TRANSCRIBING CONVERSATIONAL EXCHANGES†

By

John J. Gumperz
Department of Anthropology
University of California, Berkeley

and

Norine Berenz
Department of Linguistics
University of California, Berkeley

Cognitive Science Program

Institute of Cognitive Studies

University of California at Berkeley

May 1990

† To be published in Edwards, J. A. and Lampert, M. D. (eds.), *Transcription and Coding Methods for Language Research*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

The production of the Berkeley Cognitive Science Report Series is supported by a grant in cognitive science to the University of California at Berkeley from the Alfred P. Sloan Foundation.

Abstract

This paper presents a system for the transcription and preliminary analysis of audio- or video-taped recordings. This transcription system seeks to capture the contextualization cues that speakers and listeners demonstrably attend to both in producing talk and in interpreting and responding to what they perceive. Phenomena commonly discussed in conversational analysis, as well as prosodic and paralinguistic cues, are covered. The basic perspective is interpretive and discourse oriented, rather than descriptive and utterance oriented.

Transcribing Conversational Exchanges

John J. Gumperz and Norine Berenz
University of California at Berkeley

In Edwards, J. A. and Lampert, M. D. (eds.), *Talking Language: A Handbook for the Transcription and Coding of Spoken Language.*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Table of Contents

Transcribing Conversational Exchanges

John J. Gumperz and Norine Berenz
Department of Anthropology
University of California at Berkeley

1. The theoretical perspective.

This paper presents a way of transcribing and otherwise preparing for systematic analysis audio- and videotaped oral performances of all kinds, ranging from chat to formal discussion and ritual event, from within a sociolinguistic and functional perspective. We include monologues on the assumption that speakers always address themselves to real or imagined listeners, frequently relying on shifts in voice in response to actual or supposed listener reaction, as well as on other rhetorical devices used by participants in multiple party encounters.

We assume, then, that talk is basically interactive in nature, typically involving two or more participants. What we seek to account for is the ongoing, cooperative process of conversing in which all participants, both speakers and listeners, are simultaneously engaged in producing talk, monitoring what is said and exchanging verbal and nonverbal backchannel signals to indicate how a turn at speaking is developing and how it is being received. Turn-taking or speaker change, the phenomenon that has received so much attention in the work of conversational analysts (Atkinson & Heritage, 1986), is clearly one of the basic strategies on which conversational cooperation depends. But speaker change does not just happen in response to structurally determinable characteristics of an exchange. Rather, it is accomplished interactively as part of the process of conversing. Conversing in turn rests on speakers' and listeners' interpretation of verbal and nonverbal signs or contextualization conventions, that is, systems of cues that guide conversational management (Gumperz, 1982, 1990). Cues marking the transition relevance points where speaker change ordinarily occurs are a subset of contextualization conventions. Other functionally similar signs are used to negotiate such matters as topic change or change in perspective, relative emphasis or salience of information, style shifting, degree of formality and the like, or to achieve the kind of rhythmic integration that often marks smooth interaction (Gumperz, 1982).

Our transcription system attempts to set down on paper all those perceptual cues which past research and on-going analyses show participants rely on in their online processing of conversational management signs. Many of the phenomena we record – for example, pausing, timing of exchanges, rhythm, prosody and other paralinguistic signs – are also dealt with in other chapters of this volume (DuBois & Schuetze-Coburn, Ehlich). What is emphasized in our approach is the rhetorical impact these signs have in affecting the *situated interpretations* on which the conduct and outcome of the exchange depends. Transcription, when seen from this basically functional perspective, is an integral part of the analysis of inferential processes. So that, in transcribing a feature such as pausing, for example, we are less interested in *absolute* duration and more in the *interpretive evaluation* relative to other matters of timing and rhythm within the event. In rendering prosodic, intonational and paralinguistic phenomena, we similarly concentrate on just those features of pitch, tune and accent that can be shown to affect situated interpretation at the interactive or relational level as well as at the level of content. In other words, we present in this paper a transcription system which focuses on speakers' and listeners' use of verbal and nonverbal signs both to convey or understand information and to maintain what Goffman calls "conversational involvement." Before turning to the transcription system as such, we will discuss briefly the analytic procedures we employ in transforming raw tapes into analyzable transcripts and the basic interpretive categories in which analysis is grounded.

2. The interpretive level of analysis.

The present transcription system was developed in the course of ethnographic fieldwork in urban institutions of various kinds and typically involves the following procedures. After an initial period of interviewing and participant-observation where basic information on the history, social characteristics, everyday practices and goals of the institution in question are collected, working hypotheses are generated and encounter types most likely to provide the information with which to test these hypotheses are identified. Tape recordings of such encounters, varying in length from a few minutes to several hours, are then collected and compared systematically with a range of other encounters both within and outside the particular institutional context. The source data for these comparisons come in large part from previously collected and partially analyzed recordings which are stored in Disclab, a computerized archive of audio- and videotaped materials located at the University of California, Berkeley. Although our primary focus has been on institutional encounters, at the present stage of its development, the methodology is a general one, applicable to interactions of all kinds.

At the initial stage of analysis, we work with raw recordings which cover everything that transpires within a certain time span. In many cases, the encounters take the form of interactive situations that carry such folk labels as *job interview, medical consultation, parent-teacher conference, management committee meeting, classroom session* and the like. Labels can tell us a great deal about the nature of the interaction and relationships among participants, and about likely topics and expected outcomes. Yet this background information is never sufficient to account for the observable features of an interaction, nor are labels and the folk knowledge associated with them identical with the schematic information, that is, the implicit or taken-for-granted background expectations, which participants rely on to guide their own actions and to interpret what transpires at any one point in the interaction. Everyday encounters, in fact, rarely if ever adhere to pre-set agendas. Classroom sessions, for example, consist partly of informal talk in addition to the planned lesson. Similarly, in medical consultations much time may be taken up by the exchange of personal information and informal anecdotes.

The lack of correspondence between labels and interpretive practices raises serious problems for a functional approach. Even relatively simple, that is, perceptually salient or instrumentally measurable and seemingly indexical, signs such as pausing, pitch register shift and the like can be seen to have many functions. Charles Fillmore provides an anecdote concerning his use of pausing in the course of giving expert courtroom testimony. When he took a little more time than usual in answering the opposing attorney's question, the attorney, addressing himself to the court reporter, said, "Let the record show that the witness hesitated before answering." Presumably, the attorney, in accordance with the established California evidence code, interpreted the pause as unwillingness or at least reluctance to cooperate in answering the question. Fillmore reports that in actuality his delayed response was motivated only by his attempt to suppress a cough. Clearly, pauses have as many possible interpretations as possible causes. Aside from those that result from the nonlinguistic demands on the relevant physical mechanisms, pauses can be rhetorically motivated or serve as turn holders to allow for speech planning, among other things. What holds for pauses is also true of other contextualization cues. Conversational interpretation is inherently highly context dependent. We therefore need to sharpen our notion of context to deal with the problems that arise.

A main goal of contextualization analysis is to make explicit the processes by which the presuppositions that affect or channel interpretation at any one point are negotiated and to show how talk at any one point coheres with the preceding and following talk. What the analysis focuses on is the process of signaling which shapes participants' notions of what the context is (see Gumperz, 1990, for more detailed illustrations). We do not seek to assign context-free functions to any one cue in isolation, nor do we seek to be exhaustive in our representation of the characteristics of the talk. Rather, we transcribe all those signs that, on the basis of the analysis as a whole, can be shown to enter into that process. It is in this sense that we can say that transcription is an integral part of an overall process of interpretive analysis that includes both the translation of oral recordings into written symbols and the evaluation or assessment of communicative intent.

