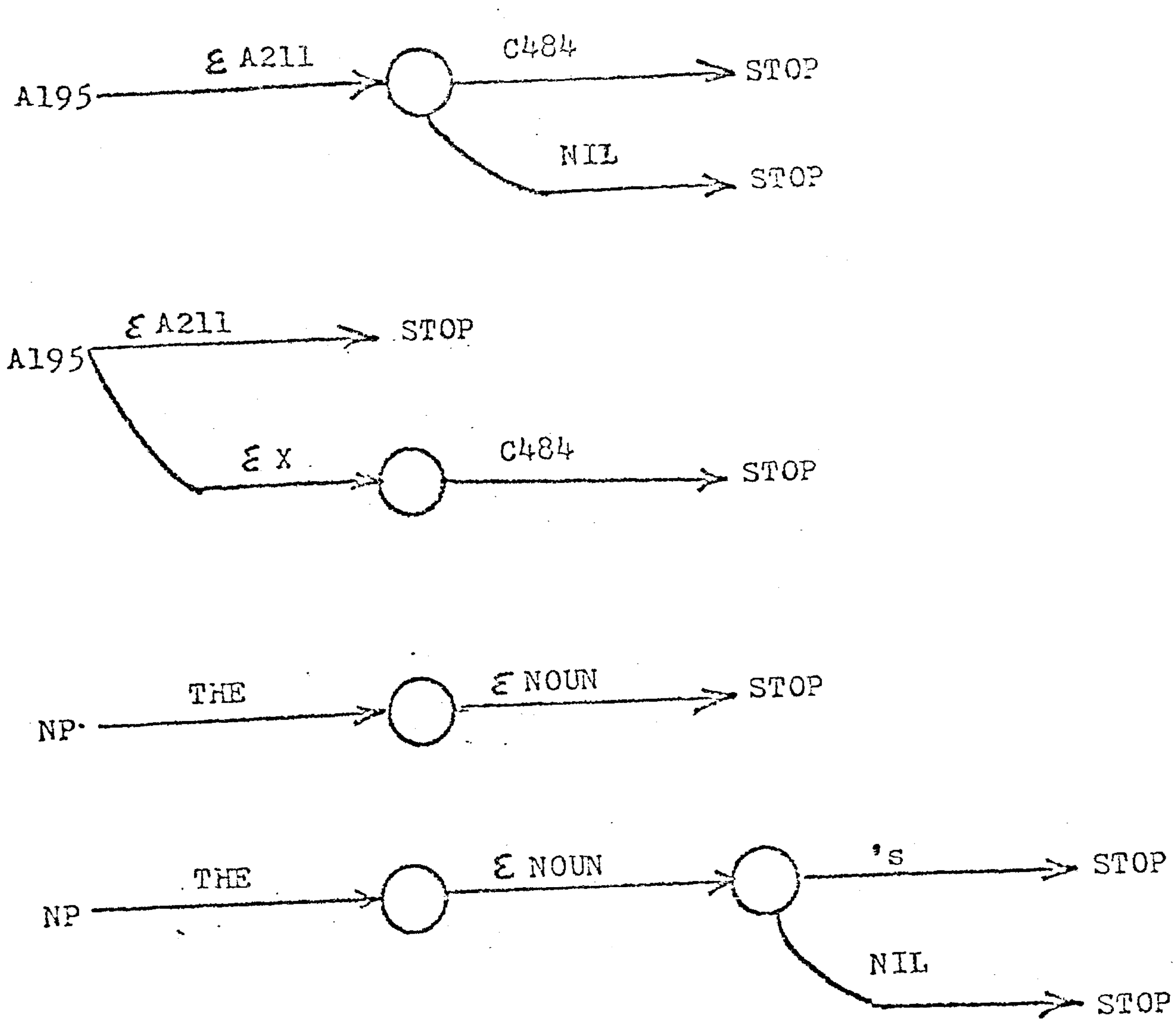
Figure 15

Some possible network grammars



Every bit as much as LAS, a child logically needs negative information to recover from overgeneralizations. The interesting question is where the negative information comes from in the case of the child. Parents do correct the child in such obvious morphemic overgeneralizations (Brown, 1973). Even today I find myself corrected (not by my parents) for my failures to properly pluralize esoteric words. The child may also use statistical evidence for a negative conclusion. In some manner he may notice that the morphemic form foots is never used by the adult and so conclude that it is wrong. Horning (1969) has formalized an algorithm for detecting such overgeneralizations by assigning probabilities to rules.

Figure 16 illustrates LAS's treatment of the last four sentences in the training sequences. These involve some three word noun phrases and also expansion of the noun phrases on the branch of the start network for RB relations. As can be seen from Figure 13, at the point of the 14th sentence LAS has expanded its grammar to the point where it will handle 616 sentences of the target language. Actually the grammar has produced some overgeneralizations--it will accept a total of 750 sentences. LAS has encountered phrases like square, square small, square red, and square red small. From this experience, LAS has generalized to the conclusion that the sentences of the language consist of a shape, followed optionally by either a size or color, followed optionally by a size. Thus the induced grammar includes phrases like squares small small because size words were found to be acceptable in both second and third positions. Interestingly, this mistake will not cause LAS any problems. It will never speak a phrase like square small small because it will never have a to-be-spoken HAM structure with two small's modifying an object. It will never hear such a phrase so and thus UNDERSTAND can not make any mistakes. This is a nice example of how an over-general grammar can be successfully constrained by considerations of semantic acceptability.

The problem of learning to sequence noun modifiers has turned out to be a source of unexpected difficulty. In part, the ordering of modifiers is governed by pragmatic factors, For instance one is likely to say small red square when referring to one of many red squares, but red small square when referring to one of many small squares. Differences like these could be controlled by ordering of links in the HAM memory structure.

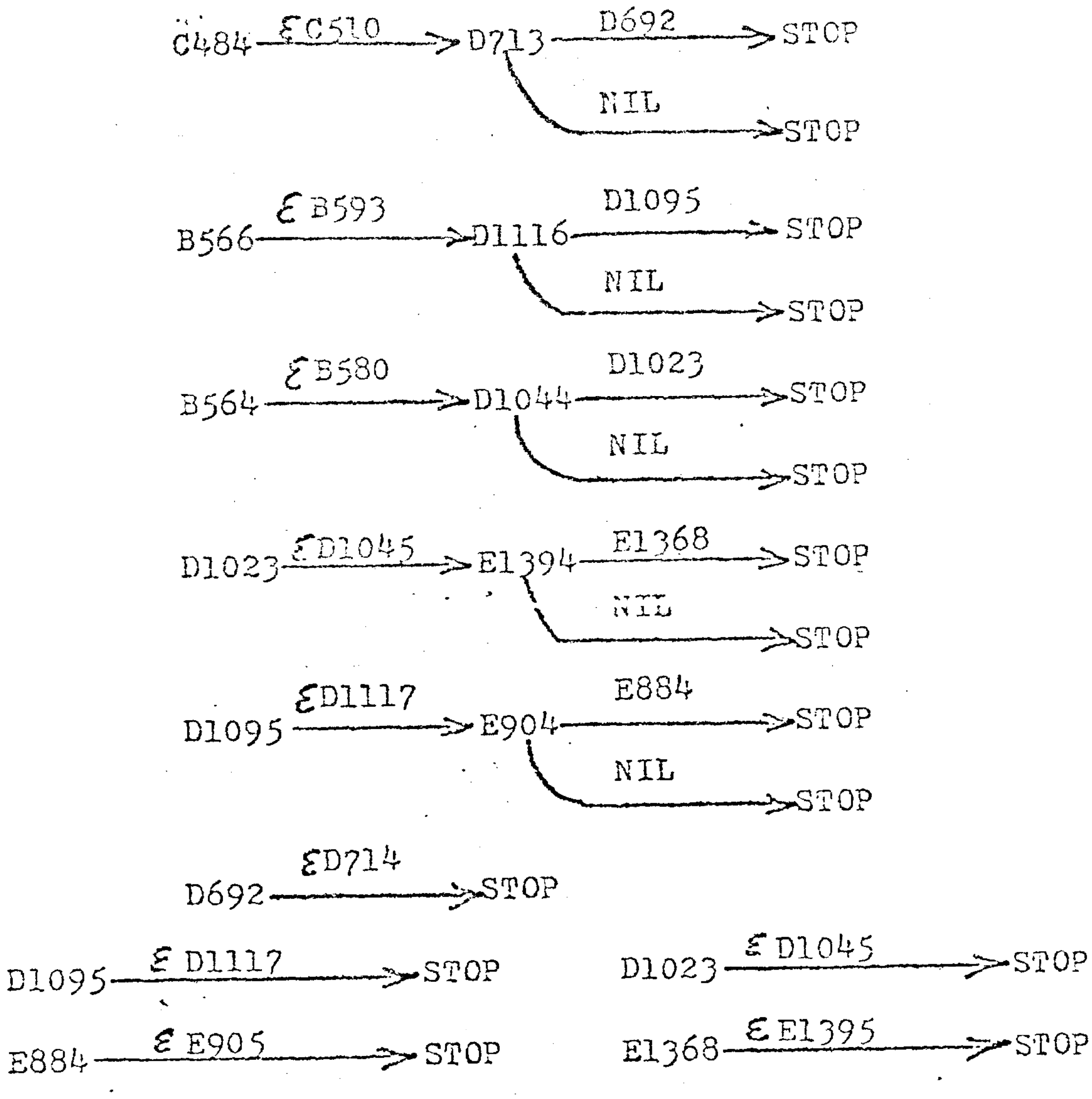
GENERALIZE.

