

8

Language presupposes an enchronic infrastructure for social interaction

N. J. ENFIELD AND JACK SIDNELL

8.1 Introduction

While some approaches to language evolution have been thoroughly linguistic yet without becoming particularly social, others have had the opposite problem. Coming out of an ethological tradition of research on primate social systems, Robin Dunbar hypothesized that language arose in our species as a way of managing or servicing social relationships, in a way analogous to physical grooming in apes and monkeys (Dunbar 1993, 1996b). The approach is grounded in research on the maintenance of social relations in complex social groups, but it has failed to gain traction in linguistics because, linguists say, the argument ‘does not say anything about the intricate grammatical structures of human languages’ (Hurford 1999: 182). ‘While language is used for social “grooming” purposes,’ says Hurford, ‘this emphasis fails to account for the impressive and subtle referential power of language’ (1999: 186). The sentiment points to a deep disconnect between the social and the linguistic in current research of relevance to language evolution. Either the researcher is handling the technicalities of language without really grasping what is going on socially, or vice versa. In this chapter, we want to draw attention to a sorely needed solution to this problem, taking as a starting point the behaviour of social interaction from a technical point of view, and seeing what language looks like from there.

8.2 Language in social interaction

Suppose that in order to get at the social origins of language we would begin by looking at language as it is employed in social interaction rather than looking at language as linguists traditionally view it. We might expect that these patterns of social interaction vary radically across cultures without significant constraint. Inuit

Language presupposes an enchronic infrastructure for social interaction 93

song duels, Wolof greetings, Iatmul Naven are just a few of the anthropologically more famous forms that human social interaction takes (Eckert and Newmark 1980; Irvine 1974; Bateson 1936). When seen in comparison with more familiar forms such as found in English courtrooms, American presidential press-conferences, and French family dinners, we may be impressed by apparently limitless diversity. However, research on the basic structures of social interaction, which serve as a chassis for linguistic behaviour, has shown that beneath such diversity is a robust, universal, generic infrastructure that exploits a range of species-specific cognitive abilities and prosocial motivations (Enfield and Levinson 2006; Levinson 2006; Schegloff 2006; Sidnell 2007; Tomasello 2008). This infrastructure is not some kind of conversational equivalent of the mental module or device that has been proposed to account for putative underlying commonalities among the grammars of the world's languages. The infrastructure being described here emerges from a combination of evolved cognitive capacities in the domain of sociality (in part specifically human, though not specific to language) and the structural patterning that emerges when social moves are made and counter-made, in the form of the kinds of few-second chunks of behaviour we call utterances (see Enfield 2013: chs 3 and 6). The precise nature of this emergence is a major topic for research. The hypothesis is that there is a universal underlying infrastructure for the use of language in interaction, even when basic styles of human interaction seem to differ radically across groups—compare, for example, norms of interaction in Japan versus Anglo America (Lebra 1976; Lebra and Lebra 1986; Wierzbicka 1991: 72ff). The technical properties of this infrastructure cannot be discovered by applying existing research tools of linguistics.

If research on the social origins of language is to be properly informed about the object of study, then what is needed, as a complement to approaches to language within the discipline of linguistics, are approaches that focus on (a) those aspects of language that are unique to multi-unit sequences produced not by a single person but by multiple people in a social interaction, i.e. conversation; and (b) those aspects of cognition-for-language that are unique to the management of social relations rather than to the conceptual management of information per se. Fortunately, these needs are already met. There are established research traditions that provide resources for studying language and cognition in just these ways. To date, however, they have tended not to be well connected with linguistics, or cognitive science more generally, partly for reasons of disciplinary affiliation: they exist as branches of anthropology and sociology. They take an enchronic perspective on language, which is to say that they look at language in the move-by-move flow of interaction, as opposed to the (also invaluable) perspectives of diachronic versus synchronic, and ontogenetic versus phylogenetic (on enchrony, see Enfield 2009a: 10, 2011: 285–91, and 2013). To review these lines of research at any length would go beyond our present scope (see e.g. Heritage and Atkinson 1984; Duranti and Goodwin 1992; Duranti 1997; Schegloff

2006, 2007; Sidnell 2010; Sidnell and Enfield 2012; Sidnell and Stivers 2012; Enfield et al. 2014). Here we merely want to point in their direction.

