Some criticisms of Sacks, Schegloff, and Jefferson on turn taking

R. J. D. POWER and M. F. DAL MARTELLO

By 'turn taking' we mean the coordination of the activities of speaking and listening in a conversation so that at any given moment there is only one speaker. Over the last fifteen hears this topic has attracted much interest. It has been investigated in two ways: first, by analyzing transcriptions of naturally occurring conversations; and second, by analyzing video-recordings, with an emphasis on the role played by nonverbal signals. The first of these methods, now usually labelled 'conversational analysis', was pioneered by Sacks and his colleagues: their major publication on turn taking is 'A simplest systematics for the organization of turn taking for conversation' (Sacks, Schegloff, and Jefferson 1974). The theory presented in that paper has had considerable influence, as can be gauged from the numerous references to it in the literature, and the fact that it has been reproduced in several textbooks (see, for example, Clark and Clark 1977, Coulthard 1977); yet so far as we know, it has never been subject to detailed criticism. We undertake such a criticism here for two reasons: first, destructively, because we think the theory has serious weaknesses; and second, constructively, because an examination of these weaknesses brings out several points of theoretical importance.

Summary of the Sacks, Schegloff, and Jefferson theory

To begin with, a few words should be said about the data upon which the Sacks, Schegloff, and Jefferson (SSJ) theory is based. The method used by SSJ is to tape-record a variety of natural interactions (telephone calls, encounter groups, coffee-room conservations, and so forth) and to transcribe them more or less as a playwright would, but with greater attention to the exact timing of pauses and interruptions. Intonation is not transcribed in detail, but is suggested to some extent by the use of commas, question marks, italicized words, and so forth. The authors illustrate their points with quotations from these trans-

cripts, but give no statistics; we have to trust that the examples quoted are typical.

SSJ use the word 'turn' ambiguously: sometimes it refers to the right to speak, as in the expression 'turn-allocation'; at other times it refers to the utterance which the rightful speaker produces, as in the expression 'turn-construction'. The principal use is the former; the latter is really an abbreviation for 'utterance produced during a turn'. Their model has two parts, a 'turn-constructional component' and a 'turn-allocational component'. A turn may be constructed from various syntactic units: it may for instance consist of a word, a phrase, a clause, or a sentence. Once an utterance is under way it should be possible for observers to guess which unit the speaker plans to use, and in this way to judge when the utterance is complete. The first possible completion point of an utterance is called a 'transition-relevance place', since, when this point is reached, the turn is reallocated and may pass to a new speaker.

The allocation of the turn proceeds as follows. The current speaker may, if he wishes, choose the next speaker by using in his present utterance a 'current-speaker-selects-next' technique such as an addressed question. This method of allocating the turn has precedence over the others. If the current speaker foregoes this option, the other participants may 'self-select' by beginning utterances of their own, the first person to speak up acquiring the turn. Finally, if the other participants let this opportunity pass, the previous speaker may, if desired, take another turn. In this case the same turn-allocation procedure occurs at the next transition-relevance place, until eventually the turn is transferred to another participant.

The turn-constructional and turn-allocational components that we have just described constitute the basis system. In addition, however, SSJ describe a number of 'repair mechanisms' which come into operation when the basic system breaks down. The commonest problem is a multiple start during self-selection; this is repaired by the abortion of all but one of the overlapping utterances. Another problem is that the first person to speak up during self-selection may preempt an utterance of higher priority; in such cases 'second-starter techniques' are used — in other words, the first speaker is interrupted.

The nature of the theory

SSJ present their theory in the form of a set of rules, and it is important to clarify straight away what these rules are supposed to represent. The crucial point is that they are not supposed to describe a pattern of

behavior to which people always conform. They have, rather, the force of a cultural convention: their purpose is to describe how people in our culture must behave if they are not to be regarded as incompetent or uncooperative. Such a theory obviously cannot be evaluated by merely considering the degree to which people in our culture conform to the postulated rules. One must also take into account how people react when the rules are broken. If people treat deviations from the rules as problems that call for special repair measures, and if persistent deviants are inferred to be incompetent or anti-social, then we have grounds for thinking that the postulated rules are indeed a convention of our society. A counter-example would have to be a case in which a deviation from the rules failed to provoke such reactions, either in the participants, or (if the incident was recorded) in subsequent observers.

