1   Composite utterances

> A theory of utterance should not begin with a division between 'speech' and 'gesture'.
>
> <div align="right">Adam Kendon, 1986</div>

In human social behaviour, interactants build communicative sequences, move by move. These moves are never semiotically simple. Their composite nature is widely varied in kind: a word combined with other words, a string of words combined with an intonation contour, a diagram combined with a caption, an icon combined with another icon, a spoken utterance combined with a hand gesture. By what means does an interpreter take multiple signs and draw them together into unified, meaningful packages? This book explores the question with special reference to one of our most familiar types of move, the speech-with-gesture composite.[1]

This introductory chapter sketches a view of how it is that interpreters may derive meaning from composite utterances. This view of meaning has emerged from the empirical studies in Chapters 2–7, but is intended to have more general application to the analysis of any kind of communicative move, regardless of whether it involves speech, gesture, both, or neither.

1.1   Meaning does not begin with language

In a person's vast array of communicative tools, language is surely unrivalled in its expressive richness, speed, productivity, and ease. But the interpretation of linguistic signs is driven by broader principles, principles of rational cognition in social life, principles which underlie other processes of human

---

[1] Following from a long line of luminaries: De Jorio (2000 [1832]), Wundt (1973 [1921]), Efron (1972 [1941]), Goffman (1963), Condon and Ogston (1967), Ekman and Friesen (1969), Birdwhistell (1970), Kendon (1972, 2004), Slama-Cazucu (1976), Schegloff (1984), McNeill (1985, 2005), Calbris (1990), Haviland (1993, 2000), Streeck (1993, 1994), Goodwin (1994, 2006), Bavelas (1994), Engle (1998), Müller (1998), de Ruiter (2000), Beattie (2003), Goldin-Meadow (2003a), Liddell (2003), Kita and Özyürek (2003), Brookes (2004), Gullberg (2008), among very many others.

judgement, from house-buying to gambling to passing people on a crowded street. So, to understand meaning, we ought not begin with language (Enfield and Levinson 2006: 28). There is meaning in language for the same reason there is meaning elsewhere in our social lives: because we take signs to be public elements of cognitive processes (Peirce 1955), evidence of others' communicative intentions (Grice 1957, 1975). Our clues for figuring out those intentions are found not only in conventional symbols like words, but in the rich iconic–indexical relations which weave threads between just about everything in sight (Peirce 1955, Silverstein 1976, Levinson 1983, Kockelman 2005). Language is just a subset of the full resources necessary for recognizing others' communicative and informative intentions.

1.2   Meaning is dynamic, motivated, and concrete

Among fashions of thinking about language over the last century, a dominant neo-Saussurean view says that meaning is a representational relation of phonological form to conceptual content: a sign has meaning because it specifies a standing-for relation between a signifier and a signified. Semanticists of many stripes agree on this (cf. Jackendoff 1983, Cruse 1986, Langacker 1987, Wierzbicka 1996, among many others). But there is reason to question whether a view of signs as static, arbitrary, and abstract is an adequate depiction of the facts, or even optimal as an analytic framework of convenience. There is reason to stay closer to the source, to see signs as they are, first and foremost: dynamic, motivated, and concrete (Hanks 1990). Standard statements about meaning such as 'the word X means Y' really mean 'people who utter the word X are normatively taken by others to intend Y across a sufficiently broad range of contexts'. We should not, then, understand dichotomies like static versus dynamic, arbitrary versus motivated, or abstract versus concrete as merely two sides of a single coin. The relation is asymmetrical, since we are always anchored in the dynamic–motivated–concrete realm of contextualized communicative signs.

Some traditions doubt whether a Saussurean 'form–meaning mapping' account of meaning is appropriate. In research on co-speech hand gesture, for example, McNeill (2005) has forcefully questioned the adequacy of a coding-for-decoding model of communication. The same point has long been made for more general reasons, in more encompassing theories of semiosis, and in theories of how types of linguistic structure mean what they mean when used as tokens in context (Grice 1975). If we need alternatives to a static view of meaning, general tools are already available for addressing specific problems raised by co-speech gesture. These tools come from two sources: (neo-)Peircean semiotics (e.g. Peirce 1955, Colapietro 1989, Parmentier 1994, Kockelman

2005) and (neo-)Gricean pragmatics (e.g. Grice 1975, Levinson 1983, 2000, Sperber and Wilson 1995 [1986], Horn 1989, Atlas 2005). Subsequent sections explore the relevant analytic tools offered by these traditions.

1.3   Meaning is a composite notion

To set the stage, we anchor the discussion with a few examples of composite signs. Figure 1.1 shows a man kneeling, atop steps, with a crowd looking on.

