1 Introduction – gesture and the origin of language

It's like seeing someone's thought.
– Mitsuko Iriye, historian, on learning to code gestures (in the 1980s)

1.1 THE ORIGIN OF LANGUAGE

The origin of language, a prodigal topic, has recently returned to respectability after a long exile.[1] Discoveries in linguistics, brain science, primate studies, children's development, and elsewhere have inspired new interest after the infamous nineteenth-century ban (actually, bans) on the topic – both the Société de Linguistique de Paris in 1865 and the Philological Society of London in 1873 prohibiting all contributions on the topic (London promising that any such would be tossed directly into the wastebasket; all of this described in Kendon 1991). The topic can be approached from many angles. Most common seems to be the comparative – differences and resemblances between humans and other primates. A related approach is to consider the brain mechanisms underlying communicative vocalizations and/or gesture. These have been recorded directly in some primate species and can be compared to humans on performance measures thought to depend on similar brain mechanisms. Or a linguistic angle – the key features of human language and whether anything can be said of how they came to be and whether other animal species show plausible counterparts. Approaches are combined in comparing human language to vocalizations, gestures, and/or the instructed sign language use of, say, orangutans or chimps. I will take a third approach, gestures, which also has its devotees, but I shall diverge from other approaches in crucial ways. I am not endorsing a popular current theory, appearing over and over in a veritable avalanche of recent books – what I dub "gesture-first." Despite the theory's name, the primatologist, neuroscientist, developmental psychologist, anthropologist, sign-language linguist, regular linguist, computer scientist, etc. proponents of gesture-first seemingly lack any serious acquaintance with gesture other than (it appears) its folk culture portrayals (so they do not recognize a key point of this book:

that language is misconstrued if it is not seen as a unity of language and gesture).

Gesture-first holds that the initial form of language lacked speech – it was a pantomimic or a sign language. I show that gesture-first (to put it delicately) is unlikely to be true because it is unable to capture the connections of speech and gesture that we, living counter-examples, display: it "predicts" what did not evolve (that gesture withered or was marginalized when speech arose) and does not predict what did evolve (that gesture is an integral part of speaking). A theory that says what didn't happen did, and what did happen didn't, can't generally be true, to say the least. That so many have adopted it I explain by the above-mentioned folk (and fabricated) beliefs about gestures.

The origins question homes in on what makes us human; how we diverge from other animal species, including our near neighbors, the Great Apes; it exposes in a fundamental way what comprises the gift of language. The approach here will ultimately synthesize various approaches to the question that modern authors have pioneered, not out of an urge to be all-encompassing but because these approaches will find a place in this approach's own inner logic.

The origin of language brought forth not only language but also new forms of action, new modes of thought, and new structures in the human brain. And these changes in action, thought, and brain are the sources of much else without which history, culture, and the human story could not have unfolded as they have. I hope by the end to clarify this story, how it began in unexpected ways, and on what it depends at a foundational level.

1.1.1 How this book differs

The approach here is to uncover the kind of mind that made the origin of language possible; and correspondingly, the kind of mind that language, once started, modified and extended. Other approaches emphasize the external aspects of the origin – communication, structure, parallels to other animal communication, all of which are valid but do not attempt to uncover the mindset of the creatures in which language came to be. My guiding idea and fundamental divergence is the following, proffered as an insight into the human mindset for language in general: Gestures are components of speech, not accompaniments but actually integral parts of it. Much evidence supports this idea, but its full implications have not always been recognized. The *growth point* (GP) hypothesis is designed to explicate this integral

linkage. It is presented briefly here, more fully with an example below, and explained in detail in Chapter 2.

Gestures offer one kind of symbol, language a different kind, and the two kinds of semiosis ("semiosis" and "semiotic" refer to the nature of symbols) are unified in GPs; in a GP symbols of these two different orders combine. A key insight is that speech on the one hand and gesture on the other, when combined in a GP, bring semiotically opposite modes of thinking together at the same time. A GP thus forms a single mental package or idea unit out of semiotically unlike components.

