A MODEL OF LANGUAGE ACQUISITION

IN THE TWO-YEAR-OLD

Jane C. Hill*

COINS Technical Report 83-08

(March 1983)

CONTENTS:

A complete description of this model may be found in "A Computational Model of Language Acquisition in the Two-Year-Old", Ph.D. Dissertation from the Department of Computer and Information Science, University of Massachusetts at Amherst, September 1982. Also to be published by the University of Indiana Linguistics Club.

A revised version of this paper will appear in Arbib, Conklin & Hill, From Schema Theory to Language.

---

* Current address: Department of Computer Science, Smith College, Northampton, MA 01063.

1. COGNITIVE DIMENSIONS OF LANGUAGE ACQUISITION

To emphasize the cognitive dimension of early child language, we briefly review Piaget's theory of schemes, stages, and invariant functions. In Piaget's theory the scheme is the basic unit of cognitive structure and is inferred from behavior. The scheme represents the properties of intelligent acts at a given stage of mental growth. Stages represent changes in structural development, and though the structures continue to develop and change over time, the invariant functions, assimilation and accommodation, comprise the processes for change at every stage. Thus higher cognitive processes evolve from consolidation and generalization of more primitive cognitive processes. The child constructs the cognitive structures through active interaction with his environment, and careful observation of performance on various tasks reveals the child's internal structure and the stage attained (Piaget 1960, 1962, 1971).

We emphasize that language is only one factor in cognitive development. In his studies of the developing child, Piaget took very little interest in language acquisition except to note it as one of the factors (and of course a very important one) in cognitive development. The sources of intellectual operations are not found in language. They are, in Piaget's view, found in preverbal sensory-motor schemes and the concept schemes. Piaget's sensory-motor period covers roughly the first eighteen to twenty-four months of an infant's life. It is a description of the period during which internal thought first develops and the child progresses from a reflexive self-centered being to one which is organized and adapted to its environment.

In Piagetian terms intellectual operations are actions that have been interiorized. Thinking is an activity. Sensory-motor schemes and thinking processes are both actions. In fact, thinking is an activity outside awareness. The body of internal events that one is at times conscious of is not the process of thinking, but rather the product of thinking, in other words symbols. The internal activity of thinking is an action which corresponds to motor schemes of an earlier period of development. Language capacity is a part of the symbolic function. The symbolic function includes imitation, imagination, play, dreaming, and language. The processes of language and conceptualization are reciprocal. Progress in conceptualization goes hand in hand with progress in language. A capacity for constructing a representation is one of the conditions necessary for the acquisition of language.

Early words represent complex schemes of actions. First words are like symbols. They are linked to symbolic play. Piaget gives, for example, an incident in which an infant used the word bow-wow one day to stand for dogs, then later for cars, then men, and eventually for anything seen out of the window. This was said

to illustrate how the first verbal schemes are intermediary between the schemes of sensory-motor intelligence and conceptual schemes. In fact, the words applied by the child to these schemes are themselves intermediary between symbolic signifiers and true signs, which words eventually become. To have words is not to have the concept. Words and concepts are attained through two separate processes. For example, the words, some, none, all, are often misused before the concept of set inclusion and set exclusion is attained.

In analyzing the child's language and cognition, it is important to realize that the child does not perceive the world as the adult does. Though, as observers, we must use our adult language to describe the child's cognitive state and his actions, we should not lose sight of the fact that the child's perception of the world is basically different from ours. It is not safe to assume that the adult and child share the same meanings for words or use them in the same way. In spite of this caveat, communication does take place. Everyone has a set of favorite examples of a child's novel usage. Here are a few that we have come across:

That's my Jane. (Claire says my daddy, my gramma, why not my Jane?)
It's getting middle-sizeder. (deVilliers and deVilliers 1978)
Sally, hello it. (Say hello to the telephone.)
I'm just gonna fall this on her. (Bowerman 1974)
Where his tight? (Having been told to hold the horse tight.)

It seems that there may be degrees of middlesizes, that hello may be a verb, that fall may be a transitive verb, that either Jane may be a common noun, or my can modify proper nouns, and that tight may be a noun meaning a part of a horse's anatomy.

Adult and child need a common interpretation of some aspect of their respective worlds in order for communication to be established. At age 14 to 20 months, when object permanence is acquired, there is seen simultaneously a sharp increase in pointing behavior and a sudden vocabulary explosion of 50 words or more. (Bates 1979, Moore and Metzoff 1978, Corrigan 1978). One can therefore infer at least a transdirectional transfer of effect between acquisition of object permanence and vocabulary explosion (Bates 1979). This is not to say that there is a causative relation between the two, but rather that they share a common need, the need for a mental representation. The various object permanence levels as explicated by Piaget are the following: (1) Objects have no existence independent from that of the child, (2) The child exhibits no interest in vanishing objects, but tracks objects with his eyes, (3) The child anticipates where an object will fall and looks for partially hidden objects, (4) The child will search in one location and chooses the location where the object was previously found, (5) The child searches systematically after successive visible displacements, and (6) The child searches after invisible displacements.

There is a wealth of empirical evidence supporting the fact that the child at first does not see the world as the adult does. One does not have to accept Piaget's theory of object permanence to accept this fact. The phenomena observed have been explained and interpreted variously as object identity: the child assumes there are multiple copies of an object (Moore and Metzoff 1978), memory deficit: the child cannot retain the object in memory (Harris 1971), a concept of space which prevents the understanding of the concept, inside (Gratch 1977), a concept which regards objects in motion as different objects from objects at rest (Bower 1977), or a concept which regards the place that an object is hidden to be unrelated to the place where an object is found (Butterworth 1974). It may well be misleading to attempt to express in adult terms, given our conceptions about the world, just what the child may or may not be thinking. It is sufficient to note that discrepancies do exist between the child's and the adult's versions of the world and that a vocabulary explosion coincides with the attainment of the adult concept of object permanence.

Piaget's classification experiments are another group of cognitive experiments which have important implications for the study of child language. These experiments, like the object permanence experiments, have been replicated by hosts of researchers who agree that the phenonmena observed by Piaget do occur though a great variety of different reasons have been proposed for the results obtained.

The classification tasks divide into three substages: (1) The child composes graphic collections of objects, (2) The child performs simple classification tasks, and (3) the child can perform multiple classification tasks, and can solve the class inclusion problem.

The free classification task consists of placing a group of objects on a table with the direction, "put together the things that go together," or merely "do something with these," or with no direction at all. The youngest child will merely group the objects into figural or graphic collections to represent an object such as a house or a design. Sometime between age two and age four children begin to classify the objects into sets of like objects, sorting on one dimension (Denney 1972, Sugarman 1981). They will, for example separate the red circles from the blue squares. Around age five, the child will sort a set of objects into an array by sorting on two dimensions, such as color and shape.

The first words which a child learns are words for the basic categories and not those for superordinate categories. He will learn cat and dog before he learns animal, and chair and table before furniture. He also learns these basic classes before he learns subordinate classes such as those of rocking chairs and kitchen chairs (Horton and Markman 1980). For the superordinate classes there is lower cue validity and greater perceptual

dissimilarity. In this area linguistic input may cause conceptual growth. Hearing such terms as furniture and animal may spur the child to form the superordinate categories. If this is true then it is plausible that these words are acquired around age five when facility in language has been acquired. Subordinate categories have greater similarity and weaker contrast, so again linguistic input may aid the development of the concept.

The process of classification has important implications for language study in another sense, because words themselves if divided into classes or categories permit generalizations about language structure. The mastery of the multiple classification task corresponds with the age at which the child is often said to have acquired language. Though the mastery of language is not complete, certainly a considerable degree of competence has been attained.

The model [Hill, 1982] which will be described in Section 3 was influenced by Piaget in several ways. In the model, templates are the basic units for language growth and these templates, like schemes, are inferred from language behavior. The model is a stage model. At least four different stages of language growth are spelled out. These stages are in no way analogous, however, to Piaget's sensory-motor, preoperational, concrete operational, and formal operational stages. In fact, all the stages of language growth to be modelled appear during the time that the two-year-old child is in Piaget's sensory-motor and preoperational stages. We call them stages because they do represent structural changes in the representation, and not just the addition of data. The invariant functions are embodied in the model's interpreter which operates on the data structures. Two invariant functions in Piaget's theory were accommodation and assimilation. Assimilation presupposes an interpretation of something in external reality to assume some kind of meaning in the subject's cognitive organization. The subject can incorporate only those components of reality which its structure can assimilate without drastic change. On the other hand new information that calls for behavior that lies beyond the scope of the subject's present level of cognitive structure induces the structures to change or accommodate themselves in order to handle the information. Assimilation is represented in the model by the addition of data to the data structures as new words and new concepts and new templates for word combinations are added. The process of accommodation is represented in the model by the reorganizing of the data as the child proceeds from the first stages until the stage of recursive language is achieved. The data structures evolve through a process of assimilation of new data until a reorganization is triggered. The reorganizations can be thought of as the process of accommodation. In the model, the data structures accommodate to three different reorganizations, each one representing a different stage.

1. the specific example level
2. reorganizing of words into abstract classes
3. reorganizing of templates so that classes are not tied to a specific position in a template
4. reorganizing of the template format to permit hierarchical representation and recursion.

I reiterate that the child's perception of the world is different from that of the adult. An important constraint on the model is that it does not depend on knowledge of word classes such as those that are obvious to the adult, such as a division between actions and objects. Though the model requires some schemes for word classification and template classification in order to grow, the actual classes used remain flexible and are inferred from the child's language behavior.

2. THE LANGUAGE OF THE TWO-YEAR OLD

In the next chapter I present the model of the language of a two-year-old child [Hill, 1982]. The model is based on analysis of the acquisition literature in general and, in particular, on the linguistic data collected [Hill, 1982] from a two-year-old child on a weekly basis, forty-five minutes per week over a time span of nine weeks. The child, Claire, was 24 months 6 days old at the start of the study.

Combining Two-Term Relations

Roger Brown (1973) in his well-known study of twelve children between the ages of 1 year 6 months and 2 years 6 months who were learning five different languages (English, Finnish, Swedish, Samoan and Spanish) drew the following two conclusions:

1. That a short list of semantic relations account for the majority of utterances at this age, and
2. That the two-term relations were combined in just the same two ways by all the children recorded.

Though there is, of course, disagreement about the list of relations and about the proper formalism for describing these relations, nevertheless it is generally agreed that children, whatever their language, begin talking about the same limited set of things (Braine 1976, Bowerman 1973, Bloom 1970, Nelson 1973).

Brown's description of the two ways of combining relations is as follows:

1. It was as if two or more relations were concatenated with all repetition terms struck out. To illustrate:

Adam hit
Hit ball
==> Adam hit ball

2. One term (always a noun phrase) was unfolded as a two-term relation. To illustrate:

Hit ball ==> Hit Adam ball

Brown emphasized the as if in (1) since he had no psychological evidence concerning what the child's actual mechanisms were.

I offer two pieces of evidence that suggest that the child may be using concatenation and deletion in combining these elementary terms, and that this process of concatenating and deleting may describe both of the ways that children combine relations. The first piece of evidence is to be found in the Claire data, and the second piece is to be found in the

psycholinguistic data presented by E. Matthei in his Ph.D. dissertation in 1979.

In the first two sessions in which Claire's language was recorded, her language consisted almost exclusively of single word utterances and two-term relations. Her only four-word sentence was

get chair this one,

and only 16 three-word sentences occurred which consisted almost entirely of patterns which she repeated:

where daddy go?
where girl go?
where pencil go?
where cow go?
what kitty doing?
what dog doing?
pick it up
Claire do it
Jane do it
there it is
look at that
knock it down

In the third session, however, there occurred a number of four-or-more-word utterances:

1. more one daddy one
2. fit it in in there
3. daddy put it in in there
4. another one fit it
5. this mommy fit it
6. little one mommy bear
7. another bear mommy bear
8. little bear baby bear
9. theres-a new one mommy chair

I propose that these are all comprised of a concatenation of the two-word relations (or one three-word and one two-word relation in the case of 2, 3, and 9). Most importantly in the case of 1, 2, 3, 7, and 8 the concatenation operation was employed but though the deletion rule could have been employed, it was not. These four-word utterances (sometimes with repeated lexical items) were rapidly transformed into three-word utterances, but the data which occurred in this short interval led us [Hill, 1982] to speculate that three-word utterances such as <u>two</u> <u>daddy</u> <u>forks</u> were arrived at by (1) concatenating the two relations

two forks

and

daddy forks

and, (2) collapsing the concatenated relations into a single three-word utterance by deleting the first occurrence of the

repeated word thus producing

two daddy forks.

Thus Brown's example

Hit Adam ball

could have been derived by concatenating and collapsing "Hit ball" and "Adam ball".

Matthei [1979] presented evidence that early in language acquisition children use flat structures where adults use hierarchical ones. In his experimental work Matthei tested the child's understanding of the phrase, the second green ball, and found that children interpreted the phrase as the ball which is second and green, as opposed to the adult interpretation which is the second of all the green balls (Figure 1). Matthei found the child's preference for flat structures to be so strong that several of the children, presented with an array in which the second ball was not green, actually rearranged the balls in order to make the situation conform to their interpretation of the words. (Of course, this is not the only evidence that children prefer flat structures. See Tavakolean, 1978, 1981, and Solan and Roeper, 1978, for a discussion of relative clauses interpreted as flat structures.)

I hypothesize, therefore, that the child (1) had two separate two-word relations for second ball green ball, (2) concatenated these relations into second ball green ball, and (3) collapsed the relations by omitting the first occurrence of the repeated word -- yielding second green ball. This would explain the child's insistence on interpreting the ball as that which is both second and green. In short, this hypothesis offers an explanation of E. Matthei's findings.

Consider the sentence, another bear mommy bear which appeared in the Claire data, concatenated and collapsed to another mommy bear. In two-year-old language, this would refer to another bear which was also a mommy bear. To the adult, however, another mommy bear means that there are at least two bears who are mommies. Presumably at some point the child will discover the discrepancy between his meaning and the adult meaning. Accommodating to this discrepancy could conceivably take the form of reorganizing the flat grammar into a hierarchical one.

E. Matthei 1979

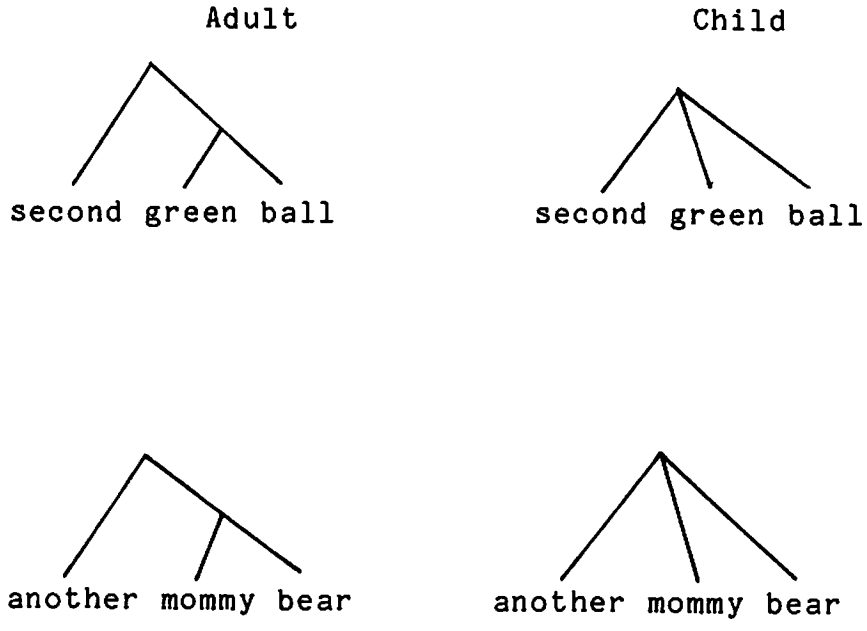


Figure 1. Adult structure contrasted with the child's flat structure

Table 2 contains a list of progressions drawn from the Claire data which shows the two-word relations, the concatenated relations, and in the cases where concatenation produced a repeated lexical item, the result of deleting the first occurrence of the repeated lexical item, which result is labelled the collapsed relation. Those concatenated relations which appear in square brackets represent hypothesized transitions whereas the others actually appeared in the data.

