

Charles Goodwin  
cgoodwin@humnet.ucla.edu

## Multimodality in human interaction Multimodalidade na interação humana

**ABSTRACT** – Human action is built by actively combining materials with intrinsically different properties into situated contextual configurations where they can mutually elaborate each other to create a whole that is both different from, and greater than, any of its constitutive parts. This has a range of consequences for the organization of language, action, knowledge and embodiment in situated interaction. Two phenomena that depend upon such distributed organization of action will be investigated here. First, Chil, a man who suffered severe damage to the left hemisphere of his brain that left him with a three word vocabulary, *Yes, No and And*, was nonetheless able to act as a powerful speaker in conversation. He did this by operating on the talk of others to lead them to produce the words he needed but could not say himself, and also by using gesture to incorporate meaningful phenomena in his surrounding environment into the organization of his utterance. Second, the processes through which archaeologists acquire the ability to see relevant structure in the dirt they are excavating, and construct the documents, such as maps, that animate the discourse of their profession are investigated. The way in which action is built through the simultaneous use of materials with diverse properties makes it possible for experienced archaeologists to calibrate the professional vision, practice and embodied knowledge of novices, and thus to interactively construct within situated interaction the cognition, ways of seeing and embodied practices of new archaeologists. Both Chil's ability to act as a speaker and the social organization of the embodied knowledge and perception required to act as a member of a scientific community are made possible through the way in which alternatively placed social actors contribute with different kinds of materials to a common course of action.

**Key words:** multimodality, aphasia, professional vision, gesture, pointing, talk-in-interaction.

**RESUMO** – A interação humana é constituída pela combinação ativa de materiais com propriedades intrinsecamente diversas, em configurações contextuais situadas, onde eles podem mutuamente elaborar um ao outro para criar um todo que é diferente e maior que suas partes constituintes. Isso gera uma série de consequências na organização da linguagem, ação, conhecimento e corporificação em interações situadas. Dois fenômenos que dependem dessa organização distribuída da ação serão investigados aqui. Primeiro, Chil, um homem que sofreu um grave comprometimento no hemisfério esquerdo de seu cérebro que o deixou com um vocabulário de três itens: *Sim, Não e E*, mas que se revela capaz de agir como um poderoso falante em interações. Chil consegue isso ao operar na fala dos outros, de forma a fazê-los produzir as palavras de que ele precisa, mas que não poderia ele próprio produzir, e também ao usar gestos na organização de seus enunciados a fim de incorporar fenômenos significativos do ambiente que os cerca. Segundo, são investigados os processos pelos quais arqueólogos adquirem a habilidade de enxergar estruturas relevantes no solo que estão escavando e de construir documentos, tais como mapas, que fomentam o discurso de sua profissão. A maneira com que a ação é construída por meio do uso simultâneo de materiais com propriedades diversas faz com que seja possível que arqueólogos experientes calibrem a visão, a prática e o conhecimento corporificado de profissionais ainda em formação e assim construam, na própria interação situada, a cognição, as formas de ver e as práticas corporificadas dos jovens arqueólogos. Tanto a habilidade de Chil, de agir como um falante, quanto a organização social do conhecimento corporificado e a percepção necessárias para agir-se como um membro de uma comunidade científica são possíveis pela forma com que atores sociais alternativamente situados contribuem com diferentes tipos de materiais para um curso de ação comum.

**Palavras-chave:** multimodalidade, afasia, visão profissional, gestos, apontação, fala-em-interação.

In this paper I will offer some perspectives for how human language, cognition, action and embodiment might be investigated as intrinsically social phenomena. To begin I will investigate how a man who was able to say only three words after severe damage to the left side of his brain, is nonetheless able to act as a powerful speaker in conversation by drawing upon resources provided by his interlocutors, and by using meaningful structure sedimented in the environment around him. It might be objected that the situation of a man with severe aphasia is special. To demonstrate the general importance

of interactive frameworks for the analysis of not only language use, but also cognition and action, I will then use video recordings of excavations at an archaeological field school, to investigate how embodied perception and knowledge — for example the professional vision that enables an archaeologist to see the remains of past activity in a patch of dirt — is organized through embodied interactive practice.

This paper draws extensively on work I have already published elsewhere. Frequently long sections of these early papers are incorporated word for word in

the present paper. The previous articles that I draw upon most extensively are Goodwin (2007b) for the analysis of how someone with aphasia can act as a powerful speaker, and Goodwin (2010) for the social organization of embodiment in archaeology. Many of my other papers are also most relevant to the arguments being made here, including Professional Vision (Goodwin and Goodwin 1992), Environmentally Coupled Gestures (Goodwin 2007a), Action and Embodiment (Goodwin 2000), and current work I am doing on how Chil, the man with severe aphasia, uses varied prosody over identical lexical items to construct very diverse forms of action (Goodwin, s.d.).

### Language complexity and the dialogic organization of language

Many models of the speaker, and of the language produced by a speaker, focus on the production of rich symbolic structures by a single individual. Thus formal linguistics asks how rich grammatical sentences can be constructed through systematic mental operations within the speaker. Even scholars such as Bakhtin (1981) and Volosinov (1973) who view language as thoroughly social use as their primary data rich language structure that makes

possible phenomena such as reported speech (Goodwin, 2007b). Here the talk produced by a prior speaker enters into a dialogic relationship with the talk of the current speaker through the way in which it is embedded within the current speaker's language and consciousness.

Using Volosinov (1973) as a point of departure Goffman (1981) developed a rich and important model that deconstructed the speaker into a range of different entities who can exist simultaneously within the scope of a single utterance.

Figure 1 is a story in which a teller quotes something that her husband said. The story is about one of the prototypical scenes of middle class society. Friends have gotten a new house. As guests visiting the house for the first time the speaker and her husband, Don, were in the position of admiring and appreciating their hosts' new possessions. However, while looking at the wallpaper in house Don asked the hosts if there were able to "pick it out" (choose their own wallpaper), or were forced to accept wallpaper chosen by the builder of the house, "take this wallpaper" (lines 13-16).

Who is speaking in lines 14 and 16? The voice that is heard is Ann's, the current story teller. However she is quoting something that her husband, Don, said, and

- (4.0)
- 1 Ann: Well- ((throat clear)) (0.4)  
 2 We coulda used a liddle, marijuna.  
 3 tih get through the **weekend**.  
 4 Beth: What happened.  
 5 Ann: Karen has this new **hou:se**.  
 6 En it's got all this like (0.2) ssilvery::g-go:ld  
 7 wwa:llpaper,  
 8 \*hh (h) en D(h)o(h)n sa(h)ys,  
 9 y'know this's th'firs'time we've seen this house.  
 10 =Fifty five thousn dollars in Cherry Hill.=Right?  
 11 (0.4)  
 12 Beth: Uh **hu:h**?  
 13 Ann: Do(h)n said (0.3)  
 14 **dih-did they ma:ke you take this-wa(h)llpa(h)p(h)er?**  
 15 Beth: hh!  
 16 Ann: =er(h)di-dju pi(h)ck i(h)t ou(h)t.  
 17 Beth: Ahh huh huh huh huh=

Figure 1. Reporting the speech of another.

moreover presenting what he did as a terrible faux pas, an insult to their hosts in the narrated scene. She is both quoting the talk of another, and also taking up a particular stance toward what was done through that talk. In a very real sense Ann (the current story teller) and Don (the principal character in her story) are both “speakers” of what is said in lines 14 and 16, though in quite different ways. The analytic framework offered by Goffman in Footing for what he called the Production Format of an utterance provides powerful tools for deconstructing the “speaker” into a complex lamination of structurally different kinds of entities (see Figure 2).

