John J. Gumperz

## SOCIOCULTURAL KNOWLEDGE IN CONVERSATIONAL INFERENCE[1]

'CONVERSATIONAL INFERENCE', as I use the term, is the 'situated' or context-bound process of interpretation, by means of which participants in a conversation assess others' intentions, and on which they base their responses. Conversational inference is ultimately a semantic process, but it is distinguished from linguists' assignment of meaning to utterances or classification of speech acts, as well as from the social scientists' measurement of attitudes. Both conventional linguistic analysis and social science measurement involve the labeling of utterances by other utterances, more often than not after the fact. Conversational inference, by contrast, is, part of the very act of conversing. One indirectly or implicitly illustrates one's understanding of what is said through verbal and nonverbal responses, by the way one builds on what one hears to participate in a conversation, rather than through talking about it in abstract terms. It follows that analysis of such processes requires different and perhaps more indirect methods of study which examine meaning as a function of the dynamic pattern of utterances and responses as they occur in conversation.

Recent studies of conversation from a variety of linguistic, psychological, anthropological, and sociological perspectives, have shed light upon a number of issues important to the study of conversational inference. It is generally agreed that grammatical knowledge is only one of several factors in the interpretation process. Aside from physical setting, participants' personal background knowledge, and their attitudes toward each other, sociocultural

Edited by
Adam Jaworski
and
Nikolas Coupland

The Discourse Reader
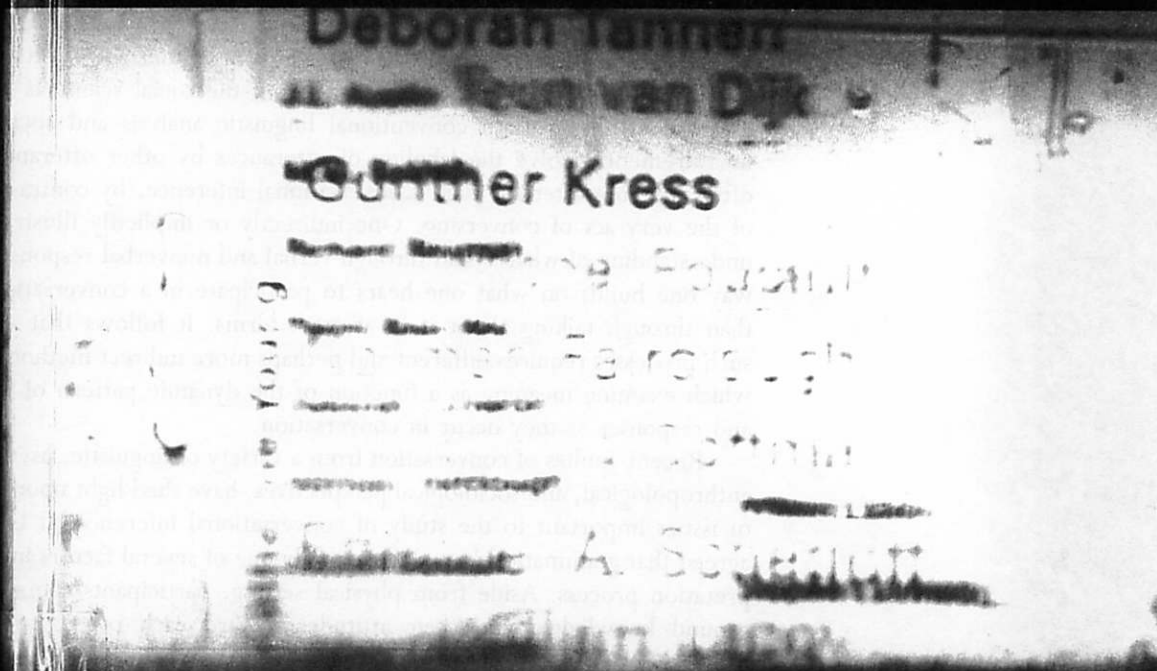
assumptions concerning role and status relationships, as well as social values associated with various message components, also play an important role. So far, however, treatment of such contextual factors has been primarily descriptive. The procedure has been to identify or list what can potentially affect interpretation. With rare exceptions, there have been no systematic attempts to show how social knowledge is used in situated interpretation. Yet we know that social presuppositions and attitudes change in the course of interaction, often without a change in extralinguistic context. Therefore, the social input to conversation is not entirely constant. Assumptions about role and status relationships vary as the conversation progresses, and these changes are signalled through speech itself (Gumperz and Cook-Gumperz 1976). The signals by which this is accomplised can be regarded as a metalanguage or a meta-signalling system. So far, however, we know very little about this metalanguage. In this paper I want to suggest at least the outlines of a theory which deals with the question of how social knowledge is stored in the mind, how it is retrieved from memory, and how it is integrated with grammatical knowledge in the act of conversing.