Two-Word Relations	Concatenated Relations	Collapsed Relations
What doing? Kitty doing?	[What doing kitty doing?]	What kitty doing?
Where pencil? Pencil go?	[Where pencil pencil go?]	Where pencil go?
Where pencil? Claire pencil	[Where pencil Claire pencil?]	Where Claire pencil?
Two forks Daddy forks	[Two forks daddy forks]	Two daddy forks
Theresa one New one	[Theresa one new one]	Theresa new one
Theresa new one Mommy chair	Theresa new one mommy chair	
Little kitty Kitty one	[Little kitty kitty one]	Little kitty one
Little one Mommy bear	Little one mommy bear	
Little bear Baby bear	Little bear baby bear	Little baby bear
Another bear Mommy bear	Another bear mommy bear	Another mommy bear
This mommy Put it in	This mommy put it in	
Another one Put in	Another one put in	
More one Daddy one	More one daddy one	More daddy one

Table 2. Table of progressions by means of concatenating and then collapsing if duplicate words occur

[] indicate hypothesized transitions

One can speculate that the reason that it is the first occurrence of the duplicated word which is deleted may be because of a memory constraint. Supporting evidence is provided in the fact that children elaborate predicates before they elaborate subjects. (Menyuk 1969). This could simply be because for young children subjects are most frequently animate, and objects most frequently inanimate; but alternatively it could be that there is a limitation on the planning process. Even for adults it is easier to tack structures on to the end of a sentence rather than planning ahead to incorporate those constructs at the beginning.

This analysis of concatenation and collapsing predicts that four-word concatenated (but not collapsed) relations will be only a fleeting phenomenon in the data collected. And this is the case in the Claire data. The sample of concatenated utterances is small, but they are taken as clues to the underlying process. Like most children's errors, these constructs can be easily overlooked if one is not specifically looking for them. If one believes that children test hypotheses and that they are sensitive to input data when they are testing these hypotheses, then most of their errors, like these of Claire's, will occur for a very brief period (Valian, Winzemer, and Errich in Tavakolian 1981, deVilliers and deVilliers 1979).

In fact, a child would not necessarily have to utter the concatenated phrases in order to use the concatenate-collapse mechanism. The mechanism predicts that the concatenated four-word utterances will appear after the two-word utterances and before the collapsed three-word utterances are common. This was in fact the case in the Claire data. Since the hypothesis has semantic explanatory power, I suggest that it is worthy of future research.

Different Children Proceed in Different Ways

There is evidence that different children proceed to learn language in different ways. (For a summary of these differences, see deVilliers and deVilliers, 1978.) In her monograph, "Structure and Strategy in Learning to Talk", K. Nelson (1973) describes some of the differences. There are those children who seem to learn to talk all at once, after months of non-verbal responses. One can only speculate about what is going on during the months of silence. Talking a lot may indicate an experimental strategy, whereas the silent child may be employing a processing strategy in which he concentrates on enlarging his repertoire of utterances understood. To model this child would consist merely of suppressing the output sentences of the model, but this would hardly serve to extend our understanding.

Nelson showed children to differ also in their use of language. She distinguished between (1) children who employ the strategy of talking about things, and (2) children who employ the strategy of talking about themselves and other people. The first group use largely object-oriented language, with much pointing and naming, and fewer phrases. The second group use largely a

socially interactive language with more phrases such as go away, stop it, don't do it, thankyou, I want it. There is a distributional difference in the lexicon associated with each type of child, the second group using more function words and more pronouns.

Since the model to be presented below is given its lexicon and proceeds to form relations and classes for language learning based on the lexicon given, it is probable that the model is capable of employing both of these different learning strategies. Claire's language seems to fall largely in the social interactive group, and therefore it is this kind of learning that has been modelled to date.

Gathering the Claire Data

Claire was 24 months 6 days old at the beginning of the study. I visited her in her home in the evening once a week for 9 successive weeks, and recorded her speech each visit for a period of 45 minutes. Her mother was always at home although seldom in the room with them. An effort was made to play each week with the same toys in order that the vocabulary used would be relatively stable, and analysis could concentrate on the structural growth of Claire's language. By the tenth week, the complexity of her speech had advanced at a pace so much faster than the development of the model that I decided to discontinue recording.

On each occasion Claire's play centered around a large Fisher-Price doll house which had two bedrooms with beds, a living room, a bathroom with a bathtub, and a kitchen with table and chairs. There were a set of male dolls which Claire referred to as daddies, a set of female dolls which she called mommies, several girl dolls, one baby doll, and several dog dolls. The house had a garage and a toy truck. When Claire tired of playing with her "peoples", she generally either chose to read a book or play with blocks.

In recording her language, her use of articles posed a problem for me, since I could not distinguish between Claire's use of a and the. At the start all articles were generally omitted. In the model we have omitted any use of articles. In light of the studies that have been done on the use of articles by even very young children (S. Carey 1978) this is a serious omission which should be rectified in a future study. In the transcriptions I included articles only where I was fairly certain of them. The choice between a and the in the transcription should not be relied upon. My own speech was transcribed only in those instances when Claire responded to what I was saying. Claire frequently ignored my remarks when she was intent on some activity, and so there is little discourse. Where exchanges occurred, I recorded them.

I relied very heavily on intonation for the process of transcription. Claire's use of question intonation was very clear and I therefore have not hesitated to use punctuation to distinguish between questions and declarative statements. Sentence boundaries were also clear from her intonation. The word boundaries are more problematical than the sentence boundaries. The problem of whether a given item is one word or two is hard to solve. I hyphenated words (e.g. right-here) only when she heard Claire use the pair exclusively. I never heard her say merely right or right followed by anything other than here. The separation of lexical items is an intuitive process, and I used the combined forms only where I felt certain that this was a correct choice. Such decisions are always open to dispute.

Claire's intonational patterns defined a single utterance, and distinguished between statements and questions. I am convinced that the process of fitting words to intonational patterns influenced her choice of word combinations. The following group of utterances illustrates one such instance. It has often been noted that the young child has a proclivity for producing longer predicates and shorter subjects. This pattern was dramatized in Claire's monologue about Humpty Dumpty. Claire and I were reading a book of nursery rhymes together.

Thats Dumpty Dumpty. Thats Dumpty Dumpty, Jane.
The Dumpty doing? Theres more Dumpty Dumpty.

It would seem that it was her intonational patterns that dictated that Dumpty Dumpty should be referred to merely as Dumpty in the third sentence, where Dumpty was placed in one of Claire's characteristic question forms:

The _____ doing?

3. AN OVERVIEW OF THE MODEL

In this section, I present in some detail a computational model, described fully in Hill [1982], of the acquisition of language by the two-year-old child. We saw in the previous chapter that children do not necessarily learn language in the same way (deVilliers and deVilliers 1978, Nelson 1973). Here I offer a general model of language acquisition in the two-year-old as evidenced in the available literature, but the model is tuned by the specific data on the particularities of how one child, Claire, was learning language.

Certainly there is no consensus about what may ultimately be necessary to explain language in all its complexity. The present approach is to look at language from the very simplest level, in detail, and in small time increments, in the hope that language at time (t) can be explained in terms of language and cognitive experience at time (t-1). I specifically do not impute to the child all the complex mechanisms hypothesized from examining adult language. No assumptions were made about the ultimate form of the adult grammar, but an attempt was made to be precise about the assumptions and processes which were found necessary for the acquisition model:

(1) the child has schemas for, and talks about, relations (Nelson 1973, Brown 1973);

(2) the child has schemas for and employs word order in his utterances (Wanner and Gleitman, 1982);

(3) concatenating and deletion rules are employed (Brown 1973);

(4) the child forms classes of concepts and classes of words (Braine 1976); and

(5) the classifying processes cause successive reorganizations of the information stored.

The model is modular in form and contains separate linguistic and conceptual domains, though interaction between the two is necessary for language growth. The process of language acquisition embodies an unconscious assumption on the part of the child that there is some orderly mapping between the sounds uttered by adults and events and actions in the child's world. Schema theory gives us a vocabulary for describing these underlying unconscious assumptions. In our model, the development of new cognitive schemas causes the child to find words to talk about these schemas, while processing linguistic input from the adult causes the child to develop schemas to undergird the unknown words.

Adult words may subdivide the child's concepts or may encompass a larger category of objects than the child's, so that learning a new word may alter the child's set of conceptual classifications (deVilliers and deVilliers 1978). Another example is learning dimensional words such as thick, thin, short, tall, which eventually provide the child with schemas for distinguishing relations on different dimensions, since at first these words are used as just so many ways of saying big or little (Carey 1978). In this manner linguistic forms may draw attention to conditions or events that the child might not otherwise have noticed. In the model, adding a word to the lexicon causes a concept to be added to the concept-space, and adding a concept to the concept-space prompts the model to ask for a lexical item corresponding to it.

The model is given words and a representation of the concepts associated with words. It develops word classes, and, starting with two-word sentences, explores the way that these may grow to form longer utterances of three, four, and five words. No information is provided to the model about word classes. The model proposes a way that words may be grouped into classes according to (1) cognitive information, and (2) the relation between words as used by the child. This method of classifying words is given the name classification through word use. Though the scheme for word classification is based at first entirely on world knowledge about the concepts which the words represent, eventually through a reorganization (or accommodation), a set of lexical classes is built up which may be separate and distinct from the cognitive categories.

The data suggest that word combinations at this stage in language development are based on concatenation of words and phrases, thus maintaining a flat rather than a hierarchical structure. We saw in Section 2 that this offers an explanation of some psycholinguistic data (Matthei 1979). It is hypothesized that hierarchical structure is linked to the recursive nature of language, both of which characteristics are lacking in the speech of the two-year-old.

I now present a quick overview of the model itself, its inputs, outputs, internal structures, and procedures, in order that the reader may understand how each process and structure fits into the overall picture. Figure 1 provides a diagram of the components of the system. All will be described in more detail in subsequent sections.

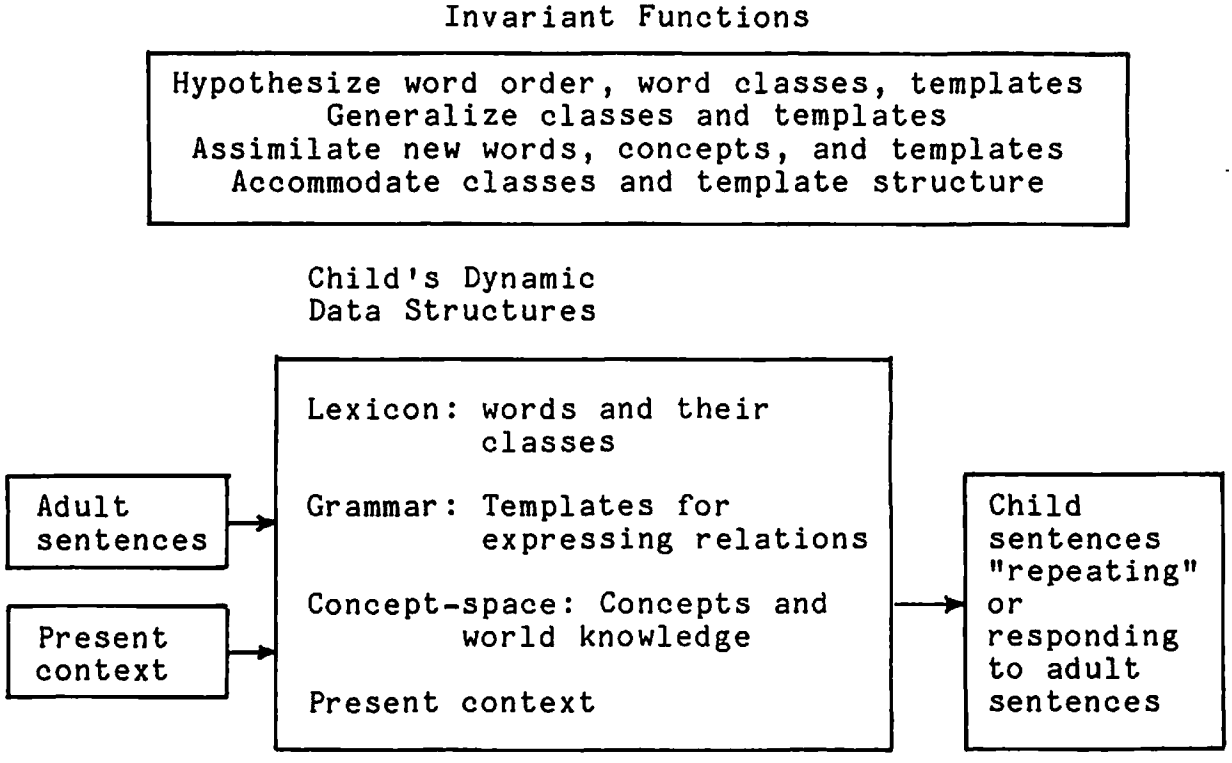


Figure 1. Components of the model

The model is a repetition and response model. One of the inputs to the model is an adult sentence, either a statement or a question. The other input is an indication (provided by the user of the model) of present context. Output of the model is a representation of a child sentence which either repeats the adult sentence or responds to it, in either case according to the child's current grammar. By accepting input in the form of adult sentences, and producing output of two or more word "sentences", we omit any consideration of phonology or of segmentation of the sound stream, and have thus tacitly assumed that the child has at the very least a word-schema. Since even the child's first two-word responses encode relations, these can be termed sentences in the child language. We have therefore granted the child a rudimentary notion of sentence (the first sentence schema), in the sense of a coherent proposition to be expressed.

One very large assumption of the model is that each word in the lexicon is linked to some meaning in the cognitive space, albeit perhaps not a correct one in the adult sense, and almost certainly not a complete meaning. In the computer implementation there are pointers between words in the lexicon and the concepts they stand for in the concept-space. This link is meant to represent the fact that the words which a child uses have some meaning for him, and represent some schema. It is important to

keep clear the distinction between a word and the concept it stands for. Having this link in place between concept and word is a deliberate simplification of the problem. We have not addressed the problem of how word meaning and words come to be linked in the first place. The model starts to understand sentences and to produce sentences at a point when a set of individual words and meanings have been acquired.

The language of the two-year-old is so simple that many different descriptions are adequate. The challenge lies in devising a formalism that is based on semantic plausibility, and yet is sufficiently flexible that it can grow into the language of the adult which has syntactic components relatively independent of their semantic basis. As I shall explain in the next section, I have used a set of 'templates' to represent the grammar at any time. It might be hypothesized that the child's comprehension may be more advanced than his production, and this should perhaps be reflected by the use of two different grammars, or perhaps a set of grammatical rules with only a subset being used for production. This would be an interesting formulation to explore in a future version of the model. At present, for simplicity's sake, the templates for comprehension and production are the same. A new template may be formed from comprehending an adult sentence, but this template will then be available for use in production.