2.1. The event.

Analysis begins with the segmentation of the interaction into thematically coherent and empirically boundable portions, that is, "events," within the encounter as a whole. In the first pass through, it is often difficult to isolate longer, thematically coherent episodes. We discover such episodes through turn-by-turn scanning that focuses on two levels of analysis: content and rhythmic organization (Erickson & Schultz, 1981; Gumperz, 1982). At the level of content, events have several characteristics commonly associated with narrative organization: namely, detectable beginnings, middles and ends. Ends are of particular interest because they provide empirical information on outcomes or on what communicative ends participants intended to achieve, whether successful or not. This outcome information can then be used as evidence on which to base hypotheses concerning (a) what participants might have inferred at any particular point and (b) how such inferences were cued.

The output of this initial phase of the analysis is a segmentation of the tape session into sequences in which the more clearly bounded events alternate with the less clearly bounded or transitional ones. In the second stage of the analysis, the individual events or portions thereof are subjected to more detailed sequential examination and transcribed using the system described below. Analysis begins with clearly bounded events. Once these are analyzed, the other less clearly bounded events are also more readily understood. Comparative analyses are based mainly on such events (Gumperz & Roberts, 1990). In what follows, we present our transcription system in such a way as to roughly reflect the stages of the transcription process.

2.2. Speaking turns and informational phrases.

We use the term "sequential analysis" to refer to the detailed examination of the local moves and countermoves that constitute a speech exchange. The basic unit of sequencing is the speaking turn – that string of utterances produced by a single speaker and bounded by other participants' turns. Yet speaker change units are not the fundamental units of speech production. Phoneticians of all persuasions agree that speech production in everyday interaction is best described not in terms of lexical items or mere strings thereof but in terms of phraselike units variously called breath groups (Couper-Kuhlen, 1986), idea units (Chafe, this volume), intonational phrases or, as we, maintaining of our functional perspective, would prefer to say, informational phrases.

The best way to characterize an informational phrase is as a rhythmically bounded, prosodically defined chunk, a lexical string that falls under a single intonational contour. Prototypically, these are set off from surrounding phrasal units by pausing and constitute semantically interpretable syntactic entities. As is true of prototype phenomena in general, not all of these characteristics need be present at the same time (Kay & Coleman, 1981). In less prototypical cases, determination of phrase boundaries depends on what divisions make sense in terms of the rhythmic and thematic organization of the surrounding discourse. Even where a phrase seems syntactically incomplete, it is interpretable if we take into account the preceding and following context. Consider turn 1L in the following sequence from a relatively heated argument analyzed in previous work (Gumperz, 1990).

Ex.1

1D: this is not a-
1L: ==of *course/ {[ac] it is not a secret//}
2D: =that it *is a secret*//=
2L: =I haven't *said= it's a secret// {[ac] I didn't say it was a secret//}

The "of course" in turn 1L is transcribed as a separate phrase (shown by the slash) although the tape recording does not show a perceptible pause. It is the shift in rhythm in the second part of the line that marks the phrase boundary. The phrasing in this example has interpretive import inasmuch as it suggests that the utterance is not to be understood as "Of course, it is not a secret," but that a different reading involving ellipsis: "Of course it isn't. It is not a secret," more adequately accounts for what transpires in the encounter as a whole. Thus, an informational phrase is basically a unit of cognitive information which, at the level of speaking, can be marked by a variety of co-occurring perceptual cues.

In the first step of the transcription process, we listen to the stream of talk attempting to isolate such units, marking phrase boundaries throughout an event or portion thereof, and then go through the recording again for in-depth analysis. The phrase division is also the main criterion for determining the makeup of transcript lines. Readability is improved by breaking up participants' turns requiring multiple lines at prosodic or syntactic boundaries, rather than running the line to the margin of the page. Some researchers structure the transcript so that each line is a separate phrasal unit. Yet readability suffers when many short phrases each occupy separate lines, thereby requiring considerable space on the page for what was only of brief duration in the saying. We therefore compromise by allowing one, two or more phrases per line, depending on the length of the lexical string. Each line, of course, must end in a complete phrase as in Ex.2 below. Rendering the turn as in Ex.3 violates our principle, not only lessening readability but also making it more difficult to follow the rhythmic organization of the turn. (For a more detailed discussion of the notions of event and activity employed in our interpretive method, see Gumperz, 1990, & in press.)

Ex.2

R: now, are the details the same as when you applied?
   you s- living at the same add(ress)?

Ex.3

R: now, are the details the same as when you applied? you s- living at the same
   add(ress)?

2.3. The lexical stream.

Along with many other discourse analysts, we render most lexical material in conventional orthography. While the use of conventional orthography contributes significantly to readability, it also presents certain problems in capturing significant features of style and dialect that may need to be represented. On the phonetic level, consistency can be difficult to achieve. In attempting to render deviations from what is assumed to be normal, some researchers use what is in effect a pseudo-phonetic form for certain words in the lexical stretch while maintaining conventional orthography for others in what appears to be an insufficiently motivated way, so that in one transcript the word "is" might be spelled in the normal way, but the word "because" might be spelled "b'cuz". This approach has been called "eye dialect" due to its comic-strip-like effect. Eye dialect not only reduces the readability of a transcript and interferes with computer retrieval of individual words and strings of text but also, and more importantly, it limits the audience for which data so represented are readily interpretable. Eye dialect relies on spelling conventions that often are not shared by all speakers of English. Readers may be fluent in English but if they are unfamiliar with comic strip conventions or the writings of modern conversation analysts, the deciphering difficulties they encounter may distract their attention from the main features of the analysis. In addition, eye dialect tends to trivialize participants' utterances by conjuring up pejorative stereotypes, while neither representing the phonetic level more precisely nor capturing detail relevant to the analysis.

We need to differentiate, on the one hand, between normal predictable pronunciation of a word for which the conventional spelling is an inaccurate representation of that pronunciation and, on the other hand, words which have variable pronunciations commonly associated with certain speech styles for which there exist relatively well-known spelling conventions. For example, the use or non-use of colloquial pronunciation may be a very salient aspect of the data. Stylistic variation in the pronunciation of "-ing" forms can reveal significant differences along the formality/informality dimension from which inferences can be made regarding participants' categorization of an interaction. Colloquial pronunciations of "going to" ("gonna," "gon," "-a") and "want to" ("wanna," "wan") can also be diagnostic of participant categorization. This kind of categorization can have a serious impact on the interpretation of the interaction, especially in cases of stylistic mismatch among participants. One option is to include both the conventional orthography and the popular spelling of the colloquialism, differentiating between the two in some regular way. For example, regularizations may be indicated as below.

Ex.4

H: i'm gonna ("going to") get a job/

Ex.5

H: ahma git ("i'm going to get") me a gig/ (Gumperz, 1982)

With this format, the transcript can be computer searched for register/style symmetry or verbal aspect, both of which may be significant to a particular analysis of the data. This is quite different from "b'cuz" where the pronunciation indicated is predictable rather than being one of several variants commonly associated with certain speech styles. Where, in a particular transcript, all participants consistently use a variety or varieties of colloquial pronunciation and that usage is relevant to the research interests for which the transcript is being developed, a table that indicates what regularizations have been made should be included in prefatory comments to the transcript. Parentheses may also be used in order to add detailed phonetic or phonemic transcription.

A similar technique of juxtaposing two representations of a particular string can be used in the transcription of code switching.

Ex.6

A: to this day he says that ... uh ... it's a shame they don't speak ... uh ... spanish/
están como burros/ les habla uno y,
("they are like donkeys, someone talks to them and,")

In this example, the speaker switches from English to Spanish within the same utterance. English appears in standard script, Spanish in italics and the translation is given using the convention for regularizations.

Unintelligible segments, frequently found in data recorded in natural settings, are conventionally represented with parentheses enclosing either a good guess or blank spaces. Where the number of syllables of unclear speech is countable, one lower case "x" per syllable can be enclosed within the parentheses. (Other conventions used in the examples will be discussed below.)

