Conversational Organization
INTERACTION BETWEEN SPEAKERS AND HEARERS

Charles Goodwin
Department of Anthropology
University of South Carolina
Columbia, South Carolina

ACADEMIC PRESS
A Subsidiary of Harcourt Brace Jovanovich, Publishers
New York London Toronto Sydney San Francisco
3. Notes on the Organization of Engagement

- Engagement Displays
- Disengagement
- Entering Disengagement
- Some Alternative Possibilities
- Talk within Disengagement
- Selecting from Coparticipation Alternatives
- Accounting for Withdrawal
- Conclusion

4. Modifying Units of Talk to Coordinate Their Production with the Actions of a Recipient

- Lengthening Units by Adding to Their Ends and Middles
- Changing the Emerging Structure of the Speaker's Sentence
- Adding Segments Repetitively
- The Use of Repairs to Lengthen a Unit
- Maintaining Focus on Talk
- Changing the Length and Meaning of Nonvocal Units
- Summary and Conclusion

5. Designing Talk for Different Types of Recipients

- Requesting the Aid of a Knowing Recipient
- Talk in the Presence of a Knowing Recipient
- Discovering New News
- Conclusion

6. Conclusion

Appendix

Data Sources

References

Subject Index

Transcription Conventions

Speech

1 2 3 4 5 6 7 8
↓ ↓ ↓ ↓ ↓ ↓ ↓ ↓

JANE: It was (-- - - - - - + - -) so: : (0.3) "incredible.
[...

MEG:

↓ ↑ ↑
9 10 11

1. A dash marks a cut-off.
2. Italics show that the talk so marked is being emphasized in some fashion.
3. Dashes within parentheses indicate tenths of a second within a silence.
4. Each full second within a silence is marked with a plus sign.
5. A colon indicates that the sound preceding it is noticeably lengthened.
6. Numbers within parentheses constitute an alternative way of showing the duration of a silence.
7. A degree sign shows that the talk so marked is spoken with noticeably lowered volume.
8. Punctuation marks indicate intonation as follows:
   A falling contour is marked with a period.
   A rising contour is marked with a question mark.
   A falling-rising contour is marked with a comma.
9. A bracket connecting the talk of different speakers shows that overlapping talk begins at that point.
10. If words are placed within parentheses, a possible but not certain hearing of that talk is indicated.
11. An equals sign indicates that no break occurs between two pieces of talk by either the same or different speakers.

Gaze

12  
16

↑

↓

12 16

SPEAKER: Brian you're gonna ha v- You kids'll have to go down closer

HEAINER:

13 14 15

The gaze of the speaker is marked above the utterance; that of the recipient(s) is marked below it.

12. A line indicates that the party so marked is gazing toward the other.
13. The absence of a line shows that that party is not gazing toward the other.
14. A dot or series of dots marks the movement that brings gaze to another.
15. A capital X connected to a specific point in the talk with a bracket shows the place where gaze reaches the other.
16. Commas mark a movement withdrawing gaze.

Preface

To engage successfully in conversation, participants are required not only to produce sentences but also to coordinate, in a meaningful fashion, their talk with the talk of others present. Such activity constitutes a pervasive—but intricate—form of human social organization, one that makes full use of both the linguistic and the cultural competence of the parties engaged in it. It is therefore not surprising that the analysis of conservation has begun to attract the attention of scholars from a number of different disciplines. The present study takes as its point of departure the work of Harvey Sacks and his colleagues on the sequential organization of conservation, and investigates some previously unexamined features of this process.

Within conservation, talk proceeds through a series of turns at talk. The most basic social identities relevant to the turn are those of speaker and hearer. This analysis focuses on how the turn is constituted through the mutual interaction of speaker and hearer. Perhaps the best way to make clear what is meant by this is to provide a brief overview of the material to be covered.

In Chapter 1, relevant research is reviewed and methodological preliminaries, such as the transcription system, are presented. Data for the analysis consist of videotapes of conversation recorded in a range of settings.

Chapter 2 investigates the work participants do to bring about a state
of affairs in which the talk of the speaker is being addressed to, and attended by, a hearer. Among the phenomena given special attention are ways in which displays of hearership affect the talk in progress, the use of phrasal breaks—such as restarts and pauses—to request gaze, and the ordering of mutual gaze within the turn.

It is also possible to produce talk without speaker and hearer displaying explicit mutual orientation. In Chapter 3, alternative types of engagement frameworks are investigated. Particular attention is paid to the organization of disengagement and also to how participants work with each other to move from one type of engagement to another.

Chapter 4 examines the ability of speakers to add new sections to units in their talk so that the talk can be precisely synchronized with relevant actions of the hearer. This process is found to be operative in the production of units on many different levels of organization and may lead to changes in the sentence being constructed within the turn.

Chapter 5 investigates ways in which speakers differentiate recipients, as well as some of the consequences this has for the production of their talk. As speakers move their gaze from one type of recipient to another, they change the sentence in progress so that it remains appropriate to its recipient of the moment.

Each chapter thus investigates a particular phenomenon implicated in the organization of action between speaker and hearer within the turn at talk. The issues posed are quite basic:

— How is a framework of mutual orientation achieved within the turn?
— What constitutes a display of hearership and how is it relevant?
— What engagement alternatives are open to participants and how do these alternatives affect the talk of the moment?
— How are the separate actions of speaker and hearer coordinated with each other within the turn?
— How do speakers make visible the appropriateness of their talk for its recipients and what consequences does this have for the structure of the talk?

An examination of the ways in which these issues are dealt with by participants serves to make visible some of the constitutive features of the turn, as well as to reveal many intricate, finely coordinated processes of interaction that occur with it.

The phenomena being investigated should be of interest to researchers concerned with social interaction and language in a number of different fields including sociology, linguistics, anthropology, psychology, and communications. The study as a whole is relevant to work in nonverbal communication, and some of the analysis in Chapter 2 complements work done in psychology and psycholinguistics on speech errors.

Acknowledgments

Most of the analysis, with the exception of Chapter 3, was done as part of my work toward a doctorate in communications at the University of Pennsylvania. I owe a tremendous debt to the people who taught me there, and would like to thank in particular Klaus Krippendorff, William Labov, Ward Goodenough, and Erving Goffman. In addition, I had the great privilege of being able to work with Harvey Sacks on some of this material. Others whose aid and suggestions were most valuable include Anita Pomerantz, Emanuel Schegloff, and John O'Neill. Two people, Candy Goodwin and Gail Jefferson, require very special mention. Without Candy's constant aid and collaboration and Gail's inspiration, teaching, and criticism, this work would not have been possible. I want to thank all of these people for what they have given me.

A slightly different version of the analysis that constitutes the beginning of Chapter 2 was published as “Starts, Pauses, and the Achievement of Mutual Gaze” in Sociological Inquiry, Vol. 50, numbers 3-4, Special Double Issue on Language and Social Interaction, Don Zimmerman and Candace West (eds.), 1980. Portions of Chapters 4 and 5 (specifically the analysis of the sentence “I gave up smoking cigarettes one week ago today actually”) have appeared as “The Interactive Construction of a Sentence in Natural Conversation” in Everyday Language: Studies in Ethnomethodology, George Psathas (ed.) (New York: Irvington Publishers, 1979). These materials have been incorporated into the present work with the permission of Jossey Bass, Inc. and Irvington Publishers.
I wish to thank Randee Falk and the editorial staff of Academic Press for the help and suggestions they gave for the final revision of this work. In addition, special thanks are owed to Joan Toms Cureton, Ginger Hollis, and Cheryl Fowler for their typing and to John Ballou for his drawings. Final preparation of the manuscript was assisted by a productive research grant from the University of South Carolina.

1 Introduction

Preliminary Definitions

Some of the phenomena to be investigated in this study will here be provisionally defined. The present definitions are provisional because I believe that precise definitions of these phenomena can only be obtained by empirical investigation of their properties.¹

CONVERSATION

Goffman (1975:36) has noted that two different approaches can be taken to the definition of conversation. One can try to capture the sense in which the term can refer to casual talk in everyday settings, or alternatively the term can be "used in a loose way as an equivalent of talk or spoken interaction." It is in this latter sense that the word "conversation" is used in this study.²

Despite the broad scope of the term when it is used in this fashion,

¹ Voloshinov (1973:45) notes that "at the outset of an investigation, it is not so much the intellectual faculty for making formulas and definitions that leads the way, but rather it is the eyes and hands attempting to get the feel of the actual presence of the subject matter."

² For a similarly broad definition of conversation, see Schegloff (1968:1075–1076).
conversation is still but a special case of what Goffman (1963:24) has called focused interaction: "the kind of interaction that occurs when persons gather close together and openly cooperate to sustain a single focus of attention." As such, it stands in contrast to unfocused interaction, the kind of communicative situation concerned with "the management of sheer and mere copresence."

Placing conversation in this typology raises some analytic difficulties. Because Goffman bounds the area of his investigation in terms of copresence, conversations between nonpresent parties—for example, phone calls—are excluded from it. Nevertheless, the distinctions he makes are valuable.

Goffman also notes (1975:33) that though conversation is defined in terms of talk, it can include behavior other than talk. In the present study, conversation is taken to include nonlinguistic as well as linguistic behavior, and both will be investigated; however, talk is seen to occupy a central place in the organization of conversation.

TURN-TAKING

A basic empirical finding about conversation, one that has been discovered independently by different investigators (see, for example, Allen and Guy 1974:30,177; Argyle 1969:201–202; Duncan 1974; Goffman 1964:135; Jaffe and Feldstein 1970:9; Sacks, Schegloff, and Jefferson 1974; Yngve 1970:1–2), and that can be seen by even casual inspection of almost any fragment of conversation, is that talk within it proceeds through a sequence of turns. Miller (1963:418) gives this phenomenon the status of a language universal but notes that it does not seem intrinsically necessary.

In the abstract, the phenomenon of turn-taking seems quite easy to define. The talk of one party bounded by the talk of others constitutes a turn, with turn-taking being the process through which the party doing the talk of the moment is changed.

A number of problems with such a definition emerge when actual conversation is closely examined. For example, both simultaneous talk and silence between the talk of different parties are regularly found. Such phenomena raise relevant theoretical questions about the proper definition of the turn's boundaries as well as the process through which it is exchanged.

4 Goffman (1953:116–117) examines in more detail some of the theoretical issues raised by such a situation. He notes that whereas directed information will be confined to a single conversational cluster, undirected information—for example, one's choice of clothes and companions—will be available to all in one's physical presence. These issues are given more extended treatment in Goffman (1963) where some of the same distinctions are examined with respect to differences between focused and unfocused interaction.

Throughout this analysis I will use the masculine pronoun to refer to speaker and hearer as generic entities. I am only too well aware of the sexist implications of such use and in fact tried to write portions of the analysis so that such pronouns were not used. I found, however, that not only did the writing become more awkward and difficult to follow but that some of the distinctions I was trying to make in the analysis were obscured or lost altogether. As there is no neuter singular pronoun in English, I have therefore reluctantly decided to continue with such use of these pronouns.

However, providing a better description of either the turn or turn-taking requires careful investigation of actual data. Such analysis is beyond the scope of the present attempt to provide preliminary definitions. Though the definition that has been given will eventually be found inadequate, it does at least locate a phenomenon that can be made the subject of further investigation. When research into the structure of turn-taking is discussed, other definitions of the turn will be examined in terms of their ability to accurately characterize the phenomena being studied.

TYPES OF PARTICIPANTS

The term "participant" will be used to refer to anyone engaged in a conversation. For example, on a busy street, several different "whites" (Goffman 1971:19–27) may be simultaneously engaged in conversation. A party is a participant to the conversation in his "with" but not a participant to conversations in other "whites." A person not part of a relevant conversation will be called a nonparticipant. Although in many cases—such as the street example—the distinction between participant and nonparticipant is quite clear, in other cases—for example, when a new member is joining a casual group—the distinction may be ambiguous and may even itself be one of the events at issue in the interaction. I wish to leave the manner in which the distinction is formulated in such cases a matter for empirical investigation. I also wish to use the term "participant" in a broad enough sense to include someone who is momentarily disattending the conversation.

A party whose turn is in progress at a particular point in time will be called a speaker. In that pauses may occur within a turn, a party may be a speaker even though he is not saying anything at the moment. Because the term "speaker" is defined in terms of the turn, in some circumstances—such as simultaneous talk—whether a party is a speaker
may be subject to dispute (for analysis of this and related issues, see Jefferson 1973).

Duncan (1974a:302) has defined an “auditor” as “a participant who does not claim the speaking turn at any given moment.” This definition seems inadequate in a number of respects. First, Schegloff (1968:1092–1093) has noted that conversation is a “minimally two-party” activity. That requirement is not satisfied by the mere copresence of two persons, one of whom is talking. It requires that there be both a “speaker” and a “hearer.” . . . . “Heapership” can be seen as a locus of rules, and a status whose incumbency is subject to demonstration . . . .

Second, a number of different types of nonspeaking participants must be differentiated. Goffman (1975:3) makes the following distinctions.

