

Primary and secondary pragmatic functions of pointing gestures

N.J. Enfield^{a,*}, Sotaro Kita^b, J.P. de Ruiter^a

^a *Language and Cognition Group, Max Planck Institute for Psycholinguistics,
PB 310, 6500 AH Nijmegen, Netherlands*

^b *School of Psychology, University of Birmingham, Birmingham B15 2TT, United Kingdom*

Received 4 March 2006; received in revised form 20 October 2006; accepted 1 March 2007

Abstract

This article presents a study of a set of pointing gestures produced together with speech in a corpus of video-recorded “locality description” interviews in rural Laos. In a restricted set of the observed gestures (we did not consider gestures with special hand shapes, gestures with arc/tracing motion, or gestures directed at referents within physical reach), two basic formal types of pointing gesture are observed: B-points (large movement, full arm, eye gaze often aligned) and S-points (small movement, hand only, casual articulation). Taking the approach that speech and gesture are structurally integrated in composite utterances, we observe that these types of pointing gesture have distinct pragmatic functions at the utterance level. One type of gesture (usually “big” in form) carries primary, informationally foregrounded information (for saying “where” or “which one”). Infants perform this type of gesture long before they can talk. The second type of gesture (usually “small” in form) carries secondary, informationally backgrounded information which responds to a possible but uncertain lack of referential common ground. We propose that the packaging of the extra locational information into a casual gesture is a way of adding extra information to an utterance without it being on-record that the added information was necessary. This is motivated by the conflict between two general imperatives of communication in social interaction: a social-affiliational imperative not to provide more information than necessary (“Don’t over-tell”), and an informational imperative not to provide less information than necessary (“Don’t under-tell”).

© 2007 Elsevier B.V. All rights reserved.

Keywords: Gesture; Pointing; Composite utterances; Interaction; Affiliation; Information structure

* Corresponding author. Tel.: +31 24 3521275.

E-mail addresses: nick.enfield@mpi.nl (N.J. Enfield), s.kita@bham.ac.uk (S. Kita), janpeter.deruiter@mpi.nl (J.P. de Ruiter).

1. Introduction

Despite long being considered the simplest, most primitive form of communicative action (Wundt, 1972/1921:74–75), the humble pointing gesture has much to reveal about the nature of human social intelligence and mind. Since more than half a century, scholarly considerations of pointing gestures have addressed questions of the very foundations of signification, from psychologists' ruminations on their production and comprehension (Wundt, 1972/1921:74–75; Bühler, 1982/1934:45), to philosophers' concerns with notorious puzzles of ostensive reference (Wittgenstein, 1955:§33; Quine, 1960:100–101; cf. Peirce 1932/1965:§2.305), to pioneering ethnographic descriptions of pointing's role in different cultural settings (Efron, 1941/1972:96–97; Sherzer, 1973; Haviland, 1993). In more recent decades, most work on pointing has focused on its cognitive and semantic status as a proto-linguistic mode of human communication. Pointing is argued to be a critical precursor to language, both in phylogeny (e.g. Hewes, 1981, 1996; Rolfe, 1996) and ontogeny (Scaife and Bruner, 1975; Bates et al., 1975; Ohama, 1984; Schaffer, 1984; Bates et al., 1987; Carpenter et al., 1988; Butterworth and Morissette, 1996; Butterworth, 2003, *inter alia*). Pointing is now understood to constitute a uniquely human mode of joint-attentional behavior (Povinelli et al., 2003; Tomasello, 2006). Recent experimental work with pre-linguistic infants implicates pointing directly in one of the defining achievements of the human mind, our capacity for shared intentionality (Liszkowski, 2006; Tomasello et al., 2005).

Our interest in pointing goes beyond its proto-linguistic status. Past pre-linguistic infancy, pointing goes on to play a significant role in fully linguistic interaction. That is, once children acquire language, they do not cease to use pointing gestures as a key tool for achieving joint attention and constructing utterances. This article examines the relation between pointing gestures and speech in structurally integrated COMPOSITE UTTERANCES (cf. Clark, 1996; Engle, 1998 on “composite signals”).¹ We analyze a set of spatially anchored pointing gestures by Lao speakers from a corpus of ethnographic interviews and naturally occurring conversational interaction (restricting the set by ruling out gestures with special hand shapes, gestures with arc/tracing motion, and gestures directed at referents within physical reach). We observe two major types of pointing gestures, which form a composite utterance with accompanying speech in two very different ways. In one type (B-points, “big” in form), the information in the gesture tends to be a primary, foregrounded component of the communicative message, along with accompanying speech and sometimes even without it (e.g. when I silently point to your keys in response to your question ‘Where are they?’). These are essentially equivalent to proto-linguistic gestures, and, unlike many types of hand gesture, are not necessarily dependent on language for their proper interpretation. A second type (S-points, “small” in form) shows a different division of labor: the speech is informationally foregrounded, the gesture is a backgrounded modifier. These smaller, backgrounded pointing gestures are of an entirely different degree of informational sophistication, being functionally stitched in to spoken utterances. What look at first to be slight variations in form (bigger versus smaller pointing gestures) turn out to indicate

¹ This systematic integration of spoken and manual representations is the general topic of research under the rubric “gesture”, a multidisciplinary field which has recently emerged from several cognitive science disciplines, including linguistics, psycholinguistics, linguistic anthropology, and conversation analysis (e.g. Ekman and Friesen, 1969; Kendon, 1972, 2004; Schegloff, 1984; McNeill, 1985, 1992; Haviland, 1993; Goodwin, 1994, 2000, 2006; Goldin-Meadow, 2003; Kita, 2003; Liddell, 2003).

qualitatively quite distinct types of phenomena, with very different underlying cognitive and communicative structures. Evidence suggests that they are acquired at opposite ends (i.e. first, and last) of the order of acquisition of gesture types. That is, what look at first sight like *a* and *b* turn out to be alpha and omega.

In what follows, we first describe our empirical study of Lao speakers' spatially anchored pointing gestures, accounting in functional terms for the differential distribution of two formally defined types of pointing gesture, which we call B-points and S-points (formally defined below). Our question is whether the variation between big versus small form of a spatially anchored pointing gesture is random, or whether this formal variation is related to the pragmatic function of the utterance in which it occurs. We suggest that form/function differences between the two types of pointing gesture reflect distinct types of constraints which interactants have to satisfy in confronting the online problem-solving task of designing utterances in face-to-face interaction. The form/function properties of B-points are relatively straightforward. They focally serve to establish joint attention, and to inform, with respect to some perceptually available, and spatially locatable object. We consider in greater detail the subtle informational complexities of S-points. These gestures arise in our data where an informational imperative comes into direct competition with a social-affiliational imperative. These two imperatives are among multiple elements contributing to a speaker's constraint satisfaction problem in the course of producing signals for communication. The informational imperative requires interactants to ensure that the calibration of their referential understanding is sufficient for current communicative purposes at any stage of the interaction—or, in a nutshell, “Don't under-tell” (Schegloff, *in press*). The requirement for such calibration is to be expected in an intelligent system of communication which follows a cumulative informational trajectory (“discourse”), and in which signals are designed for their intended recipients/audiences (Sacks and Schegloff, 1979; Clark, 1992, 1996). But at the same time, since every exchange in human interaction is socially consequential (Sacks, 1992; Heritage and Raymond, 2005; Enfield, 2006), we can expect utterance production to be as much constrained by social-affiliational imperatives as it is by informational imperatives. We learn from S-points in the Lao data that social-affiliational imperatives outrank informational ones when the two are in conflict. In later sections of the article, we consider related phenomena described in other cultural settings, and we look more broadly at these secondary functions of pointing gestures.