In terms of the categories offered by Goffman Ann is the Animator, the party whose voice is actually being used to produce this strip of speech. However, the Author of this talk, the party who constructed the phrase said, is someone else, the speaker’s husband Don. In a very real sense he is being held accountable as not only the author of that talk, but also its Principal, a party who is socially responsible for having performed the action done by the original utterance of that talk. Goffman frequently noted that the talk of speakers in everyday conversation could encompass an entire theater. And indeed here Ann is putting Don on stage as a character in the story she is telling, or in Goffman’s terms animating him as a Figure.

Moreover there is a complex laminated and temporal interdigitation among these different kinds of entities within the space of Ann’s utterance. Thus it would be impossible to mark this as a quotation by putting quotation marks before and after what Don said. In addition to the report of this talk, the utterance also contains a series of laugh tokens, which are not to be heard as part of what Don said, but instead as the current speaker’s, Ann’s, commentaries on what Don did through that talk. Through her laugh tokens Ann both displays her own stance towards Don’s utterance, formulating his talk as something to be laughed at, and, through the power of laugh tokens to act as invitations for others to join in the laughter (Jefferson, 1979), invites others to join in such treatment. Ann thus animates Don as a figure in her talk while simultaneously providing her own commentary on what he said by placing her own laugh tokens throughout the strip of speech being quoted.

In brief, in Footing Goffman provides a powerful model for systematically analyzing the complex theater of different kinds of entities that can co-exist within a single strip of reported speech. The analytic framework he develops sheds important light on the cognitive complexity of speakers in conversation, who are creating a richly inhabited and textured world through their talk.

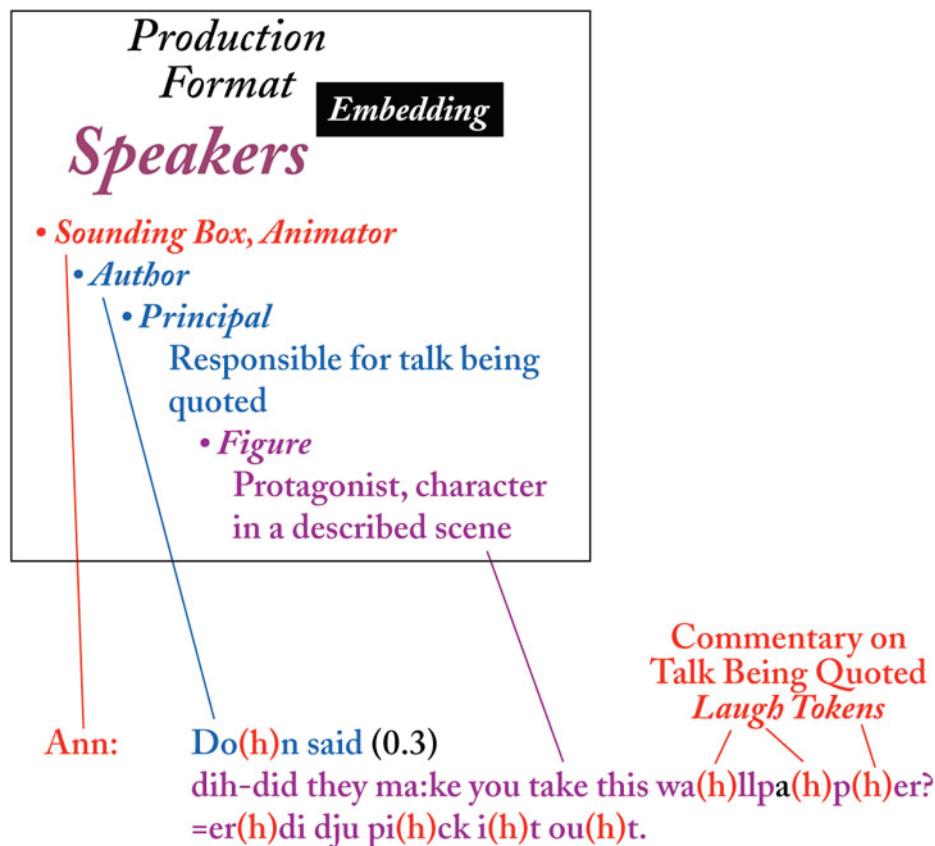


Figure 2. A laminated speaker.

In addition to producing a meaningful linguistic sentence, Ann, within the scope of a single utterance, creates a socially consequential image of another speaker. His talk is thoroughly interpenetrated with another kind of talk that displays her stance toward, and formulation of both what he said (e.g., as a laughable of some type), and the kind of person that would say such a thing. Goffman's deconstruction of the speaker provides us genuine analytic insights, and tools for applying those insights to an important range of talk.

Goffman's speaker, a laminated structure encompassing quite different kinds of entities who co-exist within the scope of a single utterance, is endowed with considerable cognitive complexity. However, no comparable semiotic life animates Goffman's hearers. In a separate section of the article they are described as cognitively simple points on an analytic grid listing possible types of participation in the speech situation (e.g., Addressee vs. Overhearer, etc.). Because of the way in which Goffman's model (and Volosinov's) focuses exclusively on phenomena within the structure of talk, the visible bodies of hearers, and the ways in which parties other than the speaker might participate reflexively in the ongoing organization of an utterance are rendered invisible. There are powerful reasons for such logocentricism. For thousands of years human beings have been grappling with the issues raised by the task of capturing significant structure in the stream of speech in writing. Writing systems, and the insights and methodological tools they have provided for the analysis of linguistic and phonetic structure, the creation of precise records that can endure in time and be transported from place to place, etc. are major accomplishments that provide a crucial infrastructure for much of research into language structure, verbal genres and more recently talk-in-interaction. However, such a bias toward what can be written renders many crucial phenomena, including the simultaneous embodied actions of hearers, invisible and inaccessible to analysis (Linell, 2005).

Contemporary video and computer technology makes it possible to repeatedly examine the bodies as well as the talk of participants in interaction, and thus to move analytically beyond logocentricism. And indeed some evidence suggests that neither talk, nor language itself, are self-contained systems, but instead function within a larger ecology of sign systems (Goodwin, 2000).

### ***Building an utterance in concert with others***

The analysis of language from the perspective of formal linguistics, as well as the social models of language of Bakhtin, Volosinov, and Goffman require as a point of departure utterances that have rich syntax, e.g., clauses in which the talk of another that is being reported is embedded within a larger utterance by the current

speaker. The necessity of rich syntax not only excludes important activities, such as many greetings which, at least in English, are frequently done with one to two word utterances (e.g., "Hi") (Schegloff, 1972), but also certain kinds of speakers.

Because of a severe stroke suffered when he was 65 years old, Chil, whose actions we will now investigate, was able to say only three words: *Yes*, *No* and *And*. It is impossible for him to produce the syntax that Goffman's Production Format and Volosinov's Reported Speech seem to require (i.e., he can't produce a sentence such as "John said X"). Someone such as Chil appears to fall beyond the pale of what counts as the competent speaker required for either their dialogic analysis or formal linguistics.