Ex.7

A: [enters the room]
H: (take a seat)/
R: hello m=ister A #surname#/=
A: =hello/=
H: ==(   )
R: ==take a seat/
A: good morning/

Ex.8

R: now, are the details the same as when you applied?
    you s- living at the same add(ress)?
K: ==y=es/=
R: =back= in road A?
K: ==yes/
R: ==yeah/

Ex.9

R: can i just check with you a few of the details? are you still living eh .. (xxx)?
B: (xxx), .. driscoe, bar mill/
R: {[p] bar mill/} bra- brookside is it?
B: yeah brookside, yeah/

In Ex.9, the fact that the first part of B's turn, although unintelligible, has the same number of syllables as the unintelligible last part of R's turn suggests that B may be echoing R, an interactively significant phenomenon.

As the above examples show, we follow the common practice of using the capitalized first letter of a person's name or some other upper case letter for identifying participants and other individuals named in the discourse while still maintaining their anonymity. This convention can be extended to other kinds of proper names, as in Ex.8, where anonymity of participants may otherwise be jeopardized. We restrict the use of upper case letters to serving as pseudonyms, having developed another convention for representing loudness as shown in Ex.10 and discussed in more detail below. If we were to use upper case letters to indicate both volume (or accent) and as pseudonyms, as is often done, then there would be no way to indicate a reference to a person uttered loudly (or with accent), for example. By restricting upper case letters to pseudonyms, this is easily done. The use of lower case letters for pseudonyms would reduce transcript readability since the participant-identifying, turn-initial cap so clearly flags turns at talk. Lower case letter pseudonyms within the text of the transcript may cause readers to confuse a reference to a person or place with false starts or other displays of nonlexical vocalizations. Like a number of other researchers, we use "X" to identify a speaker where it is not certain which participant is speaking. Colon follows the participant identifying cap.

Ex.10

B: {[ff] K,} oh .. i didn't see you there/

For purposes of computer-assisted analysis, it may be desirable to be able to isolate and collect or count instances of direct address and references by proper name – perhaps for purposes of identifying and analyzing rapport-building strategies. Reserving caps for use as pseudonyms facilitates that task. Conventions for representing some common permutations of names will be dealt with below.

Reading transcribed material is made easier if the relationship between the symbol and its assigned meaning is as mnemonic and transparent as possible, and if, all things being equal, those symbols which have been used conventionally with a certain assigned meaning in previous work by researchers in the field can be adapted to the current work. Internal consistency is important as well. Part of the interpretive task of transcription is the categorization of both the lexical (e.g., the representation of the speech stream in standard orthography or otherwise) and the nonlexical (e.g., voice quality, laughter, coughing, applause, etc.) phenomena present in the raw data. Internal consistency requires that similar phenomena be represented in similar ways and, conversely, that dissimilar phenomena be represented in dissimilar ways, as far as it is feasible to do so. In what follows, we describe our conventions. (See the appendix for a summary of these.)

2.4. Intonationally marked phrase boundaries.

An informational phrase can be described as a stretch of speech that falls under a single intonational contour or envelope and ends in an intonational boundary marker. These boundary markers indicate how the phrase in question relates to what precedes and what follows, and also provide information on such matters as whether the speaker intends to keep or to relinquish the floor, how definite a statement is, etc. The most common boundary markers are falling and rising intonation, both of which can be either slight or strong. Occasionally, it also becomes necessary to mark level endings, that is, where pitch neither rises nor falls. We use double slash "//" to indicate a final fall, which, depending on context, can be interpreted as indicating the conclusion of a stretch of talk, as a signal that the speaker is ready to yield her/his turn or as a marker of emphasis or definiteness. In some systems, period or dot "." is used for this because of the relationship between final falling intonation contour and the typical declarative sentence intonation contour. We prefer the double slash since dot serves as the pause marker. We use single slash "/" to indicate a contour terminating in a slight fall, which in some contexts ends a turn and in others suggests that there is more to come. The question mark "?" indicates terminal rise that signals uncertainty, as for example in yes/no questions. It can also serve to elicit listener feedback or to hold the floor. Comma "," indicates a terminal slight rise, which can, among other things, suggest that more is to come, as in the intonation typical for listing a series of objects, events, etc. or act as a turn holding signal. We use underscore "_" to indicate level ending. The last type of phrase final marker, dash "-", will be discussed below under truncation.

Another way of marking boundaries which may be mnemonically more useful is to adopt "/" for slight rise, "//" for strong rise, "\" for slight fall, and "\\" for final fall. This would also avoid the interpretive problems that arise when listeners automatically associate "?" with the speech act of questioning. For practical reasons, however, we have kept to the earlier usage since so much that has been published follows it. Some keyboards, moreover, may not have the "\" character.

The following example, taken from an informal narrative of a speaker reporting on her encounter with a waitress in a coffee shop, illustrates our use of phrase boundary markers. The narrator had found a hair in her sandwich and was informing the waitress of this fact.

Ex.11

...
1. so she said, {[f] [hi] well, whe:re *is it?}
2. ... and i said, whe:re's, {[ac] where's *what/}
3. well, {[very hi, shrill] where's the {[f] ˜hair/}}
4. .. {[hi] i can't re*turn it, unless i show them the {[f] ˜hair/}}
5. ... and i said, well {[hi] i- i don't, ˜know/}
6. you know, when i *noticed,
7. and {[laughing] {[lo] i just started laughing,}
8. you know, when i noticed what it was, i just {[f] threw it a*way/}
9. ... and i- and i'm: . beaming up, ==*at her,
10. and ... {[very lo] no go/}
11. [laugh] and she's staring down at me/
12. so i said, {[lo] well ˜look,} i- it's {[lo] somewhere/}
13. i'd- {[ac] it's in my} *na::pkin,
14. and {[ac] i start-} .. {[ac] [hi] and now you (know what-)}
15. {[f] i'm *picking up ˜boo:ks,}
16. {[hi] i'm *looking under the ˜pla::te,}
17. {[ac] i'm looking all over/}
18. ==and i finally look on my lap/
19. =={[lo] and i see this little curly hair//} [laughs]

...

Line 1, spoken in a high pitch register, terminates in a rise. The first two phrases in line 2 end in a comma-like, slight rise, while the third phrase ends in a slight fall. Terminal contours in the rest of the narrative are marked either by slight falls or slight rises that, in this context, both suggest that more is to come. In the concluding phrase, pitch falls and volume decreases before the final low fall "//", which indicates that the turn has ended.

2.5. Timing: pausing, overlap, latching and truncation.

Informational phrases are often bounded by pauses. We have adopted the convention of indicating short pauses by a series of dots. Generally, two dots indicate a half second or shorter pause; three dots indicate a pause between a half second and one second. It may be useful to enlarge this repertoire to include one dot for a brief but perceptible dysfluency and four dots for pauses longer than two seconds which are not timed (Tannen, 1984). It is, of course, well-known that pause phenomena may be significant to the assessment of the degree of conversational synchrony and cooperation within an interaction. We can easily do computer searches and counts for pauses since we do not use the period to indicate phrase-final falling intonation.

Ex.12

J: so you did motor mechanicking, what sorts of things were you doing/
M: ... i- i can't remember {[hi] all} things/

Ex.13

K: i think he was coming here on .. prospects/

We are aware that our treatment of pauses differs from that frequently used in conversation analysis, where all pauses are carefully timed (DuBois & Schuetze-Coburn, this volume; Ehlich, this volume). In view of our interpretive approach to transcription, what we see as important about pausing is what it signifies or how it enters into participants' inferential processes. Following Erickson and Shultz (1982), we assume that when participants are involved in an interaction, they automatically seek to establish a more or less regular rhythm and that, once established, this rhythm becomes the standard against which pausing is perceived. Rhythmic coordination in talk, when seen in this perspective, has somewhat the same character as rhythmicity in music, where deviations from the established beats, as perceived by musicians and audience, have both perceptual impact and signaling value in the performance as such. Our transcription seeks to capture this phenomenon. (See Gumperz, 1990, for a discussion of the signaling value of pausing.)

