Chapter 10

Elinor Ochs

TRANSCRIPTION AS THEORY

Naturalistic speech as a database

[. . .]

A pervasive sentiment among those who draw from performance data is that the data they utilize are more accurate than intuition data: their data constitute the real world – what *is* as opposed to what *ought* to be. There are many issues that could be entertained concerning this orientation. Here I would like to address the problem of what in fact are the performance data for such researchers: even here the internal issues are manifold. There is the issue of data collection: the means of observing and recording, the conditions (setting, time, etc.) under which the data are collected, and so on. The influence of the observer on *the* observed is, of course, a classic concern within the philosophy of science (Borger and Cioffi 1970; Popper 1959).

The utilization of mechanical means of recording may appear to eliminate some of these problems. An audiotape recorder registers a wide range of sounds and a video-tape recorder registers visual behavior falling within its scope. (We are ignoring for now the problem of camera placement; use of zoom versus wide-angle lens, and so on.) A stand taken in this chapter is that the problems of selective observation are not eliminated with the use of recording equipment. They are simply *delayed* until the moment at which the researcher sits down to transcribe the material from the audio- or video-tape. At this point, many of the classic problems just emerge.

Source: Elinor Ochs, 'Transcription as theory', in Elinor Ochs and Bambi B. Schieffelin (eds) *Developmental Pragmatics*, New York: Academic Press, 1979, 43–72.

A major intention of this chapter is to consider with some care the transcription process. We consider this process (a) because for nearly all studies based on performance, *the transcriptions are the researcher's data*; (b) because *transcription is a selective process reflecting theoretical goals and definitions*; *and* (c) because, with the exception of conversational analysis (Sacks, Schegloff and Jefferson 1974), *the process of transcription has not been foregrounded in empirical studies of verbal behavior*. The focus of this discussion will be on the nature of transcription for child language behavior. [. . .]

One of the important features of a transcript is that it should not have too much information. A transcript that is too detailed is difficult to follow and assess. A more useful transcript is a more selective one. Selectivity, then, is to be encouraged. But selectivity should not be random and implicit. Rather, the transcriber should be conscious of the filtering process. The basis for the selective transcription should be clear. It should reflect what is known about children's communicative behavior. For example, it should draw on existing studies of children's cognitive, linguistic, and social development. Furthermore, the transcript should reflect the particular interests – the hypotheses to be examined – of the researcher.

One of the consequences of ignoring transcription procedure is that researchers rarely produce a transcript that does reflect their research goals and the state of the field. Furthermore, developmental psycholinguists are unable to read from one another's transcripts the underlying theoretical assumptions.

Yet, these skills are critical in understanding and assessing the generalizations reached in a particular study. As already noted, the transcriptions are the researcher's data. What is on a transcript will influence and constrain what generalizations emerge. For example, the use of standard orthography rather than phonetic representation of sounds will influence the researcher's understanding of the child's verbal behavior. One area of behavior that is 'masked' by the use of standard orthography is sound play (Keenan 1974). The use of standard orthography forces a literal interpretation on utterances that otherwise may be simply objects of phonological manipulation. The use of standard orthography is based on the assumption that utterances are pieces of information, and this, in turn, assumes that language is used to express ideas. In sound play, the shape rather than the content of utterances is foregrounded and the function of language is playful and phatic (in the case of sound-play dialogue) rather than informative: where the researcher uses standard orthography, not all instances of sound play can be easily seen. This assumes importance when a case of sound play is reported in the literature, as in my own situation. It is difficult to assess whether its rare appearance in the literature

reflects the nature of children's verbal behavior or the nature of psycholinguistic transcription procedures.

[. . .]

Page layout

A first item to attend to in organizing and appraising a transcript is the way in which the data are physically displayed on each page. As members of a culture, we, the transcribers, bring into the transcription process a biased spatial organization. We display our data with the cultural expectation that certain items will be noticed before *others* and that certain items will be seen as part of particular units and categories (e.g., utterances, turns at talk).

Top-to-bottom biases

Across many cultures, there is a convention whereby written language is decoded from the top to the bottom of each inscription. The reading of conversational transcripts takes no exception to this norm, and, generally, the history of a discourse is unfolded in a downward direction. Utterances that appear below other utterances are treated as occurring later in time.

