

44. A “Composite Utterances” approach to meaning

1. Introduction
2. Composite utterances
3. Sign filtration: Triggers and heuristics
4. Semiotic analysis of gestures
5. Conclusion and prospects
6. References

Abstract

This chapter argues for a composite utterances approach to research on body, language, and communication. It argues that to understand meaning we need to begin with the utterance or speech act as the unit of analysis. From this perspective, the primary task in interpreting others' behaviour in communication is to infer what a person wants to say. In order to solve this task, an interpreter is free to consult any and all available information, regardless of the sensory modality in which that information is gathered (e.g., vision versus hearing), and regardless of the semiotic function of that information (e.g., iconic/indexical, symbolic/conventional, or some combination of these). Having recognized that another person has an intention to communicate, an interpreter takes the available relevant information (e.g., vocalizations, facial expressions, hand movements, all in the context of synchronic knowledge of linguistic and cultural systems, and other aspects of common ground) and looks for a way in which those co-occurring signs may simultaneously point to a single overall message of the move that a person is making. This is helped by the binding power of social cognition in an enchronic context (that is, the sequential context of turn-by-turn conversation), in particular the assumption that people are not merely saying things but making moves. The chapter focuses on co-speech hand gestures, and also discusses implications of the composite utterances approach to research on syntax, and on sign language.

1. Introduction

In human social behavior, people build communicative sequences move by move. These moves are never semiotically simple. Their composite nature is widely varied in kind: they may consist of 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 chapter explores the question with special reference to one of our most familiar types of move, the speech-with-gesture composite, a classical locus of research on body, language, and communication (see other chapters of this handbook relating to gesture, and many references therein). The central question is this: How do gestures contribute to the meaning of an utterance? To answer this, we need to situate research on gesture within broader questions of research on meaning.

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 ultimately driven by broader principles, principles of rational cognition in social life, principles which underlie other processes of human judgment, from house-buying to gambling to passing people on a crowded street. So, to understand meaning in human utterances, 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 (Kockelman 2005; Levinson 1983; Peirce 1955, Silverstein 1976). 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 different kinds agree on this (see Cruse 1986; Jackendoff 1983; 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). To explicate this point: 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 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, 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). Thus, alternatives to a static view of meaning are available for dealing with the specific problems of co-speech gesture. These come from two sources: (neo-)Peircean semiotics (e.g., Colapietro 1989; Kockelman 2005; Parmentier 1994; Peirce 1955) and (neo-)Gricean pragmatics (e.g., Atlas 2005; Grice 1975; Horn 1989; Levinson 1983, 2000; Sperber and Wilson 1995). Subsequent sections explore the relevant analytic tools offered by these traditions.

1.3. Meaning is a composite notion

When people say things they typically do so by combining words with images. A relatively simple example of a composite sign is the image-with-caption format typified by photographs and artwork. What makes this kind of thing 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; see 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. Interpreting such composites is done by means of a general heuristic of semiotic unity: when encountering multiple signs which are presented together, take them as one. This example illustrates essentially the same thing we find in the co-occurrence of expressive hand movements with speech: context-situated composites of multiple signs, part conventional, part non-conventional. Consider Fig. 44.1, 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. (Note: This example and the following one are from a corpus of video-recorded talk collected in Laos since 2000; as should be obvious, the point I am making here is not specific to the Lao data, and could be illustrated with comparable data from any other culture.)



Fig. 44.1: Man (left of image) speaking of preferred angle of a drainage pipe under construction: “Make it steep like this.”

The discussion in the context of Fig. 44.1 is about construction works under way 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 low 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 Fig. 44.1. The meanings of his words and his gesture are tightly linked, through at least three devices:

- (i) their tight spatiotemporal co-occurrence in place and time (both produced by the same source),
- (ii) the use of the explicit deictic expression “like this” (sending us on a search: “Like what?”, and leading us to consult the gesture for an answer),
- (iii) the use of eye gaze for directing attention.