This "unity of opposites," as I will call it, creates a new form of human cognition that animates language and gives it a dynamic dimension. The semiotic opposition in a GP is intrinsically unstable; it seeks a resting place. The instability and the processes initiated to stabilize or resolve it, which I call "unpacking the GP," propel thought and speech forward, hence provide a dynamic dimension of language. All of these features of language were built in by how language began.

1.2 WHAT IS "GESTURE"?

1.2.1 Definitions of "gesture"

Gesture plays a central role in the arguments of this book. It is taken seriously and I need to explain what I mean when I refer to it. I cannot deny that the word is problematic. A journalist's cliché portrays a gesture as trivial, irrelevant, and slightly contemptible. It uses "gesture" to label something that a public figure, a politician or a magnate, has done as ungenuine and feckless; as sterile, futile, pointless, unfruitful, and untruthful, made for show and not effect. The cliché is worse than irrelevant. It positively obstructs understanding. Given the word's ragged appearance I would have preferred not to use it at all but there is no avoiding it; a suitable alternative simply does not exist in our language. I once concocted a term, "temaniosis," made from a Japanese root for "imitation in the hand"[2] and a Greek suffix for "of or relating to, of the nature of," which I thought would get close to the sense of "gesture" that I am using – but discovered that it is a combination so broad linguistically, exact though it is, that it offends some readers' sensibilities. And in any case it is vain to invent a word that will not gain general currency, and I judged there was little hope of that.

Adam Kendon (2004) placed gestures in the category of "actions that have the features of manifest deliberate expressiveness." I adopt this

definition but with one qualification and one proviso. The qualification is that gesture cannot be deliberate; as we regard them "gestures" are unwitting and automatic, and anything but deliberate. (Kendon may have meant by "deliberate" non-accidental, and with this I agree; but the word also conveys, "done for a purpose," and with that I do not agree.) The proviso concerns "action." In the sense that we intend (gesture as a special or what I later call a "new" kind of action) movements are orchestrated by significances created by the speaker him- or herself, not movements to attain external goals (goals lead to practical actions, not gestures). So our definition, based on Kendon's but excising "deliberate" and specifying the kind of action (and far from tripping off the tongue), is this:

A gesture is an unwitting, non-goal-directed action orchestrated by speaker-created significances, having features of manifest expressiveness.

Very often I use "gesture" still more restrictively to mean all of the above, plus:

A gesture is a manifestly expressive action that enacts imagery (not necessarily by the hands or hands alone) and is generated as part of the process of speaking.

1.3 THE GESTURE CONTINUUM

The remainder of this chapter is organized around Figure 1.1, The Gesture Continuum, a continuum of manifest expressiveness modes, all differing but all termed "gestures," annotated to show where these definitions and other important concepts apply. Later, I give examples of the Continuum and describe in detail certain features of it, especially the gesticulation pole – the focus of this book. The Gesture Continuum plays an important role as well in sorting out different explanations of the beginning of language, as Chapter 3 explains.

To begin, as the Continuum shows, the word "gesture" is problematic not only because of the ragged aspects but also because it is seriously ambiguous. It covers very different phenomena. The gestures of concern to us are integral components of speech, not substitutes, accompaniments or ornaments. Such gestures appear at one end of the Continuum, called by Kendon (1988b) "gesticulations."[3] These gestures are synchronous and co-expressive with speech, not redundant; and not signs, salutes, or what are called emblems (see below). They are by far the most frequent – in descriptive speech about 90% of utterances are accompanied by them