For some purposes, it may be desirable to provide an approximate numerical value for long pauses. We use angle brackets to display timed pauses (usually pauses over one second long). The angle bracket convention facilitates computer searches and frequency counts since angle brackets can be specified in the search command.

One further point: it is our practice, as well as the practice of many other researchers, to enter on a separate line with no participant identifier a lengthy pause which occurs between speakers' turns, thereby not attributing it to a specific participant (Tannen 1981). We do not usually do so, however, if the same speaker continues after the pause. In this we differ from some other researchers who note that the long pause may be indicative of a missed or passed turn, and that the same speaker self-selecting in the absence of another speaker taking over the floor is more accurately viewed as a separate turn, rather than a continuation of the same turn. However, it is not only long pauses which may indicate a missed or passed turn; many other features may do so as well. Among these are lengthening of a segment, rising intonation, reformulation of the proposition of the preceding informational phrase, a short pause, or some combination of these features. Compare the next two examples of talk between an interviewer (T) and a candidate for paid job training (A).

Ex.14

T: so how old are you now, twenty::? ... twenty-two/
A: yeah/

Ex.15

T: so how old are you now, twenty::?
A: ...
T: twenty-two/
A: yeah/

The second of the two examples more clearly reflects the interpretation that emerges from a detailed analysis of the event as a whole, that is, that A fails to collaborate with T, who is trying to get him to construct a narrative account of his work experience. Where this insight is relevant to the analysis, a researcher may choose either to represent A's missed or passed turns as in the second example or to address the issue instead through an analysis of T's turn internal pauses, lengthenings, rising intonation contours and reformulations. Conversely, a turn internal extended pause may be due to a nonvocal but otherwise unremarkable activity of the speaker which cannot be determined from the audiotape recording. It is no less true of long pauses than of other features that we cannot assign a constant interpretation to any single feature. It is rather a matter of how we choose to represent analytically relevant aspects of interaction.

Traditionally overlap, that is, simultaneous production by more than one participant, has been indicated by enclosing the overlapped portions of multiple participants' turns within long brackets, manually added to the text. For technical reasons, this convention is not feasible for computer stored and distributed transcripts.

In addition to bracketing, spacing may also be used to cue the reader to overlap since the second participant's contribution (utterance, laugh, nod, etc.) is begun at whatever point in the current speaker's turn the second participant begins. Spacing alone, however, is insufficient to indicate where the overlap ends since what was produced simultaneously during the interaction will likely be reproduced in the transcript with different numbers of typographical characters, thereby failing to accurately represent the phenomenon. Also, spacing alone cannot indicate turn initial overlap. While maintaining the spacing convention, we have added single equal sign "=" before and after the overlapped portions of each participant's contribution. Thus the beginning and end of the overlap is unambiguous and nothing need be added manually.

Ex.16

R: you haven't {[f] been} to the skill cen=ter,=
K: =no,=
    =i haven't/=
R: =so you have=n't seen the .. the workshop there?
K: no/
R: <1.2> no?

This convention can also be used to represent overlap of nonlexical material with both lexical and nonlexical material. (Conventions for representing nonlexical material are discussed in more detail below.)

Ex.17

R: so that's ok =with you then?=
L: ={[nod] sure/}=

Ex.18

R: so that's ok =with you then?=
L: =[nod]=

Ex.19

R: =[laughing]=
L: =[laughing]=

With respect to latching, that is, the immediate initiation of a turn following the termination of the previous turn, it has been noted that in much conversation a brief pause regularly precedes each turn at talk. Chafe (1987) has argued that this pause may result from cognitive constraints on language processing, and therefore it is significant and should be represented in the transcript. He indicates this turn initial pause in the same way as a turn internal pause would be indicated. In his system then, latching is shown solely by the absence of turn initial dots. Since the appropriate time interval between turns is (sub)culturally and situationally specific, in our own transcription system we are guided by the pacing characteristic of the event under analysis as a whole and by other relevant empirical findings in determining the time range within which the initial pause is predictable. Taking this predictable initial pause as the default case, we mark the *deviations* from it. Where participants' communicative histories differ as regards the default time interval between turns, one default interval may be chosen to serve as the standard, and that choice noted in prefatory comments. Although this practice obviously biases the transcript to some degree, when it is done consciously and noted explicitly, it can be revelatory of the implicit yet systematic differences among different communities of speakers.

The relationship between overlap and latching is a further consideration for explicitly marking latching. Latching, like overlap, results from participants' judgments about the content and intent of current speaker's turn. When latching is indicated only by the absence of the dots indicating a turn-initial pause, then the relationship of latching to overlap is obscured. Furthermore, it would be more difficult to isolate and count instances of latching since in same speaker multiple line turns, lines subsequent to the first line may or may not begin with a pause indicator. To search for latching, one would have to construct more complex commands rather than simply searching for a turn initial marker.

Researchers working in the conversation analysis tradition or CA, as it is sometimes called, (Atkinson and Heritage, 1986) use single equal sign "=" at the end of an utterance which is immediately followed by a second speaker beginning his/her turn, and single equal sign at the beginning of that second speaker's utterance. Because we use single equal signs for overlap, we indicate latching by double equal sign "==" at the beginning of the turn which is latched to the previous turn. In this way, we avoid confusion, while indicating the relationship between the two in an iconic way. In order to use space efficiently the double equal sign marker and the contribution it precedes are entered into the transcript at the left margin immediately following the participant identifier rather than beginning under the previous turn at the point where the first participant ends, as is often done. This has the additional benefit of maintaining in the transcript a sense of the rapid turn exchange noted in the recording.

Ex.20

R: have you visited the skill center?
A: yeah once/
R: ==you have?
A: ==before this, yeah/
R: was that on a tuesday?
A: yeah/

To indicate truncation, that is, (a) a speaker perceptibly breaking off his/her own utterance and reformulating or (b) a speaker breaking off his/her utterance as a second speaker begins, we use dash "-", maintaining the meaning commonly assigned this marker. In addition to level ending, truncation is marked by glottal closure and often also acceleration of the last few syllables.

Ex.21

T: it was about t- ah: nine thirty ten, when we got there/

Ex.22

B: so don't you think it would stand you in *better stead
to com**plete the engineering course,
complete the last ten weeks of *that course,
rather than start out on something you actually know very little a*bout//
R: .. from {[hi] your} point of view is ok, what you're saying is ok, bu-
B: ==and as T has already *told you, the preentry test is quite a *stiff one/

To facilitate computer searches and counts, we use the dash only for the representation of truncation and avoid its use word internally, even in words like "pre-entry" where standard orthography might use it. However, modifications of the search command to restrict a search to dash followed by a space or by end of line would allow the use of word internal dash to be distinguished from truncation where readability of the transcript is judged to suffer.

2.6. Characteristics of the whole phrase: pitch register, rhythm and tempo.

Like turn initial pauses, and many of the other signs we discuss, the signs in this section indicate inherently relational phenomena. They do not have absolute values. Listening to an entire event segment, the transcriber determines what is "normal" for individual participants, then judges given stretches of talk against this norm. We indicate a stretch of a participant's talk relative to other stretches of his/her own or others' talk as higher or lower in pitch register by preceding the relevant lexical stretch with [hi] or [lo] and enclosing the pitch register shift marker and the lexical stretch in curly brackets, as shown in the examples below.

Phoneticians and discourse analysts have described these phenomena under the headings of "rhythm" and "tempo." The communicative significance of this is that it indicates the extent to which speakers are in tune or in sync with each other, that is, whether or not they are in agreement (Erickson & Shultz, 1982). We use a similar convention to indicate a stretch of talk which is faster or slower in relation to what precedes, follows or has otherwise been determined as the "norm," using [ac] and [dc] respectively.