While the kneeling posture may have an intrinsic, ethological basis for interpretation, this particular token of the behaviour has had a deeply enriched meaning for many who have seen it, because it was performed by this particular man, at this time and place. The man is Willy Brandt, chancellor of
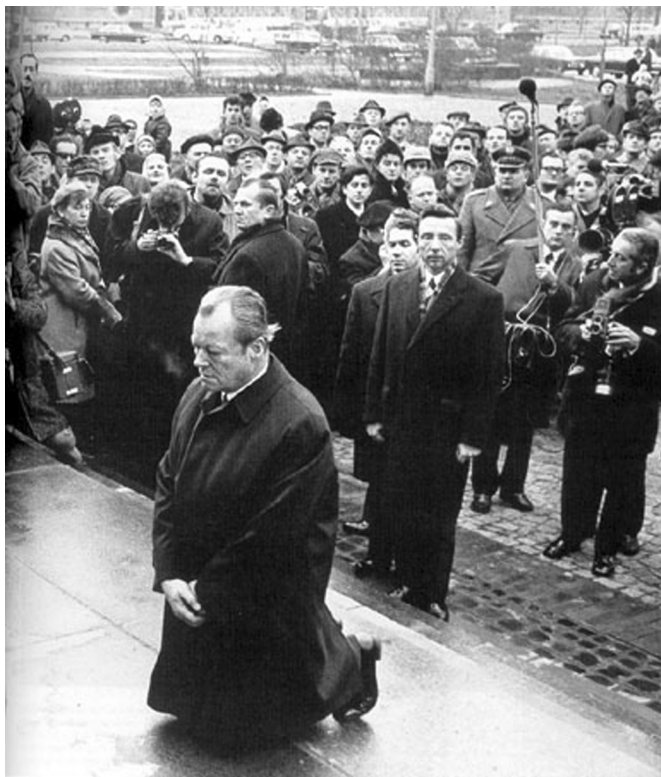

Figure 1.1 Man kneeling atop steps, with crowd looking on.

West Germany. Once you know just this, the act already begins to take on enriched meaning. It is not just a man kneeling, but a man whose actions will be taken to stand for those of a nation's people. It is 7 December 1970, a state visit to Warsaw, Poland. These new layers of information should yet further enrich your interpretation. To add another layer: the occasion is a commemoration of Jewish victims of the Warsaw Ghetto uprising of 1943. Brandt later described the moment: 'On the abyss of German history and carrying the burden of the millions who were murdered, I did what people do when words fail them.' The body posture illustrated in Figure 1.1 is a composite sign in so far as its meaning is partly a function of its co-occurrence with other signs: in particular, the role being played by its producer, given the circumstances of its time and place of production. The behaviour derives its meaning as much from its position on these coordinates as from its intrinsic significance. As Wittgenstein put it, 'Only when one knows the story does one know the significance of the picture' (Wittgenstein 1953: I–§663).

   Brandt's *Kniefall* is special partly because it was not accompanied by speech. Most composite utterances, including the speech-with-hand-movement utterances discussed in this book, do include a linguistic component. A relatively simple example of a composite sign with words is the image-with-

Figure 1.2 William Henry Fox Talbot, *Scene in a library*, 1845.

caption format typified by photographs and artwork, as in Figure 1.2. This photograph, titled *Scene in a library*, features wooden shelves with books on them. What makes this a composite sign is that the visual image and the string of words are taken together as part of the artist's single overall intention (Preissler and Bloom 2008, cf. Richert and Lillard 2002). The image and the words are different types of signs, but they are presented together, and taken together, in a composite.

As with any artwork's title, Talbot manipulates our attention to the image. Even if he had given the work a more directly descriptive title like *Books*, this would still invite us to attend differentially to what we actually see. A title *Books* would omit mention of the shelves, in line with the asymmetry in the image (the shelves are not visually foregrounded either). The title Talbot actually used – *Scene in a library* – does not narrow in on any part of the image, in fact it draws our attention to an imagined larger context which is not visible at all. We take the work to represent a scene in a library, and we trustingly presume the photograph to have actually been taken in a library, thanks to the verbal instruction embodied in the work's title. This presumption

Figure 1.3 William-Adolphe Bouguereau, *The wave*, 1896.

is easily made, under a general heuristic of semiotic unity: when encountering multiple signs which are presented together, take them as one.

This presumed bond of word and image is manipulated for wry effect in Figure 1.3.

Strangely resembling a pin-up from an auto magazine ostensibly featuring a new make of car, the title of this image makes a perverse claim about what is being presented. Despite the strong attention-directing force of the nude's blatant centrality in the image (cf. Clark *et al*. 1983), the composite sign's linguistic component directs our attention elsewhere. With the image, Bouguereau gives us a nude, but with the title he purports to depict not a nude, not even a nude by the sea, but a wave.[2]

These three examples illustrate essentially the same phenomenon as we find in the co-occurrence of expressive hand movements with speech: context-situated composites of multiple signs, part conventional, part non-conventional. Compare them to Figure 1.4, an image from a video-recording showing three Lao men sitting in a village temple, one of them thrusting his arm forward and down, with his gaze fixed on it.

The discussion is about construction works underway in the temple. The man on the left is reporting on a problem in the installation of drainage pipes from a bathroom block. He says that the drainage pipes have been fixed at too shallow an angle, and they should, instead, drop more sharply, to ensure good run-off. As he says *haj5 man2 san2 cang1 sii4* 'Make it steep like this', he thrusts his arm forward and down, fixing his gaze on it, as shown in Figure 1.4. The meanings of his words and his gesture are tightly linked, through at least three devices: (1) their tight spatiotemporal co-occurrence in place and time (both produced by the same source), (2) the use of the explicit deictic expression 'like this' (sending listeners on a search: 'Like what?'), (3) the use of eye gaze for directing attention.

