Gesture and Thought

DAVID MCNEILL

(2005)

University of Chicago Press CHICAGO AND LONDON

CHAPTER 1

Why Gestures?

INTRODUCTION

This book is a companion to *Hand and Mind*, which appeared in 1992. The key ideas were planted in that earlier book and in numerous ways have been developed and extended in this one. In 1992 the emphasis was on how gestures reveal thought; now it is how gestures fuel thought and speech. The new step is to emphasize the 'dynamic dimension' of language—how linguistic forms and gestures participate in a real-time dialectic during discourse, and thus propel and shape speech and thought as they occur moment to moment. As in the earlier book, gestures, language, and thought are seen as different sides of a single mental/brain/action process. They are integrated on actional, cognitive, and ultimately biological levels. The difference is that now I present gestures as active participants in speaking and thinking. They are conceived of as ingredients in an imagery-language dialectic that fuels speech and thought.

The gestures I mean are everyday occurrences—the spontaneous, unwitting, and regular accompaniments of speech that we see in our moving fingers, hands, and arms. They are so much a part of speaking that one is often unaware of them, but if you look around and watch someone talking in informal terms you are likely to see the hands and arms in motion. Why? This is the question I propose to answer, ultimately in evolutionary terms.

To obtain an answer, in part, I carry forward an approach introduced by Vygotsky in the 1930s. Vygotsky is celebrated as an alternative to Piaget and, for many, as an antidote to a kind of sterile asocial cognitivism they imagine (not altogether inaccurately) dominates current linguistics and cognitive

psychology. But Vygotsky had other themes. He argued for a different kind of psychology, one that is antireductionist, holistic, dialectical, and grounded in action and material experience. It is this sometimes overlooked Vygotsky that this book carries forward in ways that were not available in his day. Foremost among these is the systematic study of gesture and language as they occur spontaneously in daily speech.

The main theme of this book is that *language is inseparable from imagery*, a statement from Damasio (1994, 1999) somewhere. The imagery in question is embodied in the gestures that universally and automatically occur with speech. Speech and gesture occupy the same time slices when they share meanings and have the same relationships to context. It is profoundly an error to think of gesture as a code or 'body language', separate from spoken language. One message of this book is that gestures are *part* of language. It makes no more sense to treat gestures in isolation from speech than to read a book by looking only at the 'g's. It is also an error, in fact the same error, to think of speech as separate from gesture—as if to focus on just the 's's. The aim of the book is to present in full the arguments for the inseparability of language/speech and imagery/gesture, and to seek explanations for why this arrangement should be so. A précis of the book is given at the end of this chapter.

I suggest that language has two dimensions, static and dynamic, that combine in every speech event via the above-mentioned dialectic. This imagery-language dialectic (materialized in gesture and speech) is an interaction between unlike modes of thinking. The disparity of these modes is the 'fuel' that propels thought and language; the dialectic is the point at which the two dimensions intersect. The central part of the book describes how this dialectic takes form, how it propels thought and speech, and what must take place to resolve the tension between these unlike modes of cognition over the very brief intervals of time (just seconds) during which utterances are conceived and produced.

I should also stress what the book is not. It is not a comprehensive review of current gesture work, and it is not a commentary on all that has been discovered about gestures, which is by now a great deal. In particular, I say little about so-called emblems, or 'quotable' gestures as Adam Kendon has termed them, the very gestures that many people think of when they hear the word "gesture." Although I cover ample ground, I have kept my goal clearly in mind, and this has guided me in what to include and what not. It is accordingly a good idea

to start by distinguishing among different kinds of occurrences that can be called "gestures," and to specify the kinds that are in focus.

WHICH GESTURES? A CONTINUUM

Adam Kendon (1988a) once distinguished gestures of different kinds. I then arranged these along a continuum that I named "Kendon's continuum" in his honor (McNeill 1992). The gestures we are primarily concerned with are the 'gesticulations'.

'Gesticulation' is motion that embodies a meaning relatable to the accompanying speech. (The nature of this relationship is analyzed in Chapter 2.) Gesticulation is by far the most frequent type of gesture in daily use, and it covers many variants and usages. It is made chiefly with the arms and hands but is not restricted to these body parts—the head can take over as a kind of third hand if the anatomical hands are immobilized or otherwise engaged, and the legs and feet too can move in a gesture mode (cf. McClave 2000).

'Speech-linked gestures' are parts of sentences themselves. Such gestures occupy a grammatical slot in a sentence—"Sylvester went [gesture of an object flying out laterally]," where the gesture completes the sentence structure.

'Emblems' are conventionalized signs, such as thumbs-up or the ring (first finger and thumb tips touching, other fingers extended) for "OK."

