kinds of ‘gestures’. I subsequently arranged the distinctions along a continuum that I named the Gesture Continuum (McNeill, 1992; later elaborated into the Gesture Continua, McNeill, 2000). Here is the original Continuum:

Gesticulation → Speech-Linked → Pantomime → Emblems → Signs

The gestures with which we are concerned are the gesticulations. As one moves along the Continuum, two kinds of reciprocal changes occur. First, the degree to which speech is an obligatory accompaniment of gesture *decreases* from gesticulation to signs. Second, the degree to which gesture shows the properties of a language *increases* over the same span. Gesticulations are obligatorily accompanied by speech but have properties unlike language. Speech-linked gestures are also obligatorily performed with speech, but time with speech in a different manner—sequentially rather than concurrently, and in a specific linguistic slot (filling in for a missing complement of the verb for example). Pantomime or dumb show by definition is not accompanied by speech. Emblems such as the “OK” sign have independent status as symbolic forms. Signs in ASL and other sign languages are obligatorily *not* accompanied by speech, and the languages themselves have the essential properties of all languages. Clearly, therefore, speech and gesticulation (but not the other points along Kendon’s Continuum) combine properties that are unlike, and this combination occupies the same performance instant. A combination of unalikes at the same time is a framework for an imagery-language dialectic.

A DYNAMIC APPROACH

McNeill (2005) presents a dynamic conception of language as an imagery-language dialectic, in which gestures provide imagery. Thinking-for-speaking appears at several places in this dialectic, with imagery for speaking the first of these. Gesture is an integral component of language in this conception, not merely an accompaniment or ornament. Such gestures are synchronous and co-expressive with speech, not redundant, and are not signs, salutes, or emblems. They are frequent—about 90% of spoken utterances in narrative discourse are accompanied by them (Nobe, 2000). The synchrony of speech forms and gestures creates the conditions for an imagery-language dialectic. A dialectic implies:

- a) a conflict or opposition of some kind, and
- b) resolution of the conflict through further change or development.

The synchronous presence of unlike modes of cognition, imagery, and language, that are co-expressive of the same underlying thought unit, sets up an unstable confrontation of opposites. Even when the information content in speech and gesture is similar it is present in contrasting semiotic modes, and a dialectic occurs. This very instability fuels thinking-for-speaking as it seeks resolution. Instability is an essential feature of the dialectic, and is a key to the dynamic dimension. The concept of an imagery-language dialectic extends a concept initiated (without reference to gesture) by Vygotsky, in the 1930s (cf. Vygotsky, 1987):

“The relation of thought to word is not a thing but a process, a continual movement back and forth from thought to word and from word to thought. In that process, the relation of thought to word undergoes changes that themselves may be regarded as development in the functional sense. Thought is not merely expressed in words; it comes into existence through them.” (1987, p. 218)

This conception also recaptures an insight lost for almost a century, that language requires two simultaneous modes of thought—what Saussure, in recently discovered notes composed around 1910, termed the ‘double essence’ of language (although he too expressed this without reference to gestures; cf. Harris, 2002; Saussure, 2002).

Gesture is naturally opposed to linguistic form; they present the same underlying idea unit in two forms. At the point where speech and gesture are synchronous they are co-expressive. The idea unit ties them together, and explains the synchrony. The opposition between them is semiotic, different ways of packaging information, and exists even when the referential content of speech and gesture is the same. In gesture, information is embodied globally, as a whole, instantaneously, and concentrates in one symbol what may be distributed across

several surface elements of speech. Simultaneously, in speech, the same idea unit is represented analytically, combinatorically and linearly. In this semiotic opposition the idea unit exists at the same moment in two semiotically opposite forms, a contrast that fuels thought and speech, animating it in a dialectic.

The smallest unit of the imagery-language dialectic is posited to be a ‘growth point’ (hereafter GP), so named because it is theoretically the initial unit of thinking-for-speaking out of which a dynamic process of organization emerges. A GP combines imagery with linguistic categorial content, and the theory is that such a combination is unstable and thus initiates cognitive events. In the GP interactions between language and imagery occur in both directions, it is not that imagery is input to language or language to imagery; the effects are mutual, but in this paper the emphasis is on imagery and how it is affected by language, in keeping with the thinking-for-speaking focus.