After taking in 14 sentences LAS has built up a partial network grammar that serves to generate many more sentences than those it originally encountered. However, note that LAS has constructed four copies of a noun phrase grammar. One would like it to recognize that those grammars are the same. The failure to do so with respect to this simple artificial language only amounts to an inelegance. However, the identification of identical networks is critical to inducing languages with recursive rules.

Figure 16

Additions to LAS's grammar after studying:

10. SQUARE BLUE SMALL TRIANGLE RIGHT-OF
11. TRIANGLE RED SQUARE BLUE LEFT-OF
12. TRIANGLE SMALL SQUARE RED SMALL BELOW
13. SQUARE BLUE TRIANGLE BLUE LARGE LEFT-OF
14. SQUARE RED LARGE TRIANGLE RED LARGE BELOW



D714 = small
D1045 = red,blue,small
D1117 = blue,red
E905 = small,large
E1395 = large

A list is kept of all the networks created by SPEAKTEST. Once the structure of these networks becomes stable, GENERALIZE is called to determine which networks are identical. It compares pairs of networks looking for those which are identical. The criterion for identification of two networks is that they have the same arc paths. Two arcs are considered identical if they have the same syntactic conditions and semantic actions. Consider what LAS would do if it had the following embedding of networks:

$$
\begin{aligned}
NP &\rightarrow \text{the } NOUN_1 \\
&\rightarrow \text{the } ADJ_1\ NP \\
NP_1 &\rightarrow NOUN_2 \\
&\rightarrow ADJ_2\ NP_2 \\
NP_2 &\rightarrow NOUN_3 \\
&\rightarrow ADJ_3\ NP_3 \\
NP_3 &\rightarrow NOUN_4
\end{aligned}
$$

That is, there are four networks, NP, NP_1, NP_2, and NP_3 whose structure is indicated by the above rewrite rules. It is assumed that LAS has only experienced three consecutive adjectives and therefore SPEAKTEST has only created three embeddings. The critical inductive step for LAS is to recognize $NP_1 = NP_2$. This requires recognizing the identity of the word classes $NOUN_2$ and $NOUN_3$ and the word classes ADJ_2 and ADJ_3. This will be done on the criterion of the amount of overlap of words in the two classes. It also requires recognition that network $NP_2 = NP_3$. Thus, to identify two networks may require that two other networks be identified. The network NP_3 is only a subnetwork of NP_2. So in the recursive identification of networks, GENERALIZE will have to accept a subnetwork relation between one network like NP_2 which contains another like NP_3. The assumption is that with sufficient experience the embedded network would become filled out to be the same as the embedding network. After NP1 has been identified with NP2 HAM will have a new network structure where NP* represents the amalgamation of NP1, NP2, and NP3.

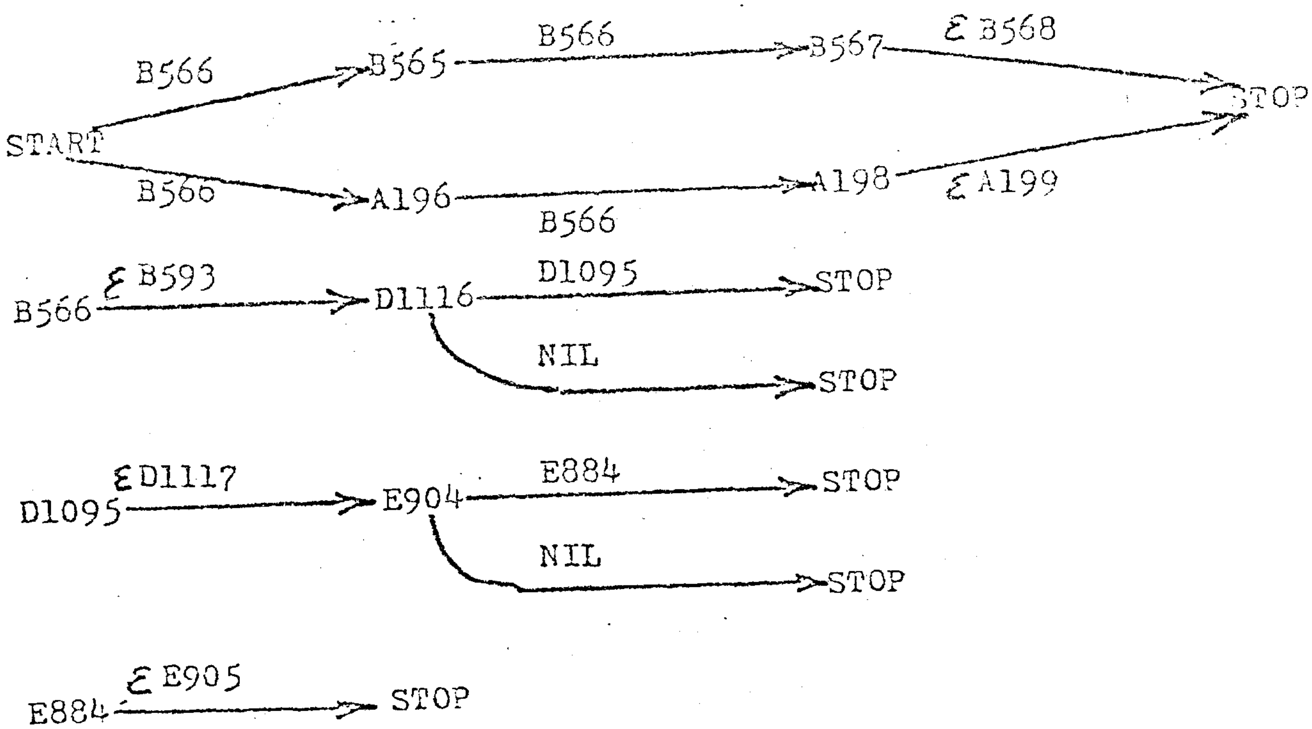
$$
\begin{aligned}
NP &\rightarrow \text{the } NOUN \\
&\quad\ \text{the } ADJ\ NP^* \\
NP^* &\rightarrow NOUN^* \\
&\quad\ ADJ^*\ NP^*
\end{aligned}
$$

Note that new word classes NOUN* and ADJ* have been created as the union of the word classes NOUN2, NOUN3, NOUN4 and of the classes ADJ2, ADJ3, respectively.

GENERALIZE was called to ruminate over the networks generated after the first fourteen sentences. GENERALIZE succeeded in identifying A195 with A197. As a consequence, network A195 replaced network A197 at the position where it occurred in the START network (see Figure 12). Similarly, B566 was identified with and replaced network B564. Finally, B566 was identified with and replaced A195 throughout the START network. The final effective grammar is illustrated in Figure 17. It now handles all the sentences of the grammar. It handles more sentences than the grammar that was constructed after the fourteenth

Figure 17

The final grammar



B568 = below,left-of
A199 = above,right-of
B593 = square,triangle
D1117 = blue,red,large,small
E905 = large,small

sentence. This is because the noun-phrase network B566 has been expanded to incorporate all possible noun phrases. Before the generalizations, none of the networks--B564, B566, A195, or A197 were complete. The network B566 became complete through merging with B564 and A195.

At this point, LAS now has a grammar adequate to speak and understand the target language. There are two major assumptions that LAS is making about the relation between sentence and referent which permit it success with these types of languages. The first is the assumption of the correspondence between the surface structure of the language and the semantic structure. This is critical to BRACKET's identification of the surface structure of the sentence which is, in turn, critical to the proper embedding of parsing networks. Second, there is the assumption of a semantics-induced equivalence of syntax. This played a critical role both in the generalization of SPEAKTEST and of GENERALIZE. It was noted with respect to pluralization that such generalizations can be in error and that children also tend to make such errors. However, I would want to argue that, on the whole, natural language is not perverse. Therefore, most of those generalizations will turn out to be good decisions. Clearly, for languages to be learnable there must be some set of generalizations which are usually safe. The only question is whether LAS has captured the safe generalizations.