A similar case is presented in Fig. 44.2, from a description of a type of traditional Lao fish trap called the *sòdn5* (see Enfield 2009: Chapter 5).



Fig. 44.2: Man describing the *sòdn5*, a traditional Lao fish trap: “As for the *sòdn5*, they make it fluted at the mouth.”

Again we see a speaker’s overall utterance meaning as a unified product of multiple sources of information:

- (i) a string of words (itself a composite sign consisting of words and grammatical constructions),
- (ii) a two-handed gesture,
- (iii) tight spatiotemporal co-occurrence of the words and gestures (from a single source), and
- (iv) eye gaze directed toward the hands, also helping to connect the composite utterance’s multiple parts.

This is subtly different from Fig. 44.1 in that it does not involve an explicit deictic element in the speech. Like the picture-with-caption examples mentioned above, spatiotemporal co-placement in Fig. 44.2 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.

A general theory of composite meaning takes Figs. 44.1 and 44.2, along with road signs, paintings on gallery walls, and captioned photographs 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, there are two important desiderata for an account of composite meaning. A first requirement is to provide a modality-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 (Engle 1998; McNeill 1992, 2005). If we want to achieve analytic generality, then a notion of holistic meaning is required not only for analyzing the meaning of co-speech hand gesture, but more generally for analyzing linguistic and other types of signs as well. 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.

2. Composite utterances

2.1. 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; see Özyürek et al. 2007). By what mechanism does this reciprocal interaction between hand and word unfold? Different approaches to analyzing meanings of co-speech gestures find evidence of a gesture’s meaning in a range of sources, including

- (i) speech which co-occurs together 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 three-part sequential structure underlies a basic trajectory model recognized by many students of human social behaviour. Schütz (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]; see Sacks 1992; Schegloff 2007 among many others).

2.2. Enchrony: The context of composite utterances

Any utterance is 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 (Atkinson and Heritage 1984; Sacks, Schegloff, and Jefferson 1974; Schegloff 1968; Goffman 1981). 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 in interaction, not historical time (for which the term diachronic is standard) but conversational time. For this I use the word *enchronic*. 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). The real-time birth and development of a composite utterance from a producer’s point of view (for which we might use the term microgenesis) is distinct from the intended meaning of enchronic here, namely the intersection of

- (i) a social causal/conditionality of related signs in sequences of social interaction and
- (ii) 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 as 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 2007) draws attention to the broader set of alternative viewpoints on systems and processes of meaning which we often need to switch between (including phylogenetic, diachronic, ontogenetic, and synchronic).

2.3. The move: A basic-level unit for social interaction

A primitive unit of an enchronic perspective is 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 2000; Schegloff 2007). The linguistic utterance is a well-studied (if idealized) type of instantiation of the move (see 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 (Chafe 1994; Pawley and Syder 2000), the turn-constructional unit (Sacks, Schegloff, and Jefferson 1974), the growth point (McNeill 1992), the composite signal (Engle 1998; see Clark 1996), and the utterance as multimodal ensemble (Goodwin 2000; Kendon 2004). 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, Camaioni, and Volterra 1975; Bates, O’Connell, and Shore 1987; Liszkowski et al. 2004; Tomasello 2006). The pointing gesture is mastered by prelinguistic infants (by around 12 months of age) and it is the first type of move to unequivocally display the sort of shared intentionality unique to human communication and social cognition (Frith and Frith 2007; Liszkowski 2006; Tomasello et al. 2005).

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 2007). 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. Much of the existing research on gesture, such as found in this handbook, examines the kinds of structure that arise when moves are built from word and hand together.

2.4. 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. 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). 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 [1916] 1959; on norms, see Brandom 1979; Kockelman 2006). 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 (Brookes 2004; Ekman and Friesen 1969). 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. These include anything that comes under the rubric of deixis (Fillmore 1997; Levinson 1983), that is, form-meaning mappings whose proper interpretation depends partly on convention and partly on context (Bühler [1934] 1982; 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 and 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 Enfield 2009, and studies of pointing in this handbook).

