Transcription and Coding for Child Language Research: The Parts are More than the Whole

Lois Bloom
Teachers College, Columbia University

1. INTRODUCTION

People have been writing down what infants and young children say at least since Darwin (1877). Two parallel lines of development in this time have influenced how we transcribe the data of children’s language for understanding how language is acquired. One of these developments has been conceptual: the different questions that researchers asked have influenced the sorts of data that were collected. At the same time, electronic innovations have provided increasingly more sophisticated equipment to supplement (but not replace) paper and pencils. The purpose of this chapter is to (a) discuss these conceptual and procedural developments that influence contemporary observational research in child language, and (b) describe the rationale and procedures for the computer-assisted transcription and coding we do in my own laboratory for the study of early language acquisition.

2. OBSERVING AND PRESERVING THE DATA OF CHILD LANGUAGE

Conceptual and procedural influences on methods of research are not independent of one another. There is an old adage: “When the only tool you have is a hammer, everything looks like a nail.” The tools that we have determine the way in which we approach a task and also determine the sorts of tasks that we consider feasible (Beckwith, Bloom, Albury, Raqib, & Booth,
the length of sentences and whether they were complete or incomplete, simple, compound, or complex, and so forth. (See McCarthy, 1954, for a thorough review of this literature and Templin, 1957, for what is no doubt the best and probably the last effort in this era of research.)

These studies were successful in that they provided the “developmental milestones” that are still widely used in pediatricians’ offices, day-care settings, well-baby clinics, and the like. Moreover, we learned certain facts about infants’ speech in the first 5 years of life that have endured and are still referred to in contemporary research. However, in the 1950s, Roger Brown began a program of research in which he pointed out that the really important questions had to do with the developments in children’s knowledge that produce the changes in what they say (e.g., Brown, 1957). That knowledge includes the grammar of language.

Linguists, at least in this century, had always operated with the assumption that language was rule-governed. But the galvanizing influence from linguistics on the search for the child’s rule system was the theory of generative transformational grammar (Chomsky, 1957). The presumption was that children are learning a grammar when they acquire language, and the grammar they are learning is a transformational one. The search for child grammar in the studies that followed sought evidence of grammatical rules from the regularities in early two-word speech (e.g., Braine, 1963; Brown & Fraser, 1964; Miller & Ervin, 1964). The lasting insight in these studies was that the early word combinations of children are, indeed, systematic. A small number of words are used frequently and in relatively fixed position, with a larger number of other words, each of which occurs relatively infrequently and without predictable word order. However, the resulting ‘grammars’ were descriptions of the regularities in frequency and word order in children’s speech. Even though the goal of research had been to uncover the child’s underlying rules of grammar, the data that were used in the endeavor still consisted of only the spoken word.

2.2. Hidden Phenomena

A linguistic fact assumes significance in relation to its “element of experience . . . content or ‘meaning’” (Sapir, 1921, p. 10). In 1968, I proposed that the meanings children express in their language learning efforts determine what they acquire of the grammar of a language. These meanings have to do with what the young child has learned and is learning about objects, events, and relations in the world:

A young child’s success in learning to talk depends on [the] ability to perceive and organize the environment, the language that is a part of the environment, and the relation between the two . . . Children learn to identify certain

The research project that is referred to here is a longitudinal study of the development of 14 infants from 8 months to about 30 months of age. The infants were first born, 7 girls and 7 boys, from varied ethnic and economic backgrounds in the New York area. Each infant and mother visited our laboratory playroom every month and were visited in their homes each month until they were 15 months old. This research has been conducted with funds provided by The Spencer Foundation and The National Science Foundation, for which we are grateful. This chapter was drawn from material in the forthcoming book, Acquiring the Power of Expression: Consciousness, Cognition, and Emotion in Transition from Infancy to Language (Bloom, in press).
grammatical relations and syntactic structures with the environmental and behavioral contexts in which they are perceived and then progress to reproducing approximations of heard structures in similar, recurring contexts. (Bloom, 1968/1970, pp. 1, 233)

If children acquire meaning from events in the context, then the context ought to be a preeminent source of information about the meanings that children express. I suggested, therefore, that we attend to and use information from the context of children's utterances in our efforts to learn how and what children are learning about language. This meant going beyond the spoken word and inquiring into the underlying meaning of the words as inferred from what children talk about. The suggestion was taken up by Roger Brown, who coined the term rich interpretation for what became the dominant method for pursuing the meanings hidden in early speech (Brown, 1973).