As our eyes move from top to bottom of a page of transcription, we interpret each utterance in light of the verbal and nonverbal behavior that has been previously displayed. In examining adult-adult conversation, overwhelmingly we treat utterances as *contingent* on the behavioral history of episode. For example, unless marked by a topic shifter (Sacks and Schegloff 1974), the contents of a speaker's turn are usually treated as in some way *relevant* to the immediately prior to turn. The expectation of the reader matches the expectation of adult speakers [see Grice, Chapter 3], and by and large inferences based on contingency are correct. These expectations and assumptions are reflected in the format in which adult conversations are typically displayed. Speaker's turns are placed below one another, as in dramatic script (from *Love's Labor's Lost*, I. xxi):

ARMADO: Boy, what sign is it when a man of great spirit grows melancholy?
MOTH: A great sign, sir, that he will look sad.

Here, for example, Moth's utterance is interpreted with respect to Armado's previous utterance. The reader makes such links as his eyes move

line by line down the page. If the reader misses a reading or has not understood an utterance, he frequently looks back to the immediately preceding line (above). Practices such as linking back (above) and linking forward (below) again reflect expectations of turn-by-turn relevance.

When we examine the verbal and nonverbal behavior of young children, important differences emerge with respect to adult communicative norms. In particular, the expectation that a speaker usually makes utterances contingent on prior talk does not match that for adult speakers. This is particularly the case in interactive situations involving a child and one or more conversational partners. Young children frequently "tune out" the utterances of their partner, because they are otherwise absorbed or because their attention span has been exhausted, or because they are bored, confused, or uncooperative.

We cannot necessarily count on an immediately prior utterance, particularly that of another speaker, to disambiguate a child's verbal act. We also cannot count on the child to signal noncontingency in a conventional manner. This means that we cannot even be certain that an utterance of a child that follows an immediately prior question is necessarily a response to that question. [. . .]

The connection between this discussion and the transcription process is that the format of a transcript influences the interpretation process carried out by the reader (researcher). Certain formats encourage the reader to link adjacent utterances and turns, whereas others encourage the reader to treat verbal acts more independently. For example, the standard "script" format described earlier tends to impose a contingent relation between immediately adjacent utterances of different speakers. Such an imposition is appropriate to the extent that it matches the conventional behavior of the speakers themselves. Such a transcript is thus far more appropriate to adult western speech than to the speech of language-acquiring children.

[. . .]

Left-to-right biases

The European culture of literacy socializes its members to encode ideas not only from top to bottom, but from left to right of the writing surface. For a page of transcription, this directionality means that within each line utterances to the left of other utterances have been produced earlier. Similarly, words to the left of other words on the same line have been uttered earlier. Leftness is linked with priority and also with inception of a statement or entire discourse.

Very close to its association with priority and inception is the link between leftness and prominence in written expression. This is clearest within the sentence in English, where subjects or topics normally appear to the left of their predicates in the declarative modality. Topics constitute the major arguments of a proposition, and subjects control verb agreement and a number of other syntactic processes.

These associations may influence the overall organization of a transcription in at least two ways. Most studies of child language involve the child interacting with just one other individual, usually an adult. In this situation, the transcriber who has opted for parallel placement of speaker turns has to decide which speaker is to be assigned to the leftmost speaker column and which to the right.

A brief review of the adult-child interaction literature indicates that, with some exception, the overwhelming tendency is for researchers to place the adult's speaker column to the left of the child's speaker column. I would like to point out here that this tendency may not be arbitrary. Rather, it may reflect perceived notions of dominance and control. That is, the researcher may be quite subtly influenced by an adult's status as caretaker or competent speaker in letting this figure assume the predominant location on the page of transcription.

The placement of the adult in the leftmost position may not only reflect but actually *reinforce* the idea of the adult as a controlling figure. How could this reinforcement come about? Recall that leftness is associated not only with prominence (e.g., placement of subject in English standard active declarative sentences), but with *temporal priority* in English-language transcripts. Each line of transcription starts at the left margin and moves towards the right. The decoding of each line as well is affected by this directionality. If the reader wants to look back at prior talk, then the eyes are orientated to the left. If the reader wants to locate the starting point of an utterance, the eyes move left until they locate the initiation of talk following a pause, interruption or final interactional boundary.