The dynamic data structures which grow as the model acquires language include representations for:

1. The lexicon
2. The grammar to which the model adds data on templates for expressing the concepts salient to the child
3. Conceptual knowledge of the child
4. Specific information about present context which is necessary to choose between a set of alternative responses to questions asked.

The processes which operate on these data structures represent invariant functions; in computer terms these processes comprise the model's interpreter. It is these processes which form and evaluate the hypotheses about (1) how to express the relations which are encoded in templates and how to combine these templates and (2) which form generalizations about linguistic classes derived from both the child's conceptual or world structure as he sees it, and his projections of lexical class based on word use.

4. THE TEMPLATE GRAMMAR

There have been many different kinds of grammars proposed for the two-word stage: pivot grammar (Braine 1963), topic-comment (Gruber 1967), semantic category (Bowerman 1973), and transformational grammar (Chomsky 1975, Roeper 1981, Brown 1973, Wexler and Culicover 1980). It has been observed that they will all work fairly well, since language at this stage is so simple. The criteria for judging them must be whether they provide a plausible foundation for the language which grows from this stage. Wanner and Gleitman [1982] even raise the question whether adult language is related to the two-word stage, or if this is actually a pre-language form of communication:

> "It is our impression that many investigators believe they are making the problem of language learning easier by pointing -- probably correctly -- to these functional semantic categories as the ones that operate in early child speech; and by supposing -- again probably correctly -- that these categories map onto the child's word orderings or inflectional markings in a simple way. But to the extent the child really makes these suppositions these can only complicate the problem of learning a language. The reason is that the supposition is false of any real language."

So though, in a sense, almost any description will do for the two-word level, it is also true that all the descriptions are somewhat unsatisfactory if we are to posit continuity in the learning process. The transformational grammar assigns to the child much more knowledge than there are empirical grounds to assume that he has from the start, and the semantic category grammar assigns to the child a grammar which is at best only a distant relative of the adult grammar.

If one assumes that early language is related in some fashion to later language (as opposed to the discontinuity theory which suggests that early language is a pre-language and unrelated to that which follows), then we are left with the problem of how one gets from one stage to the next. I have adopted the position, as Wanner and Gleitman do not, that a grammar similar to a semantic category grammar can be a springboard for the learning of syntactic categories through processes of successive reorganization.

Templates defined. We have posited a template grammar based on child language at the two-word stage, and have tried to keep it free of any characterization of the more adult grammar which will emerge, but is not yet present.

The grammar consists of templates which represent relations. Templates are patterns for understanding and producing speech. The first templates are specific instantiations of language use which the child gleans from the input data because of the immediate salience of certain concepts to the child. In the beginning every template consists of one invariant word, the relation word, and one slot with an example slot-filler.

Armed with a specific example-template, in which the order of words is encoded (i.e. want milk) and in which want is the relation word and milk is the slot-filler, the child may produce an entire set of two-word sentences in which any object of his desire for which the child knows the lexical label may be substituted for the word milk. Thus he may express want doll, want blocks, want juice; the number of utterances available to him being limited only by his vocabulary for objects which he may want.

The slot will be found to permit a small set of slot fillers, based on the meaning of the relation word. The slot may either precede or follow the relation word. As the model acquires more knowledge the templates will grow to permit more than one slot, and to include more than one relation word. The important characterization of templates is that they are flat. They do not consist of replacement rules. They may be concatenated and collapsed, but they do not contain any hierarchical elements. And they are originally expressed in terms of specific examples.

> [Children] could conceivably make an individual rule for each verb specifying that 'the one who drinks', 'the one who drives', 'the one who sings', etc. appears before the verb, whereas 'that which is drunk' and 'that which is driven' appears after the verb. They might also make the abstraction at some intermediate level between the initiators of individual verbs and the concept of agent and formulate rules which apply to classes of verbs which are semantically similar in some respect.
>
> [Bowerman 1973, p. 190]

The model takes this hypothesis seriously and begins by forming a different template not just for individual verbs but for every individual relation word.

At the two-word level, almost all the child's utterances express one of a few classes of relations [Brown 1973, Bowerman 1973, Nelson 1973]. The templates encode the child's grammar for expressing these first relations and encode the order of the lexical items. At this early stage the child is assumed to employ a set of specific templates to express the concepts he wishes to express. Figure 2 summarizes the first templates.

Statements		Questions	
s-1	{give / want / read} object	q-1	where {person / object / animal} ?
s-2	object {on / off / up}	q-2	{person / animal} doing ?
s-3	{big / little / hot} object	q-3	{person / object / animal} go?
s-4	{Mommy / daddy / animal} object	q-4	what doing?
s-5	{theresa / thatsa / thisa} object	q-5	what that?

Figure 2. First Templates Defined

From the start the child is not assumed to have generalized these templates into the different types of relations which are described in Figure 2. They are separated in this fashion for expository purposes, but no differentiation is made in the model, except on cognitive or semantic grounds. The individual lexical items included in brackets are not meant to represent an exhaustive list, but the number of first relations is not large in number. The verbs in type s-1 must be simple transitive verbs which express concepts salient to the child, as must the adverbials in type s-2 and the modifiers in type s-3. The concepts which he encodes in the templates will express relations of interest to the child which are related to his needs or which describe instances of movement or change which attract his attention (See Kagan, discrepancy principle, 1971, also Bower 1977).

The words which are available to fill the slots in the templates are the names of concrete objects which are salient to the child including food, clothing, toys, and body parts (Brown 1973, Keil 1979, Bloom 1973, Nelson 1973).

Gleaning templates. How do we characterize salience for the child of 19 to 24 months? In his dissertation, J. Conklin (1983) describes a set of guidelines for determining salience which determines the linear mapping of verbal descriptions people use to

describe the unordered data presented in pictures of such scenes as houses in suburbia, farmyard scenes, etc. Though size and centrality are extremely important concerns, they can be overridden by the importance attached to humans, animals, and action. Though the people in a picture may be small, they are assumed to be important.

Children are known to pay particular attention to people, other children, and pets. Their interest is attracted to action, to things that move, and to things that change. They talk about food and toys and the people in their immediate environment. The relations between concepts are particularly salient, and each two-word template encodes some conceptual relation which is salient to the child. In his linguistic environment the child is especially attentive to stressed words. Stressed constructs are acquired first. Children are also highly sensitive to order. (See Wanner and Gleitman [1982] for an overview.)

I propose that the child, in possession of a relation he wants to express, and in possession of a lexical item which he has associated with that relation, listens to the input data until he chances upon an instance of the relation expressed in the data, then stores away an example-template for expressing the relation. He is, in effect, gleaning examples from input data to express the concepts and relations he wants to express. For example, if given the cognitive information that wanting is directed at a physical object, the model waits for a sentence to come in with want in it, chooses a word which represents an object, and encodes the order information about the word want and its object in an example template. This template approach, then, reflects the hypothesis that children are predisposed to pay attention to relations and to word order.

In the model, this process is implemented via a gleaning list. The model is given a list of salient relations, and the input data are matched against the relations which have been placed on this list for template learning. When a sentence contains a relation word which is on the gleaning list, then an example template is extracted from the sentence for expressing that relation. Thus linguistic data are gleaned from the input data to express salient concepts.

The model begins with the rote learning of example-templates (two-or-three-word templates) and proceeds to grow templates and assign classes to words based on these first templates.

Note that at this level the data forces us to view the names for people and pets as relation-words just as verbs, adjectives, and adverbials are viewed as relation-words. This is not an entirely satisfactory use of the word relation, but these concepts do express a relation between the person and an object. Names, of course, have several roles, since they may be used as slot-fillers in templates other than the "possession" template, and words such as mommy can be both someone's name, and an attribute in such phrases as mommy bear.

Certainly, rote learning plays a very small role in language learning in the sense that language is a creative process even for very small children, so that even the smallest do not repeat verbatim the sentences of others. They are creative in their use of language. However, much memorizing does take place. Words are memorized. In languages such as French and German where the gender of nouns has no semantic correlation with the world, the gender is memorized with the noun. Idioms are memorized. Gross (1979) has emphasized the rote learning necessary for classification of simple predicates in French.

From the specific to the more general. The model begins its processing at the one or two-word sentence level normally attained by a child of one and a half to two years of age. The model progresses through several qualitatively different levels of processing which mark the transition from one stage to another. For this reason I have chosen to refer to stages of the model. Because these stages represent reorganization of data structures, they may be seen as the result of the process of accommodation. I do not claim, however, that in the child the transition is accomplished suddenly at a given time. Original hypotheses about sentence structure are obtained by forming templates based on the input sentences as example sentences. These templates will represent hypotheses about language structure and eventually evolve from example-templates to abstract-templates to more general templates as is described below.

Thus far I have described only the two-word example-template. Even at this stage, some generalizing has taken place in the form of a set of alternative lexical items which may be used to fill a slot, based on the meaning of the relation word and the world objects which may meaningfully combine with the relation.

As templates for expressing relations are added to the grammar, the items in the lexicon are tagged according to the way in which they might combine with relation words. In this way word classes based on potential for word use are projected. This process is detailed in the following section. It is, however, by means of the word classification process that templates are generalized. Figure 3 illustrates the three template stages. At Stage I the templates are example templates, but at Stage II, the relation words themselves are generalized according to the sets of words which may serve as slot fillers, and of course according to semantic distinctions as well. We shall see in the next section that one of the first groupings of templates by the child is to group together those which represent entity and attribute. In this way the templates grow from specific examples to abstract patterns. The third stage in Figure 3 represents the stage at which the child has generalized the same word class as a

Stage I
Example Template
Daddy kiss Steven

Stage II
Abstract Template
word-from-class(i) relation(j) word-from-class(k)

Stage III
Generalized Template
word-from-class(i) relation(j) word-from-class(i)

Figure 3. Template stages.

Note that in Stage III, abstract word class is no longer restricted to a fixed position in the template.

slot-filler for more than one position in a template. Stages I and II may exist at the two-word level, but Stage III requires templates of three or more words. Stage II in the model is triggered by the unwieldiness of the example notation as data is added. The classificatory process is triggered by the accumulation of data until the representation becomes unwieldy. The reorganization thus triggered achieves a higher level of language competence, and the clasification effects a change in the templates. Stage III is triggered by the collection of mostly three-word templates, and a need once again to consolidate. This time it is the templates which are consolidated.

The template grammar is a flat grammar, and of course we know that adult grammar is both hierarchical and recursive. Since the model only attains the language level of a two-year-old speaking in four-or-five-word sentences, it has no need of a recursive grammar. There are several ways that the template grammar might evolve into a recursive grammar. An illustration is the route of repeated adjectives. The child may talk of a "big big balloon," or a "big, big, big balloon," or a "big, big, ..., big balloon." Once a child attains the (unconscious) realization that there is no limit on the number of big's he may use, then this fact must be reflected in his grammar. Now he has need of an altered template form which can permit

relation-word(i) [e.g., big]
-->relation-word(i) relation-word(i)

or alternatively he has need of a name (a non-terminal symbol) for

this template:

relation-word(i) [e.g., big] class(j) [e.g., balloon]

Let the template be called T(k). Then T(k) can be rewritten as

T(k)-> relation-word(i) T(k)

Whatever the correct formulation, once the need for recursion arises, the template grammar must accommodate to permit it. At this point we have, of course, achieved a rewrite rule, a hierarchical system, and a traditional phrase structure grammar. A similar route to recursion could be described by the use of many and's or or's:

I want a cookie and a banana and a ...

or by many dependent clauses

I saw a cat that chased a cat that chased a cat...

Whatever the actual trigger(s) for recursion might be, it is clear that a significant reorganization of the grammar has occurred when the template evolves into a rewrite rule from a flat pattern.

The present model, however, is concerned with the language of the two-year-old, and we have seen (Section 2) that there is much evidence that his grammar is neither recursive nor hierarchical. The phrase structure grammar is just too powerful to be a good description of the two-word level. It is for this reason that Braine (in his 1976 monograph) proposed a "limited scope formula analysis" of children's two-word utterances, a formulation which the template grammar resembles{1}. A general rule, for example:

NP --> ADJ N

would permit a child to use adjectives in the subject position of sentences as readily as in the predicate, and it is simply not the case that he does this. He will say

The book is red

long before he will say

The red book is on the table.

(For a discussion of this, see Menyuk 1969.) The same is true of relative clauses which are used to modify the predicate long

---

1. The work of Koenraad Kuiper (1980) on the oral formulaic aspects of stock auctioneering indicates that templates have their use in adult language as well as child language, at least in certain circumstances.

before they are used to modify the subject.

The use of templates predicts that there will be very little variability in the earliest utterances. A few templates will suffice. The nineteen-month-old child has habits of speech which can be recorded and which can be related to classes of specific lexical items. Errors of over-generalization can occur only after the abstract level is attained.

5. THE PROCESS OF CLASSIFICATION

The process of understanding the world can be said to be inseparable from the process of sorting objects, ideas, events and states into sets and subsets. We are constantly noting that one thing is like or unlike another in one or more respects. "Without classes we could not slice the world up into manageable collections of objects. We would have to deal with every object in isolation." (Brainerd 1978). It is safe to assume that the child is engaged in this continuing process of sorting out the world at the same time that he is beginning to use language (Sugarman 1981). In the model, it is the process of classification of concepts and of words which makes possible the generalizing of templates.

Posit a great many word classes. Most analysts of language would agree that there must be some partitioning of words into word classes in order for the structure of language to be described. Traditionally, in English, there are thought to be eight syntactic parts of speech (more or less) which certainly include noun, adjective, and verb. Yet as detailed linguistic analysis of language proceeds, each of these partitions must be subdivided many times again. We have common nouns, proper nouns, count nouns and mass nouns; transitive verbs, intransitive verbs, and an enormous variety of subcategorizations of verbs; just to name a few of the classifications necessary to explain the structure of adult language. The correct partitioning is of course an unresolved question, and the solution depends very much on the correct analysis of language, which is an even larger unresolved question. It may also, to some extent, differ from individual to individual.

There is much data to be found in linguistic research to support the need for many sub-classes within the large classes of parts of speech. The adjective offers an excellent illustration. Those linguists who have examined the use of adjectives in English (Suppes 1978, Siegel 1976) have found it necessary to subdivide adjectives into intersective and non-intersective classes (among others) in order to capture the semantic distinction between on the one hand adjectives which can function as simple predicates denoting classes:

That's a blue wagon.

(The object designated by that is in the intersection of the set of blue things with the set of wagons.) and on the other hand those which do not:

John is the principal investigator.

(Here principal combines with investigator and is used in a non-intersective sense.) We cannot say

*The investigator is principal.

Neither can we capture the meaning of the sentence by intersecting the set of principal things with the set of investigators.

Intensive adjectives such as big, and old represent still another class of adjectives. According to Suppes' analysis (1978) people say little old bear as opposed to old little bear because we consider the set of bears and then impose a partial ordering on them to select the little ones.

little old bear

If we were to select the set of all the little bears, we might well not find any old ones in the set.

Classification through word use. There are two totally distinct ways in which we may hypothesize that the child may achieve a useful partitioning of lexical items. He may start with one large class, words, and then proceed to subdivide this class into noun, verb, adjective, and then continue in this way to subdivide again until the correct partitioning is reached for the production and comprehension of adult language. Alternatively he may begin with as many word classes as he has words, having learned his initial words each as an instantiation of an individual concept. Then as his verbal understanding and expressive power grow, he may begin to combine words into word classes by grouping words according to his own notion (admittedly incomplete and sometimes even wrong in the adult sense) of the way in which words can be combined. It is this second hypothesis which is explored in the model. We call this second hypothesis the Classification by Word Use (CWU) hypothesis.