2. An empirical study of Lao speakers' pointing gestures

Data for this study are from video-recordings of conversation and other discourse by Lao speakers in villages in and around Vientiane, the capital city of Laos (cf. Enfield, 2001, 2003a, 2003b, 2004, 2005). Included among these materials are video-recordings of the Historically Anchored Locality Description Task, a semi-structured outdoor interview on the history of a locality (Wilkins et al., 2001). Four speakers participated in such an interview. From the recordings of these interviews, we isolated all composite utterances which featured a spatially anchored pointing gesture. We defined a pointing gesture as *a communicative bodily movement which projects a vector whose direction is determined, in the context, by the conceived spatial location, relative to the person performing the gesture, of a place or thing relevant to the current utterance*. We excluded from consideration those pointing gestures which featured (a) hand shapes other than those with a single protruding digit (since it is known that these can have special, elaborated communicative functions such as politeness, denoting features of the target,

etc.; Kendon and Versante, 2003; Wilkins, 2003), or (b), arc or tracing movements of the articulator at phase of full extension, apparently representing motion (e.g. as occurred in our data in references to local roads), since other factors are likely to be involved in their use. All 102 gestures which were isolated by this procedure featured origo and target on the same horizontal plane. That is, none of the gestures pointed to a referent above or below the speaker.² They all involved articulation of the gesture by one arm only (no two-handed deictic gestures occurred), and all featured a single digit (index finger or thumb) as vector-projecting articulator.

Each utterance in the set was coded for form of the gestural action. (Coders did not understand Lao, and were not aware of the research question.) Coding was for manner of articulation of the gesture, distinguishing formally between B-points (gestures in which the whole arm is used as articulator, outstretched, with elbow fully raised), and S-points (gestures in which the hand is the main articulator, the arm is not fully straightened, typically with faster and more casual articulation). We also coded for orientation of the head (towards versus not towards the target). Information about head orientation allowed us to infer direction of eye gaze, a deictic signal of importance to pointing; see section 2.1.1, below, for discussion. The two types of gesture occurred with similar frequency in our set of examples (46 B-points vs. 56 S-points). Examples are illustrated in Figs. 1 and 2.

We wanted to know what the functional distinction was, if any, between these forms of co-speech pointing in our Lao corpus. We found, first, that B-points are characteristic of “location-focus”



Fig. 1. B-point (whole-arm articulation) with head oriented to target (speaker – in white shirt – pointing to something forward and to his left).

² Also, none of the gestures pointed to a referent within physical reach. Had we observed such gestures in our data set, we would have excluded them, since it is known that the form of a pointing gesture can be formally adapted in special ways to the fact that a referent is small and/or nearby (independent of information structure considerations).



Fig. 2. S-point (hand/lower-arm only articulation) with head not oriented to target (speaker – in white shirt – pointing to something directly behind him).

utterances (see section 2.2.2, and Fig. 8, for discussion and quantitative evidence). We first discuss these cases, before turning to S-points.

2.1. B-points in location-focus utterances

We use the term *location-focus* to describe a property of composite utterances in which the location of the target/referent is focal (in the sense of Lambrecht, 1994:Ch. 5), and nothing in the speech imports a presupposition of specific location of the referent. In other words, the primary function of the utterance is to tell the addressee where something is, in such a way that the ‘where’ information is given by the situational context (possibly a gesture), not the speech (see Fig. 1). This typically means that there is a deictic element in the speech, such as *here* or *there*, which obliges a listener to ‘go looking’ in the context for what is intended. Often it is a hand gesture that supplies the key information, but it need not be—there are other means to fill in the ‘blank’ that a deictic term opens up, the key mechanism being that a certain solution is made ‘mutually salient’ (Clark et al., 1983).³

In all of the location-focus utterances in our data set which featured a pointing gesture (34 in total), that gesture was a B-point (see Fig. 8). The canonical case is when an

³ In defining location-focus utterance for current purposes, we exclude cases in which the speech itself imports a presupposition of specific location of the referent, the main type of case being the use of proper names. Thus, *It's in Sunnydale* relies on the addressee already knowing where Sunnydale is, while *It's there* relies on the addressee to recognize in the context a mutually salient solution to the problem of where “there” must be referring to. Utterances such as *It's in Sunnydale* do focus on location, but these do not concern us here. Those location-focus utterances which include deictics require reference to context, but note that they do not entail the use of a hand gesture. Nor do they specify, if a gesture is used, what form it should take. This is the question we consider in the following section.



Fig. 3. “The District (Chief) Ngoon [at that time, his paddy fields were still here].”

utterance answers a ‘where?’ question with a demonstrative like ‘here’ or ‘there’ in combination with the gesture⁴:

- (1) I *juu1 thaang2 daj3 hoong2-hian2*
 be.at way which school
 ‘Where is it, the school?’
 D *hoong2-hian2 juu1 thaang2 [suan3-maak5 thaang2 phii4 dêø]*
 school be.at way garden-betel way right.here PCL
 ‘The school’s in the direction of [the betel garden this way right here]_{FIG. 1.}’

The gesture by D in (1) is illustrated in Fig. 1.

In another kind of example of a location-focus utterance, the locational information is volunteered independently by the speaker (rather than supplied in answer to a ‘where?’ question). The utterance is similarly primarily telling the addressee where something is. In this example, the speaker is talking about the history of his locality, and is discussing various *tasèng3* ‘district chiefs’ who had formerly governed:

- (2) *tasèng3 ngoon1 [khaaw1 nan4, naa2 phen1 ñang2 juu1*
 district.chief N period that, paddy 3SG.HON still be.at
 nii4 dêø –]
 here PCL
 ‘The District (Chief) Ngoon [at that time, his paddy fields were still here]_{FIG. 3.}’

⁴ Here, and in subsequent examples, the peak of the gesture overlaps with the bracketed section of speech. Subscript references are to figures depicting gestures which occur during the square-bracketed talk. Durations in seconds (in brackets) are silences. Transcription of Lao follows International Phonetic Association convention, except for *q* (glottal stop), *ñ* (palatal nasal), *ng* (velar nasal), *ê* (mid front vowel), *è* (low front vowel), *ù* (high back unrounded vowel), *ò* (low back vowel). There is no standard romanization of Lao. Lexical tones are marked by numerals at end of each syllable: 1 (mid level), 2 (high rising), 3 (low rising), 4 (high falling), 5 (low falling), ø (unstressed). Glosses of examples include 1/2/3 (first/second/third person), ACHV (achievement), HON (honorific), NEG (negation), PCL (particle), PL (plural), Q (question), SG (singular).

The use in speech of a demonstrative such as *nii4* ‘here’ or *phii4* ‘right here’ requires the addressee to consult the context in order to determine the intended referent. One common aid to this search is an accompanying pointing gesture. De Ruiter and Wilkins (1998:605) call these “obligatory” gestures, since the spoken part of the utterance “requires a gesture to go along with it for it to be codable and semantically complete”.⁵ These are distinguished from “supportive” gestures, in which the spoken part of the utterance “could stand on its own in context (with the primary proposition intact)” (De Ruiter and Wilkins, 1998; see section 2.2, below). The “obligatory” gesture helps determine the mutual salience required for a speaker and listener to converge on the intended referent of an indexical symbol such as a demonstrative (e.g. English *this*; Clark et al., 1983; Enfield, 2003b). We are concerned in this article to determine the relation between the form of pointing gestures (B- and S-points) and their semiotic function. While the notions of obligatory and supportive are an important beginning, that distinction focuses on the “degree of importance” the gesture has in the composite utterance. In this article, we try to focus more on the qualitative difference between roles that gestures may play in composite utterances, and, in addition, whether formal differences align with these, and if so, why.

When a referent is distant, both B-points and S-points occur, but for an utterance with this location-focus function it is a B-point that serves, articulated with the whole arm, and usually with head and gaze also oriented to the target (see next section). This ensemble of features is characteristic of infants’ primordial pointing gesture whose function is to inform or otherwise purposely establish focal joint attention (Liszkowski, 2006; Tomasello, 2006). Let us consider why it is that our location-focus composite utterances featured B-points.