Current thinking about face-to-face interaction from an enchronic perspective has been influenced by a variety of lines of research including linguistic pragmatics, Peircean semiotics, as well as research in anthropology, psychology, and other disciplines (see Enfield 2013, and many references therein). Here we want to highlight an approach to interaction that emerged in the work of Harvey Sacks, Emanuel Schegloff, and Gail Jefferson and that has come to be known as conversation analysis (Sacks 1992; Sacks et al. 1974; Sidnell 2010; Sidnell and Stivers 2012). Both Sacks and Schegloff were students of Erving Goffman, a transdisciplinary scholar who, although trained as a sociologist, had a major impact upon, and indeed was himself strongly inspired by, anthropology (see especially Goffman 1964, 1967, and 1971). Goffman was perhaps the first, and certainly the most eloquent, defender of the view that face-to-face interaction constituted its own phenomenon, that it had properties which were *sui generis* and not reducible to individual psychology or broader social processes (Goffman 1964). Sacks, Schegloff, and Jefferson incorporated this idea, and it may be understood as the first pillar of conversation analysis (see especially Schegloff 1968).

When Sacks and Schegloff were studying with Goffman at Berkeley, they were influenced by the highly original studies of Harold Garfinkel and the approach he developed known as ethnomethodology (Garfinkel 1967). The goal of Garfinkel's early studies was to uncover the underlying practices of reasoning which members of a society use in accomplishing everyday activities and which ultimately make society possible. A major part of Garfinkel's investigations was taken up with the question of how a person makes sense of another's conduct including their talk. This concern was incorporated into conversation analysis, as a kind of second pillar: the idea that participants in social interaction engage in practical reasoning both to produce their own talk and to understand the talk of others (see Heritage 1984). Both Goffman and Garfinkel thus provided inspiration for a new and distinctive approach to the study of ordinary social interaction. Others were left with the task of inventing a method by which ordinary social interaction might be systematically studied.

Sacks, Schegloff, and Jefferson took up this challenge. They studied social interaction by listening to audio recordings of telephone calls as well as co-present interaction, and there they found a locus of intricate order. Early studies showed that an interaction can be analysed into parts and that these parts consist of definable practices of speaking which have systematic effects and which together form orderly sequences of action in interaction (Schegloff 1968; see also Schegloff 2006). This order is not the product of statistical regularities or of categorical imperatives but rather of a persistent and pervasive orientation by the participants to a set of structures or norms. Like any set of norms in this sense, the norms that organize social interaction do not determine conduct but rather provide a framework through which it is

intelligible and through which it is evaluated (see Heritage 1984; Stivers et al. 2011). Membership in any particular social group (e.g. German, Surgeon, Skateboarder) requires that one be aware of, and unable to plausibly deny the existence of, a certain set of norms, and thus to be accountable to those norms at every step. Participants in interaction can then be seen—not only by analysts but, in the first place, by other participants—as following a rule, deviating from it, attempting but failing to follow it, or simply violating it flat out. These alternatives generate further informative inferences about what that participant intends or means by behaving in that way (Sidnell 2010). The orderliness of interaction then is an endogenous product that is achieved by participants in interaction in each and every one of its local instantiations through the application of regular practices of reasoning.

Since the establishment of conversation analysis as a field of empirically based research on the structures of social action, a significant literature has been produced. The recent *Handbook of Conversation Analysis* (Sidnell and Stivers 2012) has a list of some 2,000 bibliographical references. But from the point of view of linguistics and other branches of cognitive science it is as if there were a hermetic seal around this literature. In even the most sympathetic work in linguistics, there is little if any connection to what is empirically known about the structure of human interaction, nor to the social cognition that underlies it nor the situated use of language as a cooperative activity. There are a few notable exceptions, for example in psycholinguistics (e.g. Clark 1996) and linguistic anthropology (e.g. Moerman 1988; Sidnell 2005), but surprisingly, in the language evolution literature that begins with the social-relational functions of language (e.g. Dunbar 1996b; Dessalles 2009) this large and directly relevant field of research is overlooked.