The CWU process causes the model to posit a multitude of intersecting classes. The classes themselves may be quite different for different children in the course of development, depending upon child's language experience. We may run the model with a given set of input data and examine the words which are classed together and make guesses as to how these classes might be identified with some adult classification scheme, and we may guess that eventually, given sufficient linguistic and conceptual growth, these classes may evolve into those which adults are thought to use; but we cannot predict the exact course for

arriving at this eventual end.

The conservative interpretation of the CWU hypothesis is that it may explain how the child, in interaction with his environment, comes to acquire the classes of noun, verb, and adjective during the early stages of language acquisition so that these classes may provide a basis for building upon.

The radical extension of this CWU hypothesis would be that the broadest classes, those of noun, verb and adjective in their full generality might not be learned at all until encountered in school. This assumes that these classes are not actually necessary in language processing. This CWU hypothesis is a radical departure from the assumption that the classes, noun, verb, and adjective are innate, and that language processing ability is acquired by building on these basic word classes. The worth of this radical interpretation of the CWU depends, in part, on the classes necessary to the formulation of adult language processing which is finally agreed upon, and in part on the route that a child uses to arrive at this formulation.

The CWU hypothesis is supported by the generalization (and, at some point in development, overgeneralization) of plural endings and past tense endings in English. (See Selfridge [1982] for an approach to modelling this phenomenon.) The hypothesis does not deny children a schema for a conceptual class of objects of which there may be more than one, or of actions which may be described as having happened in the past, but these are only two partitions out of the many overlapping partitions which the child may use, and as linguistic analysis has repeatedly shown, both of these devices are hopelessly inadequate as rules for describing the adult classes of noun and verb. I am reluctant to use the terms noun and verb here because of all the implications they carry in adult terms. There is, of course, a physical object class, but this is not a noun class in the general sense. It does not include people, or water or air or fire. It is not even a "person, place or thing" class, and this definition is often cited as an inadequate one for defining noun. It is correct to say that a child makes use of a noun class and a verb class if we are careful to limit the definition of noun to those words which the child pluralizes by the adding of s and we limit the definition of verb to those words to which he adds ed in order to form a past tense. These are, in fact, two examples of exactly the kind of usage rule based on meaning which the model uses to form its many intersecting word classes. However if we do use noun and verb in this sense, then already some confusion has arisen because these two classes do not exist before the child has generalized plural and past tense endings, and once in existence, the noun class, for example, will not contain proper nouns when these come to be distinguished from common nouns, or mass nouns when these come to be distinguished from count nouns (milks), or gerunds (the hurrying), and the verb class, for example, will not include the infinitival (tenseless) forms or the verb to be (copula). There is evidence that at first the child adds ed to action verbs only (deVilliers and deVilliers, 1978). My hypothesis is the

relatively radical one that the child may acquire a multitude of small word classes by projecting word classes based on word use.

Noun, verb, and adjective classes are far too powerful for the young child's language and, if encoded in the grammar, would permit all sorts of constructions that the child would not say. In addition, these classes as formulated for the adult would exclude things that the child does say. There are many examples of word combinations that are present in adult language but which are lacking in child language even though all the words may exist in the child's lexicon. An example of this is that if the template, verb noun were provided, this would produce want help, in the sense of asking for assistance, but Claire, for example, did not at first combine want with anything except a word for a physical object. (Note that if the child did say want go, and want block then the model would hypothesize a lexical class containing block and go.) This is only one example out of many. For this reason it seems best to continue with our paradigm for word classification based on word use. Additional evidence for this paradigm is to be found in Braine (1976) and Ewing (1981).

This procedure is different from the distributional analysis approach explored by Harris (1964) in that it relies on the cognitive and semantic information encoded in the concept space and linked to the templates for forming word classes.

If it can be shown that the word classes can be built up through language experience and language use, then there is no need to assume that the linguistic classes are innate, although a predisposition to form classes is, we suggest, innate. This reflects the underlying assumption that an explanation of the growth of word classification is preferable to an assertion that word classes are simply built-in. The model seems to offer a plausible explanation of how word classes might grow. Instantiation of the template gives a projection of classes. Without being given any information about word classes per se, the model projects a set of classes based on word meaning and word use which are useful to model the language used by children in the age range of 18 to 24 months. These classes are used in the progression from the two-word level to the four-word level. I hypothesize that this process may remain useful to the child as his language abilities grow.

Note that the classes assigned to words in the lexicon depend critically on the model's language experience in the following sense. At one point class(i) may represent both the set of objects which can be shown and the set of objects which can be given (which might be only things that can be handed over). When the model learns the word fire, however, if fire for it refers to the fire in the fireplace then this fire can be shown but not given, and so a new set class(j) will be formed

class(j) = class(i) U {fire}

Alternatively if the model learns fire to stand for a lighted match, then no need to form a new class should arise, since fire, too can be given, and so fire will simply be added to the existing class(i). In keeping with the cautious approach, relation words are the last to be classified at the abstract level, and so should only be classed together when all the classes of their slot-fillers are the same. This means that all slot-fillers must be classified in abstract classes before relation words can be classified. This is another area in which the same body of input data will effect different results depending upon the language level that the model has so far achieved.

The way in which children learn word meaning has been carefully studied and the facts in this area may provide some insight into the question of word classification. It has been well documented in the literature that children both over-generalize some word meanings and under-generalize others (deVilliers and deVilliers, 1978). It may well turn out to be the case that in determining word classification, children use two processes, one which progresses from the specific to the more general and another which progresses from the more general to the more specific. In short, one may speculate that two processes are at work simultaneously in word classification as well as in the processes of assigning word meaning. Be that as it may, the present model is aimed at determining the viability of proceeding from the specific to the more general (the CWU hypothesis).

The role of the concept-space in the model. In the model, the concept-space is contained in a space distinct from the lexicon. The information encoded in the concept-space influences:

1. How words will be classed in the lexicon
2. Which concepts may be chosen to fill slots, and
3. How templates can be grouped

The concepts may be grouped according to their properties within the concept-space. The features used to group classes represent basic ontological categories (Keil 1979, Nelson 1973). The model makes no claims as to what these categories actually are, and permits the user of the model to group the concepts in the concept-space and to specify features in that space as he chooses. Concepts for which no features are specified are taken as primitives by the model. The model uses features, whatever they may be, in order to assign classes in the lexicon, and to determine which concepts may fill slots in templates.

The trouble with semantic features in the usual sense is the danger of using the adult view of the word. There has been a great deal of research on the subject of ontological categories and feature representations of meaning. (See Keil [1979] for one point of view.) In particular, a very common feature to specify is +/- animate; yet we are reluctant to attribute this knowledge to

the young child in view of Piaget's findings that children think clouds and bicycles and all sorts of things which move are alive (Piaget 1960). The child does distinguish between people and objects and between young people and older people. Such features may be encoded in the concept-space of the model, and their choice will contribute to the partitioning of words into lexical classes and the classes in turn will determine which words may fill slots. (See the section on gleaning for a discussion of what salience is to the child.)

The model has need of some feature system, but rather than to choose a specific system it seemed wiser to view the model as a vehicle for experimenting with different systems. For this reason no specific system is built-in and any features used must be specified by the user. At the example level the features seem to be, is it showable? giveable? wantable? In other words, these earliest relations are taken to be primitives.

One anecdote from my data emphasizes the danger of attributing adult meanings to the child's words. Nathaniel (19 months) placed a chess piece on its side and said, castle sleeping. To Nathaniel, sleeping may simply mean lying down. Then anything that may be placed on its side may sleep. Almost certainly Nathaniel lacks the concept of consciousness which is needed to define sleep in the adult sense.

To Nathaniel, the orange balloon was a specific balloon. It was the name of the balloon. It happened to be blue. What did orange mean to Nathaniel? It is not safe to assume he merely had the color wrong. The attribute he had in mind may not have been a color attribute at all. He may have had no attribute in mind.

Children are known to use mommy for woman, and daddy for man. Claire used man for the postman; no one else. Surely the problem of choosing features for this early lexicon is frought with danger.

Susan Carey's (1978) scheme for haphazard examples and missing features in the immature lexicon might be equally appropriate for representing meaning in the concept-space. In some instances it seems that even features totally different from the adult's are required.

Overview of the Model Implementation. The computer implementation of the model is in the language LISP, using GRASPER 1.0, a programming language extension providing graph processing capabilities that was developed in the Department of Computer and Information Science at the University of Massachusetts at Amherst by John Lowrance (1978).

As shown in Figure 1, the data structures of the model are divided into four separate 'spaces': lexicon, grammar, concept-space, and present context. The structures are encoded in the four spaces in the form of GRASPER 1.0 graphs. The structure

of these spaces can be seen in more detail in Figures 4, 5 and 6. In the next subsection, I shall describe how the states shown in these figures change through the process of classification. Here I want to make clear the graphical conventions used.

In the lexicon, there are at first only nodes for words (Fig. 4), but as sentences are processed and class information is gathered, nodes representing lexical classes are added (Figs. 5, 6). Membership of a word within a class is indicated in the lexicon by a pointer marked cl (or ex-cl or ab-cl, depending on the stage of the model).

In the grammar-space, the nodes represent templates. Each template consists of at least one keyword and a set of slots. Slots are distinguished from keywords by a dollar-sign prefix. Confidence factors are stored as the value of the template node in the grammar-space. Selection indices are stored as the value of the template node in the universal space.

In the concept-space the nodes represent concepts, relations, and features. The concept-nodes are linked to word-nodes in the lexicon by edges labelled meaning. The relation-nodes are linked to word-nodes in the lexicon by edges labelled r-ptr. Edges between concept-nodes and feature nodes in the concept-space are labelled feature. Edges between relation-nodes in the concept-space and template-nodes in the grammar-space are labelled template. The edges which connect a template-node in the grammar-space to a concept-node or a feature-node in the concept-space bear the name of the slot in the template which the concept may fill, as in $obj in Fig. 4.

In the context-space the nodes encode information about the sentence which is currently being processed: whether it is a statement or a question, and what object, if any, is being pointed to. For examples of context-space contents, see Figures 13 and 14 in Section 6.

The process of classification in the model. At Stage I the classes are example classes and tag a word such as juice as one which may occur as an object of want, because the object juice is indeed something which may be desired. Hence at Stage I words are classified according to the properties of the concepts they represent, and according to their potential for use in combination with other words. Thus if an object may be owned or given or shown, according to the child's conception of what it is to own or give or show, then the word which represents the object will be classified as one which can fill a slot in the templates associated with words own, give, and show. A new word coming in will be linked to some concept and the properties of this concept will determine which words the new word may be combined with.

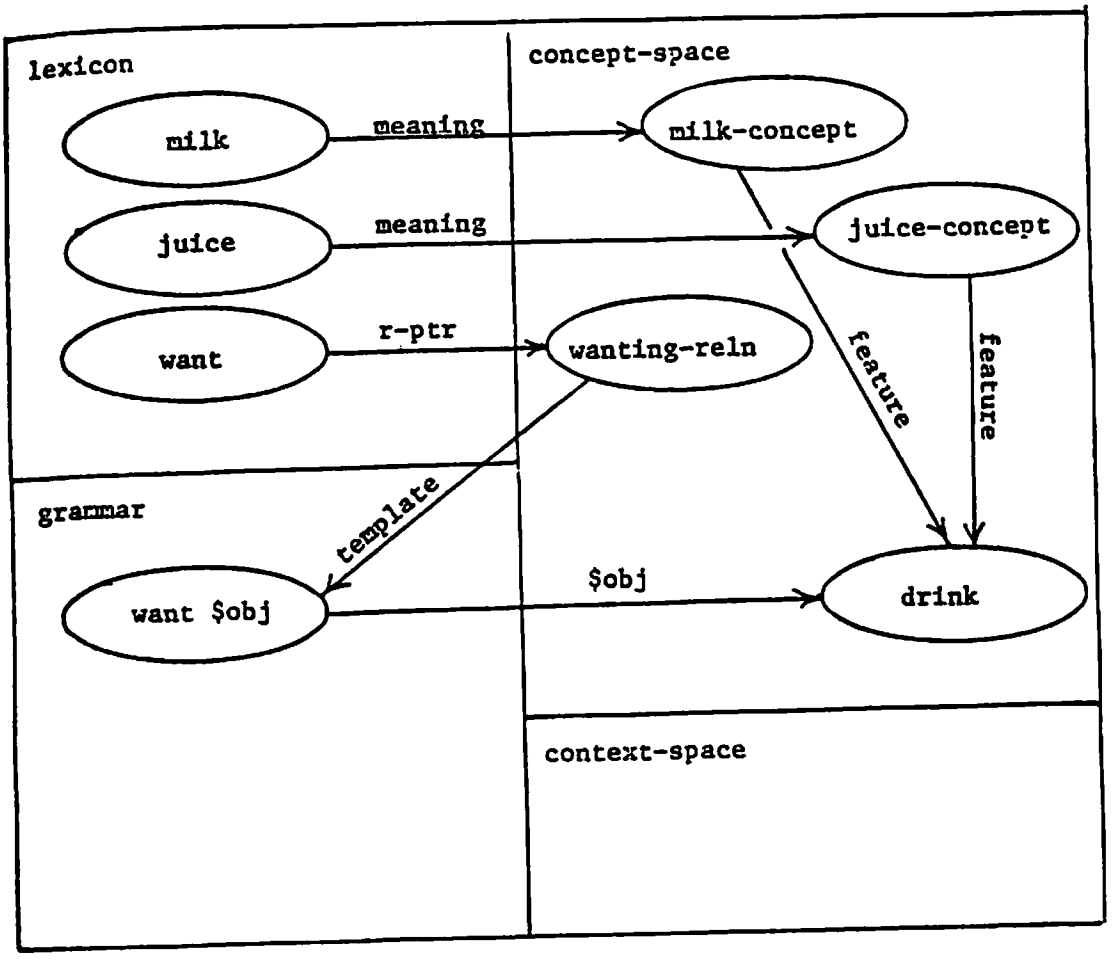


Figure 4. No lexical classification

A threshold called the complexity-constant is provided in the model (the actual value of the threshold may be varied by the user) for the number of example concepts to which a given word may be linked before a general class is formed. When this threshold has been reached, a reorganization is triggered for this word, and the word is assigned to an abstract class which is linked to one or more templates. Figures 4, 5 and 6 illustrate this process, with Figure 4 illustrating the data-spaces of the model before any classification data has been gathered, Figure 5 illustrating the accumulation of specific lexical class data, and Figure 6 showing the data-spaces of the model after an abstract-class has been formed and a direct link set up between the grammar and the lexicon. (In Section 6 below, we shall follow the processing of an example sentence in detail.)

At the earliest level, then, connection to the word is made from the grammar to the lexicon by way of the concept-space (Figs. 4, 5). A new word, when entered, derives its class from its conceptual feature set, and from the template in which it is used. Once the abstract-class level has been triggered for a word, a lexical class is formed with direct links to the lexicon from the grammar (Fig. 7). (If such a class already exists, the word is added to the existing abstract-class that is the smallest class which connects to the appropriate set of conceptual features and templates.) This abstract lexical class, though formed from the conceptual information, is now independent of it, and words may be dropped or added to the class, causing the class to be a true grammatical class and different from a conceptual one. This will be a necessity if the model is to serve as a basis for more mature language. As an example, consider the classes of mass nouns and of count nouns. These might be classed together by the young child, but they must be placed in separate classes as soon as the rules forming their plural forms are learned. Consider for example peas and corn:

some corn
*some corns
*some pea
some peas

As stated before, this process of classification results in a great many word classes and we assume that adults and children both use a large number of word classes in their grammars. A given word may fall into several classes and these classes may overlap, as is illustrated for a fragment of a lexicon in Figure 7.