The importance of semantics to child language learning has been suggested in various ways recently by many theoreticians (e.g., Bloom, 1970; Bowerman, 1973; Brown, 1973; Schlesinger, 1971; and Sinclair-de Zwart, 1973), but there has been little offered in the way of concrete algorithms to make explicit the contribution of semantics. LAS. 1 is a first small step to making this contribution explicit.

Conclusion

This concludes the explanation of the algorithms to be used by LAS.1 for language induction. In many ways the task faced by LAS. 1 is overly simplistic and its algorithms are probably too efficient and free from information-processing limitations. Therefore, the acquisition behavior of LAS. 1 does not mirror in most respects that of the child. Later versions of this program will attempt a more realistic simulation. Nonetheless, I think LAS.1 is a significant step forward. The following are the significant contributions embodied so far in LAS. 1.

1. The transition network formalism has been interfaced with a set of simple and psychologically realistic long term memory operations. In this way we have bridled the unlimited Turing-computable power of the augmented transition network.

2. A single grammatical formalism has been created for generation and understanding. Thus, LAS only needs to induce one set of grammatical rules.

3. Two important ways were identified in which a semantic referent helps grammar induction. These were stated as the graph deformation condition and the semantics-induced equivalence of syntax conditions.

4. Algorithms have been developed adequate to learn natural languages with a simple semantics.

B. Specific Aims

The general mode of developing the program LAS is as follows: A language learning situation is specified by a set of conditions. In LAS. 1 it was specified that LAS already know the meaning of the words and that it be given, as input, sentences with HAM representations of their meaning. The semantic domain was specified to be that constituted by geometric shapes. Once a set of conditions is specified, a set of goals is specified. In LAS. 1 there was only one real goal: to learn any natural-like language that described the domain. Once a set of goals is specified a plan of attack is sketched out. However, the problem is such that the details of that plan only evolve as we attempt to implement the plan as a computer program. Indeed many interesting problems and ideas that were not initially anticipated in LAS. 1 were discovered in attempting an implementation. This is part of utility of computer simulation in theoretical development.

The LAS. 1 program operated in a task domain which was similar, but by no means identical, to that of a natural language learning situation. Its behavior was similar to that of a human learning a language, but again by no means identical. In the next two years I propose to create a program LAS. 2 which comes considerably closer to simulating natural language learning. It has a more elaborate set of goals than did LAS. 1:

1. The program will incorporate realistic assumptions about short-term memory limitations and left-to-right sentence processing.

2. The program will learn the meanings of words.

3. The program should use semantic and contextual redundancy to partially replace explicitly provided HAM-encoding of pictures.

4. The program should handle sentences in a more complex semantic domain.

5. The program should be elaborated to handle such things as questions and commands as well as declarative sentences.

The general methods for achieving these goals in the LAS. 2 program will be sketched out in the proposal section. Also in that section I will propose some experiments to evaluate the LAS program. While it is true that the task faced by LAS. 1 is not really natural language learning, it still is a learning task at which human subjects apparently can succeed. The experiments will determine whether humans have the same difficulties in such tasks as does LAS and whether they make the same generalizations. However, I regard these experiments as of secondary importance relative to program development. It is more important to further articulate our understanding of what algorithms are adequate for natural language learning.

It is probably inevitable that the question will be asked as to whether it is really necessary to expend the resources necessary to construct a computer program. Could not the model just be specified conceptually? The reason why this is not possible has to do with the complexity of any theory that addresses the details of natural language. There is no other way to test the predictions of the theory or to assure that it is internally consistent. The experience with large transformational grammars hand-written for natural language is that they have hidden inconsistencies. These were only exposed by trying to simulate the grammars on a computer (e.g., Friedman, 1971). Consider the description given of LAS. 1 in the preceding section: Although lacking in many details, it was complex and lengthy. Could the reader establish for himself from this description whether the model is really internally consistent? A computer program provides a proof of the consistency and a means of determining the model's behavior. The stated goals of this project are to develop explicit algorithms for natural language learning, specify the relevant details of these algorithms, and evaluate empirically the psychological viability of these algorithms. Without the use of computer simulation none of these goals could be achieved.

C. Methods of Procedure

First I will describe the proposed extension of the LAS program. Then I will describe some experimental tests. In reading the specific extensions proposed for LAS, the reader should keep in mind that they have as their intent achieving the goals set forth in the preceding section.

The Semantic Domain

The first matter to settle upon in the new program is some semantic domain. The LAS. 1 world of shapes, properties, and geometric relations is too impoverished for further work. The following is proposed as a suggestion although there is nothing critical about its exact form. It is critical, however, that some semantic domain be chosen. It is only when there is a specified domain that an explicit goal for success in the program can be specified. The program will be regarded as successful if it can learn any natural language describing this domain.

I have chosen to look at a world close to that of a young child although there is perhaps nothing sacred about this domain. This world is set forth in Table 5. There are three people in this world. In addition to these there are four categories of objects--locations, containers, supporters, and toys. These objects can have four types of properties--number, color, size, and quality. Thus, LAS will have to deal seriously with problems of sequencing adjectives. It will also have to deal with number as a property of objects. The objects permit a much richer variety of relations than in the world of LAS. 1. This will provide a demanding test for the learning of complex multi-argument relations. There can be sentences like Mommy traded Daddy the car for a ball. In this world, people, containers, supporters, and toys can be in locations. People can change their location and that of toys. People and toys can be on supporters, toys can be in containers. People can possess toys, containers, and supporters.

TABLE 5

Categories in the World of LAS. 2

PEOPLE	LOCATIONS	CONTAINERS	SUPPORTERS	
Mommy	bedroom	box	table	
Daddy	kitchen	closet	chair	
LAS	den	dresser	bed	
TOYS	**NUMBERS**	**COLORS**	**SIZES**	**QUALITIES**
dolly	one	red	big	dirty
car	two	blue	medium	pretty
ball	three	green	small	shiny

Thus the different categories of objects enter differently into different types of relations. This fact will prove important to the predictive parsing facilities that I will want to introduce into LAS. 2.