A GP is an empirically recoverable idea unit, inferred from speech-gesture synchrony and co-expressiveness. An example recorded in an experiment (offered in part because of its ordinariness) is a description by a speaker of a classic Tweety and Sylvester escapade, which went in part as follows: “and Tweety Bird runs and gets a bowling ba[ll and drops it down the drainpipe].”² Speech was accompanied by a gesture in which the two hands thrust downward at chest level, the palms curved and angled inward and downward, as if curved over the top of a large spherical object (see Figure 1).³ At the left bracket, the hands started to move up from the speaker’s lap to prepare for the downward thrust.

² Notation for indicating gesture phase timing in relation to speech: [is the onset of the gesture phrase, when the hands move from rest or a previous gesture into position to perform the stroke;] is the end of the gesture phrase; **boldface** is the gesture stroke itself, the meaning-bearing phase of the gesture, performed with effort, and the only phase that is obligatory; **underlining** is a pre- or poststroke hold, a brief cessation of motion that tends to ensure the synchrony of stroke and targeted speech. Gesture phrases can occur inside other gesture phrases and this is marked by a double ‘[[’ and ‘]]’ (cf. line (2) of Ex. 1). **The preparation phase** is the interval between the onset of motion ‘[’ and the beginning of the stroke or prestroke hold; **the retraction phase** is that between the end of the stroke or poststroke hold and the end of motion ‘]’. In the speech transcript, a ‘/’ is a silent pause and a ‘*’ is a self-interruption. **The onset of preparation** is the first indication the idea unit in the stroke has come to life—in this example, with the word “ball” in the preceding clause. **A prestroke hold** suggests the linguistic material co-occurring with the stroke was targeted. **A poststroke hold** suggests the stroke and its speech are not merely co-occurring but are a single production. Finally, **the end of retraction** can be seen as the switching off of the idea unit.

³ Computer art in this and all following figures by Fey Parrill.



Fig. 1. Gesture stroke accompanying “it down” in the sentence “and drops it down the drainpipe”. From McNeill (2005). Used with permission.

Then her hands, at the very end of “drops,” paused briefly in the curved palm-down position, frozen in midair (the first underlining). Next was the gesture stroke—the downward thrust itself—timed exactly with “it down” (boldface). Movement proper ceased in the middle of “down,” the hands again freezing in midair until the word was finished (the second

underlining). Finally, the hands returned to rest (up to the right bracket). The two pauses or holds and the continuing preparation phase itself reveal that the downward thrust was targeted precisely at the “it down” fragment: the downward thrust was withheld until the speech fragment could begin and was maintained, despite a lack of movement, until the fragment was completed. Significantly, even though the gesture depicted downward thrusting, the stroke bypassed the very verb that describes this motion, “drops,” the preparation continuing right through it and holding at the end.

The fragment, “it down,” plus the image of a downward thrust, was the GP. It is impossible to fully understand the source of any GP without elaboration of its relationship to context. This relationship is mutually constitutive. A GP cannot exist without a context, because it is a point of differentiation within it; and the context is created, in part, to make the differentiation possible. While context reflects the physical, social, and linguistic environment, it is also a mental phenomenon, a representation; the speaker *constructs* it in order to make the intended contrast, the GP, meaningful within it. Theoretically, a growth point is a psychological predicate in Vygotsky’s (1987) sense, a significant contrast within a context (also Firbas, 1971).

A further concept, the catchment, provides an empirical route for finding this context. A catchment comprises gestures in a discourse stretch with recurring form features, and reveals the theme or field of oppositions from which the GP is differentiated. To identify the catchment in the “it down” case, we look for other gestures during the narration in which the hands are shaped and/or move similarly to the target gesture, and ask if these gestures comprise a family with thematic continuity. We find such a family; in the speaker’s rendition, all such two-handed gestures had to do with the bowling ball conceptualized as an antagonistic force, directed contra-Sylvester. The whole episode, of which the case study is a part, was construed by this speaker not merely as a cinematic episode but as a confrontation of antagonistic forces—Sylvester vs. Bowling Ball. We can thereby

further specify the “it down” GP: it was a psychological predicate specifying how the bowling ball *was* this antagonistic force:

Ways of Thwarting Sylvester: Bowling Ball Down

So the field of oppositions (Ways of Thwarting Sylvester) was differentiated as an image of a downward path, and categorized as “it” and “down”.