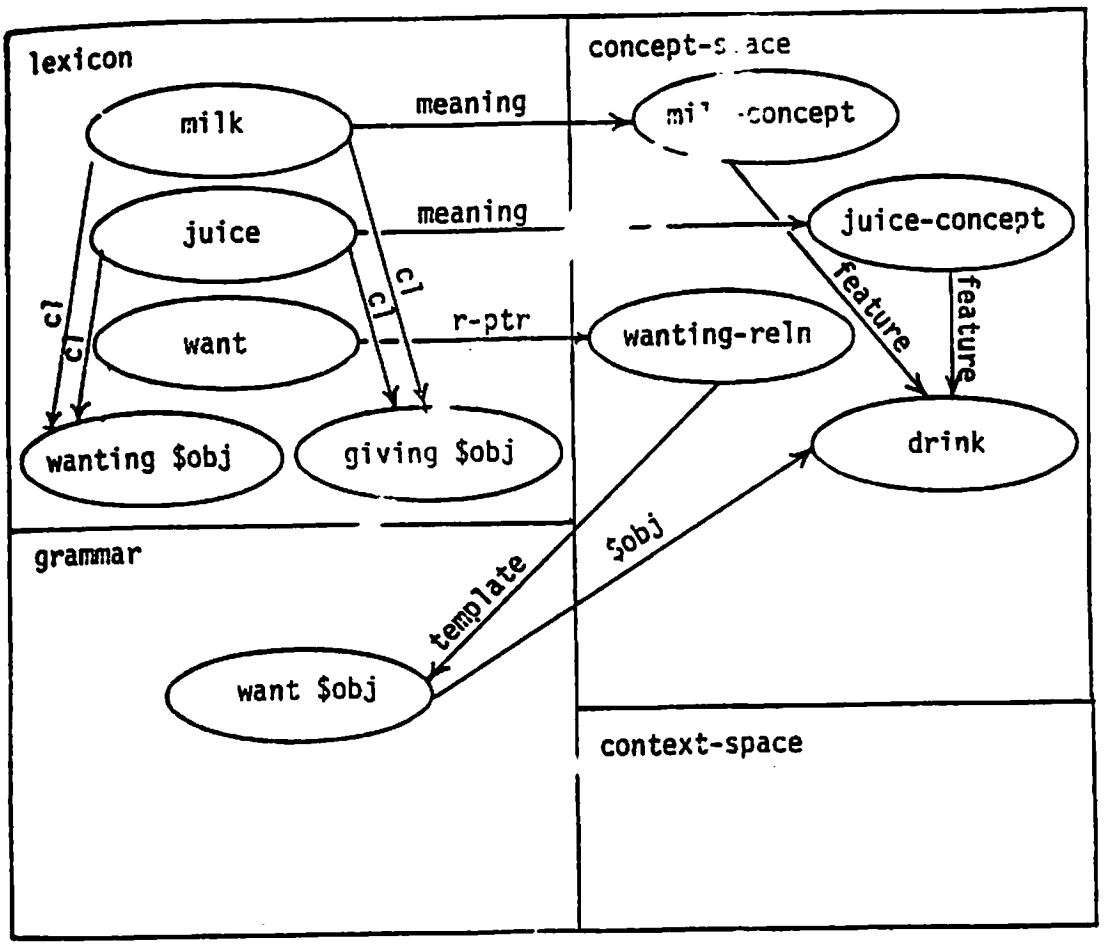


Figure 5. Accumulating specific lexical class data

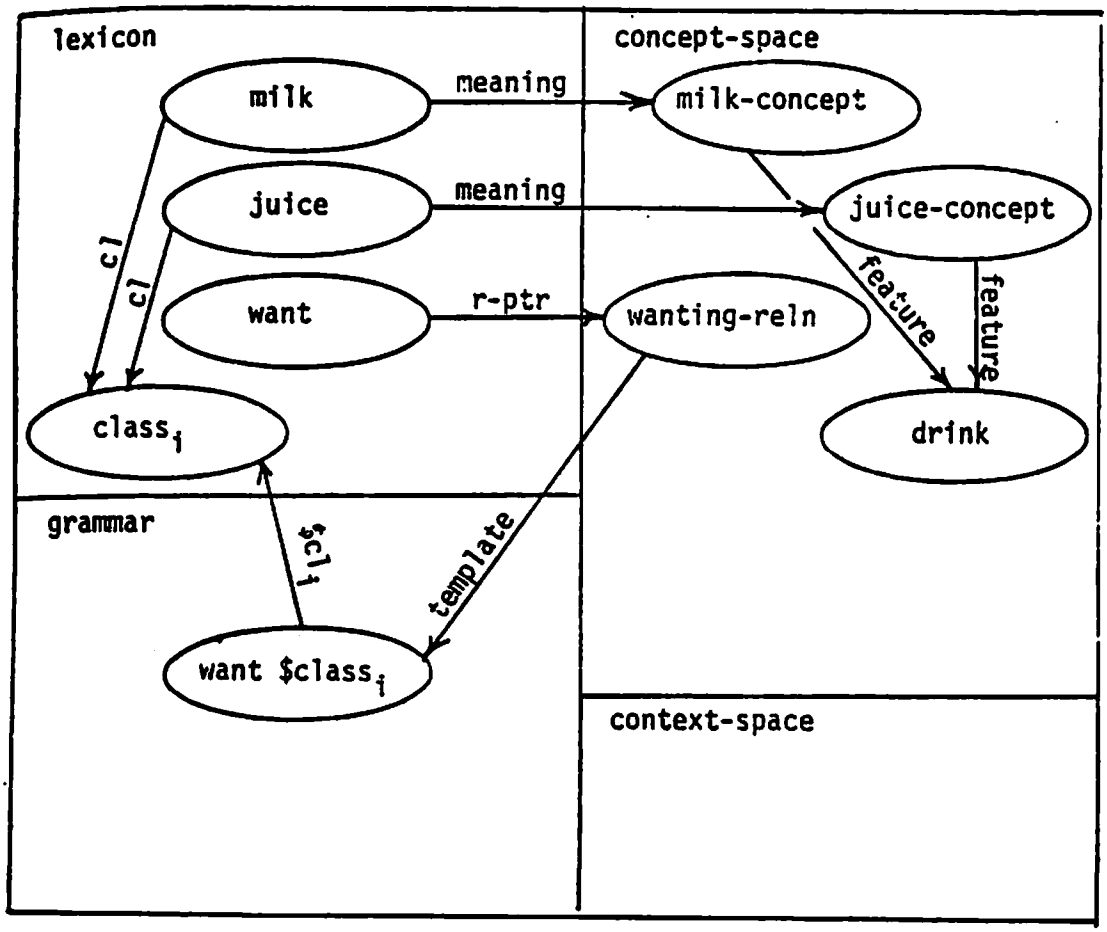


Figure 6. Direct link setup between grammar and lexicon after an abstract-class has been formed

	owner	owned object	given object	shown object	loved object
daddy	x			x	x
mommy	x			x	x
kitty	x	x	x	x	x
dog	x	x	x	x	x
baby	x		x		x
book		x	x	x	
picture		x	x	x	
truck		x	x	x	
fire		x		x	
ladder		x	x	x	
light		x		x	
car		x	x	x	

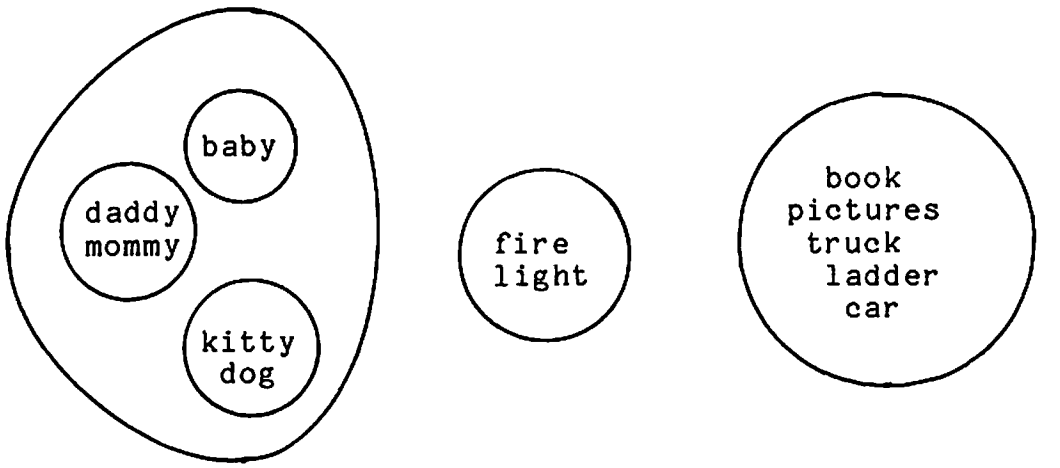


Figure 7. Classifying a fragment of the lexicon.

According to the adult conception of language, all the words in Figure 7 would be classified as nouns, but there is no reason to suppose (1) that the child at this language level is aware of or has need of this superordinate grouping or (2) that adult rules use the class, noun, rather than such groupings.

Note that the columns in Figure 7 represent the example-classes while the rows represent a grouping of the slots in templates which a word-class may fill. Each column contains words for all those concepts that by the child's definition may be owned, may be an owner, may be given, etc. It is the unconscious process of noticing differences as well as similarities (what classes a word may not belong to) that gives the information in the rows and offers the sets pictured below the matrix.

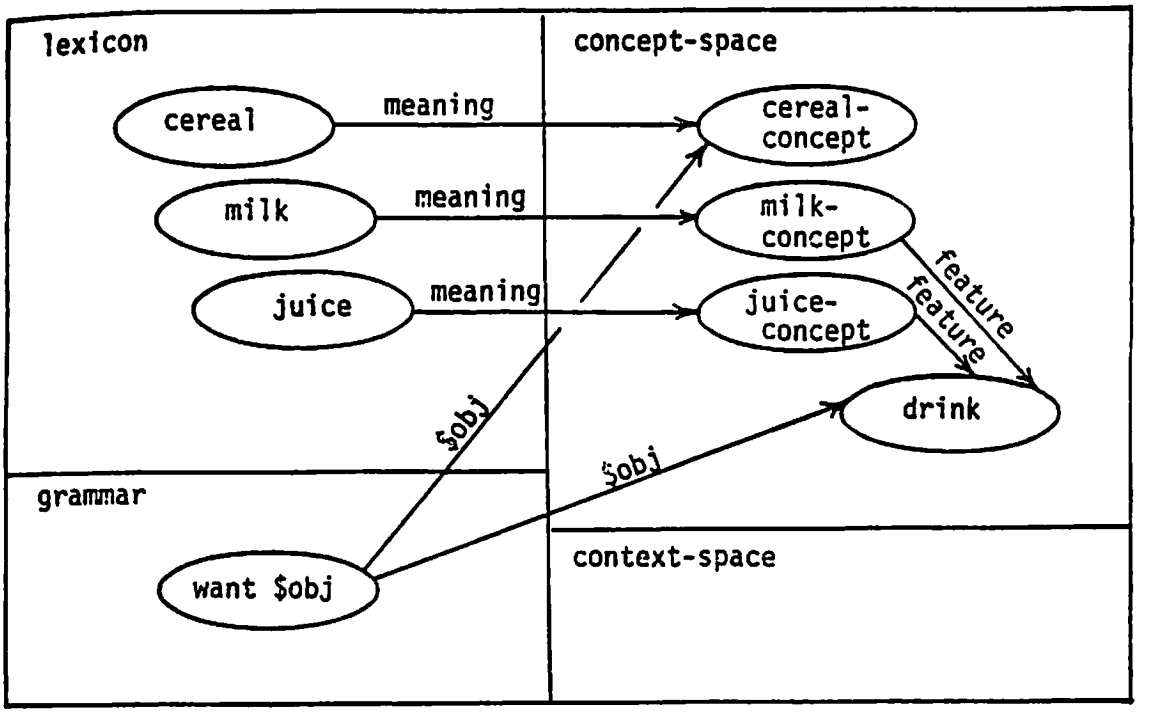
The model classifies both by column and by row. Figure 8 illustrates the state of the model's data spaces both before and after column classification, and Figure 9 illustrates the state of the lexicon both before and after classification by row. As Figure 8 shows, the process of column classification in the lexicon causes the grammar to be modified and direct links between the concept-space and the grammar to be broken. On the other hand, the process of row classification pictured in Figure 9 affects only the lexicon, though the information being accumulated will eventually be used in the forming of template classes which will enable the system of constraints on the concatenation of templates.

Rules for Template Growing

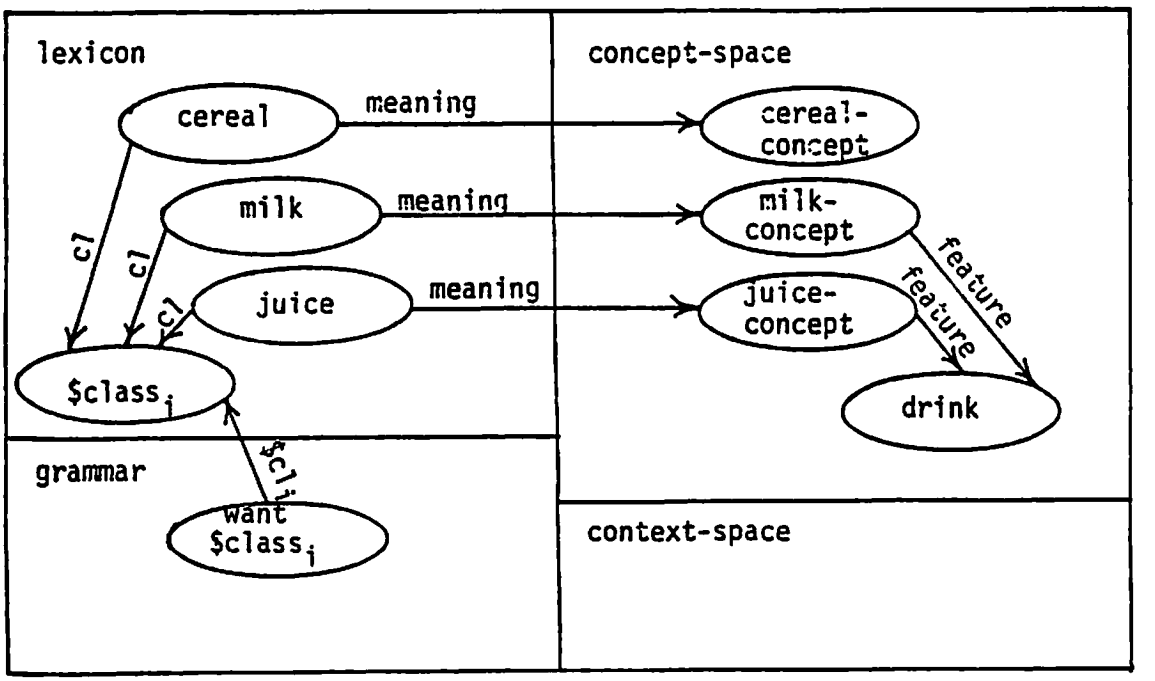
The model, having gleaned a set of two-word example-templates from the input data, will proceed to generalize these templates into abstract and then general templates. Simultaneously it will combine templates to produce longer templates. By template growth I mean the process whereby two-word templates are combined to produce three- or four-word templates. Several processes may be identified which the child employs to progress from the two-word level to the three-and-four-word level. He acquires more example-templates of a richer variety, some two-word and some three-word. He concatenates templates to form longer utterances and in some instances he collapses longer sentences to form shorter ones.