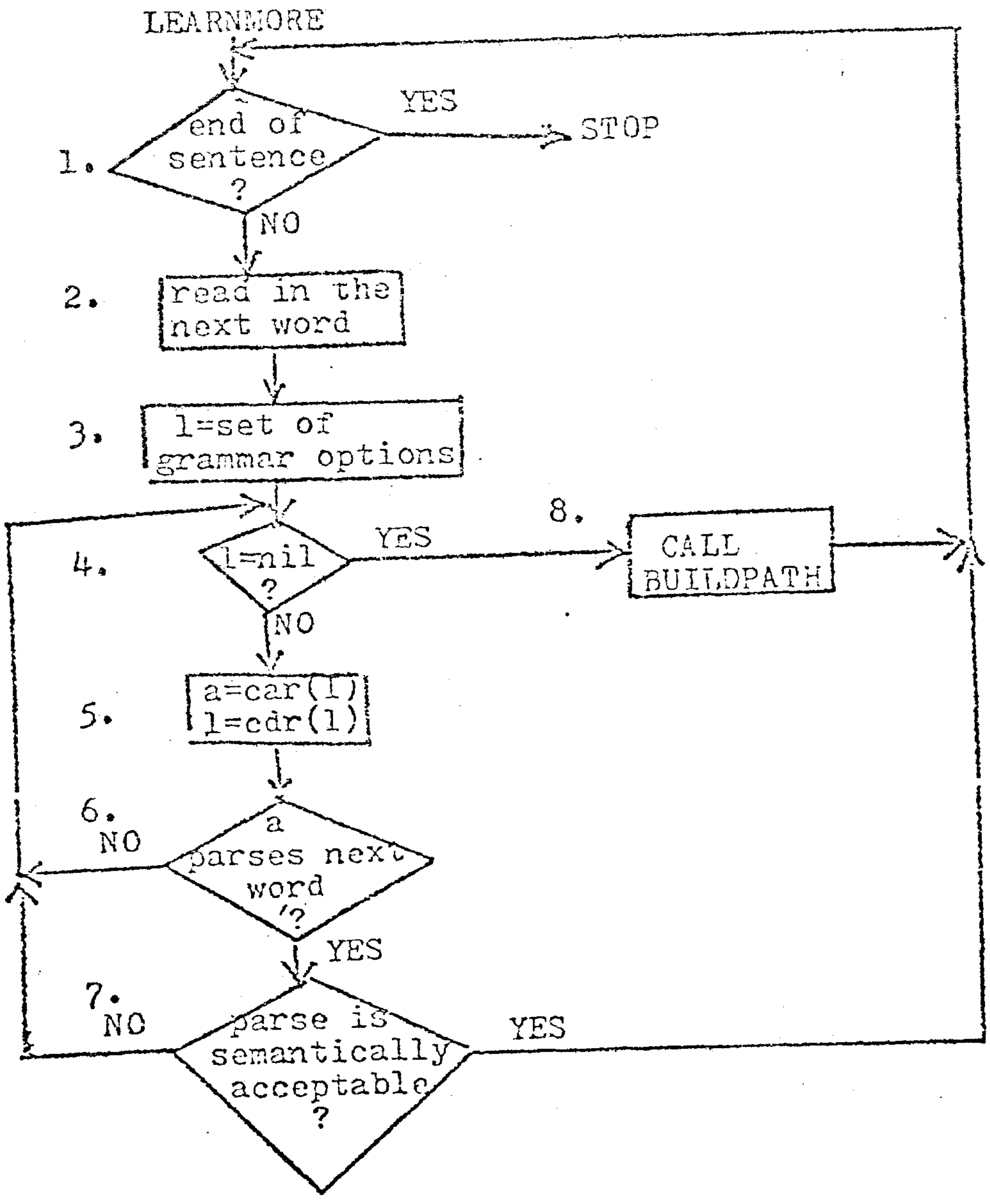
Left-to-Right Processing

Children learn language auditorily. Thus, their induction algorithms must process incoming material in a left-to-right manner. The current LEARNMORE program does not do this. BRACKET completely processes the sentence before SPEAKTEST even begins to work on it. Clearly, BRACKET and SPEAKTEST should be integrated so that the beginning of the sentence is bracketed and considered by SPEAKTEST before the end of the sentence is considered by either. Introducing this left-to-right processing is a preliminary to introducing short-term memory limitations into the induction situation.

Figure 18 illustrates in highly schematic form the left-to-right algorithm proposed for LEARNMORE. Words are considered as they come in from the sentence. LEARNMORE, as in UNDERSTAND, tries to find a path through its network grammar to parse the sentence. The difference between LEARNMORE and UNDERSTAND is that LEARNMORE has available to it a HAM conceptual structure to enable it to better evaluate various parsing options. Suppose LEARNMORE is at some point in processing the sentence. It will also be at some point in a parsing network. Let us consider how it would process the next word. At box 2 it would read in the word. At box 3 it would set l to the various grammatical options (arcs) at that node in the network. Boxes 4 through 7 are concerned with evaluating whether any of these options can handle the current word. Box 4 checks whether there are any options left. Box 5 sets a to the first option and resets l to the remaining options. Box 6 checks whether the word would be parsed by a and box 7 considers whether the action associated with that arc corresponds to a HAM structure. If a passes the tests in 6 and 7, LEARNMORE advances to considering the next word. Otherwise it tries another arc. If it exhausts all the arcs, it will call BUILDPATH (box 8) to build a new arc from the current node.

Figure 17

Flowchart for the proposed new LEARNMORE program



The work currently assigned to BRACKET will have to be assigned to box 7. That is, box 7 will have to determine when an arc should involve a push to an embedded network and when it should pop back up to an embedding network. This will be done by consulting the information in the semantic structure. It would also be possible to consult the pause structure of the sentence for information about phrase structure boundaries.

Note that certain sentences which the old LEARNMORE system could handle will not be handled by this system. For instance, consider the sentence The square that is above the triangle is right-of the square. After the first two words it would not be clear which square it was that we were referring to, the object or the subject of the right-of relation. Thus, buildpath could not assign an appropriate action to the path. In the old LEARNMORE this ambiguity about the referent of square was resolved by letting the whole sentence come in before dealing with it. Presumably, however, children would have difficulty learning from such sentences.

Lexicalization

In this system it will not be assumed that LAS knows the meaning of the words. Rather this will be something that LAS will have to learn from the pairing of sentences with conceptions. First let's discuss the learning of words whose reference is a simple concept or object, e.g., box or mommy, and postpone discussion of complex relational terms like trade. Logically, the task of lexicalization is quite simple and it would not require complex algorithms to succeed. For instance, consider this algorithm: LAS is given a sentence with n_1 words and a conceptualization it describes with m_1 concepts. Store with each word the m_1 concepts. The next sentence that comes has n_2 words and its conceptualization consists of m_2 concepts. If a word in this sentence is new, store with it the m_2 concepts. If the word is old, store with it the intersection of the concepts previously stored with it and the new m_2 concepts. Eventually, ignoring problems of polysemy, a word will become pared down to zero or one concepts. Those with zero concepts are function words and those with one concept have that concept as their meaning.

Of course, this algorithm will run into trouble if LAS does not always conceptualize all the concepts referred to by the sentenee. This can be remedied by having the algorithm wait for a sequence of disconfirming pieces of evidence before rejecting a hypothesized meaning. Incidentally, subjects behave just this way in concept attainment situations (see Bruner, Goodnow & Austin, 1965), not taking negative evidence as having its full logical force about the meaning of the word.

The basic problem with this algorithm is that it makes unreasonable assumptions about the information processing capacities of humans. In pilot research of my own, I have found that adult subjects can learn the meanings simultaneously of a number of words in a sentence. However, they do suffer difficulties when there is high ambiguity about what a word means. Presumably, children would have even greater difficulties extracting word meanings from complex sentences. Broen (1972) and Ferguson, Peizer, & Weeks (1973) report that new items of vocabulary seemed to be introduced through use in set sentence frames such as Where's ..., Here comes ..., There's ... known as deitic phrases. The noun tends to be heavily stressed and repeated. The parent frequently points to help reduce possible ambiguity of reference.

Presumably, later in lexicalization words can be learned by appearing in more complex sentence frames, provided the child knows most of the words and the grammatical structure of the sentence. To combine these various considerations, I propose the following addition to the flow chart in Figure 18 to deal with the reception of words with unknown meaning. In box 2, when an unknown word is read in, LEARNMORE will make a guess about its meaning using knowledge about context and about the word's position in the grammar. It will commit this guess to memory and stick with the guess unless later disconfirmed. The program will only hazard a guess in circumstances of low uncertainty. Thus, it will only guess if it can otherwise parse the grammatical structure in which the word appears. It will not guess if the word is preceded or followed by other words it does not know. Thus, the program, much as adults appear to, will learn on the basis of minimal contrasts between grammatical pattern and a current sentence. Thus, if the program knows the grammatical rule NP → determiner adjective noun and encounters the phrase the glick box it will suppose that glick refers to some property of the box.

Thus, the program will have to acquire its initial vocabulary by means of simple frames, as do young children. With this initial vocabulary information, it can begin to learn grammatical rules. Once in possession of grammatical rules, it will no longer need simple frames to learn new lexical items.

One interesting question is how function words are ever identified as non-meaning-bearing in this scheme. Presumably, this is done on the basis of failing to obtain a constant correlation between the word and any semantic feature. This could be detected by noting how many mistaken guesses had been associated with a word.

Concept Identification and Relation Words

So far I have assumed that all concepts are constructed before language acquisition takes place and that the only problem is to link up these concepts with words. But this is very unrealistic. Consider the verb give in the sentence Mommy gives the dolly to Daddy. The meaning of give is something like to do something which causes one to cease to possess an object and someone else to begin to possess the object. It seems very implausible that a child comes into a language learning situation with such a concept ready made. What probably happens is that he sees Mommy pushing the doll to Daddy or Mommy handing the ball to baby. With these experiences he hears sentences like Mommy gives the dolly to Daddy or Mommy gives the ball to baby. From these examples he induces the appropriate meaning of give. Concept attainment in these situations can be achieved by using the sort of concept identification used by Winston (1970) for inducing geometric concepts. That is, each use of the word give is paired with a HAM network structure given the meaning of the sentence. Winston's heuristics allow us to extract what these network structures paired with give have in common. The concept give, as verb, is then attached to the common structure. For this sort of algorithm to succeed, LAS must be set to regard certain configurations of propositions, interlinked by causal terms, as being associated with a single relational term in the language.

Note also that the effect of such an induction scheme would be to encode the meaning of complex relational terms into the network grammar. That is, in parsing the sentence Mommy gives the dolly to LAS, the network would specify that UNDERSTAND set up, as the meaning of the sentence, a HAM representation of the form Mommy does something which causes her to cease to possess the doll and LAS to begin to possess the doll.

BADEAR--The Telegraphic Perception Hypothesis

The clearest discrepancy between the behavior of LAS. 1 and a child is that LAS. 1 generates no ungrammatical sentences. In contrast, at first the child only generates ungrammatical sentences. The child's early speech has been crudely characterized as telegraphic. That is, children speak in two and three word utterances. To condense messages into such short utterances it appears that children have omitted most function words and subordinate constructions. One explanation of the origin of telegraphic speech which is appealing from the point of view of LAS is the following: Suppose that LAS did not receive as input to its induction routine complete sentences, but rather telegraphic sentences. Then, it would quite naturally induce a telegraphic grammar. It seems reasonable to suppose that a child cannot hold in immediate memory the total sentence he has heard but rather only a depleted version of that sentence. If so, then his induction algorithms would be receiving telegraphic sentences as their basic data. Let's call this the telegraphic perception hypothesis.

