she not originally have learned the correct answer from nonlinguistic cues emitted by knowledgeable trainers but then at length committed to memory the tokens that would bring reward in the presence of this or that problem configuration. Probably even Hans could have learned certain routine responses to repeated and familiar questions; cases where he was not dependent on the trainer’s cues. With the chimpanzee, and with no counting involved, the possibility seems much greater. I do not feel that the Clever Hans controls Premack has reported definitively rule out the possibility of trainer cues in the original learning and so am not certain that the data mean what Premack takes them to mean.

The Study of Adam, Eve, and Sarah

In the fall of 1962 Ursula Bellugi, Colin Fraser, and I began a longitudinal study of the development of English as a first language in the preschool years of two children, whom we have called Adam and Eve. A third child, whom we call Sarah, joined the company somewhat later. The children were selected from some thirty who were initially considered. Adam, Eve, and Sarah were selected primarily because they were all just beginning to speak multi-word utterances, had highly intelligible speech, and were highly voluble which meant we would not have to sit around forever to get usefully large transcriptions. And because the investigators undertaking primary responsibility for each child (Ursula Bellugi for Adam; Colin Fraser for Eve; Gloria Cooper for Sarah at first, and later Courtney Cazden) felt comfortable with the child and the parents.

All the children were only children at the start of the study. Adam is the son of a minister who lived at first in Cambridge and later in Boston. Eve is the daughter of a man who was at the time a graduate student at Harvard and who lived in Cambridge. Sarah is the daughter of a man who worked as a clerk, at the start of the study, and their home was in Cambridge. The parents of Adam both had college degrees; Eve’s father had a college degree and her mother a high school degree; the parents of Sarah both had high school degrees.

The principal data of the study are transcriptions of the spontaneous speech of the child and his mother (occasionally the father and others) in conversation at home. For each child we have at least two hours of transcription for every month, but within these limits the schedules have varied somewhat. For Adam and Eve a

1. The first five years of this work were supported by Public Health Service Grant MH-7088 from the National Institute of Mental Health, and the second five years by Grant HD-02908 from the National Institute of Child Health and Development. We are deeply grateful for the generosity of this support and the intelligent flexibility with which both grants have been administered. Miss Esther Sorocka has been executive secretary to the project from the beginning, and her importance to the success it has had is very great. The parents of Adam, Eve, and Sarah, and the three children themselves by their unfailing, welcoming friendliness made the whole project possible.
two-hour visit every second week was the basic schedule. For Sarah it was one half-hour each week. These are minimum schedules; when interesting things seemed to be happening fast much more speech was recorded. We found that the visits required two persons, the main investigator and one assistant (Richard Cromer, Gordon Finley, Courtney Cazden, and Melissa Bowerman have all served as assistants). One member of the team devoted himself to a written transcription, on the scene, of the speech of the child and mother (and any others) together with notes about important actions and objects of attention. The other took on the role of playmate for the child and also tended the tape recorder. All conversations were taped. In the case of Adam and Eve the microphone was in a fixed position, and all concerned simply tried to keep interaction within the microphone’s range. For Sarah we required a record of higher fidelity because her records were to be phonetically transcribed in a narrow notation including prosodic and paralinguistic ‘expressive features. Accordingly we sewed a microphone into a garment she was always asked to wear, and her speech was transmitted wirelessly to the tape. The final transcriptions, which constitute the primary data of the project, were made by the investigator principally responsible for each child working from the tape recording in conjunction with the on-the-spot transcription to make a single best record. These were made as soon after the visit as possible. Our experience is that transcription from tape of the speech of children at an early age, even when it is relatively intelligible child speech, needs the assistance of memory of the scene and a written record made on the scene.

The transcriptions of Adam and Sarah are simply at the morphemic level, that is, if a meaningful element was sounded well enough to be recognized, it was recorded in normal English spelling with no effort being made to render the particularities of the child’s pronunciation. However, the transcribers, Ursula Bellugi and Colin Fraser, took great pains over grammatically significant and phonetically minimal features like inflections, prepositions, articles, and contracted auxiliaries. It is a tribute to their immense care, I know, that the data described in Stage II are so remarkably orderly. Gloria Cooper and Melissa Bowerman did the phonetic transcriptions for Sarah. These have been of considerable value for checking on points where one worried whether the transcriptions of Adam and Eve were accurate. They have always proved to be so, on their own level. Sarah’s transcriptions have not been used as yet for a study of the development of phonology but would be valuable for that purpose. Anyone planning to undertake a phonetic record must be warned that it is an immense labor.