Of course, at the same time that old templates are being combined to form longer utterances, a set of new templates is being acquired. All these processes are at work simultaneously. Thus the three-to-four-word levels include an enriched set of two-word templates as well as some three-word templates. As more concepts are added to the concept space and more words are added to the lexicon, the templates grow in richness. (This fact suggests a correlation between size of vocabulary and complexity of language structure. Whether this is true is a subject for future research. Certainly different word types like prepositions, articles, pronouns, and conjunctions must result in more complex structure.)

Concatenating and collapsing templates. For the convenience of the reader I repeat in Figure 10 the figure which defines the template types.

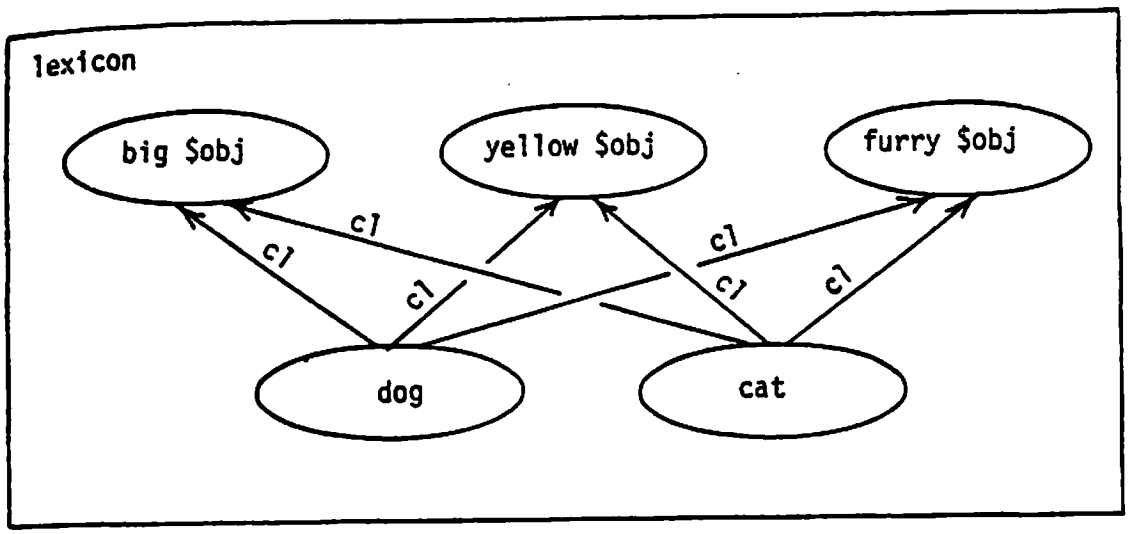


Before

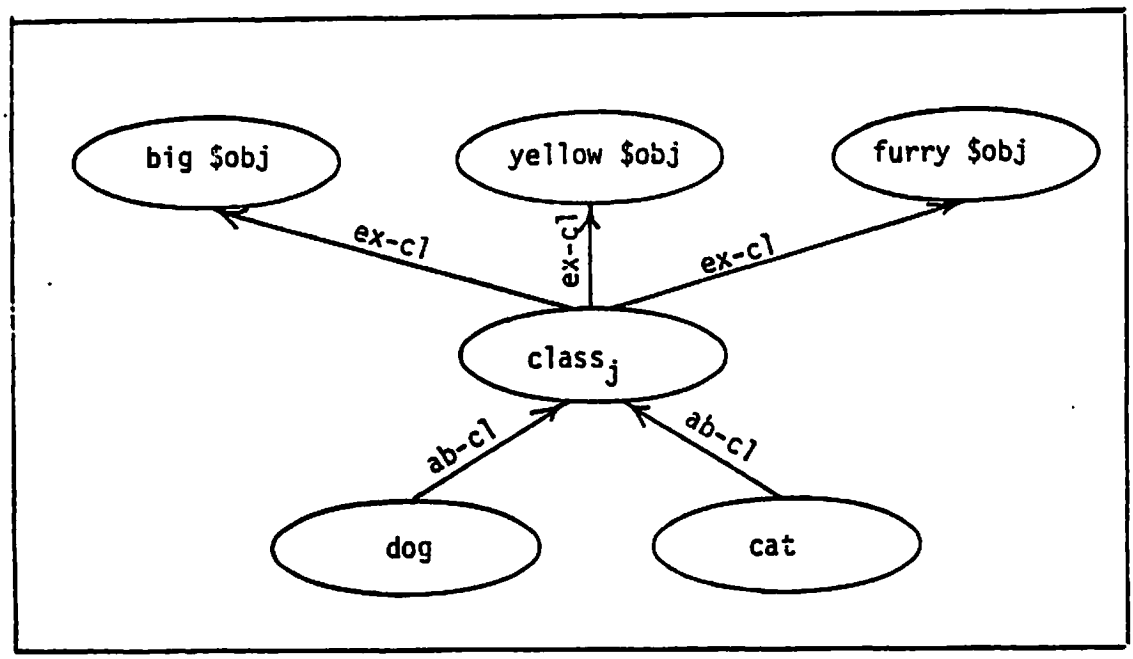


After

Figure 8. column classification: grouping words which may fill template slots



Before



After

Figure 9. Row classification: grouping template slots which may be filled by word classes

Statements		Questions	
s-1	{ give / want / read } object	q-1	where { person / object / animal } ?
s-2	object { on / off / up }	q-2	{ person / animal } doing?
s-3	{ big / little / hot } object	q-3	{ person / object / animal } go?
s-4	{ Mommy / daddy / animal } object	q-4	what doing?
s-5	{ theresa / thatsa / thisa } object	q-5	what that?

Figure 10. First templates defined.

From a logical point of view one would expect template types s-3 and s-4 to grow to represent noun phrases, and template types s-1 and s-2 to to make use of template types s-3 and s-4 as slot fillers to represent the beginning of (1) sentence predicates (e.g., want big box) and (2) longer locative phrases (e.g., Claire sock on). These expectations are summarized in Figure 11. In fact, however, none of these constructs were found in the data except for the combination (s-1(s-3)) which was found only rarely (e.g., want cottage cheese, and make another one). Instead, there were apparently two procedures employed by the child:

1. Linear concatenation of templates
2. Expansion of two-word templates to three-word templates.

Figure 11 summarizes the types of combinations which were found and which were not found.

No Embedded Templates Were Found

(want (sock off))
(s-1 (s-2))

(want (big bear))
(s-1 s-3))

(want (daddy shoe))
(s-1 (s-4))

((blue sock) off)
((s-3) s-2)

((mommy shoes) off)
((s-4) s-2)

Concatenated Templates Were Found

(two forks daddy forks)
(s-3.s-4)

(theres-a one new one)
(s-5.s-3)

(theres-a one mommy chair)
(s-5.s-4)

(little bear baby bear)
(s-3.s-3)

Expanded Templates Were Found

(Jane read porcupine)
(s-1)

(daddy in chair)
(s-2)

Figure 11. Template combinations which were not found and those that were, in the Claire data.

A number of instances of the concatenation of template types s-3, s-4, and s-5 occurred during session three of gathering the Claire data:

more one daddy one (s-3.s-4)
theres-a new one mommy chair (s-5.s-4)
little bear baby bear (s-3.s-3)

As was discussed previously, these longer templates (sometimes with repeated lexical items) were rapidly transformed into three-word utterances. The data which occurred in this short interval led us to speculate that three-word utterances such as two daddy forks were arrived at (1) by concatenating of two templates (two forks and daddy forks) and (2) by collapsing the concatenated templates into a single three-word template (two daddy forks).

The other procedure at work was the expansion of template types s-1 and s-2 by the addition of a single word. For template type s-1 an agent was added preceding the relation-word, thus expanding this template to a subject, verb, object construction (Jane read porcupine). To template type s-2 an object was added following the relation-word thus expanding this template to a noun with modifying preposition phrase (daddy in chair). Thus type s-1 relation-word slot-filler and type s-2 slot-filler relation-word both went to a form of slot-filler1 relation-word slot-filler2. As discussed above, these expanded type s-1 and type s-2 templates can also be hypothesized to be instances of concatenation and collapsing.

Template growing, then, proceeds (1) by concatenating and collapsing two templates to form a new three-word template and (2) by simply adding a word to an existing two-word template.

What are the rules which determine which of these processes may apply?

If at the example-template stage, two-word level it is true (as I hypothesize) that the templates have not been separated into different classes, but merely all exist as relation-word slot-filler combinations with information regarding slot-filler relation-word order being gleaned from the input data, then we must conclude that by the time the templates are concatenated some differentiation must have taken place. There is certainly no logical or semantic barrier to such concatenations as little dog chase cat (s-3.s-1) yet it is generally agreed that this construction occurs much later than dog chase cat or dog chase little cat.

Template types s-3 and s-4 can be distinguished from s-1 and s-2 on semantic grounds in that they represent an entity and attribute. Template s-5 could be distinguished by its initial deictic or pointing word. Since these are the templates which are concatenated and collapsed we tentatively hypothesize that the statement templates should be separated into three groups {s-1, s-2}, {s-3, s-4}, and {s-5} for the sake of template growing. Of the question templates, q-1 and q-3 can concatenate, and q-4 and q-2 can concatenate. This requires, therefore, that at least the abstract template level must be reached before templates can be restricted to the concatenating pairs shown in Figure 12. Note that cognitive differentiation of these template types is required for this grouping, because if one were to look only at word combinations, as in the classification process described previously, most things which are given can be also big or little, and no differentiation would occur. Thus at the three-to-four-word level the templates grow to represent those shown in Figure 12.

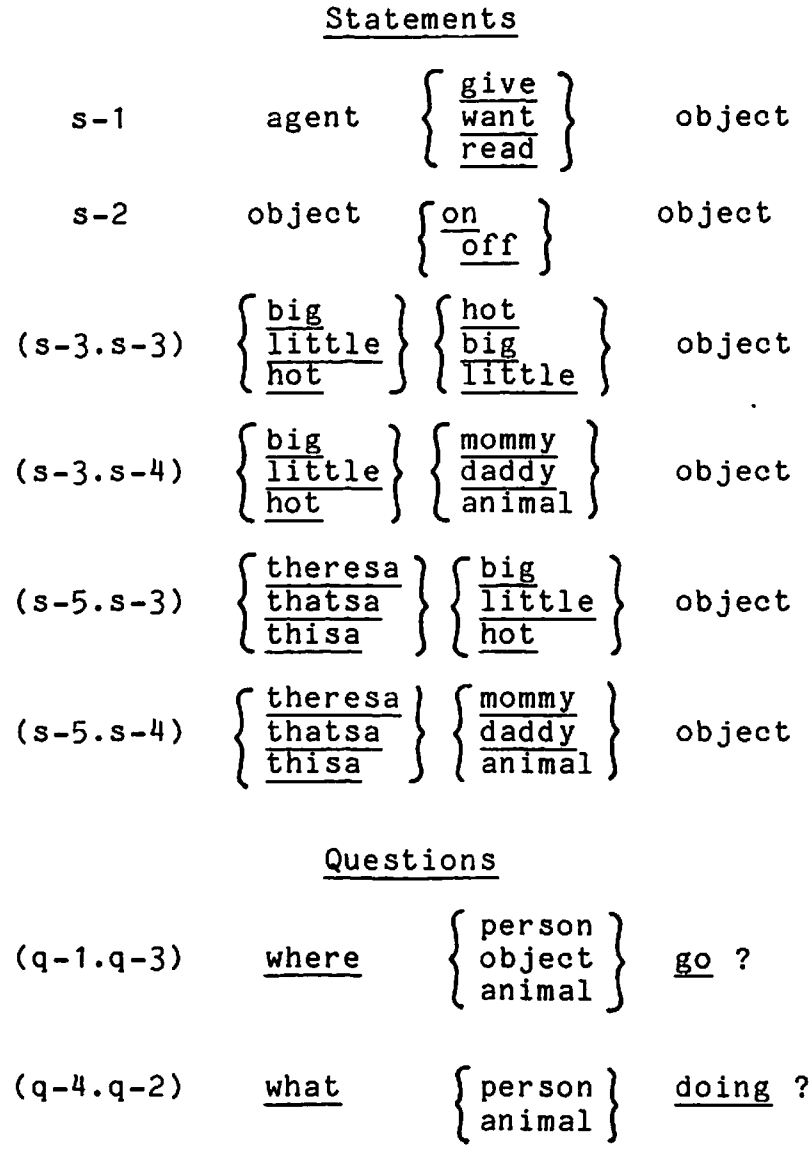


Figure 12. First templates concatenated and expanded.

The specific rules for template concatenating, then, are as follows:

1. There and that if present always start the sentence.

2. Attention to meaning bars such constructs as mommy chair daddy chair (though one might expect to find such a concatenated template pair in a suitably contrived context).

3. Concatenation (s-3.s-3) needs another restriction. The use of such combinations as big big bear when they appear (as discussed previously) have a totally different meaning than just another mommy bear which is an (s-3.s-3) concatenation. Repeated

adjectives must not collapse. They were not generated through concatenation of templates as the others were. When the adjective is repeated for intensifying purposes, this is semantically different from expressing two different characteristics of an item such as second ball and green ball or another bear and mommy bear. I have hypothesized that this repetition for intensification requires mechanisms which may later provide the bridge to recursive rule combination.

Given the constraints on template concatenation as spelled out above, the collapsing rule becomes simply: if a single word occurs twice in a concatenated pair, omit the first occurrence of that word. In Section 2 we discussed hypotheses about why it is the first of the repeated words which is deleted.

In "'A Little Linguist' Model of Syntax Learning", Valian, Winzemer, and Erreich [1981] offer a transformational explanation of a set of sentences with repeated words:

Whose is that is?
What did you did?

Their explanation is that the child has employed a copying tranformation, but has yet to learn the deletion transformation. The explanation for these particular sentences depends on (1) the use of transformations by the two-year-olds, and (2) that there is a rule of wh-movement. Sentences of this type did not occur in the Claire data. The authors do not argue for the correctness of the rules selected for their illustrations, but rather argue for this kind of analysis as a valid way of interpreting children's errors. There is nothing in my model to bar the introduction of transformations at a later stage of development but, on the other hand, there is nothing in the Claire data which requires such an analysis at her stage of language development.

In the model, then, the input data are used to grow the templates according to the rules described here. To summarize, expand templates 1 and 2 or concatenate them as specified in Figure 12 with the added restrictions listed above, and then if a common word occurs in two concatenated templates, collapse them by deleting the first occurence of the repeated word as long as this and that are not deleted from the first position. Once a three-word template has been grown, it may be added to the template set.

The use of confidence factors in the model. How does the model actually deal with erroneous hypotheses? While children are attentive to examples, they tend to be impervious to explicit correction. Adults rarely correct the syntax of children, and when they do, children tend to ignore the corrections. If adult and child understand each other, or think they do, the adult is usually willing to ignore syntactic divergences. Thus the child appears to learn his language from positive evidence alone. Negative evidence (i.e., corrections) in general are not available

to him and are not used if available (deVilliers and deVilliers 1978, Wanner and Gleitman, 1982). Of course, if the utterance is deviant enough that the child is not understood at all, or if he is misunderstood, then this evidence is probably important to the child.