Rich interpretation is not without its detractors. In the mid-1960s, when I was doing the original research on which my 1968 dissertation was based, no one wanted to admit evidence about meaning into the study of language. In linguistics, from Bloomfield through Chomsky, the domain of meaning had been avoided like the plague. And still today we have those who feel that what is hidden in the child is not admissible evidence for the language acquisition enterprise. Nevertheless, we do know that individuals provide us with a variety of signals that let us know what they are thinking and feeling, and what individuals think and feel is what they express in language (Bloom, in press; Bloom & Beckwith, 1986). We make use of these kinds of signals quite readily and easily in all of our everyday interactions with one another. Indeed, once context was introduced into the study of language acquisition, not only the meaning of an individual utterance, but developments in pragmatics and the unfolding of discourse between individuals could be studied by attention to "language in context" (Bates, 1976). Moreover, the contexts of language acquisition include not only the immediate circumstances that surround acts of expression and interpretation, but, indeed, the larger cultural world view of the individuals in a society (Schieffelin, 1979).

In sum, contemporary research that takes a developmental perspective is conducted with the fundamental assumption that language is acquired in connection with other developments and events in the life of the young language-learning child. Children acquire the sounds, words, structures, and discourse processes of a language against a background of other cognitive, social, and affective developments. In short, we now recognize the importance of paying attention to a great deal else beside the spoken word in our efforts to understand language development. This recognition has paid off in that we know much more than ever before about the language learning process. However, we have also had to deal with important conceptual issues concerning our methods for deciding what we use as data and how.

2.3. Selectivity

Individuals observe and interpret what children do every day, by necessity, in order to interact with them, and they rarely think about it. But as researchers we have to establish a certain distance from what children do so that we can think about it, describe it, and, hopefully, contribute to explaining it. One reasonable goal might be to approach the task without regard to any expectations that we might have. "On the observational level, the main—one could say the only—rule is that all facts would be carefully observed and described, without allowing any theoretical preconception to decide whether some are more important than others" (Levi-Strauss, 1963, p. 272).

However, our preconceptions cannot help but create and influence our expectations, as many people have pointed out.

For both logical and practical reasons, there can be no such thing as pure observation... one's ideas evolve with one's research, reading and thinking. ... [Trying to put oneself at sufficient distance for clear vision is like trying to leap over one's shadow. ... [Accordingly, we have to] start out with selection of one out of an infinite number of possible descriptive strategies, in accordance with whatever one's wits and experience offer as the best bet (Beer, 1973, pp. 49, 54).

We are then, as researchers, the products of our own intellectual histories, and these cannot help but influence our view of the evidence. Such selectivity, on the whole, is not only to be expected, but even encouraged, as Ochs (1979) pointed out. How else is our current and future work to benefit from what we already know about children, language, and the acquisition process? However, the researcher needs to be aware of the filtering process and take such inherent selectivity into explicit account. Moreover, "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 videotape" (Ochs, 1979, p. 44).

As soon as we make a recording we have begun a process of data reduction, which is another sort of selectivity. Whether recorded by hand or by audio or video electronics, something is necessarily left out of the record. The microphone and the camera, much less the eye, the ear, and the hand, can never preserve the detail, nuance, and complex circumstances of events. Transcription reduces electronically recorded data even more drastically and provides a serious constraint on the available information. Quite simply,
copying the richness of tone and detail that can be preserved on tape, as reduced as it is from the original event, is an impossible task. The process of transcription, then, provides the real moment of truth for the observer.

In sum, transcription presents two problems. The first is the set of biases and distortions that creep in because of the necessarily selective view of the observer. The second is the massive data reduction that results from the sheer physical limitations, through no fault of their own, of electronic devices that make the record to preserve the data and of persons who do the transcribing. The system of transcription I describe below was created in the effort to minimize these problems.

2.4. Separation of Covaribles

In effect, a rich interpretation depends on "lean transcription." Not only is it hopeless to attempt to capture everything on a recording, but too much detail in a transcript produces clutter and distraction (Ochs, 1979). Further, the transcription must, necessarily, aim to represent a description of events, rather than an interpretation (Bloom & Lahey, 1978). Admittedly, the bounds for knowing where description ends and interpretation begins are probably not definable. Indeed, one might well say that any description is by nature a form of interpretation. The point is, however, to aim at preserving the data in such a way as to allow for different analyses and different resulting interpretations. The following example is trivial, but only on its surface:

A 1-year-old picks up a small block, says more, puts it on top of a bigger block, smiles, and looks at her mother, who smiles back.