A similar case is presented in Figure 1.5, from a description of a type of traditional Lao fish trap called the *sòòn5* (see Chapter 5).

Again we see a speaker's overall utterance meaning as a unified product of multiple sources of information: (a) a string of words (itself a composite sign consisting of words and grammatical constructions), (b) a two-handed gesture, (c) tight spatiotemporal co-occurrence of the words and gestures (from a single source), and (d) eye gaze directed toward the hands, also helping to connect the composite utterance's multiple parts. This is subtly different from Figure 1.4 in that it does not involve an explicit deictic element in the speech (cf. the overt 'like this' element in Figure 1.4 which obliges us to consult the

---

[2] He may of course also be inviting us to find qualities in common between a wave and the human figure depicted.

Figure 1.4 Man (left of image) speaking of preferred angle of a drainage pipe under construction: 'Make it steep like this.'

gesture to complete the utterance's meaning). Like the image-with-caption examples in Figures 1.1–1.3, spatiotemporal co-placement in Figure 1.5 is sufficient to signal semiotic unity. The gesture, gaze, and speech components of the utterance are taken together as a unified whole. As interpreters, we effortlessly integrate them as relating to one overall idea.[3]

A general theory of composite meaning takes Figures 1.1–1.5 to be instances of a single phenomenon: signs co-occurring with other signs, acquiring unified meaning through being interpreted as co-relevant parts of a single whole. A general account for how the meanings of multiple signs are unified in any one of these cases should apply to them all, along with many other species of composite sign, including co-occurring icons in street signs, grammatical unification of lexical items and constructions, and speech-with-gesture composites.

In studying speech-with-gesture, we should register two desiderata for an account of composite meaning. A first requirement is to provide a modality-

---

[3] As Freud argued, with many since, there may be leakage of unintended, apparently unrelated information, particularly through modalities over which a sender has less control.

Figure 1.5 Man describing the *sòòn5*, a traditional Lao fish trap: 'As for the *sòòn5*, they make it fluted at the mouth.' (See Chapter 5.)

independent account of 'gesture' (Okrent 2002). While we want to capture the intuition that co-speech hand gesture (manual–visual) conveys meaning somehow differently to speech (vocal–aural), this has to be articulated without reference to modality. We need to be able to say what makes speech-accompanying hand movements 'gestural' in such a way that we can sensibly ask as to the functional equivalent of co-speech gesture in other kinds of composite utterances; for example, in sign language of the Deaf (all visual, but not all 'gesture'), or in speech heard over the phone (all vocal–aural, but not all 'language').

A second desideratum for an account of meaning in speech-with-gesture composites is to capture the notion of 'holistic' meaning in hand gestures, the idea that a hand gesture has the meaning it has only because of the role it plays in the meaning of an utterance as a whole (McNeill 1992, 2005, Engle 1998). Consistent with an aim for analytic generality, I argue that a notion of holistic meaning is required not only for analysing the meaning of co-speech hand gesture, but more generally for analysing linguistic and other types of signs as well (including wordless moves like Brandt's *Kniefall*). This results from acknowledging that an interpreter's task begins with the recognition of a signer's communicative intention (i.e. recognizing that the signer has an informative intention). The subsequent quest to lock onto a target informative

intention can drive the understanding of the composite utterance's parts, and not necessarily the other way around.

1.4   The anatomy of meaning in composite utterances

1.4.1   Contexts of hand gesture

One view of speech-with-gesture composites is that the relation between co-expressive hand and word is a reciprocal one: 'the gestural component and the spoken component **interact** with one another to create a precise and vivid understanding' (Kendon 2004: 174, original emphasis; cf. Özyürek *et al.* 2007). By what mechanism does this reciprocal interaction between hand and word unfold? Different approaches to analysing meanings of co-speech gestures find evidence of a gesture's meaning in a range of sources, including (i) speech (coterminous) which co-occurs with the hand movement, (ii) a (prior) stimulus or cause of the utterance in which the gesture occurs, (iii) a (subsequent) response to, or effect of, the utterance, or (iv) purely formal characteristics of the gesture. These four sources (often combined) draw on different components of a single underlying model of the communicative move and its sequential context, where the hand-movement component of the composite utterance is contextualized from three angles: A. what just happened; B. what else is happening now; C. what happens next. This is illustrated in Figure 1.6.

The three-part sequential structure illustrated in Figure 1.6 underlies a basic trajectory model recognized by many students of human social behaviour. Schutz (1970), for example, speaks of actions (at B) having 'because motives' (at A) and 'in-order-to motives' (at C; e.g. 'I'm picking berries [B] because I'm hungry [A], in order to eat them [C]'; cf. Sacks 1992, Schegloff 2007b among many others).


```
        A                    B                    C

Stimulus/cause  →    ⌜ Speech           ⌝   → Response/effect
                     | Hand movement    |
                     ⌞ (other)          ⌟
```
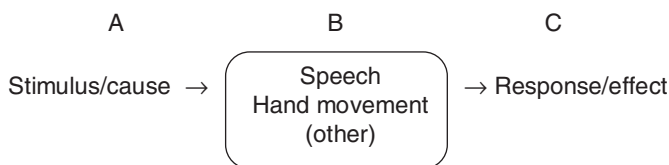
Figure 1.6 Three contexts of hand movement, in sequential interaction: at B, composite utterances may include multiple simultaneous signs; a preceding stimulus/cause at A determines a sign's **appropriateness**; a response determines its **effectiveness**.