'Pantomime' is dumb show, a gesture or sequence of gestures conveying a narrative line, with a story to tell, produced without speech.

At the other extreme of the continuum, **'signs'** are lexical words in a sign language (typically for the deaf) such as ASL. Sign languages have their own linguistic structures, including grammatical patterns, stores of words, morphological patterns, etc. The linguistic code of ASL is quite unlike that of English. Sign languages have evolved without the requirement of being coordinated with speech. In fact, hearing signers find that producing speech and signs simultaneously is disruptive to both.

As one moves along Kendon's 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*. Gesticulations are obligatorily accompanied by speech but have properties unlike language. Speech-linked gestures are also obligatorily performed with speech, but relate to speech in

a different manner—sequentially rather than concurrently and in a specific linguistic role (standing in for a complement of the verb, for example). Signs are obligatorily *not* accompanied by speech and have the essential properties of a language. Clearly, therefore, gesticulations (but not the other points along Kendon's continuum) combine properties that are unalike, and this combination occupies the same psychological instant. A combination of unalikes at the same time is a framework for an imagery-language dialectic.

FROM CONTINUUM TO CONTINUA[1]

On reflection, however, we can see that Kendon's continuum is actually a complex of separate continua, each based on an analytically distinct dimension along which the types of gestures (gesticulation, emblems, etc.) can be differentiated. I shall explain the points along the continua by reference to Figure 1.1.

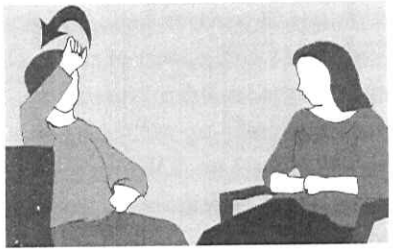

Fig. 1.1. Spontaneous iconic gesture with "bends it way back."

The speaker was saying, "he grabs a big o[ak tree and he **bends it way back**]," with his hand moving through an arc, as shown. His hand rose from the armrest of the chair as he said "oak" (left bracket), reached its apex with "he," at which moment there was a brief prestroke hold (underlining); the hand then moved downward and to the side during the boldface section (the stroke—the part of the gesture depicting the actual 'bending back': the phase shown). At this point there was a poststroke hold and a new gesture began (not shown). Gesture transcription is explained in detail in the appendix.

During the stroke phase, the hand appeared to grasp and bend back an object with some thickness. Such a gesture has clear iconicity—the movement and the handgrip; also a locus (starting high and ending low)–all creating

1. An earlier version of this section appeared in the introduction to McNeill (2000).

imagery that is analogous to the event being described in speech at the same time (a comic book character bending back an oak tree).

Continuum 1: *relationship to speech*

Gesticulation →	Emblems →	Pantomime →	Sign Language
Obligatory presence of speech	Optional presence of speech	Obligatory absence of speech	The same

The first continuum controls the occurrence of gesture with speech. The bends-it-back gesture is meaningful only in conjunction with the utterance of "bends it back." An OK emblem can be made with speech or not. Pantomime, by definition, does not accompany speech (lack of speech is therefore trivial). With sign language, while it is possible to produce signs and speak simultaneously, doing so has a disruptive effect on both speech and sign. Speech becomes hesitant and sign performance is disrupted at the level of the main grammatical mechanisms of the language that utilize space rather than time for encoding meanings (Nelson et al. 1993).

Associated with the speech continuum is another continuum that reflects the presence vs. absence of the characteristic semiotic properties of a linguistic system. This is a second continuum on which gesticulation and sign language hold down the extremes, while pantomime and emblem have exchanged places:

Continuum 2: *relationship to linguistic properties*

Gesticulation →	Pantomime →	Emblems →	Sign Language
Linguistic properties absent	The same	Some linguistic properties present	Linguistic properties present

The bends-it-back gesture lacks all linguistic properties. It was nonmorphemic, not realized through a system of phonological form constraints, and had no potential for syntactic combination with other gestures. We can demonstrate the inapplicability of linguistic properties through a thought experiment. Imagine another person saying the same thing but with "it" meaning the corner of a sheet of paper. Then, rather than the hand opening into a grip, the thumb and forefinger would come together in a pinch; rather than the arm moving forward and slightly up, the pinching hand would be held slightly

forward and down; and rather than pull the arm back, the pinching hand would rotate outward or inward. Also, this gesture would naturally be performed with two hands, the second hand 'holding' the paper that is being bent back. That is, none of the form properties of the first gesture would be present in the second gesture, bends-it-back though it is. Neither gesture in fact obeys constraints within a system of forms; there are only constraints that emerge from the imagery of bending itself—an oak tree versus a tab of paper. The handshape and position are creations of the moment and reflect the speaker's imagery—of a character from a story reaching up and forward to pull back a tree, of someone turning down the corner of piece of paper.