This analysis explains why the verb “drops” was *excluded* from the GP. The verb describes what Tweety did, not what the bowling ball did (it went down as the antagonistic force), and thus “drops” was not a significant contrast in the field of oppositions involving the bowling ball. The core idea at (2) was the bowling ball and its action, not Tweety and his. The origin of the verb in this case is explained by separate unpacking (see McNeill 2005).

GPS IN FOUR LANGUAGES

This dynamic approach can also be applied cross-linguistically. We observe that idea units are not independent of the language spoken, even when referential content is the same; given the same objective reality idea units can differ across languages. The languages to be described cover a range of types—English, Mandarin, Spanish, and the Deaf Sign Language of Taiwan (TSL).

METHODOLOGICAL PRELIMINARIES

Sources of Data

We have collected narrations of a Tweety & Sylvester cartoon stimulus in some 20 languages, with substantial collections in English, Spanish, and Mandarin. The English speakers were students at the University of Chicago. The Mandarin narrators were mostly students or spouses of students at the University of Chicago, many recent arrivals. The Spanish speakers were monolinguals recorded in Guadalajara, Mexico.⁴ The TSL narrations were recorded in Taipei by Susan Duncan. Both male and female speakers participated in all languages. The narrators viewed a 6-minute film (“Canary Row”) and retold it immediately from memory to a listener who had not seen it; narrator and listener were told that the listener would be asked to retell the story, a provision to encourage a full and clear description from the primary speaker. There was no mention of gesture, the emphasis was on storytelling. The resulting stories have coherence, including those by non-English speaking subjects. (The cartoon was selected in part because it makes limited use of speech and has highly repetitive a storyline with amusing

⁴ The recordings in Guadalajara were conducted by Lisa Miotto and Karl-Erik McCullough. The gesture motion events project was carried out with Susan Duncan.

surface variations.) Because we use a standard stimulus we are able to compare retellings across languages of the same episodes, thus holding referential content constant.

Coding

All narrations were transcribed, translated where necessary into morpheme-by-morpheme glosses by bilingual transcribers, plus idiomatic English, and coded for gestures, with the emphasis on the exact temporal location in relation to speech of the preparation, stroke, and retraction phases, plus any pre- and poststroke holds. Motion event content (both speech and gesture) was coded using Talmy's (2000) motion event semantic components—figure, path, manner and/or ground (gestures frequently combine several components).

COMPARISON ACROSS LANGUAGES

In the following analyses, languages are compared for a specific cartoon episode that involves the following drama: Sylvester the cat, pacing on a sidewalk, is attempting to reach Tweety, a canary, perched tantalizingly in a window high above. He decides to use a drainpipe running up the side of the building. The pipe conveniently ends just at Tweety's window. Sylvester tries this twice, each time with catastrophic results. His first effort is on the outside of the pipe, climbing it like a rope. He reaches Tweety but is battered off the windowsill by Tweety's fierce protector, Granny. On his second try (the case study episode), Sylvester climbs the pipe on the inside, hoping for concealment. Tweety nevertheless sees him, rushes off screen and returns with an enormous bowling ball, which he releases into the pipe. The ball and Sylvester meet explosively mid-pipe. He is next seen shooting out the bottom of the pipe, the bowling ball now inside him. A living bowling ball, he rolls (or is rolled) down a sloping street, legs spinning helplessly at his side, and disappears into a bowling alley. After an ominous pause, we hear tenpins being knocked over. This collection of motion events and how they are packaged comprises points of comparison across languages.