[. . .]

Ethnomethodologists have gone a long way toward producing a theory which treats conversation as a cooperative endeavor, subject to systematic constraints. However, a number of important questions still remain to be answered. A social view of language such as the one ethnomethodologists advocate must be able to account for interspeaker difference, yet, so far, only the pan-cultural aspects of conversational control mechanisms have been dealt with. A sociolinguist needs to know how speakers use verbal skills to create contextual conditions that reflect particular culturally realistic scenes. Furthermore, how is speakers' grammatical and phonological knowledge employed in carrying out these strategies? For example, if regular speaker change is to take place, participants must be able to scan phrases to predict when an utterance is about to end. They must be able to distinguish between rhetorical pauses and turn-relinquishing pauses. Although speaker overlap is an integral part of interaction, conversational cooperation requires that speakers not be interrupted at random. To follow the thematic progression of an argument, moreover, and to make one's contribution relevant, one must be able to recognize culturally possible lines of reasoning. To account for all these phenomena, it is necessary to show how the ethnomethodologists' control mechanisms are integrated into other aspects of speakers' linguistic knowledge.

To this end, we will look at two examples of actual conversation. [. . .] We will analyze two sequences which occurred in public situations.

They are representative of a much larger body of data we have collected, both by chance, as in these examples, and in connection with systematic programs. The first interaction is one which any native speaker of English would be able to interpret. The second constitutes an interethnic encounter, and we will show how habitual conversational inferences led to a misinterpretation of intent.

The first incident occurred when I was sitting in an aisle seat on an airplane bound for Miami, Florida. I noticed two middle-aged women walking towards the rear of the plane. Suddenly, I heard from behind, 'Tickets, please! Tickets, please!' At first I was startled and began to wonder why someone would be asking for tickets so long after the start of the flight. Then one of the women smiled toward the other and said, 'I TOLD you to leave him at home'. I looked up and saw a man passing the two women, saying, 'STEP to the rear of the bus, please'.

Americans will have no difficulty identifying this interchange as a joke, and hypothesizing that the three individuals concerned were probably traveling together and were perhaps tourists setting off on a pleasure trip. What we want to investigate is what linguistic knowledge forms the basis for such inferences, and to what extent this knowledge is culturally specific.

The initial utterance, 'Tickets, please', was repeated without pause and was spoken in higher than normal pitch, more than usual loudness, and staccato rhythm. For this reason it sounded like an announcement, or like a stock phrase associated with travel situations. My first inkling that what I heard was a joke came with the woman's statement to her friend, 'I TOLD you to leave him at home'. Although I had no way of knowing if the participants were looking at each other, the fact that the woman's statement was perfectly timed to follow the man's utterance was a cue that she was responding to him, even though her comment was addressed to a third party. Furthermore, the stress on *told* functioned to mark her statement as another stock utterance, contributing to the hypothesis that she and he were engaging in a similar activity. If the statement of the man or the woman had been uttered in normal pitch and conversational intonation, the connection between them might not have been clear. Only after I was able to hypothesize that the participants were joking, could I interpret their utterances. My hypothesis was then confirmed by the man's next statement, 'Step to the rear of the bus, please'. This was also uttered in announcement pitch, loudness, and intonation. In retrospect, we may note that both of the man's utterances were formulaic in nature, and thus culturally specific and context bound. He was exploiting the association between walking down an aisle in a plane and the similar walk performed by a conductor on a train or a bus. In identifying the interaction as a joke, I was drawing

collected,
systematic
of English
encounter,
. misinter-

seat on an
en walking
, 'Tickets,
onder why
ight. Then
ou to leave
en, saying,

as a joke,
bably trav-
trip. What
sis for such
fic.

: pause and
dness, and
ent, or like
s that what
friend, 'I
knowing if
e woman's
was a cue
s addressed
o mark her
esis that she
the man or
intonation,
after I was
erpret their
t statement,
nouncement
hat both of
ally specific
en walking
a conductor
was drawing

on the same situational-association knowledge, as well as on my awareness of the likelihood of joking among travelers bound for Miami.

Thus, suprasegmental and other surface features of speech are crucial to understanding the nature of an interaction. Such features have been extensively discussed in the linguistic literature, but treatments have dealt with the referential meaning of individual sentences. When seen in isolation, sentences can have many intonation and paralinguistic contours, without change in referential meaning. The prevalent view is that these features add expressive overtones to sentences. Moreover, the signs by which listeners recognize these overtones tend to be seen as language independent. If, however, we look at conversational inference rather than referential meaning, we see that paralinguistic and intonation contours play an important role in the identification of interpretative frames.