Broadly speaking, there are three kinds of listeners to talk: those who overhear, whether or not they are ratified participation is inadvertent and whether or not it has been encouraged; those who are ratified participants but (in the case of more than two-person talk) are not specifically addressed by the speaker; and those ratified participants who are addressed, that is, oriented to by the speaker in a manner to suggest that his words are particularly for them, and that some answer is therefore anticipated from them more so than from the other ratified participants. (I say “broadly speaking” because all sorts of minor variations are possible—for example, speaker’s practice of drawing a particular participant into an exchange and then turning to the other participants as if to offer him and his words for public delectionation.)

In describing participants to the turn, it is useful to distinguish three different levels of organization.

First, the activity of conversation provides a set of positions for the participants, the most salient being speaker and hearer. These positions have an ongoing relevance to the conversation in that different kinds of actions such as speech and silence are appropriate to each.

Second, distinct from the positions provided by the activity are the actions of individual participants displaying incumbency or nonincumbency in these positions. How participants display their occupancy of the positions provided by the activity of conversation, especially the position of hearer, is one of the topics to be investigated in this study.

Though events on this level of organization are performed by single individuals, they are nonetheless social and include a projection about the other as well as a display about the self. Consider the case of one party, A, addressing an utterance to another, B, who is, however, attending a different speaker, C. In order to adequately describe A’s action, one has to include the projection of B as an addressee; that description is unaffected by whether B displays hearership to A. The actions of B relevant to the position of hearer can be described separately. Further, a display of hearership on B’s part includes a projection of the party he is attending as speaker.

The term “hearer” can thus refer to three quite different objects. First, it might designate the complementary position to “speaker” provided by the activity of conversation. Second, it might refer to the addressee of an act by a speaker. Third, it might designate a party performing acts in his own right relevant to the position of hearer. If these distinctions are not kept in mind, confusion results, since, for example, a party may be an addressee without acting as a hearer.

A third level of organization is provided by events that can only be described in terms of the actions of more than one individual. For example, the exchange of turns in conversation requires action by at least two parties, one who changes his behavior from speaking to hearership and another who moves from hearership to speaking. The actions of either alone are insufficient to provide for an exchange of turns. Within the turn, events such as the address of the speaker toward the hearer and the orientation—or lack of it—of the hearer toward the speaker are defined on this level of organization. What Goffman speaks of as “ratification” has been made clear, one of the inappropriate parties may begin to attend the speaker as a hearer. The speaker might then emphasize who his addressee is (for example, with an address term) while avoiding the inappropriate hearer. The latter, upon recognizing that the utterance is not being directed to him, might then actively turn his attention elsewhere. In such a situation, both nonheapership and nonaddressee have been carefully displayed.

For other discussion of different types of listeners see Bales (1970:6) and Philips (1974:162–163); for an early statement on the importance of conceptualizing an utterance as being addressed to a recipient with specific characteristics see Voloshin (1971:85–86).

It must be recognized that displays of incumbency can be as carefully and relevantly constructed as displays of nonincumbency. For example, a speaker might begin an utterance addressed to a specific party and inappropriate to others present. Before the recipient of
perception” also fits here.11 The identity assumed by one party is ratified, not by his own actions, but by the action of another who assumes a complementary identity toward him. For example, it is quite common in conversation that while a speaker is addressing an utterance to one party, another, who has not been attending him, will also begin to orient to him. In such circumstances, speakers frequently address a subsequent part of their turn to the new party, thus ratifying him as a hearer. The term “collaborative action” has been given to events on this level of organization by Sacks and his colleagues, who have provided extensive investigation of their structure in conversation (Jefferson 1973; Sacks et al. 1974; Schegloff and Sacks 1973).

It should be noted that the terms “speaker” and “hearer” are being used here in a slightly different way than they are usually employed in linguistics. Whereas the present emphasis is on the complementary positions they describe in a particular social arrangement, in linguistics the social character of these terms is usually not given much attention. Rather, the speaker is conceptualized primarily as an entity capable of constructing sentences and, as such, is not generally distinguished analytically from his listener, who is assumed to possess a similar competence in order to be able to comprehend sentences. Thus Chomsky (1965:3) refers to “an ideal speaker–listener.”

**UNITS OF TALK**

Linguistics and allied fields such as kinesics have provided a rich technical vocabulary for describing the units regularly found in conversation. This vocabulary is not, however, without its problems. To begin with, it has been developed within two separate linguistic paradigms, structuralism and generative grammar. These paradigms make very different assumptions both about the nature of the phenomena being examined and about what a proper theoretical description of that phenomena consists of. Therefore, classifications of phenomena formulated within these different theoretical frameworks are not likely to be consistent. For example, Scheflen (1974:19) defines a sentence as follows: “A syntactic sentence is not identified according to a grammatical structure; it is instead that unit of speech that is marked off by certain traditional behaviors that accompany the stream of speech.” Such a definition of the sentence would not be accepted within the framework of transformational grammar. Indeed, Lyons (1972:61) argues that from the perspective of contemporary linguistics, “sentences never occur in speech.” Rather,

as a grammatical unit, the sentence is an abstract entity in terms of which the linguist accounts for the distributional relations holding within utterances. In this sense of the term, utterances never consist of sentences, but of one or more segments of speech (or written text) which can be put into correspondence with the sentences generated by the grammar [Lyons 1969:176].

For my analysis I will find it necessary to examine the details of actual speech as well as abstract linguistic units which do not stand in a one-to-one relationship with the sounds in the speech stream.12 I will use the term “utterance” to refer to the stream of speech actually produced by a speaker in conversation, and the word “sentence,” as well as related terms such as “phrase” and “clause,” to refer to abstract entities capable of describing distributional relationships within and between utterances. Bloomfield (1946:170) defines a sentence as “an independent linguistic form, not included by virtue of any grammatical construction in any larger linguistic form.” Although the structural independence of the sentence can be called into question,13 this definition remains useful. In defining “utterance” as the actual stream of speech, I mean to include the entire vocal production of the speaker—that is, not only those sounds which could be placed in correspondence with elements of sentences, but also phenomena such as midword plosives, inbreaths, laughter, crying, “uh’s,” and pauses. I also do not wish to separate a speaker’s speech into subordinate utterances in terms of sentence-like properties. Rather, I wish to leave units on these different levels of analysis conceptually distinct and admit the possibility of an utterance containing several sentences as well as the possibility of a sentence being constructed through several utterances. The utterance can, however, be divided into subsections in terms of units appropriate to its own level

---

11 A discussion of the reciprocal quality of ratification is found in Goffman (1964:35) and Goffman (1967:34).

12 For example, the word “put” occurs twice in the following fragment of speech but only once in the sentence produced through that speech:

*He put uhm, (.7 sec pause) Tch! Put *crabmeat* on the bottom.*

**Were** I unable to distinguish these different levels of organization, or were I committed to a theoretical framework that recognized the analytic validity of only one, my ability to adequately analyze the structure of conversation would be seriously compromised.

13 The work of Sacks and his colleagues on the sequential organization of conversation has provided some analysis of structures organizing separate sentences relative to each other (see, for example, Jefferson 1973; Sacks 1978; Schegloff 1968). Within linguistics, ties between different sentences have been examined by Gunter (1974), Hiz (1969), and in the work on discourse analysis to be discussed in what follows.
of organization such as the "phonemic clause" or "breath-group." For clarity, I also wish to restrict the use of the term "utterance" to vocal phenomena and not, as Grice (1969:147) does, include the possibility of "sentence-like" nonvocal phenomena such as hand signals.

The definition given the sentence also differentiates "discourse analysis" from the analysis of conversation. George Lakoff (1972:130) defines a "discourse" as "essentially a string of English sentences." In view of the distinction discussed earlier between utterance and sentence, the study of discourse, as it is conducted within the framework of contemporary linguistics, emerges as quite different from the study of conversation. And, in fact, most work on discourse in linguistics has not examined sequences of actual talk, restricting itself to the study of hypothetical sentences. The structure of speech acts, rather than turn-taking, has emerged as the central theoretical problem in this analysis. Moreover, in part because of the particular definition given discourse, analysts of it have not generally viewed events smaller than the sentence as within the scope of their inquiry, whereas analysts of conversation have devoted considerable attention to such phenomena (see, for example, Jefferson 1974; Sacks 1972a, Sacks et al. 1974). The analysis of discourse is thus not the same as the analysis of conversation. However, as the work of Labov (1972a, 1972b; Labov and Fanshel 1977) has demonstrated, much fruitful work can be done from a perspective that makes use of both approaches. In addition, some linguists working from a perspective somewhat different from that of discourse analysis have made important studies of certain phenomena that tie together units larger than the sentence. See, for example, the work of Halliday and Hasan (1976) on cohesion, Gunter's (1974) work on intonation, and van Dijk's (1977) approach to text analysis.

The units of talk considered until this point have all been vocal. However, the definition of conversation provided at the beginning of this chapter was left broad enough to include other types of behavior. Indeed, this interdependence is so strong that the boundary between language and nonlanguage emerges as a difficult theoretical problem. For example, Lyons (1972) notes that the concept of "non-verbal communication" should properly include intonation and stress, which are nonetheless essential components of "verbal" signals, and that the term "paralinguistics" may well include many gestures, facial expression, and eye movements. In order to deal with such issues, Lyons finds it useful to distinguish the different types of behavior that can be found in talk in terms of overlapping, rather than mutually exclusive, categories. The classification he develops is more accurate and useful than the more frequently made distinction between verbal and nonverbal behavior. However, since many of Lyons's distinctions are not relevant to the present analysis, they will not be described in detail here.

Birdwhistell (for example, 1970) provides very detailed description and analysis of the different kinds of nonvocal behavior that can occur in talk and examines the relationship of that behavior to speech. He has stated (1970:xiii) that his goal "was to develop a methodology which could exhaustively analyze the communicative behavior of the body."

In the present work, my primary analytic concern is not with nonvocal phenomena per se, but with rather limited aspects of the structure of the turn at talk. I will therefore examine only a very small part of the nonvocal behavior that occurs in conversation, principally whether a participant is gazing toward a specified other. My decision to limit myself to this very narrow aspect of nonvocal behavior emerges in large part from my recognition of the complexity and intricate order Birdwhistell has demonstrated to be operative in this area.

**Phenomena To Be Investigated**

The analysis in the present study will focus specifically on interaction between speaker and hearer within the turn. It will be argued that one way in which a nonspeaking party can indicate whether he is acting as a hearer is by gazing at the speaker. Hearership can of course be demonstrated in other ways (this technique would obviously not be applicable to telephone conversations), but this is the only method that will be systematically investigated here. A speaker can use gaze to indicate that the party being gazed at is an addressee of his utterance. Other techniques available to the speaker for indicating that his utterance is directed to some specified recipient will also be examined, especially in Chapter 5.

Chapter 2 will investigate some of the ways in which speaker and hearer achieve a state of mutual orientation at the beginning of the turn. It will be shown that speakers who do not obtain the gaze of a hearer may perform specific actions, such as the production of restarts and pauses. After such phrasal breaks, non-gazing hearers generally begin to move their gaze toward the speaker; if they do not, the speaker may
continue to produce phrasal breaks until he obtains an appropriate response. The data support the possibility that the actions of speaker and hearer together constitute a particular type of summons-answer sequence. When the criteria for choice between alternative actions capable of requesting the gaze of a hearer are investigated, it is found that the gaze of the speaker is also relevant to this process. Thus the task of achieving mutual orientation within the turn both provides organization for the bodies of the participants and leads to the production of a range of phenomena in the stream of speech.

It is also possible to produce talk without speaker and hearer displaying explicit mutual orientation; in Chapter 3, alternative types of engagement frameworks will be investigated. First, some of the ways in which disengagement is organized will be examined, and it will be shown that, although during disengagement the participants are explicitly displaying nonorientation toward each other, each is in fact paying close attention to what the other is doing. Analysis will then turn to how participants move from a state of talk to a state of disengagement. After disengagement has been entered, talk is still possible, but this talk has both a different sequential organization at its boundaries and a different structure of coparticipation in its course than talk produced during full engagement. The presence of such engagement alternatives has implications for processes of reengagement, including the phenomena examined in Chapter 2, and also permits participants to negotiate about the type of orientation they are prepared to give a particular piece of talk. Thus the coparticipation status that a strip of talk is seen to have might be the product of an active process of interaction between speaker and hearer as it is being spoken. The organization of engagement integrates the activities of the participants’ bodies into the organization of their conversation, and in so doing has numerous consequences for the structure of their talk.

Chapter 4 investigates the ability of participants to add new sections to their emerging vocal and nonvocal actions. Such ability is found to constitute a resource for the achievement of social organization within the turn, enabling one participant to coordinate the units he is producing with the relevant actions of a coparticipant. Specific phenomena examined include the lengthening of sound articulation within a phoneme, the addition of phrasal breaks of various types to an utterance, the addition of new words and phrases to a sentence, the addition of sentences to a turn, and, finally, the addition of new sections to the nonvocal actions of the participants. Reasons displayed by a participant for the addition of a new segment to a unit are also examined.