2.1.1. Why B-points are formally well fitted to location-focus contexts

Location-focus utterances have the primary purpose of telling an addressee where something is. B-points serve this function better than S-points for at least three reasons. First, they are BETTER FOR ATTENTION-DIRECTION, since they take up more of the addressee’s visual field, they move further, and they move faster. Second, the greater effort of physical movement involved in a B-point is a POTENTIAL INDEX OF GREATER IMPORTANCE OF THE LOCATIONAL INFORMATION (i.e. the information expressly conveyed by the gesture) to the primary assertion of the speaker’s utterance (Levy and Fowler, 2000). Third, B-points are MAXIMALLY ACCURATE as pointers (in the absence of long sticks and laser pointers). They have longer, straighter, and typically longer-maintained vector-projecting form. These three formal properties of B-points mean that they are better fitted to convey focal locational/directional information. In location-focus utterances, where the context is critical for comprehension of locational reference, the B-point does it best.

In addition to these location-focal features of the form of B-points’ hand/arm articulation, B-points also tend to occur with accompanying orientation of head to target, significantly more often than S-points do (Fisher’s exact test, $p < .001$). In addition to coding of gesture “size”, we coded for orientation of the head either to, or away from, the target of the pointing gesture at the peak of the gesture’s articulation. The two binary parameters of manner of S-point versus B-point articulation and towards versus away head-orientation yield four formal variants, distributed in the data as follows (Fig. 4):

⁵ To say that the gesture is literally “obligatory” is too strong, since the reference of a deictic can be found in the context through other means, as long as a “mutually salient” solution is available (Clark et al., 1983).

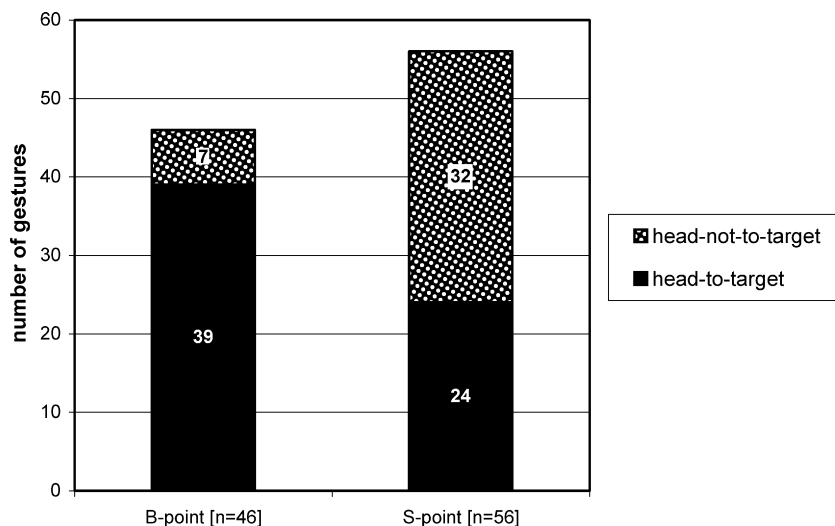


Fig. 4. Distribution of B-point (whole arm) and S-point (hand/lower-arm only) gestures to orientation of head to target.

Two generalizations are possible:

- (3) a. When a B-point is performed, the head will likely be oriented to the target of the pointing gesture (39/46 = 85%). (By contrast, S-points only weakly prefer head-not-to-target.)
- b. When the head is NOT oriented to target, a pointing gesture will likely be an S-point (32/39 = 82%). (By contrast, head orientation to target only weakly prefers B-points.)

In short: If the pointing gesture is a B-point, then head (and gaze) will be oriented to the target of the gesture. It is well established that the direction of eye gaze (inferable from head orientation in the absence of reliable eye-gaze data) is a powerful component in attention direction (Baron-Cohen, 1995).

To summarize, the formal features of B-points, along with the aligned eye gaze which typically accompanies them, account for their good fit with the function of focusing on ‘where’, and therefore on drawing focal attention to ‘where’ at the utterance level. The kinds of utterances which include these gestures are informationally relatively straightforward. They are not much different to the proto-linguistic “informing” utterances that 12-month olds perform with pointing gestures (Liszkowski, 2006). (This is of course not to deny the problematic nature of interpreting ostensive gestures, regardless of whether they are accompanied by speech; Wittgenstein, 1955:§33; Quine, 1960:100–101; Goodwin, 2006). We now turn to the more complex matter of S-points.

2.2. S-points in insecure reference environments

In contrast to the straightforward distribution of B-points to a single utterance function, the kinds of situations in which S-points occurred in our data appear to be more varied. Here are the four classes of utterance type which feature S-points in our data set:

- (4) (a) Request for clarification: The speaker is asking for clarification that this referent was the intended referent of a previous utterance (e.g. *You mean [the school]_{S-POINT} for secondary students?*)
- (b) Supply of clarification: The speaker is clarifying/confirming that this was the intended referent of a previous utterance of their own; in response to addressee's request for clarification (e.g. *What I mean is [the school]_{S-POINT} for secondary students.*)
- (c) Very general reference form: The speaker's linguistic utterance appears to contain insufficient information for unambiguous recognition of the intended referent, e.g. due to ellipsis or pronominalization (e.g. *[It's]_{S-POINT} for secondary students.*)
- (d) Initial mention of a recognitional term (e.g. a place name): The speaker introduces the referent with a proper name for the first time in the discourse, where there is possible reason to think that the linguistic term alone might not be sufficient for reference to be secured (*My kids go to [Sunnydale School]_{S-POINT}.*)

We argue that there is a functional unity to these seemingly diverse utterance types. In each case, the speaker makes linguistic reference to a uniquely recognizable entity in a situation where the referent seems LIKELY BUT NOT CERTAIN to be recognizable for these interlocutors in this context. There is apparent (if low) risk that the addressee is not going to be able to achieve reference on the basis of what's being said in the talk. We refer to these as insecure reference environments. The best way to clarify what is intended here is with examples.

2.2.1. Three examples of S-points in insecure reference environments

2.2.1.1. S-point with request for clarification: 'The steel bridge'. In this example, the interviewer wants to ask about a particular bridge he has in mind (line 1), not aware that there are multiple bridges he could be taken by this recipient to be referring to. The utterance in line 1 is insufficient to secure common reference to a single bridge in the local area. This is made apparent in line 2, when a stretch of more than a second's silence reveals an absence of uptake or grounding of the interviewer's attempted reference (Jefferson, 1989; Clark, 1996:268). In line 3, the interviewer resumes, voicing a signal of disfluency. Here, the addressee takes over and begins a request for clarification (lines 4–6), proposing a candidate referent for the bridge referred to by the interviewer in line 1. His S-point (illustrated in Fig. 2) co-occurs with his spoken description of this candidate referent in line 6, the 'steel bridge':

- (5) 1 I: *suan1 qanø khua3 juu1 thaang2 lum1 hanø dêø*
as.for um bridge be.at direction below there PCL
'As for, um, the bridge down below there?' [No gesture]
- 2 (1.3s)
- 3 I: *qanø...*
'Um...'
- 4 D: *khua3...*
'(The) bridge...'
- 5 (.3s)
- 6 D: *khua3 [mak-lêk1 (.3s)] khua3 phaan1 maa2 niø vaa3*
bridge steel bridge cross come PCL PCL,Q
'The bridge of [steel]_{FIG. 2}? The bridge (you) pass coming (here)?'



Fig. 5. “(Ø = you) don’t find (it) delicious?”.

The S-point marked in line 6 with square brackets is timed to peak with articulation in speech of the identifying feature of the bridge—i.e. that it’s the steel one, and not one of a number of other bridges made of other material, in other locations near the village.