Evidence for this hypothesis comes from studies of child imitation of adult speech. It is found that these imitations, while longer than the child's own productions are also telegraphic in nature (e.g., Brown & Fraser, 1964). Blasdell and Jensen (1970) found that children tend to repeat those words which are stressed and those words which occur in terminal positions. The semantically important words tend to be stressed in adult speech. Scholes (1969, 1970) found that children tended to omit words that had unclear semantic roles or unknown meanings. What I find striking is that these are just the variables which control what I can repeat back after hearing a French sentence--a language I know quite imperfectly. Of course, the variables of serial position, perceptual isolation, and meaningfulness all have well established effects in verbal learning experiments on immediate memory.

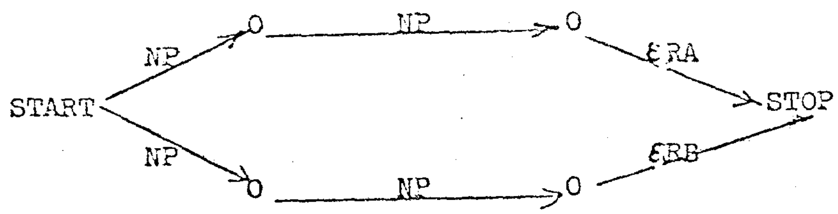
I propose to introduce telepraphic perception into LAS through an aspect of LEARNMORE called BADEAR. The BADEAR program will simulate the variables of stress, meaningfulness, and serial position in providing LAS with a depleted version of the sentence. The locus of the effect of BADEAR will be between boxes 4 and 8 in the flowchart of Figure 2. Basically it will not pass all words onto BUILDPATH. Rather some words will "slip from consciousness" after failing to be parsed. It will tend to omit words when: (a) they are unstressed, (b) their meaning is not known, (c) a critical number of new words in the sentence have already been passed to BUILDPATH. I suspect this critical number is something like one or two.

Factors (a) and (b) would generate the effects of stress and meaningfulness. Factor (c) would yield good memory for the first words of the sentence. What good memory children do show for last words in phrases probably reflects short-term acoustic memory.

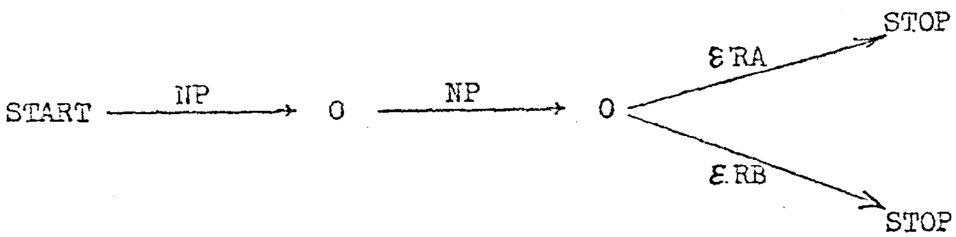
An interesting feature of BADEAR is that, as the grammar is expanded, LAS would be able to receive more of the sentence. Thus, its productions and imitations would grow as does a child's. This would be providing an explicit mechanism for an idea suggested by Braine (1971), Olson (1973, and others. Inducing a grammar from degenerate sentences presents an interesting problem. How is it that LAS ever comes to abandon its rules for generating telegraphic speech? Merely because LAS has learned rules for generating fuller sentences, it does not follow that the old rules are wrong. After all, language permits multiple means for expressing the same thoughts. Perhaps then mechanisms should be incorporated that will strengthen some grammatical rules relative to others. Rules to be strengthened would be those that could be successfully used by UNDERSTAND and that could successfully be used by SPEAK. We might think that the arcs out of a node in a parsing network are ordered on a stack to reflect their relative utilities. Subjects would try rules on the top of a stack first. Ineffective rules like the original ones for two and three word utterances would descend to the bottom of the stack and so become unavailable. This strength mechanism is the same as used to order links in the HAM memory model. This is a different way to bring negative information to bear in grammar induction than that proposed for RECOVER. That is, rather than seeking explicit disconfirmation of rules, the rules are gradually weakened out of existence as more adequate rules take over the roles the old rules used to occupy in sentence understanding and generation.

Grammar Optimization

Note that the START network induced by LAS was one with the following form:



This grammar requires considerable backup if the sentence does not have an RA relation. As suggested earlier it would be more efficient if LAS were given the power to transform the grammar into the following form:



Given that there are serious time problems (see introduction of proposal) in parsing, it is critical that methods be incorporated in the learning program for optimizing the grammar. The merging of arcs, besides making the grammar more efficient, would be another form of generalization. It could be used to further merge and build up word classes.

Further Use of Semantics in Language Acquisition

There are at least two further ways that semantics can be used to aid language acquisition in addition to those embodied in LAS. 1. One concerns using conceptual information as a further aid to word class formation. Words in a particular class tend to have a common semantic core. LAS could use this fact to adjust its threshold for merging words into a class. For instance, suppose LAS considers merging two word classes because they share certain syntactic properties, both of which contain color names. Currently, LAS. 1 makes this decision on the basis of the amount of overlap between the members of the two classes. LAS should lower its overlap threshold because these word classes do share a strong semantic property.

Another use of semantics would be to lessen LAS's reliance on explicitly given semantic interpretations of sentences. It should sometimes be able to guess these interpretations. For instance, suppose a sentence came in with the words ball, box and in. Because of the conceptual constraints between these, LAS should be able to guess their connection. This use of conceptual constraints in the semantic domain could also be used by UNDERSTAND to permit predictive parsing along the model of the Schank's (1972) system. That is, as an alternate to understanding a sentence by use of syntactic information, it is possible to look for conceptual constraints to predict what the interpretation of the sentence should be. This prediction can then be checked for syntactic correctness by use of the network grammar. It would be profitable to try to place a predictive parsing system like Schank's within the rigors of the Woods' network formalisms.

A Procedural Semantics

So far LAS has been principally concerned with representing the meaning conveyed by a declarative sentence. However, language has other purposes than just to communicate meanings from one speaker to another. Consider commands and questions. For instance, consider the sentence Put the dolly in the box. Currently, UNDERSTAND might retrieve the sentence's meaning as Speaker requests of LAS that it put the dolly in the box. This is the declarative meaning of the sentence. However, in addition LAS should evoke an action that causes it to comply or at least take an action to decide whether to comply. This is the procedural meaning of the sentence. The procedural meaning of declarative sentences is very simple: store this sentence. This is already assumed in LAS's treatment of the sentence. However, the procedural meanings underlying other types of sentences are more complex. A large part of the success of Winograd's system is that it was adequately able to deal with the procedural aspects of various sentences' semantics. It is important that LAS begin to deal with these too.

What this would mean, in terms of LAS's network grammars, is enriching the set of actions that can be stored. Currently, the only actions are ones that result in the creation of pieces of HAM structure, i.e., declarative knowledge. LAS will have to store other internal actions that specify what it does with the declarative knowledge. These will include commands to answer the question or obey the order. HAM already has commands that direct it to answer questions but executing orders would be something new. As part of the HAM project, I am working on methods for incorporating procedural knowledge into a network system. It is unclear yet what success I will have here.

It is interesting to note other aspects of natural language whose semantics are procedural. These are well documented in Winograd. Consider for instance the difference between the definite and indefinite article--the red ball versus a red ball. The former indicates an object which the listener knows. Thus the listener's response to the definite article should be to search his memory for the referent of the noun phrase. In contrast, the listener's response to an indefinite noun phrase should be to construct a new representation for it. This difference can be nicely handled in the current HAM system by whether a call to the MATCH routine is evoked.

Winograd has argued convincingly that the semantics of pronouns and other indexicals should be represented by procedures to determine their referents. This is particularly true for terms like you whose meaning is totally relative to speaker and context. Since the referent of you completely changes with speaker, a child would be lost if he tried to associate its meaning with some HAM memory node. He must be prepared to treat it as having as meaning a procedure for determining the referent.

Provided that LAS has the facilities for representing and evaluating procedures, there seem no difficulties in learning those aspects of language which are heavily embued with procedural semantics. Language learning will continue to arise from pairing sentences with semantic interpretations. However, semantic interpretations will now contain a procedural as well as a declarative aspect. Again language learning will consist of learning mappings between sentences and the now-enriched semantic representations.

Experimentation

As stated before, I do not think that experimental research should yet be the principal focus of the project. There is still much further research that needs to be done in the way of specifying algorithms that are capable of language induction. Nonetheless, in parallel with this research, I would like to perform experiments to get some initial assessments of the viability of the proposed algorithms. The type of information relevant to evaluating LAS is only acquired by looking at artifical languages. With these artificial languages it is possible to test LAS's predictions about language learnability and generalization.