8.3 Linguistic structures in social interaction

When we look at language in the context of its functions and distributions in social interaction, the phenomenon takes on a new light. Over the last 20 years or so, a group of linguists with functionalist affiliations have begun to yield new insights by incorporating analytic concepts developed in conversation analysis, particularly those of turn-taking and conversational sequence, in the analysis of traditional linguistic topics. For example, Ford (1993) examines the grammatical structures known as sentence-final adverbial clauses in English and shows ways in which their structure and distribution are not arbitrary but may be ‘interactionally generated’ (Ford 1993: 108 and *passim*). In the following example (1) the speaker S poses a question, and when no uptake is forthcoming from the recipient R in the subsequent line, S then adds the increment (in line 3), which in turn elicits the uptake from R (in line 4) that lets S know they are being understood.

- (1) 1 S: Y'know when it- (.) came from the:: I think (a) air conditioning system, it
drips on the front of the ca:rs?
2 (0.1 second pause)
3 S: if you park in a certain place?
4 R: mm-hm

The increment produced in line 3 is grammatically fitted to what came before it in line 1, but clearly lines 1 and 3 were not produced as a whole unit, such as they might appear to have been if cleaned up and written down on the page (*... it drips on the front of the cars if you park in a certain place?*). Rather, we see that in the real-time context of the dyadic interaction, at line 2 some sort of confirmation of understanding appears to have been due from R, but there was silence instead. By then producing the increment in line 3, S can continue as if the turn was not yet finished, thus getting another opportunity to elicit a confirmation of understanding, which indeed comes from R in line 4 (see also Sidnell 2012). This type of study complements linguistic research on grammatical structures by giving us a sense of the functions that such structures can have in the enchronic social context of the speech event, as distinct from their referential or representational relations to the event being narrated or described.

Another area is the study of the 'procedural resources' that languages provide for managing talk (see Blakemore 1987, 2002; Schiffrin 1988; Clark 1996). All languages have various kinds of 'feedback markers' (*mm, uh-huh*), 'discourse markers' (*oh, so, well, etc.*), 'trouble markers' (*um, uh*), 'editing expressions' (*I mean etc.*), 'repair initiators' (*what? huh?*), etc. Despite their ubiquity and importance in regulating the flow of interaction, these kinds of items are seldom studied in typological or cognitive scientific approaches to language, where the tendency is to focus on the referential functions of language. Upon inspection of data from conversation, it can be shown that these types of items have distinct distributional properties in relation to formal patterns of language use. For example, with regard to so-called 'newsmarks' in English (such as *Really?*), Jefferson (cited in Heritage 1984: 340) shows that there are definable and recurrent multi-turn sequences like the following: (Move 1) announcement of some piece of new information by Speaker A (e.g. that she didn't smoke any cigarettes); (Move 2) *Oh really?* by Speaker B; (Move 3) reconfirmation by A; and finally (Move 4) an evaluation or 'assessment' by B (e.g. *Very good*). This kind of pattern is not a construction in any standard linguistic sense, but it clearly points to a kind of interactional grammar associated with these procedural linguistic elements.

In some research, procedural items are not considered to be linguistic at all. Levelt (1989: 484) looks at *er* as a marker of disfluency (similar to *um, uh, etc.*), and while he shows that *er* has a specifiable function—signalling that 'at the moment when trouble is detected, the source of trouble is still actual or quite recent'—he concludes that it is 'a symptom, not a sign'. By contrast, Jefferson (1974: 184) suggests that *uh* perhaps has 'the status of a word in the English language'. The same has been argued by Clark and

Fox Tree (2002), who found in a corpus study that *um* and *uh* show different functional distributions in English. In addition, a quick look across languages reveals that the precise form of such items is locally conventionalized: English has *um* while Lao has ‘*un*’ (Enfield 2007: 314). These kinds of procedural items do not have the referential functions that linguists tend to privilege, but they are no less linguistic for that.

In the sorts of approaches we have just reviewed, the researcher is either looking at a familiar grammatical structure in the unfamiliar light of conversational sequence, or is looking at a well-known but oft-marginalized element that can hardly be studied at all without consulting conversational data. Then there are domains of structure that are off the linguistic map altogether. These possibly universal and arguably generic underlying components of the infrastructure for interaction, already alluded to above, are organized into partially independent or semi-autonomous domains or systems. We now want to draw attention to two of the most central of these systems, which will need to be properly defined and handled by any natural approach to language as a form of social behaviour.