2.2.1.2. *S-point with very general reference form: ‘ø not find (it) delicious?’.* Two women are sitting at an outdoor table (Fig. 5; Da is on the left, Tui on the right). They are eating a snack food which has been shared around to them and a number of children and visitors who are milling around. The discussion turns to whether the children are enjoying the food, and at a certain moment, Da formulates an utterance whose ambiguity is foreclosed only by the S-point (directed at her addressee Tui)⁶ which accompanies it (Fig. 5):

- (6) *bòø [sèèp4] vaa3*
 NEG find.delicious PCL.Q
 ‘(Ø = You) don’t find (it) delicious?’

The spoken part of the utterance is ambiguous because the subject of *sèèp4* (here meaning ‘to find something delicious’) could be the addressee Tui, or a child Khim, who is currently topicalized in the conversation. The pointing gesture disambiguates the overly general reference form in the utterance it accompanies, by filling in the gap in linguistic structure left by the “zero anaphor” (cf. Levinson, 1987:114–118; Haviland, 1993:17). Examples like this make explicit one of the key cognitive challenges in both producing and comprehending linguistic utterances: namely, binding speech and gesture into a unitary informational structure.

2.2.1.3. *S-point with initial mention of recognitional: ‘Saythani District’.* The interviewee is describing various changes that have taken place in his locality over the years:

⁶ Since the gesture is performed in the casual, fleeting manner typical of S-points, it is difficult to depict the gesture in a still image captured from video. Fig. 5 shows Da’s hand at full extension, but the hand is in motion at the moment depicted.

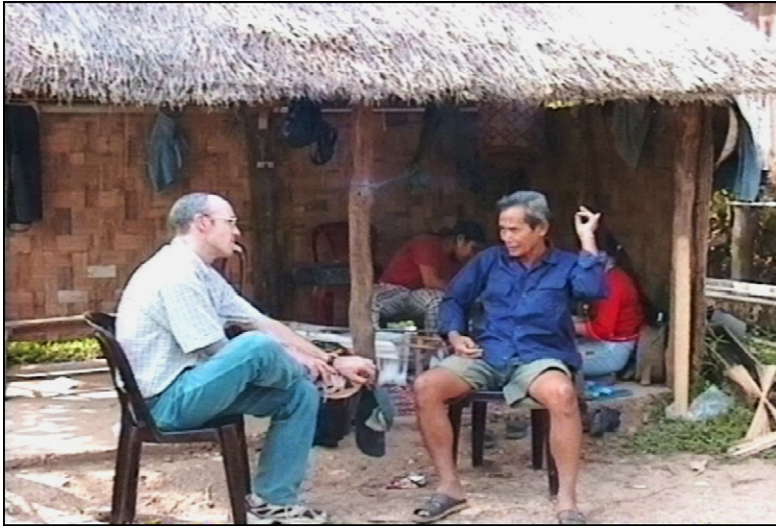


Fig. 6. “Before, our town and village still depended on [Saythani] District.”

- (7) *tèè1 kòòn1 mùang2 baan4 haw2 niø ñang2 khùn5 kap2*
 from before town village 1PL PCL still depend with
mùang2 [saj2thani2]
 district Saythani
 (pause)
bòø dajø khùn5 kap2 mùang2 [saj2sêêt5thaa3] dêj2
 NEG ACHV depend with district Saysettha PCL
 ‘Before, our town and village still depended on [Saythani]_{FIG. 6} District. It
 didn’t depend on [Saysettha]_{FIG. 7} District, you know.’

The first pointing gesture (Fig. 6) is to behind the speaker, to his left. The second gesture (Fig. 7) is to his right, forward. Both gestures are S-points. The locations to which these gestures aim are where the offices of the district administrations are physically located (each several kilometers away). These references are insecure in our sense because while the speaker has reason to believe that his addressee will know what is meant by the place names *Saythani* and *Saysettha* (the interviewer is known by the speaker to have been familiar with the area for some years), there is some reasonable doubt that these terms will be fully understood (since the interviewer is, after all, not a local).

2.2.2. When informational and social imperatives conflict: the S-point as a safety net for over-supposition

Of the set of 102 pointing gestures in our set, around two thirds (68) occurred in insecure reference contexts. Of these, more than four fifths (56) were S-points (a significantly greater proportion than in the location-focus context, where S-points were not observed; Fisher’s Exact test, $p < .001$). Further, B-points occur more often in location-focus utterances than in insecure reference utterances ($\chi^2(1) = 5.26$, $p < .05$). In other words, S-points are associated with insecure reference utterances, and B-points with location focus utterances.

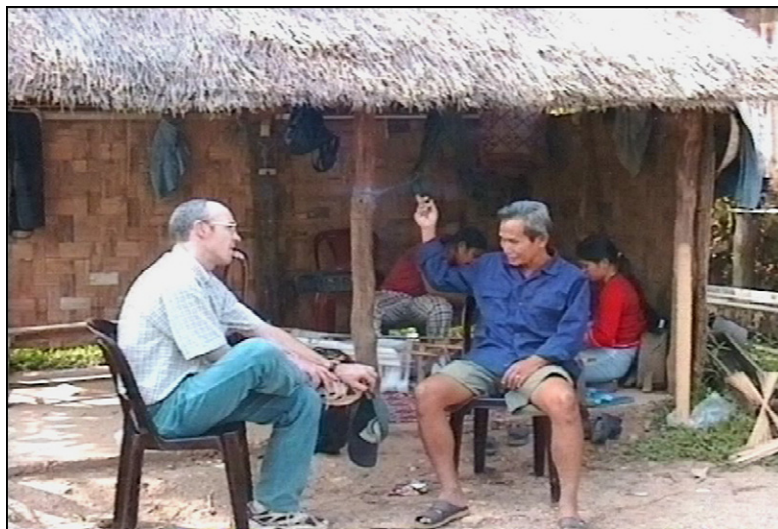


Fig. 7. “It didn’t depend on [Saysettha] District, you know.”

We propose that these S-points are selected in a class of interactional situations in which speakers face a dilemma: ‘I want to formulate a likely identifiable referent as recognitional (e.g. *You come in past Somsamai Village*), but I’m not completely confident that you will recognize the referent by the linguistic formulation I have in mind’. The speaker is stranded between options for linguistic formulation. If her linguistic formulation supposes too much – e.g. by introducing a place by proper name (as in example (7), above) when it turns out the recipient doesn’t recognize the place by name – the risk is failure to achieve reference. This fails to satisfy an informational imperative, resulting in failure of speaker and addressee to converge in referential understanding. Under-telling in this way is likely to result in a request for (and subsequent supply of) clarification, disrupting the progressivity of the interaction (Stivers and Robinson, 2006), at some cost for all concerned (cf. Heritage, 2007; Schegloff, 2007).

If the only operative imperative in an intelligent system for communication were to provide sufficient information for referential success, then when a given referential formulation is thought by its producer to be possibly inadequate for its intended recipient, one would presumably simply reject it in favor of something more explicit. So, instead of *Somsamai Village*, whose definiteness presupposes recognition, a speaker in any doubt might formulate the place as indefinite (*a village upstream of here*), or refer to it less succinctly with a complex description (*the village with the steel bridge, on the way here*). But these informationally more secure options run a different kind of risk, since despite some doubt about the referential adequacy of the selected linguistic formulation (e.g. a proper name), the speaker might at the same time also not be confident that the recipient wouldn’t achieve reference from the recognitional linguistic expression being supplied (e.g. a simple place name). If the speaker’s formulation supposes too little common ground, there is a danger of OVER-telling, being more explicit than was actually necessary, thereby failing to heed a social-affiliational imperative (Schegloff, 2007). In other words, if a speaker formulates his or her utterance so as to put “on-record” (Brown and Levinson, 1987) a presumption that the addressee does not possess the requisite knowledge to achieve reference – e.g. that they wouldn’t recognize a particular place name on first mention – the risk is