1.4.2   Enchrony: an underlying ontology for the context
of composite utterances

The structure in Figure 1.6 directs our attention to an ontology of the composite utterance as a situated unit of social behaviour with causes (or conditions) and effects (Goffman 1964, Schegloff 1968). An intentional cause and interpretive effect are as definitive of the process of meaning as the pivotal signifying behaviour itself. Any communicative move may be seen as arising more or less appropriately from certain commitments and entitlements, and in turn bringing about new commitments and entitlements (Austin 1962, Searle 1969), for which interlocutors are subsequently accountable. As an analytical framework, this remedies the static, decontextualized nature of Saussure's version of meaning (Kockelman 2005). But this is not merely because it recognizes that meaning arises through a process (McNeill 2005), it is because it recognizes the causal/conditional and normative anatomy of sequences of communicative interaction, where each step brings about a new horizon, with consequences for the people involved (Schegloff 1968, Sacks *et al.* 1974, Goffman 1981, Heritage and Atkinson 1984). Accordingly, we need a term for a causal, dynamic perspective on language whose granularity matches the pace of our most experience-near, moment-by-moment deployment of utterances, not historical time (for which the term *diachronic* is standard) but conversational time. For this I invent the word *enchronic*.[4] While diachronic analysis is concerned with relations between data from different years (with no specified type or directness of causal/conditional relations), enchronic analysis is concerned with relations between data from neighbouring moments, adjacent units of behaviour in locally coherent communicative sequences (typically, conversations). McNeill (2005) uses *epigenesis* for the real-time birth and development of a composite utterance from a producer's point of view. This is distinct from the intended meaning of enchronic here, namely the intersection of (a) a social causal/conditionality of related signs in sequences of social interaction and (b) a particular level of temporal granularity in a conditionally sequential view of language: conversational time. An enchronic perspective adopts the sequential analytic approach whose application in empirical work was pioneered by Schegloff (1968) and Sacks (1992), following earlier work in sociology. To call it enchronic rather than merely sequential (in the technical sense of Schegloff 2007b) draws attention to the broader set of alternative viewpoints on

---

[4] This is an adjective, whose nominal form is *enchrony*. The prefix *en-* refers elsewhere to causal/conditional relations and to the notions of increment and change of state (e.g. *endear*, *enfold*, *enliven*, *enrich*, *encage*).

systems and processes of meaning which we often need to switch between (phylogenetic, diachronic, ontogenetic, synchronic).

1.4.3 The move: a basic-level unit for social interaction

An enchronic perspective takes as a primitive unit the communicative move (Goffman 1981). A move may be defined as a recognizable unit contribution of communicative behaviour constituting a single, complete pushing forward of an interactional sequence by means of making some relevant social action recognizable (e.g. requesting the salt, passing it, saying *Thanks*). In communication, a richly multimodal flux of impressions is brought to order by these joint-attentional pulses of addressed behaviour (e.g. bursts of talk) marked off in the flow of time and space, yielding sequences of co-contingent social action (Goodwin 2000a, Schegloff 2007b). The linguistic utterance is a well-studied (if idealized)[5] type of instantiation of the move (cf. Austin 1962, Searle 1969). With this basic-level status, the linguistic move will be homologous with usage-based analytic units of language such as the clause (Foley and Van Valin 1984), the intonation unit (Pawley and Syder 2000, Chafe 1994), the turn-constructional unit (Sacks *et al.* 1974), the growth point (McNeill 1992), the composite signal (Engle 1998, cf. Clark 1996), and the utterance as multimodal ensemble (Kendon 2004, Goodwin 2000a). But whatever its physical form, the move is a single-serve vehicle for effecting action socially.

An important argument in favour of the move's primitive or basic-level status is its role in the acquisition of communicative skills in children. Before learning their first words, children master the move, beginning with its prototype, the pointing gesture (Kita 2003). A line of research in developmental psychology has identified the onset of the pointing gesture as a watershed moment in the development of human social cognitive and communicative capacities, both ontogenetically and phylogenetically (Bates *et al.* 1975, 1987, Liszkowski *et al.* 2004, Tomasello 2006). The pointing gesture is mastered by prelinguistic infants (at around twelve months of age) and is the first type of move to unequivocally display the sort of shared intentionality unique to human communication and social cognition (Tomasello *et al.* 2005, Liszkowski 2006, Frith and Frith 2007).

The move is therefore a starting point, a seed, a template for the deployment of signs in interaction. On the one hand, the move is a brick for larger structures, building up and out, into conversational sequences and other kinds of coherent discourse structure (Halliday and Hasan 1976, Schegloff 2007b).

---

[5] I say it is idealized because there is always more than just language available for an interpreter.

On the other hand, it is a frame or exoskeleton within which internal semiotic complexity may appear, building down and in, yielding phrase distinctions, morphosyntax, information structure, and logical semantics. The work of this book is to examine the kinds of structure that arise when moves are built from word and hand together, and we shall witness both their internal elaboration (down-and-in) and their role in structuring higher-order sequences (up-and-out).

