UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
A Computational Model Which Addresses Errors of Over-Generalization and Their Subsequent Disappearance in Early Child Language Acquisition

Permalink
https://escholarship.org/uc/item/13t3r4hz

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 8(0)

Author
Hill, Jane C.

Publication Date
1986

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

A COMPUTATIONAL MODEL WHICH ADDRESSES ERRORS OF OVER-GENERALIZATION AND THEIR SUBSEQUENT DISAPPEARANCE IN EARLY CHILD LANGUAGE ACQUISITION

Jane C. Hill
Computer Science Department
Smith College

ABSTRACT

The model discussed here is offered as a prototype of the use of a computational model to explore alternate hypotheses and to suggest possible answers to some of the questions which have been addressed in the study of language acquisition. Why does not the child end up with an overly generalized grammar or lexicon? There is much evidence concerning the kinds of generalizations and over-generalizations that children make. However if we permit no overt and specific correction of the child's errors, then how is it that errors of over-generalization do not persist into adult speech? One answer to this question is proffered by attaching a system of weights to hypotheses. There are two related problems to be solved. Some mechanism in the model must allow erroneous hypotheses to be corrected; in addition there must be a way that more mature constructs can replace earlier ones. The model accomplishes these two tasks by means of a system of weights which represent confidence values and recency values. By this system more frequently matched constructs are preferred over less frequently matched constructs, and more recent hypotheses are favored for testing. This learning paradigm is illustrated by a set of procedures for learning the past tense of verbs in English. The scheme has the advantage that for a period of time when confidence factors are approximately in balance two or more constructs can co-exist. Thus we need not talk of rules or individual cases which have been learned or have not yet been learned but rather of a continuum in which rule schemas are either strong or weak.

WHY MODEL?

The development of computational models of language aquisition is a science which is in its infancy. The model discussed here is offered as a prototype of the use of a computational model to explore alternate hypotheses and to suggest possible answers to some of the questions which have been addressed in the study of language acquisition. Computational models such as ours may be used as a tool for linguists and psychologists in the task of exploring and contrasting alternative theories about the way in which children learn their native language. We would claim that such models enable the theorist to achieve a degree of explicitness which is virtually impossible to attain when working only with pencil and paper. Equally important are the questions which such models raise which might in theory be overlooked but which must be addressed in the implementation of a computational model. Such models encourage a comparison of the results of a theory to actual data which in turn aids in refining and evaluating the theory.

There are of course many different approaches which a computational model of language acquisition may take. One approach is to proceed from a characterization of adult grammar and work backwards to see how the child might arrive at this characterization. An alternative approach, and the one embodied in our model, is to work forward from the evidence that the child provides toward some characterization of the adult language. Our model must ask not only under what conditions a language can be learned by a computational model, but also does the model learn the language in the same way that the child does? Particular attention must be paid to the errors that children make since child speech that differs from adult speech yields clues concerning the processes that the child is using to understand and produce speech. An equally important clue to the child's processes are those errors which the child typically does not make. One frequently noted phenomenon is the over-generalization of the plural of nouns and of the past tense of verbs in English. Such phenomena need explanation. Our computational model of language acquisition therefore must not only learn to understand and generate sentences of ever greater complexity as does the child, but to provide a satisfying explanation of the course of language acquisition the model must make the same kinds of errors that the child makes and eventually correct the errors after further learning has occurred. We have recently extended our model to attend to and learn past tense forms of verbs. It is gratifying to observe that the same processes which enabled the acquisition of a simple grammar extend very naturally to the learning of past tense verb forms.

In this paper we emphasize the somewhat different sorts of answers which computational models may suggest to traditional questions, principally because of an approach to language acquisition as acquiring a set of dynamic processes as opposed to acquiring a set of rules.