Some criticisms of the theory

Each of the next five sections criticizes a different aspect of the theory. The first discusses a general point: the failure of the theory to assign any role to nonverbal signals. The others examine four specific components of the theory: turn construction; current-speaker-selects-next techniques; self-selection; and repair mechanisms.

Nonverbal signals

The 'simplest systematics' theory ignores the empirical literature (already considerable in 1974) concerning the influence of nonverbal signals on turn taking. The only possible defense for this omission would seem to be that SSJ intended their theory to have maximum generality — to apply not only to face-to-face interactions but also to cases like telephone conversations in which visual signals are impossible. If this was the intention, however, the theory should have been constructed in such a way that the special techniques of face-to-face interaction could be added to it later. As it is actually presented, the theory allows no role at all to nonverbal signals, and thus implies that they are irrelevant.

A natural way of accommodating nonverbal signals within a general model would be to allot them the role of displaying how strongly a participant wished to speak (or to listen). In a telephone conversation a strong wish to speak can be signalled only by starting early or by raising one's voice; in face-to-face interactions, on the other hand, additional signals are available. In effect, SSJ consider only one method of displaying

an eagerness to speak, that of starting early, and they ignore the evidence suggesting that other methods are important: in particular, overloudness (Meltzer et al. 1971), gaze aversion (Kendon 1967), and gesticulation (Duncan 1972).

Turn construction

Before we look in detail at the turn-constructional component, we want to make some preparatory remarks about the function this component is supposed to serve. The basic observation which must be explained is that changes of speaker occur overwhelmingly at natural boundaries in the discourse. There are relatively few places at which another participant can begin to speak without seeming to interrupt. In SSJ's terminology these are called 'transition-relevance places', and the utterance in progress at a transition-relevance place is said to be 'possibly complete'. The difficulty, of course, is to state what counts as a possibly complete utterance. Several issues arise here:

a. Is there a clear-cut distinction between complete and incomplete utterances, or are there merely varying degrees of completion?

b. Do judgments of utterance completion depend upon the utterance itself, or on the nonverbal behavior of the speaker, or both?

c. Which features of an utterance do judgments of completion depend upon: prosodic, syntactic, semantic, or pragmatic? (See Rosenfeld 1977: 298–303.)

There seems to us to be just one reasonably certain result, namely, that an utterance is never judged complete while a tone group is still in progress.

With these points in mind we return now to the turn-constructional component of the 'simplest systematics'. SSJ describe this component in the following way:

There are various unit-types with which a speaker may set out to construct a turn. Unit-types for English include sentential, clausal, phrasal, and lexical constructions. Instances of the unit-types so usable allow a projection of the unit-type under way, and what, roughly, it will take for an instance of that unit-type to be completed .... The first possible completion of a first such unit constitutes an initial transition-relevance place. (Sacks, Schegloff, and Jefferson 1974: 702–703)

The first point to note about this passage is that although it implies that the relevant units are syntactic, it goes on to say, in effect, that an utterance may be constructed from almost any syntactic unit. Actually even this weak claim is dubious: an intuitively complete utterance can be constructed from a word-string which cannot be described by any one

syntactic category. Consider, for example, the words 'The red' uttered in reply to 'Which pen is yours, the red or the green?'. Of course, the person uttering 'The red' would show that he was eliding the noun by putting a terminal contour on 'red' — but in the account of turn construction quoted above, intonation is not mentioned.[1]

The second, and crucial idea in this passage is that it is possible to 'project', from the speaker's opening words, the syntactic unit with which he intends to construct his whole turn. The suggested sequence of events seems to be this: (a) the speaker selects a syntactic unit which suits his message — a noun phrase perhaps; (b) he begins this unit; (c) the listener guesses which unit the speaker has selected; (d) when this unit is completed, the listener judges the utterance to have reached a possible completion place. No details are given of how or when the listener infers which unit the speaker has selected. Presumably he would have to reach at least a partial conclusion almost immediately, since the utterance might be no longer than a word or a short phrase.