This event might be interpreted in many ways. At the minimum, we might want to say something like "initiates activity," "builds a tower," "expresses recurrence," or "expresses pleasure" to capture what happened. However, these or any other evaluations of the event must come from what we have preserved in our transcript after the fact. If all we recorded in the transcript was something like "initiates activity" or "expresses recurrence," we would not be able to use the transcript to make other interpretations. For example, the timing of the baby's smile, in relation to saying the word more and the actions with the blocks, is highly relevant to several theoretical concerns. The baby's smile could be interpreted as, again at the minimum, an act of social referencing or, alternatively, a smile of recognition or accomplishment. The fact that the baby did not smile before or at the same time as saying more is relevant to understanding the way these two systems of expression, affect and speech, come together in the single-word period (Bloom & Beckwith, 1989). In other words, if one transcribes a piece of behavior only according to its meaning, its function, or its effect, then the information about the event that might contribute to an additional or an alternative interpretation is lost.

When we began to use audiotape recorders, and our research was confined to only the spoken word, we worried about the accuracy with which we represented what was said. For example, we had to decide whether to transcribe orthographically or phonetically, and if phonetically, how broad or narrow a phonetic transcription. When we began to include information from the context in the transcription so that we could interpret something about meaning, we took notes at the time of recording, and then tried to fill in the contextual details surrounding what was said at the time of transcription. The introduction of the video recorder seemed, at first, to solve all our problems. We could use the video data instead of taking notes or trying to remember events and circumstances in the context. Our first video transcriptions were, in effect, modeled after our audio transcriptions and differed only in the added detail (e.g., Bloom, 1973; Bloom & Lahey, 1978).

However, a video record presents other sorts of problems. The amount of detail that is preserved on video tape is, quite literally, enormous, even though the information present in the original event is necessarily reduced in the sense described above. This detail can be overwhelming to the researcher in trying to decide what to include and what to leave out of the transcription. In the effort to preserve a description in the transcript rather than an interpretation, we could easily be engulfed by the details.3

In our current research, we have not attempted a full transcription of all relevant behaviors and accompanying contexts, which has been the standard operating procedure with audio and video data. Instead, we have pulled the video record apart, exploded it, so to speak, and separated out the variables of interest according to one or another research question. The advantage of having the video record is that it is impervious to these operations that we perform on it. This was pointed out as long ago as 1935 by Gesell, who pioneered in using film to study infant development:

The behavior record becomes as pliant to dissection as a piece of tissue. Any phase or strand of behavior may be exposed to view. If the view is an intricate one it may be repeated numerous times without in anyway damaging the original record. Here the dissection of behavior forms has a striking advantage over anatomical dissection. Bodily tissue suffers from the scalpel, but the integrity and conformation of behavior cannot be destroyed by repeated observation. A behavior form can be dissected over and over again in

---

3The level of detail preserved on video can also be seductive in leading the transcriber down one or another garden path. The effect of watching a small scrap of an interaction over and over again is to reveal some incredibly fine details. The result is often a severe narrowing of focus and excessive attention to what may turn out to be, in the final analysis, irrelevant.
increasing detail without loss of form. (Gesell, 1935, p. 6; quoted in Beckwith et al., 1985)

Thus, more than 50 years ago, Gesell anticipated the way in which we have exploited the technology available to us today. We can separate the variables for examination to study the relationships between them in many different ways without sacrificing the integrity of the original record.

If we return to the mother–infant episode above, we can see what this means. Our research project concerns developments in cognition and affect expression in relation to the transition from infancy to language. The relevant variables in our research so far have included child speech, mother speech, the situation accompanying mother speech, child affect expression, mother response to child affect expression, child object play, child object search, and so forth. Certain of the questions we ask have centered on object play as a window on developments in cognition (e.g., Lifter & Bloom, 1989), and on affect expression in relation to the emergence of expression through speech (e.g., Bloom, Beckwith, Capatides, & Hafitz, 1987). With data such as in the example above, we separate the covariables in independent passes through the video record and transcribe or code (a) child speech (more), (b) child affect expression (a smile with +1 intensity), (c) mother response to child affect expression (a smile with +2 intensity), (d) child object constructions (putting one block on another), and so forth. Independent coders are assigned to transcribe or code only one variable at a time. For instance, one person would only transcribe child speech; another would only code child affect expression. The beauty of the system is that we can go on to code or transcribe however many variables we would need, bounded only by the conceptual issues that we pursue (and the vagaries of available funding).