8.4 Turn-taking and sequence organization in conversation

An organized system of *turn-taking* provides for the orderly distribution of opportunities to participate in talk-in-interaction. Sacks et al. (1974) define the system as having two components. A turn constructional component defines the units out of which a possible turn can be constructed and by extension allow participants in interaction to anticipate the possible/probable extent and shape of any actual unit and thus to project or predict its completion. A turn allocation component specifies an organized set of practices by which transition from a current speaker to a next speaker is managed. Together, these two components and the rules that organize the relation between them, provide for the orderly or systematic nature of turn-taking in interaction. Sometimes the system operates in such a way that we see seamless transitions between speakers in a sequence. The following example from a telephone call shows tight temporal alignment in ‘floor transition’ between the two speakers:

- (2) (Rahman corpus A:1:VM:(4); transcription slightly simplified; transition measurements in secs from de Ruiter et al. 2006)

(Begin call)

Mat: ‘lo Redcah five o’six one?,

Ver: [+0.15s] Hello Matthew is yer mum there love.

Mat: [+0.13s] Uh no she’s, gone (up) t’ town

Ver: [+0.24s] Al:right uh will yih tell’er Antie Vera rang then.

Mat: [-0.03s] Yeh.

Ver: [+0.13s] Okay. She’s alright is she.

Mat: [+0.10s] Yeh
Ver: [+0.07s] Okay. Right. Bye bye luv,
Mat: [+0.02s] Tara,
(End call)

The turn-taking model suggests that people in a conversation will actively minimize the amount of overlap in speech (i.e. they will avoid having two or more people speaking simultaneously), and at the same time they will actively minimize the amount of silence between turns in a sequence, as we have just seen. But having such imperatives or ‘rules’ does not entail that all conversation actually proceeds one speaker at a time. Overlaps and gaps occur constantly (Sacks et al. 1974; see also Schegloff 2000: 47–8, n. 1), and these exceptions can have functional effects, arising from the very fact that people perceive them as exceptions. It is sometimes suggested that in this or that culture or social setting, a completely different system applies, or that there is no system at all (for instance, ‘In Language/Culture X, people all talk at the same time’; see Stivers et al. 2009: 10587). However, there are to date no systematic empirical studies of informal conversation that provide counterexamples to the claim of a one-speaker-at-a-time normative design for the regulation of conversational turn-taking (Schegloff 2000: 2). As in many domains of linguistic analysis, impressions and intuitions turn out not to be supported by empirical data (see de Ruiter et al. 2006; Stivers et al. 2009). Sidnell (2001) tested Reisman’s (1974) claim that everyday conversation in Antigua follows a ‘contrapuntal’ style, and found that in fact the data are compatible with a one-speaker-at-a-time model.

It is common to observe overlapping speech in conversation, but rather than constituting exceptions to the turn-taking rules, such cases usually provide evidence in support of it. When people self-select as speakers of next turns, the system dictates that they should begin speaking at a point where the previous speaker’s turn is possibly complete and where transition to a new speaker is pragmatically relevant. In example (3) below, left square brackets mark the onset of overlapping talk. Our focus is on what happens at lines 4 and 5. After Old Man says ‘The funfair changed it’, this is a point of possible completion of his utterance, and it is a point where transition to a new speaker would be relevant. Parky twice attempts to begin his turn ‘That changed it’ (before it is eventually produced in full at line 6), yet finds himself in overlap on both occasions (line 5). Notice the split-second timing here with Parky attempting to come in at just those points where Old Man has reached possible (though obviously not actual) completion of his current turn. Clearly, in order to come in at just these points, Parky must have anticipated where Old Man would reach possible completion of his current turn.

- (3) Parky
1 Tourist: Has the park changed much,
2 Parky: Oh yes,

Language presupposes an enchronic infrastructure for social interaction 99

- 3 (1.0)
4 Old man: Th' *Funfair* changed it'n [awful lot [didn'tit.
5 Parky: [Th- [That-
6 Parky: That changed it,
(example from Sacks et al. 1974; transcription slightly simplified)

An important and widely underappreciated point is that this turn-taking system operates independently of whatever social actions are being accomplished in and through the talk it organizes—that is, whether people are requesting, inviting, questioning, answering, agreeing, disagreeing, complaining, excusing, insulting, or whatever else they do in turns-at-talk constructed and distributed through the turn-taking system. All of this supports the idea that the turn-taking system is part of an infrastructure that operates ‘underneath’, and independent from, the goal-directed social behaviour that people are effecting with their context-situated usage of language.

