Gesture, verb aspect, and the nature of iconic imagery in natural discourse

Susan D. Duncan
University of Chicago

Linguistic analyses of Mandarin Chinese and English have detailed the differences between the two languages in terms of the devices each makes available for expressing distinctions in the temporal contouring of events — verb aspect and *Aktionsart*. In this study, adult native speakers of each language were shown a cartoon, a movie, or a series of short action sequences and then videotaped talking about what they had seen. Comparisons revealed systematic within-language covariation of choice of aspect and/or *Aktionsart* in speech with features of co-occurring iconic gestures. In both languages, the gestures that speakers produced in imperfective aspect-marked speech contexts were more likely to take longer to produce and were more complex than those in perfective aspect speech contexts. Further, imperfective-progressive aspect-marked spoken utterances regularly accompanied iconic gestures in which the speaker's hands engaged in some kind of temporally-extended, repeating, or 'agitated' movements. Gestures sometimes incorporated this type of motion even when there was nothing corresponding to it in the visual stimulus; for example, when speakers described events of stasis. These facts suggest that such gestural agitation may derive from an abstract level of representation, perhaps linked to aspectual view itself. No significant between-language differences in aspect- or *Aktionsart*-related gesturing were observed.

We conclude that gestural representations of witnessed events, when performed in conjunction with speech, are not simply derived from visual images, stored as perceived in the stimulus, and transposed as faithfully as possible to the hands and body of the speaker (cf. Hadar & Butterworth, 1997). Rather, such gestures are part of a linguistic-conceptual representation (McNeill & Duncan, 2000) in which verb aspect has a role. We further conclude that the noted differences between the systems for marking aspectual distinctions in spoken Mandarin and English are at a level of patterning that has little or no influence on speech-co-occurring imagistic thinking.

Gesture 2:2 (2002), **183–206.**
ISSN 1568–1475 / E-ISSN 1569–9773

Keywords: gesture, aspect, *Aktionsart*, Mandarin Chinese, iconicity, imagery in language

Introduction and overview

In unrehearsed discourse, speakers typically gesticulate as they speak. This appears to be universal across languages and cultures in face-to-face interaction. Following Kendon (1972, 1980, 2000) and McNeill (1985, 1992), such gesturing is regarded as being fully a part of language along with speech. Here we assume that detailed analyses of the forms and semantic contents of synchronized phases of speech and gesture production reveal the categorial-linguistic and analog-imagistic starting essentials of utterances (McNeill & Duncan, 2000). By combining analysis of gestures with analysis of spoken forms we obtain an 'enhanced window' on the linguistic-conceptual representations involved in language production.

We report observations of correspondences between co-occurring spoken and gestured forms during natural language production. We focus on a particular linguistic-conceptual domain: the temporal contouring of events. This is expressed in speech via distinctions in verb aspect and *Aktionsart*. The language samples consisted of unrehearsed narrative discourses collected in response to audio-videotaped eliciting stimuli. Speakers were unconstrained in their narrations other than with respect to the event content to be described. The latter was held constant across speakers in order that we could compare speakers' gestures when they made different choices among categorially-contrastive linguistic aspects, in the process of describing the same visually-presented images and events.

All languages possess the means by which to express the locations of events in time. They also all have the means to express the different ways in which events can play out in time. Our comparison of Mandarin Chinese and English takes its cue from various linguistic analyses that highlight the differences between these two languages in terms of the means each language provides for the expression of temporality; that is, the locations and structure of events in time. According to Binnick (1991: 446), while Mandarin is a, "classic tenseless language," it marks a range of aspectual distinctions both grammatically and periphrastically. In lieu of morphological tense marking, Mandarin locates events in time relative to one another by means of adverbials and explicit temporal references. English has tense-marking morphology, but is impoverished

in aspect-marking morphology, compared to Mandarin. The expression of the aspectual view of an event in spoken English may emerge more via lexically- and periphrastically-encoded *Aktionsart* than is true in Mandarin, which has a set of grammatical morphemes marking the major aspectual distinctions, as well as morphemes for marking some *Aktionsart* distinctions.

Our study focuses on three linguistic aspect distinctions that may be unambiguously marked in Mandarin. Each of these aspectual views may be signaled by a single-syllable morpheme: imperfective-progressive 'zai,' imperfective-durative 'zhe,' and perfective 'le.' The gestures that our sample of Mandarin speakers produced in speech contexts containing these linguistic markers were examined and compared with gestures observed in the English speakers, where similar meanings were being expressed. These particular three verb aspects were chosen because, in the story-telling and short-descriptive discourses examined here, they occurred with frequencies sufficient to accumulate reasonably large data sets for analysis and comparison. They were not the only aspectual distinctions expressed by the speakers.

A comparison of linguistic aspects and *Aktionsarten*

One widely accepted way of thinking about the linguistic category of aspect is that it concerns a speaker's view of an event in time (Comrie, 1981). According to Comrie, "aspects are different ways of viewing the internal temporal constituency of a situation," (1981:3). Many linguists hold that the most fundamental aspectual distinction exhibited across all languages is that between perfective and imperfective aspects (Binnick, 1991). For the purposes of this study, the simple framework for aspect and *Aktionsart* that is depicted in Table 1 was assumed. Perfective aspect is said to express an external viewpoint; imperfective aspect, an internal viewpoint that takes some account of an event's extension in time.

In spoken expression, a speaker's choice of aspect represents a decision either to expand on the, "internal temporal constituency of an event," that is, to open a view onto its internal workings, or, alternatively, to view the event from the exterior as a simple whole. In the present study, a further distinction was recognized in relation to the imperfective, or internal, aspectual view. That is the distinction between the imperfective progressive and imperfective durative aspects. Although the latter terms for imperfective aspects are sometimes used interchangeably, this analysis assumed a distinction between them, illustrated in Table 1, with examples taken from Binnick (1991). According to this

Table 1. Aspect and *Aktionsart* distinctions.