Ex.23

W: {[hi] what do you mean/} i mean why are you saying that/

Ex.24

K: {[ac] if you don’t ask now,} it’ll be too late/

Ex.25

J: {[dc] i suppose} that could be right after all/

The above features can also be nested one within another, extending over different stretches of the utterance.

Ex.26

J: {[lo] {[dc] i suppose} that could be right after all/}

2.7. Phrase internal signs: accent, loudness and syllable lengthening.

The terminal markers discussed in section 2.4 above are only one aspect of intonation. Intonation, along with other signaling mechanisms such as vowel lengthening, increase or decrease in volume or pitch level also enters into the signaling of communicatively significant prosodic prominence or accent (Auer, 1988; Bolinger, 1986; Couper-Kuhlen, 1986, 1989; Couper-Kuhlen & Auer, 1988; Ladd, 1980). We indicate accent placement within the phrase in the following ways: (a) normal prominence indicated by a single asterisk "*" preceding the syllable, (b) extra prominence indicated by double asterisk "**" preceding the syllable, (c) fluctuating pitch over a single word indicated by tilde "˜" preceding the syllable. Asterisk and tilde were chosen for their respective tasks because they convey visually something of the quality of the voice as it is perceived auditorally on the tape. These conventions can also be used along with the upper case letters for person identification within the turns at talk. Our experience has shown that these markings, along with the terminal markers, reflect the most important discourse level signaling characteristics of intonation.

This set of markers conflates some of the detail captured in the systems worked out by British phoneticians (Trim, 1971). For example, we use "˜" to indicate fluctuating intonation, without distinguishing rise-fall from fall-rise while that system uses "ˆ" to indicate rise-fall and "ᵛ" to indicate fall-rise. The Trim system also allows for the marking of direction of pitch movement where that movement is simple rather than complex, as with the fluctuating intonation. Thus, " ’ " indicates rise, " ‘ " indicates fall. When more detail is needed, the British system can be used or intonational curves can be drawn and superimposed on the text. Examples of this can be found in the above references.

Ex.27

N: that’s ridiculous, **P gets to stay out until midnight on weekends/
J: i don’t care if ˜P gets to stay out until four in the morning, **you can’t/

Ex.28

R: he doesn’t have any idea what it’s all a*bout/

Ex.29

C: there’s no ˜way he’s got all the **right end of the stick/

Some researchers feel that when accent is predictable, it is unnecessary to represent it, as for example, when the accent falls on the last content word of the informational phrase, as in Ex.28. We tend to follow this convention whenever predictable accent is not relevant for a particular analysis. However, there are instances when what is predictable given the larger context may not be immediately obvious to the reader, yet accenting is relevant to the analysis. In these cases, it becomes necessary to cue the reader as to which word has received accent. The following examples illustrate this. In Ex.30 "swamper" is highlighted and is treated as new information, and thus marked with accent, while "S Construction," which also carries some stress but is perceived as less prominent, is not. In the next line, "bricklaying" is used contrastively, and we use double asterisk to mark it off both from "swamper" and from "too."

Ex.30

H: i see you were a *swamper for S Construction/
B: yeah, i did some **bricklaying for them *too/

Ex.31

H: i see you *also worked for S Construction/
B: yeah, i did some bricklaying for *them too/

Of course, since the object of study is an interactional sequence as a whole and not isolated utterances, the disambiguating property of context will always be available to the reader.

There is one further consideration, however: the importance of providing support for the claim that dialectal differences between speakers with regard to the prosodic characteristics of talk have significant impact on the success or failure of the interaction. Large corpora of transcripts where a wide range of prosodic features, including accent, are represented (e.g., London/Lund corpora, see Svartvik & Quirk, 1980) facilitate computer-assisted comparison among speech communities which employ lexically and syntactically similar systems within their linguistic repertoires. Talk among people who share ethnicity is the basis for establishing norms which then enter into the analyses of interethnic communication. Furthermore, the explicit representation of predictable accent and many other prosodic features, as well as turn-initial pausing (discussed above), would minimize the implicit influence of the transcriber who, in listening to the data, must determine what is predictable and what is deviant. Limits on time allowable for transcribing may mitigate against marking these prosodic features, despite the desirable end of developing such corpora. Also, since no transcription is final, we assume that in comparative analyses investigators will always want to go back to the tape recordings to validate findings.

For louder or softer stretches of a participant’s speech, we use [f] and [p] respectively, in the same way as was described above for indicating shifts in pitch register and tempo. These two symbols are taken from musical notation and indicate, by analogy with music, "loud" (forte) and "soft" (piano) relative to normal volume level (Tannen, 1984). If caps were used to indicate loud speech, as is done by some, there would be no comparable way of easily indicating soft speech or shifts in pitch register and tempo, all of which seem to be similar phenomena. It may be useful to be able to isolate, collect and count such changes in voice quality in an analysis of rhetorical style, or as evidence in support of an evaluation of an interaction as a heated exchange, and so forth. Note that in this system shouting could be indicated by "ff" (fortissimo = very loud) and whispering by "pp" (pianissimo = very soft) so that the transcriber need not make the evaluative judgment implicit in designating a stretch of speech as shouting or whispering. These changes in voice quality (hi, lo, ac, dc, f, ff, p, pp) are easily isolated and counted by searching for two or fewer characters within the square brackets. It is for this reason and for improved readability that it was suggested above that similar phenomena be represented by similar conventions, dissimilar phenomena by dissimilar conventions.

Segments of a word are sometimes lengthened to increase the salience of the word and sometimes to allow the speaker additional processing time for formulating the rest of his/her utterance. The use of colon ":" to indicate lengthening is conventional in much conversation research and we adopt that usage here. Although not originally chosen with regard to computer-assisted analysis, colon works well for this purpose because it can be suppressed when searching for instances of specific words. The use of reduplicated segments to represent lengthening results in spelling irregularities which complicate searching. We reject some symbols because they look too imposing when several of them are necessary to indicate (impressionistically) various degrees of lengthening. In our system, the only other use of colon – following the single upper case letter indicating who is speaking, coughing, etc. – will not be confounded with lengthening since the first two fields of a line are reserved for participant identification.

Ex.32

B: yeah, i’ve been to the class, and it looks::-
R: .. yeah,
B: ==it looked ok really/

The lengthening convention can be used with the person identifying upper case letter to indicate that lengthening occurred in the pronunciation of the name even though the particular lengthened segment is not specified. This maintains the anonymity of the referent while loosing little interactively relevant material.

Ex.33

G: did you mean F::? you’re talking about D’s son?
R: <1> no?

One other signaling mechanism that needs to be discussed under this heading is *rhythmicity*, that is, the spacing in time of successive prosodic prominences, accents or rhythmic beats. One of the first discussions of this can be found in Erickson and Shultz (1982). The phonetic realizations of rhythmicity have recently been described in a series of detailed studies by Auer, Couper-Kuhlen and their associates (1986, 1988). Rhythmicity can characterize a single person’s speech, and it may also occur across different speakers’ turns in an interaction. Rhythmic synchronization of participants’ turns may play an important part in marking the relative smoothness or communicative success of an interaction or in signaling conversational repairs. One way to mark rhythmicity is by placing the left square bracket "[" at the beginning of the rhythmic unit, as in the following example adapted from Couper-Kuhlen. Couper-Kuhlen and Auer use "/" to mark rhythmic units and employ " . " to mark falling intonational boundaries. To be consistent with our system, we have substituted "[" for "/".

Ex.34

T: I [didn't even [know him in [nineteen [twenty three [did I [Ernest/
L: he [wasn't   [born in   [nineteen [twenty three//

Note that the rhythmic units here are aligned to indicate rhythmic integration across speakers. Lack of synchrony can also be indicated by spacing.