The arrangement of linguistically conducted social actions into *sequences* represents a distinct domain of organization in interaction, yet it presupposes an underlying turn-taking mechanism (Schegloff 1968, 2007). Many social actions that are carried out through the use of language come in pairs, for example request and granting (or rejection), invitation and acceptance (or refusal), complaint and excuse (or denial), and so on. These pairs are linked together by a relation of conditional relevance whereby, to paraphrase Schegloff, given a first action (such as a request, invitation, or complaint), a second action is made expectable. Upon the occurrence of a second it can be seen to be a second item to the first (rather than an independent turn) and upon its non-occurrence it can be seen to be absent (where an infinite number of other things did not occur but are not absent in the same way; Schegloff 1968). Conditional relevance thus establishes a relation between a first and second action that has both a prospective and a retrospective dimension. The prospective dimension ensures that the doing of a first action will activate a norm making the doing of the second action relevant and noticeably absent if not produced. This norm draws on a cooperative assumption in social interaction. The retrospective dimension allows the speaker of the first to see if, and how, she was understood. For example, if someone produces a responsive utterance that is recognizable as an *excuse*, this will reveal to the first speaker that she was apparently heard to be complaining or accusing, whether that was her intention or not. Thus the production of actions within sequences constitutes an architecture of intersubjectivity by which understandings are publically displayed and ratified incarnately, *en passant* in the course of whatever business the talk is occupied with (Heritage 1984; Clark 1996).

Episodes of talk-in-interaction can, typically, be described in terms of base sequences (often adjacency pairs) and their frequently multiple pre-, insert-, and post-expansions (Schegloff 2007). Such expansions are also made up of sequences and these may be the loci of their own expansions. Thus a maximally simple ordering

of utterances into adjacency pairs can nevertheless result in sequences of considerable complexity, again implying a kind of structured grammar of interaction, of clear importance to the social origins of language and yet unknown to the science of linguistics.

If we take the elements of the turn-taking system and the action sequences that ride upon it and incorporate them into our understanding of language structure, this can give us new insights into the function and distribution of the kinds of lexico-grammatical structures that linguists normally want to explain. One of the central features of conversational sequences, as just described, is the notion that utterances are not just produced as conversational moves in a flat string, but rather are related to each other in specific ways. Most important among such inter-move relations is the relation between *initiating* utterances and *responsive* utterances. A simple example of an initiating utterance is a question: it initiates one person's local project (e.g. it is a means to achieve a person's goal to get a piece of information that they need prior to some next action, such as when I ask you the time so I can know whether I need to run for the train or not), and obliges another person to join this local project and, if possible, cooperate by providing the required information (see example (3), Q = 'Has the park changed much,' A = 'Oh yes,'). Many other kinds of two-turn sequences have this kind of structure, where the first move puts an obligation on another person to provide a next move of a certain type (see discussion above).

Linguists seldom invoke turn or sequence structure in carrying out core linguistic research, but recent attempts to do just this have yielded good results. Gipper (2011) examines a set of grammatical markers of evidentiality in Yurakaré (a language isolate of Central Bolivia), of the kind that have proven notoriously difficult to pin down in semantic and pragmatic analysis. Gipper draws on data from Yurakaré conversation, which allows her to use sequential position as a factor in the analysis. She distinguishes cases in which a certain evidential is used in an initiating type of turn in a conversational sequence from cases in which it is used in a responding type of turn. She finds that the different effects or meanings of evidential markers can be defined in terms of the conversational position (and, relatedly, the action type) they occur in.

Earlier research used these pair-part asymmetries of conversational sequence as a way to account for the distribution of certain kinds of lexical and grammatical structures, while bringing in more explicit reference to the social elements of speakers' rights and duties in producing different kinds of utterances. Pomerantz (1984) discovered some ways in which an utterance can set up a *preference* structure that constrains the ways in which another person can respond (Pomerantz and Heritage 2012). For example, when a person issues an invitation, the linguistic structure of the response will be different depending on whether it is an acceptance or a rejection. In the case of accepting, which we can regard as the socially preferred, cooperative response to an invitation, the response normally comes without delay

and in simple form, as shown in example (4a). By contrast, in the case of declining, which we can regard as the relatively non-cooperative, socially marked response to an invitation, the response has a set of formal properties that are not observed in the acceptance, namely delay, prefacing with ‘discourse markers’ like *well*, markers of disfluency or hesitation, and the provision of accounts or reasons for the declination, as shown in example (4b):

- (4) (a) i. Do you want to go for a drink?
ii. Sounds good!
- (b) i. Do you want to go for a drink?
ii. (pause) Um, well, I kinda still have work to do, so maybe um. . . .