The ASL sign TREE (shown in Figure 1.2) in contrast *is* constrained by the phonological properties of the ASL language system.[1] The 5 handshape is a standard one of the language; the sign could not be formed and remain intelligible with a handshape that is not part of the language. While the 5 handshape has recognizable iconicity, it is a standardized selection of iconic features that other sign languages, with signs equally iconic, do not use (Danish Sign Language, for example, traces an outline of a tree). And the sign is what Okrent calls 'nonspecific' in that it is used equally well for all kinds of trees and tree shapes, not just trees with long bare trunks and fluttering leaves.

Fig. 1.2. ASL sign for TREE.

Pantomime, like gesticulation, does not seem to obey any system constraints (not considering theatrical pantomime, which does have its own traditional forms and rules; see Fischer-Lichte 1992). For example, showing what a vortex is with pantomime could be done by twirling a finger or by rotating the whole hand, and neither version would be unintelligible or seem to be a violation of a system constraint.

Emblems, on the other hand, do show system constraints. There are differences between well-formed and not-well-formed ways of making the gesture.

1. I am grateful to Arika Okrent for this example and analysis.

Placing the middle finger on the thumb results in a gesture with some kind of precision meaning, but is not recognizable as the OK sign. The OK gesture, like a word, is constrained to assume a certain 'phonological' shape. Yet these constraints are limited and don't by any means amount to a full language. There is no way to reliably reverse the OK sign, for example. Forming it and waving it back and forth laterally (another emblem that, on its own, conveys negation) might convey "not OK," but it also might be seen as meaning the opposite of negation—waving the hand could call attention to the OK sign, or to suggest that many different things are OK—a flexibility that is basically not linguistic in character.

Comparing Continuum 1, 'relationship to speech', to Continuum 2, 'relationship to linguistic properties', we see one of the basic facts of gesture life: the gesticulations, with which speech is obligatorily *present*, are the least languagelike; the signs, from which speech is obligatorily *absent*, have linguistic properties of their own. This is not so paradoxical as it may seem. It reveals that 'gesture' has the potential to take on the traits of a linguistic system, but as it does so it ceases to be a component of the spoken language system. This is the conclusion of thirty years of investigation of the sign languages of the deaf (see, for example, the collection of papers in Emmorey & Reilly 1995; Liddell 2003a,b). It is also the conclusion of research on the deaf children of hearing, nonsigning parents. These children are exposed to neither a sign language nor speech (in that they cannot hear the speech their caretakers produce) and they develop their own means of gestural communication, termed 'home signs', that manifest a number of important linguistic properties, such as a lexicon and basic syntax (Goldin-Meadow & Mylander 1984; home signs and cultural sign languages in general are discussed in Chapter 4.3). In effect, their gestures move to the right of the continuum. The conclusion is that nothing about the visual-manual modality per se is incompatible with the presence of linguistic properties. Yet gestures combined with speech lack linguistic properties.

The comparison of the first and second continua thus shows that when the *vocal* modality has linguistic system properties, *gesture*, the manual modality, does not take on these properties. And when it does not, speech tends to be *obligatory* with gesture. This is certainly one of the more interesting facts of gesture. It implies that speech and gesture combine into a system of their own in which each modality performs its own functions, the two modalities supporting one another. This book operates upon this premise.

Continuum 3: *relationship to conventions*

Gesticulation	→	Pantomime	→	Emblems	→	Sign Language
Not conventionalized		The same		Partly conventionalized		Fully conventionalized

Convention means that the forms of gestures and the meanings with which they are paired meet some kind of socially constituted or collective standard. It is because the gesture is ruled by convention that only forefinger and thumb contact are recognizable as OK. At the gesticulation end, in contrast, a lack of convention is a sine qua non. The bends-it-back gesture is conventional only in the broadest sense (e.g., that gesturing is acceptable in storytelling contexts). There are no conventions telling the speaker what form bending back is to take. The TREE sign, however, *is* constrained by the conventions of ASL. It must meet form standards according to which only an upright arm with a 5 handshape is TREE.

The fourth continuum concerns the semiotic differences between gesticulation and linguistic codes of all kinds (spoken as well as signed). This dimension further shows the richness that can emerge from combining gesticulation with speech in a unified speech-gesture system, a system that places contrasting kinds of semiotic properties into one vessel (in what sense *sign*-gesture systems might also exist is a topic of much current interest; see Liddell 2000, 2003a,b, and the discussion in Chapter 4.3).