Aspect	
Perfective	external view of an event in time
Imperfective	internal view of an event
— progressive	a dimensionless point within an ongoing action: "The comet is coming."
— durative	the temporal extent of an ongoing action: "The comet comes ever nearer."
Aktionsart	
Punctate	hit, snap, pop, shoot, catch
Durative	meander, deteriorate, slide, ooze, wring
Iterative	crackle, chatter, wag, pepper, pummel
Completive	fill up, land, reach, oust, break

framework, imperfective progressive aspect is conceived of as momentary, a single instant within an extended temporal interval, whereas imperfective durative addresses the temporal extent of an event.

Distinguishable from, though related to, aspectual view as it is used here, is the notion of *Aktionsart*. As opposed to the view-of-event as expressed in choice of aspect-encoding linguistic morphemes applied to the verb, *Aktionsart* is a feature of event type itself, as encoded in the verb or periphrastically. According to Klein (1994) *Aktionsart* refers to the, "inherent temporal features of the lexical content of verbs, and more complex constructions." Particular verbs and phrases of every language capture event types with certain inherent temporal features (Vendler, 1967). For instance, the English verb 'hit' expresses a punctate event structure. There is a single instant within a larger event context that may be properly labeled by this verb. The verb 'reach' (one sense of it) is telic; the verb 'deteriorate,' durational; the verb-particle combination 'fill up,' completive. Aspect and *Aktionsart* are distinct analytic categories; in fact, they can both be explicitly distinguished in a single utterance, as the following example from this study's Mandarin sample illustrates:

(1) ta ting-dao le
he listen-COMPL PFVE
he hears this

In this example, the morpheme, "dao" following the verb, "ting" (hear, listen) explicitly signals completive or resultative *Aktionsart*, creating a so-called 'resultative verb complement' construction. The same *Aktionsart* distinction is captured in English in the lexical item 'hear' as opposed to 'listen.' Mandarin

has only the lexical item 'ting' and so the *Aktionsart* distinction is captured by the presence or absence of the completive marker 'dao.' In (1), the speaker further assigns an external aspectual view to the event of hearing by applying the Mandarin perfective aspect marker, "le" to the construction.

We assume that a speaker's choice of perfective as opposed to imperfective verb aspect in an interval of discourse flows from how the speaker is thinking about an event at the moment of speaking. We expect therefore that something reflective of this distinction will likely be evident in the speaker's gestures as he speaks about the event. The definition of the difference between the two major verb aspect categories, given above, suggests the sort of gestural difference we may expect to see. For instance, choice of imperfective aspect in describing an event would mean that the speaker is focusing on — and perhaps visually imagining while speaking — details of the event's unfolding in time. We expect some gestural expression of the process to flow from such elaborated mental imagery. In contrast, we expect choice of perfective aspect to accompany gestures that manifest less elaborated imagery.

A comparison of Mandarin and English

Mandarin Chinese and English have in common the fact that aspect-expressive morphological marking on the verb of an utterance is not grammatically-obligatory. The two languages have been analyzed as differing, however, in how aspectual distinctions are encoded (Binnick, 1991). Mandarin is thought to be less ambiguous, by virtue of possessing a set of single-syllable grammatical aspect-marking particles, two of which are illustrated in the following examples.

(2) a. *Imperfective*
mao zai pa
cat PROG climb
the cat is climbing

b. *Perfective*
nei-zhi mao pa-shang-qu le
the-CL cat climb-up-go PFVE
the cat climbed up

Durative aspect is also indicated by a single-syllable morpheme, "zhe," that follows directly on the main verb of the utterance.

Expression of aspect in English depends more on periphrasis. While imperfective progressive is systematically indicated with the 'be … -ing' construction, as in the English gloss of (2a.), durativity is indicated with a diversity of constructions, such as, "while he was climbing … ," "as he climbed … ," "he keeps on climbing," and so on. In English the expression of perfective aspect is confounded with the simple past tense. Since Mandarin, as mentioned above, altogether lacks morphological tense, there is no possibility of such confounding.

Aspect versus *Aktionsart* in the present study

Though they are analytically distinct, the present study made use of the fact that aspect and *Aktionsart* intersect as linguistic categories in the sense that verbs or phrases expressive of a particular *Aktionsart* often conceptually align more naturally with one linguistic verb aspect than with another. Sampling from the English narrative data made use of such natural alignments, since aspect is not always unambiguously marked in English.

To see how aspect and *Aktionsart* may be aligned but still distinct as linguistic analytic categories, consider the earlier Mandarin example (1). Leaving the perfective aspect marker off can make the utterance sound anomolous to native-speaker ears in many speech contexts. This indicates that completive *Aktionsart* strongly favours a perfective aspectual view of the event. Correspondingly, 'ting-dao' (hear) resists application of either of Mandarin's imperfective aspect-marking particles as illustrated in (3).

(3) a. ʔta zai ting- dao
he PROG {hear/listen}-COMPL
he is hearing

b. *ta ting- zhe dao
he {hear/listen}-DUR COMPL
while he listen-hears

Such examples, however, do not mean that there are no separate and non-redundant systems in Mandarin for marking aspect and *Aktionsart*.

As for English, consider verbs such as 'meander' or 'deteriorate,' whose inherent temporal structure, or *Aktionsart*, aligns with an imperfective aspectual view. Durative *Aktionsart* verbs such as these invite consideration of an event's internal workings. In contrast, taking an internal view of an event expressed by

a verb of punctate *Aktionsart* such as 'hit,' requires a distinct effort of imagination. One might think in terms of a slow motion film clip of a single punch. Embedding a punctate verb in a progressive imperfective aspectual construction, as in, "he's hitting it," changes the *Aktionsart* to iterative (Comrie, 1981a). The imperfective aspect in such a usage occurs at the level of the construction and remains outside the event structure expressed by the verb itself. To focus linguistic expression on the internal constituency of a single iteration of a punctate event — that is, to impose an imperfective aspectual view on the action expressed by such a verb — it is necessary to elaborate the expression. Perhaps a phrase like, "as he is in the act of hitting it for the third time," approaches expression of such a temporal contour. The effortfulness of this exercise is a demonstration of the conceptual alignments that can exist between the linguistic categories aspect and *Aktionsart*.