This identification of specific conversational exchanges as representative of socioculturally familiar activities is the crucial process I call 'contextualization'. It is the process by which we evaluate message meaning and sequencing patterns in relation to aspects of the surface structure of the message, called 'contextualization cues'. The linguistic basis for this matching procedure resides in 'co-occurrence expectations', which are learned in the course of previous interactive experience and form part of our habitual and instinctive linguistic knowledge. Co-occurrence expectations enable us to associate styles of speaking with contextual presuppositions. We regularly rely upon these matching processes in everyday conversation, but they are rarely talked about. In fact, they tend to be noticed only when things go wrong, and even then, the conclusions drawn are more likely to be about the other person's attitudes than about differences in linguistic conventions. Yet, as our next example shows, contextualization expectations are highly culturally specific; that is, they are dependent upon interactants' ethnic or communicative background.

The second incident I am going to relate took place in London, England, on a bus driven by a West Indian driver–conductor. The bus was standing at a stop, and passengers were filing in. The driver announced periodically, 'Exact change, please', as London bus drivers often do. When passengers who had been standing close by either did not have money ready or tried to give him a large bill, the driver repeated, 'Exact change, please'. The second time around, he said 'please' with extra loudness, high pitch, and falling intonation, and he seemed to pause before 'please'. One passenger so addressed, as well as others following him, walked down the bus aisle exchanging angry looks and obviously annoyed, muttering, 'Why do these people have to be so rude and threatening about it?' Was the bus driver really annoyed? Did he intend to be rude, or is the passengers' interpretation a case of cross-cultural misunderstanding?

To understand what happened here and why it happened, it is necessary to go into some more detail about the nature of contextualization cues and their function in conversation. The term 'contextualization cue' refers to any aspect of the surface form of utterances which, when mapped onto message content, can be shown to be functional in the signalling of interpretative frames. In the examples given in this paper, the cues are largely prosodic and paralinguistic, but many other signalling mechanisms can function as contextualization cues, including lexical or phonological choice; use of idiomatic or formulaic expressions such as greetings, openers, interjections, or frozen sequences; or code-switching (Gumperz 1976; Gumperz and Cook-Gumperz 1976). In the present discussion, however, we concentrate on prosody (i.e., intonation and stress) and paralinguistics (pitch register, rhythm, loudness, etc.) since some aspects of these features are always involved in conversation.

[. . .]

Prosody consists of three basic signalling mechanisms: tone grouping; tonic or nucleus placement within a tone group; and tune, the direction of the tonal change which characterizes the nucleus. Paralinguistic features include, among others, pitch register, loudness, rhythm, and tempo, and apply to the tone group as a whole, rather than to parts thereof.

Among prosodic cues, tone grouping refers to the use of intonation and stress to chunk larger stretches of speech into separable bits of information that are to be processed as single units. Our example:

Exact change please //

could be uttered as a single chunk, as it was the first time the driver said it, or as two chunks:

Exact chạnge / pleäse //

as he said it the second time. To treat *please* as a separate bit of information implies that it is worthy of separate attention. We recognize two types of tone group boundaries: minor tone group, (/) which suggests that the preceding message portion is semantically related to others within a larger whole, and a major tone group, (//) which suggests finality.

The second element of prosody, tonic or nucleus placement, refers to the selection of one or another of the stressed syllables in a tone of group as the nucleus, or the part on which the tonal shift occurs. Nucleus placement is predictable in many types of sentences. Normally, it identifies that

portion of the message that is to be regarded as new, as compared to what can be assumed to be shared or given. Note, however, that this is not merely a matter of syntax or lexicon but also a matter of culturally specific practice. If I say

> I'm giving my paper //

it is the object, *paper*, which is assumed to carry the new information. In

> I'm cancelling my paper //

the verb is normally stressed, since *cancelling*, in our culture, is not a customary activity in relation to paper giving.

The third prosodic mechanism, tune, refers to the fall and rises in tone such as are associated with the intonational contrast between questions and answers. We furthermore distinguish two levels on which the fall or rise can occur: high or low.

> Please //
> please //

A shift to high level generally calls special attention to the segments so marked; a shift to low level often indicates that an item of information is known or expected.

Note that in English, tune is also important in signalling thematic progression. It is used, for example, to show the distinction between dependent and independent clauses.

> Because I'm busy / I don't want to be interrupted //

If *busy* were spoken with a fall rather than a rise, this sentence would sound odd.

Paralinguistic cues, finally, are the relative pitch level or loudness of an entire tone group, rather than part of the group as in nucleus placement, and the rhythm or tempo of the utterance. In English, these cues usually signal special discourse functions, such as distinctions in degree of formality; they can also mark quotes, interjections or asides, or indicate, for example, announcing style as in *Tickets, please*.

To be understood at all, all sentences must carry some kind of tone grouping, nucleus placement and tune. When these are in keeping with expectations based on content, no additional meanings are signalled. However,

there are also certain optional uses of prosody to highlight unexpected information which function to suggest indirect inferences. For example, isolating an utterance segment as part of a separate tone group, as the bus driver did in my second example, assigns it special importance and invites the listener to infer the reason. Note, however, that in British as well as in American English, tone grouping options are constrained by pragmatic rules. Of the following examples, (1), (2), and (3) are all possible.