2.8. Overlays of the lexical stretch.

Overlays, as we call them, are nonlexical phenomena produced simultaneously with the lexical stretch by a single participant, as opposed to overlaps which are the simultaneous productions of more than one participant. When some nonlexical phenomenon is discernible over the lexical stretch, it should be represented in the transcription in a way similar to other such phenomena. For example, if in a given interaction a participant, during one of his/her turns, pounds on a table with his/her fist punctuating the lexical stretch at certain intervals, some convention should be employed to signal to the reader what was done and when relative to the spoken discourse.

Ex.35

M: {[fist pounding on table = "+"] we've +asked you .. +over and +over again/}

Likewise, if sometime during an interaction a participant punctuates his/her lexical stretch at certain intervals with fingersnapping, some similar convention, suitably adapted, should be used.

Ex.36

L: {[fingersnap = "+"] i +know you +want to +go, so +blow/}

Nonlexical phenomena which overlay continuously rather than intermittently some lexical stretch are similarly represented. For both intermittent and continuous overlays, the lexical material included within the curly brackets is understood as overlaid by the nonlexical material represented by the label within the nested square brackets.

We employ the same convention to represent the superimposition over the lexical stretch of nonlexical phenomena that are discernible (i.e. visible and/or audible) and nonvocal (Ex.37–Ex.39) with nonlexical phenomena which are discernible and vocal (Ex.40).

Ex.37

R: {[fingersnap] right/}

Ex.38

S: {[nod] sure/}

Ex.39

P: {[handing paper to doctor] is this it?} is this the one you mean?

Ex.40

M: {[laughing] are you sure of that?}
because this wouldn't be the first time someone made that mistake/

2.9. Vocal and nonvocal interruptions

In the above section, we discussed the representation of simultaneity of lexical and non-lexical elements within a single speaker's turn. We now turn to interruptions of the lexical stretch. *Interruption*, as the term is being used here, labels nonlexical phenomena produced sequentially (rather than simultaneously, as with overlays) with the lexical stretch by a single speaker; interruptions by other speakers, whether they are realized as overlapped utterances or latched utterances, were dealt with above. Where lexical and nonlexical elements are sequential rather than simultaneous, this can be indicated by the absence of the curly outer bracketing.

Ex.41

T: i asked her to [cough] come in when she can/

Ex.42

L: come in/ [door closing] take a seat/

For timed interruptions, a number following the descriptor indicates its duration in seconds.

Ex.43

T: i asked her to [cough .6] excuse me, to come in when she can/

Ex.44

L: come in/ [door closing 2.5] take a seat/

2.10. Comments about computer searches.

For practical purposes, if, for example, one wanted to have a computer tabulate instances of laughter, a search for "[laugh" would retrieve both the simultaneous and the sequential. Similarly all instances of simultaneous realization could be counted by searching for "{"; more specific types of simultaneous realization could be counted by searching for "{[laugh", "{[lo", etc. All instances of sequential realization could be counted by searching for " [" (where "space" is understood as a character by the computer as it is in UNIX); more specific types of sequential realization could be counted by searching for " [laugh", " [cough", etc. In general, we try to limit our use of symbols that have special meanings for editing and search programs to avoid complicating the transcription process.

2.11. Background information.

Although background information can precede or follow the body of transcribed text, there may be times when it is desirable to include it at the point within the text where it is relevant to the unfolding of the interaction, particularly noting situations in which some important information is known to some but not all of the participants (Ervin-Tripp, personal comment). Background information can be included in the turn at talk where it becomes relevant by flagging it in some way so the reader can easily recognize that it is not part of the recording itself. The pound sign "#" before and after the background information works well to set it off from the transcribed text. Where the information serves as background for the whole turn at talk and is rather lengthy it can precede the utterance. As in the next example, a dotted line is used to separate the lengthy background information from preceding and following transcribed talk.

Ex.45

Family dinner table conversation on a Sunday evening among a husband (J)
and wife (L) who are both lawyers and their teen-age son (T).
L has just returned to the table following a telephone call.

J: who was it/
L: {[sigh] a **client/}
#J knows that one of L's clients has been phoning her at home
and that L is annoyed by his calls#
J: {[laugh] i think he must be in love with you/}
T: who was it mom, got a new boyfriend?

This convention can be extended, as in Ex.46, to specify what kind of name the upper case participant identification letter represents: surname, given name, nickname, diminutive, and what grammatical function the name plays: vocative or referential. Such information can be important in determining how participants categorize the interaction along the formality/informality continuum, for looking at rapport-building strategies, and the like. It is however insufficient for representing the innovative permutations of names that occur in word play, a frequent and significant phenomenon in child language acquisition data. For this type of data the use of phonologically similar pseudonyms is preferred for obvious reasons.

Ex.46

M: ok, and thank you mr R #vocative surname#/
i'll let you have a chat now with mr C #referential surname#/
C: morning A #vocative given name#/
R: morning/

2.12. Transcription of languages other than English.

For transcribing languages other than (standard American) English, we use the following format: text line (t), morpheme by morpheme line (m), English translation line (e), each line beginning with the participant-identifier and the line type identifier.

Ex.47

Rt: [text]
Rm:[morpheme line]
Re: [English translation]
Bt: [text]
Bm:[morpheme line]
Be: [English translation]

A space between speakers can be added to improve readability for long turns although this complicates the use of spacing as one of the indicators of overlap. However, since overlap is also indicated by "=" before and after the overlapped sections, readability should not suffer too much.

Ex.48

```
Rt:  você ta ("está") estudando pra ("para") =o exame de amanhã?=
Rm: you are           studying for            the exam of tomorrow
Re:  are you           studying for            =tomorrow's exam?=
Bt:                                            =quê droga, ne?=
Bm:                                            what drug not-it-is
Be:                                            =what a drag, isn't it?=
```

With this format, one can select to search just the English, just the language of the interaction, all of one speaker's utterances, etc.

As an alternative to the three-line format, parentheses can be used as in the case of regularization discussed above, but only for brief sequences since readability suffers if the turns are long. Here the morpheme by morpheme translation would most likely be omitted. These parentheses should, in any case, be parentheses with quotes. Recall that parentheses without quotes are for unintelligible sequences. As in Ex.5 adapted here as Ex.49, it may, in fact, be preferable to use this convention instead of the tripartite format for regularizations of dialects to the standard form in the language of the interaction.

Ex.49

W: ahma git me a gig/ ("i'm going to get myself a job")

2.13. Nonverbal phenomena.

We have dealt with a variety of common nonverbal signs in various sections. Most common nonverbal signs can be indicated by means of square bracketed comments. The following extract from a recent analysis of a videotape will illustrate this.

Ex.50

```
 1. A: i went down to (xxxxx) on that site,
 2. B: {[head tilted down but gaze on A] yeah/}
 3. A: in eh-  B #place name#) .. i've got to come here now/
 4. B: [bends forward towards A]
 5. A: ... i've been in estimate, eh- if number seventy-six is on the list/
 6. B: .. yes/
 7. A: it is- .. it isn't/ ... for plastering//
 8. B: [slight nod, no other response although A is clearly expecting something more]
 9. A: ... and eh- {[hi] when i came down here before,} just a few weeks ago,
       .. {[lifts her hand and points with a jabbing motion
       to someone standing behind A] that fellow there tell me/}
       {[f] it should have been done in march//}
10. B: =={[ac] i'll just check on it okay?} [gets up and walks away]
```

Applause is an additional important signaling device. It enters into conversation in several ways. When used by individual participants, it may either overlap, latch or otherwise function as a conversational turn. Atkinson and Heritage (1986) suggest alternately using lower and upper case "x" to indicate increasing or decreasing volume. We suggest using upper case "X", along with the system of signaling loudness described above, since lower case "x" is reserved for representing unintelligible lexical material. "X" is repeated to approximate either the length or the manner of the turn. Thus, X indicates a single clap, XXXXX is repeated clapping, X-X-X- signals spasmodic or hesitant clapping. The use of upper case "X" turn internally will not overlap with its use as a participant identifier since the participant cap is always in the first field of a line.

