The Whorfian hypothesis has alternately attracted and repelled 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” (p. 435).¹ 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 it that has taken place since then. I will demonstrate gestures at work in thinking-for-speaking in four languages. To provide a theoretical framework, I will first define “gesture”, and then present a dynamic theoretical 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 Kendon’s Continuum (McNeill, 1992; later elaborated into Kendon’s 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 unalike, 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 new conception of language: 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 new 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 expressed this without reference to gestures; cf. Harris, 2002; Saussure, 2002).

Gesture is naturally opposed to linguistic form. At the point where speech and gesture are synchronous they are co-expressive; they present the same underlying idea unit. This idea ties them together, and causes the synchrony. The opposition between them is semiotic, different ways of packaging information, and exists even when (on rare occasions) 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. Synchronously, in speech, the idea unit is represented analytically, combinatorically and linearly. In this semiotic opposition, the idea exists at the same moment in two opposite forms, and the contrast fuels thought and speech, animating it, in a dialectic.²

The smallest unit of the imagery-language dialectic is posited to be a ‘growth point’ (hereafter abbreviated as 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 initiates cognitive events. In the GP model, 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].”^3

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).^4 At the left bracket, the hands started to move up from the speaker’s lap to prepare for the downward thrust.

![Fig. 1. Gesture stroke accompanying “it down” in the sentence “and drops it down the drainpipe”](image)


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 GP was the fragment, “it down,” plus the image of a downward thrust. Both sides of the GP are essential. They are opposed dialectically in that the linguistic components have combinatoric potential and categorize the image; the imagery component embodies these categories as an instantaneous whole. These different modes are simultaneously active (for the speaker and the listener). That one idea exists in two different modes is the motive force for the dialectic and its consequent unfolding.

The model also must explain how the dialectic is resolved by unpacking a GP into a stable outcome, a full linguistic form, in such a way that the form presents the GP; how this may require further meaning generation with its own context; how emerging linguistic intuitions provide ‘stop orders’; and the incorporation of the context of speaking into the GP itself. These aspects of the GP are explained at length in McNeill (2005). In this paper the focus is on imagery and the dialectic, so that we can see how language differences foster dialectic differences, and the rest of the theory can be set aside. However, it is impossible to understand the GP fully 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; the speaker constructs this context 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 or, as I will say, a field of oppositions. The GP, including its imagery, is a point of high communicative dynamism at the moment of speaking (Firbas, 1971 initiated the concept of communicative dynamism)—newsworthy information within a field of oppositions. The underlying assumption is that verbal meaning is non-associative, it is a product of differentiation.

A further concept, the catchment, provides an empirical route for finding this field of oppositions. 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. As noted earlier, exclusion is evidenced in the fact that the stroke did not synchronize with the verb; in fact, it was withheld from the verb by continued preparation and the possible brief prestroke hold. The verb however 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 the later unpacking of the GP into a causative construction (see McNeill 2005).

GPS IN FOUR LANGUAGES

Now this dynamic approach can be applied cross-linguistically. We observe that core thinking-for-speaking idea units are not independent of the language spoken, even when referential content is the same; idea units, given the same objective reality, differ across languages. The languages to be observed 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. 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 and this includes the retellings by non-English speaking subjects. (The cartoon was selected in part because it makes limited use of speech and has a storyline that is highly repetitive with amusing surface variations.) Because we use a standard
stimulus we are able to make comparisons in narrations across languages at specific points of the same episodes, thus holding constant referential content.

Coding

All narrations were transcribed, translated where necessary by bilingual transcribers into morpheme-by-morpheme glosses, plus idiomatic English, and then coded for gestures with the emphasis on the exact temporal location of the preparation, stroke, and retraction phases plus any pre- and poststroke holds in relation to speech. 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, tantalizingly perched in a window high above. He decides to use a drainpipe running up the side of the building. The pipe ends conveniently just at Tweety’s window. Sylvester tries this twice, with catastrophic results each time. 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 fearsome 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 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 between languages.

Important in the comparison of English and Spanish is Talmy’s (2000) motion events typology, according to which the languages (and many others) differ systematically 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 verb particle, or a 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 \textit{inside} the verb and \textit{manner} is \textit{outside} it in a new verb or gerund (cf. Sp. “sale volando” ‘exits flying’), or is omitted altogether. In what follows, we see how gestures can differ in these two types of languages.

Gestural Paths Tend to be in Segments in English, Wholes in Spanish

S-type and V-type languages induce different imagery modes. Given the GP theory, this implies distinct thinking-for-speaking approaches. In S-type English, imagery of path or direction is broken into one or more 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 the whole.

In English.

Mostly the path gestures of English speakers synchronize with path satellites, with a few others linking to ground/landmark elements. In all cases, complex curvilinear paths tend to be broken 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):

\textbf{Example 1}

(1) [/ and it \textbf{goes down}]
(2) but [[it \textbf{roll}][\textbf{s him out*}]]
(3) [[\textbf{down the} /\!\!\!/]]
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.
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, as is expected from the kind of analytic path-satellite treatment directionality receives in S-type languages.