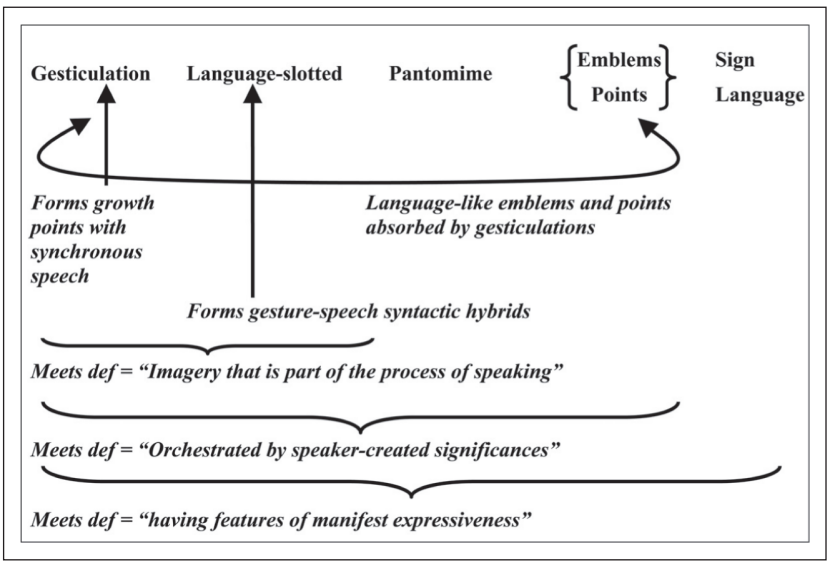

Figure 1.1 The Gesture Continuum annotated.

(Nobe 2000) and they occur in similar form and numbers across many languages.

1.3.1 Dimensions

Underlying the Continuum are three dimensions – how necessary speech is to the gesture; how language-like is the gesture; and how conventionalized is its form, so as one goes from gesticulation to sign the relationship of gesture to speech changes:

- The obligatory presence of speech decreases.
- Language-like properties increase.
- Socially regulated conventional signs replace self-generated form-meaning pairs.

1.3.2 Semiotic packages

We see the changes in how the positions along the Continuum form their own characteristic semiotic packages. At the gesticulation end (our concern) a *dual semiotic* prevails, imagery and linguistic encoding of the

same underlying idea in one package, the GP. At the language-slotted point a gesture is absorbed into its sentence, changing the relationship to language by losing co-expressivity and becoming *a constituent of the sentence itself*. At pantomime the semiosis is *reenactment*, and at the emblem/pointing point a gesture is *partly encoded* in itself. Finally, at the sign language pole gesture is *fully encoded* (cf. Klima and Bellugi 1979, Bechter 2009). The examples below illustrate these semiotic packages.

1.3.3 Timing

Also characteristic of each position is a different speech–gesture temporal arrangement (difficult to indicate in a linear layout). At the gesticulation end, the significant part of the gesture – the "stroke" – and its co-expressive speech are synchronous; at the language-slotted position, gesture slots into a vacancy in the sentence; at the pantomime and emblem/point position, gesture and speech have loose temporal relationships and speech may be completely absent. Sign language, finally, is freestanding and without speech.

1.3.4 Examples (from the most to the least language-like)

An example of an American Sign Language (ASL) *sign* is TREE – the dominant arm extended upward at the elbow, the fingers extended outward (a "5-hand") and the hand rotating back and forth at the wrist. The subordinate hand is extended flat and prone under the dominant hand elbow. The sign obviously depicts a schematic tree – trunk, leaves fluttering in the wind and the ground – but the iconicity is conventionalized and must include these specific features. A signer does not make up a new sign for each occasion. Arika Okrent (pers. comm.) calls it "non-specific," in that it is used equally for all kinds of trees and tree shapes, not just those with long bare trunks and fluttering leaves. This too is part of its conventionalization. Sign Languages such as American Sign Language of the deaf and others around the world are socioculturally maintained linguistic codes that have arisen where vocal/auditory communication is impossible. The most established are full language systems in their own right. While iconicity is present it too is conventionalized. The ASL sign is an iconic depiction, but it is a *standardized* selection of iconic features that other sign languages, also with signs that are iconic and regulated, may not use at all (Danish Sign Language traces an outline of a tree).