Chil in fact acts as a powerful speaker in interaction, and moreover one who is able to include the talk of others in his utterances. Describing how he does this requires a model of the speaker that moves beyond the individual. The sequence in Figure 3 provides an example. Chil's son Chuck and daughter-in-law Candy are talking with him about the amount of snow the winter has brought to the New York area where Chil lives. After Candy notes that not much has fallen "this year" (which Chil strongly agrees with in talk omitted from the transcript), in line 11 she proposes that such a situation contrasts markedly with the amount that fell "last year." Initially, with his "yeah-" Chil seems to agree (in the interaction during the omitted talk Chil was strongly agreeing with what Candy was saying, and thus might have grounds to expect and act as though that process would continue here). However, he ends his agreement with a cut-off (thus visibly interrupting and correcting his initial agreement) and moves to strong, vivid disagreement in line 13. Candy immediately turns to him and changes her "last year" to "the year before last." Before she finishes Chil (line 15) affirms the correctness of her revised version.

Despite his severely impoverished language Chil is able to make a move in the conversation that is both intricate and precise: unlike what Candy initially proposed in line 10, it was not "last year" but the "the year before last" when there was a lot of snow. Chil says this by getting someone else to produce just the words that he needs. The talk in line 14 is semantically and syntactically far beyond anything that Chil could say on his own.

Though not only spoken, but constructed by Candy, it would be clearly wrong to treat line 14 as a statement by her. First, just a moment earlier, in lines 10 and 12, she voiced the position that is being contradicted here. Second, as indicated by Chil's agreement in line 15, Candy is offering her revision as something to be accepted or rejected by Chil, not as a statement that is epistemically her own. Line 14 thus seems to require a deconstruction of the speaker of the type called for by Goffman in Footing, with Candy in some sense being an animator, or "sounding box" for a position being voiced by Chil. However, the



## The participants have been discussing snow in the area where Chil Lives

1 Candy: You haven't had **that** much this year have you.

...((8 lines ommitted))

10 Candy: But **la** **st** year. Whoo!  
 11 Chuck: °mm  
 12 Candy: [ In the l [ ast year-  
 13 Chil: °yeah- [ No No. No.:  
 14 Candy: er the year before la [ st.  
 15 Chil: [ Yes.



**Figure 3.** Using the language abilities of another.

analytic framework offered in Footing does not accurately capture what is occurring here. Though Candy is in some important sense acting as an Animator for Chil, he is not a cited figure in her talk, and no quotation is occurring. Intuitively the notion that Chil is in some sense the Author of line 14, and its Principal, seems plausible (what is said here would not have been spoken without his intervention, and he is treated as the ultimate judge of its correctness). However, how could someone completely unable to produce either the semantics or the syntax of line 14 be identified as its author?

Clarifying such issues requires a closer look at the interactive practices used to construct the talk that is occurring here. Chil's intervention in line 13 is an instance of what Schegloff, Jefferson and Sacks (1977) describe as Other Initiated Repair. With his "No No. No.:" Chil forcefully indicates that there is something wrong with what Candy, the prior and still current speaker, has just said. She can re-examine her talk to try and locate what needs repair, and indeed here that process seems straightforward. In response to Chil's move Candy changes "last year", the crucial formulation in the talk Chil is objecting to, to an alternative "the year before last".

Such practices for the organization of repair, which are pervasive not only in Chil's interaction, but in the talk of fully fluent speakers as well (Schegloff *et al.*, 1977), have crucial consequences for both Chil's ability to function as a speaker in interaction, and for probing the analytic models offered by Goffman and Volosinov. First, through the way in which Chil's Yes's and No's are tied to specific bits of talk produced by others (e.g., what Candy has just said) they have a strong indexical component which allows him to use as a resource detailed structure in the talk of others, and in some sense incorporate that

talk into his own, linguistically impoverished utterances. Thus in line 13 he is heard to be objecting not to life in general, but to precisely what Candy said in line 12, and to be agreeing with what she said in line 14. Second, such expansion of the linguistic resources available to Chil is built upon the way in which his individual utterances are embedded within sequences of dialogue with others, or more generally the sequential organization of interaction. However, this notion of dialogue, as multi-party sequences of talk, was precisely what Volosinov (1973, p. 116) worked to exclude from his formulation of the dialogic organization of language. Nonetheless, Chil's actions here provide a clear demonstration of the larger Bakhtinian argument that speakers talk by "renting" and reusing the words of others.

Third, what happens here requires a deconstruction of the speaker that is relevant to, but different from, that offered by Goffman in Footing. What Chil says with his "No" in line 13 indexically incorporates what Candy said in line 11, though Chil does not, and cannot, quote what she said there. Instead of the structurally rich single utterance offered in Goffman's model of multiple voices laminated within the complex talk of a single speaker, here we find a single lexical item, a simple "No", that encompasses talk produced in multiple turns (e.g., both lines 11 and 13) by separate actors (Candy and Chil). Unlike Ann's story in Figure 1 Chil's talk cannot be understood or analyzed in isolation. Its comprehension requires inclusion of the utterances of others that Chil is visibly tying to.

Rather than being located within a single individual, the speaker here is distributed across multiple bodies and is lodged within a sequence of utterances. Chil's competence to manipulate in detail the structure of emerging talk

by objecting to what has just been said, that is to act in interaction, constitutes him as a crucial Author of Candy's revision in line 14, despite his inability to produce the language that occurs there. Though not reporting the speech of another Candy speaks for Chil in line 14, and locates him as the Principal for what is being said there. All of this requires a model of the speaker that takes as its central point of departure not the competence to quote the talk of another (though being able to incorporate, tie to, and reuse another's talk is absolutely central), but instead the ability to produce consequential action within sequences of interaction.

Fourth, the action occurring here and the differentiated roles parties are occupying within it are constituted not only through talk, but also through participation as a dynamically unfolding process. As line 13 begins Candy has turned away from Chil to gaze at Chuck. Chil's talk in line 13 pulls Candy's gaze back to him (her eyes moves from Chuck to Chil over the last of his three No's). Such securing the gaze of an addressee is similar to way in which fluent speakers use phenomena such as restarts to obtain the gaze of a hearer before proceeding with a substantive utterance (Goodwin, 1980, 1981).

In this case, however, it is the addressee, Candy, rather than Chil, the party who solicited gaze, who produces the talk that follows. Nonetheless, through the way in which he organizes his body Chil displays that he acting as something more than a recipient of Candy's talk, and instead sharing the role of its speaker. Typically gestures are produced by speakers. Indeed the work of McNeill (1992) argues strongly that an utterance and the gesture accompanying it are integrated components of a single underlying process. Line 14 is accompanied by gesture. However it is performed not by the person speaking, Candy, but instead by Chil (see Figure 4).

Chil thus participates in Candy's utterance by performing an action usually reserved for speakers, and in so doing visibly displays that he is in some way acting as something more than a hearer. The gesture seems to

provide a visual version of what Candy was saying, and specifically to illustrate the notion that one unit (which can be understood as a "year" through the way in which the gesture is temporally bound to Candy's talk) has another that precedes it. As Candy says "the year" Chil raises his hand toward her with two fingers extended. Then as she says "before last" he moves his gesturing hand down and to the left (see Figure 6). Even if this interpretation of the gesture must remain speculative (for participants as well as the analyst) because of Chil's inability to fully explicate it with talk of his own, the gesture is precisely coordinated with the emerging structure of Candy's talk, and vividly demonstrates Chil's participation in the field of action being organized through that talk.

Rather than being constituted through rich symbolic structures lodged within the mental life of an individual, the speaker found hear is distributed across multiple bodies, and the signs used to build the utterance and action found here, extend into embodied action beyond the stream of speech. The utterance, and the proposition it expresses, is both multi-party and multi-modal.