In the context of these three kinds of sign, it is important to be mindful of the distinction between type and token (Hutton 1990; Peirce 1955). 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 [1892] 1960), “timeless meaning” (Grice 1989) or “semantic invariance” (Wierzbicka 1985, 1996). Conventional signs are pre-fabricated signs, already signs by their very nature. 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 such as 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 section, it is worthwhile registering a common inconsistency in discussions of the meaning of hand movements in composite utterances. The problem is an inconsistent 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*_[GESTURE-up-and-down-away] later re-describes it as *the cat bounces out the pipe*. 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”; concerning 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”. 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 _[GESTURE-up-and-down-away] together point to a single idea “bounces out”. While word recognition has no analogue in the interpretation of the iconic gesture here (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.

2.5. 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 Fig. 44.3 (see Hanks 1990: 51ff; Levinson 1983: 14, 131).

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*

Fig. 44.3: Sources of composite meaning for interpretation of communicative moves. “Encoded” = conventional sign components. “Enriched” = non-conventional token meanings drawing on context.

In Fig. 44.3, “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.). “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; 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).

Thus, 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.

3. Sign filtration: Triggers and heuristics

The taxonomy of elements of composite signs in Fig. 44.3 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 in the first place. 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 Figs. 44.1 and 44.2, above). An example of the latter is discussed in Enfield (2009, 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 (Grice 1957; Peirce 1955). 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 an actual jar in your hands, the lack of conceivable practical aim is likely to act as a trigger for implicature (Gergely, Bekkering, and Király 2002; Levinson 1983: 157).

Data of the kind presented throughout this handbook do not usually present special 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 utterance interpretation, whether or not we take them to have been intended to communicate.

Note the kinds of heuristics that are likely being used in solving the problem of sign filtration. (On heuristics and bounded rationality in general see Gigerenzer, Hertwig, and Pachur 2011 and references therein.) By a convention heuristic, if a form is recognizable as a socially conventionalized type of sign, assume that 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 contextually 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). For further elaboration on the application of a heuristic model to the interpretation of speech-gesture composites, see Enfield (2009: 223–227).

4. 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. In the example of a fingerprint on the murder weapon, the print is iconic and indexical. It is iconic in that the print 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

- (i) it was directly caused by the fingertip making an impression on the weapon (thus a sign standing for an event of handling it), and
- (ii) 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 (Kendon 2004; McNeill 1992; inter alia) are fully explicable in terms of these types of semiotic ground, as shown in Fig. 44.4.

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 arc meant to look as if they were interacting with the referent;
 - in *mimetic* enactment, the hands arc moving as if they arc doing something to or with the referent
 - in *holding*, the hands arc shaped to look as if they arc holding the referent

Modeling:

- semiotic function: *iconic*; the hands arc meant to look as if they arc 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) arc meant to look as if they were tracing the shape of some salient feature of the referent, such as its outline.

Fig. 44.4: Sketch of some semiotic devices used in illustrative co-speech gestures (see Kendon 1988; Mandel 1977; Müller 1998).

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 (see Kockelman 2005: 240–241). This will entail teasing apart the large set of distinct semiotic dimensions which hand movements incorporate (Talmy 2006; see 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 pitch and loudness 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. Hand 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 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)” (Wilkins 2006: 132). A similar example is a Lao speaker’s conventional way of talking about sizes of fish, 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 as a way of illustrating or diagramming (Enfield 2009: Chapter 6; Kendon 1988; Mandel 1977). 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. 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.

Most important is the collaborative, public, socially strategic nature of the process of constructing composite utterances (Goodwin 2000; Streeck 2009). 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 a speaker, an addressee, and what the speaker is trying to say.