Continuum 4: *character of semiosis*

Gesticulation	→	Pantomime	→	Emblems	→	Sign Language
Global & synthetic		Global & analytic		Segmented & synthetic		Segmented & analytic

Global refers to the fact that the determination of meaning in a gesticulation proceeds in a downward direction. The meanings of the 'parts' of the gesture are determined by the meaning of the whole. This contrasts to the upward determination of meanings in sentences. In the bending back gesture, we understand from the meaning of the gesture as a whole—a character bending something back—that the hand (one of the 'parts') equals the character's hand, the movement (another part) equals the character's movement, and

the backward direction (a third part) equals the character's backward movement. These are not gestural morphemes. It is not the case that the hand in general means a hand or movement backward must always mean movement backward. Özyürek (2000) observed gestures for the same 'outward' meaning performed in different directions, dependent on the social context—whether the speaker was addressing one listener or two. The pairing of the outward meaning and the form (direction) of the gesture was not fixed, hence it is not listable as in a morphology, but varied depending on the context of communication. In speech, on the other hand, the character bending back the tree was constructed out of independently meaningful words or segments organized according to a standardized plan or syntax. The top-down global semiotic of gesticulation contrasts with the bottom-up mapping of the sentence. The linguistic mapping is *segmented*. Pantomime also appears to be global. A twirling downward pointing finger is understood to be a swizzle stick because the gesture, as a whole, has the meaning of a vortex. ASL signs are clearly segmented in their semiotic principles. The approbation meaning of the OK sign is not composed out of separately meaningful parts, but the precision image of the first finger–thumb contact is a component of the emblem, so it is probably correct to say that this gesture is analytic and segmented, in the sense that the meanings of at least some of the parts have an independent status.

Synthetic refers to the fact that a single gesticulation concentrates into one symbol distinct meanings that might be spread across the entire surface of the accompanying sentence. The single bends-it-back gesture displayed the actor, his action, and the path of the tree he acted upon. In the accompanying sentence, these same semantic functions were separated—"he," "bends," and "back"—and a meaning of extension was added ("way"). The mode of the sentence was analytic. Like English, ASL is also analytic. Emblems on the other hand are synthetic, like gesticulations. The OK meaning, bundled into one gesture, can spread over the full surface structure of a spoken equivalent ("a job well done," for example). Pantomime may also be analytic, though the lack of definition of pantomime makes the attribution uncertain. The twirl of a vortex gesture is the translation of a lexical term, which suggests analyticity. The issue of whether gesticulations characteristically map onto single lexical items or are not so constrained is a matter of some dispute (see Kita 2000, de Ruiter 2000, and Krauss et al. 2000). The issue is discussed in Chapter 2.

Summary of Differences along the Continua

Gesticulation accompanies speech, is nonconventionalized, is global and synthetic in mode of expression, and lacks languagelike properties of its own. The speech with which the gesticulation occurs, in contrast, is conventionalized, segmented, and analytic, and is fully possessed of linguistic properties. These contrasting modes of structuring meaning coexist in speech and gesture, a fact of importance for this book's understanding of thought and language in general, and how they function in communication.

Signs in ASL, like words in speech, are conventionalized, segmented, and analytic, and possessed of language properties, while they are obligatorily not performed with speech. The presence or absence of speech with gesture is thus correlated with the absence or presence of conventional linguistic properties.

Emblems are at an intermediate position on the various dimensions of contrasting gestures. They are partly like gesticulations, partly like signs. For many individuals, emblems are the closest thing to a sign 'language' they possess, although it is crucial to emphasize the *non*-linguistic character of these gestures: they lack a fully contrastive system (there is no contrasting 'partly OK' gesture made with the second finger and thumb, for example) and syntactic potential (the impossibility of combining two emblems into a gesture sentence).

Pointing requires special mention. It has a form that is standardized within a given culture. Thus in North America, the standard is the G handshape (the index finger extended, the other fingers curled in), but in other societies one might use two fingers or an entire hand, or the face, lips, nose, etc. (cf. Enfield 2001). At the same time, however, the prescribed form is not required for the pointing act to take place—in appropriate circumstances, the two-fingers or full-hand alternatives (and even lip or nose pointing) would be understood in North America. Thus, pointing is less constrained than the OK sign. In some contexts the presence of speech with pointing may seem obligatory (de Ruiter 2000) but pointing is fully interpretable without speech as well. Haviland (2000) analyzes the semiotic foundations of pointing and describes in detail its use in the Guugu Yimithirr Aboriginal culture in Australia and the Tzotzil Mayan culture in Mexico, where it is an integral part of linguistic performance—in one case because the coding of directionality is obligatory in the language, in the other for the opposite reason, because such coding is outside the normal resources of the language.

