Language in Society **41**, 259–266.
doi:10.1017/S0047404512000073

REVIEW ARTICLE

Where the action isn't

Nick J. Enfield, *The anatomy of meaning: Speech, gesture, and composite utterances*. (Language, culture and cognition) Cambridge: Cambridge University Press, 2009. Pp. 264. Hb. $124.

Reviewed by Michael Lempert
Department of Anthropology, University of Michigan
101 West Hall, 1085 S. University Avenue
Ann Arbor, MI 48108, USA
mlemp@umich.edu

ABSTRACT

Enfield's *The anatomy of meaning* is a pathbreaking exploration of multimodal communication based on Lao video recordings and fieldwork. Of special interest is his notion of COMPOSITES, through which he addresses the question of cross-modal integration—how it is that signs in different modalities cohere, speech and gesture being his focus. For whom, to what degree, and through what means do composites coalesce? Does this notion denote a KIND of cross-modal orderliness or the achievement of orderliness—perhaps through heuristics and inference-making? As I detail the book's engagement with this issue and compare it with others—especially recent discussions of "textuality" in linguistic anthropology and older structuralist approaches to behavioral events—I suggest that Enfield's notion is best viewed as a methodological operator. It is a notion that performs a kind of methodological displacement, pushing against an entrenched logocentrism while spurring us to view communicative events as multiplex, cross-modal assemblages. (Multi-modal, cross-modal, gesture, interaction, video, Lao)

A half-century ago, at the close of the pioneering 1962 conference on "Paralinguistics and kinesics" at Indiana University in Bloomington, Margaret Mead grasped for a word that could sum up the conference's aims. That word was SEMIOTICS: "[W]e are, I think, conceivably working in a field which in time will include the study of all patterned communication in all modalities" (Sebeok et al. 1964:275). In time, quite a few of us have, indeed, moved toward modalities other than language, though not at the pace anticipated by Mead, and not always under the banner of semiotics. Still, if Enfield's book is a signpost, the 1962 participants—which included Ray Birdwhistell, Erving Goffman, Edward Hall, and Thomas Sebeok—would be pleased at how far we've come. Drawing on extensive fieldwork and a vast store of Lao video data, Enfield explores SPEECH-WITH-GESTURE COMPOSITES, and the lucid, methodical manner of his investigation—whose phenomena range from lip-pointing to exquisite hand-gestures that depict

© Cambridge University Press, 2012 0047-4045/12 $15.00

https://doi.org/10.1017/S0047404512000073 Published online by Cambridge University Press

woven conical fish traps and the logic of sibling-exchange marriages—makes *Anatomy of meaning* an exciting work. Exciting not just for the way it ventures far beyond linguistics' blinkered sentence-sized units, but for the care it takes en route. It's one thing to invoke multi- or cross-modal relations. It's quite another to show precisely how these relations manifest themselves in communicative events—and without losing oneself, or one's readers, along the way. (See Birdwhistell's (1970) baroque transcripts in *Kinesics and context* for a reminder of what can happen when a descriptivist impulse goes unchecked.) *Anatomy of meaning* is, simply put, in the vanguard of the science of communication. It deserves wide attention.

Those who have spent time poring over video recordings of face-to-face interaction will find themselves nodding in agreement when Enfield rails against the logocentrism that has cordoned off language as an autonomous object, as he does in his opening salvo: "The payoff of studying speech-with-gesture is not only to understand gesture as a phenomenon of interest in itself, but—of greater consequence—to help shake linguists and other students of meaning from the view that language is an encapsulated system for conveying meaning" (ix). "It isn't," he answers flatly. This, together with his indictment of (neo)-Saussurian approaches to meaning, which privilege the "static, arbitrary, and abstract" rather than the "dynamic, motivated, and concrete" (2), is still news for some, but for fields like linguistic anthropology and sociolinguistics these are pedestrian claims—commonplaces, really—for these traditions have long taken as their foil the Saussurian legacy and its uprooting of language from the matrix of society, culture, and history. In fact, in Enfield's words one hears echoes of Vološinov's (1986) prescient critique of Saussurian structuralism from the 1920s, though readers will find no mention of that genealogy here.