A series of studies have subsequently examined these issues of preference, where the social-interactive factors at play at a given moment in the interaction can account for why a particular grammatical structure has been selected at all, accounting thus for both the specific function and distribution of a grammatical device which might otherwise go unexplained (Pomerantz 1984; Pomerantz and Heritage 2012; Heritage and Raymond 2005; Sidnell and Enfield 2012, among many others). These kinds of features of preference structure have two important properties that put them outside the scope of most linguistic research. Firstly, they are explained in terms of inherently social-interactive factors such as the degree to which an utterance constitutes a cooperative action, as opposed to an action that resists the trajectory that another person has set out on. Secondly, they are inherently enchronic, being defined in terms of specific positions in conversational sequence—initiating versus responsive—which cannot be studied without looking at data from conversation. Conversation is a type of data that is seldom studied in linguistics, including the kinds of linguistics that have currency in research on language evolution.

8.5 Repair

A second system of practices for the use of language in interaction is the system for *repair* (Schegloff et al. 1977). In using language in human interaction, there is always the possibility that troubles will arise in speaking, hearing, and understanding (Levelt 1983; Clark 1996; Schegloff 2006; Hayashi et al. 2013). An organized set of practices of repair constitutes a natural, interactive system by which such troubles may be addressed at or near the point where they occur, and may be potentially resolved more or less immediately before there is intolerable divergence of the participants’ intersubjective understandings of what is going on in the interaction. Research on the practices of repair have shown it to be a structured and systematic domain, as if repair had a grammar all its own; though again, this system has remained outside of the usual purview of linguistics.

The analysis of repair in interaction makes reference to several kinds of distinguishing structural features. Firstly, there are distinctions in the *components* of the repair sequence, listed in (5):

- (5) Components of a repair sequence
- A. **Trouble source or repairable** = some element of an utterance that constitutes a source of trouble, for any of a range of reasons including problems of production, word choice, hearing, and appropriateness, relevance, among other possible problems.
 - B. **Signal of trouble** = a sign that there is a source of trouble, including the manifest nature of the trouble itself as well as techniques for drawing attention to the trouble.
 - C. **Initiation of repair** = a sign that the trouble will be, or should be, fixed; often is formulated in such a way as to identify and/or characterize the trouble source.
 - D. **Repair** = a correction or redoing of the trouble in (A).

Secondly, there are distinctions in the *personnel* involved in the repair sequence. The speaker of the trouble source turn is often referred to as 'self' (e.g. in the term *self-repair*), though this person may or may not be the one to produce some other component of the sequence; for instance, a major class of repair sequences is called *other-initiation of repair*. This refers to cases in which a problem in one person's utterance is signalled by another person in a subsequent turn, and is subsequently repaired by the original speaker, as shown in the following example:

- (6) Other-initiation of repair in a 3-turn sequence
- | | |
|-----------------------------|--|
| <i>Trouble source</i> | B: Oh Sibbie's <u>sistuh</u> had a <u>baby boy</u> |
| <i>Initiation of repair</i> | A: Who? |
| <i>Repair</i> | B: Sibbie's sister. |
- (example from Schegloff et al. 1977: 376–8; transcription slightly simplified)

Thirdly, there are distinctions in the possible *positions* in a 'repair opportunity space' at which the different components can be observed, and where different personnel can be involved:

- (7) Positions in the repair opportunity space: '3 turns long', 4 main positions
- A. **Same turn** = within the turn that includes the 'trouble source' (T₁)
 - B. **Transition space** = at completion of T₁, when turn transition is relevant/possible
 - C. **Next turn** = in the turn that follows T₁ (T₂)
 - D. **Third turn/position** = in the turn that follows next turn (with complexities left aside here)

Language presupposes an enchronic infrastructure for social interaction 103

Together, these three sets of distinctions—the components, personnel, and positions of a repair sequence—define a possibility space for how any repair sequence might emerge or be generated in a given stretch of social interaction. A solid literature has begun to make inroads into mapping out the many possible structures and functions of repair sequences, mostly with a focus on English. Much more work needs to be done, not only on continuing to describe the grammar of repair and determine the extent to which it constitutes a system, but also to carry out the necessary cross-linguistic comparative work (see Hayashi et al. 2013, and references therein, for steps in this direction).