There must be some mechanism in the model that will allow erroneous hypotheses about word order and word classes to be corrected at some time. In addition, there must also be some way that more mature constructs can replace earlier ones. The model accomplishes this by means of a system of confidence factors and selection indices. The use of such factors is by no means unique. Similar weighting schemes have been employed in many models (cf., Samuel 1959, Kelley 1967, Feigenbaum 1977, to name just a few). Even at the example-template stage it is possible for several different templates to be acquired for one relation-word. For this reason a confidence factor is associated with each template and this confidence factor is increased each time a template is matched to the input. In this way, more frequently matched templates can be preferred over (given a higher confidence factor than) less frequently matched templates for any given relation. A template is never disconfirmed and never modified, but new templates are added. The value by which confidence factors are incremented is based on recency, in the sense that the value of the increment grows over time. For lack of a better measure, the passage of time in the model is equated to sentence count. Hence the value of the increment is boosted every n sentences (tentatively n=10). Each instantiation of a template causes the confidence factor for that template to be incremented so that the confidence factors are a function of the number of instantiations of a template weighted by recency. The rationale for this procedure is that growing language capacity will value recent instantiations of a template more highly than less recent ones. When a sentence is processed and a relation-word is found, all the templates for that relation-word are retrieved. They are then ordered based on their confidence factors, and that template with the highest value is selected for matching. If the match fails, then the template with the next highest confidence factor will be selected for matching. This process is meant to simulate the concept that templates must be reinforced to survive, and that templates which are not reinforced may eventually be "forgotten" simply because they fall so far to the end of the list.

No psychological significance is claimed for the actual formula used in the program; many other formulae would do as well. The functioning of the confidence factors is such that if the model has a template for Mommy sweater (from that's Mommy's sweater) and another for sweater Mommy (from that's a sweater for Mommy) both would be retained in the model and their relative confidence factors would depend on how often each structure was matched. Since templates are added but not revised, it is the reinforced templates which survive.

The system as described above would work well for a template set of similar "age," but what of a newly hypothesized template for a given relation? If it is given a confidence value of zero, then it will fall to the end of the list and never be selected for matching. On the other hand, one does not want to give a high confidence factor to an untested hypothesis. For this reason the model employs not only confidence factors, but selection indices as well. When a new template is hypothesized for a relation, the highest selection factor (where selection factor is the product of confidence factor times selection index) is calculated for the set of templates for the relation and then the newly hypothesized template is given a very low confidence factor, but it is given a selection index equal to the highest selection factor for the set of templates plus the recency based increment described above. Thus the desired effect is attained, of forcing newly hypothesized templates to the front of the ordered list of a template set so that newly hypothesized templates will be selected for matching soon after their creation.

6. PROCESSING AN EXAMPLE SENTENCE

In previous sections the templates have been described in detail, as well as the processes of word classification and template growth through concatenation and collapsing. Here I illustrate the functioning of the various modules by describing the processing of an example sentence. I will then close the chapter by exhibiting the way in which the system can respond to questions.

Input sentence: I'll put this daddy right-here.
Context: (daddy-concept isa toy-doll)

The sentence is chosen because it will fit three templates of different sorts. The daddy referred to in the sentence is one of the male dolls in a Fisher-Price toy house set, and daddy is the way that Claire normally referred to the dolls. There were several daddies. Figure 13 provides a graphical representation of the context as well as the following initial knowledge:

In the lexicon:

put
this
daddy
right-here

Templates:

put $obj
$obj right-here
this $obj
$owner $obj

(Recall that the words preceded by $ represent slots.)

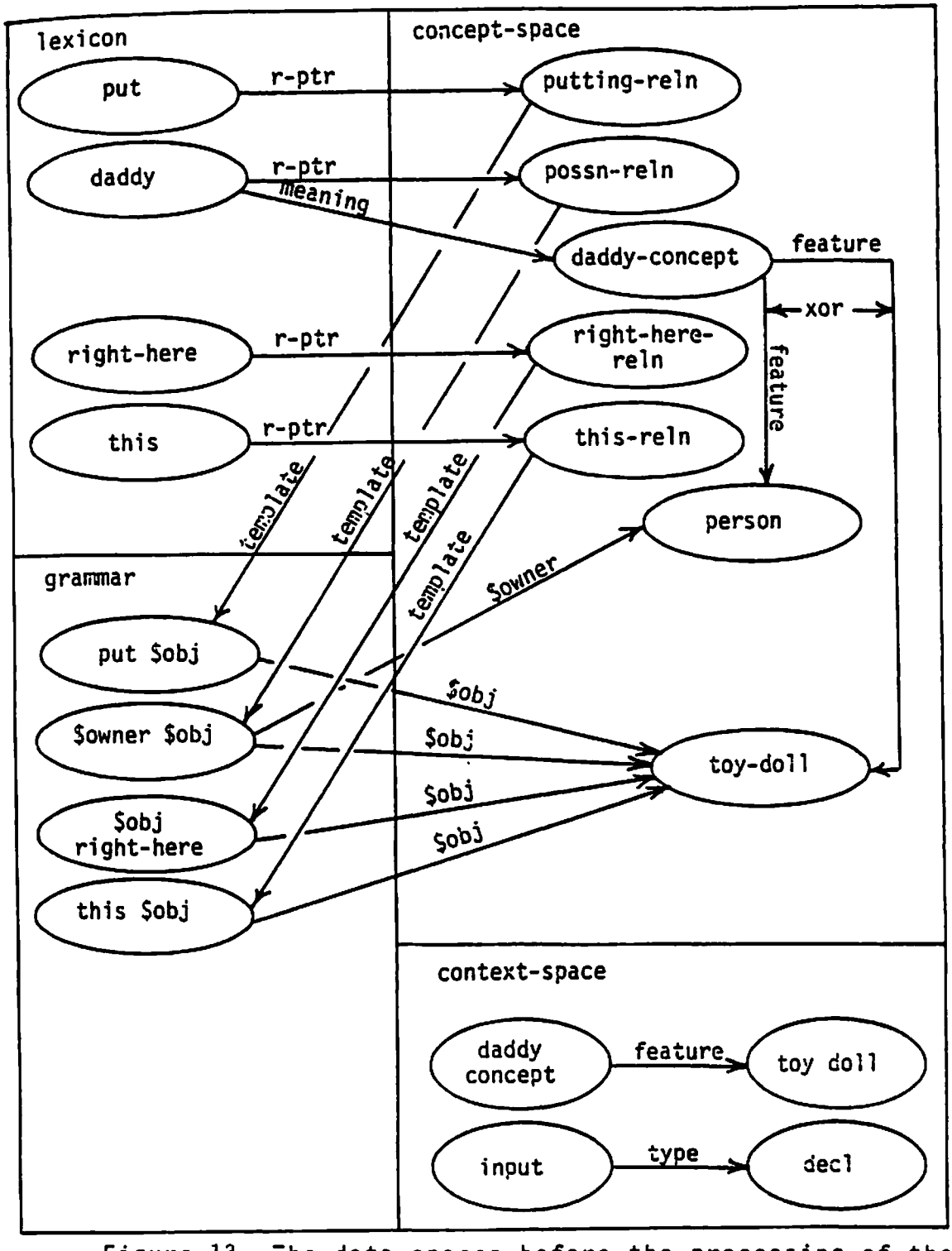


Figure 13. The data-spaces before the processing of the sentence

Cognitive knowledge:

what it is to put.
what is meant by right-here.
what is meant by this.
that toy dolls can be put
that toy dolls can be pointed to as this
that toy dolls can be right-here.

At the first pass, let us begin at the two-word level, example-template stage (Stage I). The sentence is processed left to right, word by word. I'll is examined by the find-template routine and ignored because it is an unknown word. The child at this point in her development had neither I nor you in her lexicon. It is not clear if it belonged in her concept-space or not.

Put is examined by the find-template routine, found in the lexicon, and determined to be a relation-word. The only template for put is discovered to be put $obj. Had there been more than one template for put, then that template with the highest selection factor would have been chosen. Recall that selection factors are formed by calculating the product of confidence factors which represent instantiations of a template times a selection index which boosts the value of a newly hypothesized template for the purpose of causing it to be tested.

Having found a template, the fit-template routine is called. The word, put is found to match the first word of the template, so fit-template checks to see if a word can be found in the sentence which would be meaningful as the object of put.

Fit-template examines this to see if it is a meaningful object of put. At this point the model uses this only as a relation-word at the start of a template, so this is rejected. At a later date, when this has a larger meaning the sentence put this might well be formed. For example purposes, however, we will reject this and go on to check daddy. Daddy, because of its toy doll meaning (the one selected by context), will be accepted as a slot-filler for the template, and put daddy will be placed in the "memory buffer" (which is temporary storage for templates which are matched in the processing of a sentence) as a possible utterance.

Apply-template is entered, and apply-template will increment the confidence factors associated with the template and with the word, daddy, as an object of put. The value of the increment is a function of recency, so that more recent instantiations are given a higher value than less recent ones. Since we are at the two-word level, no larger template can be gleaned. For the sake of simplicity we will assume that no abstract class is formed at this point. Daddy will be entered as a pointed-to entity in the context-space. Then the sentence is checked to see if more words remain. There are more words to be processed in the sentence so control reverts to find-template where the word this is found.

Again, this is found to be a relation word, and similar processing takes place in apply-template. This time (this daddy) is added to the memory buffer which now contains (put daddy)(this daddy).

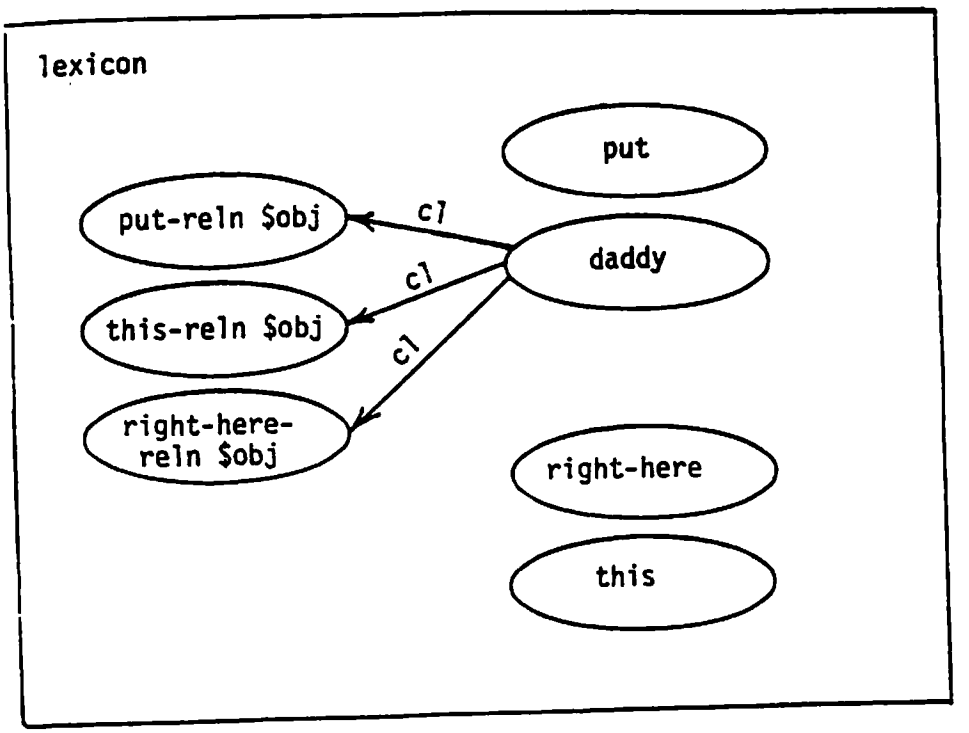
Now daddy is examined as a relation-word, but there is no word found to be acceptable to fit the owned-obj slot, so no new sentence is added to the memory buffer, and right-here will be entered in the context-space as the place associated with daddy.

Right-here is examined. It is a relation-word and after the processing of the template in apply-template, (daddy right-here) is placed in the memory buffer. The buffer now contains (put daddy) (this daddy) (daddy right-here). Because we are at the two-word level, only (daddy right-here) is output by the model. It is chosen because it is the last sentence in the buffer.

Figure 14 shows the class information that has been added to the lexicon as a result of processing this sentence (top), and the information that has been added to the context-space (bottom).

On a subsequent pass of the model, if the sentence-length-limit has been raised by the user to permit concatenation and collapsing, then the same sentence might produce the output (this daddy right-here) (3-2) provided that the templates have been generalized into classes, and the model permits concatenating and collapsing of these two types. Once this sentence has been processed, the function collapse could form a new template type (s-3.s-2) This $obj right-here. On the next iteration of this sentence the three word template would cause the buffer to hold (put this) (this daddy right-here). Finally at some future stage, the model would be able to output the sentence (put this daddy right-here).

Responding to Questions. As we have just seen, the model will produce different results at different levels depending on the language level attained. To illustrate this fact Table 3 presents a list of the sets of responses which the model is capable of producing in reply to the questions posed.



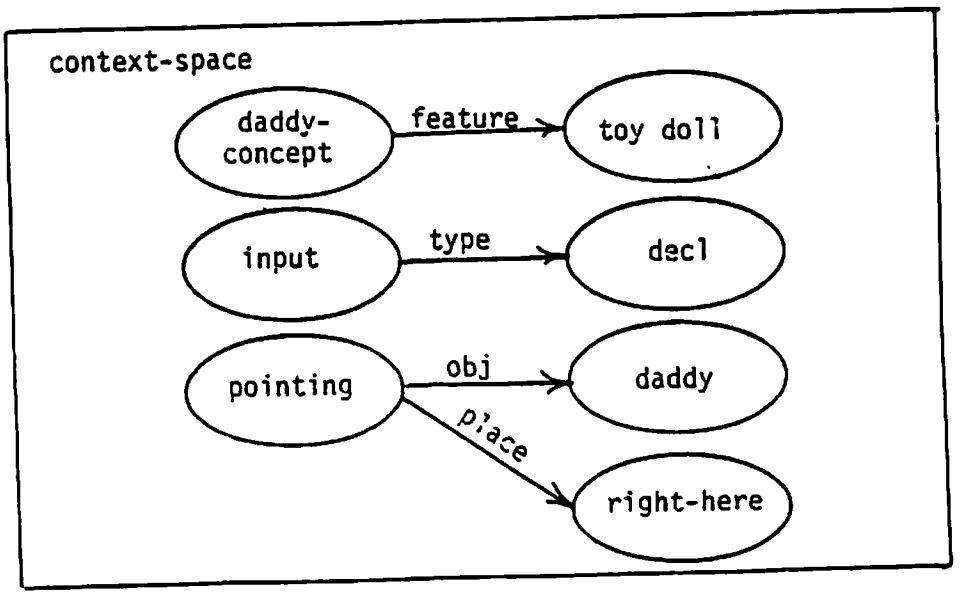


Figure 14. The lexicon and context spaces with the information which was added as a result of processing the example sentence

Input sentence	Claire's response	The model's responses
Oh Claire, is your house shut up?	house shut up	house shut house shut up house open house open up
What's the kitty cat doing?	what doing?	cat doing? kitty cat doing? what kitty cat doing? what doing? kitty riding kitty riding bicycle
Where's the doggie	where doggie?	where doggie doggie outside
What's this?	What this?	what this thatsa rooster
Where's Daddy?	right-here	where daddy? daddy right-here
What room is this?	bed	what this what room thatsa bed

Table 3. Table of Claire's responses contrasted with sets of possible model responses at different stages of the model.