Crucial for comparing English and Spanish is Talmy's (2000) motion events typology, according to which these two languages (and many others) differ in how they package motion event semantic components. In **satellite-framed (or 'S-type') languages** (including English, German, Scandinavian languages, Chinese), *path* is encoded *outside* the verb, in a so-called satellite or preposition—cf. the different directions of walking in “walk in/out/across/through, etc.” *Manner* in contrast is encoded *within* the verb—cf. the different ways of getting across in “walk/run/stride/stagger/sidle, etc. across”. In contrast, in **verb-framed (or 'V-type') languages** (such as Spanish, French, Italian, Turkish, Japanese,

ASL), *path* is encoded *inside* the verb and *manner* is *outside* either in a new verb or gerund (cf. Sp. “sale volando” ‘exits flying’), or is omitted altogether. In what follows, we see how gestures differ in these two kinds of languages.

Gestural Paths Tend to be Straight-Line Segments in English and Unbroken Wholes in Spanish

S-type and V-type languages typically induce different imagery modes (cf. Özyürek, et al. 2005). Given the GP theory, this implies distinct thinking-for-speaking approaches. In S-type English, imagery of path or direction is broken into straight-line segments. In Spanish, the V-type, path is more often a single unbroken whole. The cross-language difference becomes clear when complex, curvilinear paths such as the bowling ball episode are compared—in the S-type, the path devolves into a series of short segments. The same path in the V-type is preserved in its full curvilinear complexity. GPs in English thus tend to focus on segments and how they relate, while those in Spanish focus on wholes. . Different contexts would tend to be constructed by speakers of the two languages to make these GPs possible—in English, where the segments of the path may have communicative dynamism, what Slobin (1996) has described as “elaborated trajectories of motion”, versus, in Spanish, contexts in which the path as a whole stands out, perhaps what he described as “elaborated descriptions of the static locations of objects” (p. 78).

In English.

The general rule seems to be that the gestures of English speakers convey path information synchronized with path satellites. The exceptions are that some path gestures align with ground/landmark elements. But overall, complex curvilinear paths break down into a series of more or less straight path segments (paths 1, 2, 3 and 5 in the following align with satellites, 4 and 6 with ground/landmark elements):

Example 1

- (1) [/ and it **goes down**]
- (2) but [[it **roll**][s him out*]]
- (3) [[down **the** / /]
- (4) [/ **rainspo**]
- (5) [ut/ **out**i][nto
- (6) **the sidew**ewalk/ into a] [bowling **alley**

The statements and their linked gestures are shown in Figure 2, and the match-up is perfect. Visuospatial cognition consisted of six straight-line segments. Such a division is expected from the kind of analytic path-satellite treatment directionality receives in S-type languages.



Fig. 2. English speaker's six path segments for Sylvester's trip down the pipe and into the bowling alley, accompanied by "and it goes down but it rolls him out down the rain spout out into the sidewalk into a bowling alley and he knocks over all the pins". Compare to Fig. 3. The hand performed two similar strokes in Path 6. From McNeill (2005). Used with permission.

In Spanish.

Spanish speakers, in contrast, often seem to represent the same scene without significant segmentation. Example 2 is a description by a monolingual Spanish speaker. In speech there was onomatopoeia, a frequent verb substitute in our Spanish-language productions:

Example 2

[**entonces SSS**]

then he-falls ONOM

then SSSS he falls

The accompanying gesture traced a single, unbroken path down and to the right (Figure 3—there are no pauses or interruptions). For the very same event, what had been segmented in English was a single curvaceous trajectory in Spanish.



Fig. 3. Spanish speaker's single continuous arc for Sylvester's trip down the pipe, accompanied by "entonces SSS" ('then SSS he falls'). Compare to Fig. 2. From McNeill (2005). Used with permission.

To quantify this cross-linguistic difference, Table 1 shows the number of path segments contained in gestures depicting path from Sylvester's encounter with the bowling ball to the denouement in the bowling alley, for both Spanish and English descriptions. All speakers, regardless of language, segment, but English speakers break the trajectory into 43 percent more segments than do Spanish speakers. On average, each English

speaker produced 3.3 segments, while each Spanish speaker produced 2.3 segments. Extremes of segmentation, moreover, strongly favor English. Five English speakers divided the trajectory into six or more segments, compared to only one Spanish speaker going so far. Thus Spanish speakers, even when they divide paths into segments, have fewer of them. And since they do not introduce lapses of gesture the segments are also broader, covering more speech (however, when speech describes a boundary or change of state—conditions that necessitate a new clause in Spanish—any accompanying path gesture stops and a new gesture or a gesture cessation ensues; English speakers, in contrast, at these same points, simply continue the path gesture, since in this language no new verb or clause accrues; McNeill & Duncan, 2000).