HOW GESTURES HAVE BEEN REGARDED: SOME HISTORY

The next step in this introduction is to set this book in a framework of past understandings of gesture and what gesture does. Recent study has emerged via a twofold shift away from a tradition that dates to Roman times, when the whole emphasis was on rhetorical gestures—the mannered performances of orators with the hands and body comprising more or less deliberate gestured embellishments on spoken public performances (Quintilian wrote a treatise on gesture for the instruction of orators in first-century Rome; see a 1977 reprint, in Latin; for a history of gesture in the theater and many useful references, see Bremmer & Roodenburg 1991).

With the first shift, commencing with Efron (1941) in the 1930s, gestures have come to be studied in life, as they occur spontaneously during conversation and other discourse modes. This new approach has been greatly enhanced, one might say made possible, by the advent of slow-motion film and now video, without which the careful study of gesture in relation to speech and thought would scarcely be possible. All aspects of this book draw from modern audio-video recordings.

In the second shift, commencing with Kendon in 1972 and continuing with ever increasing vigor into the present day, gestures are regarded as parts of *language itself*—not as embellishments or elaborations, but as integral parts of the processes of language and its use. The development of this line offers new insights into the nature of speaking, thinking, remembering, and interacting with words in a social context. This book takes the point of view that language and gesture are integral parts of a whole and regards this multimodal unit as language itself.

For histories of gesture study, see, in addition to Bremmer & Roodenburg (1991), Kendon (1981) and Müller (1998, 2000). There is also a brief survey in McNeill (1992). An interesting discussion of gestures as inferred from classical texts and material artifacts is in Boegehold (1999). A major publishing event is Kendon's translation and introduction to Andrea de Jorio's *La mimica degli antichi investigata nel gestire napoletano* (Gesture in Naples and gesture in classical antiquity, 1832) (Kendon 2000). De Jorio, an antiquarian and archeologist in early nineteenth-century Naples, was greatly struck by the similarities of the gestures he saw around him in daily Neapolitan life and the gestures he found depicted in the wall paintings and other illustrative materials in the

then recently excavated ruins of Pompeii and Herculaneum. He collected examples of current gesture use and organized them into a kind lexicographic compendium, together with a commentary whose foresight and modernity Kendon celebrates in his introduction.

The following sketch presents a few highlights from Cornelia Müller's (1998, 2000) account of the history of gesture study. She identifies five themes that have emerged throughout the two-thousand-plus-year history of commentary on gesture. Behind the themes are deeper underlying beliefs and attitudes, a number of which are still with us today.

First to appear is the 'domestication' of gesture and speech. Writers since antiquity have been concerned to suppress extensive and spontaneous gesture use. The underlying belief has been that gestures compete with language and therefore must be controlled. We shall see in the next chapter that, far from competing with language, gesture and language rise and fall in complexity and fluency *together*—this ancient belief, on current evidence, is mistaken.

Second, and almost as ancient, is the prescription of public and monologic gestures for oratorical use. The Quintilian text exemplifies this tradition. Again, there is behind the theme the belief that gesture and language compete; thus gesture must be regulated and defined prescriptively. Modern political speakers seem to be coached in the same vein, although in much simpler terms than their Roman counterparts. As recently as the nineteenth century, there were still elaborate expositions on the public performance of gestures and what gestures would be appropriate during speeches and sermons (Austin 1806, for example, and before him in the seventeenth century Bulwer 1974 wrote such guides for English-speaking audiences). This interest has declined in modern times but the ancient yet erroneous belief in language-gesture competition remains.

Third, since the Renaissance, there has been an interest in the private and dialogic use of gestures in conversation. This interest arose out of books on manners, and again there was a strong prescriptive component. The purpose was to state the things to do and not do in polite company (such as not losing control of your hands while speaking). However, rather than reject gesture out of hand, a concept of a suitably *restrained* gesture led to its use as an expression of education and nobility.

Fourth, dating from ancient times but becoming significant in the Enlightenment, philosophers such as Condillac and Diderot invoked the idea of gesture as the natural or original language of man (see Harris & Taylor 1989).

This idea has echoes in modern times in the gesture-first theory of language origins (e.g., Corballis 2002), which I consider at the end of the book.

Fifth, and last to arise, is the study of gestures in everyday speech. De Jorio pioneered this approach, along with much else, but it was not until the current wave of interest in gesture study that it has truly blossomed.