3. A PLAN FOR COMPUTER-ASSISTED TRANSCRIPTION

3.1. The computer system

Because the variables in the stream of activity do, in fact, covary, we had to have a way of putting them back together again. This is the distinctive feature of the system. The system we devised uses state of the art (circa 1981) equipment and is schematized in Fig. 6.1. The hardware is either still commercially available or reproducible in some other way. At the time of the original observation, the audio signal from the interaction is recorded on one sound track of the stereo video tape while a computer-readable audio time code is recorded on the other. Videotape runs at 30 frames per second and the timecode generator (SMpte FOR-A) lays down a unique discrete audio signal for each frame, that is, 30 times every second. At the time of playback, the videodeck (Sony stereo Betamax) is interfaced with a microprocessor (Apple II Plus) via a timecode reader. This entails a simple multiplexing circuit made to connect the 32-bit timecode readers with the 8-bit Apples.

The Apple controls the videodeck, turning it on, moving it forward or backward, slowing or speeding the playback, freezing the frame, and so on. More important, the Apple can also read the timecode, giving it the ability to find any particular frame in the recorded observation.

A coder sits at the Apple keyboard watching the video monitor and decides when a behavior occurs that is relevant to the variable being coded. The coder uses the keyboard to stop the playback and then must decide on which frame the behavior begins and/or ends. This is done by telling the computer to move the videotape forward or backward a specific number of frames until the desired frame is targeted. Accuracy in determining onset and offset times, after training, is remarkably high. For example, the mean discrepancy between pairs of independent coders was 2 video frames (.07 second) for speech onset time and 5 video frames (.17 second) for speech offset. The accuracy in finding onset time of an affect expression was somewhat less due to the fact that several kinds of cues were used to code affect (i.e., facial expression, body tension, affective vocalization). The mean discrepancy in

---

*This section contains material presented originally in Beckwith, et al. (1985).*
locating affect onset was 16 frames (.50 second). (The offset of an affect expression was the onset of the next expression because affect was coded as a continuous variable.)

When the onset or offset frame is found, the coder enters an edit mode and types in the data. Data can be entered as a string of up to 255 characters and can include anything that can be typed. The four columns on the computer monitor in Fig. 6.1 (here labeled 001, 002, and so forth) represent any four of the 135 separately coded variables that the system is capable of handling. At the time of transcription, the researcher can call up any three variables in order to code a fourth variable. For instance, in coding the situation that mothers’ sentences were about (Beckwith, 1988), the coder asked the computer to locate each mother sentence in column 002 and then typed in the situation code in column 050.

Once entered, the data are stored in files that are ordered sequentially according to time code. An extract from a data file is presented in Fig. 6.2. Each line in the file represents one record. Each record has three fields. The first field is a hexadecimal number (two columns) that identifies the particular variable coded (child speech is 01, mother speech is 02, and so forth). The second field in the record is a hexadecimal number (six columns) that represents the frame count for hour, minute, second, frame. The third field consists of text of variable length (up to 255 characters). This text may be transcription of speech or any one of a number of codes that we have devised for representing the information from different data variables.

The coded data can be manipulated to (a) generate a hard copy transcript for traditional sorts of descriptive analyses and (b) perform a variety of quantitative analyses for statistical treatments.

3.2. Generating a Transcript

Because the data are stored electronically, they are flexible for manipulation before output. Thus, transcripts can be set up in any number of ways with different combinations of the separate coding passes. Moreover, because the computer can read the timecode and every coding decision is associated with a time of onset and/or offset, the separate codes can be merged sequentially. The merging produces an integrated account with the relevant behaviors lined up according to the original temporal relations between them. An example of such a computer-generated transcript is presented in Fig. 6.3 using the data from the example in Fig. 6.2. The codes represented in the data displayed in the two figures are explained briefly in Table 6.1.

---

![FIG. 6.2. Data file example. (Note: The duration of this example is 15 seconds, 13 frames. See Table 6.1 for a description of these codes.)](image)
Table 6.1
Explanations of Example Codes in Data File and Transcription

19 CHILD SPEECH CONTENT
19 005325 v:a-go(bff in (slide-in))>d/n
=a child volition (v:) that is achieved (a) to put the boy family figure (bff) into (in) the interior space of the slide (slide-in); said during the action (d) and not directed (n) to the mother.