***

If this isn't news, what is? The book's significance lies in its notion of COMPOSITES—"context-situated composites of multiple signs, part conventional, part non-conventional" (6). Enfield presents this notion in his introduction, then copiously illustrates it in six main chapters that fall into two symmetrical parts. These two parts are defined by semiotic function rather than modality—no accident, for Enfield argues (with Okrent 2002) that we need a modality-independent account of gesture rather than an account that just presumes modalities are different; to this end he also draws inspiration from Peircian semiotics. (Peirce did not have Saussure's preoccupation with language; his more expansive semiotics has been used to explain everything from honeybee communication to human speech.) Part 1 discusses indexical functions and offers chapters on linguistic deixis, lip-pointing, and hand-pointing, respectively. Iconic, "illustrative" functions command attention in Part 2, with chapters on cross-modal acts of modeling, diagramming, and editing. In each, Enfield treats us to close empirical investigations of Lao video data, which he presents with characteristic lucidity and care. The first-rate

https://doi.org/10.1017/S0047404512000073 Published online by Cambridge University Press

quality of the analysis, aided by a judicious use of video frame grabs, figures, and uncluttered transcript segments, makes these chapters not just well argued but admirably accessible.

Still, the book's signature contribution is also its most unsettling question, for it is not always easy to say what composites are, well, made of. Enfield is trying to address a major question here, that of cross-modal integration. For starters, Enfield recounts a remark by Kendon, who suggested that speech and gesture work reciprocally: "the gestural component and the spoken component interact with one another to create a precise and vivid understanding" (Kendon 2004:9, cited in Enfield 2009:9; cf. Agha 1996). Enfield rightly asks what "INTERACT" means. Mere spatio-temporal co-occurrence—the fact that speech and gesture co-occur? With his avowed neo-Gricean leanings, Enfield answers this by supplying heuristics—tacit, pragmatic rules-of-thumb that people use to make sense of signs. In this he is not frugal. While presumptive meanings have been neglected in fields like linguistic anthropology (Levinson 2000), Enfield piles heuristic upon heuristic, making the problematic of cross-modal integration feel somewhat cluttered—as cluttered as John Austin's (1962) felicity conditions in his classic lectures on performativity. (Remember how it felt like the whole universe had to be analytically reconstructed, from sincerity conditions to procedure to personnel, in order for a metonym like *I hereby bet you sixpence*—an "illocutionary one-liner," as Silverstein (1992:51) wryly put it—to be felicitous, to count as an act?) Enfield posits five heuristics, though we can always add more. For example, there's the orientation heuristic: "if a signer is bodily oriented toward you, most obviously by body position and eye gaze, assume they are addressing you." There's the UNIFIED UTTERANCE-MEANING HEURISTIC: "assume that contextually associated signs point to a unified, single, addressed utterance-meaning." There's an AGENCY HEURISTIC: "if a sign has greater control over a behavior, assume [all things being equal] that this sign is more likely to have been communicatively intended" (17).

At this point we may wonder whether Enfield sees heuristics as the epoxy that holds composites together. When he speaks of a "general theory of composite meaning," he means "signs co-occurring with other signs, acquiring unified meaning THROUGH BEING INTERPRETED as co-relevant parts of a single whole" (7; emphasis mine). Does Enfield, with Engle (1998), whom he cites favorably, thus want to argue that multi-modal composites are composites by virtue of HEURISTICS that presume integration?

Indeed, this is what Enfield often suggests, but surely we should distinguish, minimally, between (a) PRESUMPTIVE cross-modal integration, where integration is a function of heuristics used by interactants to construe signs; and (b) empirically DEMONSTRABLE cross-modal integration, that is, the gradient and fractional manner in which co-occurring semiotic resources can be shown to articulate with one another in some event of communication. How interactants sort through and stitch together signs is a vital topic, but this should not come at the expense of a heuristic-independent account of cross-modal integration. It's not that Enfield

https://doi.org/10.1017/S0047404512000073 Published online by Cambridge University Press

utterly neglects the latter, because he does, for instance, argue that people rely on "triggers"—signs that guide participants in their effort to draw together otherwise disparate signs. As I address below, at other times he even seems to insinuate that composites stick together through their intrinsic compositionality.

On balance, though, Enfield relies on heuristics to explain cross-modal integration, and this has its costs. After all, researchers have found all sorts of cross-modal relations between speech and gesture, like COMPENSATORY relations, where gesture adds information critical to, say, successful reference—information not conveyed by co-occurring speech; or ADDITIVE relations where one modality contributes increased resolution, as it were; CONFIGURATIONAL relations where the gesture-speech composite conveys meaning globally and synthetically, and hence cannot be resolved into its parts and their combination. And so on.