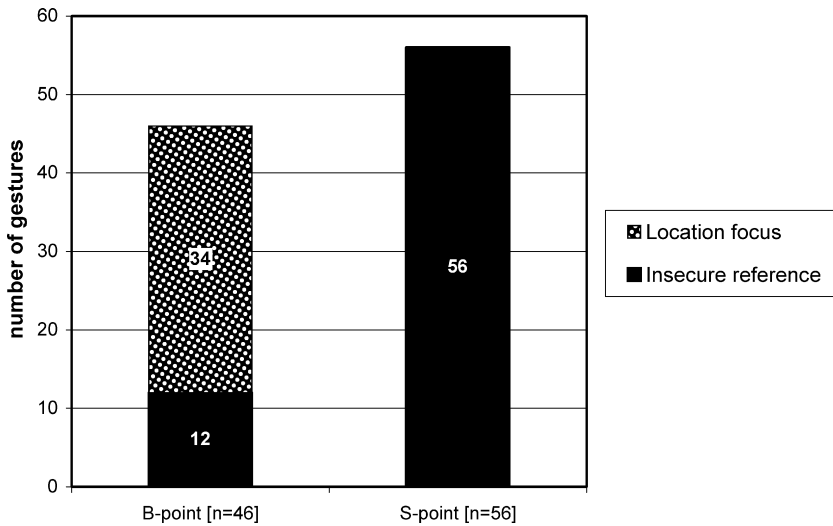


Fig. 8. Distribution of B-point (whole-arm) vs. S-point (hand/lower-arm only) in the two utterance functions. (No S-points occurred in the location-focus condition.)

socially distancing the interlocutor by treating them as not sharing knowledge with the speaker (Enfield, 2006).

So, while the risk of under-telling is informational/referential failure, the risk of over-telling is epistemic misjudgment and social/affiliational failure. The insecure reference utterances present the speaker, or better, utterance designer, with a dilemma of constraint satisfaction. According to our analysis, the Lao speakers' S-points give them a half-way solution: maintain the symbolic, focal, linguistic component of the utterance as it should be, assuming informational calibration, and add a pinch of extra information (here, the spatial information indexed by the gesture) in a relatively low-cost, off-record, non-intrusive, potentially pre-emptive, compensatory form. (Or: Risk under-telling in the speech, risk over-telling in the gesture.) For affiliative reasons, speakers will show a preference to err on the side of under-telling (or, "over-supposing", as Schegloff, 2007 puts it) rather than over-telling. When one is teetering in the middle, a well-placed S-point is an informational safety net.

2.2.3. Why S-points are formally well fitted to insecure-reference contexts

Think about the kind of utterance which is home to the S-point. This is an utterance whose purpose is to converge with the speaker on referential coordination when the adequacy of the linguistic formulation alone is not guaranteed. As laid out in the previous section, the S-point component of the utterance functions to provide a bit of extra information "just in case", but where this information should not be perceived as even necessary, since this active attention to a perceived informational risk of under-telling must be balanced against the desire to avoid a social-affiliational risk of over-telling. Here, as we saw for B-points, above, form fits function. We argue that the reduced physical movement involved in the articulation of S-points is indexical/iconic of "not primary assertion" (cf. Levy and Fowler, 2000), via a Gricean pragmatic rule of quantity (i.e. given that the option of doing the bigger gesture is available yet not selected; Grice, 1987; Levinson, 2000). More importantly here, the speaker's formulation of the gesture as casual and small is an iconic means to overtly downplay the need for its even

being there. The S-point's casual, reduced, fast articulation means that the locational information it provides is vulnerable to not being easily interpreted by the recipient, and this vulnerability is itself part of the signal. This property is neatly fitted for the function we have illustrated (and we expect, given the naturalness or non-arbitrariness of these form-function mappings, that our findings will more or less generalize across cultural settings). The speaker indicates by this very manner of formulation that the information in the gesture is not especially important, as if to say "you probably don't need this extra information, because you probably know what/where I mean, so I'm not going to too much trouble to make it available to you, since it's probably not going to matter if you don't receive it". In this sense, the addition of locational information by the S-point gesture is less on-record, and a possible threat to face is minimized.

3. A broader consideration of "secondary pointing" phenomena

If our Lao speakers' B-points represent an adult form of proto-linguistic gesture-focal utterances with speech more or less tacked on, the S-points, by contrast, have an inverted structure. In the S-point utterance, speech is primary. The hand gesture component retreats into the finer information structure of the utterance. Let us now consider S-points within a broader category of *secondary pointing utterances* reported in other studies.

Bavelas et al. (1992) describe a type of gesture called "citing", one of a number of gestures they label "interactive", since these gestures function to manage the interpersonal/interactional context of speech. This type of gesture is familiar to us (the authors) from everyday life, for instance in scientific discussion: a speaker proffers a small pointing gesture, formally identical to the Lao S-points described above, in the direction of someone present, while offering a comment which makes reference to something this person has previously or elsewhere said independently. (Imagine you raise the topic of the general theory of relativity while Albert Einstein is in the room; we bet you will accompany this with a little point or a nod in his direction.) Like a Lao S-point, this form of citation also deals with social-affiliational considerations rather than focal informational content. Our example works because of a socially consequential notion in the cultural environment of scientists (with its analogue in every imaginable domain of social expertise), that different individuals have differential rights to claim authority, according to their real or rightful area of knowledge. The problem is, if you make a statement unprompted – i.e., in "first position" in conversation or other discourse – you are thereby making an implicit claim of having primary authority to make that statement (Heritage and Raymond, 2005). And if you happen not to be the one in the room with the collectively acknowledged highest authority to say it, you will have to negotiate the mismatch between your own (actual) secondary authority and the (implicit) claim of primary authority via the "first position" placement of your contribution.

Like the Lao speakers' S-points, these small "citing" pointing gestures provide a neat solution to the dilemma. A wave of the index finger by the speaker in the direction of the present higher authority is sufficient to explicitly defeat any implicature of primary authority arising from the speaker's first-position placement of the utterance. The common link with the insecure-reference environments in the Lao study is that the small pointing gesture is performed in the service of heading off a possible problem at the social-affiliational level. This type of gesture is socially regulated. Its function is not primarily to achieve informational alignment, or even some kind of structural alignment in the interactional sequence. Its function is to deal with interpersonal affiliation. We see again the speaker's systematic

attention to satisfying a social-affiliational constraint, as distinct from their efforts to simply achieve referential calibration.⁷

Another type of secondary pointing gesture is observed among the “abstract deictic gestures” identified by McNeill and colleagues (McNeill, 1992, 2003; McNeill et al., 1993) in data from American English speakers. Abstract deictic gestures are defined formally by their orientation to “empty space”.⁸ The chunks of space indicated by such gestures are temporally endowed with representational value for the purpose of discourse (McNeill, 1992:113ff, 171, 199–200; cf. Enfield, 2003a, 2005). We are interested here in a subset of these, which are used for higher-level discourse-organizational tasks, at “metanarrative” and “paranarrative” levels. (We shall refer to these together as “supernarrative”.) Like our S-points, supernarrative abstract pointing gestures are “reduced” and “restricted” in form (McNeill, 1992:199–200). They function to track both concrete and abstract components of a discourse at informationally backgrounded levels. Furthermore, McNeill and colleagues note, in their observations of English and Georgian conversations, that such pointing gestures tend to appear with the initial few references to a newly introduced entity in a discourse (McNeill, 1992:215–217, McNeill et al., 1993:15–17). This is analogous to the function of S-points in the Lao data, discussed above. At the moment of potential insecurity in converging on a referent, abstract supernarrative pointing gestures anchor the referent in a gesturally created space, and S-points anchor the referent in the physical environment that surrounds the speaker and the addressee.