05 CHILD OBJECT PLAY
05 005357 CO-SN-IM-GE-NA-NA
=a construction (CO), without support (SN), that was an imposed (IM), general (GE) relation; (NA) = not applicable

04 CHILD SEARCH
04 0054D5 LO-XC-NA-NA-MR-AR-GE
=locating one object (LO), in order to construct a relation (XC), which mother had suggested (MR), when the child was attending to another object (AR), and the resulting relation was a general one (GE); (NA) = not applicable

14 CHILD AFFECT
14 005342 N
=neutral affect expression; only onset times coded because affect expression was coded as a continuous variable (onset of an expression = onset of the previous expression)

14 CHILD AFFECT
14 0054AC +1
=affect expression with positive valence and low level intensity

32 MOTHER SENTENCE/SITUATION
32 0054DD (GO)(ENTITY = mff&fff&cff)(CAUSE = child)(SRCE = NIL) (PATH = NIL) (DEST = slide-in)
=the child (CAUSE) moving (GO) the mother, father, and child family figures (ENTITY = mff&fff&cff) to the interior of the slide (DEST = slide-in); (SRCE) = source; (PATH) = path; (NIL) = not applicable
This example of a transcript reproduces the five separate columns containing (a) transcription of child speech, (b) the code for the underlying cognitive representation attributed to child speech, (c) mother speech, (d) the situation code for mothers’ sentences, and (e) child object play. Each entry in each column is associated with the times of onset and offset (for child and mother speech) or the times of onset (for the two codes associated with child and mother speech and for object play). The five columns are integrated with one another sequentially according to the timestamp. This restores the temporal relations that existed among child and mother speech and child object play at the time of the observation in the original video recording.

A transcript such as this one can be used in much the same way as a traditional transcription. We can examine the unfolding of events in regard to one or another research question, perform descriptive analyses, generate new hypotheses, and so forth. In addition, however, because the data are electronically stored, the computer can also perform many of the data analyses we need.

### 3.3. Quantitative Analyses to Assess Qualitative Domains

The coded data are transferred from the Apples onto an IBM-PC and stored in a standard form—delimited ASCII—that allows access from any number of different programs. Most recently, we have developed software for converting the data into database format using DBASE+ (Ashton-Tate) run by the Clipper Compiler (Nantucket). This allows us to run a variety of data manipulation programs.

Most simply, the different codes within a variable can be counted to look at relative frequencies and contingent probabilities. We know, for example, as a result of several such analyses, how frequently our subjects expressed positive and negative affect; the relative time spent in nonneutral, positive, and negative affect expression; and their average duration (reported in Bloom, Beckwith, & Capatides, 1988). We also know how the children’s ages at certain language milestones (a) correlate with these measures of affect expression (reported in Bloom & Capatides, 1987), and (b) correspond to developments in object play (reported in Lifter & Bloom, 1989). We have looked at the relative frequency of expression of different kinds of meaning content through affect and words in the single-word period (reported in Bloom, Beckwith, Capatides, & Hafitz, 1987). In studies of later developments in the children’s multiword speech, we have reported how they acquired the distinction between transitive and intransitive verbs (Rispoli & Bloom, 1988), and the correspondence between the sentence types in mother and child speech (Beckwith, 1988). In still other studies of how the mothers in our sample have influenced their children’s development, we have studied the mothers’ behaviors that surround their children’s affect expression (Capatides, 1990) and play with objects (Harner, Bloom, & Gronell, 1991).

But, in addition, we can look very carefully at the timing of one behavior relative to another. Thus far, we have done this for expression through affect and speech and been able to show how the two systems—one of which (affect) has been in place since early infancy and the other (speech) is just emerging—come together in the single-word period (reported in Bloom & Beckwith, 1989). The system would also allow us to look at this contingency in the multiword period, as well as other contingencies, such as between mother speech and child speech and gaze. Most recently, we have looked at the profiles of speech and affect expression in the moments before and after episodes of object play (Bloom, Tinker, & Beckwith, forthcoming).

To be sure, very many qualitative phenomena will always resist such quantitative treatment. This is to be expected whenever we delve into the domain of human expression. Nevertheless, in our efforts to understand how young children acquire the power of expression we must, of necessity, make the relevant phenomena accessible for study. This means recording the phenomena so as to preserve them in the first place, and transforming the recording for the analyses that we do. The plan for computer-assisted transcription described here has made use of the technology that is currently available in order to respond to the problems of selectivity in transcription cited earlier.