It is telling that while Enfield says that composites involve the interplay of conventional and nonconventional sign-vehicles[1] —a fair point—we never really learn what specific empirical forms this interplay takes. On this issue Agha's (2007) discussions of "text" and "textuality" (cf. Hanks 1989; Bauman & Briggs 1990; Silverstein & Urban 1996; Silverstein 1997) are instructive. Like Enfield, Agha offers a well-defined place for the categorial values of linguistic form-types, such as deictic expressions, honorifics, and lexicogrammatical resources. Agha demotes these values by calling attention to their precarious status in discourse as "text-defaults," "in the sense that the item regularly conveys a set of effects in the absence of interference from co-textual effects" (Agha 2007:39). Linguistic forms so conceived supply participants with default values, values that are modulated by emergent effects projected by the signs with which they co-occur as elements of their co(n)text, forming a "text-configuration." Semiotic effects produced concurrently by distinct functional principles (that is, as part of text-configurations) can "illuminate each other during usage," and "are reciprocally reflexive when considered in relation to each other" (Agha 2007:470). Agha appreciates the place of the conventional while detailing a few key ways that a text-configuration can diverge from its text-partials. A text-configuration featuring deictic forms, for instance, may:

> (a) **specify** more fully the categorial effects of a deictic token; or
> (b) render **defeasible** (i.e., partly deform or cancel) *some* dimension of categorial content, such as one or more of its text-defaults; or
> (c) constitute effects **notionally distinct** from deictic-categorial content. (Agha 2007:49)

A simple illustration of the broader principle:

> The so-called 'use of a form' is never just the use of a single sign; the idiom typically describes a text-pattern in which many other signs make a contribution, and these contributions may partially cancel the categorial effects of the form. Thus, a yes-no question categorially indexes the expectation of a 'yes or no'

https://doi.org/10.1017/S0047404512000073 Published online by Cambridge University Press

> response by a co-present addressee. In English, such a question is marked by a specific organization of sentence form (e.g. subject-verb inversion, sentence-final intonation rise). What we call a 'rhetorical use' of such a question, is the case where some sign (whether linguistic or non-linguistic) that co-occurs with the question token cancels the effect that a response is expected. For example no-one may be co-present or ratified as an addressee, as in the case of a dramatic soliloquy; or the predicate which occurs in the question token may characterize a state of affairs manifestly known only to speaker, thus canceling the expectation that a response is expected from anyone co-present. To speak of the 'rhetorical use' of the question is to speak of the effects of a sign-configuration of which the question marker (the syntactic-prosodic marker) is a fragment. (Agha 2007:53)

While Agha does not focus on gesture-speech composites, his sensitivity to the interaction effects of co-occurring signs is critical for another reason. When we have multiple semiotic resources in play we inevitably encounter what Agha calls "fractional congruence"—and that is the stuff of tropes. Consider, for instance, Ochs, Gonzales, & Jacoby (1996), who detail how the physicists in their study alternated between physicist-centric constructions like *And then I lower the temperature* and constructions like *If I go below in temperature*, where the scientists seem to adopt the perspective of the phenomena under investigation. The patterned deviation from expected co-occurrence patterns helps motivate "transpositional" interpretations—as if the zero-point of deictic reckoning shifts from observer to observed (cf. Hanks 1990; Haviland 1996; Perrino 2005). Or consider the familiar cases of tropic departures from second-person singular address forms. (Plural *vous* for singular addressee in French, a number trope in which one refers to addressee as if s/he were 'greater than one'; or, in Italian, the displacement to third-person singular feminine *Lei* (Silverstein 2003:208). Denotationally "incorrect" reference—again, a matter of fractional congruence—motivates tropic readings of social relations (Agha 2007). It is not that gesture-speech "mismatches" (cf. McNeill 1992) need attention (they are apparently quite rare), but that composites generally—especially ritualized ones—can involve juxtapositions of seemingly incompatible elements and thus flout conventions of "unity" for pragmatic effect; we risk missing this when we focus on heuristics and privilege cross-modal unification.

***

Anatomy of meaning may rely too much on heuristics to address the question of cross-modal integration, but on a few occasions it seems to suggest that there's something about composites THEMSELVES that makes their signs stick together. Composites, in a word, can exhibit compositionality.