On what basis, then, could the listener infer the syntactic unit of the whole utterance from the first few words? So far as we can see, the only possible strategy would be to infer, taking into account the context of the utterance, what type of message the speaker intended to convey, and then to select that syntactic unit which was (a) most suited to the inferred nature of the message, and (b) consistent with the utterance to date. For instance, if Mary had just asked John 'Who did you have lunch with today?', and John's reply began 'The __, then Mary might reason as follows: the context requires John to identify a person (or persons); the syntactic unit best suited to this purpose is the noun phrase; noun phrases can begin with 'the'; therefore I shall assume that a transition-relevance place has arrived as soon as John completes a noun phrase.

It is easy to see, however, that this policy might lead to absurdity. Suppose that in the example just cited John intended to reply 'The man from the insurance company', would Mary consider it appropriate to begin a new turn after the word 'man'? Of course not, since in this context 'the man', although a complete noun phrase, fails to identify a particular person. And this example leads us to final point: if it is necessary in any case to infer the nature of the speaker's message, why not use *this* as one's criterion of utterance completion, instead of speculating about how such a message might be realized syntactically?

Current-speaker-selects-next techniques

SSJ mention four techniques by which the current speaker can determine who speaks next:

1. Addressed questions (or, more generally, initial parts of 'adjacency pairs'[2]). Example: What's for lunch, mum?'
2. One-word questions referring to the preceding utterance. Example: 'Who?'
3. Utterances ending in a tag question. Example: 'That's John, isn't it?'
4. Utterances which, given the social relationships among the participants, could only be appropriately answered by one of them (or perhaps, one of a subset of them). Example: in a conversation between two couples, the Smiths and the Browns, if Mr Smith says 'Would you like to see a film?', one of the Browns must respond. (This example is adapted from Sacks, Schegloff, and Jefferson [1974: 718].)

Before criticizing this part of the theory, we want to make it clear that we accept the data on which it is based. That is to say, we see no reason to doubt that in most cases an addressed question is promptly answered by the addressee, a one-word question is answered by the previous speaker, and so forth. (One would prefer, even so, to see some figures.[3]) If, then, SSJ were content to say that the current speaker could, in these various ways, influence who spoke next, there would be no reason to dissent. But they make a much stronger claim than this. They state that the above techniques place the selected individual under an *obligation* to speak next, so that failure to do so would be regarded as a breach of normal conduct requiring special explanation. So that there may be no doubt on this point, we quote their actual words:

If the turn-so-far is so constructed as to involve the use of a 'current-speaker-selects-next' technique, then the party so selected has the right and is obliged to take next turn to speak; no others have such rights or obligations, and transfer occurs at that place. (Sacks, Schegloff, and Jefferson 1974: 704)

We believe that this claim is false, and can be refuted by counterexamples. We give below a composed dialogue containing a counterexample to each of the four selection techniques listed above; the reader may judge whether, at any point, a participant would have been perceived to be speaking out of turn. The scene is the house of John and Jane, who are being visited by another married couple, Mike and Mary. The labels (1), (2), etc. indicate which selection technique has just occurred.

(*John, Jane, and Mary are in the living room. Mike is outside.*)

John: Have you seen Mike, Jane? (1) I want him to help me move the piano.
Jane: Mary was the last person to see him.
John: Where? (2)
Mary: In the garage.

(*Mike is called and enters the living room.*)

John: You don't mind helping me, do you? (3) Just for a few minutes?