Table 1. Segmentation of paths by English- and Spanish-speaking adults

Number of gesture segments	Number of speakers	
	English (N=21)	Spanish (N=18)
0	0	1
1	3	5
2	7	6
3	3	4
4	2	1
5	1	0
≥ 6	5	1

Gestures Expand Manner in Spanish, Modulate Manner in English

Manner is the other diagnostic motion event component, along with path, separating the S-type and V-type languages. Manner imagery is accessible to speakers of both types of language, but enters into GP formation in different ways. Table 2 provides an overview of gestural manner with a specific verb, “rolls” in English, “rodar” in Spanish. Perhaps surprisingly, Spanish has a higher incidence of gestural manner with this verb.

Table 2. Gestures With Spoken References to Rolling

	Manner in Gesture	Non-Manner in Gesture
Spanish Adults (18)	80%	20%
English Adults (21)	43%	57%

In Spanish.

In Spanish speech, manner requires a second verb or gerund and is often omitted altogether (Slobin, 1996, 2004). Nonetheless, a manner gesture can combine imagery with other motion components in speech, typically the ground/landmark. The result is an expansion of the Spanish sentence to include manner without actually lexicalizing it. The effect seems to be that different kinds of surfaces, situations, etc. imply their own manners of action; we shall see an example of this constraint below.

One result of the expansion is a ‘manner fog’ — a blanketing of manner via gesture when there is no manner in speech. An example is shown in Example 3, a description of Sylvester climbing the inside of the pipe:

Example 3

(1) e entonces busca la ma[nera (**silent pause**)]

and so he looks for the way

*Gesture depicts the shape of the pipe: **ground**.*

(2) [de entra][r / / se met][e por el]

to enter REFL goes-into through the

*Both hands rock and rise simultaneously: **manner + path** combined (left hand only through “mete”)*

(3) [desague__] [/ / si?]

drainpipe...yes?

*Right hand circles in arc: **manner + ground** (shape of pipe).*

(4) [desague entra /]

drainpipe, enters

*Both hands briefly in palm-down position (clambering paws) and then rise with chop-like motion: **manner+ path** combined.*

Manner gestures appeared in the second, third, and fourth lines, despite a total absence from speech. Thus while manner may seem to be absent when speech alone is considered it may be present, even abundant, in visuospatial thinking. In this example manner is categorized as motion along a path and/or as a ground element (the pipe). In the ground (pipe) GPs, the speaker seems to have conceived of the pipe and its shape as constraining a certain *kind* of manner. The swirling gesture with “desague” in (3) is an illustration—circling around, a constraint on manner arising from the interior contour of the pipe⁵ (English speakers sometimes convey this as well, but use a verb for it, e.g., “barreling”).

In English.

In English, imagery modulates lexical manner, either emphasizing it or downplaying it. Such a role is correlated with the obligatory presence of manner in S-type verbs. A verb like “rolls” contains manner regardless of communicative dynamism. If manner is part of the GP manner is likely to appear in a gesture

⁵ Observation due to Sue Duncan.

synchronized with the verb; but if manner is not the point of differentiation the gesture can lack manner, emphasize path or some other motion event component, and may not synchronize with the verb at all (or there can be no gesture of course). Thus *gesture modulates manner*. Whereas a manner fog adds manner when it is lacking from speech, modulation adjusts the manner that is present in speech. The direction in which the modulation goes—enhancement or minimization—can be traced to the communicative weight given to manner in the context of speaking. The following examples, from different speakers, show enhancement and minimization, respectively, and how they correspond to different contextual weightings of manner.

The enhancement example was at the end of a series of references to the bowling ball, where its manner of motion would plausibly have been highlighted (cf. Parrill, 2008). The gesture contains manner and synchronizes with the manner verb, “rolls.” This content and co-occurrence with the gesture highlight manner and suggest that it was part of the psychological predicate, as shown in Example 4.