1.4.4   Conventional and non-conventional components of composite utterances

Three types of sign are important in interpreting composite utterances: conventional signs, non-conventional signs, and symbolic indexicals.[6] A conventional sign is found when people take a certain signifier to stand for a certain signified because that is what members of their community normatively do (Saussure 1959 [1916]).[7] This kind of sign allows for arbitrary relations like /kʰæt/ referring to 'cat', by which the cause of my taking [kʰæt] to mean 'cat' is my experience with previous occasions of use of tokens of the signifier /kʰæt/. Examples of conventional signs include words and grammatical constructions, idioms, and 'emblem' hand gestures such as the OK sign, V for victory, or the finger (Ekman and Friesen 1969, Brookes 2004). Non-conventional signs, by contrast, are found when people take certain signifiers to stand for certain signifieds not because of previous experience with that particular form–meaning pair or from social convention, but where the standing-for relation between form and meaning comes about by virtue of just that singular event of interpretation. Examples include representational hand gestures (in the sense of Kita 2000), that is, where the gesture component of an utterance is a token, analogue representation of its object.

   The symbolic indexical is a hybrid of the two types of sign just described, having properties of both.[8] These include anything that comes under the rubric of deixis (Fillmore 1997 [1971], Levinson 1983), that is, form–

---

[6] For convenience, I simplify the analysis of sign types employed here. A full anatomy of sign types would lay out the logical possibilities first mapped by Peirce (1955), and most accessibly interpreted by Parmentier (1994) and Kockelman (2005). The notion of conventional sign here corresponds to Peirce's symbol; non-conventional sign includes his icon and index. The Peircean type/token distinction (Hutton 1990) cuts across these (see below).

[7] By saying that a behaviour is normative, I mean that carrying out the behaviour under the appropriate conditions will be effective, and will not evince justified surprise or sanction (cf. Brandom 1979, Kockelman 2006).

[8] Because symbolic indexicals have both conventional and non-conventional components, statements about those kinds of signs also hold for the relevant components of symbolic indexicals.

meaning mappings whose proper interpretation depends partly on convention and partly on context (Bühler 1982 [1934], Jakobson 1971, Silverstein 1976). Take for example *him* in *Take a photo of him*. Your understanding of *him* will depend partly on your recognition of a conventional, context-independent meaning of the English form *him* (third person, singular, male, accusative) and partly on non-conventional facts unique to the speech event (e.g. whichever male referent is most salient given our current joint attention or common ground). Symbolic indexicals play a critical role in many types of composite utterance, since their job is to glue things together, including words, gestures, and (imagined) things in the world (see Part I of this book).

In the context of these three kinds of sign, it is important to be mindful of the distinction between type and token (Peirce 1955, Hutton 1990). All of the signs discussed above occur as tokens, that is, as perceptible, contextualized, unique instances. But only conventional signs (including conventional components of symbolic indexicals) necessarily have both type and token identities. That is, when they occur as tokens, they are **tokens of types**, or what Peirce called replicas. It is because of their abstract type identity that conventional signs can be regarded as meaningful independent of context, as having 'sense' (Frege 1960 [1892]), 'timeless meaning' (Grice 1989) or 'semantic invariance' (Wierzbicka 1985, 1996). Conventional signs are pre-fabricated signs, already signs by their very nature. They have 'I am an addressed, relevant sign' stamped on their foreheads. By contrast, non-conventional signs (including non-conventional components of symbolic indexicals) are tokens but not tokens of types. They are singularities (Kockelman 2005). They become signs only **when taken as signs in context**. This is the key to understanding the asymmetries we observe in composite utterances like speech-with-gesture ensembles. A hand gesture may be a conventional sign (e.g. as 'emblem'). Or it may be non-conventional, only becoming a sign because of how it is used in that context (e.g. as 'iconic' or 'metaphoric'). Or it may be a symbolic indexical (e.g. as pointing gesture, with conventionally recognizable form, but dependent on token context for referential resolution). Hand gestures are not at all unique in this regard: the linguistic component of an utterance may, similarly, be conventional (e.g. words, grammar), non-conventional (e.g. voice quality, sound stretches), or symbolic indexical (e.g. demonstratives like *yay* or *this*). Ditto for sign components of graphs, diagrams, and other illustrations. Sensory or articulatory modality is no obstacle to semiotic flexibility.

Before concluding this chapter, it is worthwhile registering a common inconsistency in discussion of the meaning of hand movements in composite utterances. The problem is an asymmetric treatment of the way meaning is attributed to words, on the one hand, and gestures, on the other. Linguistic