For the purposes of this study, aspectual view and *Aktionsart* are understood as categories that interact in expression to create different senses of the temporal contours of events. In compiling the sample of English-language utterances used in the comparisons described below, we relied on the linguistic-conceptual alignments that exist between *Aktionsarten* and verb aspects. However, aspect itself is the target of the comparative analyses. Therefore, in sampling, overt grammatical aspect marking was given precedence when inferring the aspectual view adopted by a speaker, including instances where such marking combined with *Aktionsarten* of different types. When overt linguistic aspect marking was absent, aspectual view was inferred on the basis of *Aktionsart*.

'Thinking-for-speaking' in different languages

It seems important to compare the gestures seen in speakers of different languages whose grammars handle the linguistic-conceptual category of aspect in systematically different ways, when these speakers are describing similar witnessed events. Slobin (1987, 1991, 1996) has hypothesized that, where languages differ in terms of the grammatical devices used to encode particular linguistic-conceptual domains, speakers of those languages will manifest different patterns of 'thinking-for-speaking' about these domains. Comparisons of Spanish-speaker and English-speaker descriptions of motion events have generated support for this hypothesis. Briefly, Slobin and colleagues have observed a number of systematic differences in motion event expression

between the two speaker groups that are in accord with differences between the two languages at the levels of lexical semantics and syntactic structure described by Talmy (1985, 1991). For example, Talmy has compared motion expressive clauses in Spanish and English and observed that the MANNER component of motion (e.g., float, bounce) tends to be expressed in the main verb of an utterance in English but in adjuncts to the verb in Spanish. Slobin analyzed motion event expression in English and Spanish narrative discourse and determined that the grammatical difference by Talmy observed contributes to an overall reduction in MANNER 'coloration' in Spanish narrative motion event description, as compared to English; that is, fewer explicit mentions of MANNER of motion. On the basis of this disparity between the two languages, Slobin inferred that, for speakers of Spanish, MANNER of motion is less conceptually salient during acts of speaking or writing about motion events; that is, Spanish thinking-for-speaking involves less conceptualization of the MANNER component of motion.

In the case of verb aspect, under consideration here, differences between Mandarin and English in whether, how richly, or how unambiguously particular aspectual distinctions are encoded might be expected to associate with differences in conceptual representation during acts of speaking. The work of Slobin and colleagues was limited to analyses of spoken and written narrations. In the study reported here we take advantage of the fact that the gestures speakers produce give the analyst additional information about the conceptualizations that are in play at the moment of speaking. Therefore, in addition to within-language gestural differences related to choice of aspectual view, we may also expect between-language gesture differences that reflect the differing means of expression available for this linguistic-conceptual domain.

In this study we compared spoken and gestured expression in Mandarin and English in response to the same set of eliciting stimuli. On the basis of such a comparison we were able to determine whether the distinction between perfective and imperfective verb aspects in speech is related to distinctions in the form and execution of gestures in both languages, when the stimulus to be described is held constant. In addition, we were able to determine whether differences between the two spoken languages, in terms of linguistic encoding of aspectual distinctions, contribute a further source of variation in gesture form. Such comparisons across aspect contexts and across languages permits us to explore thinking-for-speaking in an expanded way, bringing in the dimension of how imagistic thinking functions in relation to linguistic categorial thinking in language production.

Methods

Participants

Fourteen native Mandarin speakers (ten female and four male) and eleven native English speakers (six female and five male), ages 18 to 55, volunteered to participate in an approximately hour-long elicitation session. The Mandarin-speaking group included individuals from both Mainland China and from Taiwan. The English-speaking group included one speaker from Great Britain.

Stimuli

Three different eliciting stimuli were used:

i. Cartoon: An approximately 6 1/2-minute American cartoon of a classic and internationally well-known type, featuring a cat and bird. The events of the cartoon and interactions of the characters are vividly depicted and there is little dialogue. Therefore, though the cartoon is in English, our Mandarin-speaking participants had no difficulty understanding and relating the story content.
ii. Vignettes: A set of 65 videotaped action sequences involving small plastic characters or inanimate objects, sometimes depicted singly executing some movement, sometimes depicted in pairs moving in relation to one another. Each vignette was about 1 to 1 1/2 seconds in length. The 65 vignettes used as eliciting stimuli in this study are part of a test battery designed to elicit American Sign Language verb morphology (Supalla et al., 1993).
iii. Movie: An early feature film by Alfred Hitchcock, one hour and twenty-four minutes in length.

Several of the participants, time permitting, contributed language samples in response to more than one of these elicitations.

Procedure

Participants watched each eliciting stimulus on videotape. Immediately afterward they were videotaped telling what they recalled seeing to a listener. Speakers were asked to relate what they had seen in as much detail as they could recall. They were assured, however, that the tasks were not tests of memory; rather, that we were interested in communicative language use. In the cartoon and movie narration tasks, speakers understood that their listeners were unfamiliar with the story content and that the listeners would subsequently be asked to tell the cartoon or movie story themselves, without having seen it. The goal was to motivate speakers to tell well fleshed-out, comprehensible stories. In

the case of the vignettes elicitation, the investigator served as listener and was present as the speaker viewed each short stimulus, advancing the stimulus video between presentations.

Participants were not told of our interest in their gestures.

Analysis

The elicited cartoon narrations ranged from about four to about eleven minutes in length; the movie narrations, about ten to twenty minutes. As for the vignettes, speakers typically took less than thirty seconds to describe each one.

Speech. The speech resulting from all the elicitations was transcribed in detail. The transcripts include notations indicating all pauses (filled and unfilled), non-speech sounds such as intakes of breath and laughter, and speech dysfluencies such as self-interruptions, self-corrections, repetitions, and prolongations. Locations in the text where spoken utterances expressive of the target aspect and *Aktionsart* distinctions were noted.