These expectations concerning where talk initiates could very well affect judgements concerning the initiation of a *sequence* of talk. A tendency for the western reader may be to turn to the left to locate such initiation points in a verbal interaction. In particular, readers may turn to talk in the leftmost speaker column as a "natural" location for opening up an interactional sequence. Looking to the right-hand column of talk, is, in this sense, a less "natural" move in the pursuit of an interactional opening.

This means that whichever speaker is assigned to the leftmost column has a better than average probability of being an initiator of a sequence of talk. In transcripts in which the adult is assigned to this speaker column, the adult becomes the more probable occupant of the initiator role.

[. . .]

Placement of nonverbal and verbal behavior

In studies of child-language development, there is an overwhelming preference for foregrounding verbal over nonverbal behavior. This is due to at least three sources:

The first and most obvious is the *goal* of the research at hand. The researcher is, after all, concerned primarily with language. Nonverbal context is usually considered to the extent that it directly relates to the utterance produced.

A second source is the *method of recording* child behavior. Child-language studies have relied upon three basic means of obtaining data: diary method, audiotape recording plus notetaking and video-tape recording. While the use of video-tape allows a relatively detailed view of nonverbal behavior and environment, it is still a relatively restricted mode of documentation within developmental psycholinguistics. [. . .]

A third source of verbal foregrounding stems from using analyses of adult communicative behavior as models. In nearly all linguistic, sociological, and psychological treatments of adult–adult speech behavior, nonverbal considerations in the immediate situation are minimized or ignored. Where nonverbal behavior is attended to [for example, see Kendon, Chapter 22], such behavior tends to be treated as a set of variables that co-occur with language but do not necessarily constitute part of the idea conveyed. By and large, the message content is considered to be conveyed through language.

One of the major advances within child language in the past decade has been the understanding of the communicative import of nonverbal behavior among young children. There are now numerous documents of the communicative skills of children before language emerges. These studies show that nonverbal behavior may be an *alternative* rather than an accompaniment to verbal behavior. Children are able to employ gesture, body orientation and eye gaze to perform a variety of communicative acts (e.g., pointing out the existence of some object, requesting some future action from the intended addressee, offering, demonstrating, etc.). The emergence of language is understood as a move away from a primary reliance on nonverbal means towards greater reliance on verbal means to convey an intention. In the course of this process, verbal means are employed conjointly with nonverbal means and *together* they convey the child's intentions.

[. . .]

A practical fact to be reckoned with is that it takes more space to represent nonverbal behavior than to represent verbal behavior. This might be minimized by a well-developed system of notation for nonverbal features. However, there are just so many features that one would want to symbolize

in code-like fashion. In the typical transcript, utterances would be surrounded by notes on nonverbal context, and the researcher would be faced with sorting out the forest from the trees in many of the analyses to be carried out.

An above or below representation of nonverbal and verbal behavior becomes increasingly unfeasible the greater the amount of nonverbal information there is to report. We should consider, before going on, the extent to which we need to deal with quantities of nonverbal data. One could argue, for example, that only in the very early stages of communication development is the detailed recording of nonverbal context critical to assessing intentionality. In looking over a great many transcripts, I see that while reliance on the immediate context lessens over developmental time, it is still the case that children continue to rely heavily on the immediate setting well into the multiword stage. This generalization holds more for certain physical and social conditions than for others. For example, where a child is talking in bed at night or in the semi-darkness of early morning, nonverbal considerations are minimized. Where a child is carrying out some activity other than talking, for example, eating, playing, in a daylight setting, nonverbal considerations take on a greater importance. Not only setting but also co-participants seem to affect the extent to which here and now is communicatively significant for the child. Whereas an adult may lead the child into discussions of past and future events, child–child interaction is rooted in the here and now. [. . .]