Example 4 (enhancement)

and he drops a [bowl]ing ball [**into** the rain spout]

[and it **goes down**]

and it* [/] ah*

you [**can't tell if the bowling ball** /]

[is un* /] [is **under** Sylvester

or inside of him]

[but it **rolls him out**]*

(= *gesture with manner: Both hands sweep to right and rotate as they go, conveying both path and manner*)

In the minimization example, despite the same verb, “rolls,” the gesture skips the verb and has no manner content of its own. It shows path, and co-occurs with the path satellite, “down.” Both the timing and the shape of the gesture thus suggest that manner was not a major element of the speaker’s intent, and that “rolls,” while referentially appropriate, was de-emphasized and functioned as a verb referring to the fact of motion with manner content downplayed (the speaker could just as well have said “goes down”, avoiding lexical manner, but this would have meant editing out what the speaker also knew—Sylvester was rolling; the manner component in “rolls” referred without adding to communicative dynamism). This situation is shown in Example 5:

Example 5 (minimization)

[the canary] # [throws*] #
[puts a # [bowling] [ball] #
into] # [the **drain** spout as the]
[cat is climbing up /and]
[it goes into his] [mouth] / (topic switch to Sylvester)
[and of course] #
[into his stomach] #
[and he rolls # **down** the drain spout] (= gesture with path but no
manner: Left hand plunges straight down, in synchrony with the satellite)
[and [across] [the street] into [the bowling] alley #]

English, Compared to Mandarin, is Enslaved to Predication

Mandarin and English are each languages of the S-type but differ in many other ways. A less-than-obvious difference emerges from the gesture data. English, more than Mandarin, seems committed to predicates as the loci of gestures.

Example 6 (English – committed)

[so it hits him on the head][d and he winds up rolling down the street]

The speaker performed a rolling down gesture as she was saying “hits him on the head”. Gesture and speech comprised a sensible combination, the gesture showing the consequence of the action that speech described. However, the hand did not return to rest but held in place waiting (in fact, twice) for the predicate; then the gesture repeated on a larger scale. The larger gesture had the character of a repair, was exaggerated, as if to correct the ‘misplaced’ previous version.

Mandarin, perhaps because it has alternate construction strategies such as topic and comment, seems less wedded to predication (also, on the English side, word order inflexibility could promote tighter linkages of gestures to predicates). We find Mandarin examples in which a gesture depicting an action co-occurs with a noun phrase referring to the instrument of that action, not the verb phrase identifying the action itself. An action-instrument combination also makes sense; in fact, much the same kind of sense as the English speaker’s cause-effect combination. The predicate, however, when it comes, is free of gesture. It is as though the predicate—far from a repair—is felt to be redundant, a repetition of

something already conveyed, and is included only to meet standards of well-formedness. An example is the following:⁶

Example 7 (Mandarin – free)

lao tai-tai [na -ge da **bang hao**]-xiang gei ta da-xia
old lady hold CLASSIFIER big stick seem CAUSE him hit down_{verb-satellite}

‘The old lady seemed to have knocked him down with a big stick’

Left hand in grip moves downward sharply from head to waist

Despite the spacing, there were no pauses or hesitations. The speaker, as she said “da bang” (a big stick), performed a gesture that seemingly held the stick while executing a downward blow. Her hand then went to rest and remained there as she continued on to the predicate, the meaning of which was close to that of the gesture. There were no further gestures. Thus unlike English, a gesture depicting an action with speech not part of the predicate was not repaired; to the contrary, the predicate was treated like a repetition.

Thus the languages show opposite inclinations toward predicates, revealed in the treatment of gestures. In terms of GPs, the English speaker’s initial gesture-speech combination, though sensible, was not a successful growth point, because it attempted to swallow what for her belonged to the predicate.⁷ Such was not a barrier for the Mandarin speaker, who in fact released the predicate from service once the instrument-action combination had been created with the gesture. Imagery is categorized differently—in English it is tied to predication; in Mandarin it may seek other partners.