OVERVIEW OF MODEL

We will now describe our model of the acquisition of English in the two-year-old. Our model is an on-going research project, but the mechanisms discussed here have been fully implemented. We can give only a brief overview of the model and its assumptions in this paper. Readers interested in details of the model should consult Hill (1982, 1983) for full particulars of the model including examples of computer output together with corresponding linguistic data collected from a two-year-old child. The psychological validity of the model is defended in Hill and Arbib (1984) and in Hill (1984). The model is described as a member of the class of schema-theoretic models in Arbib, Conklin, and Hill (1986). A discussion of the use of the model in investigating the child's understanding and formation of coordinate structure is to be found in Hill (1985). It is characteristic of our model that the internal representation of the learning which takes place is more important than the output of the system, so the model must be described in terms of the

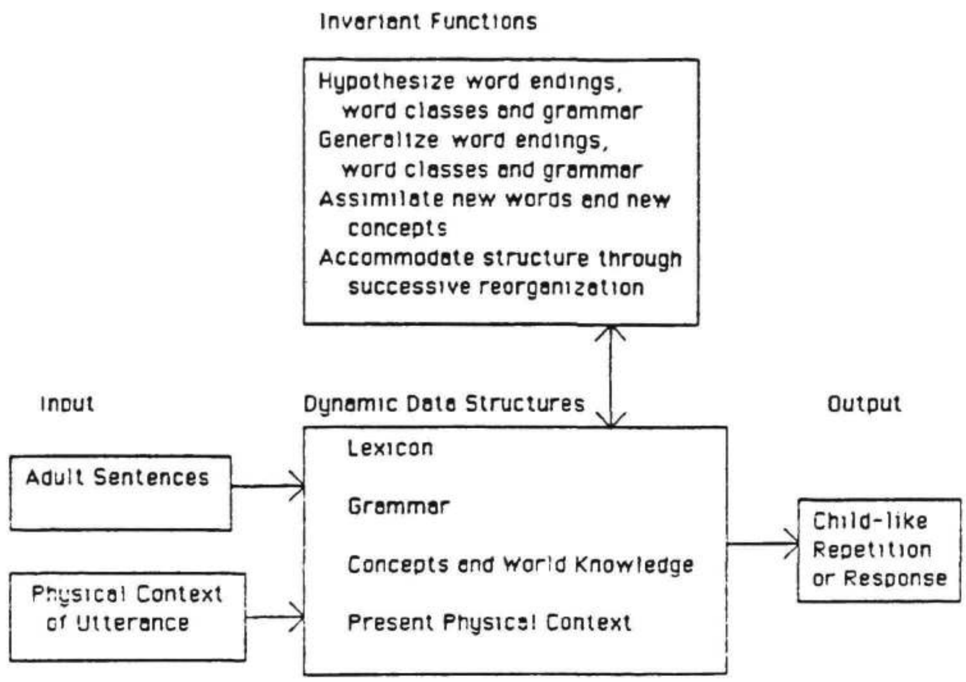


Figure 1
Components of Our Model of Language Acquisition in the Two-Year-Old

knowledge structures which are built as the model acquires language as well as in terms of the output.

Figure 1 provides a diagram of the components of the model. The model takes as its input adult sentences together with indications provided by the user, where relevant, of the physical context in which the sentences are uttered. Output from the model is a representation of child-like sentences repeating or responding to the adult input in accordance with the current state of the model's linguistic capacity. The child's knowledge is represented by dynamic data structures encoding the child's lexicon, the child's grammar, the conceptual knowledge of the child, and the physical context of the dialogue. The model is given a basic lexicon and a set of concepts with a mapping between the two. No assumptions have been made about the ultimate form of the adult grammar nor about what must be built-in to the model, but a precise account is kept of the knowledge and processes which are found to be necessary to be built-in to the model even for this elementary level of language understanding and production. Processes attend to the adult input and use rules of salience to focus on examples within the adult data which are used as the basis for language growth. The model is written in LISP using the semantic net language GRASPER (Lowrance 1978). The world knowledge is encoded in a semantic net as are the grammar templates and the lexicon.

The model uses its language experience to build a grammar which is at first a flat template grammar but which eventually evolves into a grammar which is actually a procedural grammar but which can be described by a set of recursive context-free phrase structure rules. The model notices and employs relations and word order, employs rules for concatenating relations and deleting words, and as the model grows word classes are formed, and new word forms are learned.