Criticisms of Experiments with Artificial Languages

For ethical reasons it is not possible to expose young children, just learning their first language, to an artificial language which LAS had identified as degenerate and probably not learnable. This means that all experimentation with artificial languages must be done on older children already well-established in their first language or on adults. Consequently, the first language may be mediating acquisition of the second language. There is evidence (see Lennenberg, 1967) that there is a critical initial period during which languages can be learned much more successfully than in later years. Lennenberg speculates that there is a physiological basis for this critical period. Thus, one might wonder whether the same processes are being studied with older subjects as in the young child. Personally, I also doubt that the mechanisms of language-acquisition are the entirely same with the young child in first language learning as with the older subject in second language learning. However, it does seem probable that there should be considerable overlap in the mechanisms for the two situations. The reason for this belief has already been stated: Both for the adult and the young child, language acquisition presents largely the same set of severe and unique information-processing demands. The algorithms that deal with induction problems therefore are probably not very different in any system that successfully learns the language.

Other criticisms (e.g., those of Slobin, 1971; Miller, 1967) of studies of artificial language learning focus on the fact that these languages are artificial. Natural language is much more complicated than an artificial laboratory language; it takes years to acquire; it serves more complex functions; the child's motivations are more complex than the laboratory subject. However, these criticisms miss the whole point of laboratory experimentation which is to isolate and study significant aspects of a complex natural phenomena. Another criticism of the past artificial languages studies (e.g., those studies of Braine, 1963b; Miller, 1967; Reber, 1969) is that they lack a semantic referent. Clearly, this makes an enormous difference to the sort of algorithms a subject can employ. The critical heuristics used by LAS would be useless without semantics. Moesser and Bregman (1972, 1973) have shown that the existence of a semantic referent has a huge effect on language acquisition. Except for control conditions, all of my experiments will involve a semantic referent.

LAS's Predictions about Language Learnability

Critical to LAS's induction algorithm is that the graph deformation condition be met concerning the relation between the surface structure of the sentence and the HAM conceptual structure. That is, the surface structure must preserve the original connectivity of concepts. In Section A5 we described languages which violated this assumption. Consider the following language:

S → NP NP relation
NP → noun (Color) (adjective) (clause)
CLAUSE → te NP relation
NOUN → square, circle, triangle, diamond
Color → red, blue
Size → small, large
Relation → above, below, right-of, left-of

This is an expanded version of GRAMMAR1 described in Table 1. (The element te serves the function of a relative pronoun like that.) An example of a sentence in this language is Square red te triangle big above circle blue small right-of. An experiment I will do compares four conditions of learning for this language.

1. No reference. Here subjects simply study strings of the language trying to infer their grammatical structure.

2. Bad semantics. Here a picture of the sentence's referent will be presented along with the sentences. However, the relationship between the sentence's semantic referent and the surface structure will violate LAS's constraints. The adjective associated with the ith noun phrase will modify the (n + 1 – i)th shape in the sentence (where n is the number of noun phrases). For example, the adjectives associated with the first noun phrase will modify the last

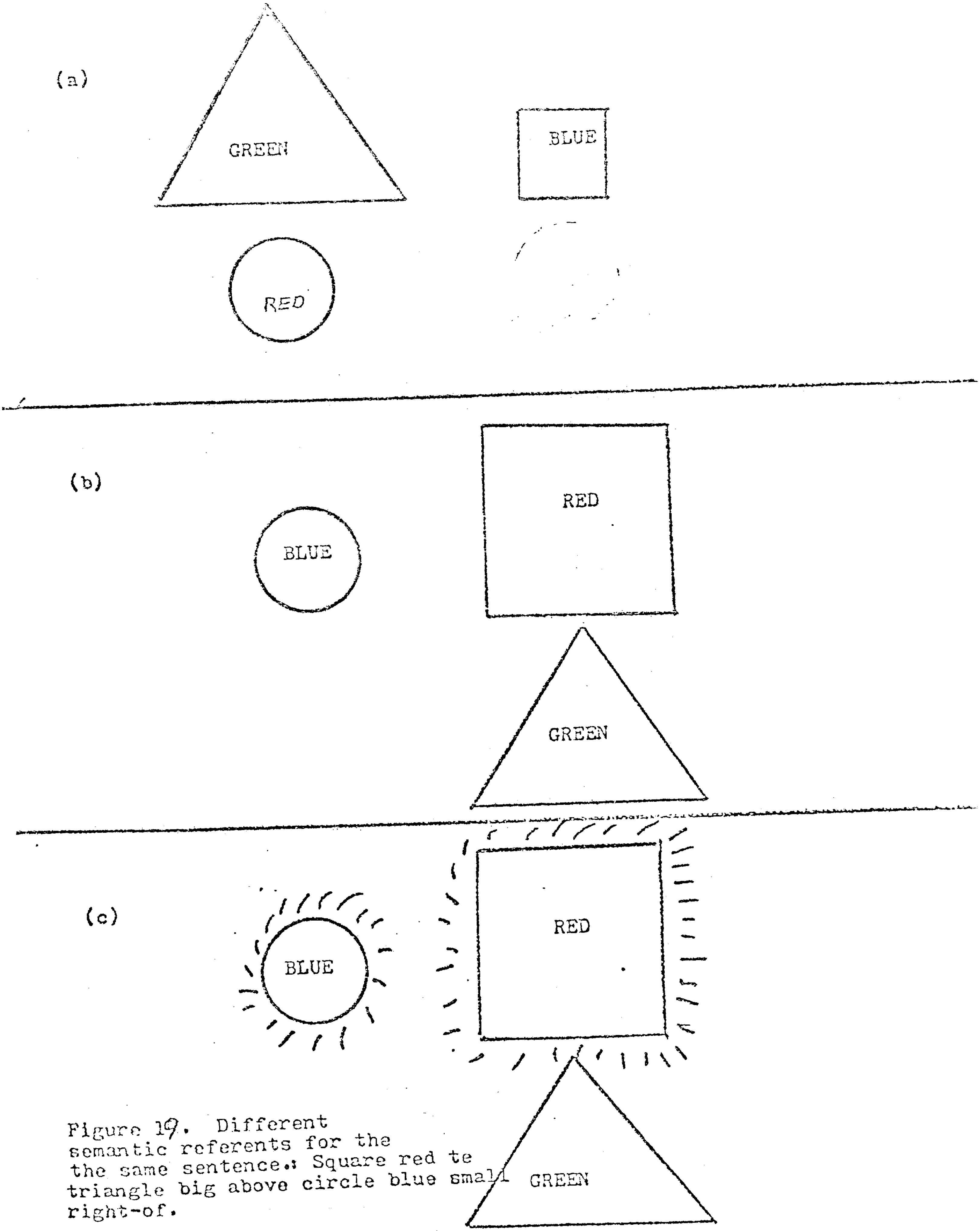


Figure 19. Different semantic referents for the the same sentence.: Square red te triangle big above circle blue small right-of.

shape. Similarly, the ith relation will describe the relation between the $(m + 1 - i)$th related pair of shapes (where m is the number of relations). So for instance the second relation right-of will describe the relationship between the first pair of shapes square and triangle. The appropriate picture for the example sentence is given in Fiuure 19a.

3. Good semantics. Here the adjective in each noun phrase will modify the noun in that phrase.. Relations will relate the appropriate nouns in the surface structure. The appropriate picture for the example sentence in this case is given in Figure 12b. LAS could bracket sentences given this picture if it could guess the main proposition.

4. Good semantics plus main proposition. The picture in this condition will be the same as in 3 but the two shapes in the main proposition will be highlighted. In this condition LAS would be guaranteed of successfully bracketing the sentence because the main proposition is given.

In some ways this experiment is like Moesser and Bregman's. However, here English words are used so that the subjects do not need to induce the language's lexicalization as well as its grammar. This corresponds to the situation faced by LAS. 1. If English words were replaced by nonsense syllables this would require a simplification of the language to make induction tractible. The predictions of LAS are, of course, that best learning occurs in Condition 4, next best in 3, and failure of any learning in 1 and 2. It would not be surprising to see subjects perform better in 1 than in 2 since in they might partially be able to imagine an appropriate semantics.

The procedure would have subjects in all conditions study the same sequence of sentences but vary the accompanying semantic information according to condition. After a study phase they would be tested for grammaticality judgments about a set of sentences, some of which violate one of the rules for generation. Since the syntax of the language is the same in all four conditions, the same sentences will be grammatical in all four conditions. Even though the syntactic information given during study will be the same in all conditions, marked differences in syntactic knowledge should appear across conditions. The current plan is to alternate sequences of study trials with sequences of test trials, so the subject might study six sentences, with the semantic information (appropriate to his condition, if any). Then he would see six test pairs, one sentence of each pair violating some syntactic rule. For each pair of he would have to choose the grammatically correct pair. By frequently alternating study and test, it would be possible to carefully monitor the growth of information in the conditions.