Gesture. On a subsequent pass through the same data, the gestures that were co-produced with the target aspect structures were examined at regular speed and in slow motion on the videotapes.[2] Descriptions of the forms and semantic contents of these gestures were added to the speech transcripts. 'Representational' gestures were flagged; that is, gestures with depictive value (iconic, metaphoric). The duration of each gesture in relation to the speech with which it synchronized was determined on the basis of the number of video frames spanned. Also, the presence of particular features of gesture production were noted, in order to build an index of degree of gestural complexity. These features included whether the gesture was executed with a single, uni-directional movement or with a multi-directional movement and whether it involved one or both hands. Gestures in which two hands were employed were coded according to whether the two hands mirrored one another in shape and position in space, or whether they contrasted with one another in these respects. Two handed 'mirror' gestures were judged to be less complex than gestures in which the two hands were in contrast with one another.

The semantic features encoded by these representational gestures were described largely in terms of the 'components of motion' introduced in Talmy (1985). Talmy's classificatory scheme includes such components as MOTION, PATH, MANNER, FIGURE, and GROUND. Consider, for example, a participant who spoke of a cat climbing up a pipe, simultaneously gesturing with both hands the alternating movements of the cat's paws. The speaker further incorporated into these alternating movements an overall upward trajectory of motion. Such a

pantomimic iconic gesture was judged to encode FIGURE (the speaker and her hands represented the cat and its paws), MANNER of motion (alternating hand motions represented a climbing manner), and PATH (the overall upward movement). In contrast, another speaker executed a very simple movement in response to the same stimulus event, where one hand flicked briefly upward. Such a gesture was judged to encode only PATH of motion.

Excluded from these analyses were gestures, whatever their nature, that occurred in the absence of speech. Such gestures may often be configured by the speaker to take the place of speech, and so may not be as much a part of the 'thinking-for-speaking' processes that are our concern here. Gestures that occur on pauses in speech may differ in important, but as yet poorly understood, ways from those that participate with speech in the co-expression of idea units.

Following Kendon (1980), whole gestural movements are termed 'phrases.' Gesture phrases are thought to comprise several 'phases,' the most significant of which, for the purposes of the present analysis, is the 'stroke' phase. This is the phase that displays semantic content related to that of the speech with which the gesture co-occurs.

(4) [the cat **rolls down** the hill] with the bowling ball inside him

Example (4) is an instance of the sort of speech-gesture co-occurrence we encountered in the data analyzed for this study. This transcripted and annotated excerpt also illustrates the typographical conventions we use to annotate features of this co-occurrence. The extent of the interval of occurrence of a gesture phrase in relation to the spoken utterance is indicated by square brackets. In (4) the utterance, "the cat rolls down the hill," is co-extensive with a right-hand gestural movement bounded at onset and conclusion by intervals during which the hand was at rest on the speaker's lap. Within the bracketed interval, there are distinguishable phases of gesture execution, referred to as preparation, stroke, and retraction phases. Thus, the left bracket signals the instant of onset of a gesture phrase in relation to speech. Between this bracket and the beginning of the words highlighted in bold font is the interval of the preparation phase. During a preparation phase, the hand or hands move into position for performance of the gesture's stroke phase. In a transcripted utterance, bold font indicates the co-occurrence of the gesture's stroke phase with some constituents of the accompanying utterance. In this case, the stroke co-occurred with the words, "rolls down." We will discuss presently how the stroke phase was distinguished from the other two phases within the gesture phrase. Finally, the interval of gestural movement in (4) that co-occurs with the

words "the hill" is the retraction phase. During this interval, the gesturing hand is in the process of returning to rest. The right bracket marks the conclusion of the entire gesture phrase relative to the accompanying speech. By this point in this utterance, the speaker's hand has returned to rest.

For the purposes of the analyses reported here, the gesture stroke phase is judged to be the interval within a gesture phrase when some feature or features of the gesture's form, position, and/or movement are semantically interpretable in relation to the accompanying utterance. The stroke may be thought of as the 'semantically-contentful' portion of a gesture production.[3] A somewhat detailed description of the gestural component of (4) will clarify the process of inference by which we identify stroke phases.

At the onset of the first word, "the," the speaker's right hand, loose and relaxed in form, began to rise from a position of rest on the speaker's thigh, moving right to left to reach a point above the left thigh by the offset of the word, "cat." With the onset of, "rolls," the index finger extended slightly from the loose hand to point away from the speaker's body, and the hand began to loop iteratively while also returning left to right on a slightly downward path of motion. During the third and final phase of motion within the bracketed gesture phrase, the hand, having returned to a position above the right thigh, reverted to a loose form while dropping vertically back to the position of rest on the right thigh.

To understand how the process of identifying gesture stroke phases proceeds, the reader is asked to consider which of the three phases of gesture production just described has features of form or execution that link most transparently to the meanings expressed in the accompanying utterance. In line with the analysis on which the findings we report here are based, comparing across the three phases of (4), we would say that the first and final phases bear less of a (if any) semantic correspondence to constituents of the accompanying utterance than does the middle phase. In the first phase the hand, rising as a loose form, does not seem to depict any meaning related to any portion of the spoken utterance. The same is true in the final phase, when it drops back to the speaker's lap. The middle phase most transparently expresses a meaning related to the accompanying utterance. Its repeating circular motion and overall trajectory is readily interpretable as depictive of rolling down an incline. It thus co-expresses an idea conveyed in speech and so is judged to be the stroke phase.

A significant source of input to the process of inference just described derives from the nature of our elicitation protocol. Participants in this study spoke and gestured about stimulus content that was known to the analyst. In

the case of (4), the rightward and downward motion trajectory of the middle phase of gesture execution, together with the repeating circular movement of hand and finger, mirrored the PATH and MANNER of motion of the cat that the participant saw in the eliciting stimulus. Such correspondences among images presented in the stimulus and images presented by gestures are the norm in narrative data elicited with the protocol we used in this study. They add certainty to the process of locating the meaningful stroke phase within each gesture phrase. The stroke is where we see most clearly the nature of whatever imagistic thinking is in play, moment-by-moment, during the act of speaking.