Let us explore the use of specific examples drawn from the input data. We will assume that gave has been identified as a relation-word and placed on the gleaning list. From an adult sentence such as "Daddy gave the toy to the boy" the model might initially respond with a single word such as toy. A subsequent presentation of the same sentence might cause the model to acquire a template for gave toy where gave would be classified as a relation-word and toy as a slot-filler. Yet another presentation of the sentence might cause the model to learn the template Daddy gave where Daddy was a slot-filler, and eventually the template (slot1 gave slot2) would be learned for Daddy gave toy. The learning is highly dynamic in that each time the same body of input is presented to the model a different set of grammar rules and additional lexical class information may be learned. What is learned in each presentation of the input depends upon the language experience of the model and what has been learned so far. No information is given the model about word classes, but hearing sentences such as "Mommy gave the toy," "John gave the book," "Sue gave the puzzle," would eventually cause the model to put toy, book, and puzzle all together in a word class meaning words which stand for possible objects of the relation-word gave. Note that it would not matter if the input sentences were far more complex than those used here for illustration. There is no requirement for correct and ordered exemplars. Typically we use as input adult sentences taken from a transcribed session of adult/child conversation. In the computer run from which illustrations have been chosen for this paper the adult sentences were taken from the Adam corpus which was collected by Roger Brown, Ursula Belugi, Colin Fraser, and Courtney Cazden prior to 1973 (Brown 1973) and which has been made available to us through the Child Language Data Exchange System. (For details see MacWhinney & Snow 1985). By the processes described above word classes are derived from the child's own ability to produce language. If the model is focussing on the word gave, then a sentence such as "Mommy gave the toy to Sue while she went into the store to buy groceries" would have just the same effect as the short sentences used above for an illustration. These processes result in a multiplicity of overlapping and intersecting word classes. The model requires schemas for word classification and template classification in order to grow, but the actual classes remain flexible. Processes of generalization eventually also permit the classifying of relation-words which might permit, for example, giving and bringing to be relation-words that could be classed together as words which have similar syntactic properties.

Successive reorganizations of the grammar and the lexicon occur as learning takes place. This process of gradual broadening of word classes and grammatical rules from applying to specific exemplars to sets of specific exemplars to more general categories has been defended in Kuczaj (1982), and Maratsos and Chalkley (1980). In this fashion the model suggests one way in which language based initially on cognitive knowledge can grow into a syntactic system which will be increasingly independent of its semantic and cognitive foundations. It is important to note that although the rules embodied in the model are simple, their interaction is complex enough to necessitate the use of a computer model.

The model to date has attained only the level of a two-year-old producing sentences of up to six words in length. Initially the grammar acquired is entirely flat and is made up of rules for forming two-word utterances expressing relations and for combining those two-word relations into utterances which may be two to six words in length. It can understand and produce coordinate structure such as that found in the Adam corpus, ages two years three months up to two years eleven months, and with the enhancements to be described below it can now acquire a subset of English suffixes and prefixes.

DYNAMIC RULE SCHEMAS AND THE USE OF CONFIDENCE FACTORS IN THE MODEL

Why does not the child end up with an overly generalized grammar or lexicon? There is much discussion in the literature concerning the kinds of generalizations and over-generalizations that children make. (See for example Brown,1973, and deVilliers and deVilliers, 1978). We believe that it is important to focus on the errors that children make because of the insights which they yield concerning the processes that the child employs in language acquisition. Bowerman (1974) states this position very clearly. A study by Bybee and Slobin (1982) presents a careful examination of the acquisition of irregular past-tense forms of verbs in English. If, however, we permit no overt and specific correction of the child's errors, then how shall we explain the fact that errors of over-generalization do not persist into adult speech? Consider the verb, <u>break</u>. It is an empirical fact that children at the earliest stage of language acquisition typically learn the word <u>broke</u> and seem to use it correctly. One may assume that such forms have been learned by rote. Then at a subsequent stage of development the child will start to use the word <u>breaked</u>. Presumably this is because the child has formed a general schema for forming the past tense of verbs. Eventually of course children learn that <u>break</u> is an irregular verb and does not obey the general rule in the form of its past tense. But the puzzle is that for a period of time, sometimes for years, both forms exist in the child's vocabulary. How can this period of imbalance between the erroneous and the correct forms be explained? This behavior cannot be explained if the language mechanism is expressed in terms of explicit rules which the child either knows or does not know.