Many readers may not be surprised by the prediction of better learning in Conditions 3 and 4. Hopefully, the significance of such an outcome would be clear. It would show that semantics is important to induction of the syntactic structure of a natural language. However, it would also show that semantics is useless if the relation between the semantic referent and the syntactic structure is arbitrary. The surface structure of the sentence must be a graph-deformation of the underlying semantic structure. Failures to appreciate the contribution of semantics to language induction and failure to understand the nature of this contribution of semantics to the induction process have been fundamental in the stagnation of attempts to understand the algorithms permitting language induction. These facts may be obvious when pointed out but they have been unavailable to the linguistic theorists for fifteen years.

There are other experiments of this variety which can be done to see how well humans can learn languages which do or do not meet the constraints demanded by LAS's induction algorithms. These constraints have the same purpose as Chomsky's (1965) proposals for linguistic universals. That is, they constrain the set of possible hypotheses about language structure so that the target language can be identified. However, the constraints used by LAS are not the same as those suggested by Chomsky. For instance, Chomsky proposed that transformations which reversed the order of words in a sentence would be unacceptable. This is because such a rule does not refer to the sentence's constituent structure. However, a language which contained sentences of a natural language and their reversals would be learnable by LAS. It would just develop one set of rules for sentences in one order and another independent set for reverse order sentences. It would be interesting to see whether human subjects could learn such a language.

In the example of the induction of GRAMMAR1 we found that there was no way for LAS to detect non-semantic contingencies between syntactic choices in the first noun-phrase and in the second noun-phrase pushed to in the main network. For instance, it is possible that a morphemic embellishment of the adjectives in the second noun phrase may depend on a choice of morphemic embellishment for the noun in the first noun phrase. Human subjects should also find it hard to detect such syntactic contingencies.

Predictions about Generalization

There are another set of predictions, besides those concerned with language learnability, which it will be useful to explore. LAS makes predictions about the situations under which humans will tend to generalize rules and when humans will not. Suppose LAS learned the following grammar:

S → VERB NP NP
NP → (PREPP) N_1 (ADJ)
PREPP → PREP N_2
N_1 → boy, girl, etc.
N_2 → room, bank, etc.
ADJ → tall, nice, etc.
PREP → in, near, etc.
VERB → like, hit, etc.

A typical sentence in this language would be Like in room boy tall girl nice which means The tall boy in the room likes the nice girl. This language is given English terms only to make its semantics clearer. Suppose, in fact, words in the language were das meaning man, jir meaning woman, fos meaning boy, and tuk meaning girl. Suppose the subject studies the following pair of sentences:

1. Like das tuk.
2. Like fos jir.

Then, it is interesting to consider his judgments of the acceptability of sentences like:

3. Like das tuk.
4. Like das jir.
5. Like jir das.

Accepting (3) only involves recalling sentence (1), but accepting (4) would involve a generalization: LAS would currently make this generalization because it would merge das and fos into a single word class and it would similarly merge tuk and jir. If the subject accepted (5) he would be making a more interesting generalization not currently predicted by LAS. He has never encountered jir in the first noun slot or das in the second noun slot. Nonetheless, he assumes they are acceptable in these positions on the basis of their semantic similarity to words which are in these classes.

Neither (4) nor (5) need be acceptable sentences. The words jir and das could, for instance, take a different case inflection when they appear in different slots. This would make (5) unacceptable. Sentence (4) could be unacceptable because jir took a different morphemic embellishment when preceded by das. It would be interesting to see how learnable a language would be that contained such violations of the potential generalizations.

One can explore other questions about generalization in this artificial language. Suppose a subject studied sentences like (6). Would he accept sentences like (7)?

6. Like in room boy tall girl
7. Like girl in room boy tall

That is, will rules generalize from the subject noun phrase to the object noun phrase. As LAS is currently constituted such generalizations would not occur until it had built up fairly stable noun phrases. Again suppose LAS had initially only encountered simple sentences such as (8):

8. Like boy man

From sentences such as (8) LAS would learn the class of nouns that occurred in first and second noun phrase slots. Suppose then sentence (9) was studied. On the basis of it, would sentence (10) be accepted as grammatical? That is, would the prepositional phrase in bank generalize to other nouns in the same class as woman?

9. Like boy in bank woman
10. Like girl in bank man

This would be an example of right generalization which does not occur in LAS. In contrast, LAS does perform left generalization. That is, after studying (11) LAS would accept (12).

11. Like boy woman nice
12. Like boy man nice

It will be interesting to see if humans show any preference for left generalization over right generalization.

It is critical that these artificial language experiments be done with a number of age groups, from young children (e.g., ages 4 and 5) to adults. While one can never really study first language acquisition with artificial languages, it is important to get an appreciation of what the developmental trends are. Since young children cannot handle written languages, much of this language training will have to be done with auditory presentation of the to-be-learned language. There has been little work done on artificial language learning by such young children, so probably much pilot research will be necessary to establish workable procedures.

D. Significance

LAS is a program with two purposes, one concerned with psychology and one concerned with artificial intelligence. I think this mixed purpose is fruitful because it promotes a cross-fertilization of ideas from two fields and so helps prevent theoretical stagnation. There is no guarantee that LAS, in the broad outline currently conceived, will ever achieve the goal of an adequate simulation of a child's acquisition of language. However, a certain outcome of this will be a clearer understanding of the information-processing demands of language-acquisition and of the role of a semantic referent in grammar induction. If LAS fails we will learn what is wrong with one explicit set of induction algorithms. Even that would be a significant contribution to the current theoretical development in a field rich in data but almost totally lacking explicit information-processing theories. I hope, of course, that the processes uncovered in the LAS project will be the same as those used by humans in language learning. A successful simulation program would constitute an enormous advance in our understanding of cognitive development.

The contributions of LAS to the artificial intelligence field are less certain and more distant. Nonetheless, generality in language understanding systems is an important goal and one for which a learning system approach seems ideal. It is therefore important to understand the contribution language learning systems can make in this field. It would be a significant advance to know in detail why a learning system approach was not the answer to language understanding or at least why LAS was not the right sort of learning system. Of course, if LAS does prove to be the basis for a viable language understanding system, its contribution to artificial intelligence will also be of considerable importance.

E. Facilities Available

I shall have available the entire facilities of the Human Performance Center, University of Michigan. My current appointment expires June 30, 1976, but can be extended for one to three years. My principal resource will be the Michigan Terminal System which supports a rich variety of programs. Most of the programming will be performed in Michigan LISP (see Hafner & Wilcox, 1974) which is a relatively economical and an error-free version of LISP.

REFERENCES

ALPAC (Automated Language Processing Advisory Committee). Language and machines - computers in translation and linguistics. National Academy of Sciences, Washington, D. C., 1966.

Anderson, J. R. Computer simulation of a language-acquisition system. In R. L Solso (Ed.) Information Processing and Cognition: The Loyola Symposium. Washington: Lawrence Erlbaum, 1975.

Anderson, J. R. and Bower, G. H. Human associative memory. Washington: Winston and Sons, 1973.

Bar-Hillel, Y. Language and information. Reading, Mass: Addison-Wesley, 1964.

Bever, T. G.. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.) Cognition and the development of language. New York: Wiley, 1970.

Bierman, A. W. An interactive finite-state language learner. First USA-JAPAN Computer Conference, 1972.

Bierman, A. W. and Feldman, J. A. A survey of results in grammatical inference. In (Ed.), Frontiers of pattern recognition. New York: Academic Press, 1972.

Blasdell, R. and Jensen, P. Stress and word position as determinants of induction in first-language learners. Journal of Speech and Hearing Research, 1970, 13, 193-202.

Bloom, L. One word at a time. The Hague: Mouton, 1973.

Bobrow, D. G. A question-answering system for high school algebra word problems. AFIPS Conference Proceedings, 1964, 26, 577-589.

Bowerman, M. Early syntactic development. Cambridge, England: University Press, 1973.

Boyer, R. S. Locking: A restriction of resolution. Ph.D. Dissertation. University of Texas at Austin, 1971.

Braine, M. D. S. On learning grammatical order of words. Psychological Review, 1963, 70, 323-348.

Braine, M. D. S. On two types of models of the internalization of grammars. In D. J. Sloben (Ed.) The ontogenesis of grammar. New York: Academic Press, 1971, 153-188.

Broen, P. The verbal environment of the language-learning child. Monographs of the American Speech & Hearing Association, 1972, 17.

Brown, R. A first language. Cambridge, Mass.: Harvard University Press, 1973.