There was a long hiatus in the twentieth century during which neither psychology nor linguistics saw gesture as a subject worthy of study, or of particular interest if they did stumble across it. I suspect that in both psychology and linguistics behaviorism was to blame for this long dry spell. Wundt discussed gesture in detail at the start of the twentieth century (Wundt 1970 [1912]) but after this strong beginning from the 'father of experimental psychology', curiosity and discussion simply disappeared for sixty years. It then reemerged in the 1970s and has been moving forward with accelerating energy ever since. Gesture studies is a developing new field at the intersection of the humanities, linguistics, psychology, social science, neuropsychology, and computer engineering/computer science. Young researchers across a wide range of disciplines are drawn to the topic. There are a new journal, *Gesture* (edited by Cornelia Müller and Adam Kendon and published by Benjamins), and a new scholarly society (the International Society of Gesture Studies, of which Jürgen Streeck is the first president) with international conferences (the first held in Austin in 2002, and the second currently planned for Lyon in 2005). Wundt's dream, a century later, is at last coming true.

THIS BOOK

The main theme of this book was stated at the start of the chapter and can be briefly repeated now at the end. It is that language is inseparable from imagery. The imagery in question is embodied in the gestures that universally and automatically occur with speech. Such gestures are a necessary component of speaking and thinking. The goal of the book is to elaborate this argument and explore it in depth. Most importantly, the book describes a mechanism of language-gesture integration, the growth point, and the moment-by-moment thinking that takes place as one speaks. The mechanism of this immediate ongoing thought process is that of an imagery-language dialectic, an intrinsically dynamic process. The concept of the growth point, in turn, enables us to branch out in a new way to the neuropsychology of gesture and language, including consideration of how the brain circuits creating growth points might have

evolved. The resulting evolution model explains why, ultimately, language and imagery *are* inseparable: a joint system with these two components was part of the evolutionary selection of the human brain. I shall present evidence that languages have their own forms of dialectic, and that comparing speech and gesture across languages reveals partially distinct patterns of thinking for speaking. The important point for now is that the two frameworks of gesture and categorial content coexist and have the potential or, better, the *necessity* to combine.

The Chapters

The argument of the book is laid out in the following plan:

Chapter 1. Why gesture? The concept of language being inseparable from imagery; types of gesture on Kendon's continuum, revised to Kendon's continua; a brief history of gesture study; a précis of the argument.

Chapter 2. How gestures carry meaning. Speech-gesture synchrony and co-expressiveness; unbreakable binding; gesture anatomy; concept of the lexical affiliate; kinds of gesture, metaphor, and morphology; the question of who benefits from gesture; and the meaning of 'image'.

Chapter 3. Two dimensions. Contrasting traditions—the static, or Saussurian, and the dynamic, or Vygotskian—and their modern continuations.

Chapter 4. Imagery-language dialectic. Four subchapters:

4.1. Dialectic and material carriers. Vygotsky's concept of signs as material carriers of meaning applied to Merleau-Ponty, the 'H-model', Werner and Kaplan's organismic foundations, embodied cognition, etc.; gesture as such a material carrier of meaning.

4.2. The growth point as a model of the dialectic; inseparability from context (and how to portray it).

4.3. Extensions of GP to other languages and to different conditions of speaking, such as gesture-speech asynchronies and mismatches.

4.4. Social-interactive content. Analysis of a critical juncture in a conversation; the social shaping of growth points; gestures and growth point mimicry.

Chapter 5. Discourse. The catchment and discourse levels in single gestures.

Chapter 6. Children and Whorf. Ontogenesis of imagery-language dialectics; the shaping of visuospatial thinking by language.

Chapter 7. Neurogesture. Brain model and the orchestration of the brain for language and gesture; Broca's aphasia, Wernicke's aphasia, right hemisphere injury, and the split-brain.

Chapter 8. The thought-language-hand link and language origins. A case study (IW), and uncovering a dedicated thought-language-hand link in the human brain; the possible evolution of this link as a way of orchestrating brain action via mirror neurons and 'Mead's loop'; scenarios and time line for this selection.

In addition, the appendix describes methods for the collection, transcription, and coding of gestures.

Synopsis

For those who profit from a capsule summary, I have prepared the following précis. For those who prefer to plunge right in, this section can be skipped without loss; all its content, and much more, will unfold over the next seven chapters. However, the text in these chapters—Part 2 especially—is not, I realize, light reading. I have tried to write comprehensively and clearly, but the ideas are what they are and, when complex and novel, can be daunting. This précis can perhaps soften the blow.