8.6 Conclusion

We have introduced a couple of the most important contributions of pioneering research on talk-in-interaction over the last 30 years or so. A key feature of this line of work is that it has effectively taken an ethological approach to the use of language in human behaviour, but has not suffered from one of the central problems of human ethology, namely the problem of proceeding as if our possession of language doesn't make enough of a difference that we should study people any differently from how we study animals. We are of course animals, and we are not excused from being studied as such (Tiger and Fox 1966). But let's face it, language changes things a lot. We have argued that to get a direct view of how language works in social terms, an enchronic perspective is required, a causal frame in which the moves we build from bits of language (and much else) are embedded in trajectories of joint activity that necessarily involve multiple parties and that necessarily have social causes and consequences. Technically defined systems for interaction such as turn-taking and repair seem to transcend language, and yet they are so closely bound up with it that we might ultimately expect some kind of a co-evolutionary account. That said, to have been able to get language up and running as we know it today, it seems to us that the fundamentals of the infrastructure for interaction would have to have been in place first. This implies, as many have suggested (e.g. Tomasello 2006), that if our closest relatives lack certain key capacities for the shared intentionality that enables the most basic sequences of human interaction, this would account for why they also don't have language, and can't get it. It seems clear that in the realm of vocal communication (see Clay and Zuberbühler, Chapter 11), other apes do not show anything like the responsive, contingent turn-taking behaviour so characteristic of human interaction (Arcadi 2000); though we note that in the realm of visible bodily behaviours such as ritualized gestures, there are patterns of behaviour that do more closely resemble the kernels of adjacency-pair sequences (Rossano, F. 2012). So what's now needed is more empirical research. The first pass that conversation analysis has carried out over recent decades was an important initial step toward developing a rigorous account of the enchronic infrastructure for interaction. By

determining the systems or domains out of which talk-in-interaction is composed, and upon which language usage rides, this work has uncovered some fundamental areas upon which entire research programmes can now be based.

Subsequent research will have to explore implications for language evolution of the idea that language *presupposes* an enchronic infrastructure for social interaction. If language is not possible without such an infrastructure, then the question of language evolution turns to the evolutionary origins of the infrastructure itself. An important issue for subsequent work will be to test the extent of human diversity in the basic structures of interaction. Recent work from an anthropological and cross-linguistic perspective has begun to ask whether the particular language being spoken has consequences for the organization of interaction as described here (see Sidnell 2009; Dingemanse and Floyd 2014). That research is in its infancy but initial results suggest that the underlying, generic structures of interaction may be inflected or torqued by the particular semiotic structures through which interaction is accomplished, as well as the local circumstances within which it operates. Sidnell and Enfield (2012) explore the idea that the distinct lexico-syntactic resources of a language can have distinct *collateral effects* on the ways in which certain types of social action can play out in sequences of conversation.

Our goal in this chapter has been to draw attention to some aspects of language that inhabit the divide between the established reference-oriented, sentence-based interests of linguistics, on the one hand, and the social-relational concerns of research on human sociality, on the other. Linguistics has achieved an enormous amount, yet it still struggles to connect with social cognition and social behaviour in ways that other disciplines can apprehend and apply. We submit that this problem can be readily solved in research on the social origins of language, thanks to the existence of a significant body of literature that provides the much-needed link between language and human sociality.

8.7 Acknowledgements

We are extremely grateful to Daniel Dor, Chris Knight, Maggie Tallerman, and Kathleen Gibson for comments on drafts of this chapter. Some sections draw on parts of the chapter ‘Face-to-Face Interaction’ by Jack Sidnell (2012, in *Encyclopedia of Theory in Social and Cultural Anthropology*; Sage Publications) and a section from p. 516 of de Ruiter, Mitterer, and Enfield (2006). We thank members of the audience at Social Origins of Language conference at UCL (London, February 2011) and the Linguistic Society of America Institute Workshop ‘Interactional Foundations of Language’ (Boulder CO, 18 July 2011) for their input. This work is supported by European Research Council (ERC Project ‘Human Sociality and Systems of Language Use’, 2010–2014), and the Max Planck Institute for Psycholinguistics, Nijmegen.