Chapter 5 will investigate one way in which possible recipients to a turn might be distinguished from one another. It will be argued that some actions in conversation (for example, reports) propose as their hearer a recipient who does not yet know about the event being described by the speaker, whereas other actions (for example, a request for information) propose a recipient who has knowledge of the event being talked about. These two types of recipients are mutually exclusive in that an action appropriate to one is inappropriate to the other. Situations will be examined in which both types of recipients are present—for example, the situation in which a speaker describes an event in the presence of both someone who has not yet heard about it and someone who himself participated in it. Analysis will focus on the problem of how the speaker can construct a turn capable of providing for the participation of both types of recipients. It will be found that the speaker has available to him a number of techniques that enable him to change an utterance appropriate to one type of recipient into one appropriate to the other. The use of these techniques produces a range of characteristic phenomena within the turn, including changes in the intonation of the utterance, changes in the type of action being constructed by the utterance, and changes in the state of knowledge proposed for the speaker as well as his recipient. In this chapter, some demonstration is provided that the speaker has the ability, not only to add new sections to his utterance, but also to change its emerging meaning so that it maintains its appropriateness for the recipient of the moment.

This study thus investigates some specific aspects of the interaction of speaker and hearer in the construction of the turn at talk. First, particular states of mutual orientation between speaker and hearer are described and are demonstrated to be relevant to the structure of the turn. Second, the participants are shown to possess specific techniques for achieving and maintaining appropriate states of mutual orientation, and the structure and operation of these techniques is described. Third, the use of these techniques is shown to both provide organization for the bodies of the participants and produce specific phenomena in their talk.

Relevance of This Research to Other Lines of Study

The research reported here is relevant to several different lines of study in the social sciences.

First, it is perhaps most relevant to the study of human interaction.
Simmel (1950:21–22) has argued that "if society is conceived as interaction among individuals, the description of the forms of this interaction is the task of the science of society in its strictest and most essential sense." Conversation is among the most pervasive forms of human interaction. However, as Goffman (1963:13) has noted: "The exchange of words and glances between individuals in each other’s presence is a very common social arrangement, yet it is one whose distinctive communication properties are difficult to disentangle." Conversation has been studied as a form of human interaction by a number of different investigators, including Goffman, Sacks and his colleagues, and Duncan (see, for example, Duncan 1974a). The organization of gaze in interaction has also received considerable attention (see, for example, Argyle and Cook 1976; Kendon 1967). The present research examines some previously uninvestigated aspects of these phenomena.

Second, the work to be presented here is relevant to several methodological and theoretical issues in linguistics. For example, the present research investigates an aspect of communicative competence relevant to the production of language—the interaction of speaker and hearer in the construction of the turn at talk—that has been almost totally ignored in traditional linguistics. Moreover, as the work in Chapters 4 and 5 will show, these phenomena are implicated in the process of sentence construction. Methodologically, most contemporary linguists do not use actual speech as a source of data for the analysis of linguistic structure. They base this position in part on the argument that the phrasal breaks, such as restarts, found in actual speech give evidence of such defective performance that the data are useless for the study of competence (see, for example, Chomsky 1965:3–4). In Chapter 2 it will be found that, when the actions of the hearer are taken into consideration, such phenomena may in fact demonstrate, not only the competence of the speaker, but also his orientation toward the production of coherent, unbroken sentences. Moreover, many of the phenomena investigated here could not have been studied if actual talk were not looked at carefully. The situation is perhaps not that actual speech restricts the analyst to

---

16 A similar position is taken by ethologists in the study of nonhuman societies. For example, Cullen (1972:101) states that "all social life in animals depends on the coordination of interactions between them."

17 Some analysis of the assumptions a speaker makes about his recipient have been provided in the study of speech acts (for example, Searle 1970) and deixis (for example, Bar-Hillel 1954). However, in such studies the hearer has been analyzed merely as an addressee and the process of interaction between speaker and hearer has not been investigated.

degenerate data, but rather that, if he refuses to look at actual talk, an important range of phenomena may be inaccessible to observation and study.

Third, some of the work to be reported here is relevant to a line of research in psychology and sociology which has investigated phrasal breaks, such as restarts and pauses, in utterances (see, for example, Allen and Guy 1974; Bernstein 1962; Cook, Smith, and Lalljee 1974; Dittman 1974; Goldman-Eisler 1961, 1972; Mishler and Waxler 1970). Details of this work will be examined where relevant in Chapter 2; for the present, it is sufficient to note that analysis has focused entirely on the speaker, and that it has been assumed that his phrasal breaks are manifestations of performance difficulty. The current study thus complements this line of research by investigating interactively phenomena that have there been studied from an individual perspective.

Fourth, the research reported here is relevant in a number of different ways to the study of human communication. Cherry (1971:12) has stated that "conversation...is the fundamental unit of human communication." Though types of interaction in which no words are exchanged are just as fundamental, conversation is certainly among the most basic forms of human communication. Analysis of the procedures through which conversation is organized thus contributes to our understanding of how human beings communicate with each other.

In addition to its importance in its own right, the analysis of conversation is also relevant to a number of theoretical issues in communications research.

First, many communications researchers have assumed that a unit smaller than the exchange of turns cannot be investigated as a communications process. For example, Coulthard and Ashby (1975:140) state that "the basic unit of all verbal interaction is the exchange. An exchange consists minimally of two successive utterances: one speaker says something and a second says something in return. Anything less is not interactive." Similarly, Rogers and Farace (1975:226) argue that "the smallest unit of relational analysis is a paired exchange of two messages," where message is defined as "each verbal interaction by participants in dialogue."

18 Such a position has been consistently taken by Goffman, who conceptualizes conversation as but one type of focused interaction and assigns equal theoretical importance to unfocused interaction (see, for example, Goffman 1963). Similarly, though the work of Sacks and his colleagues has been directed specifically to conversation, they state explicitly that "this is not because of a special interest in language, or any theoretical primacy we accord conversation [Sacks 1975:291]."
Second, the turn has been employed to locate relevant units in many category systems constructed to study interpersonal communication (some examples are provided by the statements quoted in the last paragraph). However, in such studies the structure of the turn itself has remained unanalyzed, with the result that analytic units are being specified in terms of a structure whose own properties are unknown (on this issue, see Sacks 1963; Sacks _et al._ 1974:701–702).

Third, but related to the point just made, a consistent problem in the study of interpersonal communication has been the location of appropriate units for analysis. In general, the objects that participants within interaction in fact construct, such as actual utterances, have not been made the primary subject of analysis. Rather, these objects have been transformed into other objects through the use of a category system, such as the ones proposed by Bales (1950), Rogers and Farace (1975), Sluzki and Beavin (1965), or Soskin and John (1963). Analysis has then focused on relationships between these categories rather than on the actual phenomena. In contrast, this study focuses on the objects actually being constructed within the interaction, such as specific sentences.

Fourth, Krippendorff (1969a) has distinguished three different analytic models for the study of communications processes: an association model, a discourse model, and a communications model. Each of these models makes different assumptions about the phenomena being studied and requires data with a different structure (the types of data required for different types of communications analysis are discussed more fully in Krippendorff 1969b). Communications models are more powerful than discourse models, which in turn are more powerful than association models. Conversation provides data of the type required by communications models, specifically a detailed protocol of ordered exchanges through time. This study provides some analysis of how the messages being exchanged by communicators are both changed by and manifestations of the constraints organizing their communication.

---

Previous Research on the Turn and Its Constituents

Both the structure of the turn and the structure of recognizable units in the stream of speech have been examined by investigators in a number of different fields.

THE TURN AND TURN-TAKING

Despite the abstract simplicity of the notions of turn and turn-taking, and the ease with which such phenomena can be recognized in conversation, providing a precise description of the turn is a difficult and elusive task. A review of attempts to describe its structure will not only provide a more accurate definition of the turn, but will also summarize most of the research on the turn relevant to this study.

In that the description of the turn is as much an empirical as a theoretical issue, in order to evaluate various proposals about its structure, it might be useful to examine them with respect to actual data. For this reason a transcript of a fragment of actual conversation has been included in what follows. The complete transcription system can be found on pages 46–53; for present purposes it is sufficient to note that numbers in parentheses mark periods of silence to the nearest tenth of a second and that a left bracket joining utterances on different lines means that these pieces of talk are being produced simultaneously. Line 1 is being shouted to someone who is driving down the street.

---

(1)20

1. **Marsh:** BYE BYE ENJOY YOUR BROCCOLI PIE:::
2. 
3. **Diane:** Broccoli pie:::
4. 
5. **Marsh:** She’s going to her sister’s house.
6. 
7. **Marsh:** (She thought–) She just couldn’t wait to get over there and get rid of this hassle right? And then she heard she was having broccoli pie and she was really ticked off she didn’t want to go,

---

20 As will be seen later in the chapter, dealing with transcribed material makes strong demands on the reader. In order to present the transcribed data extracts in as clear and uncluttered a fashion as possible, on occasion the sources of each extract (i.e., the setting and tape number it is drawn from) appear, not in the text itself, but rather in a separate appendix, where they are listed according to chapter and example number.
It can be observed that the talk in this fragment does proceed through a sequence of turns. The two parties alternate in their production of talk, and while one is speaking the other is generally silent. Nonetheless, the delineation of the unit being exchanged—the turn—poses problems. Are Lines 5 and 7, in which the same party speaks after a period of silence, different parts of the same turn or two different turns? The same situation occurs in Lines 18 and 20, but there the sentence begun in Line 18 is not completed until Line 20. Are these cases different or the same? Is the silence in Line 4 part of any particular turn and if so which one? Is this silence the same type of object as the silence in Line 19? Line 14 occurs simultaneously with the end of Line 13. Whose turn is in progress at that point? All of Line 26 is produced simultaneously with part of Line 25. Does Line 26 constitute a turn?

Though the unit being examined has not always been called a turn, questions such as these have occupied the attention of linguists, communications researchers, and anthropologists, as well as researchers explicitly investigating conversation. Thus, Harris (1951:14) defines the utterance as "a stretch of talk, by one person before and after which there is silence on the part of the person." According to such a definition, Lines 18 and 20, as well as Lines 5 and 7–11, would be different units. In contrast, Bernstein (1962:38), by defining an utterance to extend "from the time subject commenced to talk until he finished," would group each of these pairs of lines into a single unit.

Taking a slightly different approach, some researchers have attempted to specify the boundaries of the turn in terms of talk on the part of the other party rather than silence on the part of the speaker. Thus Fries (1952, cited in Jaffe and Feldstein 1970:10) defines the utterance as "all the speech of one participant until the other begins to speak." However, this definition runs into problems when simultaneous talk is considered. According to the definition, Dianne's utterance in Line 15 ends before she has finished pronouncing her sentence. Norwine and Murphy's definition of "talk-sprint" (1938:281, cited in Jaffe and Feldstein 1970:12) encounters similar problems.

Jaffe and Feldstein (1970:19) avoid the conceptual ambiguities of their predecessors and produce a set of rules and categories so clear that it enables a computer to code some turn-relevant features of audio records of conversation without human intervention. Their approach is to ignore the content of what is said, examining the process of exchanging turns purely in terms of the sequence of sounds and silence of the different participants. Thus, their definition of possession of the floor marks its boundaries in terms of both speech by the next speaker and silence by the previous speaker:

The speaker who utters the first unilateral sound both initiates the conversation and gains possession of the floor. Having gained possession, a speaker maintains it until the first unilateral sound by another speaker, at which time the latter gains possession of the floor. The conversation terminates at its last sound [Jaffe and Feldstein 1970:19].

The very success of their project raises the question of whether constructing an internally consistent set of categories capable of unambiguously coding any relevant data presented to it is in fact what is at issue in defining the phenomena being investigated. Jaffe and Feldstein themselves admit that, rather than revealing the order in terms of which the data are structured, their category system sometimes—for example, in dealing with simultaneous speech—imposes order on the data by fiat.

[Other patterns] especially those involving simultaneous speech, are so complex that some rule is called for to bring order out of the chaos. The "speaker switching rule" 

21 Frake (1972:91) proposes a similar definition: "The constituents of exchanges are utterances: stretches of continuous speech by one person."
used in defining possession of the floor... resolves, by fiat, all these complex
patterns that defy classification [1970:114].

The precision of their categories thus obscures, rather than clarifies, the
phenomena being investigated through use of those categories. Simultaneous
speech has been approached as a phenomenon worthy of study in its own right by other investigators (see, for example, Jefferson 1973),
and they have found it to be not chaotic but, rather, precisely ordered.

Similar problems arise with the way Jaffe and Feldstein classify silence in
conversation. Silence between the talk of different parties is assigned to
the turn of the party who was speaking before the silence (Jaffe and
Feldstein 1970:19). However, as Sacks and his colleagues (Sacks et al.
1974:715) have pointed out, silence after a question is regularly heard as
being part of the next speaker's—the answerer's—turn. (Consider,
for example, the silence after a teacher asks a student a question.)

It is conceivable that the problems with Jaffe and Feldstein's system
are mere weaknesses, which could eventually be eliminated by successively
refining their definitions. However, this does not appear to be the
case. Closer study reveals that any category system that unambiguously
divides a stretch of observed conversation into a single set of distinct
objects will suffer similar problems.

Consider the categorization of the silence that occurs in the following
fragment:

(2) JOHN: Well I, I took this course.
   (0.5)
ANN: How to quit?
   [ ]
JOHN: which I really recommend.