Linguistics traditionally has focused on what can be termed the static dimension of language. In this tradition, language is regarded as an object, not a process. From Saussure on, the assumptions at the foundation of linguistics have been synchronic, and the goal has been to uncover the organization of the system of *langue*, or (in its modern version, with changes) competence. The static is a genuine dimension of language, experienced by speakers and listeners most directly through intuitions of well-formedness. However, there is another dimension, the dynamic. In this tradition, language is regarded as a process, not a thing. The figure most clearly exemplifying the dynamic approach is Vygotsky, but there are other historical figures as well—Wundt, Werner, Saussure himself at the end of his life–and the phenomenological tradition, especially Merleau-Ponty. The dynamic is also a real dimension of language, and gestures are a special route to accessing it.

I treat the static and dynamic therefore as real dimensions. Both are valid and necessary for a complete understanding of what Saussure called *langage*, or human language. Some phenomena are more accessible on one dimension, some on the other—linguistic forms, clearly, inhabit the static dimension but the linkage of language to discourse, narrative, and the focal consciousness of the speaker are phenomena more visible on the dynamic dimension. The conceptual challenge, after identifying the two dimensions, is to see how

they combine in the cognitive and linguistic performance of speakers and hearers.

My starting point, as noted earlier, is that language appears to be inseparable from imagery. The imagery in question is 'gesture'. The first step, in this chapter, has been clarification of what kinds of gestures these are and how to characterize them. In Chapter 2, I describe how gestures carry meanings. Then, in Chapter 3, I analyze the two dimensions of language. In Chapter 4, I introduce the main concept for combining the static and dynamic, that of a dialectic, an idea first proposed by Vygotsky. Vygotsky lacked the critical data that modern gesture study provides and the dialectic model was accordingly not completed by him.

This dialectic is analyzed in various ways in the subparts of Chapter 4, but the chief approach is in terms of 'growth points'. A growth point, or GP, is a minimal unit of dialectic in which imagery and linguistic content are combined. A GP contains opposite semiotic modes of meaning capture—instantaneous, global, nonhierarchical imagery with temporally sequential, segmented, and hierarchical language. The GP is a unit with demonstrable self-binding power (attempts to disrupt it, for example, with delayed auditory feedback do not succeed), and the opposition of semiotic modes within it fuels the dialectic. The key to the dialectic is that the two modes are simultaneously active in the mental experience of the speaker. Simultaneously representing the same idea unit in opposite modes creates instability, a 'benevolent instability' that is resolved by accessing forms on the static dimension—constructions and lexical choices, states of repose par excellence. The GP and the unpacking of it into constructions and lexical items is how the two dimensions of language combine—the unstable growth point, itself a combination of imagery and linguistic content, is unpacked into an increasingly well-formed, hence increasingly stable, structure on the static dimension. This process continues until, eventually, a 'stop order' occurs (it stops only temporarily: a new cycle begins immediately or might overlap the earlier one). A stop order is an intuitively complete (or complete enough) static structure (intuitions of well-formedness being how one experiences the static dimension). Thus gesture and the imagery it embodies are an integral step on the way to a sentence.

Images vary materially from no apparent gesture at all to elaborate multidimensional displays; but, hypothetically, imagery is ever present. What varies is the amount of materialization. Materialization runs from little to much, depending on the predictability/continuity of the specific GP with its context

of speaking. The concept of a 'material carrier' of meaning is also taken from Vygotsky. It is extended via the 'H-model', after Heidegger, according to which the materialization of one's meaning in a gesture (and speech) is, for the one speaking, not a representation but an updating the speaker's momentary state of cognitive being. The greater this contribution, the more the materialization, hence the more developed the gesture. The listener, in turn, inhabits the same meaning by updating in parallel his or her own momentary being, communication being a matter not only of signal exchange but of social resonance and inhabitance in the same 'house of being'.

Context, a notoriously slippery concept, is conceptualized as a field of oppositions. The more unpredictable/discontinuous the growth point in its field of oppositions is, the more elaborately materialized it tends to be in both speech and gesture; conversely, the more predictable/continuous the GP, the less the materialization, ultimately down to no gesture at all.

The GP model is congruent with Wundt, who, in a famous statement given in full in Chapter 4.1, emphasized that "the sentence is both a simultaneous and a sequential structure" (Wundt 1970, p. 21). In Chapter 4.2, I develop a detailed case study of a GP that, among other things, shows how a GP analysis incorporates the context of speaking as an integral part of the process of forming and unpacking growth points, and this leads to a way of modeling context and bringing it into a systematic relationship to growth point formation.