During the first year of the project a group of students of the psychology of language met each week to discuss the state of the children’s construction process as of that date. The regular participants were: Jean Berko Gleason, Ursula Bellugi (now Bellugi-Klima), Colin Fraser, Samuel Anderson, David McNeill, Dan Slobin, Courtney Cazden, Richard Cromer, and Gordon Finley. We had wonderfully stimulating, light-hearted discussions. Anyone in developmental psycholinguistics
looking over the membership of this seminar will realize how bounteous that year was. In the seminar small experiments or near-experiments were often suggested and then tried by the main investigators. The results were sometimes useful but never conclusive; the difficulties of experimentation on language with small children are considerable, and we put the transcription schedule first.

At the end of the first year the project suffered the kind of blow to which longitudinal studies are liable: Eve’s family had to move to Nova Scotia, and our 20 two-hour transcriptions were all we would be able to obtain from her. We continued taking data from Adam and Sarah for another four years. But as it happened the fact of Eve’s withdrawal has shaped my role in the project even until the present time. I decided to concentrate on just the developmental period for which we had data from all three children. As it happened Eve’s speech developed so much more rapidly than that of Adam and Sarah that 10 months of her transcriptions equaled about 20 months for Adam and Sarah.

Long before the end of the first year the children got way ahead of the seminar. Their records were far too rich to be analyzed in a two-hour session. It became clear that a fine-grained analysis was a big job and had to be undertaken by one person. Even then only a fraction of the data could be examined. Still I was determined to make the effort because I had not set out to create an immense archive that no one would ever use.

It is sensible to ask and we were often asked, “Why not code the sentences for grammatically significant features and put them on a computer so that studies could readily be made by anyone?” My answer always was that I was continually discovering new kinds of information that could be mined from a transcription of conversation and never felt that I knew what the full coding should be. This was certainly the case and indeed it can be said that in the entire decade since 1962 investigators have continued to hit upon new ways of inferring grammatical and semantic knowledge or competence from free conversation. But, for myself, I must, in candor, add that there was also a factor of research style. I have little patience with prolonged “tooling up” for research. I always want to get started. A better scientist would probably have done more planning and used the computer. He can do so today, in any case, with considerable confidence that he knows what to code.

Our three children were not at the same chronological age when we began our study; Eve was 18 months; Adam and Sarah were 27 months. We had not equated for age because we knew, from much earlier work, that children acquire language at widely varying rates. We had rather equated them from the length of their utterances, both the mean length (MLU) and the upper bound or longest utterance. The mean length of utterance (MLU) is an excellent simple index of grammatical development because almost every new kind of knowledge increases length: the number of semantic roles expressed in a sentence, the addition of obligatory morphemes, coding modulations of meaning, the addition of negative forms and auxiliaries used in interrogative and negative modalities, and, of course,
Table 7. Rules for calculating mean length of utterance and upper bound

1. Start with the second page of the transcription unless that page involves a recitation of some kind. In this latter case start with the first recitation-free stretch. Count the first 100 utterances satisfying the following rules.

2. Only fully transcribed utterances are used; none with blanks. Portions of utterances, entered in parentheses to indicate doubtful transcription, are used.

3. Include all exact utterance repetitions (marked with a plus sign in records). Stuttering is marked as repeated efforts at a single word; count the word once in the most complete form produced. In the few cases where a word is produced for emphasis or the like (no, no, no) count each occurrence.

4. Do not count such fillers as mm or oh, but do count no, yeah, and hi.

5. All compound words (two or more free morphemes), proper names, and ritualized reduplications count as single words. Examples: birthday, rackety-boom, choo-choo, quack-quack, night-night, pocketbook, see saw. Justification is that no evidence that the constituent morphemes function as such for these children.

6. Count as one morpheme all irregular pasts of the verb (got, did, went, saw). Justification is that there is no evidence that the child relates these to present forms.

7. Count as one morpheme all diminutives (doggie, mommie) because these children at least do not seem to use the suffix productively. Diminutives are the standard forms used by the child.

8. Count as separate morphemes all auxiliaries (is, have, will, can, must, would). Also all catenatives: gonna, wanna, hafia. These latter counted as single morphemes rather than as going to or want to because evidence is that they function so for the children. Count as separate morphemes all inflections, for example, possessive [s], plural [s], third person singular [s], regular past [d], progressive [iŋ].

9. The range count follows the above rules but is always calculated for the total transcription rather than for 100 utterances.

embedding and coordinating. All alike have the common effect on the surface form of the sentence of increasing length (especially if measured in morphemes, which includes bound forms like inflections rather than words). By the time the child reaches Stage V, however, he is able to make constructions of such great variety that what he happens to say and the MLU of a sample begin to depend more on the character of the interaction than on what the child knows, and so the index loses its value as an indicator of grammatical knowledge.