items like words are often described merely in terms of what they conventionally **encode** (as standing for lexical **types**), while gestures are typically described in terms of what they non-conventionally **convey** (as standing for utterance-level **tokens** of informative intention). In other words, the interpreter's problem of comprehending word meaning is taken to be one of **recognition** (from token form to type lexical entry), while the problem of comprehending gesture meaning is taken to be one of **interpretation** (from token form to token informative intention). The inconsistency here is that it overlooks the fact that comprehension of the linguistic component **also** involves interpretation, yielding token informative intentions. In interpreting the meanings of words, we do not stop with mere recognition of type lexical entries, but, just like with gestures, we **also** use them for recognizing a speaker's token informative intention. To illustrate, take an example cited by McNeill (2005: 26), in which a speaker says *and he came out the pipe* while doing an 'up-and-down away' hand gesture (the hand is moving away from the body as it is moved repeatedly up and down). Hearing *came out*, an interpreter recognizes these sounds to be tokens of types (i.e. with the meaning 'came out'). He or she may also enrich this meaning 'came out' in using it as a clue for figuring out the speaker's informative intention in producing this composite utterance. They may of course exploit the accompanying gesture in this process of enrichment. In the experiment described by McNeill, a subject who heard the first speaker's description of the scene as *and he came out the pipe*[GESTUREup-and-down-away] later re-describes it as *the cat **bounces** out the pipe*.[9] This shows that **both** the gesture and the words are enriched by their co-occurrence in that context, being taken to be co-occurring signs of a single informative intention. *Came out* and [GESTUREup-and-down-away] together point to a single idea 'bounces out'. While word recognition has no analogue in the interpretation of the iconic gesture (since the gesture is a token but not a token of a type), attribution of overall utterance-intention of words **does** have an analogue in the interpretation of the gesture. When examining gesture, as when examining any other component of composite utterances, we must carefully distinguish between token meaning (enriched, context-situated), type meaning (raw, context-independent, pre-packaged), and sheer form (no necessary meaning at all outside of a particular context in which it is taken to have meaning). These distinctions may apply to signs in any modality.

---

[9] Note that the re-teller not only enriches *came out*[**GESTURE** up-and-down-away] as 'bounces out', he also enriches *he* as 'the cat'; regarding the pronoun *he* in the original utterance, the subject must have both recognized *he* as a token of the type 'he', which stands in this case for a token informative intention 'the cat'.

I. Encoded

    I.1. Lexical (open class, symbolic)

    I.2. Grammatical (closed class, symbolic indexical)

II. Enriched

    II.1. Indexical resolution

        II.1.1 Explicit (via symbolic indexicals, e.g., pointing or

            demonstratives)

        II.1.2 Implicit (e.g., from physical situation)

    II.2. Implicature

        II.2.1 From code

        II.2.2 From context

Figure 1.7 Sources of composite meaning for interpretation of communicative moves. 'Encoded' = conventional sign components. 'Enriched' = non-conventional token meanings drawing on context.[10, 11]

1.4.5   Elements of composite utterances

Based on the discussion so far, we may define the composite utterance as a communicative move that incorporates multiple signs of multiple types. Sources of these types of sign are given in Figure 1.7 (cf. Levinson 1983: 14, 131, Hanks 1990: 51ff.).

Composite utterances are interpreted through the recognition and bringing together of these multiple signs under a pragmatic unity heuristic or co-relevance principle, i.e. an interpreter's steadfast presumption of pragmatic unity despite semiotic complexity.

---

[10] Encoded meaning encompasses both lexical and grammatical meaning. Grammatical signs show greater indexicality because they signify context-specific ties between two or more elements of a composite utterance (e.g. grammatical agreement, case-marking etc.) or between the speech event and a narrated event (Jakobson 1971; e.g. through tense-marking, spatial deixis etc.).

[11] Indexical enrichment refers to the resolution of reference left open either explicitly (e.g. through symbolic indexicals like *this*) or implicitly (e.g. by simple co-placement in space or time; thus, a 'no smoking' sign need not specify 'no smoking **here**'). Enrichment through implicature refers to Gricean token understandings, arising either through rational interpretation based on knowledge of a restricted system of code (i.e. informativeness scales and other mechanisms for Generalized Conversational Implicature; cf. Levinson 2000), or through rational interpretation based on cultural or personal common ground (e.g. Particularized Conversational Implicatures such as those based on a maxim of relevance; Sperber and Wilson 1995 [1986]).

1.5   Sign filtration: triggers and heuristics

The taxonomy of elements of composite signs in Figure 1.7 presupposes that an interpreter can solve the problem of sign filtration, i.e. that they can parse out from a flux of impressions those things that are to be taken as signs. This filtration is assisted by triggers which direct us to lock on to certain signs, constraining the search space. An important trigger is that a perceptible impression must be recognizable as addressed, that is, being produced by a person for the sake of its interpretation by another. Conventional signs like words have this addressed-ness by their very nature. But other perceptibles are only potential signs, and their addressed-ness needs to be specially marked. This can be achieved by means of attention-drawing indexicals (hand pointing, saying 'like this' etc.), by sheer spatiotemporal co-occurrence, or by special diacritic marking (see Figures 1.1–1.5, above). An example of the latter is discussed in Chapter 3, where movements of the face and head can serve as triggers for eye gaze to be interpreted as pointing, not merely as looking. In yet other cases, interpreters can employ abductive, rational interpretation to detect that an action is done with a communicative intention (Peirce 1955, Grice 1957). For instance, if you open a jar I may be unlikely to take this to be communicative, but if you carry out the same physical action without a jar in your hands, the lack of conceivable practical aim is likely to act as a trigger for implicature (Levinson 1983: 157, Gergely *et al.* 2002).

The data in this book do not present particular difficulties for interpreters in detecting communicative intention or identifying which signs to include when interpreting a composite utterance. Mostly, the mere fact of language being used triggers a process of interpretation, and the gestures which accompany speech are straightforwardly taken to be associated with what a speaker is saying (Kendon 2004). Hand gestures are therefore available for inclusion in a unified interpretation, whether or not we take them to be intended to communicate.