In all the above respects, this analysis is profoundly unlike the typical approach of psycholinguistics, which I regard as far too static in its assumptions and goals. The incorporation of context by GPs is one salient difference. Whereas context is integrated into the dialectic theory, it is the opposite—an external input—in these psycholinguistic models. Context is incorporated into the dialectic model automatically, because each growth point is what Vygotsky termed a psychological predicate—a point of differentiation of newsworthy content from a background; the background, or context, is an integral part of the growth point, without which it does not exist. Context is formalized in the GP model, as mentioned, as a field of oppositions. It is empirically recovered via 'catchments'—thematic discourse units realized in an observable thread of recurring gestural imagery. How, in GPs, idea units incorporate context, in the form of fields of oppositions, is one the central themes of the book.

I apply this theoretical framework to a very wide range of situations—discourse and gestures in different languages (Turkish, Spanish, and Mandarin, as well as my primary source, English); the gestures of children at the earliest

stages of development; the Whorfian hypothesis, arguing that the impact of language on imagery is often a dynamic dimension effect that has been concealed by concentration on the static dimension; linguistic impairments (aphasia; right-hemisphere damage, which impairs discourse cohesion; and the split-brain state, all of which were described in *Hand and Mind* but are now integrated into a new neurogestural model). An important new source of observations is the case of IW. This is a man who suffered, as a young adult, complete deafferentation from the neck down and has relearned movement control, including gestures with speech, which he can do to perfection even without vision, a condition where it is difficult for him to carry out nongesture actions. His case suggests a partial dissociation in the brain of the organization of gesture from the organization of instrumental action and the existence of a dedicated thought-language-hand link that would be the common heritage of all humankind.

The book accordingly ends with an attempt to provide 'the ultimate answer' to the question of an imagery-language dialectic and why it exists at all, by proposing a theory of language evolution, focusing on this thought-language-hand link. I develop the hypothesis that what made us human crucially depended at one point on gestures. Without gestures, according to this hypothesis, language could not have evolved; some of the brain circuits required for language could not have evolved in the way they apparently have. The integration of gesture with language we observe in ourselves today is an essential part of the machinery that evolved. Gesture is not a behavioral fossil but an indispensable part of our current ongoing system of language. The theory is based on the IW case and the neurogestural model and employs recently discovered 'mirror neurons', supplemented with what I am calling Mead's loop, to explain how and under what conditions a thought-language-hand link could have evolved. According to the Mead's loop theory, what had to be selected is a capacity, not present in other primate brains, for the mirror neuron circuit to respond to one's own gestures as if they belonged to someone else (this produces the apparent social framing of gestures, even when they are invisible—talking on the phone, a blind person talking to another blind person). This reconfiguration of circuitry provided the thought-language-hand link and a way for language to co-opt the machinery of Broca's area. Contrary to the gesture-first theory, a model that has become popular with Corballis (2002), I am arguing that evolution selected the ability to *combine* speech and gesture under a meaning, and that speech and gesture emerged in evolution together. This combination was the essential property evolution chose; there would not have been a gesture-first step. Just as speech could not have evolved without gesture, gesture could not have evolved without speech.

TERMINOLOGICAL TANGO

Saying (as I do) both that gesture is 'part' of language and also that there is a 'language-gesture' dialectic led one reviewer of the book in manuscript to suggest that I was saying something absurd—that a part can be in a dialectic with its own whole. This is not what I mean, of course, but my locutions could be open to such a construal. The problem is terminological; the word 'language' is being used in two ways. It would be tedious to signal the shifts between them, so I use this section to call the reader's attention to the double usage and trust that so flagging it will suffice to keep the two senses apart. Nothing hinges on the ambiguity, and allowing it to stand seems harmless. First, I use 'language' in a technical-linguistic way, to refer to those static structures of language consisting of grammar, words, etc. In this sense, there is a 'language-imagery' or 'language-gesture' dialectic, implying that the dynamic meets the static in this process. Second, I also use 'language' in a traditional nontechnical way, to indicate what it is we know when we say we 'know a language' or what we 'use' when we speak, listen, read, etc. In this way, gesture is 'part' of language, implying that language consists of more than words, sentences, etc., and also includes spontaneous, speech-synchronized gestures. The reviewer suggested the term 'grammar' for the technical sense, but 'grammar' is too narrow. The static dimension includes many levels besides 'grammar', and such a term risks making nonsense of the second sense, since gesture is certainly not 'part' of grammar (it is linked to the context of speaking in ways that grammar, because it depends on repeatability, cannot capture). 'Speech' is also too narrow, and implies an unwanted division of speech and sign, contradicting what I regard as the deep similarities in their underlying dynamics (see Chapter 4.3). So I shall simply request that the reader tolerate and move along with the ambiguity—when a 'dialectic of imagery and language' is meant, language is understood in the technical sense; when 'gesture is part of language', it is in the traditional sense.