

Cambridge University Press

978-0-521-88064-0 - The Anatomy of Meaning: Speech, Gesture, and Composite Utterances

N. J. Enfield

Excerpt

[More information](#)

1 Composite utterances

A theory of utterance should not begin with a division between ‘speech’ and ‘gesture’.

Adam Kendon, 1986

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.

Cambridge University Press

978-0-521-88064-0 - The Anatomy of Meaning: Speech, Gesture, and Composite Utterances

N. J. Enfield

Excerpt

[More information](#)

2 Composite utterances

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. Semanticians 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

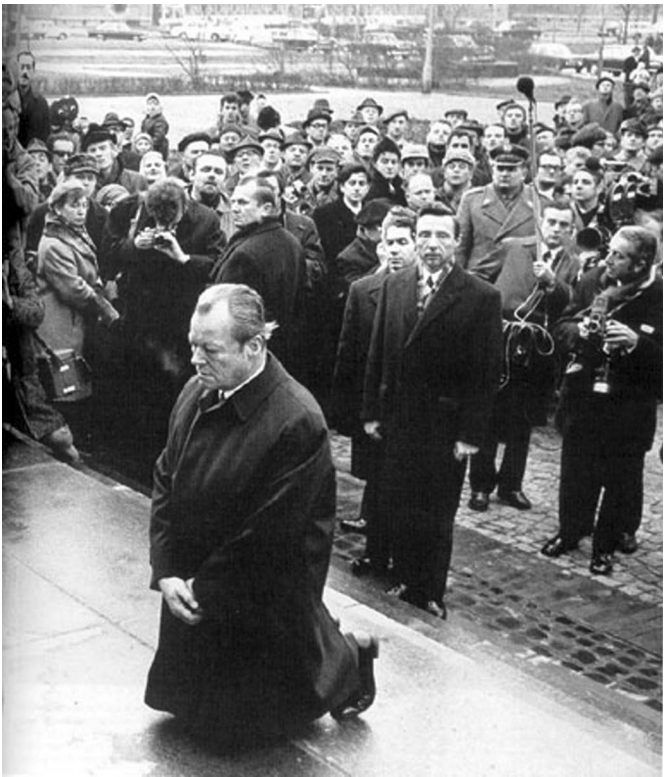


Figure 1.1 Man kneeling atop steps, with crowd looking on.

Cambridge University Press

978-0-521-88064-0 - The Anatomy of Meaning: Speech, Gesture, and Composite Utterances
N. J. Enfield

Excerpt

[More information](#)

4 Composite utterances

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.

Cambridge University Press

978-0-521-88064-0 - The Anatomy of Meaning: Speech, Gesture, and Composite Utterances

N. J. Enfield

Excerpt

[More information](#)

6 Composite utterances

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 *sòdn5* (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.

Cambridge University Press

978-0-521-88064-0 - The Anatomy of Meaning: Speech, Gesture, and Composite Utterances

N. J. Enfield

Excerpt

[More information](#)

Figure 1.5 Man describing the *sòdn5*, a traditional Lao fish trap: ‘As for the *sòdn5*, 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).

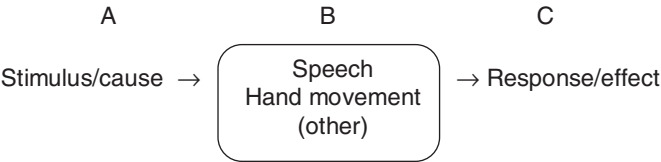


Figure 1.6 Three contexts of hand movement, in sequential interaction: at B, composite utterances may include multiple simultaneous signs; a preceding stimulus/cause at A determines a sign’s **appropriateness**; a response determines its **effectiveness**.

Cambridge University Press

978-0-521-88064-0 - The Anatomy of Meaning: Speech, Gesture, and Composite Utterances

N. J. Enfield

Excerpt

[More information](#)

10 Composite utterances

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*, *engage*).