### 4. CLOSING THE GAP

This chapter has continued the dialog, begun in Bloom (1974) and taken up by Ochs (1979) and Beckwith et al. (1985), concerning the accountability of evidence in child language research. The problems of selectivity in transcription will always be with us. However, we believe that we have succeeded in making at least modest progress toward closing the gap between an act of expression and the record we make of it in our efforts to understand it.

On the face of it, separating the covariables that are contained in an act of expression in order to attend to only one variable at a time would seem to be a radical reduction of the data indeed. However, separation of covariables is not the same as isolating the variables, because we necessarily attend to whatever surrounds the target variable in making coding decisions regarding it. Furthermore, separating a variable for the purpose of coding or transcription assures us that we have preserved the integrity of that particular variable beyond what would be possible if we tried attending to many or even several aspects of behavior at the same time. The effect of narrowing the focus in this
way is to enlarge the view we have of the target variable, enabling us to see it more clearly and consider it more carefully. In addition, transcription problems having to do with training and reliability are far more manageable than they would otherwise be. But most important, we can always add variables. Our transcript is never limited by whatever questions or concerns motivated our research at one particular time. Thus, separation of covariables reduces the data only to enlarge our view of the evidence.

Two other factors contribute to our confidence that we avoid at least some of the susceptibility to investigator bias inherent in observational research. The use of coders and transcriptioners who are uninformed as to the research questions and hypotheses in one or another study reduces the threat that we will tend toward interpretation rather than description in the decisions that are made during transcription. Another factor is the delicate balance in confidence we have between persons and machines for making decisions regarding timing in events. Persons make the original decisions regarding onset and/or offset of a particular piece of the action. But machines are far more adept than persons at putting the pieces of the action back together again.

In sum, the system for coding and transcription described here is an example of how conceptual developments in the field and developments in technology came together in the last two decades of child language research. Studying language in context, and studying the development of language in the context of other developments in the child, require that we preserve far more than just the spoken word in the record we make of the data we collect. Nonindustrial video recording became available at just about the time that context was introduced into the study of child language in the late 1960s. A video record confronts us with an overwhelming amount of information, even recognizing how much is lost in the inevitable reduction of the data that occurs through recording. The resulting problems of selectivity for transcription have begun to seem more manageable with the development of the microprocessor and personal computers in the last 10 years. We wonder how the ways in which we study children's language might change in the 21st century as a result of what we learn from our present research and the developments in technology to come.

REFERENCES


Appendix A.1
CONVENTIONS FOR TRANSCRIPTION OF
CHILD LANGUAGE RECORDINGS

1. All speech by the child and all speech to the child or within the child's hearing is fully transcribed on paper divided by a vertical line. Utterances by the child appear on the right side. Utterances by other speakers appear on the left. The person is identified by an initial (M for Mommy, L for Lois, D for Daddy, etc.). Information about the situational context also appears on the left and is enclosed in parentheses.

(M takes cookie from bag; offering it to A)
M: Look what I have/
(A taking cookie)
cookie/

2. An action or event that occurs simultaneously with the child utterance appears on the same line with that utterance.

(E banging blocks together) crash/

3. When an utterance precedes or follows an action or event, the utterance appears on the preceding or succeeding line.

(E throws block) no block/
(E picks up another block) more/

4. Note the differential use of verb tenses in describing the situations: progressive for simultaneous action; simple present for actions or events that precede or follow an utterance.

5. For situational information accompanying utterances by someone other than the child, use the same verb tense conventions, but utterances and description can, of course, succeed one another on different lines since there is rarely enough space to put both on the same line.

6. Utterances that succeed each other immediately—WITH NO CHANGE IN SITUATION—follow each other on the same line.

(G reaching for box of cookies) more/ more/ cookie/
If there is any change in situation, the utterances appear on different lines.

(G reaching for box of cookies; taking box off counter; reaches in; pulling out cookie) cookie/
When in doubt about the situational context, use separate lines.

Punctuation

7. For utterances of child and other speakers, the usual sign of utterance boundary is a slash (/). The boundary is determined by length of pause before the next utterance and by its apparent terminal contour. The judgment is sometimes very difficult to make. With older children and adults, the slash may be considered equivalent to a period, but it is important to make each judgment carefully and as objectively as possible.

8. Utterances by adult or child may be followed by an exclamation mark. When a child utterance is exclamatory, it should be followed by both an exclamation mark and the usual slash.