There is general agreement among investigators that silence should be
classified differently according to whether it occurs within the turn of
a single speaker or between the turns of two different speakers. (See,
for example, Goffman 1975:10; Sacks et al. 1974:715. Even Jaffe and
Feldstein 1970:19, who did not include the content of speech in their
analysis, found it necessary to distinguish different kinds of silence in
these terms.) For convenience, a within-turn silence is frequently referred
to as a 'pause,' whereas a between-turn silence is called a 'gap.'

When Ann begins to talk, the silence in this fragment is placed between
the turns of two different speakers. It thus constitutes a gap rather than
a pause. However, John's talk a moment later continues the production
of the unit in progress before the silence began. The silence is now
placed within the ongoing talk of a single speaker. As such it is a pause
rather than a gap. Thus, the same silence yields alternative classifications
at different moments in time and from the perspective of different
definitions. This is not to say that either the silence or the rules for produc-
ing it are ambiguous. The types of objects—pauses and gaps—construc-
ted by the alternative structural descriptions remain conceptually
distinct. Further, at the point where Ann begins to talk, the data provide
no evidence to support the classification of the silence as a pause rather
than a gap. Though John subsequently demonstrates that he has not
finished talking and that the silence should therefore be categorized as
a pause, this does not change the reading of the situation available at
the time Ann began to act.

In short, no single classification of this silence is available to the
analyst, who, instead, must deal with it as an event emerging through
and thus capable of ongoing transformation. Much the same point
can be made with respect to the definition of the turn. When Ann begins
to talk, John may be seen as having constructed a complete turn. (Ann's
action of beginning her talk where she does provide some evidence that
participants within the conversation itself see the turn as having been
completed.) However, when his later talk is produced, his earlier talk
comes but the beginning of the turn eventually constructed. (Note that
the talk in the later unit is a subordinate element of the earlier unit and
thus cannot be seen as the beginning of a new unit.) At the time Ann
begins to talk, her turn is positioned as the next turn after John's. How-
ever, when John resumes talking, Ann's talk becomes placed in an "in-
terruptive" position, beginning not after but in the middle of another
party's talk.

This example provides some insight into why obtaining an accurate
and analytically relevant definition of the turn has proved so elusive.

First, almost all of the definitions considered have been concerned
with the problem of accurately defining the boundaries of the turn. How-
ever, it appears that in actual conversation the boundaries of the turn
are mutable. Different boundaries can be specified for the same unit at

---

22 For clarity, the issue here has been oversimplified. In fact, it might be argued that,
when John produces his second piece of talk, the participants are proposing competing
definitions of what is occurring. As Jefferson (for example, 1973) has demonstrated,
participants have available to them techniques for negotiating such issues.

23 Bloomfield's distinction (1946:170) between "included position," "a linguistic form
that occurs as part of a larger form," and "absolute position," a linguistic form "not
included in any larger (complex) linguistic form," is relevant here. John's second piece
of talk is included position with respect to his first and thus cannot be seen as the
beginning of a new sentence.
different points in the sequence. Even the issue of whether or not some turn follows another may have different answers at different points in time. Thus, a definition of the turn as a static unit with fixed boundaries does not accurately describe its structure; rather, the turn has to be conceptualized as a time-bound process.

Second, some of the data considered (for example, Ann’s beginning to talk where she does) suggest that the location of turn boundaries is not simply a problem for the analyst but one of the issues the participants face in arranging the exchange of turns. If this is correct, then the delineation of the turn is not properly an analytic tool for the study of conversation, but rather part of the phenomena being investigated and as such should be approached empirically (for more complete discussion of this issue, see Sacks et al. 1974:728–729; Schegloff and Sacks 1973:290).

Third, insofar as the boundaries of the turn mark points of speaker change, an accurate definition of the turn is not independent of a specification of the process through which turns are exchanged. It thus does not seem possible to first define the turn and then work out how it is to be exchanged. Rather, intrinsic structural elements of the unit being exchanged—its boundaries—seem implicated in the process of exchange itself.

The organization of turn-taking in conversation has been most extensively investigated by Sacks, Schegloff, and Jefferson (1974). The turn-taking system they describe provides a way to deal with the problematic aspects of the turn noted earlier and to specify its structure more adequately. Because this work constitutes the point of departure for the present study, it will be examined in some detail.

The system that Sacks, Schegloff, and Jefferson describe consists of two components and rules that operate on those components. A first component describes the type of units that can be utilized to construct a turn. A key feature of such turn-constructed units is that they “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 [Sacks et al. 1974:702].” Many different types of speech units—from single words to sentences—have this feature. The property of recognizable completion has several consequences. First, it specifies where in the turn transition to a new turn can occur. Second, it specifies the limits of the speaker’s current right to talk. Initially, a speaker is entitled to one such unit; at the completion of that unit, a place occurs where speaker transition becomes relevant. A second component, which allocates next turn, includes two groups of procedures: In one group, current speaker selects next speaker; in the other, next speaker self-selects.

The system also contains Rule-set 1a–c and Rule 2. The former, operating at an initial transition-relevance place, provides for three possibilities: (a) that if a “current speaker selects next” allocation technique is used, then transfer to the party so selected occurs at this place; (b) that if such an allocation technique has not been used, then self-selection is permitted, but not required, at this place; and (c) that if another does not self-select, then current speaker may, but need not, continue. Rule 2 provides that, in those cases where current speaker continues into a new turn-constructed unit, the rule set reapsplies at the next transition-relevance place and others that follow it until transfer to a new speaker occurs.

In specifying how turns are exchanged, these rules also describe significant aspects of the structure of the turn itself. For example, they avoid the problems of approaches that conceptualize the turn as a static structure by explicitly providing (for example, through Rule 2) discrete but mutable boundaries. These rules also lead to alternative classifications of silence, as well as the possibility of one type of silence being transformed into a different type. For example, a gap can be transformed into a pause if the silence is ended by further talk by the same speaker (Sacks et al. 1974:715, footnote 16).

On a more general level, both turn-taking and the turn itself can be characterized as being “locally managed, party-administered, and interactionally controlled” [Sacks et al. 1974:727].” Turn-taking is locally managed because the system deals with single transitions at a time in a comprehensive, exclusive, and serial fashion. It is party administered because control over its operations and products is vested in the participants to the conversation themselves (p. 726). Finally, and of particular relevance to the present work, by virtue of the options it gives both speaker and hearer, this system provides for the interactive construction of the turn:

A speaker can talk in such a way as to permit projection of possible completion to be made from his talk, from its start, allowing others to use its transition places to start talk, to pass up talk, to affect directions of talk etc.; and . . . their starting to talk, if properly placed, can determine where he ought to stop talk. That is, the turn as a unit is interactively determined [pp. 726–727].

Such a view of turn-taking stands in contrast to many other approaches (for example, Taylor 1970) which have sought structure in conversation (or in the groups conversing) by trying to find repetitive multiturn sequences.
The structure of the turn-taking system also provides for the interactive organization of a number of more specific types of phenomena in particular types of turns. For example, stories routinely contain many sentences before they come to their completion. However, the turn-taking system only allocates one turn-constructional unit (of which the sentence is a particular type) to the speaker at a time. The systematic production of stories without interruption is possible only if Rule 1b, granting others the right to begin talk at each transition-relevance place, can be suspended until the end of the story. Such a suspension requires the agreement of the hearer since it is he who would invoke Rule 1b. This dilemma shapes the production of stories in conversation into a particular format. First, the speaker produces a single-unit turn containing an offer to produce a multisentence turn (a turn of this type is frequently referred to as a "story preface"). The hearer then responds with an acceptance (or rejection) of the offer and only then does the speaker proceed to construct his multisentence turn. The preface routinely provides information enabling the hearer to recognize when the story has been completed so that the suspension of Rule 1b can be lifted at the appropriate moment. The particular structure stories take in conversation is thus organized in part by the orientation of participants to the features of the turn-taking system.

The features of the turn-taking system can also provide for the systematic production of a particular class of restarts. Consider the following:

KEN: You wanna hear muhn—eh my sister told me a story last night.

By producing a correction here, the speaker is able to begin a new unit without overrunning the completion point of his initial unit (for more detailed analysis of this phenomenon see Sacks 1974:342).

In turns that contain more than a single turn-constructional component, the distribution of components within the turn is frequently organized by the properties of the turn-taking system. Many adjacent turns in conversation take the form of particular types of utterance-pairs, for example, question-answer, greeting-greeting, accusation—denial, complaint—rejection. Despite differences in particular pair types, all such pairs have many organizational features in common (for example, the first element in the pair sets constraints on what can be done in the turn following it) and therefore can be analyzed as a single class. For convenience, the members of this class are referred to as "adjacency-pairs." 2

The interactive structure of stories in conversation receives extensive analysis in Sacks's unpublished lectures of spring 1970 and fall 1971. The use of story prefaces to provide for the production of multisentence turns is analyzed explicitly in the lecture of April 9, 1970. A published synopsis of some of this work, including the points discussed here, can be found in Sacks (1974).

(Schegloff and Sacks 1973). The first element in a pair is called a "first pair part" and the second, a "second pair part." 27 A turn may contain many components in addition to a first or second pair part. 28 However, in such multicomponent turns, the first pair part will be placed in a particular position—at the end of the turn—since it invokes Rule 1a. Similarly, if a turn contains a second pair part, it will be placed at the beginning of the turn. A speaker can thus employ a first pair part—for example, a tag question—to specifically mark that his turn has come to completion and that someone else now has the obligation to talk.

The structure of turn-taking is thus implicated in the organization of many different types of conversational phenomena from corrections (Jefferson 1974a) to stories (Sacks 1974) and even, as Jefferson (1979) has shown, the syllable-by-syllable production of laughter. Most relevant to the present study is the interactive organization turn-taking provides for the structure of the turn.

The position turn-taking occupies in conversation permits a more precise definition of conversation itself. Although not all conversation occurs in "single conversations," 29 turn-taking does seem to be central to conversational activity (Schegloff and Sacks 1973:292). Other forms of talk, such as debates, meetings, and ceremonies, can be distinguished from conversation in terms of explicit differences in the structure of their turn-taking. 30 Indeed, it appears that these speech-exchange systems represent systematic transformations of the turn-taking system for conversation (Sacks et al. 1974:729). The organization of turn-taking as analyzed by Sacks and his colleagues thus permits more accurate and precise definitions of both conversation and the turn than those provided at the beginning of this chapter. 31

Other investigators have provided different analyses of how turn-taking
might be achieved in conversation. Jaffe and Feldstein (1970:17) provide the simplest version of what is perhaps the most common hypothesis, the proposal that turn-transition is cued by a discrete signal on the part of the speaker:

An explanation for the switch of roles is still required, however. We look to the cues operative at the boundary between time domains. The utterance of each speaker is presumably terminated by an unambiguous “end of message” signal, at which point the direction of the one-way channel (and the transmitting and receiving roles) are simply reversed.

In essence, conversation is argued to be like short-wave radio communication, with the production of some equivalent of “over” at the end of each turn signaling to the recipient that he should now take the floor. A common candidate for such a signal is a pause.\(^2\)

The turn-taking system proposed by Duncan (1974a, see also Duncan and Fiske 1977) is essentially of this type. In this system, the speaker cues his recipient that he is about to relinquish the floor by producing a “turn-yielding signal” (Duncan 1974a:302). On the basis of empirical observation, Duncan describes six specific turn-yielding signals: rising or falling (but not sustained) pitch at the end of a phonemic clause, elongation of the final syllable of a phonemic clause, the termination of a hand movement used during the turn, a number of stereotyped expressions such as “you know” which may be accompanied by a drop in pitch, and the termination of a grammatical clause. Though the hearer may take the floor after one or more of these signals, he is not required to do so (Duncan 1974a:303). The more signals displayed at a specific moment, the greater the probability of the hearer taking the floor (p. 308). However, the speaker has the ability to neutralize any floor-yielding signals he is displaying with an “attempt-suppressing signal.” This signal consists of the speaker maintaining gesticulation of his hands during the turn-yielding signals (p. 304). Duncan’s work thus provides detailed and important analysis of many phenomena occurring at points of speaker transition.

However, because of its focus on a set of discrete signals, Duncan’s turn-taking system does not organize in terms of a small set of specific rules the range of conversational phenomena that the system of Sacks and his colleagues does. For example, it confines its analysis to the termination points of turn-constructional units and does not examine either their projectability or the ability of the speaker to delay or avoid their reaching termination. Different types of turn allocation techniques, such as adjacency pairs, are not included, and no sharp distinction is drawn between a current speaker selecting a next at a specific point (so that the selected party is located as the one who has the floor even if he is silent) and self-selection by the next speaker. Sacks’s system provides for the systematic possibility of overlap (for example, two parties may invoke Rule 1b simultaneously) at the positions where it characteristically occurs (transition points), whereas, for Duncan, such a situation means that “the turn-taking mechanism may be said to have broken down, or perhaps to have been discarded, for the duration of that state [1974a:320].” Neither is gap between turns analyzed by Duncan, who states that it did not occur in his data. It would, however, seem that insofar as gap is one of the basic sequential possibilities arising at turn-transition (a structural alternative to both smooth transition and overlap), it cannot be ignored in any general theory of turn-taking. Because of its power and generality, and because it provides a more accurate description of the detailed phenomena actually found in conversation (for example, gap and overlap), the approach to turn-taking of Sacks and his colleagues will be followed in this study.