Table 7 presents a copy of the rules we used in calculating mean length of utterances (MLU) and upper bound or longest utterance for a transcription. These rules take account of things we learned about child speech in the first year of the study, for example, the fact that compound words are not analyzed as such and the fact that the irregular pasts that occur early are not used with semantic consistency or contrasted with present forms. Still no claim can be made that these are just the right rules. They have, however, served all of us well as a simple way of making one child’s data comparable with another’s, one project with another, and in limited degree, development in one language comparable with development in another.
When I say that the indices have served us well I mean simply that two children matched for MLU are much more likely to have speech that is, on internal grounds, at the same level of constructional complexity than are two children of the same chronological age. We know that we are going to run into serious inconsistencies and uncertainties with some foreign languages, and these are discussed in Stage I. However, the MLU may be effectively redefined or we may find some other, almost equally simple index, preferable. In any case we are getting beyond the point where a single index is vital because we are accumulating knowledge about the acquisition order of general construction types and their meanings, and it is the order of knowledge we really care about.

I calculated the MLU's and upper bounds for all sample transcriptions for all children. The results when MLU is plotted against chronological age for just the period in which Eve participated in the study appear as Figure 1. The values rise quite consistently with age, for Eve most amazingly so. It was almost impossible to fail to find an increment every time two weeks had elapsed. As I remember it the one downward jog came on a day when Eve had a cold. This stretch of development, common to the three children, is what I undertook to analyze in some detail — a good many years ago.

![Figure 1](image)

Figure 1. Mean utterance length and chronological age for three children.
How to proceed? There was far too much data in even this interval to be exhaustively analyzed. I decided to divide the total shared developmental stretch at five points as nearly as possible equidistant from one another in terms both of MLU and upper bound (UB) and draw 713 consecutive complete utterances from each child at each point for detailed linguistic analysis. The odd number, 713, was an accidental consequence of the size of the transcriptions from which the first samples were drawn.

Table 8. Target values and approximations attained for mean length of utterance and upper bounds

<table>
<thead>
<tr>
<th>Stage</th>
<th>Target Value</th>
<th>MLU</th>
<th>Upper Bound</th>
<th>Approximation Attained</th>
<th>Maximum Distance from MLU</th>
<th>Maximum Distance from Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1.75</td>
<td>5</td>
<td>.31</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>2.25</td>
<td>7</td>
<td>.05</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>2.75</td>
<td>9</td>
<td>.25</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>3.50</td>
<td>11</td>
<td>.20</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>4.00</td>
<td>13</td>
<td>.06</td>
<td>1.67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 8 describes my target values which for MLU begin at 1.75 and end at 4.00 with increments of 0.50 (except for III to IV). The upper bounds begin at 5.0 and proceed by increments of 2.0 to 13.0. Of course I could not hit these targets exactly for the samples of varying size from each child. And so Table 8 also describes the widest departures from the target values that the data ever forced me to accept. These are never very great.

Figure 2 is just like Figure 1 except that the Roman numerals I-V and corresponding horizontal lines mark the points at which fifteen 713 utterance samples were taken. These were my preliminary Stages I to V. Then the work of analysis began. I decided to press for an explicit generative grammar for each sample but, because I knew any such grammar would have to remain indeterminate at countless points, I undertook also to write extensive annotations describing alternative formulations, gaps in the evidence, and so forth. Of course the data of performance have long ago been pronounced (Chomsky, 1964) an inadequate base for a grammar that attempts to represent competence or knowledge. I agree that it always is but I venture to say that not many people know how much can be milked from mere performance in the case of small children—especially conversational performance in which you can track relations between sentences.

I have found the process of grammar writing a continual discovery of new things to look at, new aspects of the data that could tell me something about the knowledge in the minds of these three children.

Why write admittedly indeterminate grammars? Simply because the requirement to be fully explicit and develop rules that will really derive the sentences you have obtained forces a kind of intense examination and continuing re-examination of the
data, and so is a good way to get to know it very well. The 15 annotated grammars took something like two months each to do, with the time longer in the later stages. I suppose they are mostly 50 pages or more long. In the years that this took, linguistic theory, of course, went on changing in response to intuition and logical arguments, and when I saw the point of a change I also shifted my procedure and formal notation. In the end I was left with 15 weighty manuscripts which not more than half-a-dozen people in the world have the knowledge, the patience, and the interest to read; nay, not so many as half a dozen.