(*Some time later ....*)

Mike: We'd better be getting along.
John: Why not come again next weekend? (4)
Jane: Yes, we'd love to have you.

If these counter-examples are acceptable, it must be conceded that the four techniques mentioned by SSJ do not in fact select the next speaker. We would indeed go further and suggest that no technique of this sort exists for conversation. The nearest case we can think of is the way in which 'Roger' and 'Over' are used when messages are passed by radio or through a computer system; but in normal conversation such devices never occur.

Although we reject this part of the SSJ model, we believe that it contains an important insight: namely, that utterances such as addressed questions powerfully constrain the subsequent course of the dialogue. SSJ may have mistaken the nature of this constraint, but they are surely correct in asserting that some sort of constraint exists. It is instructive to pursue this point a little. Consider again the case of an addressed question. Suppose that there are three participants, John, Mary, and Bill, and that the following sequence occurs:

John: Do you have the time, Mary?
Bill: Where are you going this summer, Mary?

Here, as SSJ predict, Bill's utterance would be judged improper — but not because it was Mary's turn to speak. The following successor to John's question is equally anomalous, although spoken by Mary —

Mary: Where are you going this summer, Bill?

while, as we have seen, it might easily be acceptable for John to continue—

John: Do you have the time, Mary? My watch has stopped.

or for Bill rather than Mary to supply the time if he was better placed to know it. The reason Bill's utterance in the original sequence is out of place is not that it violates Mary's turn to speak, but that it violates *John's* turn to achieve one of his goals. The relevant constraint, then, is that an unresolved topic should not be interrupted unless the new topic is more urgent.

We have pursued this example because it contains an important lesson: that if we use the concept of 'turn', we must take care to distinguish the speaking turn from other kinds of turns such as turn to contribute to a topic, or turn to introduce a topic. We shall point out another example of this possible confusion when we discuss repair mechanisms.

Self-selection

The self-selection rule states, in effect, that if the current speaker has foregone his option to select the next speaker, then the first of the other participants to speak up acquires the right to the turn. This rule has two important implications:
1. If X begins an utterance, thereby pre-empting another participant Y, then Y should postpone his utterance and listen to X.
2. If in this situation Y does not hear X in time to check his own utterance, then Y, not X, should withdraw.
(It is assumed in either case that Y has no special justification for interrupting.)

Though we have some reservations about the self-selection rule, it seems to us to be a plausible proposal, and we want to begin by quoting some points in its favor. First of all, it appears to be true (although again, SSJ supply no statistics) that utterances are rarely interrupted during the first few words. In most contexts repeated interruptions of this kind would not be regarded as proper. There is also experimental evidence (Meltzer et al. 1971, Morris 1971) that when overlap does occur, a defensive raising of the voice by the previous speaker is particularly effective, much more effective than a corresponding raise by the challenger. The authors interpret this asymmetry as evidence that the person already speaking is judged to have the more legitimate claim to the turn.

The main weakness of the self-selection theory is the implication that the destiny of the next turn (assuming no current-speaker-selects-next technique) should depend soley upon who manages to speak up first. We have already cited, in the section on nonverbal signals, evidence that there are other ways of displaying an eagerness to speak (gaze aversion, hand gestures, etc.). There is also evidence that negotiations regarding the next turn can be conducted nonverbally during the current utterance. For example, Duncan (1974) found that when the listener produces a response of the 'mm-hm' type before a phonemic clause juncture (rather than at the juncture, as is normal), the speaker is more likely than usual to emit floor-retaining signals such gaze aversion and gesticulation; this suggests, as Rosenfeld (1977: 314) points out, that an early listener response is taken

to mean that the listener has already understood what the speaker is going to say, and might be preparing to speak himself.