Using professional-grade VCRs (Sony, model EVO-9650), the synchrony of the gesture phrases and phases in relation to the speech with which they occurred was assessed to the degree of within-syllable accuracy.[4] The bulk of gesture-descriptive annotation on which these analyses rest was verified by more than one experienced gesture coder.[5]

Sampling. From each language, Mandarin and English, 100 utterance-gesture pairs were sampled from the data provided by the speakers of that language. The sampling procedure was simply to scan through each speaker's narrative data sequentially from the beginning until an utterance-gesture pair meeting the sampling criteria was encountered. Recall that the criteria were that the spoken utterance contain one of the target aspect or *Aktionsart* usages and also be accompanied by a representational gesture. This unit of speech-transcripted and gesture-annotated narrative production would be extracted, added to a separate document file of accumulated utterance-gesture excerpts, and held for subsequent analysis. The sequential scan through the data would then continue. The data elicited from each speaker were sampled in roughly equal proportion to that of any other speaker. No fewer than two and no more than five utterance-gesture pairs of any one aspect or associated *Aktionsart* type were extracted from any one speaker.

The results described below are based on this two-language sample of 200 utterance-gesture pairs. Within each language's sample of 100 spoken utterances, 50 utterances were of the perfective aspect-marked type and 50 of the imperfective aspect-marked type. Within the latter, imperfective, category for each language were 25 spoken utterances each of the progressive and durative aspect types.

Results

Comparisons within and across the Mandarin and English narrations revealed systematic within-language covariation of linguistic aspectual view and/or verb *Aktionsart* and features of co-occurring gesture. In both languages, gestures that occurred in imperfective aspect-marked speech contexts were statistically more likely to: (*i*) take longer to produce and (*ii*) be more complex than those that occurred in perfective aspect-marked speech contexts. No between-language differences were observed on the dimensions analyzed.

Following a summary of the results obtained for all three aspect categories, the feature of gesture form that we described as 'stroke agitation' is treated in more detail.

Durations of gesture strokes

Figure 1 shows that the durations of gesture strokes that occur in perfective- and imperfective-marked speech contexts differ. The strokes of gestures that occur in perfective aspect-marked speech contexts are much shorter on average than those that occur in imperfective contexts. This is true for both Mandarin and English. The means and standard deviations for the duration data are displayed in Table 2.

Within each language, the difference in stroke durations of gestures accompanying speech marked for these aspectual distinctions is statistically highly significant. An additive model, two-way analysis of variance comparing stroke durations between the two languages yielded a significant effect for aspect [$F(1) = 104.31$, $p < .0001$], and no effect for language [$F(1) = 0.50$, $p < .480$].

Complexity of gesture strokes

Table 3 gives an overview of the major differences that were observed in the

Table 2. Mean durations in seconds, and standard deviations, of gesture strokes accompanying utterances with perfective versus imperfective aspect: Mandarin versus English.

	Mandarin	***English***
Perfective	mean = 0.274 (std. dev., 0.186) *n* = 50	mean = 0.249 (std. dev., 0.114) *n* = 50
Imperfective	mean = 1.018 (std. dev., 0.934) *n* = 50	mean = 1.158 (std. dev., 0.623) *n* = 50

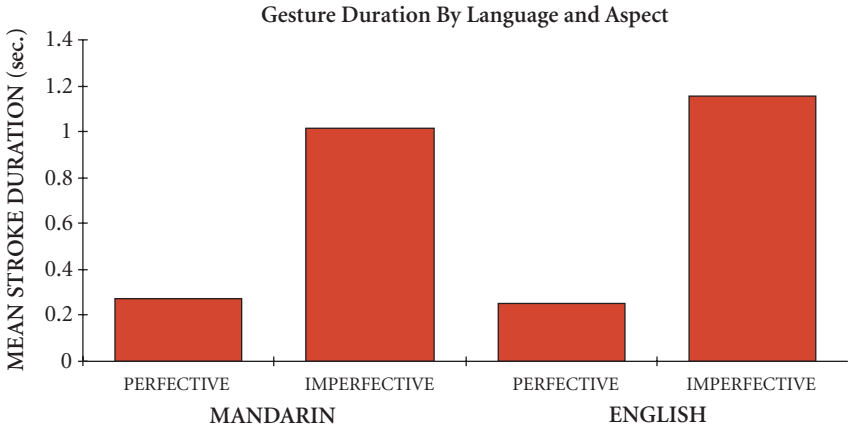

Figure 1. Mean durations of the stroke phases of gesture for Mandarin and English, by choice of aspectual view.

complexity and inferred semantic content of gestures produced together with the different aspect categories in speech. In terms of the components of motion encoded by gesture form and execution, almost 100% of gestures in perfective contexts omitted all event dimensions other than PATH. Often the depicted trajectory itself was quite reduced, resulting in a very minimal gesture form. There was little representation of MANNER, FIGURE, or GROUND in the sample of gestures that accompanied perfective aspect-marked speech. The minimal character of these gesture forms may be in part a consequence of their very short durations; that is, the shorter the interval of execution, the less opportunity there may be to execute highly complex, multi-element, or repeating movements.

Table 3. Features of gesture form and semantic content typical of aspect contexts for both Mandarin and English, indicating degree of complexity.

	Perfective	*Progressive*	*Durative*
Form	– 1- or 2-hand/mirror – simple motion strokes	– 1- or 2-hand/mirror – complex motion strokes	– 1- or 2-hand/contrast – motion and hold-strokes
Semantic content	– 1 event – PATH	– 1 event – PATH – MANNER – (FIGURE)	– 2 events or entities – 2 PATHS – GROUND – FIGURE – (MANNER)

In contrast, around 85% of gestures accompanying spoken progressive aspect-marked utterances in both Mandarin and English were multi-directional, 'agitated' motions. In the case of speech-gesture descriptions of stimulus motion events, stroke agitation was often interpretable by the analyst as expressive of MANNER of motion, on the basis of the criteria discussed in relation to example (4). Of the three aspect categories compared, gestures in progressive aspect speech contexts showed the greatest amount of MANNER-depiction.