Let us be clear about what this argument is not. This is not an attempt to scramble back to the rarefied heights of high structuralism. This is not, say, Kenneth Pike's (1967) *Language in relation to a unified theory of the structure of human behavior*,

https://doi.org/10.1017/S0047404512000073 Published online by Cambridge University Press

where all sorts of behavioral events were resolved into constituents. In Pike's analysis of a church sermon, for instance, he raced from "articulatory movements" to "sounds" to "words" to "lines" to "stanzas" to the "singing of a hymn" (Pike 1967:78)—as if all were amenable to distributional analysis and subject to constraints within syntagmatic "slots." He culled metadiscursive words and expressions, organized them into a partonymy, then assumed that lexicosemantic orderliness offered a direct window onto the sermon's cultural constituency.

Enfield does not repeat such an exercise, but his "anatomy" trope, trumpeted in the title and echoed by his notion of the composite utterance, does lure us into thinking that there is a SPECIFIC orderliness that makes composites cohere. Enfield knows perfectly well that some composites are not, strictly speaking, "compositional"—as in sense-compositionality in language; rather, some are configurational and global-synthetic (McNeill 1992, 2005), as in holophrastic lexations like *kick the bucket* ('die'), where one cannot predict the meaning of the whole by studying the meaning of the parts and their combination (Agha 2007). Still, he is at pains to demonstrate how some gestural sequences can exhibit qualities of compositionality familiar in language. If one were to ask "whether non-linguistic components of composite utterances show linear segmentation and combinatoric dependencies" (113), the answer is yes—they can. In cases of "symmetry-dominance" forms of gestural modeling, detailed in Ch. 5, interactants first hold hands in one position, then one hand does something new against that background. There's nothing surprising about the appearance of something like segmentation and combinatorics in cospeech gesture, but we should take care not to read these qualities into all co-speech gesture nor collapse distinctions among kinds of composites. At times Enfield does exercise care. About the symmetry-dominance gestures, for instance, he adds: "The extent to which this is equivalent to a grammatical construction in speech is open to discussion. It has some of the elements" (148). Linguistic anthropologists have often shown how ritualized discourse tends to exploit the "poetic" function (Jakobson 1960; Silverstein 2004). Rituals tend to have a high degree of cross-modal integration, so that they seem segmentable into constituents, like a whole in which subsidiary moves and phases subtend. (Compare with Conversation Analysis, where one adjacency pair can seem nested within another, higher-ranking adjacency pair. See Schegloff 2007; Sidnell 2010.) This emergent text-structure exhibits a kind of virtual compositionality that depends on indexical-iconicity (resemblance across co-occurring chunks of text), but this doesn't mean it's the same as constituency in grammar.

In the final analysis, let us read Enfield's composites not as an argument about any one KIND of cross-modal orderliness—even though he sometimes invites this reading through his choice of tropes and through showcasing instances of gestural compositionality. Let us read it methodologically, where it is most provocative. Like many uses of "discourse" and "text," the term "composite" targets, interdiscursively, the autonomization of language as denotational code and its blinkered sentence-size units. If we leave "composite" as an empirical placeholder for

https://doi.org/10.1017/S0047404512000073 Published online by Cambridge University Press

different forms of cross-modal orderliness—and disorderliness—then it is free to do what it does best, and that is to perform a methodological displacement (cf. Latour 2005), reminding us where the action isn't. It pushes us to see communicative phenomena as Enfield himself sees them—expansively, as multiplex, cross-modal assemblages—and this alone makes *Anatomy of meaning* an instructive, and indispensable, work.

NOTE

[1]Enfield creates space for hybrid, partly-conventionalized signs, namely, "symbolic indexicals" (especially deictic expressions), but he tends to speak of conventionalization in stark, categorical terms. There is no space to address this problematic here, but conventionalization is better viewed as gradient, admitting of degrees. On the very notion of "conventionality" we should also exercise care, given the way social-contract theory often underwrites this notion (see Parmentier 1994).