One answer to this question is proffered by our computational model which attaches a system of weights to hypotheses about word forms and grammar rules. These weights are meant to represent the relative strength of the various hypotheses. (An alternative model of this phenomenon is offered in the connectionist model of Rumelhart & McClelland, 1987, to appear.) There are two related problems to be solved. Some mechanism in the model must allow erroneous hypotheses to be corrected; in addition there must be a way that more mature constructs can replace earlier ones. Our model accomplishes these two tasks by means of a system of confidence values and recency values. A confidence value is associated with each hypothesis and this confidence value is increased each time the hypothesis is instantiated in the adult speech input; the confidence value is increased to a lesser degree each time the hypothesis is instantiated in the child speech output. In this way more frequently matched constructs come to be preferred over (given a higher confidence factor than) less frequently matched constructs. Hypotheses must be reinforced to survive. If new hypotheses, however, are to start with very low confidence values they will have trouble "catching up" with earlier hypotheses. For this reason separate recency values are employed whose function it is to cause more recent hypotheses to be favored for testing. We will refer to The combination of confidence values and recency values as _confidence factors_. The use of weights to direct learning in computational models is by no means unique. Similar weighting schemes have been employed in many computational models. (See, for example, Kelley, 1967.) By means of the use of confidence factors our model attains the desired effects.

To illustrate, if the model were given a lexical pair such as _tie_ and _untie_ and the knowledge that _untie_ is the reverse process of _tie_, then the model might hypothesize a set of similar process pairs such as _cover_/_uncover_, _drop_/_undrop_, _hang_/_unhang_. It might be that only after a great deal of language experience will the hypothesized lexical entries _unhang_ and _undrop_ be forgotten. This general scheme has the advantage that for a period of time when confidence factors are approximately in balance two or more constructs can co-exist, as, for example, in the case of the past-tense verb _broke_ and the over-generalized form _breaked_ for the verb _break_. Thus we need not talk of rules or individual cases which have been learned or have not yet been learned but rather of a continuum in which rule procedures are either strong or weak.

LEARNING PAST-TENSE VERB FORMS IN ENGLISH

It is especially interesting to explore the use of verbs in English in the developing language of the child since learning English is intimately tied to the learning of verbs. DeVilliers (1985) has found evidence that input language has a significant impact on the child's developing language with respect to verbs. The mother's use of verbs is a high predictor of the child's use of verbs. Note that it is not the frequency of the mother's use, but rather the variety of verb forms in the mother's use of a particular verb which is a significant predictor of the variety of forms the child will use for that same verb. This is interpreted to mean that the child is monitoring the input for clues about the grammatical prototypicality of forms of individual verbs. Wide differences in the use of verbs between subjects were found in the samples considered in her study. Verbs with a variety of heard uses were used with greater confidence by the child even in unheard contexts. Our model simulates this monitoring process. DeVilliers' analysis did not address the issue of over-generalization but it does lend credence to our processes which rely on the information gleaned from the input by focussing on different constructs at different times for the learning of the forms.

The learning paradigm is as follows:

1. Observe a correlation between past-tense verbs and **-ed** endings

2. Place -**ed** at the end of all action-verbs and give a modest confidence factor to all these past-tense forms

3. Proceed to modify the confidence factors depending on the experience of the model as follows.

   - add a small increment to the confidence factor of a given form each time the model produces that past-tense form

   - add a larger increment to the confidence factor of a given form each time the model notices that form in the adult input

Figure 2
Learning Past-Tense Forms of Verbs in English Through Use of Dynamic Rule Schemas and Confidence Factors

Figure 2 summarizes our model's paradigm for learning past-tense forms of English verbs. In order to observe the correlation between past-tense and -ed endings, the model must be given a representation of time-past in its cognitive knowledge, and the ability to identify action verbs in its lexicon. The model simulation begins by forming past-tense entries in its lexicon for all action verbs simply by adding -ed endings. Each of these forms is initially given a modest confidence factor. The model then proceeds to modify the confidence factors of the past-tense forms depending upon its language experience. A small increment is added to the confidence factor of a form each time that the model produces a past-tense form; a larger increment is added each time that the model finds a past-tense form in the adult input.