The common function of these various kinds of secondary pointing gesture – our S-points, Bavelas et al.’s citing gestures, McNeill et al.’s supernarrative gestures – is that unlike primary pointing gestures (our B-points) they function at levels of organization which are subordinate to the focal information in the overall utterance (i.e. as expressed by the speech). Here’s the “omega” end of the pointing spectrum. The two major types of gesture we identify in this study are not just “smaller” versus “bigger”. They are qualitatively quite distinct. The type we are calling secondary pointing gestures are complex, subtle, and dependent, and have an entirely different life to the proto-linguistic B-points.

McNeill (1992) raises the issue of the acquisition of children’s ability to perform supernarrative gestures. Concrete pointing is the first of all types of gesture acquired (McNeill, 1992:24, 300; cf. Liszkowski, 2006), coming at about one year of age, significantly earlier than language. By contrast, the use of abstract deictic gestures with metanarrative function “increases from age 8” (McNeill, 1992:322), while abstract deictic gestures with paranarrative function are not acquired until around age 11, very late (McNeill, 1992:326). These late-acquired supernarrative functions of gesture presuppose great sophistication in handling information structure at both the utterance and discourse levels. McNeill argues that the late acquisition of supernarrative abstract pointing is due to the lateness of acquisition of clearly distinct narrative levels in the general organization of discourse (McNeill, 1992:319). “The lateness of abstract pointing is particularly striking given that concrete pointing is the first of all gestures to emerge; abstract pointing is, however, one of the last” (McNeill, 1992:319–321). Our account – from alpha to omega – fits this picture.

⁷ Dealing with face is one of the central computational problems for a speaker, given that face is linked to social intelligence as well as emotion, both central to human cognition. If we are correct, then interestingly, the social imperative outranks the informational imperative when the two come into conflict. A question for further research is whether that ranking is universal.

⁸ The definition of abstract pointing in McNeill et al. (2003) includes pointing at empty space in both narrative and supernarrative (metanarrative and paranarrative) levels. However, McNeill (1992) has a slightly narrower definition, which excludes narrative level pointing that utilizes an iconically established spatial layout (p. 173).

4. Discussion: distributing informational and affiliational tasks in utterance formulation

Any cognitive scientific interest in complex external representations in human communication must incorporate the full range of constraints and resources for communicative/representational action in face-to-face interaction. Our investigation raises implications for our understanding of how humans represent meaning in face-to-face interaction, and the kind of constraints and affordances our habitual mode of representation (fully multimodal) presents us with.

For solving the challenging computational problem of making one's communicative intentions recognizable to others, people have a range of resources at their disposal. While spoken lexico-syntax is a linear, digital medium, other semiotic resources allow additional information to be represented simultaneously to speech. In composite utterances, made up of lexico-syntactic structure, prosody, hand gestures, facial expressions, etc. in concert,⁹ the symbolic component (i.e. the lexico-syntax) has a special role, being the most "on-record" component. If we want certain parts of our overall message to be less on-record, then it is in our interest to keep them out of the lexico-syntax. The solution is to employ semiotic resources which are relatively unobtrusive. A good example is intonation contour, which can be varied independent of lexico-syntax (cf. De Ruiter et al., 2006). Directly analogous to one of the S-point functions discussed in this article is the try-marker (Sacks and Schegloff, 1979/2007). Suppose you want to say *Yesterday I saw John*, but you're not entirely sure if your addressee will recognize who it is you mean by *John*. You can "try-mark" the proper name by overlaying it with rising intonation (like a question). This has the effect of "asking" the hearer to supply confirmation that she understands who is intended, while allowing you to maintain a grammatical formulation which presupposes recognition. A: *Yesterday I saw John?* B: *Uh-huh* (indicating recognition, go-ahead). A: *He was (etc. . .)*. On paper, the formulation "Yesterday I saw John" is done in such a way as to display an assumption that the addressee should know who is meant by "John". The pragmatic modulation achieved by the try-marker does not compromise the informationally optimistic (and therefore socially affiliative) framing of the utterance's lexico-syntactic component.¹⁰ The semiotic structure of try-marking directly parallels the Lao S-points. In both cases, information is added to the utterance in a form which is manipulable independent of, and simultaneously with, the lexico-syntax. Like the Lao S-point, the try-marker is used in initial mentions of recognitionals (i.e. proper names). The common mechanism is that the addition of information which might be a threat to face (since it betrays the perceived possibility of a current lack of common ground; Enfield, 2006) is done in such a way that it does not cause the speaker to have to reformulate the more optimistically phrased lexico-syntactic structure.

5. Conclusion

We have described two ways in which speech and pointing are combined in composite utterances. The two kinds of pointing can be related to the fundamental properties of pointing as a

⁹ Some traditions in research on meaning in cognitive science and linguistics have all but ignored the formal complexity of composite utterances, focusing on linguistic semantic structures which underlie lexical (and sometimes constructional) resources. On the other hand, both anthropologists and cognitive scientists have looked seriously at multimodality in perception and cognition where external artefacts are involved (Goody, 1977; Larkin and Simon, 1987; Norman, 1991; Glenberg and Langston, 1992; Hutchins, 1995; Zhang, 1997; Oestermeier and Hesse, 2000; cf. Enfield, 2005).

¹⁰ Along the same lines, Hayashi (2005) describes the try-marker as a technique for preserving the integrity of the syntax in the best interests of conversational productivity (Stivers and Robinson, 2006).

basic mode of human communicative action (Tomasello et al., 2005; Tomasello, 2006; Liszkowski, 2006). The B-point (big in form) is hardly different functionally to the pointing gestures performed by pre-linguistic infants. It is essential for comprehending the basic meaning of the composite location-focus utterance, and in some cases may inform all by itself. Typically, added speech is merely supportive/elaborative of the B-point gesture, and if anything is dispensable it is the speech. S-points (small in form) are of a different order. They are more subtle, more dependent, more hidden in the information structure of the composite utterance. While a B-point is doing the primary work of the utterance, with speech playing a supporting role, an S-point adds a backgrounded modifier to an utterance in which speech is central. A speaker's addition of an S-point to a spoken utterance is a low-cost, low-risk communicative action, which might save against a potentially high social-interpersonal cost, i.e. the undesirable situation of being on record as having underestimated common ground (Enfield, 2006). It is a way of striking a middle ground when two preferences conflict—a socially driven preference against over-telling versus an informationally driven preference against under-telling. While the B-point is an openly foregrounded informative action, the S-point is a backgrounded, subtle, pre-emptive strike in the ongoing struggle to maintain intersubjectivity.

Acknowledgements

We thank the Lao speakers who participated in this study by agreeing to be interviewed on video, and allowing us to analyze and publish the results. We would also like to acknowledge David Wilkins' contribution to development of our thoughts on the semiotics and ethnography of pointing (cf. Wilkins et al., 2001). Thanks to Alex Dukers for his help with the illustrations. We have benefited from helpful commentary by audiences at the 1st Meeting of the International Society for Gesture Studies (Austin, Texas, 2002), the Cognitive Science Colloquium Series, Northwestern University (Chicago, Illinois, 2005), the 104th Meeting of the American Anthropological Association (Washington, DC, 2005), and the MPI Nijmegen; we thank in particular Ric Ashley, Chuck Goodwin, John Haviland, Steve Levinson, and Jürgen Streeck. This work was conducted within the Gesture Project (now Nijmegen Gesture Centre) and the Multimodal Interaction Project at the Max Planck Institute for Psycholinguistics in Nijmegen. Enfield thanks the departments of Psychology and Linguistics at the Max Planck Institute for Evolutionary Anthropology in Leipzig for their hospitality during a visit in April 2005 when a first draft of this article was written. We gratefully acknowledge the Max Planck Society for field work, technical, and research support.