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 SSS 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 one curvaceous trajectory in Spanish.7

To quantify this cross-linguistic difference, Table 1 shows the number of path segments contained in gestures depicting the path running 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).
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.

<table>
<thead>
<tr>
<th>Number of gesture segments</th>
<th>Number of speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English (N=21)</td>
</tr>
<tr>
<td></td>
<td>Spanish (N=18)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>≥6</td>
<td>5</td>
</tr>
</tbody>
</table>

Gestural Manner Expands Spanish, Modulates 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>
<thead>
<tr>
<th>Gesture</th>
<th>Manner in Gesture</th>
<th>Non-Manner in Gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish Adults (18)</td>
<td>80%</td>
<td>20%</td>
</tr>
<tr>
<td>English Adults (21)</td>
<td>43%</td>
<td>57%</td>
</tr>
</tbody>
</table>

In Spanish.

The greater tendency of Spanish to include manner in gesture extends to cases where there is no mention of manner at all in speech, creating 'manner fogs'—a blanketing of manner via gestures alone. An example is shown in Example 3, a description of Sylvester climbing the inside of the pipe:

Example 3

(1) e entonces busca la manera (silent pause)8
    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 REFLE 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 / /]
    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 pipe9 (English speakers sometimes convey this as well, but have a verb for it, “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 English verbs. The verb contains manner regardless of its communicative dynamism manner in context, but if manner is part of the GP it also may appear in a gesture synchronized with the verb; but if manner is not the point of the contextual differentiation, a gesture without manner can emphasize path or some other motion event component and may not synchronize with the verb at all (or there could be no gesture at all, of course). Thus gesture modulates manner in English. Whereas a manner fog adds manner when it is lacking from speech, modulation adjusts manner present in speech to fit the point of focus. 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 reinforcement example was at the end of a series of references to the bowling ball, where its manner of motion would plausibly have been highlighted. 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 [bowling 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 the 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 was rolling; the manner
component in “rolls” has a referential function 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 and synchronized with path satellite)
[and [across] [the street] into [the bowling] alley # ]
```

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)**

```
so it hits him on the head 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 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, less word order flexibility 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 purely syntactic standards of well-formedness. An example is the following.10
Example 7 (Mandarin)

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’

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 that appear outside them. 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. Such was not a barrier for the Mandarin speaker, who in fact released the predicate from further service once the instrument-action combination with gesture had been created. Imagery is categorized differently—in English it is tied to predication; in Mandarin it can seek other partners.

GPs 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 it is a regularized kind of image. Sign languages utilize space and motion as the linguistic medium. This conventionalization alters the semiotic quality of manual mode from that of gestures.

Sign language imagery is analytic, segmented and combinatoric—the semiotic of the language side (cf. Note 2)—and would seem to leave little room for an imagery-language dialectic. But it would peculiar if sign languages alone do not have GPs. So what would the gestural side of a sign GP look like? Klima & Bellugi (1979) pointed out the ubiquity of mimetic gestures in regularized sign production, but did not address the problem of how gestures combine with signs proper. Liddell (2003) has presented a new conception of ASL that places this merging at the heart of the language. Liddell cites pronouns and directional verbs (“give”, “see”, “talk to”, and the like) as the prime cases of the gesture-sign interface (incidentally connecting the two ends of Kendon’s Continuum). These ASL forms have implicit slots into which deictic-iconic gestures must go, as a requirement of form. The slots are part of the grammatical structure, but the gestures that enter them are graded, non-morphemic, and unlistable. Thus a non-
first-person pronoun has, as part of its structure, a deictic-iconic gesture indicating the locus and other properties (such as size) of the referent. A pronoun referring to a very tall individual is different gesturally from one referring to a child, but in either case it indicates the location of the referent (or, more precisely, the targeted part of the referent, the eyes). Suppala (2003) has defended the opposite view, that these gestures, so-called, are in fact not gestural but are morphemic in ASL. It is difficult for an outsider to adjudicate this dispute, but the Liddell hypothesis is attractive as a way to explain (rather than to define away) the co-presence of signing and gesturing.

But gesture in sign language production may be more widespread. Other gestures, not just those with grammatically determined slots in pronouns and directional verbs, may also co-occur with ASL signs. This is what the GP hypothesis predicts. If so, one sign can be both spontaneous and conventional, and create within itself the potential for a dialectic opposition. Duncan (2005) proposes that a signal of a sign language gesture is deformation of the sign in a specific way so as to incorporate discourse significant information of the kind that psychological predicates incorporate. She has recorded this phenomenon in Taiwan sign language (TSL). Sign distortions appear at points of newsworthy content, the same points where speaking narrators also perform gestures. The deformations modify the sign to highlight what is significant the immediate discourse context. Figure 4 illustrates signs using a classifier in TSL for animate beings with more than two legs (animals and 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. However, when describing Sylvester going up the pipe on the outside, the signer modified the classifier by having the first and second fingers, instead of extending outward motionlessly, ‘walk up’ in space (the left hand was also curved, apparently to show the pipe, but it did not provide the surface of the ‘walking’). The thumb was also extended, which reinforces the conclusion that what we see here is a modified sign (extension of the thumb is part of the canonical sign form). Also note that the fourth finger was extended, which was not part of the canonical form, but was perhaps added to make the third leg. If so, the gesture was modified by the sign’s conventional content as well as its form, thus showing an impact of language on gesture as well as of gesture on language. The GP was something like the image of climbing, linguistically categorized as being done by a creature-with-more-than-two-legs. This was a combination significant in a field of oppositions:

Ways of an Animate Creature With 3 or 4 Legs Getting Tweety: 
Climb the Pipe.
Describing Sylvester’s second attempt, on the inside (Figure 4 right panel), the signer used the same sign but reduced the canonical shape to just the extended first finger, which he pushed under his left hand (similar to the use of a single finger placed in the envelope of the other hand in ASL and in some spontaneous gestures of hearing speakers in our data). The GP was now the point of differentiation in a new context:

Ways of Getting Tweety Using the Pipe: On the Inside

The modified sign highlighted interiority, the significant point of departure, which was categorized again as involving a creature with more than two legs.

These deformations occurred at just the discourse junctures where hearing speakers also perform gestures, and plausibly are actual gestures in TSL co-occurring, in the deformations, with conventional signs on the same hand. A semiotic contrast of linguistic form and imagery is thus present: signs are conventional, combinatoric, and repeatable; gestures the opposite on each front. There is the potential for tension of the two cognitive modes, and thinking-for-speaking in the form of GPs could occur (not to minimize the impact that having imagery on both sides of the dialectic can have). Not enough is known of the structure of TSL to infer cross-linguistic differences (or similarities) with, say, spoken English, Spanish or Mandarin or for that matter signed ASL, but it seems special to TSL that the image of Sylvester in the first ascent emphasized a third leg.

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 both in a regulated way, as in pronoun directionality, and through distortions of signs to include imagery in 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 the 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 only during acts of speech, hence is ‘weaker’. Apart from the 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 the impact of language on thought is readily observed.
REFERENCES


NOTES

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.

2 The terms ‘global’, ‘synthetic’, ‘combinatoric’ and ‘co-expressive’ have technical uses in GP theory. Global means that the parts of the gesture gain meaning from the meaning of the whole. They do not have independent meanings, as do linguistic signs. Synthetic means that one gesture gathers meanings that, in speech, are spread over the surface sentence. Combinatoric means that a symbol has built into it the potential to combine with other symbols; in Saussurian terms (1966) it has syntagmatic value. Gestures of the gesticulation type lack this property, whereas it resides in a
linguistic sign intrinsically. Even a single word possesses it (consider *down*—though alone, it has the potential to combine with verbs as a satellite, does not have the potential to combine as a subject, etc.). Adding a second hand to a gesture is a kind of combination, but not of the linguistic type. It builds a more elaborate image (e.g., one hand rising up, the second forming an open space around it for the inside of a drainpipe) but not a syntagmatic value. So, in total, the semiotic opposition in a dialectic is between packages of meaning in gestures that are global, synthetic and non-combinatoric vs. in speech that are analytic, segmented and combinatoric. These modes are *simultaneously active* where they are *co-expressive*—the fourth technical term, which means that synchronous speech and gesture embody one idea unit (possibly different aspects of it). So a single idea exists in two opposite modes that may capture different sides of it.

3 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; boldface is the gesture stroke itself, the meaning-bearing phase of the gesture, performed with effort, and the only phase that is obligatory; ] is the end of the gesture phrase; 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 "[]" (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.

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

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

6 English speakers of course also perform curved path gestures, for example with “around” and verbs such as “to swing”. Still, English speakers break down complex, multi-turn curvilinear trajectories into a series of small mostly straight segments, while Spanish speakers favor single curvilinear gestures. Incidentally, speakers of Japanese tend not to have curved gestures for a cartoon event in which Sylvester swings on rope, while English speakers, for the same event, have a curved gesture. The reason for the Japanese straight-line gesture has nothing to do with the V-type status of the language but is because Japanese lacks an intransitive verb that means to swing—there is only a transitive verb, “to swing something”. When this accidental gap applies, as in the rope scene, speakers tend to perform straight-line gestures (Kita & Özyürek, 2003). This phenomenon also demonstrates that visuospatial cognition in gesture and with it the realization of an imagery-language dialectic is not independent of the language being spoken.

7 It is interesting that the cartoon, designed and conceptualized by animators from an S-type linguistic culture (viz. Hollywood), has a satellite-framed-like division of scene shifts, new camera angles, fadeouts, etc. The path is never shown as a whole. Such path segmentation makes this unbroken V-type representation all the more remarkable.

8 Kendon (personal communication) observes that when a gesture occurs during a speech pause, listeners tend to look at it, and this may be a way of calling attention to the gesture. However, such a device does not appear to have been in use here (speech amplitude during the pause trailed off, for ample, rather than ceasing abruptly, suggesting not deployment of a conscious device but rather an involuntary delay in forming the upcoming idea unit).

9 Observation due to Sue Duncan.
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.

Some risk of overstatement in speaking this way. Gestures in English do occur outside predicates, but as Exs. 6 and 7 suggest the gravitational pull of the predicate is greater in English.

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