Note the kinds of heuristics that are likely being used in solving the problem of sign filtration. By a **convention heuristic**, if a form is recognizable as a socially conventionalized type of sign, assume it stands for its socially conventional meaning. Symbols like words may thus be considered as pre-fabricated semiotic processes: their very existence is due to their role in communication (unlike iconic–indexical relations which may exist in the absence of interpretants). By an **orientation heuristic**, if a signer is bodily oriented toward you, most obviously by body position and eye gaze, assume they are addressing you. By a **contextual association heuristic**, if two signs are contextually associated, assume they are part of one signifying action. Triggers for contextual association are timing and other types of indexical proximity (e.g. placing caption and picture together, placing word and gesture

together). By a **unified utterance–meaning heuristic**, assume that context-
ually associated signs point to a unified, single, addressed utterance–meaning.
And by an **agency heuristic**, if a signer has greater control over a behaviour,
assume (all things being equal) that this sign is more likely to have been
communicatively intended. Language scores higher than gesture on a range of
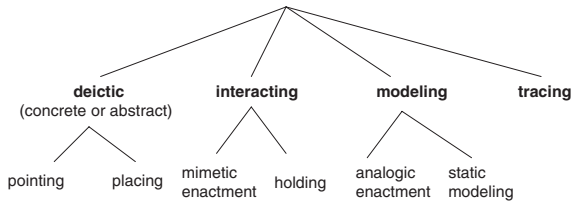measures of agency (Kockelman 2007).

1.6   Semiotic analysis of gestures

Like any signs, hand movements can stand for things in three essential ways
(often in combination), referred to by Peirce (1955) as types of ground:
iconic, indexical, symbolic. These crucial yet widely mishandled distinctions
are defined as follows. A relation of a sign standing for an object is iconic
when the sign is taken to stand for the object because it has perceptible
qualities in common with it. The sign is indexical when it is taken to stand for
an object because it has a relation of actual contiguity (spatial, temporal, or
causal) with that object. The relation is symbolic when the sign is taken to
stand for an object because of a norm in the community that this sign shall be
taken to stand for this object. These three types of ground are not exclusive,
but co-occur. A fingerprint on the murder weapon is iconic and indexical. It is
iconic in that it has qualities in common with the pattern on the killer's actual
fingertip and in this way it is a sign that can be taken to stand for the
fingertip. It is indexical in that (a) it was directly caused by the fingertip
making an impression on the weapon (thus a sign standing for an event of
handling it), and (b) the fingertip of the killer is in contiguity with the whole
killer (thus a sign standing for the killer himself). Standard taxonomies of
gesture types (McNeill 1992, Kendon 2004, *inter alia*) are fully explicable in
terms of these types of semiotic ground (Figure 1.8).[12]

   An exhaustive analysis of the semiotics of hand gestures will need to
systematically explore their values on the many parameters along which
signs differ: Formal segmentability, stability across populations, evanescence
or persistence in time from production, symmetry of perceptual access for
producer and interpreter, relative immediacy of the processes of production and
interpretation, portability, combinatorics, information structure (cf. Kockelman
2005: 240–241). This will entail teasing apart the large set of distinct
semiotic dimensions which hand movements incorporate (Talmy 2006).[13] Hand

---

[12] Figure 1.8 presents only some types of hand gesture of interest in this book. See references in
   note 1, above, for discussion of these and numerous other types such as 'beats', 'adaptors', and
   'quotable gestures'.

[13] The notion of semiotic dimension is distinct from that of sensory modality. Any sign medium
   within a single sensory modality may have numerous distinct semiotic dimensions, meaning
   any distinguishable, simultaneously variable perceptible component of a sign vehicle which

Deictic:

● semiotic function: *indexical* (in that the directional orientation of the gesture is determined by the conceived location of a referent), and *symbolic* (in that the form of pointing can be locally conventionalized); the hands are used to bring the referent and the attention of the addressee together;

- in *concrete deixis*, the referent is a physical entity in the speech situation, while in *abstract deixis* the referent is a reference-assigned chunk of space with stable coordinates
- in *pointing*, the attention of the addressee is directed to the referent by some vector-projecting articulator (such as the index finger or gaze)
- in *placing*, the referent is positioned for the attention of the addressee
 (Nb.: *Gaze* plays an important role in deictic gestures; it projects its own attention-directing vector which may (a) reinforce a deictic hand gesture by providing a second vector oriented towards the same referent, and (b) assist in the management of attention direction during production of other gestures.)

Interacting:

● semiotic function: *iconic* (in that the hands imitate an action) *and indexical* (in that the shape of the hands is not the shape of the referent, but is *determined by* the shape of the referent); the hands are meant to look as if they were interacting with the referent;

- in *mimetic enactment*, the hands are moving as if they are doing something to or with the referent
- in *holding*, the hands are shaped to look as if they are holding the referent

Modeling:

● semiotic function: *iconic*; the hands are meant to look as if they are the referent

- in *analogic enactment*, the hand's movement imitates the movement of the referent
- in *static modeling*, the hand's shape imitates the shape of the referent

Tracing:

● semiotic function: *iconic* (in that the gesture imitates drawing) *and indexical* (in that only part of the referent is depicted, but the whole is referred to); the hands (more specifically, the fingers) are meant to look as if they were tracing the shape of some salient feature of the referent, such as its outline.