GPs With Manner in Sign Languages

What of sign languages? It goes without saying that a sign language uses imagery for speaking in one sense of image; however this is a regularized kind of imagery. Sign languages conventionalize space and motion (utilizing space and motion as the linguistic medium). Conventionalization alters the semiotic quality of signs. A sign is analytic, segmented, and combinatoric—the semiotic of the language side. But it would be odd if sign languages did not have GPs. So what would the gestural side of a signed GP look like? Duncan (2005) proposes that *sign deformation* signals a gesture occurrence. She has recorded this phenomenon in Taiwan sign language (TSL) at points of newsworthy content, the same points

⁶ Example, transcription, and translation due to Sue Duncan. Though we don’t have numerical data, Duncan has often noticed such combinations in daily Mandarin speech.

⁷ Gestures in English do occur outside predicates, but as Exs. 6 and 7 suggest the gravitational pull of the predicate is greater in English.

where speaking narrators also perform gestures. The deformations modify the sign to highlight what is significant the immediate discourse context. Thus we can examine TSL gestures in signing to see if, here too, manner imagery is adapted to linguistic form in the environment of a sign. Figure 4 (left panel) illustrates a sign using a classifier in TSL for animate beings with more than two legs (animals, crawling babies). The standard form of the classifier is the thumb, the index finger, and the middle finger extended and spread apart, the other fingers curled in, with the palm down. However, when describing Sylvester going up



Fig. 4. Possible gesture (left panel) with the TSL sign for 'animate being with more than two legs', deformed for iconicity at the same discourse junctures where gestures also occur by hearing speakers (recorded by S. Duncan). Compare to English speaker's gesture for the same event (right panel). Both illustrations from McNeill (2005) are used with permission.

the pipe on the outside, the signer modified the classifier by having the first and second fingers, instead of extending outward motionlessly, alternately 'walk up' in space (also the left hand, at leg level, was curved, apparently to show the pipe, but this location was not a locus for the 'walking'). The thumb was extended, which indicates that we see here the sign modified. What is key is that also the fourth finger was extended, an extension that is not part of the canonical form but was perhaps added to be the *third leg* as specified in the canonical sign (unmoving but extended).⁸ If so, imagery for speaking (signing) was modified to meet the sign's conventional form. The GP was something like the image of climbing, linguistically categorized as having been done by a creature with more than two legs. The right panel of Fig. 4 shows a hearing English speaker's version of the same cartoon event. Here, the gesture emphasizes interiority and ascent, just as did the TSL gesture/sign, but lacks protuberances for legs (the upright fingers creating a bowl shape) and in fact does not depict clambering at all. So perhaps the emphasis in TSL, but not in English, on climbing was another case of thinking for speaking adapting to the affordances of the language, due to linguistic features of the sign finding special significance in images of clambering feet (*presumed* feet, since they were not in fact visible in the stimulus).

⁸ This interpretation supercedes the one offered, with puzzlement, in McNeill (2005).

GENERAL SUMMARY

The different uses of gesture in these four languages provide evidence that visuospatial cognition during thinking-for-speaking differs systematically across languages. Four major ways in which gestures co-occur with speech have been observed:

(a) Gestural paths tend to be broken into straight-line segments in English and remain unbroken curvilinear wholes in Spanish.

(b) Gestural manner tends to expand the encoding resources of Spanish and to modulate them in English, following the packaging of manner information in these languages—unavoidable manner in English, all-too-avoidable in Spanish.

(c) Gestures reveal a ‘tyranny of predication’ in English GPs that is lacking or minimal in Mandarin.

(d) Gestures combine with signs through distortions to provide imagery, and create signed GPs.

CONCLUSION: A STRONG/WEAK PARADOX

We began this paper with the Whorfian hypothesis. Now it is time to expose a paradox in the ‘strong’ and ‘weak’ versions of this hypothesis. Depending on what we mean by ‘strength’, the weak version is stronger. I would prefer not to use the muscular analogy at all: it misses the real distinction, which is that the ‘strong’ version is *static*, the ‘weak’ version, in its thinking-for-speaking form, is *dynamic*. The strong version, so-called, refers to habitual thought, and proposes that the effects of language on cognitive dispositions do not require the production of speech to activate them—‘strong’ in this sense means that the influence of language does not depend on the unfolding of a process, namely speech. Thinking for (and while) speaking, on the other hand, occurs during acts of speech only, hence is ‘weaker’.