(Peter takes tire off car) there! finish/

9. Adult questions are indicated by question marks. For the child utterance, however, there are two different ways of indicating that an utterance has question form. For Wh questions, a question mark may be used.

(P looking in toy bag) where's a car?/
When a child utterance seems to be a question because it has rising intonation, it should be followed by a rising arrow (↑) instead of a question mark.

(K meeting L at door) did you bring the toys today ↑/

1 Prepared in collaboration with Lois Hood, and Patsy Lightbown.
In either case, a slash should also be used to mark the utterance boundary clearly.

10. A pause within an utterance is indicated by a dot (·).

(E trying to fit peg in hole) put this one in/

11. A long pause between utterances within the same general situation is indicated by horizontal dots across the center line.

(P tries to get wheel on car) wheel goes in there/

(P succeeds) there!

12. A long pause between utterances where there is a change in the general situation is indicated by three vertical dots on the center line.

(G trying to stack blocks) Gia make a house/

(G running to kitchen) juice!/Gia drink juice/

13. A colon is used to indicate that an utterance or word is drawn out.

(E trying to fit large block inside small one) no:/

14. A curving arrow is used when there is some kind of utterance boundary, but the utterance sounds unfinished, such as when the child is counting or "listing".

one→/ two→/ three→/

15. Stress marks indicate strongly emphasized words.

L: Do you want this one? (L giving G a blue disc) no!/ (G reaching for red one L is holding) that one/

Capitalization

16. Names are capitalized. Initial letter of child utterance is not. Initial letter of adult utterance may be.

Other "Punctuation"

17. An utterance may be followed by falling arrow (↓) when it is important to emphasize the fact that the utterance had falling terminal contour.

(P looking in toy bag; pulls out tire for car) wheel ↑ /

18. When a child or other speaker suddenly interrupts their own utterance—apparently leaving the utterance unfinished—a line (____) indicates the abrupt stop.

L: Do you want some ____/

(E picks up cup and spills juice)

19. When a child or other speaker interrupts their own utterance apparently to change or correct it, a "self-correct" symbol (sic) is used.

L: Those are your /, my toys/ don't /, a want toys/

20. An unintelligible utterance or portion of an utterance is indicated by three dashes (---). If possible, a phonetic transcription is used instead.

(E pushing over house of blocks, making loud crash) no more/---/ house/

Abbreviations

21. When a child or other speaker repeats their own utterance completely and exactly, an X is used to show the repetition. Any change in the utterance must be indicated, including clear changes in intonation.

L: Be careful/ X/

(P touching tape recorder) open/ X/ X /

XI

22. When an adult repeats a child’s utterance, an equal sign (=) is used to show the repetition. When a child repeats an adult utterance, however, the child’s utterance is written in full, even if the repetition is exact. An equal sign can never represent a child utterance, although an equal sign may be placed next to the utterance to indicate that it is a repetition of an adult utterance.

M: =/ I only see one in there/ two cookies/

23. The symbol # may be used to indicate that there is material on the tape that is not transcribed. It can only appear on the left side and usually represents conversations between adults. The symbol is only used when it is reasonable to assume that the child is not attending to or, in fact, does not hear the conversation.

24. (If) laugh (wh) whisper (cr) cry (wm) whimper (wn) whine (y) yell (gr) grunt

These abbreviations may be useful for behavior that occurs fairly frequently. The abbreviation should appear on the left side of the line.
Labeling

25. Pages should be numbered front and back, with numbers in upper right corner.

26. In order to make it easier to locate material on the tape, a number should be placed in the right margin every time the counter on the tape recorder registers a multiple of 50.

27. Every time a new tape or a new side of a tape is started, the tape number, side number (1 or 2) and the date and time of the recording session (if different from the previous tape or side) should be indicated.

Appendix A.2
ADDENDUM TO CONVENTIONS FOR TRANSCRIPTION OF CHILD LANGUAGE RECORDINGS: PROCEDURES FOR TRANSCRIBING VIDEO RECORDED DATA

The aim is to transcribe all speech by the child (column 3) and to the child or within his or her hearing (column 2), to describe the accompanying nonverbal behaviors of the child (column 4) and of other speakers (column 1), and to record information about the context in which the verbal and nonverbal behaviors occur (column 1). The format is illustrated below. Persons other than the child are identified by an initial (e.g., D for Debbie, J for Jerry). Descriptions of nonverbal behaviors and context are enclosed in parentheses.