To understand the role of eye gaze, gesture, action and setting in peer interaction, consider the following scene, involving Toby and David Keenan at age three years and five months. While earlier months of recording involved the children interacting in their bedroom in near darkness between 6 a.m. and 7 a.m., at this time of the year, the morning light was considerably brighter. The children made greater use of stuffed animals and blankets and played in a number of locations within the room. The piece of recording which we are examining shows Toby and David sitting face to face on Toby's bed. David is sucking his thumb, holding a toy rabbit and security blanket. Toby holds a monkey wrapped inside his security blanket. Prior to the moment at hand, Toby had announced that his blanket was a *steamroller* and David had agreed. Both Toby and David are looking down at Toby's blanket. At this moment, David begins to hum, where upon Toby interrupts, saying *yeah Im gonna make car/*.

In the course of his utterance, Toby performs a series of actions. In the course of *Im gonna make* he moves his blanket and monkey to his right side. (His blanket unfolds in the process.) Between *make* and *car* there is a slight pause (a 'beat'), and in this pause, Toby begins pushing his blanket into a ball, completing the process as he utters the word *car*. Following this sequence

of actions, Toby says *heres here thats handle/*. In the course of this utterance, yet another series of actions is performed. Immediately following *heres*, Toby picks up a section of the blanket, holding it in the air for one 'beat' between *thats* and *handle*. While uttering *handle*, he pushes the section down to the bed. Immediately he says *and thats people*, picking up another section of the blanket in the space of a beat between *thats* and *people*. Following the uttering of *people*, Toby drops that section of the blanket. Actions and utterances of similar character follow.

This description indicates the amount of nonverbal data that needs to be recorded to assess the nature of reference and other speech acts carried out by the child. For example, without indicating accompanying nonverbal behavior, we would not know if *steamroller* and *car* named the same referent, and we would not know the referents for the deictic terms *heres*, *here*, and *thats*. The detailed recording of accompanying movements and eye gaze is, then, not superfluous to an analysis of communicative competence.

This description as well indicates the difficulties of integrating verbal and nonverbal behavior. It indicates the amount of nonverbal data that needs to be reported for a small number of utterances. (In the preceding description, four utterances are examined.) While the situation is reported in "prose style", it indicates the difficulties in following exactly what is happening across both nonverbal and verbal modalities when both are reported in the same descriptive space.

The situation just examined illustrates yet a further feature of nonverbal and verbal behavior that is not captured in any of the transcripts written for or by developmental psycholinguists. This feature is that of *interoccurrence*. Verbal behavior may occur one or more times in the course of some other action carried out by a participant. Alternatively, nonverbal actions may be carried out one or several times in the course of any one single utterance. [. . .] Careful observation shows that typically utterances and actions do not start at the same point in time. An utterance usually precedes or follows the initiation of some nonverbal act. For example, in the situtation reported above, the action of picking up a section of the blanket overlaps but *precedes* the utterance of *here thats handle*. Alternatively, the same action occurs in the *middle* of the subsequent utterance *and thats people*.

The initiation points of utterances and actions provide clues concerning the organization of a communicative act. For example, in the utterances treated above, the relation of verbal and nonverbal behavior differs. In the first case (*here thats handle*), verbal behavior makes reference to and predicates something about an object that is already a focus of attention. The verbal act identified an object previously indicated through nonverbal means. In the second case (*and thats people*), reference is expressed initially through

verbal means and only subsequently through nonverbal means. Here nonverbal means clarify what object is being referred to by the lexical item *that*. In these two utterances, then, nonverbal and verbal behavior may carry out different types of communicative work. [. . .]

Ideally, we want our transcript to meet practical as well as theoretical considerations. We want our transcript to express the relation between nonverbal and verbal behavior as accurately as possible: we want it to encode not only prior and subsequent behaviors, but co-occurrent and interoccurrent behaviors as well. We do not want a transcript that discourages the reader from integrating verbal and nonverbal acts. On the other hand, we want a readable transcript, one that displays clearly and systematically utterances and contexts.

One possible solution to these demands is to display verbal and nonverbal data in separate locations but to use *superscripts* to locate where verbal and nonverbal acts occur. In so doing, utterances and nonverbal information would be distinguishable, yet, through superscripting, would be integrated. Where children are young, where the setting is light (daytime), where actions are varied and frequent, nonverbal information should be given prominence. In these situations, nonverbal behavior should be reported to the *left* of a participant's verbal behavior. Both nonverbal and verbal behavior of a participant are placed within that participant's behavior column.