The following provides another example of how the position of speaker is distributed across multiple bodies, and lodged within the sequential organization of dialogue. Here Chil's daughter Pat and son Chuck are planning a shopping expedition. Once again Chil intercepts a speaker's talk with a strong "No" (lines 6-7 in Figure 5). Pat is talking about the problem of finding socks that fit over Chil's leg brace, since the store where she bought them last went out of business.

What occurs here has is structurally similar to the "Year Before Last" sequence examined in Figures 5 and 6. After Chil uses a "No" to challenge something in the current talk, that speaker produces a revision, which Chil affirms. Once again Chil is operating on the emerging sequential structure of the local dialogue to lead another speaker to produce the words he needs. However, while Candy in Figure 5 could locate the revision needed through a rather direct transformation of the talk then in



**Figure 4.** Separate participants produce talk and its accompanying gesture.

- 1 Pat: So you always need the black socks:  
 2 Chuck: °umm.  
 3 (1.7)  
 4 Pat: It's too bad we found **those** there **too**  
 5 at the store that went outa busi [ness.  
 → 6 Chil: [Na- °dee **No:=**  
 7 =dih dih dih-  
 8 (0.5)  
 → 9 Pat: Yeah.=**You** went to **Bergenfield**.  
 10 Chil: **Ye:s**.

Figure 5. Talk alone is inadequate.

progress (changing “last year” to “the year before last”), the resources that Pat uses to construct her revision are not visible in the transcript. How is she able to find a completely different store, and moreover locate it geographically? When a visual record of the exchange is examined we find that in addition to talk, Chil produces a vivid pointing gesture as he objects to what Pat is saying. Pat treats this as indicating a particular place in their local neighborhood, a store in an adjacent town in the direction Chil is pointing (see Figure 6).

Chil constructs his action in lines 6-7 by using simultaneously a number of quite different meaning making practices that mutually elaborate each other. First, as was seen the “last year” example in Figures 6 and 7, by precisely placing his “No” (again overlapping the statement being challenged), Chil is able use what is being said by another speaker as the indexical point of departure for his own action. His hearers can use that talk to locate something quite specific about what Chil is trying to indicate (e.g., that his action concerns something about the place where the socks were bought). Nonetheless, as this example amply demonstrates, such indexical framing is not in any way adequate to specify precisely what Chil is attempting to say (e.g., in lines 4-5 there is no indication of a store in Bergenfield). However, Chil complements his “No” with a second action, his pointing gesture. In isolation such a point could be quite difficult for an addressee to interpret. Even if one were to assume that something in the environment was being indicated, the line created by Chil’s finger extends indefinitely. Is he pointing toward something in the room in front of them, or as in this case, a place that is actually miles away? However, by using the co-occurring talk a hearer can gain crucial information about what the point might be doing (e.g., indicating where the socks being discussed were bought). Simultaneously the point constrains the rather open ended indexical field provided by the prior talk by indicating an alternative to what was just said. By themselves both the

talk and the pointing gesture are partial and incomplete. However when each is used to elaborate, and make sense out of the other, a whole that is great than the sum of its parts is created (see also Wilkinson *et al.*, 2003).

The ability to properly see and use Chil’s pointing gesture requires knowledge of the structure of the environment being invoked through the gesture. As someone who regularly acts and moves within Chil’s local neighborhood Pat can be expected to recognize such structure. A stranger would not. Chil’s action thus encompasses a number of quite different semiotic fields (Goodwin, 2000) including his own talk, the talk of another speaker that Chil’s “No” is tied to, his gesturing arm, and the spatial organization of his surroundings. Though built through general practices (negation, pointing, etc.), Chil’s action is situated in, and reflexively invokes, a local environment that is shaped by both the emerging sequential structure of the talk in progress, and the detailed organization of the lifeworld that he and his interlocutors inhabit together.

One pervasive model of how human beings communicate conceptualizes the addressee/hearer as an entity that simply decodes the linguistic and other signs that make up an utterance, and through this process recovers what the speaker is saying. Such a model is clearly inadequate for what occurs here. To figure out what Chil is trying to say or indicate Pat must go well beyond what can actually be found in either her talk or Chil’s pointing gesture. Rather than in and of themselves encoding a proposition the signs Chil produces presuppose a hearer who will use them as a point of departure for complex, contingent inferential work. Chil requires a cognitively complex hearer who collaborates with him in establishing public meaning through participation in ongoing courses of action.

The participation structures through which Chil is constituted as a speaker are not lodged within his utterance alone, but instead distributed across multiple

## Multiple Contextual Fields

**Local Talk**    **No**

**Pointing Gesture**    **Structured Space of the Local Lifeworld**

1 Pat:        So you always need the black socks:

2 Chuck:     °umm.

3                (1.7)

4 Pat:        It's too bad we found *those* there *too*

5                at the store that went outa business.

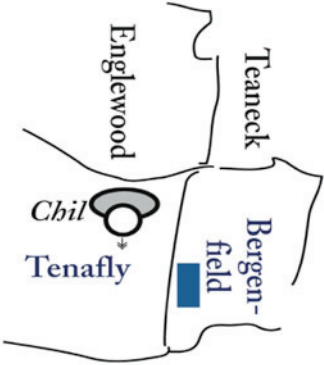
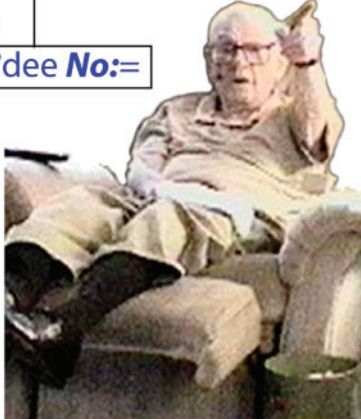
6 Chil:        Na- °dee **No:=**

7                =dih dih dih-

8                (0.5)

9 Pat:        Yeah.=**You** went to **Bergenfield**.

10 Chil:       **Yes.**

**Recipient Gloss in Next Turn Provides Talk that Explicates Gesture**

**Figure 6.** Building meaning with different kinds of signs that mutually elaborate each other.

utterances and actors. In line 9 Pat responds to Chil's intervention by providing a gloss of what she takes him to be saying "You went to Bergenfield". Chil affirms the correctness of this with his "Yes" in line 10. If this action is analyzed using only the printed transcript as a guide it might seem to constitute a simple agreement with what Pat said in line 9. However when a visual record of the interaction is examined Chil can be seen to move his gaze from Pat to Chuck as he speaks this word (Figure 7).

Chuck, who is visiting, lives across the continent. He is thus not aware of many recent events in Chil's life, including the store in Bergenfield that Pat has just recognized (though Chuck, who grew up in this town, is familiar with its local geography). With his gaze shift (and the precise way in which Chil speaks "Yes" which is beyond my abilities to appropriately indicate on the printed page) Chil visibly assumes the position of someone who is telling Chuck about this store. Chil thus acts as not only the author, but also the speaker and teller, of this news. He has of course excellent grounds for claiming this position. A moment earlier, in lines 4-5 Pat said something quite different, and it was only Chil's intervention that led her to produce the talk he is now affirming. Within the single syllable of line 10 Chil builds different kinds of action for structurally different kinds of recipients: first,

a confirmation of what Pat, someone who knows about the event at issue and now recognizes it, has just said and second, a report about that event to Chuck, an unknowing recipient.