<table>
<thead>
<tr>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Column 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context and behaviors of other speakers</td>
<td>Utterances of other speakers</td>
<td>Child utterances</td>
<td>Child behaviors</td>
</tr>
<tr>
<td>D giving block to S</td>
<td>D: here's the block/</td>
<td>block/</td>
<td>(S accepting block)</td>
</tr>
</tbody>
</table>

Recording Nonverbal Behavior

1. Nonverbal behaviors consist primarily of actions and gestures. The abbreviation [gs] is used for gestures followed by a description of the type of gesture, for example,

   ([gs] arms straight up over head)

Facial expressions may be included, but are optional.

1 Prepared by Peggy Miller, Lorraine Rocissano, Karin Lifter, and Ellen Tanouye.
2 The following includes frequently used descriptive terms.
   (1) Inspect—to explore visually.
   (2) Examine—to explore visually and manually.
   (3) Extend—to hold out arm at shoulder level, with or without an object; often an object is extended toward someone in a "display" or "showing" gesture.
   (4) Give—to transfer an object from the giver's hand to a recipient; "give" is distinguished from "extend" as follows: (a) an object is always involved; (b) the action is "completed"; that is, another person received the object.
2. The temporal ordering of utterances and behaviors is represented in the following ways.

a. Utterances or behaviors occurring simultaneously appear on the same line; two or more simultaneous behaviors performed by the same person appear on the same line, separated by commas.

b. Utterances or behaviors occurring successively appear on successive lines.

(M entering room) M: Where's your truck?/ on table/ (putting block aside) (approaches table, smiles)

3. Direction of the child's gaze is noted whenever possible, using the following set of symbols.

a. Gaze directed toward an object: \( \text{doll} \).

b. Gaze directed toward a person: \( \text{D} \), where D is the person's initial.

c. Gaze directed toward camera: \( \text{O} \).

d. Shifting gaze: \( \text{S} \).

e. Cannot be determined: \( \text{?} \).

Each utterance is followed immediately by one of the above symbols, indicating the direction of gaze that accompanies the utterance. “Accompanies” is defined as “roughly simultaneous” with the utterance. Finer temporal discriminations (e.g., shift of gaze immediately prior to onset of utterance versus during utterance) are not made.

(M sitting at table) M: you want a cookie?/ (stringing beads) (turning toward M)

4. Optional. Record direction of gaze of other speakers, using procedures in number 3 above.

5. Direction of child's gaze following (in response to) another speaker's utterance is noted whenever possible, immediately following the speaker's utterance using the following set of symbols.

a. Gaze directed toward an object: \( \text{doll} \).

b. Gaze directed toward a person: \( \text{D} \), where D is the person's initial.

c. Gaze directed toward camera: \( \text{O} \).

d. Shifting gaze: \( \text{S} \).

e. Cannot be determined: \( \text{?} \).

If a pointing gesture accompanies a child utterance, the appropriate symbol is written above the utterance in column 3. “Accompanies” is defined as “roughly simultaneous” with the utterance. Finer temporal discriminations (e.g., pointing immediately prior to onset of utterance versus during utterance) are not made.

If a pointing gesture does not accompany a child utterance, the appropriate symbol is written in column 4, along with other nonverbal behaviors.

Note here that the action (turning) and the pointing gesture (→) are simultaneous.
Note here that the action and pointing gesture are successive.

7. Optional. Record pointing gestures made by other speakers, using procedures in number 6 above.

### Recording Utterances

1. An idiosyncratic pronunciation of a word is indicated with an asterisk as in the following example where "truck" is pronounced "kuk".

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>baby eat/</td>
<td>(putting cookie to doll's mouth)</td>
<td>(turning toward juice)</td>
<td></td>
</tr>
</tbody>
</table>

2. An adult's misinterpretation of a child's utterance is indicated with a check.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>wait a minute/</td>
<td>want mine ♤</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Guesses about the form of unintelligible utterances are followed by a question mark and are enclosed in parentheses above the utterance.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(read book?)</td>
<td>-- --</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Miscellaneous

1. An arrow is drawn from one line to the next when an utterance is too long for the space provided.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>do you want</td>
<td>go to the zoo?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

2. The exact repetition of a nonverbal behavior is indicated by an "X" enclosed in parentheses. If the actor is someone other than the child, the actor's initial is also given.

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>bear/</td>
<td>(shakes bear)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(M shakes head negatively)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>poor bear/</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(M X)

3. A small diagram of the room and of the subjects' positions relative to one another is drawn at the top of each page, as in the following example.

[Diagram of room and subjects' positions]