An alternate system for nonverbal signs, particularly useful for representing gaze throughout the transcript, is that based on the work of Charles Goodwin. An exposition of this can be found in Atkinson and Heritage (1984).

3. Sample transcripts and analyses.

By way of further illustration of how we use the transcription system, we conclude with a few, somewhat longer examples from texts we are currently working on. The first example is taken from a consultation session between a medical student and her supervising physician. The medical student has just examined the patient and is reporting on her preliminary diagnosis to the physician who questions her, and in the end (not shown here) suggests an alternate diagnosis.

Ex.51

1. A: {[ac] [p] so what do you think is going on?}
2. S: .. {[dc] {[f] well,} i *think her: .. her *leg pain,}
3. i mean, her **knee pain, it doesn't seem to be *any te- any tears/
4. and doesn't seem to be any u:m *fractures/ {[p] at all/
5. and no patella- and no patella dislocation/ or anything/}
6. so, {[f] *i think} this *may be like, kind of *post traumatic/
7. ==she hasn't *taken anything, *for it/
8. A: ={[pp] mhm/=
9. S: =any= aspirins, or any um .. NXID's, or anything like that/
10. ==that may reduce the inflammation, or the swelling there/
11. ==so, u:m, .. i'd be inclined to maybe:, {[dc] see how she-}
12. to give her some aspirins, see how she d- if the pain resides on that/
13. A:==mhm/
14. S: ==um so, i think this may be just post traumatic/
15. ==because it's- {[f] it's been about three months/}
16. but the pain's been redu- {dc] has been reducing/}
17. ==it's not- definitely not *increased/
18. A:={[pp] [hi] (mhm)/}=
19. S: ={[p] (since)= she . haven't-} hasn't had any um .. major problems with tha::t/}
20. =={[f] and she gets it *mostly on physical exertion/}
21. ==it's not something that comes on .. at rest/
22. =={[p] or, while she's, not doing anything, really, u:m}
23. A: =={[p] mhm/}
24. S: ==real tiring/

In line 1, our use of [ac] and [p] indicates increased rhythm and quieter tone in comparison with the preceding talk. Similarly in line 2, [dc] and [f] mark slower rhythm and more fortis (more pronounced) tone. The curly brackets in line 1 show that the voice quality markers are to be understood as affecting the whole line. In line 2, the fortis "well" is nested within the curly brackets delimiting the lexical stretch marked by deceleration, indicating that "well" is both louder and slower, while the rest of the utterance is only slower than the preceding talk. In this way it becomes possible to locate the sequential positioning of rhythmic shifts, relate them to other cues such as pausing and make more explicit the rhythmic organization of a turn at speaking. Depending on context, these features of rhythmic organization may be quite important in distinguishing between such conceptual level phenomena as planning, hesitation, uncertainty and the like.

The two dot initial pause in line 2 indicates that S hesitated slightly longer than the predictable turn initial pause. The colon following the "r" of "her" shows that this segment was lengthened. This pausing, deceleration and lengthening are likely to be significant to an analysis of the interaction since they suggest that S is carefully planning her turn. For the same reason, the normal accent on "think" is included to suggest our hypothesis that, in accenting "think," S is framing her diagnosis as tentative.

We chose to include the normal accent on "leg" to mark the contrast with the stronger accent on "knee" in the following line. This is also true for the normal accent on "any." This accent is significant to the analysis because "tears" is not accented. It can therefore be understood as being treated as old information despite the fact that no explicit mention of tears had been made before this point in the interaction.

The "at all" in line 4 and the "or anything" in line 5 are treated as separate tone groups primarily on the basis of intonational contour; they are not preceded by detectable pauses. Communicatively, this contouring may say something about the nature of the hedge.

"She hasn't" in line 7, "that" in line 11 and "so" in line 12 follow immediately on the preceding informational phrase without a pause. At the risk of complicating our search procedures, we have adapted our convention for representing latching as discussed above to the representation of this intra-turn latching. While rarely described in the literature, intra-turn latching is a common turn-holding device in such diagnostic sessions, where speakers work hard to create the space to make their point. The fact that A's latched "mhm" in line 14 is countered with a latched "um" by S confirms our hypothesis. Note also that in line 18 A raises her pitch register, a common attention getting device, but her turn is again overlapped by S. A seems to be trying to take over, perhaps to stop S before she commits herself to a potentially embarrassing, inaccurate diagnosis, but gives up when S fails to yield the floor. In these and other ways, the transcription system displays the evidence for our hypothesis concerning participants' conversational management strategies.

The next example, taken from a recording of a televised 1960's debate between the then mayor of San Francisco and a local African-American community leader, will illustrate how the transcription system can be used to illustrate communicatively significant differences in conversational style.

Ex.52

[Responding to B, who has objected to what he has said.]
1. A: ==( ) {[hi] you'll} see those things/ {[hi] you'll} see those things/
2. but the *point **i:s,
3. .. that the *black community in San Francisco Hunter's Point/
4. is *not giving me the story that they're being excluded

5. from any kind of positions of responsibility/
6. .. and i **have *told them, if they’re gonna par*ticipate in these things/
7. that there’re gonna be *no, .. bequeaths from Mt Olympus down *to them//

[One turn by a third speaker omitted.]

8. C: =={[lo] i would just like to say *this/}
9. that uh .. the mayor, ... is the mayor of San Fran*cisco/
10. he’s not the mayor of the United *States//
11. .. and uh, that’s his area of responsi*bility//
12. ==an::d, i would like to take exception to uh,
13. ... the glowing *picture that he’s just painted,
14. .. as to the relationship between uh,
15. .. his *office, and his admini*stration, and his *government,
16. .. and the black community//
17. now it may well *be that he has,
18. a good rapport with certain people in Hunter’s Point//
19. .. {[hi] at the same *ti:me,}
20. .. the black community is very well a*ware/
21. ... that San Francisco, .. *under Mayor A-
22. ... {[ac] A, excuse me/}
23. A: {[pp] alright//}
24. C: uh ... the same as a::ll,
25. .. uh city governments across this *country/
26. ... {[dc] are sharply escalating/}
27. ... the war against the black community//

Note that A relies mainly on accent and terminal contours to achieve rhetorical effect, and we indicate that by using the appropriate symbols. On the other hand, C relies on phrasing, pausing and sometimes pitch register shifts. He regularly uses these devices to break up syntactic units in order to highlight his points. There is some suggestion here that the speakers rely on somewhat different contextualization conventions which, depending on power relations and the nature of the audience, may affect rhetorical effectiveness and the outcome of the debate.

In the last example, the first speaker is a native speaker of English and the second, a native speaker of a North Indian language whose English style shows some signs of language interference. The two speakers are arguing about admission to a community college course. The extract begins shortly after the beginning of the conversation.