Apart from the time-honored but dubious distinction between ‘language’ and ‘speech’, there is another way to think about this comparison, and in this alternative the weak outpowers the strong. The question is, on which dimension of language is the influence on thought stronger? And here the dynamic dimension is clearly the place: the impact of language on thought is readily demonstrated in the dynamic unfolding of language where thinking-for-speaking takes place. To judge from the years of controversy and waves of endorsement followed by rejection, the impact of language on thought on the static dimension is hard to discern: this is the paradox, but also the remedy. Slobin’s thinking-for-speaking hypothesis is

the breakthrough that reveals the dynamic dimension of the Whorfian hypothesis, where an impact of language on thought is readily observed.

REFERENCES

- Duncan, S. (2005). Gesture in signing: A case study from Taiwan Sign Language. *Language and Linguistics*, 6, 279-318
- Firbas, J. (1971). On the concept of communicative dynamism in the theory of functional sentence perspective. *Philologica Pragensia*, 8, 135-144.
- Harris, R. (2002). Why words really do not stay still. *Times Literary Supplement*, July 26 issue, p. 30.
- Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207-227). The Hague: Mouton and Co.
- Lucy, J. A. (1992a). *Language diversity and thought: A reformulation of the Linguistic Relativity Hypothesis*. Cambridge: Cambridge University Press.
- Lucy, J. A. (1992b). *Grammatical categories and cognition: A case study of the Linguistic Relativity Hypothesis*. Cambridge: Cambridge University Press.
- McNeill, D., & Duncan, S. D. (2000). Growth points in thinking-for-speaking. In D. McNeill (Ed.), *Language and gesture* (pp. 141-161). Cambridge: Cambridge University Press.
- McNeill, D. (1992). *Hand and mind*. Chicago: University of Chicago Press.
- McNeill, D. (2000). Catchments and contexts: Non-modular factors in speech and gesture production. In D. McNeill (Ed.), *Language and gesture* (pp. 312-328). Cambridge: Cambridge University Press.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Nobe, S. (2000). Where do most spontaneous representational gestures actually occur with respect to speech? In D. McNeill (Ed.), *Language and gesture* (pp. 186-198). Cambridge: Cambridge University Press.
- Özyürek, A.; Kita, S.; Allen, S. E.M.; Furman, R.; Brown, A. (2005). How does linguistic framing of events influence co-speech gestures?: Insights from crosslinguistic variations and similarities. *Gesture* 5, 219-240.
- Parrill, F. (2008). Subjects in the hands of speakers: An experimental study of syntactic subject and speech-gesture integration. *Cognitive Linguistics*, 19, 283-299.

- Saussure, F. de (1966). *Course in general linguistics* (C. Bally & A. Sechehaye, eds., in collaboration with A. Riedlinger; W. Baskin, trans.). New York: McGraw-Hill.
- Saussure, F. de (2002). *Écrits de linguistique générale* (compiled and edited by S. Bouquet & R. Engler). Paris: Gallimard.
- Slobin, D. I. (1987). Thinking for speaking. In J. Aske, N. Beery, L. Michaelis, & H. Filip (Eds.), *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society* (pp. 435-445). Berkeley: Berkeley Linguistic Society.
- Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge: Cambridge University Press.
- Slobin, D. I. (2004). The many ways to search for a frog: Linguistic typology and the expression of motion events. In S. Strömquist & L. Verhoeven (Eds.), *Relating events in narrative, Vol. 2: Typological and contextual perspectives* (pp. 219-257). Mahwah, NJ: Lawrence Erlbaum Associates.
- Talmy, L. (2000). *Toward a cognitive semantics*. Cambridge, MA: The MIT Press.
- Vygotsky, L. S. (1987). *Thought and language*. (Edited and translated by E. Hanfmann & G. Vakar; revised and edited by A. Kozulin). Cambridge: MIT Press.