Table 10.1 illustrates the use of superscripting with a re-reporting of the situation outlined earlier. Certain symbols will be used to describe nonverbal actions and frames, as well as matters of timing. These symbols are explained in the following section.

Transcription symbols for verbal and nonverbal behavior

The orthographic representation of utterances will vary according to the goals of the research undertaken. Scollon's work (1976) indicates that utterances at the single-word stage should be transcribed phonetically. As the child's pronunciation approaches adult norms, use of phonetic representation should be less critical. However, there are situations in which the speech of older participants is best represented phonetically. These include instances of sound play (Keenan 1974) and instances of unintelligible speech. Furthermore, strictly standard orthography should be avoided. Rather, a modified orthography such as that adopted by Sacks *et al.* (1974) [see Chapter 9] should be employed. A modified orthography captures roughly the way in which a lexical item is pronounced versus the way in which it is written. For example,

Table 10.1 Numbered actions are explained in the nonverbal column

David		*Toby*	
Nonverbal	*Verbal*	*Nonverbal*	*Verbal*
[1]sucks thumb ⇓ Toby's blanket	(1.2)[1] mm//mm	[2]moves blanket & monkey towards rt., >	//yeah][2] Im gonna make, ([3]) car[4]
		[3]blanket, blanket reaches rest loc.	
		[4]pushes blanket into ball	heres/
		[5]picks up part of blanket	[5]here thats? ([6])
		[6]holds part, ⇓ part	[7]handle≠
		[7]pushes part down	
		[8]picks up another part, holds it	and thats, ([8])
		[9]releases part	people≠[9]

modified orthography includes such items as *gonna, wanna, whazat, yah see?, lemme see it*, and the like.

The conventions will be presented in the form of detailed tables (see Tables 10.2 and 10.3). In these tables, three types of information will be provided. First, the tables will present each behavioral property to be represented in the transcript. Second, the convention for representing each of the properties will be displayed, along with an illustration of its use. Third, the tables will briefly point out the motivation for marking this property, its significance in an assessment of communicative competence.

Do our data have a future?

The discussion of transcription and theory presented here is to be taken as a first venture into a vast wilderness of research concerns. Many issues have not been addressed. Furthermore, certain transcription conventions invite modification by others with expertise in the field. [. . .]

Table 10.2 Verbal transcription

What to Mark	*How to Mark*	*Why*
1 Utterance boundary	/placed at end of utterance example: *don't make ears funny/he cry/like that/*	Utterance = basic unit in assessing and measuring communicative development.
2 No gap (latching)	=placed between utterances with no time gap example: *look ≠ look ≠ look ≠ look*	Utterances should have a single intonation contour and single breath group, but there are cases in which more than one intonational contour appears in single breath unit. Each contour may correspond to an informational unit. To mark contours linked in this way (no gap), we use "latch marks" (=).
3 Pause length	(.3) placed before utterance; utterances separated by significant pauses should be placed on separate lines example: *and/lettuce/man's eating lettuce/* (5) *one day/was little rabbit/called Lucy/* (.) indicates very slight pause example: *gonna (.) throwit (.) fields/*	a partly defines utterance boundary b partly define "turn" (turn = utterance bounded by significant pause or by utterance of other participant) c number of utterances per turn may be measure of control d may signal end of topic sequence or propositional sequence e may signal leavetaking of floor, elicit feedback from next speaker f may signal distress (cognitive, linguistic, disagreement)

Table 10.2 continued

What to Mark	*How to Mark*	*Why*
4 Overlap	//placed at beginning of overlap,] placed at end of overlapped utterances overlapped utterances go on same line example: A *steamroller's stuck //now]/* B *oh dear] dear /*	a like pause length, indicates sensitivity to turn and utterance units; may show child or caretaker sensitivity to informational units b frequency and placement of overlap may be variable in caretaker speech c may be important in assessing cultural differences in language socialization
5 Self-interruption	– placed at point of interruption example: *want some – all of it/*	a may reflect trouble spots in interaction (see 3); trouble can be cognitive, sociological, etc., e.g., can't get reference established, can't get attention of addressee b extent to which speaker can reformulate utterance indicates ability to (1) self-correct, (2) paraphrase
6 Intonation prosodic quality	, marks low rise ? marks high rise . marks low fall (only use in adult speech) ! marks exclamatory utterance place, ?.! at end of utterance capital letters mark increased volume: example: *YOU SILLY/* _____ marks stress	a may mark new information b may mark hearer selection (e.g., self or other, human versus toy, etc.) c may mark communicative act d may mark utterance boundary