Both Volosinov's analysis of Reported Speech and Goffman's deconstruction of the speaker focused on the isolated utterance of a single individual who was able to constitute a laminated set of structurally different kinds of participants by using complex syntax to quote the talk of another. By way of contrast the analytic frameworks necessary to describe Chil's speakership in line 18, must move beyond him as an isolated actor to encompass the talk and actions of others, which he indexically incorporates into his single syllable utterance in line 10. Moreover grasping his action requires attending to not only structure in the stream of speech but also his visible body, and relevant structure in the surround. Chil's speakership is distributed across multiple utterances produced by different actors (e.g., Pat's talk in both lines 13 and 17 is a central part of what is being reported through his "Yes"), and encompasses non-linguistic structure provided by both his visible body and the semiotic organization of the environment around him. His talk is thoroughly dialogic. However analysis of how it incorporates the talk of others in its structure requires moving beyond the models for reported speech and the speaker provided by Volosinov



- 4 Pat: It's too bad we found **those** there **too**  
 5 at the store that went outa busi-ness.  
 6 Chil: Na- °dee **No:=**  
 7 =dih dih dih-  
 8 (0.5)  
 9 Pat: Yeah.=**You** went to **Bergenfield**.  
 10 Chil: **Yes.**  
 11 Chuck: °uh huh. uh huh. ((Nods to Chil))  
 12 Pat: To the men's shop.

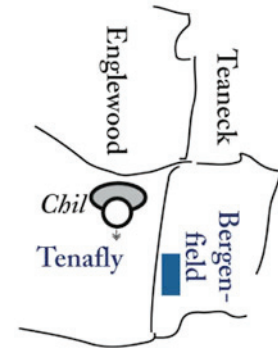
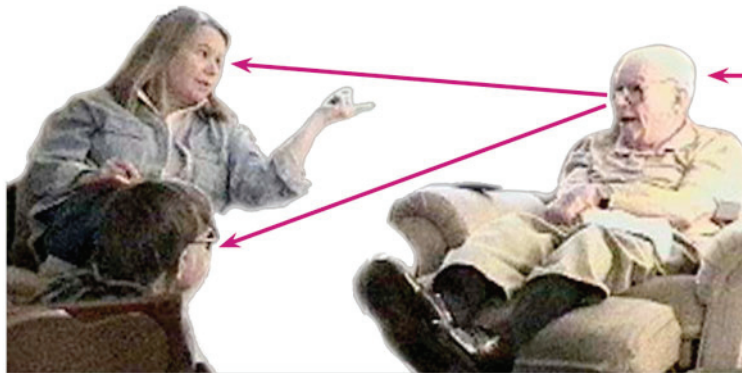
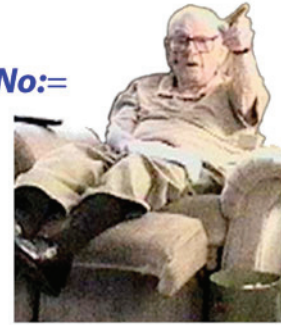


Figure 7. A speaker distributed across multiple actors and sign structures.

and Goffman.

### The interactive construction of cognitively rich, embodied actors

The multimodal, interactive organization of language and the body that was central to Chil's ability to act as a speaker and build action in concert with others is not restricted to special cases, such as aphasia. Moreover these same practices can encompass not just talk and the bodies of speakers and hearers, but also objects in the world, and the organization of embodied practice that sits at the center of professional skill.

To examine such phenomena we will now look video recordings of interaction between archaeologists involved in the process of seeing and mapping relevant structure that becomes visible to their professional eyes in the dirt they are examining. The recordings were made a field school (the analysis being presented here is taken from Goodwin, 2010). A senior archaeologist is guiding the developing vision of a newcomer who is trying to see and map the remains of ancient buildings which are now visible only as faint color patches in the dirt she is examining. The professional vision (Goodwin, 1994) that is being developed here, specifically the ability to see the world as an archaeologist and to trust others within that

community to see the world in the same way, is central to the social organization of (i) both perception and action, (ii) to embodiment as something that is socially organized, and (iii) to the forms of perspectival seeing that are central to a range of analytic concepts, including the notion of culture in anthropology.

Sitting at the heart of the anthropological notion of culture is the observation that different social groups see and classify the environment, and the things found within it, in radically different ways. Cultural anthropology provides many rich descriptions of the varied category systems found in diverse cultures. However, the possibility of such diversity raises the question, not simply of difference, but rather of how it is possible, without some form of mind reading, for the separate individuals within a community (such as the profession of archaeology) to reliably locate the same objects within the complex perceptual environments that are the focus of their group's scrutiny, and to classify what they see in a congruent fashion. How do archaeologists not only see phenomena of interest to them, such as post molds and plow scars, in the amorphous field of subtle color differences provided by the dirt they are examining, but also trust other archaeologists, but not outsiders, to reliably see the same thing? The proper classification of such abilities is not something that is lodged within the

mental life of the individual. Rather, the task of separate individuals seeing, classifying and working with the things that are the focus of their work in a congruent fashion is posed by the necessity of accomplishing joint action in collaboration with each other.

In my own work I have found that a useful place to investigate such issues is provided by the settings of apprenticeship within which newcomers become competent members of professions such as archaeology, surgery and chemistry. I will now briefly examine some of the work done by a young archaeological student, Sue, on one of the very first days of her first fieldschool. She is faced with the task of defining an archaeological feature (a general term for a structure visible in the earth that is analytically relevant to the work of archaeologists) by using the tip of her trowel to outline the shape of a post mold that supported the roof of an ancient house so that it can be drawn on a map. The map is a most necessary record since the post mold is only visible in the color patterning of the dirt now being worked with, and the shapes that constitute it will be destroyed as that dirt is removed to excavate deeper. Her task thus encompasses three of the mundane objects that provide the material and cognitive infrastructure of archaeology as a profession: first a feature, the material traces of the activities of an earlier human society; second, a tool, in this case a trowel, that is being used to reveal such features in the dirt that constitutes, quite literally, the primordial ground for archaeological practice; and third a map, a portable record of what was to be seen in a dirt surface that was later destroyed.