In other work Duncan (1974b) provides detailed and interesting analysis of some processes of interaction between speaker and hearer occurring within the turn and notes how these processes might segment the turn into subordinate units. Though the particular phenomena examined (auditor back channel behaviors and speaker cues that elicit such behaviors) will not be investigated in the current study, their importance to the analysis of the interactive organization of the turn is recognized.

UTTERANCE UNITS

In addition to research on turn-taking as a phenomenon in its own right, some of the phenomena that occur within the turn have received extensive attention from investigators in a number of disciplines. The unit that has perhaps been most studied is that which has come to be called the phonemic clause: “a phonologically marked macrosegment which, according to Trager and Smith, contains one and only one primary stress and ends in one of the terminal junctures /i/ or /#/ [Boomer 1965:150].” This unit has been important not only in the analysis of the natural units into which the stream of speech, the utterance, is divided, but also in the investigation of intonation, kinesics, and the psychological study of speech encoding. Though the phonemic clause fell into some disrepute when Chomsky’s positions initially gained ascendency in linguistics, a
closely related unit, the “breath-group,” was subsequently reintroduced into linguistics within the framework of transformational grammar by Lieberman (1967). Lieberman’s work has not been generally accepted (for a critique, see Gunter 1976), and recent work by Goffman (1981) indicates that the structure of the utterance, including its intonation, is far more complex than the work done on the phonemic clause would indicate. Nevertheless, the unit has been quite important to a number of very diverse approaches to the study of a range of phenomena occurring within the turn. Research on its structure will therefore be examined in some detail.

Analysis of the phonemic clause stems from Pike’s (1945) work on intonation. Pike distinguishes two different patterns of intonation that can terminate units. Rising intonation indicates “uncertainty or finality” and is found “in hesitation and after almost all questions [Pike 1945:32].” Falling intonation marks “finality” and “occurs most often at the end of statements [Pike 1945:33].” These terminal contours thus divide utterances into two different classes, roughly corresponding to statements and questions. Building on Pike’s work, Trager and Smith (1951) distinguished three terminal junctures characterized by falling [\#], rising [/], and sustained [/] pitch. These are of course the three terminal junctures that, with the requirement that there be one and only one primary stress, define the phonemic clause.

In introducing the “breath-group” within the framework of transformational grammar, Lieberman argued that, if extra articulatory effort is not expended, both pitch and amplitude naturally fall at the end of a unit of talk. A unit with such a terminal contour is called an unmarked breath-group. It stands in contrast to the marked breath group in which the tension of the muscles in the larynx is increased during the last 150–200 msec. of phonation with the effect that the terminal contour does not fall.

Although Lieberman’s breath-group is not identical to the phonemic clause, the two units have much in common. First, both lines of research are in agreement that it is possible to clearly demarcate comparatively large units in the stream of speech. Second, in both, the intonation contour at the end of these units, roughly the final 150–200 msec., is found to be particularly important. Third, differences in the ending intonation contour are categorized in approximately the same fashion. A primary distinction is made by all investigators between falling and nonfalling intonation, with some investigators further subdividing nonfalling into sustained and rising. Fourth, despite very different theoretical points of departure, investigators in both traditions agree that falling intonation at the end of a unit marks finality and is found at the termination of declarative statements (see, for example, Pike 1945:33 and Lieberman 1967:38–39), whereas nonfalling intonation either marks a question or indicates that the utterance being produced has not yet come to completion (see for example, Pike 1945:32 and Lieberman 1967:60, 168). The study of such phenomena is quite relevant to the analysis of turn-constructional units. Indeed, Duncan’s work utilizes the phonemic clause explicitly (1974a:301), and the first of his turn-yielding cues—a phonemic clause ending on either raising or falling intonation (1974a:303)—is based directly on the work of Trager and Smith. However, the work on the phonemic clause is not sufficient to provide an adequate characterization of turn-constructional units; for example, it fails to take into account the projectability of such units. Nevertheless, this work constitutes one of the major studies of the natural units constructing the turn.

One important reason for paying close attention to the phonemic clause is that it has been used as an analytic resource in disciplines other than linguistics; it has been found to organize nonvocal as well as vocal phenomena within the turn and to be relevant to the study of speech encoding. The study of kinesics is based explicitly on the methods of structural linguistics (Birdwhistell 1973:97). Schefflen (1964:320) reports unpublished work of Birdwhistell demonstrating that the junctures marking the phonemic clause are regularly accompanied by similar movements of the body. In brief, “if pitch is raised, the eyelids, head, or hand will be elevated slightly. When pitch is lowered, such bodily part is lowered [Schefflen 1974:20].”

The relevance of the phonemic clause to the psychological study of speech encoding was investigated by Boomer (1965), who found that pauses in speech most frequently occur after the first word of a phonemic clause. Boomer argued that this provided evidence that speech encoding is organized in terms of the phonemic clause rather than proceeding word by word as some earlier studies (for example, MacKay and Osgood 1959) had implied. This work led to a second line of investigation relating speech to body movement through the phonemic clause. Building on Boomer’s work, Dittman (1974:174) found that body movement, as well as pauses in speech, occurred near the beginning of the phonemic clause (see also Dittman and Llewellyn 1969). In addition, the phonemic clause was found to organize the actions of the hearer as well as those of the speaker. Dittman and Llewellyn (1967:342) report that hearer’s listening

\( \text{Rising intonation is not, however, a definitive question marker since, on the one hand, it can occur in the absence of a question (for example, to mark nontermination), and, on the other, questions constructed with special particles, such as wh-words, are terminated with falling intonation (on this issue, see Lieberman 1967:132–133). For further problems with the notion of a "question" see Schegloff (1979).} \)
responses occur at the boundaries of phonemic clauses rather than within them. Such a finding is obviously relevant to the description of the turn at talk, as it provides an approach for specifying the distribution of one party's talk within the turn of another. The structure of the phonemic clause was also used to differentiate two different types of pauses: juncture pauses, which occur at its boundaries, and hesitation pauses, which occur within the clause (Boomer 1965:151, 153–154).

Work in both kinesics and psychology thus provides some demonstration that a number of different aspects of talk, including both vocal and nonvocal phenomena, may be organized in terms of a single unit, the phonemic clause or breath-group. Similar findings have been made with respect to units on other levels of organization. Condon and his associates (for example, Condon and Ogston 1966; Condon and Ogston 1967; Condon and Sander 1974) have shown that the boundaries of body movements of both speaker and hearer coincide with syllable and other boundaries in the stream of the speaker's speech. Condon and Sander (1974) even found that the movements of 1-day-old infants were precisely synchronized with the articulatory segments of human speech (whether English or Chinese, live or taped) but not with disconnected vowel or tapping sounds. The stream of speech thus seems to provide a (perhaps innately recognized) reference signal capable of synchronizing the behavior of separate participants. (An analogy that comes readily to mind is the music that trapeze artists use to coordinate their separate actions. However, in conversation, the signal used to synchronize the action of the participants, the stream of speech, is itself a product of their coordinated action, much as if the music in the circus was not a preformulated melody but rather an emergent product of the coordinated actions of the performers and simultaneously a resource employed to achieve that very coordination.) This work provides a strong demonstration that language is not simply a mode of expression for the speaker but rather constitutes a form of social organization, implicated in the coordination of the behavior of the different parties present.

Condon and Ogston (1967:227–229) note that speech and body movement become more independent in sequences larger than the word. The method they use for finding a relationship between speech and body movement—congruent boundaries for these different types of action—must therefore be used with caution when analyzing units as large as the phonemic clause. For example, Lindenfeld (1971) has sought to determine just how much relationship exists between syntactic units and units of body movement. She argued (1971:228) that body movements whose boundaries coincided with syntactic boundaries were related to speech whereas body movements whose boundaries fell in the middle of syntactic units were not. However, when language and body movement are considered with reference to the process of turn-taking, an alternative possibility emerges. Specifically, in order to indicate that though a possible turn-transition place is being marked syntactically the floor is not being yielded, the speaker might position his body movement so that it bridges a syntactic boundary, beginning shortly before the termination of one turn-constructional unit but not ending until a new unit is underway. (In such a case the body movement would constitute what Duncan [1974a:304] has analyzed as an "attempt-suppressing signal.") From this perspective, a close relationship between kinesics and syntax would be demonstrated precisely in the lack of congruence between syntactic and kinesic boundaries. Some of Lindenfeld's own examples—including the following (1971:231)—are consistent with this line of analysis:

There was nobody I could talk to and no . . . no . . . etc.

I didn't go for [that . . . And uh] every . . . one, etc.

In both of these examples, the speaker begins his body movement just before the next transition point of his turn and continues the movement until a new turn-constructional unit has been begun. Such positioning is quite consistent with the argument that the speaker is placing his body movement so as to indicate that he is not prepared to yield the floor at the syntactic boundary in his utterance marking the termination of a turn-constructional unit.

The analysis of the natural units into which the stream of speech is divided thus supports Goffman's conceptualization of talk (for example, the definitions of conversation cited at the beginning of this chapter) as an interactionally sustained form of social organization, achieved through the coordinated action of multiple participants and including within its scope nonvocal, as well as vocal, phenomena.

Gaze

The aspect of nonvocal behavior to be examined most intensively in this work is gaze. The glances of individuals toward other individuals, and especially their mutual gaze upon each other, has in fact been the subject of considerable study in the social sciences. Simmel (1969:358)
argues that "the totality of social relations of human beings, their self-assertion and self-abnegation, their intimacies and estrangements, would be changed in unpredictable ways if there occurred no glance of eye to eye." Of special importance to the present study is the fact that gaze is not simply a means of obtaining information, the receiving end of a communications system, but is itself a social act (see, for example, Simmel 1969:358–359 and Goffman 1963:92). Within conversation, the gaze of the participants toward each other is constrained by the social character of gaze and this constraint, rather than purely informational issues, provides for its organization and meaningfulness within the turn. Thus, the gaze of a speaker toward another party can constitute a signal that the speaker's utterance is being addressed to that party. Similarly, the gaze of another party toward the speaker can constitute a display of hearership. Such social attributes of gaze provide for its ordered distribution within the turn. The structure of this distribution will be one of the main subjects investigated in Chapter 2.

The movement of gaze within conversation makes relevant some consideration of how participants arrange themselves for conversation. Schefflen (1964:326–327) notes two basic patterns: side-by-side or face-to-face, this latter being referred to as a vis-à-vis arrangement. He argues that these different arrangements are typical of different kinds of activities; the vis-à-vis provides for interaction between the participants whereas side-by-side arrangement involves mutual orientation toward some third party or object. In conversations with more than two participants, both arrangements are typically found—for example, two side-by-side listeners vis-à-vis a speaker. Participants sometimes orient different parts of their bodies in different directions so that the same party can be in vis-à-vis arrangements with two different others. The exact orientation of participants toward each other within a vis-à-vis requires more precise specification. Sommer (1959:250–251) found that people who had a choice preferred to seat themselves corner-to-corner rather than face-to-face. Ekman and Friesen (1974:276–277), reporting much the same preference, note that such a seating arrangement is implicated in the organization of gaze, since it makes gazing at the other a marked act. They also note other aspects of the arrangement of the participants that are relevant to the organization of the conversation. For example, the order in which a speaker generally addresses different recipients may be constrained by the details of their seating arrangement. However, although phenomena related to arrangement of participants are important subjects for further research, they are beyond the scope of the present study.48

Kendon (1967) has provided the most extensive analysis of the function of gaze within conversation. He reports a particular distribution of gaze over the course of an utterance (a term he uses in roughly the sense of turn at talk). A speaker looks away at the beginning of his utterance but gazes steadily toward his addressee as the utterance approaches termination, whereas a hearer at this point looks away from the speaker.49

---

45 That is, positions such as a and A in the following diagram were preferred over positions such as b and h:

```
E    D    C    B
F    G    H    A
```

46 It may, however, be reported that the data are generally consistent with the findings of Sommer and of Ekman and Friesen, but that very frequently the physical structures available for seating made achievement of the preferred arrangement difficult or impossible. For example, most picnic tables have benches along the side but do not have chairs at the end. Thus, when participants were seated at picnic tables, only face-to-face or side-by-side positions were available to them, though they could—and did—modify this somewhat by turning their bodies in appropriate directions. In a dyadic conversation that was not constrained in such a fashion (the participants were seated in individual lawn chairs), the participants arranged themselves in just the positions described by Ekman and Friesen.