Our other reservation about the self-selection theory is more fundamental. We think it can be plausibly argued that the data on which the theory is based can be explained by general principles of rationality and cooperativeness, without invoking turn-taking conventions at all. We intend to develop this point in another publication, but a sketch of the argument might be useful here. The basic observation to be explained is that once an utterance is under way it is not usually interrupted without special reason; other participants who were planning to speak usually withdraw. The general principle of rational behavior applying to this situation is, we suggest, that one should try to avoid wasted effort. This principle applies with even greater force in social situations in which effort has been invested by another participant, since in this case any action causing an effort to be wasted would be not just inefficient but also inconsiderate. It seems clear that in most cases, the interruption of an utterance during the first few words would oblige the speaker either to abandon it or to begin it again, thus relegating the part already delivered to wasted breath. The advantage of this line of explanation, apart from economy, is that it also accounts for those cases in which interruption is acceptable. If the speaker's utterance is unnecessary, or ineffective, or not directed to the goal of highest priority, principles of rationality and cooperativeness may warrant or indeed oblige interruption.

Although we believe the above case is a strong one, it is not decisive. It is possible, for example, that the general principles referred to above have given birth to specific conventions, rather like rules of thumb, which apply to specific activities such as conversation, and which, in accordance with general principles in normal circumstances, are easier to apply. Whether this is so in the case of turn taking is an empirical question. It should be possible to investigate this issue by examining cases in which specific convention and general principle disagree.

Repair mechanisms

SSJ do not give a systematic treatment of repair mechanisms, but we want to comment on one of their illustrations, in which, it seems to us, they fail to draw the above-mentioned distinction between the speaking turn and other kinds of turn. They state: 'No special theoretical motivation is needed to observe that questions such as "who, me?" ... are repair devices directed no troubles in the organization and distribution of turns to talk.' (Sacks, Schegloff, and Jefferson 1974: 723), and quote the following instance:

MOTHER: Wha'd are you doin.
DAUGHTER: Me?
MOTHER: Yeh, you (etc.)

The point, of course, is that the daughter's 'Me?' is not concerned with whose turn it is to speak, but with who should answer the mother's question. Such utterances can only be spoken properly and understood if the issue of the speaking turn has been settled already. Another typical case is that in which X and Y begin utterances simultaneously, then both stop, then X says 'Sorry, go ahead'. Again it would be mistaken to regard X's utterance as repairing a problem of floor apportionment; on the contrary, it is possible because the problem of the speaking turn has already been solved. What it does repair is the problem of who should contribute next to the main topic: it proposes that the next contributor should be Y. This is not the same as saying that Y should speak next. There would be no inconsistency in X's continuing to speak, provided that he confined his remarks to the side topic rather than the main topic. Thus 'Sorry, go ahead. I thought you'd finished' is not at all anomalous.

The source of this confusion is probably that in ordinary usage 'turn to speak' is usually taken to mean 'turn to contribute to the main topic'; this concept might arise, for example, in a debate in which it was necessary to allow both points of view a fair hearing. In scientific writings, however, expressions like 'speaker' and 'turn to speak' have come to have different meanings and should be regarded as technical terms.

Conclusions

Our conclusions specifically concerning the 'simplest systematics' theory are as follows:
a. The turn-constructional component is inadequate. There is no evidence that judgments of utterance completion are based on syntactic projection; moreover, no plausible mechanism of syntactic projection has been proposed, and it is doubtful whether any could be. It would seem more plausible to seek an explanation based on the goals of the discourse, though such an explanation would admittedly be difficult to formulate precisely. The empirical literature suggests that intonation is also relevant.
b. The proposal that the current speaker can decide who speaks next by using one of a set of conventional techniques is also, on a strict interpretation, mistaken. It may be based on a failure to distinguish the concept of 'turn to speak' from that of 'turn to contribute to the main topic'.

c. The self-selection component of the theory has more plausibility: speaking up early certainly seems to be an important way in which a person can try to gain the floor. The trouble with this part of the theory is that it oversimplifies: it ignores the considerable empirical literature showing that people can also struggle or negotiate for the floor by means of nonverbal signals.