The example speech-gesture productions below illustrate some features common to gestures in imperfective progressive-marked speech contexts. Example (5) was an utterance elicited by a vignette in the Supalla et al. (1993) ASL verb morphology battery:

(5) a construction m[an **is rolling across the screen** /]

The progressive-marked speech here is accompanied by a rather elaborate gesture in which the speaker's forearm, representing the prone FIGURE in the eliciting stimulus, moves laterally while oriented roughly perpendicular to his torso. Hand and arm repeatedly loop while at the same time moving across the speaker's gesture space in a PATH and MANNER similar to that of the small plastic man's rolling motion depicted in the eliciting stimulus. This gesture, illustrated in Figure 2, is exemplary of the many cases in which extendable, agitated motion of the gesture stroke iconically represents the MANNER component of a motion event.

Examples (6) and (7) illustrate the fact that two speakers who view the same stimulus event may choose different aspectual views when describing it. That is, the motion events depicted in the eliciting stimuli are amenable to description according to more than one aspectual frame, depending on the speaker's immediate discourse purpose. These excerpts, illustrated in Figure 3, are from the narrations of two different Mandarin speakers, each narrating a cartoon event in which the cat rolls down a hill, propelled by a bowling ball that he has swallowed. Each of the speakers made the same choice of main verb, "gun" (roll). These examples further illustrate how iconic gestures covary with aspectual view, when both image and verb choice are the same.

(6) *Perfective aspect*:
[ying-gai shi **gun-xi**a-lai le]
should be roll-down-come PTVE
it is probable that (he) rolled down
(iconic gesture, 2-hand/mirror, simple-stroke, PATH only)

Figure 2. A complex gesture stroke encoding figure, path, and manner, produced with imperfective progressive aspect-marked speech.

(7) *Imperfective progressive aspect*:
[ran-hou ne jiu zhi-**jie gun**][−xia-lai jiu **nei qiu yi-zhi za**]i gun
after NE then direct roll-down-come then the ball continuous PROG roll
so then, (it) rolls right down then the ball is continuously rolling
(iconic gesture, 2-hand/mirror, agitated-stroke, PATH+MANNER)

The speaker shown on the left in Figure 3 used the verb, "gun" (roll) with the Mandarin perfective aspectual marker, "le." His speech was accompanied by a brief, two-handed, PATH-encoding gesture that moved from his left to right. The speaker shown on the right narrated the same event and also used the verb, "gun" (roll), but chose progressive aspectual view instead. Her two mirroring hands also moved from left to right, but with fingers wiggling, such that her gesture depicted MANNER and PATH simultaneously.

Spoken durative aspect

As for the gestures that speakers produced in synchrony with durative aspect-marked speech, what was most noticeable in both Mandarin and English was their sustained quality and further, the way these gestures maintained more complex spatial arrangements. The long duration of these gestures would seem to facilitate the greater complexity of depiction.

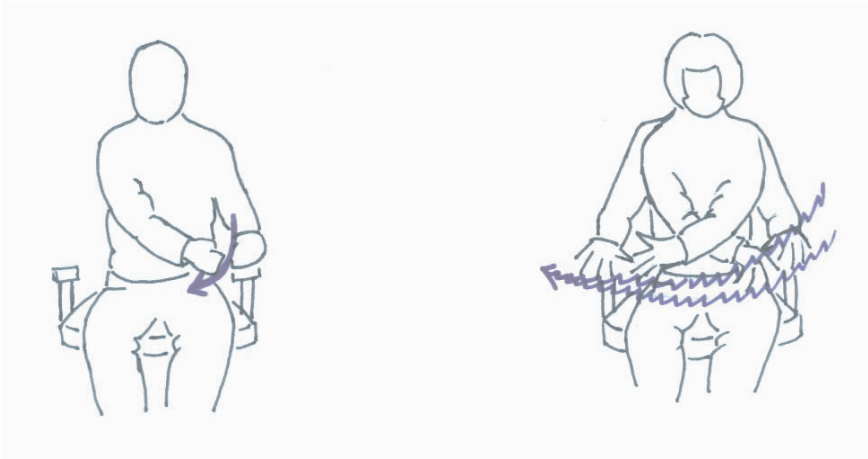

Figure 3. Two gestures depictive of the same stimulus content. Left: A simple PATH stroke, produced in the context of perfective aspect-marked speech; right: a complex PATH and MANNER stroke produced with imperfective progressive aspect-marked speech.

(8) *Imperfective durative aspect*
[$_{RH}$and **/ as he's com**ing up and the bowling ball's coming down]
and / as he's coming up [$_{LH}$and the **bowling** ball's coming down]

The speaker in example (8) gestures about two events that occur simultaneously in the cartoon.[6] Our data indicate that this is a typical discourse environment for choice of durative aspectual view. Durativity facilitates foregrounding and backgrounding of events in relation to one another. In addition to showing contrasting handshapes and positions, the two hands execute their stroke phases at different times relative to the accompanying speech, all the while maintaining a unified, well-specified configuration in gesture space.

Gesture stroke agitation

As Table 3 and the examples above show, in the case of motion event descriptions, stroke agitation in progressive aspect-marked speech contexts often works to depict MANNER of motion. For example, if a speaker describes a figure that rolls, the gesture stroke may consist of repeating circular movements expressive of a rolling MANNER of motion. Similarly, in responses to other stimulus events depicting various types of MANNER, we observed, for instance, gestured running, climbing, or bouncing MANNERS, in the context of progressive aspect-marked speech. Initially, it appeared that stroke agitation in all cases

might be nothing more than iconic representation of MANNER. Of significance for our interpretation of these data, however, are two further observations: (*i*) stroke agitation sometimes distorted the iconic depictive relationship between gestures and the witnessed events, and (*ii*) stroke agitation appeared as well when speakers described events of stasis (e.g., 'stand,' 'listen') or used mental state verbs (e.g., 'worry'). Since it seems unreasonable to think of MANNER of motion as part of the lexical semantic specifications of such verbs, these facts may be taken as evidence that stroke agitation is representationally partially abstract in nature. In what follows, we discuss the possibility that stroke agitation arises in part as a result of speaker choice of the progressive aspectual view.