47 Analysts investigating gaze from an individual, rather than interactive, perspective have found that, after being asked a question, a subject turns his head to the side in characteristic directions (for example, left versus right) depending on the content of the question (for example, whether it deals with verbal or mathematical material). The argument here is that lateral orientation is controlled by frontal centers in each hemisphere of the brain and that "when the effects of the two centers are equally balanced, attention is directed straight ahead [Kinsbourne 1972:539]." However, the brain is asymmetrical with respect to certain cognitive functions, with language processes occurring predominantly in the left hemisphere, in contrast to spatial and temporal processes, which are localized in the right hemisphere. It is proposed that when a person engages in processes requiring the use of a specific hemisphere, for example, a verbal task, "the verbal activation overflows into the left-sided orientation center, driving attentional balance off center and to the right [Kinsbourne 1972:539]." In such experiments, the person asking the stimulus question is seated behind the subject. Gurt (1975) investigated what happened when subject and experimenter were seated face to face. She found that the same subjects who would
Thus, when turn-transition occurs, the new speaker is gazing away from his recipient, as is expected of a speaker near the beginning of his utterance. The sequencing of gaze at turn-beginning studied in Chapter 2 of this study is consistent with the pattern described by Kendon and supports his findings.

Kendon also finds (1967:26) that the hearer gazes at the speaker more than the speaker gazes at the hearer. The pattern of gazes is also somewhat different for each position. Hearsers give speakers fairly long looks broken by comparatively brief glances away, whereas speakers alternate looks toward their recipients with looks away from them of about equal length (pp. 27,33). The looks of the speaker toward the hearer occur at the ends of phrases (p. 40). At points of hesitation, the speaker looks away from his recipient, gazing back at him when fluent speech is resumed (p. 41); for more extensive analysis of the relationship between gaze and hesitation and the possible relationship of such phenomena to underlying processes of speech production, see Beattie (1978b, 1979).

Mutual gaze between speaker and hearer is found to be quite short, in most cases lasting less than a second (p. 28).

According to Kendon (pp. 52–53), an individual's perceptual activity within interaction functions in two different but interrelated ways: as a means of monitoring and as a means of regulation and expression. These functions account in some measure for the positioning of gaze within interaction. Thus, the places where a speaker gazes at his recipient—utterance endings and phrase boundaries within the utterance—are choice points, places where the future action of the speaker is contingent on the subsequent action of his hearer. By looking at his recipient at these points, the speaker can both monitor the recipient's response and signal that a response is desired (p. 4).

Kendon also suggests (p. 60) that the characteristic gaze patterns at utterance endings may be used to signal the willingness of each party to effect turn-transition and thus help facilitate a smooth exchange of turns. Such a function for gaze in the process of turn-taking has not been supported by subsequent analysis (see, for example, Rutter et al. 1978 turn their eyes in different task-related directions when not facing another would, when facing the questioner, move "their eyes predominantly in only one direction, either right or left, regardless of problem type [1975:751]."

This supports the possibility that "an experimenter's presence before the subject affects the lateralization of underlying cerebral activities in lawful and meaningful ways [p. 752]." Gur concludes that "situational variables interact with variables related to cerebral activity in producing gaze aversions as well as in determining their direction [p. 756]." For focusing on a particular situational variable, processes of interaction between speaker and hearer implicated in the construction of the turn at talk, the present work complements this line of investigation. For a more complete summary of such work see Argyle and Cook (1976:21–23).

Data

Data for the analysis to be reported here consist of approximately 50 hours of videotape of actual conversation recorded by myself and Marjorie Goodwin in a range of natural settings. The term "actual conversation" is meant to contrast the data used in this work with, on the one hand, data consisting of reports about conversation^40 (as might, for example, be obtained by questioning people about what they do in con-
Schefflen (1964:319), arguing for the importance of studying events in context, observes that "the chance to determine experimentally the function of an element is lost if the system in which it functions is scrapped." The importance of natural data for the study of the hearer, one of the main subjects investigated in this study, has been emphasized by Kendon. For example, after reviewing some existing research on the hearer, Kendon (1974:150) states

In all these cases, however, the investigator has studied only those features of the listener’s behavior he has determined in advance. The listener is always giving a controlled performance, where what he does and when he does it has been decided upon beforehand, as part of the experimental design. We know remarkably little in a systematic way, about what it is that listeners ordinarily do, and how what they do is related to what speakers do.

Argyle (1969:22) notes that, even within the naturalistic approach, investigators have largely taken their data from psychotherapy sessions and laboratory groups, and he states that "it would be most valuable to have similar material on sequences of interaction in families, workgroups, etc."

An emphasis on the importance of natural data is not confined to analysts of human interaction. It has come to be recognized in linguistics—in large part through the work of Labov—that the study of language requires data drawn from the actual situations of everyday life. Thus, Labov (1972b:xiii) states that there is a growing realization that the basis of inter-subjective knowledge in linguistics must be found in speech-language as it is used in everyday life by members of the social order, that vehicle of communication in which they argue with their wives, joke with their friends, and deceive their enemies.

Labov’s theories about the type of data appropriate for the study of speech were a strong influence on the naturalistic approach to data collection taken in this study. The data for this study consist of conversations recorded in the following situations:

4 Schefflen (1972:432) makes a similar argument about the weaknesses of hypothetical data: “A central reason for frowning on invented data is that while it can be easily invented, it is invented only from the point at which it is relevant to the point being made, thereby eliminating a central resource members use in hearing it, i.e., its placement at some ‘here’ in a conversation, after X, in short, by eliminating its conversational context.”

4 One tape, a half-hour dinner conversation, was not recorded by me, but rather by George Kuemeyer, and I am indebted to him and his parties on the tape for permitting me to use it.
Both the types of events that could be recorded and the usefulness of the material obtained were heavily constrained by the technical requirements of the recording process. Some consideration of this process will both clarify the nature of these constraints and provide a more precise description of the data utilized in this study.

All data were recorded on one-half inch videotape (EIAJ Type-1 standard) in black and white. Although film could have provided data suitable for the analysis in this study, tape was chosen for the following reasons: To begin with, one-half inch videotape equipment is much less expensive than a 16 mm film camera and tape recorder capable of recording a film with a synchronized soundtrack, and videotape is much less expensive than film. Furthermore, a comparatively long period of time (slightly over half an hour on the equipment I used) can be recorded without interruption; equipment capable of doing this in 16 mm is both expensive and bulky. However, it should be noted that, for work of the type done in this study, film is in certain respects superior medium to tape. It provides greater resolution, thus permitting the recording of finer detail, is more permanent than tape, and is capable of being easily viewed at a great many different speeds. The decision to use black and white rather with a characteristic structure when recipients with both states of knowledge are copresent. This feature is quite sensitive to other aspects of human social organization, serving, for example, to mark in moment-to-moment talk the distinction between parties who share much of their experience in common, such as spouses, and those who do not, such as acquaintances. It might thus seem that the organization of conversation should be analyzed in terms of social attributes of the participants such as their marital status. This approach would not, however, accurately characterize the phenomena under investigation. On the one hand, differential states of knowledge can be used to invoke the relevance of a very broad range of social attributes (for example, even in a situation where spouses are present, talk by army veterans about common service experience may locate them as parties who share knowledge of events that their spouses lack—description of the participants in terms of particular attributes thus does not necessarily indicate how the parties are being classified within the conversation); on the other, a particular ordering of information states is not consistent within specific social relationships (for example, when husband tells wife what happened at the office, the spouses’ states of knowledge are not equivalent). The structure of this feature is thus independent of the particular social identities invoked by it within specific situations. Such considerations show the value of examining conversation in a broad range of situations and events (the generality and structural variety of its procedures can be more clearly investigated), but indicate that the attributes of such situations are not necessarily organizing features of the conversation occurring within them.

It should be noted that some work in sociolinguistics has followed a quite different approach. For example, Ervin-Tripp (1973:66) states that “for most sociolinguistic analyses the important social attributes of participants will be sociological attributes. These include the participants’ status in the society, in terms such as sex, age, and occupation; their roles relative to one another, such as employer and employee, a husband and his wife; and roles specific to the social situation, such as host-guest, teacher-pupil, and customer-salesgirl.”
than color was made both for reasons of cost and because the recording equipment was more reliable and versatile. What was lost by not having color cannot be assessed. Smith's work on tongue displays (Smith, Chase, and Lieblich 1974) suggests that the color difference between the tongue and lips and the rest of the face might be an important signal in interaction, one that is quite possibly relevant to the work in this study.

Because of the focus of this study on conversation, securing a high quality record of the participants' speech was a primary concern in data collection. The video camera I used (a Sony AVC-3400) had a microphone built into the camera. This microphone was not, however, adequate for my purpose. It recorded a high-pitched hum generated by other electrical equipment in the camera, and, being at camera position, it was some distance from the participants. Tests at the time I was beginning to record data showed that the main influence on sound quality, even more important than the quality of the microphone used, was the distance of the microphone from the participants. The closer the microphone, the better the sound. The best sound is obtained by actually attaching a lavaliere microphone to the speaker. Because of the quality obtained, this method is regularly used by linguists to obtain samples of speech.

Such a procedure would, however, pose serious problems for the present study. It would necessitate attaching wires to the participants, which would severely constrain their movements within the group. Those wishing to leave altogether would have to disentangle themselves from the microphone, and any new participant would have to be wired-up before he could join. Furthermore, anyone looking at another participant would have his attention directed to the recording situation. This is quite different from the issue of the participants' awareness that they were being recorded. Gazing at the other is an integral part of conversational activity, and, indeed, one of the principal phenomena investigated in this study. To obtain good sound, the lavaliere microphone would be placed quite close to the mouth, constituting an unusual, noticeable, and distracting object just at the point when gaze at the other was initiated. In view of these problems, it is not surprising that use by linguists of the lavaliere typically takes place in a special situation, the interview, where the single party wearing the microphone is confined to a restricted place and does not see anyone else so encumbered. In such circum-

stances, obtaining samples of other than formal speech styles is a difficult problem.46

Some of the liabilities of the lavaliere can be avoided by using a highly directional "shotgun" microphone which is capable of obtaining fairly good sound at some distance from the speaker. This is, in fact, the method used to obtain sound in natural situations by many documentary filmmakers. However, such a microphone would not be suitable for the present work. Precisely because it is so highly directional, its position must be constantly shifted to keep it pointing at the speaker of the moment. Further, it records the speech of the person it is being pointed at better than it records the speech of other participants. A microphone of this type would thus be both extremely intrusive and would produce a poor record of many basic conversational phenomena such as overlap.47

In view of these considerations, I recorded speech by positioning a stationary microphone with the participants but not attached to them. The microphone was centrally placed and located as close to the participants as possible without being excessively intrusive. The placement that produced perhaps the best results was over the center of the group, slightly above the heads of the participants. It seems that within conversation our eyes do not glance equally in all directions but gaze predominately in front of us or downward. Thus, though a microphone might be only a foot or so from a person, if it is overhead it will remain relatively unobtrusive. A standard microphone stand with a flexible gooseneck was capable of placing the microphone in this position; however, the arrangement was much less intrusive if a stand was not placed within the group. Outdoors, the best arrangement consisted in hanging the microphone from a tree and running the cable through the branches and along the trunk of the tree. Indoors, the microphone could be hung from some fixture on the ceiling or placed on a stand positioned on a high object such as a refrigerator.

This method of obtaining sound imposed strong constraints on the type of conversation that could be recorded. Most notably, because of

---

46 At one point in my data collection, I used a lavaliere, placing it on the prospective bride at a bridal shower. She reported being quite aware of the microphone and the attendant sitting next to her said that she was reminded of the fact that what she said was being recorded every time she looked at the bride. Both of these participants felt that the presence of the lavaliere constrained their talk.

47 Use of this microphone is in fact quite congruent with the behavior of listeners in conversation. For example, it shifts attention from participant to participant as speakership changes. Indeed, I have observed that a sound-man manipulating this device relies on many of the same conversational cues examined in this study: for example, moving to a new speaker after a restart, and thus producing a clear record of the sentence begun after the restart. It is precisely the ability of this microphone to adapt to conversational structures and human participation in them that makes it a poor tool for the analysis of such phenomena.
the cables, it was difficult to record people who moved from place to place. Recording was most successful when the conversation occurred in a fixed place of limited size, such as at a table. Moreover, because of the intrusion caused by hanging cables, it was desirable to have the microphones placed before the participants arrived. Thus, many conversations were chosen to be recorded, not on the basis of participants, who were not known when the choice was made, but rather, for technical reasons, on the basis of location.

To offset the limitations of being confined to a single location, three microphones were used. Sometimes several microphones were placed in different locations so that the camera could move from one to another as circumstances demanded. More frequently, one microphone was hung in a fixed location while one or two of the others were mounted on stands so that they could be moved when needed. Although this arrangement provided some flexibility for moving from location to location, it did not make it possible to record moving groups.

The technical requirements for obtaining a picture of adequate quality also constrained the types of events that could be recorded. The most important factor governing picture quality was the amount of light available. The video camera used would produce a picture with ordinary room lighting. However, the picture was grainy, lacked some detail, and was not of sufficient quality to produce good copies. Although some early data were obtained under these conditions, whenever possible an attempt was made to provide sufficient light to produce a good picture. This could be done in a variety of ways. For some situations floodlights were directed toward the participants. However, such lights have the strong disadvantage of being quite intrusive, even when bounced off the ceiling. Some of the liabilities of floodlights can be avoided by placing higher powered bulbs (at least 200 watts) in the existing light fixtures of the setting. Although it changes the light level, this method does maintain the normal lighting arrangement of the setting and it is far less intrusive than movie lights. This arrangement works best when high overhead fixtures are available. It was used in preference to floodlights whenever possible. The least intrusive way of obtaining sufficient light consists of choosing a location where the existing lighting is adequate. Indoors, this can best be accomplished by choosing a room well illuminated by fluorescent lighting, but the best, as well as the least intrusive, lighting can be obtained by taping outdoors. For this reason, much of the data used in this study were recorded outdoors.