References

- Baron-Cohen, S., 1995. *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, Cambridge, MA/London.
- Bates, Elizabeth, Camaioni, L., Volterra, V., 1975. The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly* 21, 205–224.
- Bates, Elizabeth, O'Connell, B., Shore, Cecilia M., 1987. Language and communication in infancy. In: Osofsky, J.D. (Ed.), *Handbook of Infant Competence (Second Edition)*. Wiley and Sons, New York, pp. 149–203.
- Bavelas, Janet Beavin, Nicole, Chovil, Douglas, A. Lawrie, Allan, Wade, 1992. Interactive gestures. *Discourse Processes* 15, 469–489.
- Brown, Penelope, Levinson, Stephen C., 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press, Cambridge.
- Bühler, Karl, 1982/1934. The Deictic Field of Language and Deictic Words. In: Jarvella, Robert J., Wolfgang, Klein (Eds.), *Speech, Place, and Action*. John Wiley and Sons, Chichester, pp. 9–30.

- Butterworth, George, 2003. Pointing is the royal road to language for babies. In: Sotaro, Kita (Ed.), *Pointing*. Lawrence Erlbaum, Mahwah, pp. 9–34.
- Butterworth, George, Morissette, P., 1996. Onset of pointing and the acquisition of language in infancy. *Journal of Reproductive and Infant Psychology* 14, 219–231.
- Carpenter, M., Nagell, K., Tomasello, Michael, 1988. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development* 63 (4).
- Clark, Herbert H., 1992. *Arenas of Language Use*. University of Chicago Press, Chicago.
- Clark, Herbert H., 1996. *Using Language*. Cambridge University Press, Cambridge.
- Clark, Herbert H., Schreuder, Robert, Buttrick, Samuel, 1983. Common ground and the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior* 22, 245–258.
- De Ruiter, Jan Peter, Wilkins, David P., 1998. The synchronization of gesture and speech in Dutch and Arrernte (an Australian Aboriginal Language): a cross-cultural comparison. In: *Oralité Et Gestualité: Communication Multimodale*, Interaction, pp. 603–607.
- De Ruiter, Jan Peter, Mitterer, H., Enfield, N.J., 2006. Projecting the end of a speaker's turn: a cognitive cornerstone of conversation. *Language* 82 (3), 504–524.
- Efron, David, 1972/1941. *Gesture, Race, and Culture: A Tentative Study of Some of the Spatio-Temporal and 'Linguistic' Aspects of the Gestural Behavior of Eastern Jews and Southern Italians in New York City Living under Similar as Well as Different Environmental Conditions*. Mouton, The Hague/Paris.
- Ekman, Paul, Friesen, Wallace V., 1969. The repertoire of nonverbal behavior: origins, usage, and coding. *Semiotica* 1, 49–98.
- Enfield, N.J., 2001. 'Lip-pointing'-a discussion of form and function with reference to data from Laos. *Gesture* 1 (2), 185–212.
- Enfield, N.J., 2003a. Producing and editing diagrams using co-speech gesture: spatializing non-spatial relations in explanations of kinship in Laos. *Journal of Linguistic Anthropology* 13 (1), 7–50.
- Enfield, N.J., 2003b. Demonstratives in space and interaction: data from Lao speakers and implications for semantic analysis. *Language* 79 (1), 82–117.
- Enfield, N.J., 2004. On linear segmentation and combinatorics in co-speech gesture: a symmetry-dominance construction in Lao fish trap descriptions. *Semiotica* 149 (1–4), 57–123.
- Enfield, N.J., 2005. The body as a cognitive artifact in kinship representations. *Hand gesture diagrams by speakers of Lao*. *Current Anthropology* 46 (1), 51–81.
- Enfield, N.J., 2006. Social consequences of common ground. In: Enfield, N.J., Levinson, Stephen C. (Eds.), *Roots of Human Sociality: Culture Cognition and Interaction*. Berg, London, pp. 399–430.
- Engle, Randi A., 1998. Not channels but composite signals: speech, gesture diagrams and object demonstrations are integrated in multimodal explanations. In: Gernsbacher, M.A., Derry, S.J. (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*. Erlbaum, Mahwah, pp. 321–327.
- Glenberg, A.M., Langston, W.E., 1992. Comprehension of illustrated text: pictures help to build mental models. *Journal of memory and language* 31 (no. 2), 129–151.
- Goldin-Meadow, Susan, 2003. *Hearing Gesture: How Our Hands Help Us Think*. Harvard University Press, Cambridge, MA.
- Goodwin, Charles, 1994. Professional vision. *American Anthropologist* 96 (no. 3), 606–633.
- Goodwin, Charles, 2000. Action and embodiment within situated human interaction. *Journal of Pragmatics* 32, 1489–1522.
- Goodwin, Charles, 2006. Human sociality as mutual orientation in a rich interactive environment: multimodal utterances and pointing in aphasia. In: Enfield, N.J., Levinson, Stephen C. (Eds.), *Roots of Human Sociality: Culture Cognition and Interaction*. Berg, London, pp. 97–125.
- Goody, Jack, 1977. *The Domestication of the Savage Mind*. Cambridge University Press, Cambridge.
- Grice, H. Paul, 1987. *Studies in the way of words*. Harvard University Press, Cambridge.
- Haviland, John, 1993. Anchoring, iconicity, and orientation in Guugu Yimithirr pointing gestures. *Journal of Linguistic Anthropology* 3 (1), 3–45.
- Hayashi, Makoto, 2005. Some ways in which grammar shapes local interactional practices. Paper Presented at 104th Annual Meeting of the American Anthropological Association, Washington DC, December 3, 2005.
- Heritage, John, 2007. Intersubjectivity and progressivity in person (and place) reference. In: Enfield, N.J., Stivers, Tanya (Eds.), *Person Reference in Interaction: Linguistic, Social, and Cultural Perspectives*. Cambridge University Press, Cambridge, pp. 226–252.
- Heritage, John, Raymond, Geoffrey, 2005. The terms of agreement: indexing epistemic authority and subordination in talk-in-interaction. *Social Psychology Quarterly* 68 (no. 1), 15–38.