We want, finally, to draw two general morals from our examination of the theory. The first concerns the meanings of the terms 'speaker' and 'turn'. We have already noted that these terms invite misunderstanding, since their scientific use differs from their ordinary use. A definition of their technical meanings is therefore necessary. We suggest the following: a participant X *has the turn* (or equivalently *is the speaker*) at time t *iff* the participants in the conversation mutually understand at t that X should say something and that the others should attend to what X is saying. Note that it follows from this definition that utterances such as 'Have you got the time, Mary?' and 'Sorry, go ahead' do not transfer the turn to another participant: if they did, then the utterances 'Have you got the time, Mary? My watch has stopped' and 'Sorry, go ahead. I thought you'd finished' would be paradoxical, which they are plainly not.

Our second moral is more fundamental: it concerns the nature of the theory rather than its detailed content. It seems to us that SSJ assume too easily that a regularity in social behavior must be due to a social convention specifically prescribing that regularity. Observing that the first person to speak up is usually allowed to continue, or that addressed questions are usually immediately answered by the addressee, they assume without discussion that these regularities are due to specific turn-taking conventions; the possibility that they are due instead to general pragmatic principles of efficiency and consideration for others is not mentioned. In making this criticism we are assuming, of course, that the rules given by SSJ are supposed to represent a cultural convention and not merely a statistical regularity. This seems to us to be clear, partly because the rules make repeated use of such concepts as 'right' and 'obligation', and also because the authors have stated in several places (e.g. in Schegloff and Sacks 1973: 290) that the rules of social behavior which they aim to discover are those already used by the participants themselves.

Notes

1. SSJ do, however, acknowledge the importance of intonation in their commentary on the model: 'When it is further realized that any word can be made into a "one-word" unit-type, via intonation, then we can appreciate the partial character of the unit-types' description in syntactic terms'. (Sacks, Schegloff, and Jefferson 1974: 722)

2. For an explanation of this term, see Schegloff and Sacks (1973). Examples of adjacent pairs are: question–answer; proposal–response; statement–acknowledgement; and greeting–greeting.
3. Yngve (1970) makes a point of the fact that, in his data, questions were not always followed by floor switches.

References

Clark, H. H. and Clark E. V. (1977). *Psychology of Language: An Introduction to Psycholinguistics*. New York: Harcourt Brace.
Coulthard, M. (1977). *An Introduction to Discourse Analysis*. London: Longman.
Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology* 9(1), 79–84.
—(1974). On the structure of speaker–auditor interaction during speaking turns. *Language in Society* 3(2), 161–180.
Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychologica* 26, 22–63.
Meltzer, L., Morris, W. N. and Hayes, D. P. (1971). Interruption outcomes and vocal amplitude: 'Explorations in social psychophysics. *Journal of Personality and Social Psychology* 18, 392–402.
Morris, W. N. (1971). Manipulated amplitude and interruption outcomes. *Journal of Personality and Social Psychology* 20(3), 319–331.
Rosenfeld, H. M. (1977). Conversational control functions of nonverbal behavior. In *Nonverbal Behavior and Communication*, A. W. Siegman and S. Feldstein (eds.), 291–328. Hillsdale, N.J.: Lawrence Erlbaum.
Sacks, H., Schegloff, E. A. and Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language* 50(4), 696–735.
Schegloff, E. A. and Sacks, H. (1973). Opening up closings. *Semiotica* 8(4), 289–327.
Yngve, V. (1970). On getting a word in edgeways. Papers from the 6th Regional Meeting of the Chicago Linguistic Society, 567–577.

Richard Power (b. 1948) is Lettore di lingua inglese in the Department of Psychology, Padua University. His principal research interest is in language comprehension. His major publications include 'Learning to count: A computational model of language acquisition' (1978), and 'The organisation of purposeful dialogues' (1979).

Maria Felicita Dal Martello (b. 1950) is a researcher in the Department of Psychology at Padua University.