Two of the microphones were medium quality electric condensors (Sony ECM 21's); a test before purchase showed that higher quality microphones did not produce a noticeable improvement in sound quality under field conditions. The third, which was used far less frequently, was a dynamic microphone (an Electro-Voice 644).

Other constraints on what could be recorded were imposed by the characteristics of the television camera I used (a Sony AVC-3400). Unlike a film camera, this camera averages all the light in a scene. Therefore, participants could not be recorded in front of a bright background, such as the sky or a window, without losing detail in their features. Further, any bright point of light in the picture produces a dark, permanent burn on the tube and must be avoided. In order to obtain the best picture, the lighting has to be comparatively even. All of these considerations limited what could be successfully taped. For example, when recording outdoors, it was desirable to have the participants in the shade and in front of some background other than sky.

Yet other limitations were imposed by the characteristics of the recorded image. First, its ability to resolve detail is limited. A great deal more can be seen about a face that fills the frame than about one that occupies only a corner of it. If the actions of several participants are to be observed simultaneously, information is lost about the finer actions of each. A choice must therefore be made. For the work being discussed here, the choice of what to include within the frame was governed by the research problems for which the data were being collected. For example, at one point, I wished to investigate how speakers animate characters within stories. Therefore, whenever a story preface occurred, I filled the frame with the face of the speaker who had produced the preface. However, most of my research focused on the process of interaction between speaker and hearer. For such analysis I needed information about the simultaneous action of all participants. Therefore, all participants were included within the frame.

In order to obtain maximum detail, the camera was panned and tilted, and a 12.5–75 mm zoom lens was adjusted, as the configuration of the group changed, or its members moved, so that the group just filled the frame. On a few occasions, it was necessary to use an 8.5 mm wide-angle lens rather than the zoom in order to include all members within the group. The camera was still panned and tilted when this lens was used.

* On this issue see Goffman (1974, Chapter 13, especially Section VI).
* This work, which occupied less than an hour of tape, is not reported in this study.
* This method of taping thus does not conform to the "locked off camera" paradigm of Feld and Williams (1975:25). However, neither does it conform to their "researchable film" paradigm where "angle and focal length changes are justified by the triggering pattern of human response and intuition in relation to the structure of the event [p. 31]."
* In the present work, particular research interests, rather than the intuitions of the moment, determined what was to be included within the frame. I am in complete agreement with Feld and Williams when they state (p. 30) that "it is essential that the researcher, who has been trained in the observation of his subject, is also the filmer."
Seating arrangements posed a second technical problem for the recorded image. If someone was behind someone else he could not be seen. In many cases, this problem could be avoided or at least strongly limited by careful selection of camera position. An attempt was made to find natural seating arrangements—such as picnic tables with benches but no chairs—that would also provide an opening for the camera. When these were not available, chairs would sometimes be moved so that visual access to the group would not be blocked. The camera was mounted on a movable cart, allowing its position to be changed easily to provide the best view of all participants as circumstances changed. Only very rarely (on two occasions, a bridge game and a family dinner) were the participants arranged specifically for the camera. As people moved within the group, it frequently happened that someone was blocked, at least temporarily. Unless this occurred at the very beginning of taping, or involved a new person sitting down, it was not called to the attention of the group. If the camera could not be moved to a better position, the problems created by this situation were accepted.

Although the technical details of the recording situation can be specified with some exactness, it is difficult to calculate the consequences on the event of the fact that it was being observed. Heider (1976:80) notes that “normal, naturally occurring conversation . . . is a relatively low-energy, fragile sort of behavior, which is easily disrupted by the camera.” Problems related to the process of observation have, in fact, emerged as important theoretical and methodological issues in several different fields. In linguistics, largely through the work of Labov, it is recognized that the most important source of data for the study of linguistic structure is the vernacular, “the style in which the minimum attention is given to the monitoring of speech [Labov 1972b:208].” However, “‘any systematic observation of a speaker defines a formal context in which more than the minimum attention is paid to speech [p. 209: italics in original].’” The problem noted by Labov, far from being confined to linguistics, seems to arise in any inquiry in which precise information about natural human behavior becomes important. Thus, Eibl-Eibesfeldt (1974:21) states that hidden cameras are “a prerequisite for any documentation of natural undisturbed behavior.” It should be noted, however, that considerable disagreement exists as to the extent of the problem. In contrast to Eibl-Eibesfeldt, Feld and Williams (1974:31) believe that the process of filming does not significantly alter the behavior being filmed.

For both technical and ethical reasons, hidden cameras were not used to collect any of the data for this study. Participants always knew that they were being recorded. The problems of observation are thus relevant to the data being used for analysis in this study.

Although most discussion of this issue has focused simply on the presence of the observer (or camera), such a concept in fact includes several different types of phenomena which must be distinguished analytically. To begin with, the behavior of the observer may organize the behavior being observed. Interviews provide a particularly clear example of this process. The actions of the interviewer shape the interaction into a particular pattern with a distinctive turn-taking structure providing different types of action for the interviewer and the party or parties being interviewed. Wolfson (1976:189ff) examines some of the problems posed by the use of such structures in linguistics. Some investigators have attempted to deal with such problems by making the actions of the interviewer, as well as the parties being interviewed, part of the final published record of the event (a particularly striking example is provided by Jean Rouch’s Chronicle of a Summer). However, although such a strategy makes accessible the actions of the observer, it does nothing about the changes in the event itself wrought by the structure of his behavior.

The observer’s actions may modify the structure of the event even though the observer does not cause any changes in the behavior of the participants. For example, after the event he can rearrange his record of it, as happens, for instance, when a film is edited. Further, the technology used to record the event in the first place will inevitably modify it in a systematic fashion. Any camera position or framing of participants involves a choice from a set of alternatives and any of the alternatives not selected would have produced a different record of the event. Similarly, using a category system, such as that of Bales (1970:92), to code the event will lose much information about the event and organize the information that remains in a particular fashion which is determined as much by the structure of the category system as by the events being categorized.

Analytically distinct from the behavior of the observer is the behavior of such human act differently toward different types of others,25 and this will have consequences on their production of talk.26 The implications for an investigator wishing to sample the speech behavior of different individuals are obvious. If the investigator is the addressee of the party he is observing, as is the case with interviews, what he will in general obtain are samples of how these different individuals talk to an academic stranger—rather than samples of how they talk to each other.

25 For some discussion of the relevance of this for the conceptualization of culture, see Goedenaugh (1963:250–251).

26 For some analysis of precisely how talk will vary in terms of its intended recipient, see Schegloff (1972).
An investigator can, however, systematically observe and record the speech of different groups of people without himself being the addressee of that talk. In his early interviews in New York, Labov (1972b:89) observed that casual speech emerged when the party being interviewed began to talk to others present. In order to obtain better data about the vernacular, he therefore began to supplement formal interviews with group sessions. In these sessions, "the adolescents behaved much as usual, and most of the interaction—physical and verbal—took place between the members. As a result the effect of systematic observation was reduced to a minimum [p. 210]."

Analytically distinct from both the behavior of the observer and his status as an addressee of the participant's action are changes in the event caused by the mere fact that it is being recorded as well as by the observable presence of the recording equipment. The issue of how participants deal with observation is in fact a somewhat subtle one. Within conversation, participants never behave as if they were unobserved; it is clear that they organize their behavior in terms of the observation it will receive from their coparticipants. For example, a speaker does not simply "forget" a word; instead, he actively displays to the others present that he is searching for a word. Thus the issue is, not what participants do when they are unobserved, but whether the techniques they use to deal with observation by a camera are different from those used to deal with observation by coparticipants. This is an empirical question requiring further research. It seems quite plausible that people may avoid discussing a variety of "sensitive" topics in the presence of a tape recorder (though the Watergate tapes provide some counterevidence), just as they avoid mentioning such topics in the presence of certain types of coparticipants. It seems far less plausible that phenomena on the level being examined in this study would be changed—that, for example, restarts would act to bring the gaze of a recipient toward the speaker when the camera was present but not when it was absent—though this remains an empirical question.

It is frequently assumed and sometimes explicitly argued (for example, Wolfson 1976) that direct participant-observation is less disruptive of the phenomena being observed than recording that phenomena with a tape recorder. This does not necessarily seem to be the case. Consider the problem of investigating the gaze of the hearer. The tool that a participant-observer would use to observe the gaze of others—his own gaze—is itself a relevant event in the interaction in which he is participating. If the observer employs his gaze in an inappropriate fashion, a noticeable event will occur which may well disrupt the process being observed. As noted by Schefflen (1973:88–89), gazing at a hearer is inappropriate: "One is to look at the speaker of the moment. . . . As a consequence we rarely get to observe the behavior of listeners and we do not ordinarily see the total bodily behavior of others in conversation." The camera, though intrusive and perhaps disruptive in other ways, does not focus attention on the gaze of either party (especially if it is not pointed at one participant in particular but includes both speaker and hearer within the frame) and is not itself an oriented-to feature of the process under observation. In this particular case, use of a camera is less destructive of the process being examined than direct participant-observation would be.

In gathering data, I tried to deal with the issues raised by the recording process in a number of different ways.

First, I attempted to limit as much as possible my interaction with the people I was taping. I could, of course, have chosen instead to become a member of the group myself. However, although such an approach would have provided a better record of my actions, it would have made more serious many of the other problems that have been discussed. For example, all the different groups I taped would have had a common addressee and my own behavior would have significantly organized the behavior of others in the group. Moreover, focus toward the camera would have been greatly increased unless I fixed it at a particular angle and focal length and left it. The strategy I chose to adopt was quite similar to that employed by Labov in recording group sessions. I accounted for my lack of engagement by displaying involvement in the technical details of recording. Thus were a participant to turn to me, he would find me studying the VU meter on the microphone mixer or checking the image in the viewfinder. I also wore earphones and gazed toward the viewfinder from a slight distance and at an oblique angle rather than pressing my eye to the camera. The camera was thus not presented as an extension of my face and body directed toward the participants but, rather as an object that was itself the focus of my attention (this was, of course, made possible by the fact that I was not peering through the lens, as is the case with a film camera, but rather looking at a very small television monitor). I was thus a person present at the event but not one immediately accessible for interaction, my involvement being directed to other tasks claiming my full attention.

Second, I tried to limit and make explicit, as far as possible, the organization imposed on the event by my recording of it. Thus, once

---

44 For a more detailed analysis of changes in speech that occur when someone other than the interviewer becomes the addressee, see Labov (1972a:207–212; 1972b:89–90).
45 The techniques employed by speakers to signal "word searches" have been extensively investigated by Sacks and his students. Though most of this work is as yet unpublished, Jefferson (1974) analyzes some aspects of this process.
46 For an analysis of such avoidance in an actual speech situation, see Thomas (1958:70–71).
the camera was set up and the participants were present, I tried to record them continuously until they left the scene. My reasons for not trying to select particular events are the same as those given by Goffman (1953:3) for his use of a similar strategy.

While in the field, I tried to record happenings between persons regardless of how uninteresting and picayune these events seemed to be. The assumption was that all interaction between persons took place in accordance with certain patterns, and hence, with certain exceptions, there was no prima facie reason to think that one event was a better or worse expression of this pattern than any other event.

From a somewhat different perspective, Margaret Mead (1973:257) has noted that

the future usefulness of field data for different kinds of exploitation, many of them unanticipated at the time the field work was done, is a direct function of the extent to which material can be collected in large, sequential and simultaneous natural lumps on which no analytical devices of selection have operated. . . . Long verbatim texts are more valuable than many short verbatim texts, tapes which contain many other kinds of information are more valuable than several hand-recorded verbatim texts. Only materials which preserve the original spatial-temporal relationships are virtually inexhaustible as sources for new hypotheses and ways of testing old hypotheses. The more material is codified by the method of selection, as when sample scenes, standard-length anecdotes, standard interviews, standard texts, are used the more immediately useful it may be in relation to some hypothesis and the less its permanent value.

The video recorder I used could record for slightly longer than 30 minutes before tape had to be changed. Except for the time lost when tape was being changed (approximately .5–1 minute), the recorder was run continuously, sometimes for more than 6 hours. In order to maintain a consistent and explicit approach toward the selection of what participants to include in the shot, the shot was framed (with several exceptions noted earlier) to include all present participants. The practice of shooting continuously for a long period of time might also have contributed toward obtaining samples of interaction in which the behavior of the participants was influenced less by the camera than by each other. This is, however, an empirical question which requires further investigation.

Transcription

Data were transcribed according to a system for capturing the auditory details of conversation designed by Gail Jefferson (Sacks et al. 1974:731–733) and a system for recording gaze direction devised by myself. Both of these will be described shortly. All transcription was checked by at least two and frequently three transcribers. Some of the tapes used in the analysis were audio-transcribed by Gail Jefferson and I am deeply indebted to her for this.