The *emblem* is the sort of gesture that appears in atlases and dictionaries of the "gesture language" of some nationality or other. And emblems indeed show systematic language-like constraints. There are differences between well-formed and not-well-formed ways of making them. Placing the middle finger on the thumb results in a gesture with some kind of precision meaning but it is not recognizable as the "OK" sign, where forefinger and thumb are in contact and the rest of the fingers spread out (see Figure 1.4). The "OK" gesture, like a word, is constrained to assume a certain "phonological" shape. Emblems also have culturally fixed meanings or functions (for the "OK" sign approval) and exclude or hide otherwise plausible meanings (such as precision). Yet, the constraints are limited and don't by any stretch amount to a full language. There is no way to reliably reverse "OK," for example. Forming it and waving it back and forth laterally (another emblem that, on its own, signals negation) might convey "not OK" but it also might equally be seen as the opposite – waving the hand to call attention to the sign, or to suggest that many different things are OK – a flexibility that is basically not linguistic in character. I will later place *pointing* with the emblems (§1.3.2).

Pantomime is an object- or action-simulation performed without speech, a dumb show (to cite the *OED*). It is a movement, often complex and sequential, that does not accompany speech but also is not part of a gesture code. If there is speech, the pantomime tends to appear during a brief pause or oppositely may extend well before, during and after the utterance; in other words, it is loosely timed or not timed at all with speech. A simple example is, in silence, moving the hand forward from the hip with pinched fingers and turning it, to depict taking a key out of the pocket and opening a lock. The same pantomime can be performed with speech (e.g., "there's only one thing to do"), and while speech and pantomime coincide they have no organic connection, quite unlike gesticulations. Pantomimes figure prominently in the discussion of gesture-first in Chapter 3.

Language-slotted gestures may look like pantomimes or gesticulations (the least language-like pole) but the distinguishing quality is how they combine with speech. They occupy a grammatical slot, become part of the syntax of the sentence, and acquire what Saussure (1959) called "syntagmatic" linguistic value (the value a word gains in combination with other words: how for example a noun becomes a "direct object" when combined with a verb, a value it does not have alone). An example is "the place was all [gesture suggesting uproar]," in which the uproar gesture has the syntagmatic value of a predicate adjective.

Gesticulation, in contrast to language-slotted gestures, is co-produced with speech. These gestures do not replace words in grammatical slots. In the

[/ and it **goe**s dOWn]

Both hands mirroring each other in tense spread C-shapes, palms toward center, move down from upper central periphery to lower central periphery; square brackets indicate a gesture phrase; boldface a stroke, underlining a hold, in which the shape and position of the stroke is maintained, and font size prosodic peaks; see Chapter 2 §2.2.1.1.2 for full details of notation.

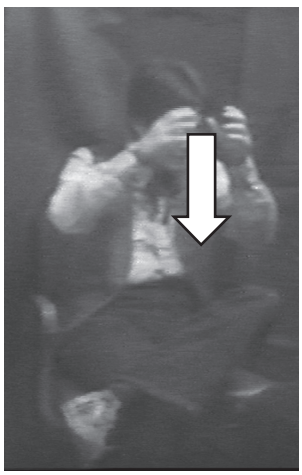

Figure 1.2 A speech-synchronized gesticulation.

narrations that provide most of the cases studied in this book, gesticulations are by far the most frequent type, occurring as often as one per second. Beats, iconics and metaphorics (all explained later) are at this gesticulation pole of the Continuum – in all cases, gesture and co-expressive speech are synchronized, gestures lack the language-like properties of recurrence and combination, and are unconstrained by conventional rules of gesture form.

Not only hand movements but the space in which the gesture occurs also can be called gesticulation. Where to make a gesture is in itself gestural and carries significance. For example, Figure 1.2 uses space iconically, the locus and direction of the gesture carrying information about the layout of the event being depicted (other gesture spaces are more metaphoric – putting one discourse theme in the left space and a contrasting theme in the right space, for example, which depicts in one layout the two themes and that they contrast).

1.4 DETAILS OF SELECTED POINTS

1.4.1 The emblem

An emblem is characterized by four related properties:

- First, an emblem is like a word of spoken language in that it is repeatable, listable, and reportable. However, unlike spoken language words,

emblems do not combine into larger units ("the" + "ball" forms a new unit, a noun phrase; that phrase plus "hit" forms another unit, a verb phrase, "hit the ball"; and that phrase plus "Ludwig" forms yet another unit, a sentence) and each of the new combinations is still a unit of the language: emblems do not have this combinatoric, hierarchic, recurrent property. One emblem, say "OK," followed by another emblem, say "no" (hand, palm forward, waving back and forth), may in some contexts look like "not OK," but the two emblems have not formed a larger emblem unit. It is, rather, still two emblems, first one, then another focused on it.