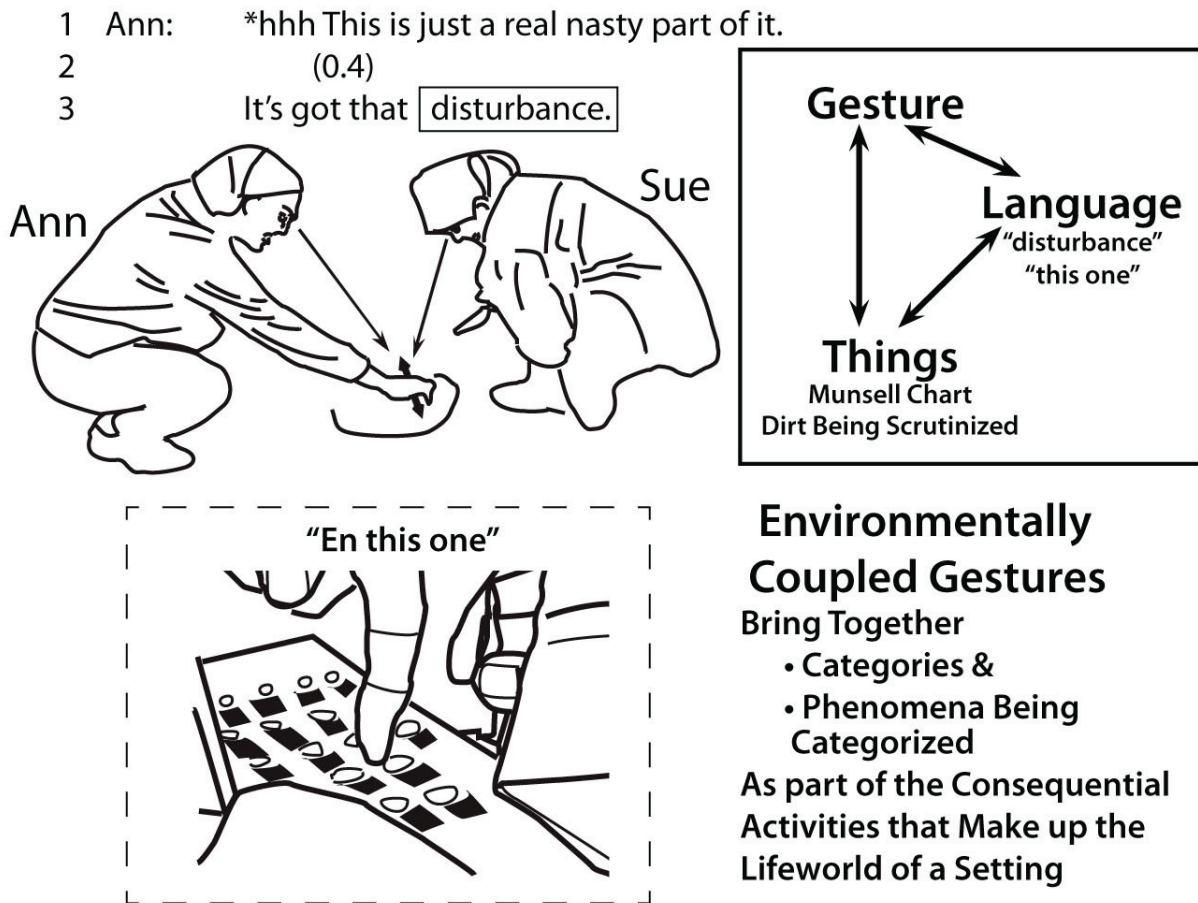
Sue has reached a place in the dirt where it is difficult to see the shape of the post mold she is attempting to outline. Ann, a senior Archaeologist who is directing the field school, traces her finger along a section of the post mold while saying “This is just a real nasty part of it” and then a moment later moves her hand over a long stripe in the dirt that she describes as a “disturbance.” As is seen in the top of Figure 7 the thumb and fingers of Ann’s hand, which are held in an inverted U shape, delineate the width of the stripe while her moving hand traces its length (Figure 8).

Ann builds her actions here through a triad of structurally different kinds of sign resources — language, her gesturing hand, and the dirt with its color patterning — that mutually elaborate each other to create a whole that is not only greater than, but different from, any of its component parts. Sue could not appropriately grasp what Ann is telling her about how to do her work if she attended to any component of this triad in isolation, for example, if she simply listened to Ann’s talk or focused only on the dirt. Such environmentally coupled gestures (Goodwin, 2007a), which link things in the world to embodied action and classifications of those things in ways that are relevant to local participants (a “disturbance” that obscures a

feature being mapped), are common, and indeed pervasive in some settings, such as archaeological excavations (the point to the Munsell chart at the bottom of Figure 7, provides another example). Why might this be the case? Note that a purely symbolic understanding of work relevant categories, such as “disturbance” or “post mold” is completely inadequate for a practicing archaeologist. Knowing in the abstract that a disturbance is something that deforms stratigraphy or features in no way provides a working archaeologist with the skills and professional vision required to competently locate disturbances with their rich physical variety — material traces of plows, burrowing rodents, etc. — in the actual dirt that it is her job to excavate. However, environmentally coupled gestures bring together in a single action package relevant categories and the actual things being categorized as part of the consequential activities that make up the lifeworld of a setting. They thus help negotiate through situated practice the gap noted by Wittgenstein (1958) between a rule (in this case a category) and its application, here the things in environment that are to be seen as instantiations of that category. Simultaneously, in instructional settings such as fieldschools, they provide resources for constituting through endogenous social practice both the things, such as post molds and maps, that are the focus of a community’s work, and the community’s embodied actors who can be trusted to appropriately recognize and work with those things in precisely the ways that are relevant to the concerns of the community (locating and mapping features for example).

The ongoing transformation of environments such as those found in Figure 8 provide crucial resources for calibrating through public practice the professional vision required to see, recognize and properly work with the things that are the focus of the work of a community. Drawing the line that outlines a feature transduces into the dirt being excavated, that is into a public arena where it can be inspected by others, the precise way in which the person drawing the outline has seen the feature, where exactly she has located its boundaries. This construction of the humanly made shape that will be later transferred to the map constitutes an act of categorization, specifically the creation of an iconic sign representing crucial aspects of the thing being attended to in the dirt. Indeed the activity of defining a feature is one central place where the raw material provided by the dirt that is the focus of archaeological scrutiny is transformed into the relevant objects, such as shapes on maps, that animate the distinctive discourse of archaeology. It is here that a natural thing, a color stain in a patch of dirt, is transformed into a cultural object that is consequential in the cognitive work of a specific community.

However, unlike the identically shaped figure on the map that will be carried away from the field site, the sign created by the outline in the dirt is situated in the midst of



**Figure 8.** Environmentally coupled gestures.

the same visual and material field as the feature it depicts. It has not yet been removed from the very color patterning in the dirt that it is representing. The liminal position of this sign, the way it is positioned simultaneously within the messy particulars of the dirt being coded as well as in the world of clean, humanly made iconic representations of archaeologically relevant objects, provides crucial resources for the calibration of professional vision and practice. Thus another archaeologist can systematically judge the accuracy and skill of the work practices of a newcomer by comparing the outline drawn with the shape that the competent practitioner sees in the dirt itself. Note that such a comparison becomes impossible once the figure is removed from the dirt and only the map can be scrutinized.