Because the transcription system makes use of basic English orthography, it might appear to the reader that sections of transcribed data can be read in much the same way as the basic text. Such material is, however, as different from the rest of the text as the statistical tables found in many journal articles. Both comprehension and evaluation of such data require that the material be attended to in quite specific ways. I will do several things to aid the reader in this task. First, relevant transcription devices, rather than being either assumed at the outset or relegated to an appendix, will be described initially here and then reintroduced progressively as the analysis develops. It is hoped that this will enable the reader both to see in detail the relevance of particular transcription conventions for the analysis and also to learn to use them before encountering new ones. Second, the transcription of the talk will be presented in a simplified fashion. Basically, only those distinctions necessary for the analysis being developed will be included. Though this does not provide the most accurate version possible of the talk as spoken, it will make the material far more accessible. Third, I have tried throughout to keep the examples as short as possible.

We will begin by examining some of the conventions used to transcribe speech. To facilitate subsequent reference to this section, symbols and key aspects of their meaning will appear in boldface. (For ease of reference a simplified version of the transcription system is also provided at the front of the book, immediately after the contents.) The following data fragment includes some of the conventions that are most important to the present study:

1 2
↓ ↓
1–uh: one–one week ago today, actually,
↑ 3 4

Researchers who have utilized phonemic systems have found them almost useless for investigating conversational phenomena. Thus, Duncan (1974:300–301) transcribed his data in terms of segmental phonemes but found that "the segmental phonemes were the least important components of the study." The Jefferson system was constructed specifically to record phenomena in the stream of speech relevant to the organization of conversation. Thus, it not only notes such sequential phenomena as the precise location of both silence and simultaneous speech, but also records changes in duration which do not distinguish segmental phonemes in English and phenomena relevant to units larger than the sentence, such as differences in time between sentences or turns. Although this system does not capture all relevant distinctions in the stream of speech, it is the system most relevant to the issues being investigated in the present analysis. For discussion of some of the theoretical issues posed by the transcription process see Ochs (1979).
1. Introduction

A colon (1) indicates that the articulation of the sound preceding it is noticeably prolonged (note that such a “sound stretch” also occurs within the word “today”). Multiple colons would indicate even greater prolongation.

A dash (2) marks a cut-off; that is, it indicates that the sound in progress is noticeably and abruptly terminated. Very frequently, the sudden closing of the vocal cords produces a glottal stop.

Italics (3) indicate some form of emphasis which may result from increases in either pitch or amplitude. Thus, in these data, a cut-off occurs at the end of the first “one,” and the second “one” is noticeably emphasized.

Punctuation marks (4) indicate intonation. A falling contour is indicated with a period, a rising one with a question mark, and falling-rising contour (the kind of intonation that one finds, for example, after items in a list) with a comma. Thus, in these data, a full stop occurs before the end of the utterance, just after the word “today.” The fact that these symbols are being used to mark intonation and not traditional orthographic distinctions should always be kept in mind. For example, because of their characteristic intonation patterns, most wh-questions should be marked with a period. Were a question mark to be used, the rising contour thus indicated would in fact sound rather unusual.

In the following, dashes within a parentheses (5) indicate tenths of seconds within a silence. Each full second is marked with a plus sign (6).

How’s uh, (--------- + ---) Jimmy Linder.

\[ \begin{align*}
\uparrow & \uparrow \\
5 & 6
\end{align*}\]

The silence can also be indicated with numbers in parentheses—for example, “How’s uh, (1.2) Jimmy Linder”; the dashes and pluses will only be used when it is necessary to show something that happens at a specific place within the silence.

The conventions that have just been described are the ones that are the most important to the present analysis. As was stated earlier, for reasons of clarity, many details of the talk that can be captured by transcription have not been included. Thus the first fragment could be more accurately transcribed as follows:

I—uh: one—one week ago t’day. acharly,

The nuances that have been omitted are certainly important for many types of analysis that could be developed from this fragment, and they should be included in any working transcript. However, they are not necessary for the presentation of the current analysis and the material is easier to follow if they are excluded.

Several other transcription symbols will also be used on occasion, though much less frequently than those already introduced.

A bracket connecting the talk of two speakers (7) indicates the point at which overlapping talk begins:

\[
\begin{align*}
A: & \quad \text{It was s::s: so : goo: d.} \\
B: & \quad \text{I love it.}
\end{align*}\]

The same phenomenon can also be indicated with double slashes (8). In this case, the talk of the second speaker is placed at the beginning of the next line of the transcript.

\[
\begin{align*}
8 \\
A: & \quad \text{It was s::s:oll: goo: d.} \\
B: & \quad \text{I love it.}
\end{align*}\]

Double brackets before the talk of two speakers (9) indicate that they start to talk simultaneously.

\[
\begin{align*}
9 & \quad \text{Yeah. Right.} \\
B: & \quad \text{And then just covered it with \ldots}
\end{align*}\]

\[\text{Sic can-ta comma, sic du-o puncta: sic ve-ro punctum.}\]

\[\text{Sic sig-num in-ter-ro-ga-tio-nis?}\]

\[\text{Lieberman (1967:129, citing Hadding-Koch 1961-96) describes "a medieval rule for liturgical recitation from Munster which states that a fall in pitch corresponds to periods, a small rise to commas, and a large rise to interro-gatives. ..." The rule was written as shown here:}\]
A degree sign (10) is used to show that the talk so marked is spoken with noticeably lowered volume.

A less than sign (11) is used to indicate a hurried start—that is, a push into the prior space:

11
↓
The first catch <1 mean Susie—

An equal sign (12) indicates that no break occurs between two pieces of talk by either the same or different speakers:

12
↓
A: ... and then, you know, =
B: = They had the new dorms. = Right.
↑
↑
12
12

Thus, in this example, B’s talk follows A’s without any noticeable gap whatsoever. The second unit of the talk is then produces is also “latched” to the first unit without any visible break.

A series of h’s (13) marks an outbreath, unless a dot precedes the h’s (14), in which case an inbreath is indicated:

h h h h (0.4) ‘h h We just want to get ...
↑
↑
13
14

Thus, in this example, the speaker first produces an outbreath, then pauses, then produces an inbreath and finally begins to speak.

H’s within parentheses (15) indicate within-speech plosives:

grease it with (h)th va(h)sel(in|h)e
↑
↑
↑
↑
15

These plosives may be associated with phenomena such as laughter, crying, or breathlessness.

A blank within parentheses (16) indicates that the transcriber was not able to recover what was said:

She was ( )ing guys up to the ...
↑
16

Words within parentheses (17) indicate a possible hearing. Two sets of parentheses containing words (18) show that alternative readings are possible:

18 — (Ours is) a hell of a discussion.
(this is)

↑
17

The marking of multiple readings might indicate either disagreement among cotranscribers, agreement to both possibilities by cotranscribers, or double readings by a single transcriber.

Although many investigators have treated transcription as unproblematic and argued that different listeners should reach agreement on what is said in a particular passage (see, for example, MacKay and Osgood 1959:25), this does not appear to be either a realistic or an appropriate way to deal with the transcription of conversation. Not only do conversations in natural settings occur in locations that are far from ideal for either hearing or recording speech, but the speech signal itself may not be entirely unambiguous. Lieberman (1967:164–165) reports a series of experiments showing that words spoken in conversation and recorded under the very best of conditions cannot be reliably identified when heard in isolation. The regularity with which a request to repeat some item occurs in conversation provides some demonstration that accurately hearing what was said is a problem faced by participants within the conversation itself. In view of such phenomena, the goal of accurate transcription would seem better served by admitting the possibility of different readings of the same stretch of speech. Accepting this possibility, as the Jefferson transcription system does, produces a more accurate record of the speech being transcribed than either settling disputed cases by flipping a coin (a method used by Buben [1976:285] to resolve differences between coders) or forcing transcribers to agree on a single hearing.

Transcription of nonvocal phenomena will be restricted to an extremely limited set of distinctions about the participants’ gaze toward each other. This is not because these distinctions are thought to be the only ones relevant to the organization of the participants’ interaction, but rather for just the opposite reason: specifically, because of a recognition of just how much the details of body movement are implicated in the organization of talk. Their importance is demonstrated by the research on kinesics discussed earlier and also became apparent whenever data were examined closely. It was nevertheless decided that if the scope of investigation were expanded, even the limited phenomena already included would not be dealt with either adequately or within a reasonable period of time. The work of McQuown and his associates (1971) demonstrates
just how much time (well over 20 years) can be devoted to the intensive analysis of a very small strip of interaction.

Gaze will be transcribed as follows: The gaze of the speaker will be marked above the utterance and that of the recipient(s) below it. A line (19) indicates that the party being marked is gazing toward the other. The precise place where gaze reaches the other is marked with a capital X (20) tied to a specific place within the talk with a bracket. Thus, in the following, speaker (the party above the utterance) is gazing at recipient from the beginning of the talk transcribed, whereas recipient’s gaze reaches speaker after the talk has begun:

```
19
↓
↓

A: ........................................
We went down to—(0.2) When we went back ...

B: ......................................
X
↑↑↑
21 20
```

The movement bringing one’s party gaze to the other is marked with dots (21), whereas the movement withdrawing gaze is indicated with commas (22). Thus, in this example, hearer begins to move toward speaker at the end of the word “down,” whereas speaker’s gaze leaves hearer at the beginning of the word “we.” For a number of reasons, the termination of a withdrawal movement cannot always be accurately marked (for example, the movement may become part of another activity). Nonetheless, it is generally possible to capture both the beginning of this movement and the beginning and the end of the approach movement.

These are the most important transcription devices for recording non-vocal behavior that will be used in this study. However, on occasion some other phenomena will also be marked. In some multiparty situations it will be necessary to indicate who in particular is being gazed at. This will be done by putting a name above the gaze line (in such cases the name replaces the X marking gaze arrival). Thus, in the following, the speaker moves his gaze from Beth to Ann; one recipient, Beth, does not gaze at all; and another, Ann, moves her gaze from Beth to the speaker:

```
JOHN: ........................................ Ann
                 ........................................ [ ]
                                  I—uh: one—one week ago toda: y. actually.

BETH: ........................................

ANN: ........................................ Beth  John
```

In Chapter 3, it will sometimes be necessary to show a move toward another that does not culminate in full gaze. This will be done by indicating its duration with dashes (23) rather than a line and marking the point of arrival with a lower case y (24) rather than a capital X. Thus, in the following, x moves toward b but stops that movement before his gaze actually reaches b:

```
24 23
↓↓
A: ........................................ y—

Could’t stand him.

B:
```

The meaning of particular transcription symbols for gaze will be noted again when the symbols are reintroduced in subsequent chapters. It is hoped that this will make the transcription as easy to follow as possible for the reader.

What exactly is being transcribed as gaze toward the other requires further discussion. Because multiple participants were included on the screen, it was frequently impossible to distinguish individuals’ eyeballs. Thus, what is being noted is the orientation of the head toward the other rather than the detailed behavior of the eyes. I am using the term “gaze” to refer to such orientation for a number of reasons. First, in view of the great number of times that I must refer to this event, a more cumbersome locution, such as “orientation of the head toward the other,” would quickly become quite awkward and make the analysis more difficult to follow. More importantly, as will be demonstrated in some detail, the participants themselves do attend to precisely this distinction in the organization of their activity. For participants—and it is their distinctions that the analyst is concerned with—orientation of the head is one of the central components of the activity of gazing. It is of course recognized that gazing also includes many other phenomena which themselves merit extensive research.

60 Despite the general acceptance of the phrase “eye contact,” research indicates that participants do not, in fact, gaze into each other’s eyes. Both Scheflen (1974:67–68) and Exline (1974:73–74) note that gaze toward another falls not precisely at the eyes but rather in a region about the face. However, as noted by Exline (1974:74), this does not pose serious problems for the analyst (or participants): “It is my belief that the validity problem is not critical, for our observations indicate that most people turn their heads and faces slightly away from the other when they break contact. Even if one looks into a zone of regard rather than the eye itself, the other reacts as if he were engaged in eye contact.” Such a view is quite compatible with the approach to gaze being taken in the present study.
The Apparent Disorderliness of Natural Speech

Natural speech is frequently considered a poor source of data for the analysis of linguistic structure (see, for example, Chomsky 1965:3–4). Specifically, sentences produced within it are regularly found to be impaired in a variety of ways. Thus, a sample of natural speech will contain not only well-formed grammatical sentences:

(1) **JOHN:** These egg rolls are very good.
(2) **CURT:** Al's a pretty damn good driver.
(3) **MARSHA:** Christ it was just gorgeous.

but also sentences characterized by phrasal breaks, false starts, long pauses, and isolated ungrammatical fragments:

Thus Chomsky (1965:58) argues that actual speech is of such "degenerate quality" that it is of limited usefulness for the study of linguistic competence. A similar view of speech production has been expressed by some psychologists, for example, Martin and Strange (1968:478), who argue that natural speech is so defective "that it is hazardous to guess at the exact constituent structure of any given utterance." It is also frequently argued that participants themselves do not perceive the restarts, pauses, and fragments in their talk (see, for example, Lyons 1972:58 and Mahl 1959:114).