- Second and related, emblems have standards of good form. The "OK" sign must be made with the tips of the forefinger and thumb in contact, the other fingers more or less extended straight out. If some other finger makes contact it may be seen as a gesture of precision but it is not the "OK" sign. Whatever the historical origin of "OK," it must meet this form standard. I consider adherence to well-formedness as one half the hallmark of an emblem, such that violations result in rejecting the gesture, even though it is meaningful as a metaphor (precision in this instance). The other half-hallmark is having culturally specified functions (here approbation), another area that is standardized. The two hallmarks correspond to what Hockett and Altmann (1968) called "duality of patterning" – both signifier form (forefinger and thumb touching) and signified content (approbation) are "patterned" (regulated) linguistically and culturally.
- The third property, in keeping with these hallmarks, is that emblems are culturally defined and maintained. Kendon's 2004 book, *Gesture*, describes the Neapolitan gesture culture in detail. Every culture has a vocabulary of emblems, usually not nearly so developed as the Neapolitan, but everywhere emblems are culturally maintained symbolic forms with specified functions – again "OK" is a convenient example. Many emblems in North America seem to have Italian or even ancient Roman sources – "OK" is one, and there are less polite others (the favorite of the road-enraged, the "finger," is said to be Roman, Morris, *et al.* 1979; whether ancient Romans used it as such is not known but it is conceivable).
- The fourth property, having to do with sources, is that many emblems can be seen to be codified versions of metaphoric gestures (these being spontaneous gesticulations that present non-spatial, non-form meanings in terms of form and space) or metonymic gestures that present meanings in terms of something else that occurs with or causes the conveyed meaning. As "raw" forms these metaphoric or metonymic gestures appear in

discourse on their own. An emblem is often (I suspect always) a culturally specified version of such a metaphor or metonym, with form and meaning constrained by standards (Ishino 2007 has studied metonymy in Japanese gestures). For example, again using "OK," finger tip–finger tip contact as a metaphor of precision in the emblem takes only a certain form (forefinger/thumb contact) and a meaning (approbation). The metaphor source is seen in that the approbation is of a certain kind, that for something "just right"; furthermore, a different metaphor source is in "thumbs-up," a different approbation emblem, not the precision of "just-right" but the general metaphor of better is higher, "up on top," this the spatial locus the upturned thumb depicts iconically). This link of approbation to precision and on-top is not only in emblems but also appears in spoken forms, "precisely – that's it!" and "came out on top" (of an exam or contest).

1.4.2 Pointing

Traditionally the point does not have its own spot on the Gesture Continuum and, indeed, it is not obvious where to put it. Almost every complex gesticulation implies some deixis. The downward thrust of Figure 1.2 indicated the location of the pipe, its position relative to the character and the bowling ball. This deixis was accomplished not with a dedicated point but was built into the gesticulation itself. A dedicated, stand-alone point on the other hand has properties that make it like an emblem. First, points also have form standards – the extended index finger is standard in North American and Northern European culture; a flat hand is standard in some British Isle uses (Kendon 2004); and lip points are standard in Laos (Enfield 2001; see Figure 1.3). All have in common an iconic vector from the zero point, or "origo," to some target of the point.

Another similarity, less obvious, is how points and emblems relate to speech. While points and demonstrative pronouns ("this," "that" etc.) can synchronize (Levelt, Richardson and La Heij 1982), and thus appear to be like gesticulations, in fact the timing is different from that of gesticulation and more like that of an emblem. The similarity appears when gesture and speech are asynchronous. For both points and emblems asynchronies are meaningful, and are so in both directions. Asynchronous gesticulations, on the other hand, are merely slovenly and are not meaningful. Say "that" and then point; or point and then say "that"; or say and point simultaneously – each combination is meaningful and different (the meanings seem