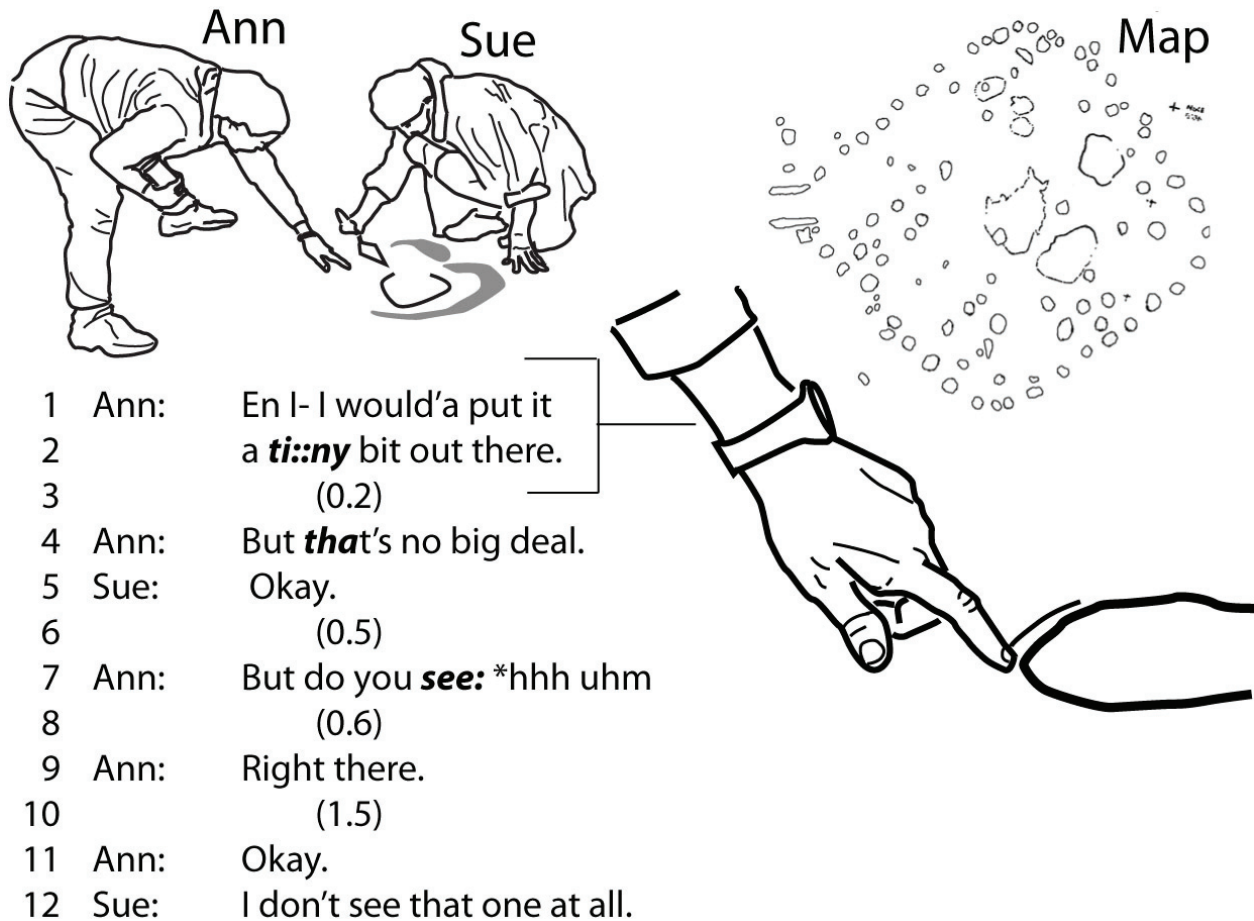
By making additional marks in the dirt the skilled archaeologist can use these same resources to make public the precise details of how she, in contrast to the newcomer, sees the shape. The sequence in Example 9 occurred when Ann, the senior archaeologist, inspected an outline that Sue had drawn.

In lines 1-2, Ann uses her finger to show precisely

where she would have drawn the outline differently, making a moving pointing gesture that leaves a slight mark in the dirt just outside Sue's circle. The work relevant seeing of the post mold being worked with is calibrated across multiple actors through systematic practices that leave visible traces in a public arena, indeed the field that contains the actual object being worked with. Such practices provide systematic resources for accomplishing the intergenerational transmission of just those ways of recognizing relevant objects and using tools to work with them (in this case rendering the object visible through the skilled use of a trowel) that constitute the cognitive infrastructure of a profession.

What is central to this process is not only the visible, material presence of the objects being worked with, and the possibility of manipulating, classifying and annotating relevant phenomena within a field of action that enables public, multi-party scrutiny, but also the organization of collaborative action within interaction. By virtue of their embodied co-presence in a relevant setting Ann is able to see not only the actual environment that is the distinctive focus of her profession's scrutiny (the dirt





**Figure 9.** The dialogic organization of shared vision.

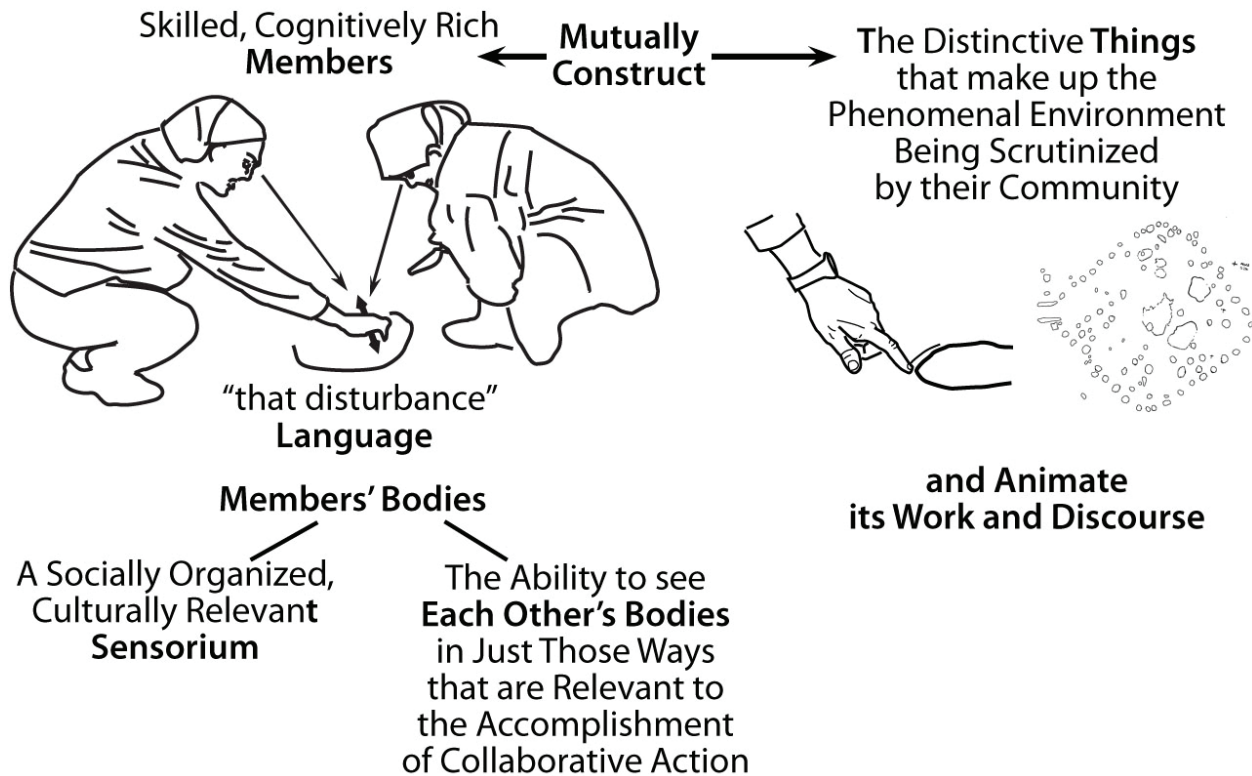
floor of an emerging excavation), but also the operations that a newcomer is performing on that environment as she attempts to locate and work with the things that any competent member should see there. Moreover, since Ann is not simply an observer, but someone engaged in collaborative interaction with Sue, she can and does use what Sue has done as the point of departure for her own next actions. The mark she makes with her finger indicating where she would have located the boundary of the feature does not stand alone as an isolated action, but is, instead, a visible next action to Sue's line right next to it. Ann's new mark, and the talk accompanying the gesture, critique and correct what Sue has done by offering an alternative to where she has visibly located the feature.