Brown, R. and Fraser, C. The acquisition of syntax. In C. N. Cofer and B. S. Musgrave (Eds.), Verbal behavior and learning: Problems and processes. New York: McGraw-Hill, 1963, 158-197.

Bruner, J. S., Goodnow, J., and Austin, G. A. A study of thinking. New York: Wiley, 1956.

Charniak, E. Computer solution of word problems. Proceedings of the International Joint Conference on Artificial Intelligence. Washington, D. C.: 1969, 303-316.

Chomsky, N. Aspects of the theory of syntax. Cambridge, Mass.: MIT Press, 1965.

Colby, K. M. and Enea, H. Inductive inference by intelligent machines. Scientia, 103, 669-720 (Jan. - Feb. 1968).

Clark, E. V. Non-linguistic strategies and the acquisition of word meanings. Cognition: International Journal of Cognitive Psychology, 1974, in press.

Coles, L. S. Talking with a robot in English. Proceedings of the International Joint Conference on Artificial Intelligence. Washington, D. C. 1969, 587-596.

Crespi-Reghizzi, S. The mechanical acquisition of precedence grammars. Report No. UCLA-ENG-7054, School of Engineering and Applied Science, University of California at Los Angeles, 1970.

Dreyfus, H. L. What computers can't do. New York: Harper and Row, 1972.

Ervin, S. M. Imitation and structural change in children's language. In E. H. Lennenberg (Ed.), New directions in the study of language. Cambridge, Mass.: MIT Press, 1964, 163-189.

Feldman, J. A. Some decidability results on grammatical inference and complexity. A. I. memo No 93.1, Computer Science Department, Stanford University, 1970.

Fergusen, C. A., Perzer, D. B., & Weeks, T. E. Model-and-replica phonological grammar of a child's first words. Lingua, 1973, 31, 35-65.

Fernald, C. Children's active and passive knowledge of syntax. Paper presented to the Midwestern Psychological Association, 1970.

Fikes, R. E., Hart, P. E. and Nilsson, N. J. Some new directions in robot problem solving. Stanford Research Institute, August, 1972.

Fillmore, C. J. The case for case. In E. Bach and R. J. Harms (Eds.), Universals in linguistic theory. New York: Holt, Rinehart and Winston, 1968.

Friedman, J. A computer model of transformational grammar. American Elsevier, 1971.

Fraser, D., Bellugi, U., & Brown, R. Control of grammar in imitation, comprehension, and production. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 121-135.

Gold, M. Language identification in the limit. Information and Control, 1967, 10, 447-474.

Green, C. D. and Raphael, B. Research on intelligent question answering systems. Proceedings of the ACM 23rd National Conference. Princeton: Brandon Systems Press, 1968, 169-181.

Hafner, C. and Wilcox, B. LISP MTS programmer's manual. Mental Health Research Communication 302 and Information Processing Paper 21. University of Michigan, 1974.

Horning, J. J. A study of grammatical inference. Technical Report No. CS 139, Computer Science Department, Stanford University, August, 1969.

Kasher, A. Data retrieval by computer: A critical survey. Hebrew University of Jerusalem, January 1966. Technical Report No 22 to Office of Naval Research, Information Systems Branch.

Kelley, K. L. Early syntactic acquisition. P-3719, the Rand Corp., Santa Monica, California, 1967.

Kelloggs, C. H. A natural language compiler for on-line data management. Proceedings of the 1968 Fall Joint Computer Conference, 473-492.

Kuno, S. The predictive analyzer and a path elimination technique. Communications of the ACM, 1965, 7, 453-462.

Lenneberg, E. H. Understanding language without ability to speak: A case report. Journal of Abnormal and Social Psychology, 1962, 65, 419-425.

Lenneberg, E. H. Biological foundations of language. New York: Wiley, 1967.

Lindsay, R. K. Inferential memory as a basis of machines which understand natural language. In E. A. Feigenbaum and J. Feldman (Eds.), Computers and thought. New York: McGraw-Hill, 1963.

Loveland, D. W. A linear format for resolution. Proceedings of the IRIA Symposium on Automatic Demonstration. New York: Springer-Varlay, 1970, 147-162.

Luckham, D. Refinemants in resolution theory. Proceedings of the IRIA Symposium on Automatic Demonstration. New York: Springer-Varlay, 1970, 163-190.

Miller, G. A. The psychology of communication. New York: Basic Books, 1967.

Minsky, M. (Ed.), Semantic information processing. Cambridge, Mass.: MIT Press, 1968.

Moeser, S. D. and Bregman, A. S. The role of reference in the acquisition of a miniature artificial language. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 759-769.

Moeser, S. D. and Bregman, A. S. Imagery and language acquisition. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 91-98.

Moore, E. F. Gedanken experiments on sequential machines. Automata Studies, Princeton, 1956.

Olson, G. M. Developmental changes in memory and the acquisition of language. In T. E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press, 1973.

Pao, T. W. L. A solution of the syntactic induction-conference problem for a non-trivial subset context-free languages. Report No. 70-19, The Moore School of Electrical Engineering, University of Pennsylvania, August, 1969.

Quillian, M. R. The teachable language comprehender. Communications of the Association for Computing Machinery, 1969, 12, 459-476.

Reber, A. S. Transfer of structure in synthetic languages. Journal of Experimental Psychology, 1969, 81, 115-119.

Richards, I. A., Jasuilko, E. and Gibson, C. Russian through pictures, Book I. New York: Washington Square Press, 1961.

Robinson, J. A. A machine-oriented logic based on the resolution principle. Journal of the ACM, 1965, 12, 23-41.

Rumelhart, D. E., Lindsay, P, and Norman, D. A. A process model for long-term memory. In E. Tulving and W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Saporta, S., Blumenthal, A. L., Lackowski, P. and Reiff, D. G. Grammatical models of language learning. In R. J. DePietro (Ed.), Monograph Series on language and linguistics. Vol. 16. Report of the 14th Annual Round Table Meeting on Linguistic and Language Studies, 1963, 133-142.

Schank, R. C. Conceptual dependency: A theory of natural-language understanding. Cognitive Psychology, 1972, 3, 552-631.

Schlesinger, I. M. Production of utterances and language acquisition. In D. I. Slobin (Ed.), The ontogenesis of grammar. New York: Academic Press, 1971.

Scholes, R. J. The role of grammaticality in the imitation of word strings by children and adults. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 225-228.

Scholes, R. J. On functors and contentives in children's imitations of word strings. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 167-170.

Schwarcz, R. M., B rger, J. F., and Simmons, R. F. A deductive question-answerer for natural language inference. Communications of the Association for Computing Machinery, 1970, 13, 167-183.

Shamir, E. and Bar-Hillel, Y. Review 2476. Computing Reviews, 1962, 3, 5.

Siklossy, L. A language-learning heuristic program. Cognitive Psychology, 1971, 2, 479-495.

Simmons, R. F. Natural language question-answering systems: 1969. Communications of the Association for Computing Machinery, 1970, 13, 15-30.

Simmons, R. F. Semantic networks: Their computations and use for understanding English sentences. In R. C. Schank and K. M Colby (Eds.), Computer models of thought and language. San Francisco: Freeman, 1973.

Simon, H. A. The sciences of the artificial. Cambridge, Mass.: MIT Press, 1969.

Sinclair-deZwart, H. Language acquisition and cognitive development. In T. E. Moore (Ed.), Cognitive development and the acquisition of language. New York: Academic Press, 1973.

Slagle, J. R. Experiments with a deductive question-answering program. Communications of the Association for Computing Machinery, 1965, 8, 792-798.

Slagle, J. R. Automatic theorem proving with renamable and semantic resolution. Journal of the ACM, 1967, 14, 687-697.

Slobin, D. I. The ontogenesis of grammar. New York: Academic Press, 1971.

Solomonoff, R. J. A formal theory of inductive inference, Part II. Information and Control, 1964, 7, 224-254.

Weizenbaum, J. ELIZA - a computer program for the study of natural language communications between man and machine. Communications of the ACM, 1966, 1, 36-45.

Wilks, Y. The Stanford MT and understanding project. In Schank and Colby (Eds.), Computer models of thought and language. San Francisco, 1973.

Winograd, T. Understanding natural language. Cognitive Psychology, 1972, 3, 1-191.

Winston, P. H. Learning structural descriptions from examples. MIT Artificial Intelligence Laboratory Project AI-TR-231, 1970.

Woods, W. A. Procedural semantics for a question-answering machine. Proceedings of the 1968 Fall Joint Computer Conference, 457-471.

Woods, W. A. Progress in natural language understanding: An application to lunar geology. AFIPS Proceedings, 1973 National Computer Conference and Exposition.

Woods, W. A. Transition network grammars for natural language analysis. Communications of the ACM, 1970, 13, 591-606.