- Hewes, Gordon W., 1981. Pointing and language. In: Myers, Terry, Laver, John, Anderson, John (Eds.), *The Cognitive Representation of Speech*. North-Holland, Amsterdam, pp. 263–269.
- Hewes, Gordon W., 1996. A history of the study of language origins and the gestural primacy hypothesis. In: Lock, A., Peters, C.R. (Eds.), *Handbook of Human Symbolic Evolution*. Clarendon Press, Oxford, pp. 571–595.
- Hutchins, Edwin, 1995. *Cognition the Wild*. MIT Press, Cambridge, MA.
- Jefferson, Gail, 1989. Preliminary notes on a possible metric which provides for a 'standard maximum' silences of approximately one second in conversation. In: Roger, D., Bull, P. (Eds.), *Conversation: An Interdisciplinary Perspective*. Multilingual Matters, Clevedon, pp. 166–196.
- Kendon, Adam, 1972. Some relationships between body motion and speech: an analysis of an example. In: Siegman, A.W., Pope, B. (Eds.), *Studies in Dyadic Communication*. Pergamon Press, New York, pp. 177–210.
- Kendon, Adam, 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press, Cambridge.
- Kendon, Adam, Versante, Laura, 2003. Pointing by hand in 'Neapolitan'. In: Sotaro, Kita (Ed.), *Pointing*. Lawrence Erlbaum, Mahwah, NJ, pp. 109–138.
- Kita, Sotaro (Ed.), 2003. *Pointing: Where Language Cognition and Culture Meet*. Lawrence Erlbaum, Mahwah.
- Lambrecht, Knud, 1994. *Information structure and sentence form*. Cambridge University Press, Cambridge.
- Larkin, Jill H., Simon, Herbert A., 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive science* 11, 65–99.
- Levinson, Stephen C., 1987. Minimization and conversational inference. In: Verschueren, J., Bertelucci-Papi, M. (Eds.), *The Pragmatic Perspective*. John Benjamins, Amsterdam, pp. 61–129.
- Levinson, Stephen C., 2000. *Presumptive Meanings: The Theory of Generalized Conversational Implicature*. MIT Press, Cambridge, MA/London.
- Levy, Elena T., Fowler, Carol A., 2000. The role of gestures and other graded language forms in the grounding of reference in perception. In: McNeill, David (Ed.), *Language and Gesture*. Cambridge University Press, Cambridge, pp. 215–234.
- Liddell, Scott K., 2003. *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge University Press, Cambridge.
- Liszkowski, Ulf, 2006. Infant pointing at twelve months: communicative goals, motives, and social-cognitive abilities. In: Enfield, N.J., Levinson, Stephen C. (Eds.), *Roots of Human Sociality: Culture Cognition and Interaction*. Berg, London.
- McNeill, David, 1985. So you think gestures are nonverbal? *Psychological Review* 92 (no. 3), 350–371.
- McNeill, David, 1992. *Hand and Mind: What Gestures Reveal About Thought*. University of Chicago Press, Chicago.
- McNeill, David, 2003. Pointing and morality in Chicago. In: Sotaro, Kita (Ed.), *Pointing*. Lawrence Erlbaum Associates, Mahwah, NJ.
- McNeill, David, Cassell, Justine, Levy, Elena T., 1993. Abstract Deixis. *Semiotica* 95 (1–2), 5–19.
- Norman, Donald A., 1991. Cognitive artifacts. In: Carroll, John M. (Ed.), *Designing Interaction: Psychology at the Human-Computer Interface*. Cambridge University Press, Cambridge, pp. 17–38.
- Oestermeier, Uwe, Hesse, Friedrich W., 2000. Verbal and visual causal arguments. *Cognition* 75, 65–104.
- Ohama, K., 1984. Development of pointing behavior in infants and mothers' responsive behavior: longitudinal study of infants from 9 to 30 months. In: Ogino, M., Ohama, K., Saito, K., Takei, S., Tatsuno, T. (Eds.), *The Development of Verbal Behavior VI*. Bulletin of the Faculty of Education, University of Tokyo, Tokyo.
- Peirce, Charles S., 1965/1932. Speculative grammar. In: Hartshorne, Charles, Weiss, Paul (Eds.), *Collected Papers of Charles Sanders Peirce, Volume II, Elements of Logic*. The Belknap Press of Harvard University Press, Cambridge, Mass, pp. 127–269.
- Povinelli, Daniel J., Bering, Jesse M., Giambrone, Steve, 2003. Chimpanzees' "pointing": another error of the argument by analogy? In: Kita, Sotaro (Ed.), *Pointing: Where Language, Culture, and Cognition Meet*. Lawrence Erlbaum, Mahwah, NJ, pp. 35–68.
- Quine, W.V.O., 1960. *Word and Object*. MIT Press, Cambridge, Mass.
- Rolfe, Leonard H., 1996. Theoretical stages in the prehistory of grammar. In: Lock, Andrew J., Peters, Charles R. (Eds.), *Handbook of Human Symbolic Evolution*, pp. 776–792.
- Sacks, Harvey, 1992. *Lectures on Conversation*. Blackwell, London.
- Sacks, Harvey, Schegloff, Emanuel A., 1979/2007. Two preferences in the organization of reference to persons in conversation and their interaction. In: Psathas, George (Ed.), *Everyday Language: Studies in Ethnomethodology*. Irvington, New York, pp. 15–21. (Reprint in press, in: Enfield, N.J., Stivers, Tanya (Eds.), *Person Reference in Interaction: Linguistic, Social, and Cultural Perspectives*. Cambridge University Press, Cambridge, 2007.)
- Scaife, Mike, Bruner, Jerome S., 1975. The capacity for joint attention in the infant. *Nature* 253, 265–266.
- Schaffer, H.R., 1984. *The Child's Entry into a Social World*. Academic Press, New York.

- Schegloff, Emanuel A., 1984. On some gestures' relation to talk. In: Atkinson, J. Maxwell, Heritage, John (Eds.), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge University Press, Cambridge, pp. 266–296.
- Schegloff, Emanuel A., 2007. Conveying who you are: the presentation of self, strictly speaking. In: Enfield, N.J., Stivers, Tanya (Eds.), *Person Reference in Interaction: Linguistic, Social, and Cultural Perspectives*. Cambridge University Press, Cambridge, pp. 123–148.
- Sherzer, Joel, 1973. Verbal and nonverbal deixis: the pointed lip gesture among the San Blas Cuna. *Language in Society* 2, 117–131.
- Stivers, Tanya, Robinson, Jeffrey D., 2006. A preference for progressivity in interaction. *Language in Society* 35 (3), 367–392.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., Moll, H., 2005. Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences* 28, 675–691.
- Tomasello, Michael, 2006. Why don't apes point? In: Enfield, N.J., Levinson, Stephen C. (Eds.), *Roots of Human Sociality: Culture, Cognition and Interaction*. Berg, Oxford, pp. 506–524.
- Wilkins, David P., Kita, Sotaro, Enfield, N.J., 2001. Ethnography of pointing questionnaire (version 2). In: Stephen, C. Levinson, Enfield, N.J. (Eds.), *Manual for the Field Season, 2001*. Language and Cognition Group, Max Planck Institute for Psycholinguistics, Nijmegen, pp. 136–141.
- Wilkins, David P., 2003. Why pointing with the index finger is not a universal (in socio-cultural and semiotic terms). In: Sotaro, Kita (Ed.), *Pointing*. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 171–216.
- Wittgenstein, Ludwig, 1955. *Philosophical Investigations*. Basil Blackwell, Oxford.
- Wundt, Wilhelm, 1972/1921. *The language of gestures*. Mouton, The Hague.
- Zhang, Jiajie, 1997. The nature of external representations in problem solving. *Cognitive Science* 21, 179–217.

Further reading

- Richerson, Peter J., Robert, Boyd, 2004. *Not by Genes Alone: How Culture Transformed Human Evolution*. University of Chicago Press, Chicago.
- Saitz, Robert L., Cervenka, Edward J., 1972. *Handbook of Gestures: Colombia and the United States*. Mouton, The Hague.

N.J. Enfield is a scientific staff member in the Language and Cognition Group of the Max Planck Institute for Psycholinguistics, Nijmegen. His main research interests are the semantics-pragmatics relationship, the language-cognition relationship, the structure and semiotics of face-to-face interaction, and descriptive field work on Lao and other Southeast Asian languages. Publications include *Ethnosyntax* (OUP 2002), *Linguistic Epidemiology* (Routledge 2003), 'The body as a cognitive artifact in kinship representations' (*Current Anthropology* 2005), and *Roots of Human Sociality* (with Stephen C. Levinson, Berg, 2006).

Sotaro Kita is a Reader at the School of Psychology at the University of Birmingham. His main research interests are cognitive psychological, interactional, and ethnographic studies of the relationship between speech and spontaneous co-speech gestures. His research on gesture spans both production and comprehension, and also adults and children. In addition to gesture, his research interests also include child language acquisition, semantics and pragmatics of spatial expressions, and cross-linguistic studies of spatial conceptualization. Publications include *Pointing* (Erlbaum, 2003), 'Children creating core properties of language: Evidence from an emerging sign language in Nicaragua' (*Science*, 2004), and 'What does cross-linguistic variation in semantic coordination of speech and gesture reveal?' (*J. of Memory and Lg.*, 2003).

J.P. de Ruiter is a scientific staff member in the Language and Cognition Group of the Max Planck Institute for Psycholinguistics, and is also involved in international research projects funded by the European Union. His main interests are in multimodal human-human and human-machine communication, gesture, mathematical modeling of human behavior, and turn-taking. Publications include "The production of gesture and speech" (Cambridge University Press, 2000), "Parallelism in the production of noun phrases: Experiments and reaction time models." *Journal of Experimental Psychology* (1999), and "SLOT; a research platform for investigating multimodal communication" (BRMIC, 2003).