In the process I formed a conception of great commonality among these unacquainted children and of a remarkably invariant order in the kinds of things they said. I learned that there were some points on which the data were simply too thin to support any sort of generalization; for example, the order of adverbs. I also formed a conception of the kinds of things about which something might reasonably be said, and these are the five constructional processes described in the beginning of the chapter. The 15 grammars I now regard as a protracted preliminary exploration, not boring to me because of the puzzle properties each one develops.
One can stand a lot of sorting and resorting when it is done with hypotheses in mind which make the outcomes exciting. I am personally reluctant to hand this process over to a computer. About two years ago I started to write a new set of Stages I-V, but sticking this time to the kinds of things about which something can be said and hoping to make myself clear to more than half-a-dozen readers.

It remains only to answer a question that was put to me, with some asperity, a few years ago after a talk about child language: “What is this work about really?” It is not about the way the child’s mind in fact processes sentences in speaking and understanding. I do not know how that is done. It is about knowledge; knowledge concerning grammar and the meanings coded by grammar. Knowledge inferred, of course, from performance, from sentences spoken, the settings in which they are spoken, and from signs of comprehension or incomprehension of sentences spoken by others. The book primarily presents evidence that knowledge of the kind described develops in an approximately invariant form in all children, though at different rates. There is also evidence that the primary determinants of the order are the relative semantic and grammatical complexity of constructions rather than their frequency or the way in which parents react to them. I believe that this knowledge must somehow be utilized in actual sentence processing, in speaking and understanding, but cannot say how. I hope the volumes will help to establish reasonably firm generalizations about the unfolding of construction knowledge in children, generalizations on which theory can build.

The Expository Plan of this Work

The plan is not quite the same for any two stages but there are several things that are constant throughout.

1. The stages are not known to be true stages in Piaget’s sense; that is they may not be qualitative changes of organization forced on the investigator by the data themselves. The original equidistant samples based on MLU were simply a device for sampling the data; a discontinuous sampling imposed upon more continuous data. My divisions I to V were rather like a sociologist’s imposition of arbitrary dividing points on a continuous distribution of incomes.

2. The original stages were points on an MLU distribution, but in this work they have become intervals. Stage II, for instance, begins as soon as the MLU rises above 1.0, when multi-word utterances begin, and ends at 2.0. This is because our discussions are not limited to the data from Adam, Eve, and Sarah. Since we started in 1962, there have been numerous studies of the development of English and other languages which started when the child’s MLU was less than 1.75 the target value for Stage I in the original analyses in our study. I have tried to put all this work together, and it is clear that construction begins before 1.75. Stage II extends from 2.0 to 2.50 because we found it possible to deal with all the data in this...
period in a certain quantitative respect. Other considerations make intervals of all the later stages.

3. A stage is named (Semantic Roles and Syntactic Relations for Stage I; Grammatical Morphemes and the Modulation of Meaning for Stage II; Modalities of the Simple Sentence for Stage III; and so on, either for a process that is the major new development occurring in that interval or for an exceptionally elaborate development of a process at that stage. However, the whole development of any-one of the major constructional processes is not contained within a given stage interval. Semantic roles go on developing after Stage I; the modulations of meaning extend from Stage II to beyond even Stage V. The germs of the major modalities of simple sentences (interrogation, negation, the imperative) are to be found even in Stage I in a syntactically rudimentary form, and there are combinations of the modalities, like the tag question, which do not appear until after Stage V. When we discuss embedding in Stage IV we shall have to go all the way back to Stage I to show why certain constructions (the possessive and the prenominal adjective) which appear in Stage I, and are analyzed by many linguists as embeddings, are not such in the child’s speech. In general there is something of interest to say about all five major aspects of construction in all five stages. In general the stage discussion deals with the construction aspect for which it is named across the full range of data.

4. As mentioned in passing, not only the data of Adam, Eve, and Sarah but all the longitudinal and experimental data available to me on a given stage are discussed. There is far more information beyond our own, on Stage I, than on any other stage, including longitudinal studies of Finnish, Samoan, Swedish, Spanish, Luo, and German, as well as a number of studies of American English. The data thin out after Stage I simply because most of the investigators in question have not had time to carry their analyses beyond this point.

5. The stages are not simply descriptive but are organized as evidence and argument for and against certain generalizations. In Stage I, for example, the question is, on what evidence, if at all can the constructional meanings of the first sentences be inferred? There is one recurrent theme in all stages, that order of development, conceived in the right abstract terms is invariant across both children and languages and is primarily determined by the relative semantic and grammatical complexity of constructions.