1   See that chair over there in the corner. //
2   See that chair / Over there in the corner. //
3   Put that chair over there in the corner, //
4   Put that chair / Over there in the corner. //

Example (4) seems odd, however, since *over there* is semantically a part of the predicate, rather than part of a separate adverbial complement.

Optional nucleus placement on an item which under ordinary conditions would count as given, is unexpected. The hearer's attempt to understand the speaker's motivation constitutes the conversational inference. The woman plane passenger in my first example uses this device in saying 'I told you to leave him at home' and, given our knowledge of similar situations and of the extralinguistic setting, we use this information to identify her utterance as formulaic.

Similarly, the use of high rise or fall when low rise or fall is expected can serve to signal special emphasis. I use the term 'normal information flow' to indicate uses of prosody which are expected and signal no indirect inferences. The term 'contrastiveness', on the other hand, refers to those cases where deviations from expected patterns are exploited conversationally.

Note that while short utterances need not show contrastiveness, longer utterances involving complex, connected discourse employ contrastiveness as an essential part of the signalling process. Only through contrastiveness can we scan utterances to determine the relative importance of various bits of information in longer messages.

[. . .]

In contrast to prosody, paralinguistic cues are somewhat more optional in English. Nevertheless, they are a regular feature of everyday conversation. In fact, as already suggested, they are our primary means of distinguishing various degrees of formality of talk and degrees of interspeaker involvement, of signalling topic changes, and distinguishing between asides and main parts of the argument. In our first example, the paralinguistic cues enabled us to identify *Tickets, please* as an announcement.

Let us now return to the second example:

Exact change, please.

As previously noted, the West Indian bus driver said this sentence twice, using different contextualization cues in each case. A speaker of British English in repeating this utterance, could optionally (a) place the nucleus on *change* or (b) split the sentence into two tone groups with two nuclei: *change* and *please*. In (a), the normal interpretation would be, 'I said, "change".' In (b), the separation of *please* would emphasize that word and call attention to the fact that a request has been made. Note that in (b) *please* must carry rising time, to suggest tentativeness and avoid excessive directness, which would seem rude. The bus driver in our example said *please* with falling intonation as well as increased pitch and loudness. Hence, for speakers using British English contextualization conventions, the conclusion of rudeness is natural.

In order to determine whether the interpretation of rudeness corresponds to West Indian contextualization conventions, we want to look at how prosodic and paralinguistic cues normally function in West Indian conversation. Examination of the contextualization conventions employed in our tapes of West Indian Londoners talking to each other, suggests that their use of prosody and paralinguistics is significantly different from that of British English or American English speakers. For example, syntactic constraints on the placement of tone group boundaries differ. West Indians can split a sentence into much smaller tone group units than British English speakers can. Furthermore, their use of rising tone to indicate intersentence connections is much more restricted. Moreover, once a tone group boundary has been established, nucleus placement within such a tone group must be on the last content word of that tone group, regardless of meaning. In contrast to other forms of English, therefore, nucleus placement is syntactically rather than semantically constrained. Finally, pitch and loudness differences serve as a major means of signalling contrastiveness rather than expressiveness. They are regularly used to indicate emphasis without any connotation of excitement or other emotional overtones. To give only one example, in the course of an ordinary, calm discussion, one speaker said,

He was selected/ MAINLY/ because he had a degree//.

The word *mainly* was separated by tone group boundaries and set off from the rest of the sentence by increased pitch and loudness. The context shows that the word *mainly* was used contrastively within a line of reasoning which argued that having practical experience was as important as formal

education. Our conclusion is that the West Indian bus driver's *Exact change / please //* was his normal way of emphasizing the word *please*, corresponding to the British English option (b). Therefore, his intention was, if anything, to be polite.

To summarize, then, we conclude that conversational inference processes such as we have discussed involve several distinct elements. On the one hand is the perception of prosodic and paralinguistic cues. On the other is the problem of interpreting them. Interpretation in turn requires, first of all, judgments of expectedness and then a search for an interpretation that makes sense in terms of what we know and what we have perceived. We can never be certain of the ultimate meaning of any message, but by looking at systematic patterns in the relationship of perception of surface cues to interpretation, we can gather strong evidence for the social basis of contextualization conventions.

[. . .]

Note

References

Gumperz, J. (1976) 'Language communication and public negotiation', in Sanday, P. (ed.) *Anthropology and the Public Interest*, New York: Academic Press.

Gumperz, J. and Cook-Gumperz, J. (1976) Papers on language and context. Working Papers No. 46, Language Behavior Research Laboratory, University of California, Berkeley.