5. Conclusion and prospects

In solving the ever-present puzzle of figuring out what others are trying to say, our evidence comes in chunks: composite utterances built from multiple signs of multiple types. These composites are produced by people in trajectories of collaborative social activity. As communicative behaviours, they are strategic, context-embedded efforts to make social goals recognizable. If we are to understand how people interpret such

efforts, our primary unit of analysis must be the utterance or move, the single increment in a sequence of social interaction. Component signs will only make sense in terms of how they contribute to the function of the move as a whole.

This chapter has focused on moves built from speech-with-gesture as a sample domain for exploring the anatomy of meaning. But the analytic requirement to think in terms of composite utterances is not unique to speech-with-gesture. Because all utterances are composite in kind, our findings on speech-with-gesture should help us to understand meaning more generally. This is because research on the comprehension of speech-with-gesture is a sub-field of a more general pursuit: to learn how it is that interpreters understand token contributions to situated sequences of social interaction (see Goffman 1981; Goodwin 2000; Schegloff 1968; Streeck 2009).

How are multiple signs brought together in unified interpretations? The issue was framed above in terms of semiotic function of a composite's distinct components (see Fig. 44.4). A broad distinction was made between conventional meaning and non-conventional meaning, where these two may be joined by indexical mechanisms of various kinds. Think of a painting hanging in a gallery: a title (words, conventional) is taken to belong with an image (an arrangement of paint, non-conventional) via indexical links (spatial co-placement on a gallery wall, putative source in a single creator and single act of creation). Speech-with-gesture composites can be analyzed in the same way. When a man says *Make it steep like this* with eye gaze fixed on his arm held at an angle (see Fig. 44.1), the conventional signs of his speech are joined to the non-conventional sign of his arm gesture by means of indexical devices including temporal co-placement, source in a single producer, eye gaze, and the symbolic indexical expression *like this*. In these "illustrative gesture" cases, hand movements constitute the non-conventional "image" component of the utterance. By contrast, in cases of "deictic gesture" or pointing, hand movement is what provides the indexical link between words and an image or thing in the world, such as a person walking by, or diagrams in ink or mid-air.

This semiotic framework permits systematic comparison of speech-with-gesture moves to other species of composite utterance. An important case is sign language of the Deaf. There is considerable controversy as to how, if at all, gesture and sign language are to be compared (see Emmorey and Reilly 1995). The present account makes it clear that the visible components of a sign language utterance cannot be compared directly to the visible hand movements that accompany speech, nor to mere speech alone (i.e., with visible hand movements subtracted), but may only be properly compared to the entire speech-with-gesture composite (see Liddell 2003; Okrent 2002). The unit of comparison in both cases must be the move. By the analysis advanced here, different components of a move in sign language will have different semiotic functions, in the sense just discussed: conventional signs with non-conventional signs, linked indexically. Take the example of sign language "classifier constructions" or "depicting verbs" (Liddell 2003: 261ff). In a typical construction of this kind, a single articulator (the hand) will be the vehicle for both a conventional sign component (a conventionalized hand shape such as the American Sign Language "vehicle classifier") and a non-conventional sign component (some path of movement, often relative to a contextually established set of token spatial referents), where linking indexical mechanisms such as spatio-temporal co-placement and source in single creator are maximized through instantiation in single sign vehicle, i.e., one and the same hand.