The model's response to questions is highly dependent on the available contextual information encoded in the context-space. At present the model can respond to the following different kinds of questions. Note that input questions are identified by their terminating question marks.

1. where x ?
2. what doing?
3. what this?
4. yes/no

In the model where questions are defined to be those that begin with the word where. What doing? questions are defined as being those that begin with the word what and contain the word doing. What this? questions are defined as those which begin with what and end with this. All others are treated as yes/no questions.

It is significant that although every effort was made to avoid giving the model a partitioning of words into action classes and object classes, or into nouns and verbs, our model was forced at least to define a pointed-to object and an action-set in order to form answers to the what and where questions. The model's definition of things that can be pointed to is somewhat different from a physical object class in that it can include such things as sky and clouds, and the model defines actions as those relations whose relation-words may end in ing. Both of these classes can be defended in terms of Claire's pointing behavior and her use of ing endings. However, that the model was forced to define these classes in order to respond to what and where questions is very interesting. Until Session 6, Claire did not respond to these questions, except to repeat them. Brown (1973) observed that young children do not respond appropriately to these types of questions, but use the pragmatics of the situation to determine their answers.

The context-space is altogether necessary to the question-answering function. The processing of each sentence causes information to be stored in the context-space. The first object in a sentence that can be pointed to is stored in the context-space as the pointed-to object for that sentence, and any place-word in the sentence is stored as the place for that pointed-to object. In addition explicit contextual information can be given the model together with the input sentence.

If the input question is a where question, then the context-space is searched and the place word for the currently pointed to entity is encoded in the response. For example

Where x?

produces a response

x place-word

If the question is a what question and the question is

What doing?

then the answer to

What x doing?

is

x doing y

where y is the action attributed to x in the context-space. If the question is a what question, then it is examined to see if the question is

What's this?

If so, the answer is

Thatsa x

where x is the pointed-to entity in the context-space.

All the other questions are assumed to be yes/no questions. The response to yes/no questions is formed as a repetition which may be negated. A negative response is formed in one of two ways. If an antonym is encoded in the lexicon for one or more words in the repetition, then the negative response is formed by replacing the first such word in the sentence by its opposite. For example

Is the big table outside?

might produce the response

little table outside

If no words with available opposites occur in the sentence, then the sentence produced is the repetition preceded by the word, no. As noted above, Claire seemed to form her negations in this way.

7. CONCLUSIONS AND QUESTIONS

Inasmuch as the model assumes a minimalist approach to language acquisition, an important goal in the process of building the model was to be aware of those assumptions and processors which I was compelled to build in -- these may be summarized as follows.

The model assumes the following characterization of two-year old speech:

1. The interaction of cognitive and linguistic growth
2. A great many word classes
3. Word classification based on word use
4. Flat structure
5. Learning that begins with specific examples which evolve into general patterns.

In order to build the model, it was necessary to build in the following processes:

1. Classification of concepts and words
2. Attention to salience as defined for the child
3. Attention to word order
4. Encoding of relations
5. Concatenation
6. Collapsing
7. Gleaning of templates from example sentences
8. Reorganizing of the data structures into different stages of competence.

By making use of the template grammar and the concatenating and collapsing paradigm I was able to show how the model might progress from two-word utterances to utterances which contain four or five content words. Moreover I was able to show that the development of language ability in the model closely paralleled the development of language which was found in the Claire data.

Another aspect of the model that is worth noting is its modularity and the ease with which it can be extended. The model is a flexible tool for experimenting with different formulations. It is not limited to the template grammar: the template module could be replaced by a phrase structure module; transformational rules could be included other than concatenation and collapsing.

A second advantage of the flexibility of the model is the ease with which the various functions can be turned on or off. The following is offered as an illustration of the sort of ideas that can be generated from experimenting with the model.

My theory from examining the empirical data was that three-word templates should be gleaned from examples in order to get such sentences as

Mommy drinks tea.

However, running the model with the three-word gleaning function inactivated, the model produced this sentence:

Mommy drink tea

from concatenating

(mommy tea)(drink tea).

This led me to speculate that the process of combining pairs of two-word templates may be a source of some of the early three-word sentences with animate subjects.

Another illustration of this sort of experimentation is to be found in the concatenation of single words. On a run in which the model was permitted to concatenate single words instead of pairs or triples of words expressing relations, the very first sentence

Oh Claire, is your house shut up?

produced on the first iteration

Claire up

and on the second iteration

Claire shut.

These are both clearly wrong, and the theory that it must be at least two-word relations that are to be concatenated is strengthened.

It is not yet clear if templates should be grown to lengths longer than three words. Should the model grow longer templates or develop more rules for combining templates, or some mix of each? It is clear, for example, that a template

slot1 relation-word slot2

is useful, but is there any point in creating a template

relation-word slot1 relation-word

where the same effect could always be achieved from processes applied to the two different relation words? The trade-off of course is between storing more templates or alternatively performing the same processes repeatedly. The answer may vary and may depend on the frequency with which a longer template is used. (See Kuiper, 1980 for a discussion of template-like constructs

used by auctioneers. See Moulton and Robinson, 1981, for a model of language that describes the organization of language in terms of pairs.) This question is clearly a subject for future research.

In the course of describing the model, several points have been suggested as interesting directions for extending the model.

1. The model should learn to use articles; the use of articles will force a reorganization of the template formalism, since articles can be viewed neither as relation words, nor as slot fillers.

2. The model should learn to generalize past tense and plural endings; noticing word endings will require a different set of learning processes than any that are currently in the model.

3. The effects of using a different grammar for production than that for comprehension could be explored.

Additionally my own interest inclines me to explore the child's use of conjunctions. I would like to collect data on the set of conjunctions first used by the child, and on the constructs first conjoined. When does the child start to employ but as well as and, and does he appreciate the different semantics of these words? Not only are conjunctions interesting from a semantic point of view, but their inclusion in the model would force the use of a recursive grammar. In addition I am interested in conjunctions because of the additional knowledge of structure needed before a child can form such sentences as

Mary and Bill played

or

Mary played and sang.

These sentence seem to be formed by using concatenating and collapsing rules on the original simple sentences:

Mary played. Bill played.

and

Mary played. Mary sang.

I would like to explore the relation of the rules needed for this process to the rules currently encoded in the model for concatenating and collapsing relations.

There is no doubt that the simple mechanisms proposed here as adequate for the language capacity of the two year old will prove to be inadequate for language of any complexity. It should be illuminating to experiment with the current formalism in order to discover just how far it can go, and when and where it breaks

down.

Of course the mysteries of the process of language acquisition will not be solved until all the data are in from the study of many different children learning the languages of many different countries. It is my hope that the data presented here and analyzed in the form of the template grammar will aid in the understanding of how a young child may begin by using a flat structure for a finite, semantically based language which may nevertheless be capable of evolving into an hierarchical structure for the infinite, formal language of the adult.

SELECTED BIBLIOGRAPHY

Arbib, M. A., Conklin, E. J., and Hill, J. C., From Schema Theory to Language, to appear.

Bates, E., Language and Context: The Acquisition of Pragmatics, Academic Press, New York (1976).

_____, "The Emergence of Symobls: Ontogeny and Phylogeny", in The Minnesota Symposia on Child Psychology, W. A. Collins, ed., 12 (1979):121-156.

Bloom, L., Language Development: Form and Function in Emerging Grammars, MIT Press, Cambridge, Massachusetts (1970).

_____, One Word at a Time: The Use of Single Word Utterances Before Syntax, Mouton, The Hague, (1973).

Bower, T. G. R., Development in Infancy, W. H. Freeman, San Francisco (1974).

_____, A Primer of Infant Development, W. H. Freeman, San Francisco (1977).

Bowerman, M. "Early Syntactic Development, A Cross-linguistic Study With Special Reference to Finnish", Cambridge Studies in Linguistics, no. 11, Cambridge University Press, London (1973).

_____, "Learning the Structure of Causative Verbs: A Study in the Relationship of Cognitive, Semantic, and Syntactic Development", in Papers and Reports on Child Language Development, E. Clark ed., no.8, Stanford University Committee on Linguistics, Stanford, California, (1974), pp.142-178.

Braine, M. D. S., "The Ontogeny of English Phrase Structure: The First Phase", in Language, 39 no.1 (1963):1-13.

_____ "Children's First Word Combinations", Monographs of the Society for Research in Child Development, University of Chicago Press, 41, no.1, (1976).

Brainerd, C. J. Piaget's Theory of Intelligence, Prentice-Hall, Englewood Cliffs, New Jersey (1978).

Brown, R. A., A First Language: The Early Stages, Harvard University Press, Cambridge, Massachusetts (1973).

Butterworth, G. E., "The Development of the Object Concept in Human Infants", D. Philosophy Thesis, University of Oxford (1974).

Carey, S., "The Child as Word Learner", in Linguistic Theory and Psychological Reality, M. Halle, J. Bresnan, G. Miller, eds., MIT Press, Cambridge, Massachusetts (1978), pp.269-291.

Chomsky, N., Reflections on Language, Pantheon Books, New York (1975).

Conklin, E. J., "Salience as a Heuristic in Planning Text Generation", PH.D. dissertation, Department of Computer and Information Science, University of Massachusetts at Amherst (1983).

Corrigan, R., "Language Development as Related to Stage 6 Object Permanence Development", Journal of Child Language, 5 no.2 (1978):173:189.

Denney, N. W., "Free Classification of Preschool Children", Child Development, 43 (1972):1161--1170.

deVilliers, J. G., and deVilliers, P. A., Language Acquisition, Harvard University Press, Cambridge, Massachusetts (1978).

deVilliers, J. G., and deVilliers, P. A., Early Language, Harvard University Press, Cambridge, Massachusetts (1979).

Ewing, G., "Word-order Invariance and Variability in Five Children's Three-word Utterances: A Limited Scope Formual Analysis", presented at the 2nd International Congress for the Study of Child Language, Vancouver (August 1981).

Feigenbaum E. A. "The Art of Artificial Intelligence: I. Themes and Case Studies of Knowledge Engineering", in IJCAI-5, MIT, Cambridge, Massachusetts, (1977):1014-1029.

Gratch, G., "Review of Piagetian Infancy Research: Object Concept Development", in Knowledge and Development, W. F. Overton and J. M. Gallagher, eds. Plenum Press, New York, vol.1 (1977), pp.59-92.

Gross, M., "On the Failure of Generative Grammar", Language, 55 (1979):859-885.

Gruber, J. "Topicalization in Child Language", Foundations of Language, 3, no.1 (1967):37-65.

Harris, P. L., "Examination and Search in Infants", British Journal of Psychology, 62 (1971):469-473.

Harris, Z. S., "Distributional Structure", in The Structure of Language: Readings in the Philosophy of Linguistics, J. Fodor and J. Katz, eds., Prentice-Hall, Englewood Cliffs, New Jersey (1964).

Hill, J. C., "A Computational Model of Language Acquisition in the Two-Year-Old", Ph.D. Dissertation, University of Massachusetts at Amherst (1982), also to be published by the University of Indiana Linguistics Club.

Horton, M. S., and Markman, E. M., "Developmental Differences in the Acquisition of Basic and Superordinate Categories", Child Development, 51 (1980):708-719.

Kagan, J., Change and Continuity in Infancy, John Wiley and Sons, New York (1971).

Keil, F., Semantic and Conceptual Development, An Ontological Perspective, Harvard University Press, Cambridge, Massachusetts (1979).

Kelley, K. L., "Early Syntactic Acquisition", Ph.D. Dissertation, University of California at Los Angeles, 1967, also published as Report no. P-3719, The Rand Corporation, Santa Monica, California (November 1967).

Kuiper, K., "Minimal Grammar and Stock Auctioneering", in Man Machine Studies, University of Canterbury, Christchurch, New Zealand (December 1980), pp29-50.

Lowrance, J., Grasper 1.0 Reference Manual, COINS Technical Report 78-20, University of Massachusetts at Amherst (December 1978).

Matthei, E., "The Acquisition of Prenominal Modifier Sequences: Stalking the Second Green Ball", Ph.D. Dissertation, Department of Linguistics, University of Massachusetts at Amherst (1979).

Menyuk, P., Sentences Children Use, MIT Press, Cambridge, Massachusetts (1969).

Moore, M. K. and Metzoff, A. N., "Object Permanence, Imitation and Language Development: Toward a Neo-Piagetian Perspective on Communicative and Cognitive Development", in Communicative and Cognitive Abilities--Early Behavioral Assessment, F. D. Minifie and L. L. LLoyd, eds., University Park Press, Baltimore, Maryland (1978), pp151-181.

Moulton, J. and Robinson, G., The Organization of Language, Cambridge University Press, Cambridge (1981).

Nelson, K., "Structure and Strategy in Learning to Talk", Monographs of the Society for Research in Child Development, ser.149, 38, nos.1-2 (February-April 1973).

_____, "Concept, Word, and Sentence: Interrelations in Acquisition and Development", Psychological Review, 81 (1974):267-285.

Piaget, J., The Child's Conception of the World, trans. J. and A. Tomlinson, Littlefield, Adams and Co., Totowa, New Jersey (1960).

_____, Play, Dreams and Imitation in Childhood, trans. G. Gattegno and F. M. Hodgson, W. W. Norton and Company, New York (1962).

_____, Biology and Knowledge, trans. B. Walsh, University of Chicago Press, (1971).

Roeper, T., "On the Deductive Model and the Acquisition of Productive Morphology", The Logical Problem of Language Acquisition, C. L. Baker and J. J. Mc Carthy, eds., The MIT Press, Cambridge, Massachusetts (1981).

Samuel, A. L. "Some Studies in Machine Learning Using the Game of Checkers", IBM Journal of Research and Development, 3 (1959):211-229.

Selfridge, M., "Why Do Children Say 'Goed'?: A Computer Model of Child Generation", published in the proceedings of the "Third Annual Conference of the Cognitive Science Society," Berkely California, August 19-21, 1981.

Siegel, M. E. A., "Capturing the Adjective", Ph.D. Dissertation, Department of Linguistics, University of Massachusetts at Amherst (1976).

Solan, L. and Roeper, T., "Children's Use of Syntactic Structure in Interpreting Relative Clauses", Papers in the Structure and Development of Child Language, H. Goodluck and L. Solan (eds.), University of Massachusetts Occasional Papers in Linguistics, vol.4, (1978).

Sugarman, S., "The Development of Language and Action in Young Children: Evidence from Spontaneous Block Play", talk given at University of Massachusetts School of Education (April 27, 1981).

Suppes, P., "The Semantics of Children's Language", in American Psychologist, 29 (1974):103-114.

Suppes, P., and Macken, E., "Steps Toward a Variable Free Semantics of Attributive Adjectives, Possessives and Intensifying Adverbs", Children's Language, K. Nelson, ed., vol.1, New York, Gardner Press (1978).

Tavakolian, S. L., "The Conjoined Clause Analysis of Relative Clauses and Other Structures", Papers in the Structure and Development of Child Language, H. Goodluck and L. Sol, eds., University of Massachusetts Occasional Papers in Linguistics, vol.4 (1978).

Valian, V., Winzemer, J., and Erreich, A., "'A Little Linguist' Model of Syntax Learning", in Language Acquisition and Linguistic Theory, Tavakolian, S. L., ed., The MIT Press, Cambridge, Massachusetts (1981).

Wanner, E. and Gleitman, L., eds., Language Acquisition: the State of the Art, Cambridge University Press, (1982).

Wexler, K., and Culicover, P., Formal Principles of Language Acquisition, The MIT Press, Cambridge, Massachusetts (1980).