Figure 1.8 Some semiotic devices used in illustrative co-speech gestures discussed in this book (cf. Mandel 1977, Kendon 1988, Müller 1998).

movements are well suited to iconic–indexical meaning thanks to their rich potential for sharing perceptible qualities in common with physical objects and events. But they are not at all confined to these types of meaning. As Wilkins writes, '[the] analog and suprasegmental or synthetic nature [of gestures] does not make them any less subject to convention, and does not deny them combinatorial constraints or rules of structural form' (Wilkins 2006: 132). For example, in some communities, 'the demonstration of the

could conceivably be taken to be a sign for something (de Ruiter *et al*. 2003). For example, upon uttering a word, the human voice can simultaneously vary many distinct features of a speaker's identity (sex, age, origin, state of arousal, individual identity etc.), along with pitch, loudness, among other things. What makes pitch and loudness distinct semiotic dimensions is that they can be varied independently of each other. But loudness is a single dimension, because it is impossible to produce a word simultaneously at two different volumes.

Figure 1.9 Man shows the size of a fish using his forearm as a measure. Note his eye gaze as a cue for the communicative relevance of what his hand is doing (cf. Figures 1.4, 1.5, above)

length of something with two outstretched hands may require a flat hand for the length of objects with volume (like a beam of wood) and the extended index fingers for the length of essentially linear objects lacking significant volume (e.g. string or wire)' (ibid.). A similar example is the Lao speaker's conventional way of talking about sizes of fish, illustrated in Figure 1.9, by using the hand or hands to encircle a cross-section of a tapering tubular body part such as the forearm, calf, or thigh. This is taken as standing for the actual size of a cross-section of the fish.

Another kind of conventionality in gestures concerns types of communicative practice like, say, 'tracing' in mid air (Mandel 1977, Kendon 1988). It may be argued that there are conventions which allow interpreters to recognize that a person is doing an illustrative tracing gesture, based presumably on formal distinctions in types of hand movement in combination with attention-directing eye gaze toward the gesture space (see Part II of this book). While the exact form of a tracing gesture cannot be pre-specified, its general manner of execution may be sufficient to signal **that** it is a tracing gesture.

1.7   Overview of the book

While each chapter deals with composite utterances, Part I focuses on deictic or symbolic indexical components, while Part II focuses on non-conventional illustrative components. Part I concentrates on signs whose main job it is to

link conventional signs with non-conventional signs, while Part II concentrates on signs whose main job it is to create new signs as (virtual) things in the world which may be talked about and pointed to.

The focus of Part I is the role in composite utterances of symbolic indexicals in attaching conventional signs (in most cases, words) to things in the world. In Chapter 2, the symbolic indexicals in question are in the spoken component of the composite utterance, in the form of demonstratives (words like *this* and *that*). In Lao, as in all languages, demonstratives form a closed set, where the members of this set play off each other in how they direct attention to things in context, and express stances toward those things. Conventional signs in a closed set of this kind not only have intrinsic, coded meanings, but also pick up enriched meanings through markedness relations within a restricted paradigm. The Lao system of demonstratives is not different in this respect to the system of pointing discussed in Chapters 3 and 4. The systematic contrast among types of bodily pointing (lip-pointing, 'small' finger-points, 'big' finger-points) demonstrate that the idea of a system of meaningful oppositions is as applicable to types of hand gestures as it is to types of words. In this sense, gestures can have grammatical properties.

Part II examines composite utterances in which hand gestures are employed to create virtual illustrations which function like models of concrete objects and diagrams of abstract sets of relations. The analyses show speakers using hand movements iconically and indexically to create novel signs which show remarkable structural persistence, both in space and through time. As these structures are established in the common field of attention, pointing plays a crucial role. The structures that the modelling and diagramming gestures create are effectively treated like physical objects. As Chapter 7 shows, speakers have to cope with the consequences of this virtual reality, being required to make explicit editing manoeuvres upon otherwise ethereal structures.

There are some principled differences between the phenomena described in Parts I and II. For example, while the deictic signs in Part I constitute closed grammatical systems, the illustrative signs in Part II are novel, open class items. And while Part I focuses more on the internal structure of composite utterance units (regarding the move as exoskeleton, fleshed out down-and-in), Part II explores complex structures which emerge through sequences of multiple moves (regarding the move as brick, building up-and-out).

These differences in focus of the two parts are offset by some thoroughgoing common themes. Most important is the collaborative, public, socially strategic nature of the process of constructing composite utterances. These communicative moves are not merely designed but designed for, and with, anticipated interpreters. They are not merely indices of cognitive processes,

they constitute cognitive processes. They are distributed, publicized, and intersubjectively grounded. Each type of composite utterance discussed in this book is regulated by its producer's aim not just to convey some meaning but to bring about a desired understanding in a social other. So, like all instruments of meaning, these composites are not bipolar form–meaning mappings, or mere word-to-world glue, they are premised on a triadic, cooperative activity consisting of you, me, and what I'm trying to say.