Another domain in which a general composite utterance analysis should fit is in linguistic research on syntax. Syntactic constructions, too, are made up of multiple signs, where these are mostly the conventional signs of morphemes and constructions (though note of course that many grammatical morphemes are symbolic indexicals). An increasingly popular view of syntax takes lexical items (words, morphemes) and grammatical configurations (constructions) to be instances of the same thing: linguistic signs (Croft 2001; Goldberg 1995; Langacker 1987). From this “construction grammar” viewpoint, interpretation of speech-only utterances should be just as for speech-with-gesture. It means dealing with multiple, simultaneously occurring signs (e.g., *That guy* may be both noun phrase and sentential subject), and looking to determine an overall target meaning for the communicative move that these signs are converging to signify. A difference is that while semantic relations within grammatical structures are often narrowly determined by conventions like word order, speech-with-gesture composites appear to involve simple co-occurrence of signs, with no special formal instruction for interpreters as to how their meanings are to be unified. Because of this extreme under-determination of semiotic relation between, say, a gesture and its accompanying speech, many researchers conclude that there are no systematic combinatorics in speech-with-gesture. But speech-with-gesture composites are merely a limiting case in the range of ways that signs combine: all an interpreter knows is that these signs are to be taken together, but there may be no conventionally coded constraints on how. Such under-determination is not unique to gesture. In language, too, we find minimal interpretive constraints on syntactic combinations within the clause, as documented for example by Gil (2005) for the extreme forms of isolating grammar found in some spoken languages. And beyond the clause level, such under-determined relations are the standard fabric of textual cohesion (Halliday and Hasan 1976).

In sum, to understand the process of interpreting any type of composite utterance, we should not begin with components like “noun”, “rising intonation”, or “pointing gesture”. We begin instead with the notion of a whole utterance, a complete unit of social action which always has multiple components, which is always embedded in a sequential context (simultaneously an effect of something prior and a cause of something next), and whose interpretation always draws on both conventional and non-conventional signs, joined indexically as wholes.

Research on speech-with-gesture yields ample motivation to question the standard focus in mainstream linguistics on competence and static representations of meaning (as opposed to performance and dynamic processes of meaning; see McNeill 2005: 64ff, Wilkins 2006: 140–141). There is a need for due attention to meaning at a context-situated token level, a stance preferred by many functionalist linguists, linguistic anthropologists, conversational analysts, and some gesture researchers. Speech-with-gesture composites quickly make this need apparent, because they force us to examine singularities, i.e., semiotic structures that are tokens but not tokens-of-types. These singularities include non-conventional gestures as utterance components, as well as the overall utterances themselves, each a unique combination of signs. This is why, for instance, Kendon writes of speech-with-gesture composites that “it is only by studying them as they appear within situations of interaction that we can understand how they serve in communication” (Kendon 2004: 47–48; see also Hanks 1990, 1996, among many others). Here is the key point: What Kendon writes is already true of speech whether it is accompanied by gesture or not. Speech-with-gesture teaches us to treat

utterances as dynamic, motivated, concrete, and context-bound, which is the stance we need for the proper treatment of communicative moves more generally. By studying gesture in the right way, we study meaning better.

Acknowledgements

The text of this chapter is drawn from chapters 1 and 8 of *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances* (Cambridge University Press, 2009), with revisions. I gratefully acknowledge Cambridge University Press for permission to reproduce those sections here. I would also like to thank Cornelia Müller for her encouragement, guidance, and patience.