Retrospectively Ann uses what Sue has done as an organizing framework for the construction of her own action. Prospectively, Ann's new mark, and the accompanying talk that categorizes that mark as, unlike Sue's, a correct delineation of the feature, creates a transformed environment for new work-relevant seeing that Sue should now perform (comparing Ann's mark with the color patterning in the dirt, "Do you see:" in

line 7), and makes relevant a reply from Sue. In line 12, after noticeably failing to see the patterning that Ann is indicating, Sue says "I don't see that one at all." What is crucial here is not Sue's honest admission that she cannot see what Ann wants her to see, but rather the way in which the sequential organization of action in interaction (Heritage, 1984; Sacks *et al.*, 1974; Schegloff, 1968) creates continuously updated public contexts within which actors use the present state of the environment as the point of departure for building a next action (for example Ann's placement of her mark adjacent to Sue's line) and in so doing creates a new or modified context that shapes what can happen after that. This architecture for intersubjectivity, lodged within ongoing interaction with both other actors and a consequential material world, provides the resources that enable calibration of the professional vision required for members of a community to recognize in common the things they trust each other to see in the environment that is the focus of their work, and to master the practices required to properly work with those things (for example recognize a post mold and transfer its shape to a map). Acquisition of the



## The Reciprocal Constitution of Things, Actors and Communities



**Figure 10.** Building communities and cognition through public interactive practice.

practices required to construct a map simultaneously constructs the relevant cognitive architecture of the archaeologists who use such maps to do their work (Figure 10).

Historically, the human sciences have partitioned study of the phenomena that constitute human life to different disciplines. Thus, Saussure claimed language as the subject matter of a distinct discipline, one that should exist should as a self-sufficient domain of research that was separate from others, such as sociology and psychology. Within Anthropology one subdiscipline, Archaeology, focuses on material objects that have endured from earlier societies, while another, Linguistic Anthropology, takes language as its subject matter, while other subfields investigate social organization, biology and culture. While great gains have been made by such disciplined inquiry restricted to specific phenomena, human action in fact transcends such boundaries. As is well demonstrated by the work of Conversation Analysts, the organization of talk-in-interaction is not simply the place where language emerges in the natural world, but an elementary form of human sociality. The study of human social organization requires intense analysis of

the details of human language use, and the analysis of language requires attention to endogenous social practices through which language is articulated as social practice in the lived social world. In this paper I have attempted to demonstrate how analysis of the actual practices used by members of specific communities to accomplish the events that make up their mundane social worlds permits the study of human language, cognition, social organization, tool use and embodiment from integrated perspective. Within this interactive field the cognitive life of someone such as Chil becomes possible, and the detailed forms of knowledge, social practice, embodied tool use, and distinctive ways of seeing and acting upon the world that constitute a profession such as archaeology emerge as lived social practice. The intrinsic multi-modality of human action, that way that is built by bringing together diverse resources to create a whole that goes beyond any of its parts — for example environmentally coupled gestures that link categories to the world in ways that make possible apprenticeship into the professional vision that sits at the heart of a cognitive life of a society — opens the possibility of analysis of human language, bodies, cognition, and social life from an integrated perspective.

## References

- BAKHTIN, M. 1981. *The dialogic imagination: Four essays*. Austin, University of Texas Press, 444 p.
- GOFFMAN, E. (org.). 1981. Footing. In: E. GOFFMAN, *Forms of talk*. Philadelphia, University of Pennsylvania, p. 124-159.
- GOODWIN, C. 2010. Things and their embodied environments. In: L. MALAFOURIS; C. RENFREW (orgs.), *The cognitive life of things*. Cambridge, McDonald Institute Monographs, p. 103-120.
- GOODWIN, C. 1980. Restarts, pauses, and the achievement of mutual gaze at turn-beginning. *Sociological Inquiry*, **50**(3-4):272-302. <http://dx.doi.org/10.1111/j.1475-682X.1980.tb00023.x>
- GOODWIN, C. 1981. *Conversational organization: Interaction between speakers and hearers*. New York, Academic Press, 195 p.
- GOODWIN, C. 1994. Professional vision. *American Anthropologist*, **96**(3):606-633. <http://dx.doi.org/10.1525/aa.1994.96.3.02a00100>
- GOODWIN, C. 2000. Action and embodiment within situated human interaction. *Journal of Pragmatics*, **32**(10):1489-1522. [http://dx.doi.org/10.1016/S0378-2166\(99\)00096-X](http://dx.doi.org/10.1016/S0378-2166(99)00096-X)
- GOODWIN, C. 2007a. Environmentally coupled gestures. In: S. DUNCAN; J. CASSELL; E. LEVY (orgs.), *Gesture and the dynamic dimension of language*. Amsterdam/Philadelphia, John Benjamins, p. 195-212.
- GOODWIN, C. 2007b. Interactive footing. In: E. HOLT; R. CLIFT (orgs.), *Reporting talk: Reported speech in interaction*. Cambridge, Cambridge University Press, p. 16-46.
- GOODWIN, C. [s.d.] Constructing meaning through prosody in aphasia. In: D. BARTH-WEINGARTEN; E. REBER; M. SELTING (orgs.), *Prosody in interaction*. Amsterdam, John Benjamins [in press].
- GOODWIN, C.; GOODWIN, M.H. 1992. Professional vision. Plenary lecture. In: CONFERENCE ON DISCOURSE IN THE PROFESSIONS, Uppsala, Sweden, 1992.
- HERITAGE, J. 1984. *Garfinkel and ethnomethodology*. Cambridge, Polity Press, 344 p.
- JEFFERSON, G. 1979. A Technique for inviting laughter and its subsequent acceptance/declination. In: G. PSATHAS (ed.), *Everyday Language: Studies in Ethnomethodology*. New York, Irvington Publishers, p. 79-96.
- LINELL, P. 2005. *The written language bias in linguistics: Its nature, origin and transformations*. Oxford, Routledge, 264 p. <http://dx.doi.org/10.4324/9780203342763>
- MCNEILL, D. 1992. *Hand & mind: What gestures reveal about thought*. Chicago, University of Chicago Press, 416 p.
- SACKS, H.; SCHEGLOFF, E.A.; JEFFERSON, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, **50**(4):696-735. <http://dx.doi.org/10.2307/412243>
- SCHEGLOFF, E.A. 1972. Sequencing in conversational openings. In: J.J. GUMPERZ; D. HYMES (orgs.), *Directions in sociolinguistics: The ethnography of communication*. New York, Holt, Rinehart and Winston, p. 346-380.
- SCHEGLOFF, E.A. 1968. Sequencing in conversational openings. *American Anthropologist*, **70**(6):1075-1095. <http://dx.doi.org/10.1525/aa.1968.70.6.02a00030>
- SCHEGLOFF, E.A.; JEFFERSON, G.; SACKS, H. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, **53**(2):361-382. <http://dx.doi.org/10.2307/413107>
- VOLOSINOV, V.N. 1973. *Marxism and the philosophy of language*. Cambridge/London, Harvard University Press, 205 p.
- WILKINSON, R.; BEEKE, S.; MAXIM, J. 2003. Adapting to conversation: On the use of linguistic resources by speakers with fluent aphasia in the construction of turns at talk. In: C. GOODWIN (org.), *Conversation and Brain Damage*. Oxford, Oxford University, p. 59-89.
- WITTGENSTEIN, L. 1958. *Philosophical investigations*. 2<sup>nd</sup> ed., Oxford, Blackwell, 592 p.

Submissão: 02/06/2010

Aceite: 28/07/2010

**Charles Goodwin**  
 UCLA - Applied Linguistics  
 3300 Rolfe Hall  
 90095-1531, Los Angeles, CA, USA