Table 10.2 continued

What to Mark		*How to Mark*	*Why*
		example: *I want <u>that</u> one/.* : : : marks lengthened syllable (each : = one "beat") example: *hello::/* (( )) marks other voice qualities, e.g., ((LF)) laugh ((WH)) whisper ((CR)) cry ((WM)) whimper ((WN)) whine ((GR)) grunt	e : : : may be tied to marking of aspect
7	Audible breathing	-h marks in-breath h marks out-breath (h) marks laughter	a may indicate utterance boundary b hesitation marker
8	Metatranscription marks	( ) unclear reading, no hearing achieved (cow) tentative reading X/repetition of prior utterance, e.g., *no/X/X/*	

Table 10.3 Nonverbal transcription

What to Mark	*How to Mark*	*Why*
1. Changes in gross motor activity	Bloom *et al.* (unpublished manuscript 1974) suggest using present progressive tense to describe action simultaneous with utterance. Use simple present tense to describe action prior or subsequent to utterance. Put action prior to utterance on line above, simultaneous action on same line, and subsequent action on line below utterance (in nonverbal column). If using videotape, mark precise overlap of action and utterance with a superscripted number above point in utterance. example: *[want^{1}cow/^{1}grabs cow* If superscripts are used, use only present tense to describe action, as simultaneity is otherwise marked.	a aids in determining reference and predication b aids in interpretation of communicative act (self-description, refusal, etc.) c aids in interpretation of interactional sequence (nonverbal means of accomplishing 1st or 2nd part of a sequential pair d provides information linking utterance to change of state or change of object. e indicates child's understanding of or ability to express tense of aspect.

Table 10.3 continued

What to Mark	*How to Mark*	*Why*
2 Eye gaze	⇑ looks up (+target: +name) ⇓ looks down (+target: +name) example: ⇑M (use initial for person) ⇓ car > towards right < towards left ▽ facing camera △ back of head to camera example: >⇓ M (looks down towards right of monitor screen at mother)	a indicates intended addressee, referent b indicates extent to which child attending c indicates extent to which speech is planned
3 Gestures	PT pointing R reaching HD holding up TG tugging OF offer	a primary means of reference b indicates communicative act (e.g., summons, offer, description)
4 Body orientation	⊂ marks direction of pelvis (bird's eye view) example: A (A and B are facing ∩ each other; A's body ∪ is facing camera, B's B back is to camera)	a provides social "frame" for talk and action b indicates extent to which participants engaged in focused interaction

A greater awareness of transcription form can move the field in productive directions. Not only will we able to read much more off our own transcripts, we will be better equipped to read the transcriptions of others. This, in turn, should better equip us to evaluate particular interpretations of data (i.e., transcribed behavior).

Our data may have a future if we give them the attention they deserve.

References

Bloom, L., Lightbrown, P. M. and Hood, L (1974) 'Conventions for transcription of child language recordings', unpublished manuscript, Teachers College, Columbia University.

Borger, R. and Cioffi, F. (1970) *Explanation in the Behavioural Sciences: Confrontations*, Cambridge: Cambridge University Press.

Keenan, E. Ochs (1974) 'Conversational competence in children', *Journal of Child Language* 1: 163–85.

Popper, K. (1959) *The Logic of Scientific Discovery*, New York: Science Editions, Inc.

Sacks, H. and Schegloff, E. (1974) 'Two preferences in the organization of reference of persons in conversation and their interaction' in Avison, N. H. and Wilson, R. J. (eds) *Ethnomethodology: Labelling Theory and Deviant Behavior*, London: Routledge & Kegan Paul.

Scollon, R. (1976) *Conversations with a One Year Old*, Honolulu: University of Hawaii Press.