6. References

- Atkinson, J. Maxwell and John Heritage 1984. *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.
- Atlas, Jay D. 2005. *Logic, Meaning, and Conversation: Semantical Underdeterminacy, Implicature, and Their Interface*. Oxford: Oxford University Press.
- Austin, John L. 1962. *How to Do Things with Words*. Cambridge, MA: Harvard University Press.
- Bates, Elizabeth, Luigia Camaioni and Virginia Volterra 1975. The acquisition of performatives prior to speech. *Merril-Palmer Quarterly* 21: 205–224.
- Bates, Elizabeth, Barbara O'Connell and Cecilia M. Shore 1987. Language and communication in infancy. In: Joy D. Osofsky (ed.), *Handbook of Infant Competence* (2nd edition), 149–203. New York: Wiley and Sons.
- Brandom, Robert B. 1979. Freedom and constraint by norms. *American Philosophical Quarterly* 16(3): 187–196.
- Brookes, Heather 2004. A repertoire of South African quotable gestures. *Journal of Linguistic Anthropology* 14(2): 186–224.
- Bühler, Karl 1982. The deictic field of language and deictic words. In: Robert J. Jarvella and Wolfgang Klein (eds.), *Speech, Place, and Action*, 9–30. Chichester: John Wiley and Sons. First published [1934].
- Chafe, Wallace 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.
- Clark, Herbert H 1996. *Using Language*. Cambridge: Cambridge University Press.
- Colapietro, Vincent M. 1989. *Peirce's Approach to the Self: A Semiotic Perspective on Human Subjectivity*. Albany: State University of New York Press.
- Croft, William 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.
- Cruse, D. Alan. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.
- Ekman, Paul and Wallace V. Friesen 1969. The repertoire of nonverbal behavior: Origins, usage, and coding. *Semiotica* 1: 49–98.
- Emmorey, Karen and Judy S. Reilly 1995. *Language, Gesture, and Space*. Hillsdale, NY: Lawrence Erlbaum.
- Enfield, N. J. 2009. *The Anatomy of Meaning: Speech, Gesture, and Composite Utterances*. Cambridge: Cambridge University Press.
- Enfield, N. J. and Stephen C. Levinson 2006. Introduction: Human sociality as a new interdisciplinary field. In: N. J. Enfield, Stephen C. Levinson, Robert J. Jarvella and Wolfgang Klein (eds.), *Roots of Human Sociality: Culture, Cognition, and Interaction*, 1–35. Oxford: Berg.
- Engle, Randi A. 1998. Not channels but composite signals: Speech, gesture, diagrams and object demonstrations are integrated in multimodal explanations. In: Morton A. Gernsbacher, Sharon J. Derry, Robert J. Jarvella and Wolfgang Klein (eds.), *Proceedings of the Twentieth*

- Annual Conference of the Cognitive Science Society*, 321–326. Mahwah, NJ: Lawrence Erlbaum.
- Fillmore, Charles J. 1997. *Lectures on Deixis*. Stanford, CA: Center for the Study of Language and Information Publications.
- Foley, William A. and Robert D. Van Valin Jr. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- Frege, Gottlob 1960. On sense and reference. In: Peter Geach and Max Black (eds.), *Translations from the Philosophical Writings of Gottlob Frege*, 56–78. Oxford: Blackwell. First published [1892].
- Frith, Chris D. and Uta Frith 2007. Social cognition in humans. *Current Biology* 17(16): R724–R732.
- Gergely, György, Harold Bekkering and Ildikó Király 2002. Developmental psychology: Rational imitation in preverbal infants. *Nature* 415: 755.
- Gigerenzer, Gerd, Ralph Hertwig and Thorsten Pachur (eds.) 2011. *Heuristics: The Foundations of Adaptive Behavior*. New York: Oxford University Press.
- Gil, David 2005. Isolating-monocategorical-associational language. In: Henri Cohen and Claire Lefebvre (eds.), *Categorization in Cognitive Science*, 348–380. Amsterdam: Elsevier.
- Goffman, Erving 1964. The neglected situation. *American Anthropologist* 66(6): 133–136.
- Goffman, Erving 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.
- Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.
- Goodwin, Charles 2000. Action and embodiment within situated human interaction. *Journal of Pragmatics* 32: 1489–1522.
- Grice, H. Paul 1957. Meaning. *Philosophical Review* 67: 377–388.
- Grice, H. Paul 1975. Logic and conversation. In: Peter Cole and Jerry L. Morgan (eds.), *Speech Acts*, 41–58. New York: Academic Press.
- Grice, H. Paul 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Halliday, Michael A. K. and Ruqaiya Hasan 1976. *Cohesion in English*. London: Longman.
- Hanks, William F. 1990. *Referential Practice: Language and Lived Space among the Maya*. Chicago: University of Chicago Press.
- Hanks, William F. 1996. *Language and Communicative Practices*. Boulder, CO: Westview Press.
- Horn, Laurence R. 1989. *A Natural History of Negation*. Chicago: Chicago University Press.
- Hutton, Christopher M. 1990. *Abstraction and Instance: The Type-Token Relation in Linguistic Theory*. Oxford: Pergamon Press.
- Jackendoff, Ray 1983. *Semantics and Cognition*. Cambridge: Massachusetts Institute of Technology Press.
- Jakobson, Roman 1971. Shifters, verbals categories, and the Russian verb. In: Roman Jakobson (ed.), *Selected Writings II: Word and Language*, 130–147. The Hague: Mouton.
- Kendon, Adam 1988. *Sign Languages of Aboriginal Australia: Cultural, Semiotic and Communicative Perspectives*. Cambridge: Cambridge University Press.
- Kendon, Adam 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.
- Kita, Sotaro 2000. How representational gestures help speaking. In: David McNeill (ed.), *Language and Gesture*, 162–185. Cambridge: Cambridge University Press.
- Kita, Sotaro 2003. *Pointing: Where Language, Cognition, and Culture Meet*. Mahwah, NJ: Lawrence Erlbaum.
- Kockelman, Paul 2005. The semiotic stance. *Semiotica* 157(1/4): 233–304.
- Kockelman, Paul 2006. Residence in the world: Affordances, instruments, actions, roles, and identities. *Semiotica* 162(1/4): 19–71.
- Kockelman, Paul 2007. Agency: The relation between meaning, power, and knowledge. *Current Anthropology* 48(3): 375–401.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Volume I, Theoretical Prerequisites*. Stanford, CA: Stanford University Press.

- Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
- Levinson, Stephen 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. Cambridge: Massachusetts Institute of Technology Press.
- Liddell, Scott K. 2003. *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.
- Liszkowski, Ulf 2006. Infant pointing at twelve months: Communicative goals, motives, and social-cognitive abilities. In: N. J. Enfield and Stephen C. Levinson (eds.), *Roots of Human Sociality: Culture, Cognition, and Interaction*, 153–178. London: Berg.
- Liszkowski, Ulf, Malinda Carpenter, Anne Henning, Tricia Striano and Michael Tomasello 2004. Twelve-month-olds point to share attention and interest. *Developmental Science* 7(3): 297–307.
- Mandel, Mark 1977. Iconic devices in American Sign Language. In: Lynn A. Friedman (ed.), *On the Other Hand: New Perspectives on American Sign Language*, 57–107. New York: Academic Press.
- McNeill, David 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.
- McNeill, David 2005. *Gesture and Thought*. Chicago: Chicago University Press.
- Müller, Cornelia 1998. Iconicity and gesture. In: Serge Santi, Isabelle Guaitella, Christian Cavé and Gabrielle Konopczynski (eds.), *Oralité et Gestualité: Communication Multimodale, Interaction*, 321–328. Paris: L'Harmattan.
- Okrent, Arika 2002. A modality-free notion of gesture and how it can help us with the morpheme vs. gesture question in Sign Language linguistics (or at least give us some criteria to work with). In: Richard P. Meier, Kearsy Cormier and David Quinto-Pozos (eds.), *Modality and Structure in Signed and Spoken Languages*, 175–198. Cambridge: Cambridge University Press.
- Özyürek, Asli, Roel M. Willems, Sotaro Kita and Peter Hagoort 2007. On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience* 19(4): 605–616.
- Parmentier, Richard J. 1994. *Signs in Society: Studies in Semiotic Anthropology*. Bloomington: Indiana University Press.
- Pawley, Andrew and Frances Syder 2000. The one clause at a time hypothesis. In: Heidi Riggensbach (ed.), *Perspectives on Fluency*, 163–191. Ann Arbor, MI: University of Michigan Press.
- Peirce, Charles S. 1955. *Philosophical Writings of Peirce*. New York: Dover.
- Preissler, Melissa Allen and Paul Bloom 2008. Two-year-olds use artist intention to understand drawings. *Cognition* 106: 512–518.
- Richert, Rebekah A. and Angeline S. Lillard 2002. Children's understanding of the knowledge prerequisites of drawing and pretending. *Developmental Psychology* 38(6): 1004–1015.
- Ruiter, Jan Peter de, Stephane Rossignol, Louis Vuurpijl, Douglas W. Cunningham and Willem J. M. Levelt 2003. SLOT: A research platform for investigating multimodal communication. *Behavior Research Methods, Instruments, and Computers* 35(3): 408–419.
- Sacks, Harvey 1992. *Lectures on Conversation*. London: Blackwell.
- Sacks, Harvey, Emanuel A. Schegloff and Gail Jefferson 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4): 696–735.
- Saussure, Ferdinand de 1959. *Course in General Linguistics*. New York: McGraw-Hill. First published [1916].
- Schegloff, Emanuel A. 1968. Sequencing in conversational openings. *American Anthropologist* 70(6): 1075–1095.
- Schegloff, Emanuel A. 2007. *Sequence Organization in Interaction: A Primer in Conversation Analysis, Volume 1*. Cambridge: Cambridge University Press.
- Schütz, Alfred 1970. *On Phenomenology and Social Relations*. Chicago: University of Chicago Press.
- Searle, John R. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Cambridge: Cambridge University Press.
- Silverstein, Michael 1976. Shifters, linguistic categories, and cultural description. In: Keith Basso and Henry Selby (eds.), *Meaning in Anthropology*, 11–55. Albuquerque: University of New Mexico Press.