Ex.53

2L: what I *said was/
.. that it was *not a suitable course/ .. for you to *apply for//
because it is ( )//
.. {[lo] now if you *want to apply for it/}
.. {[hi] of *course/} you can do what you *want//
but/ {[hi] if you are *doing the twilight course at the *moment/}
.. {[lo] it was *not something which-}
.. Mrs N and Mr G *thought/ *originally/
that it was a course to carry *on/ *with the *twilight course/
but this is {[ff] not} the case//
3D: no// what you- you take *one thing at a time*//
this case// that whatever {[f] *they know*//}
I get that even .. hmm// for a D .. *me*//

{[lo] and I am *student in E *College*//}
and Mr W *knows me*// he// .. I am student in the *same school*//
[f] he knows *my qualifications*/} and what- whether I'm suitable *or not*//

L's use of prosody is typical for British or American native speakers of English. She makes extensive use of accenting to highlight certain aspects of the message and of pitch register shift over a complete syntactic clause to mark such things as the contrast between main information and asides, as well as of terminal contour to indicate interclausal relations. In D's speech, on the other hand, accent does not fall on a single syllable. Typically, his clauses are divided into two strings, one of which is set off from the other by a combination of tempo, rhythm and loudness. This contrasts with L's use of typically English cues, and is most probably a reflection of significant differences in communicative background. More research is needed to understand the details of the prosodic and paralinguistic signals characteristic of D's and many other South Asian natives' English and to determine how these cues enter into his contextualization strategies. Yet, for the purposes of the analysis at hand, it seems sufficient simply to mark the differences between the two speakers' strategies. For this, we chose to use italics in D's case because it is less cumbersome and less intrusive than the bracketing convention, while at the same time conveying something of what the difference between styles is like. Where this font change causes difficulty in transferring across systems, other conventions can be used.

4. Conclusion

The transcription system we have described and illustrated can be used for conversations as well as for monologic texts. In conversation, the focus is on the interactive aspects of conversational management, and on the perceptual cues that speakers and audiences rely on in signaling what they intend to convey and in assessing what they hear. In monologue, it is well-suited for revealing the use of prosodic and paralinguistic signs to establish and maintain coherence.

We are aware that transcription is by its very nature inherently selective and that what is selected depends in large part on analysts' background knowledge and theoretical perspective. Much has been left out and many of the questions we have touched on are far from resolved. As we said at the outset, our main goal is to reveal the functioning of communicative signs in the turn by turn interpretation of talk, not to record everything that can be heard or provide exact measures of duration and pitch. Yet, at the same time, we seek to remain as comprehensive and attentive to detail as possible in showing what the phenomenological or perceptual bases of our interpretation are.

5. Acknowledgements

This transcription system was first developed for use in the University of California, Institute of Cognitive Studies, Disclab, an on-line text archive, which serves as a database for studies of language use in a variety natural contexts, ranging from institutional settings such as formal interviews, doctor-patient interactions, seminars or academic lectures to family interaction, children's talk and informal conversation. We are grateful to members of the Disclab group, Susan Ervin-Tripp, Wallace Chafe, Dan Slobin, Jane Edwards, Paula Chertok, Sarah Freeman, Martin Lampert, and Meryl Siegal, for their many helpful comments. We would also like to thank Peter Auer for his helpful comments on an earlier version of this paper. The following basic sources have been useful. The approach to pausing, latching and overlap owes a great deal to Gail Jefferson's transcription system, described in Atkinson and Heritage (1984). Our notion of prosody rests in part on that developed by John Trim (1971). The perspective on rhythm and conversational synchrony is based on that of Erickson and Shultz (1982). We have also profited greatly from reading Couper-Kuhlen (1986).

Appendix

Symbol	Significance
//	Final fall
/	Slight final fall indicating temporary closure (e.g. more can be said on the topic)
?	Final rise
,	Slight rise as in listing intonation (e.g. more is expected)
-	Truncation (e.g. what ti- what time is it/)
_	Level ending

..	Pauses of less than .5 second
...	Pauses greater than .5 second (unless precisely timed)
<2>	Precise units of time (= 2 second pause)
=	To indicate overlap and latching of speakers' utterances e.g. R: so you understand =the requirements?= B: =yeah, i under=stand them/ ~~~~~~ R: so you understand the requirements? B: ==yeah, i understand them/ R: ==and the schedule? B: yeah/ with spacing and single "=" before and after the appropriate portions of the text indicating overlap and turn initial double "=" indicating latching of the utterance to the preceding one.

::	Lengthened segments (e.g. wha::t)
˜	Fluctuating intonation over one word
*	Accent; normal prominence
**	Extra prominence
{[ ]}	Nonlexical phenomena, both vocal and nonvocal, which overlays the lexical stretch e.g. {[lo] text//}
[ ]	Nonlexical phenomena, both vocal and nonvocal, which interrupts the lexical stretch e.g. text [laugh] text//

( )	Unintelligible speech
di(d)	A good guess at an unclear segment
(did)	A good guess at an unclear word
(xxx)	Unclear word for which a good guess can be made as to how many syllables were uttered with "x" = one syllable
(" ")	Regularization (e.g. i'm gonna ("going to") come soon/)
# #	Use cross-hatches when extratextual information needs to be included within the text (e.g. R: did you ask M #surname# to come?)

Bibliography

Auer, P. (1988, April). Rhythmic integration in phone closings. *KontRI Working papers.* Konstanz, Germany: University of Konstanz, Department of Linguistics.

Atkinson, J. M. and Heritage, J. (Eds.). (1984). *Structures of social action.* Cambridge: Cambridge University Press.

Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford: Stanford University Press.

Bolinger, D. (1989). *Intonation and its uses: Melody in grammar and discourse*. Stanford: Stanford University Press.

Chafe, W. (in press). Prosodic and functional units of language. In J. A. Edwards & M. D. Lampert (Eds.), *Transcription and Coding Methods for Language Research*. Hillsdale, NJ: Lawrence Erlbaum.

Chafe, W. (1987). Cognitive constraints on information flow. In R. S. Tomlin (Ed.), *Coherence and grounding in discourse*. Philadelphia: John Benjamins.

Couper-Kuhlen, E. (1986). *An introduction to English prosody*. London: Edward Arnold.

Couper-Kuhlen, E. (1988, April). Contextualizing discourse: The prosody of interactive repair. *KontRI Working papers.* Konstanz, Germany: University of Konstanz, Department of Linguistics.

Couper-Kuhlen, E., & Auer, P. (1988, April). On the contextualizing function of speech rhythm in conversation. *KontRI Working papers.* Konstanz, Germany: University of Konstanz, Department of Linguistics.

DuBois, J. W., & Schuetze-Coburn, S. (in press). Outline of a system for discourse transcription. In J. A. Edwards & M. D. Lampert (Eds.), *Transcription and Coding Methods for Language Research*. Hillsdale, NJ: Lawrence Erlbaum.

Ehlich, K. (in press). HIAT: A transcription system for discourse data. In J. A. Edwards & M. D. Lampert (Eds.), *Transcription and Coding Methods for Language Research*. Hillsdale, NJ: Lawrence Erlbaum.

Erickson, F., & Shultz, J. J. (1982). *The Counselor as gatekeeper*. New York: Academic Press.

Gumperz, J. J. (1982). *Discourse strategies.* Cambridge: Cambridge University Press.

Gumperz, J. J. (1990). Contextualization and understanding. In A. Duranti & C. Goodwin (Eds.), *Rethinking context*. Cambridge: Cambridge University Press.

Gumperz, J. J. (in press). Contextualization revisited. In P. Auer & A. Di Luzio (Eds.), *Contextualization*. Amsterdam: John Benjamins.

Gumperz, J. J., & Roberts, C. (1989). Understanding in intercultural encounters. In J. Verschueren (Ed.), *Proceedings of the 1987 Meetings of the International Pragmatics*

Association. Amsterdam: J. Benjamins.

Kay, P., & Coleman, L. (1981). Prototype semantics: The English word *lie*. *Language*, **57**, 26-44.

Ladd, R.D. (1980). *The structure of intonational meaning.* Bloomington, IN.: Indiana University Press.

Svartvik, J. and Quirk, R. (1980). *A Corpus of English conversation.* Lund, Sweden: C.W.K. Gleerup.

Tannen, D. (1984). *Conversational style*. Norwood, N.J.: Ablex.

Tannen, D. (1981). Review of: *Therapeutic discourse: Psychotherapy as conversation*. By William Labov and David Fanshel. *Language,* 57(2).481-6.

Trim, J. (1971). *English intonation*. Cambridge: University of Cambridge, Department of Linguistics.