Stroke agitation in depicting events of motion. Example (9), from the Mandarin data, illustrates how speaker choice of aspectual view appears to set part of the framework for iconic gestural depiction of motion events. This speaker was describing a vignette from the Supalla, et al. (1993) battery in which a ruler slides smoothly and uni-directionally across the surface of a table. In speech, he chose to frame his description with the progressive aspectual view.

(9) [yi-ge chi-zi zai zhuo-zi-shang] [/ **zai dong**]
one-CL ruler on table-surface / PROG move
a ruler is moving on top of a table

In the first part of the speaker's utterance, his gestural movement does represent a unidirectional motion. A straight path gesture was produced during the first interval of bracketed speech. But then the speaker performed a second gesture in synchrony with a phrase comprising the progressive aspect particle, "zai" and the verb, "dong" (move). The latter utterance synchronized with a repeating form of agitated motion stroke that, in zig-zagging back and forth, was in explicit contrast to the smooth MANNER of motion depicted in the eliciting stimulus.

Stroke agitation in depicting events of stasis. An English speaker described a different vignette from the Supalla, et al. battery, in which a ruler stands on its end, completely still, for an interval of time and then falls over. The speaker says, "a ruler is standing on its end," accompanied by a gesture in which she holds her forearm vertically out in front of herself, an iconic representation of the ruler. The speaker then *shakes* the forearm slightly in synchrony with the imperfective progressive aspect-marked phrase, "is standing."

Similar examples of motion-suggestive gestures generated in the absence of any such motion in the eliciting stimulus come from speakers narrating a scene in the cartoon stimulus where the cat, partially concealed in a hiding place, listened for information that would lead him to the bird. In the eliciting stimulus, the act of listening was depicted by the cat perking up an ear and holding quite still for a moment while listening for the information. There was no movement associated with the listening pose in the cartoon scene the speakers viewed, yet some speakers described this event of stasis by saying, "he's listening," while at the same time slightly shaking a cupped hand, palm forward, next to one of their ears.

Examples like these highlight the conceptually abstract nature of stroke agitation: it can occur in contexts when no 'agitated' MANNER of motion is being described. Thus it is not produced solely in service of iconic depiction of witnessed events. What shows up as MANNER-marking in many motion gestures and as non-iconically depictive agitation in others may derive in part from a more kinesthetic, as opposed to visual, representation of a state of being or an action in progress, as something that is 'enlivened.'

Aspect and metaphor. The gestural indices of aspect outlined here were not confined to iconic gestures depicting concrete events and actions. Metaphoric gestures displayed the same range of features. Several instances were found in the data examined for this study. For example, in one of the movie narrations, the speaker described the protagonist's troubled state of mind by saying that she's, "sitting there worrying." The utterance accompanied a gesture in which the speaker's two hands rotated around one another alternatingly, stroke agitation expressive of some kind of processing metaphor. The protagonist in the movie scene, however, was physically completely still.

Other examples of aspect-marked metaphoric gestures were those found at episode boundaries and at the conclusions of narrations. At such narrative junctures, speakers often made some metanarrative statement to the effect of, "that's it for that part," "that's how it ends," or a simple paranarrative/interactive, "OK?" synchronized with an inquiring glance at the listener. In Mandarin, such statements usually carried the perfective aspect marker, "le." In speakers of both languages, they were frequently accompanied by small gestures that appeared to quickly sweep something away at the edge of gesture space, or by conduit metaphoric gestures that opened up in a quick, single stroke, appearing to release their contents.

Such examples show that the gestural indices of linguistic aspectual view are pervasive in narrative discourse, that they are separable from iconic depiction of witnessed events, and that they occur on multiple discourse levels.

Summary and conclusions

We have identified several distinctions between the gestures that occur in perfective as opposed to imperfective aspect-marked speech contexts. These distinctions in relation to choice of verb aspect were evident even when we compared narrative descriptions of the same witnessed events across speakers. We saw instances in which gesture features that covary with verb aspect were incorporated even when doing so partially distorted iconic representations of the witnessed events. Three things are clear. The first is that speaker choice of aspectual view is to a large extent independent of the nature of the event being described. Second, in the process of becoming gestural representations, stimulus images retained by the speaker are modulated in relation to the categorial linguistic choices the speaker makes. At least this seems true in the domain of categorial aspect distinctions considered here. In other words, aspect appears to set a significant part of the framework for gestural iconic depiction. Third, that no significant differences were found between Mandarin and English speaker's gestures suggests that the domain tapped by linguistic aspect is very fundamental in human cognition. The differences in grammatical form that have been described for the two languages appear, on the basis of this analysis, to be notational variants that allow speakers to accomplish very similar conceptual and expressive ends.

The nature of iconicity in gesture

The findings concerning aspect-related features of gesture form make it necessary to think carefully about the nature of iconic depiction in gesture. They prompt us to ask how some observed features of gesture may relate to more abstract levels of linguistic-conceptual organization, beyond the visual images a speaker retains as a result of viewing an event. We saw that gestures in spoken perfective contexts typically reduced event depiction to a minimum of detail. A witnessed motion event, such as a cat rolling down a hill with its legs flailing, can be collapsed into an 'iconic' gesture that encodes nothing but a reduced PATH of motion. This leaves out many components of motion that are easily expressed in gesture. We see that they are routinely expressed in other

speech contexts. Alternatively, the event may be encoded in a complex gesture that preserves many more of its components, or even *adds* motion that is not present in the stimulus.

Hadar and Butterworth claim that gesture comes from visual imagery via a, "direct route," that it is, "the motor manifestation of imagistic activation" (1997: 167). It seems reasonable to think of gestures as manifestations of some kind of imagistic thinking that speakers experience as they describe events. Yet, we have demonstrated gestural diminishment as well as augmentation of known stimulus images that we can assume with some confidence speakers had in mind as they spoke. This is evidence that features of production, such as very minimal movement strokes and stroke agitation, are not always nor simply iconically depictive. We must consider how such images come to differ from those we can assume the participants visually encoded at the time of stimulus presentation and then retained for later description.