- Sperber, Dan and Deirdre Wilson 1995. *Relevance: Communication and Cognition* (2nd edition). Oxford: Blackwell.
- Streeck, Jürgen 2009. *Gesturecraft: The Manu-Facture of Meaning*. Amsterdam: John Benjamins.
- Talmy, Leonard 2006. The representation of spatial structure in spoken and signed language. In: Maya Hickmann and Robert Stéphanie (eds.), *Space in Languages: Linguistic Systems and Cognitive Categories*, 207–238. Amsterdam: John Benjamins.
- Tomasello, Michael 2006. Why don't apes point? In: N. J. Enfield and Stephen C. Levinson (eds.), *Roots of Human Sociality: Culture, Cognition, and Interaction*, 506–524. London: Berg.
- Tomasello, Michael, Malinda Carpenter, Josep Call, Tanya Behne and Henrike Moll 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28(5): 664–670.
- Wierzbicka, Anna 1985. *Lexicography and Conceptual Analysis*. Ann Arbor, MI: Karoma.
- Wierzbicka, Anna 1996. *Semantics: Primes and Universals*. Oxford: Oxford University Press.
- Wilkins, David P. 2006. Review of Adam Kendon (2004). *Gesture: Visible action as utterance*. *Gesture* 6(1): 119–144.

N. J. Enfield, Nijmegen (Netherlands)

45. Towards a grammar of gestures: A form-based view

1. Introduction
2. Principles of meaning creation: From form to meaning
3. Simultaneous structures of gesture forms: Variation of formal features
4. Linear structures of gesture forms: Combining gestures
5. Conclusion
6. References

Abstract

Departing from Kendon's (2004) notion of "features of manifest deliberateness" and their particular movement characteristics: "Deliberate expressive movement was found to be movement that had a sharp boundary of onset and offset and that was an excursion, rather than resulting in any sustained change of position" (Kendon 2004: 12), the chapter presents a form-based approach to gesture analysis, which regards gestures as motivated signs and considers a close analysis of their form as the point of departure for reconstructing their meaning. Furthermore, by considering gestural meaning not only as visual action but also as a form of dynamic embodied conceptualization, the approach takes a cognitive and interactive perspective on the process of ad hoc meaning construction in the flow of a discourse. By discussing principles of meaning creation (sign motivation via semiotic and cognitive processes) and the simultaneous (variation of formational features and gesture families) and linear structures (combinations within gesture units) of gesture forms, the chapter explicates individual aspects of a "grammar" of gestures. It is concluded that in gestures we can find the seeds of language or the embodied potential of hand-movements for developing linguistic structures.