REFERENCES

Agha, Asif (1996). Tropic aggression in the Clinton-Dole presidential debate. *Pragmatics* 7(4):461–97.
——— (2007). *Language and social relations*. Cambridge: Cambridge University Press.
Austin, John (1962). *How to do things with words*. Ed. by J. O. Urmson. London: Oxford University Press.
Bauman, Richard, & Charles L. Briggs (1990). Poetics and performance as critical perspectives on language and social life. *Annual Review of Anthropology* 19:59–88.
Birdwhistell, Ray L. (1970). *Kinesics and context: Essays on body motion communication*. Philadelphia: University of Pennsylvania Press.
Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances, language, culture and cognition*. Cambridge: Cambridge University Press.
Engle, Randi A. (1998). Not channels but composite signals: Speech, gesture, diagrams and object demonstrations are integrated in multimodal explanations. In M. A. Gernsbacher & S. J. Derry (eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, 321–26. Mahwah, NJ: Erlbaum.
Hanks, William F. (1989). Text and textuality. *Annual Review of Anthropology* 18:95–127.
——— (1990). *Referential practice: Language and lived space among the Maya*. Chicago: University of Chicago Press.
Haviland, John B. (1996). Projections, transpositions, and relativity. In John J. Gumperz & Stephen C. Levinson (eds.), *Rethinking linguistic relativity*, 271–323. Cambridge: Cambridge University Press.
Jakobson, Roman (1960). Closing statement: Linguistics and poetics. In Thomas Sebeok (ed.), *Style in language*, 350–77. Cambridge, MA: MIT Press.
Kendon, Adam (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
Latour, Bruno (2005). *Reassembling the social: An introduction to actor-network-theory*. Oxford: Oxford University Press.
Levinson, Stephen C. (2000). *Presumptive meanings: The theory of generalized conversational implicature, language, speech, and communication*. Cambridge, MA: MIT Press.
McNeill, David (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.
——— (2005). *Gesture and thought*. Chicago: University of Chicago Press.
Ochs, Elinor; Patrick Gonzales; & Sally Jacoby (1996). "When I come down I'm in the domain state": Grammar and graphic representation in the interpretive activity of physicists. In Elinor Ochs, Emanuel

https://doi.org/10.1017/S0047404512000073 Published online by Cambridge University Press

A. Schegloff, & Sandra A. Thompson (eds.), *Interaction and grammar*, 328–69. Cambridge: Cambridge University Press.

Okrent, A. (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 R. P. Meier, K. Cormier, & D. Quinto-Pozos (eds.), *Modality and structure in signed and spoken languages*, 175–98. Cambridge: Cambridge University Press.

Parmentier, Richard J. (1994). Naturalization of convention. In Richard J. Parmentier (ed.), *Signs in society: Studies in semiotic anthropology*, 175–90. Bloomington: Indiana University Press.

Perrino, Sabina M. (2005). Participant transposition: Text and trope in Senegalese oral narrative. *Narrative Inquiry* 15(2):345–75.

Pike, Kenneth (1967). *Language in relation to a unified theory of the structure of human behavior*. The Hague: Mouton.

Schegloff, Emanuel A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.

Sebeok, Thomas A.; Alfred, S. Hayes; Mary Catherine Bateson; & Indiana University, Research Center in Anthropology, Folklore, and Linguistics (eds.) (1964). *Approaches to semiotics: Cultural anthropology, education, linguistics, psychiatry, psychology; transactions*, (Janua linguarum, series maior 15). The Hague: Mouton.

Sidnell, Jack (2010). *Conversation analysis: An introduction*. (Language in society) Malden, MA: Wiley-Blackwell.

Silverstein, Michael (1992). Metapragmatic discourse and metapragmatic function. In John Lucy (ed.), *Reflexive language: Reported speech and metapragmatics*, 33–58. Cambridge: Cambridge University Press.

——— (1997). The improvisational performance of "culture" in real-time discursive practice. In R. Keith Sawyer (ed.), *Creativity in performance*, 265–312. Greenwich, CT: Ablex.

——— (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication* 23 (3–4):193–229.

——— (2004). "Cultural" concepts and the language-culture nexus. *Current Anthropology* 45(5): 621–52.

Silverstein, Michael, & Greg Urban (1996). *Natural histories of discourse*. Chicago: University of Chicago Press.

Vološinov, V. N. (1986). *Marxism and the philosophy of language*. Trans. by L. Matejka and I. R. Titunik. Cambridge, MA: Harvard University Press.

(Received 29 July 2011)

https://doi.org/10.1017/S0047404512000073 Published online by Cambridge University Press