In our study, the gestured manifestation of images differed systematically in terms of their complexity and duration in relation to choice of verb aspect, suggesting that aspect plays a significant role in the content and structure of expression. On the basis of such evidence, we conclude that gestures are part of linguistic-conceptual representations, rather than being manifestations of purely visual imagery, processed somehow independently of spoken forms during intervals of language production. This is not the same as saying that gestures merely mirror linguistically codified aspect contrasts. Rather, different verb aspects appear expressive of fundamental distinctions in the ways we can 'cognize' an event during acts of speaking. We claim that synchronized elements of speech and gesture are co-expressions of a unitary representation at this fundamental level. Our findings and conclusions are thus in accord with theories of language production (McNeill, 1992; McNeill & Duncan, 2000) that claim, on the basis of such semantic co-expressivity as well as the tight synchrony of the two modalities in production, that speech and gesture forms flow from a single source in the process of thinking-for-speaking.

Notes

1. Hui-fang Hong (personal communication).

2. The procedures for transcribing and annotating narrative speech and gesture production are described in McNeill (1992) and in Duncan et al. (1995).

3. Note that this is not the same definition originally offered by Kendon, who defined the stroke phase of gesture production as, "an effort peak," or, "a moment of accented movement" (1980: 212). The analyses of speech-gesture co-occurrence described here are largely meaning-driven. The discussion of example (4) that follows illustrates how it is possible to make an assessment of the meanings of gesture phases as these relate to synchronized speech.

4. The degree of precision of the assessments of speech-gesture synchrony is illustrated in the annotated transcription excerpts given as examples (4)–(9) in this paper. Among these, the reader will note instances where gesture phrase (bracketing) or stroke phase (boldface) onsets or offsets occur within a syllable. Such synchrony assessment is possible only using VCRs of the type that permit slow-motion and frame-by-frame video play with accompanying audio. All of the results reported in this study were based on narrative data analyzed in this way so as to obtain very precise and accurate assessments of gesture-speech synchrony.

5. We thank Karl-Erik McCullough, Inge Eigsti, Hui-fang Hong, and Desha Baker for their contributions to perfecting the gesture coding on which these analyses were based.

6. Underlining is the convention for annotating the presence of gestural holds. These are intervals of time where the hands remain motionless in space, maintaining marked handshapes and spatial arrangements.

References

Binnick, Robert (1991) *Time and the verb: A study of tense and aspect.* Cambridge: Oxford University Press.

Comrie, Bernard (1981a) *Aspect.* Cambridge: Cambridge University Press.

Comrie, Bernard (1981b) *Language universals and linguistic typology*. Oxford: Blackwell.

Duncan, Susan D., David McNeill, & Karl-Erik McCullough (1995) How to transcribe the invisible — and what we see. In Daniel C. O'Connell, Sabine Kowal, & Roland Posner (Eds.), *KODIKAS/CODE* (Special issue on signs for time: *Zur Notation und Transkription von Bewegungsabläufen)*, *18.*

Hadar, Uri & Brian Butterworth (1997) Iconic gestures, imagery, and word retrieval in speech. *Semiotica*, *115, 1/2*, 147–172.

Kendon, Adam (1972) Some relationships between body motion and speech. In Aaron W. Siegman & Benjamin Pope (Eds.), *Studies in dyadic communication*, 177–210. New York: Pergamon Press.

Kendon, Adam (1980) Gesticulation and speech: Two aspects of the process of utterance. In Mary Ritchie Key (Ed.), *The relation between verbal and nonverbal communication* (pp. 206–227). The Hague: Mouton Publishers.

Kendon, Adam (2000) Speech and gesture: Unity or duality? In David McNeill (Ed.), *Language and Gesture.* Cambridge, Cambridge Unversity Press.

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

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

McNeill David & Susan Duncan (2000) Growth points in thinking-for-speaking. In David McNeill (Ed.), *Language and Gesture*. Cambridge: Cambridge Unversity Press.

Slobin, Daniel I. (1987) Thinking for Speaking. In Jon Aske et al. (Eds), *Proceedings of the Annual Meeting of the Berkeley Linguistic Society*, pp. 435–445.

Slobin, Daniel I. (1991) Learning to think for speaking: Native language, cognition, and rhetorical style. In John Gumperz & Stephen C. Levinson (Eds.), *Rethinking linguistic relativity*. Cambridge: Cambridge University Press.

Slobin, Daniel I. (1996) Two ways to travel: Verbs of motion in English and Spanish. In Masayoshi Shibatani & Sandra A. Thompson (Eds.), *Grammatical constructions: Their form and meanings*. Oxford: Clarendon Press.

Talmy, Leonard (1985) Lexicalization patterns: Semantic structure in lexical forms. In Timothy Shopen (Ed.), *Language typology and syntactic description, volume III: Grammatical categories and the lexicon* (pp. 57–149). Cambridge, Cambridge University Press.

Talmy, Leonard (1991) Path to realization: A typology of event conflation. In L. A. Sutton, C. Johnson, & R. Shields (Eds.), *Proceedings of the 17th annual meeting of the Berkeley Linguistics Society*, pp. 480–519. Berkeley, CA: Berkeley Linguistics Society.

Vendler, Zeno (1967) *Linguistics in philosophy*. Ithaca, New York: Cornell University Press.

Author's address

Susan D. Duncan
Psychology Department
University of Chicago
5848 S. University Ave.
Chicago, IL 60637
USA

E-mail: deng@uchicago.edu

About the author

Susan Duncan is a Research Associate in psychology at the University of Chicago, Chicago, IL USA (PhD in psychology from the University of Chicago in 1996). She has held postdoctoral fellowships at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands, as well as at National Yang Ming University, Taipei, Taiwan, ROC where she was a National Science Council Fellow and university lecturer. Her research has focused on cross-language comparative analysis of speech and gesture in adults and children and in individuals with neurological damage or disorders affecting language production. Recently her work has extended to analysis of prosody in sign language, focusing on Taiwan Sign Language.