The choice of past-tense form is entirely dependent upon the input sentences used, so no conclusions can be drawn about the specific past-tense forms which are learned, but depending upon the input data

- The model may keep an erroneous -ed ending

- The model may proceed through a period of instability in which it will vacillate between -ed and irregular forms

- The model may discard the erroneous -**ed** form and replace it by the irregular form

Figure 3
Observed Behavior of the Model with respect to past-tense verb forms in English

Sample Model Output -- First Iteration

Don t you think you should pick it up and put it in the basket?
Basket

See the other one and put them together?
Together

we ll take his coat and hang it up all right?
Up

when you open the door and go outside what do you put on?
On

Oh Dale sit in the chair and watch the game?
In chair

Won t you please go over and take Ursula s pocketbook to her?
Pocketbook

Do you want Mommy to take it and put it up?
Up

Do you want Mommy to take the whistle and put it up?
Up

Why don t you pick the books up and take them over to Ursula?
Over

Yes, you fell down and broke your horn didn t you?
Horn

Yes, you put them in your bank, remember, and it broke your bank?
Bank

Do you want Mommy to take it and put it up?
Puted up

Remember we took the basket to her?
Basket

If you fall down you ll break the horn
Horn

Sample Model Output -- Second Iteration

Don t you think you should pick it up and put it in the basket?
Picked up and puted in basket

See the other one and put them together?
Together

We ll take his coat and hang it up all right?
Taked coat

When you open the door and go outside what do you put on?
Puted on

Oh Dale sit in the chair and watch the game?
In chair and watched game

Won t you please go over and take Ursula s pocketbook to her?
Taked pocketbook

Do you want Mommy to take it and put it up?
Puted up

Do you want Mommy to take the whistle and put it up?
Puted up

Why don t you pick the books up and take them over to Ursula?
Picked books

Yes, you fell down and broke your horn didn t you?
Broked horn

Yes, you put them in your bank, remember, and it broke your bank?
Broked bank

Do you want Mommy to take it and put it up?
Taked up and puted up

Remember we took the basket to her?
Tooked

If you fall down you ll break the horn
Falled down and broke horn

Sample Model Output -- Final Iteration

Don t you think you should pick it up and put it in the basket?
Picked up and put in basket

See the other one and put them together?
Together

We ll take his coat and hang it up all right?
Taked coat

When you open the door and go outside what do you put on?
What puted on

Oh Dale sit in the chair and watch the game?
In chair and watched game

Won t you please go over and take Ursula s pocketbook to her?
Taked pocketbook

Do you want Mommy to take it and put it up?
Taked up and puted up

Do you want Mommy to take the whistle and put it up?
Taked whistle and puted up

Why don t you pick the books up and take them over to Ursula?
Picked books

Yes, you fell down and broke your horn didn t you?
Broke horn

Yes, you put them in your bank, remember, and it broke your bank?
Broke bank

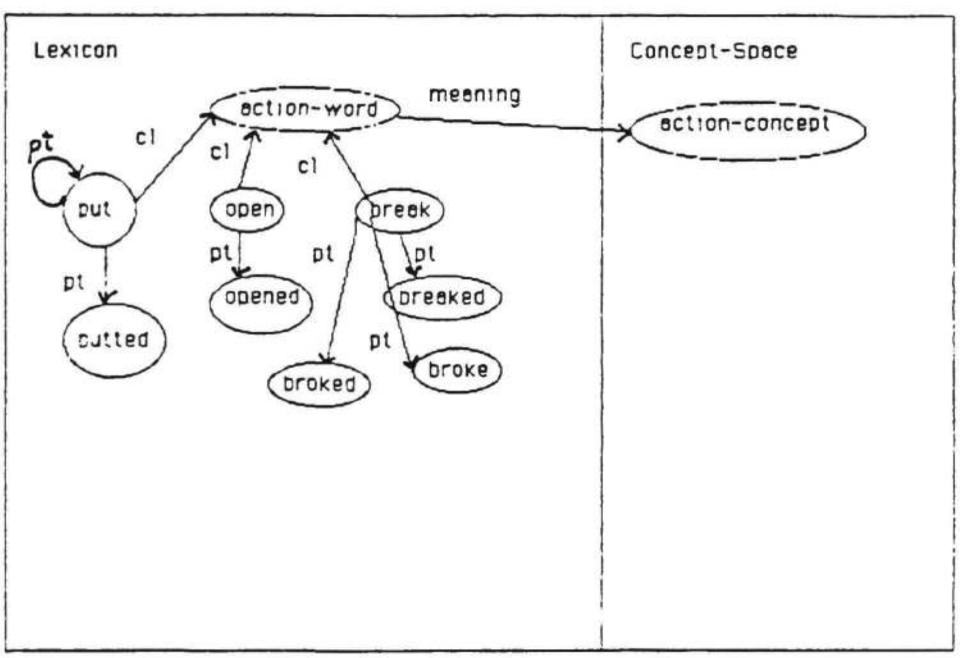
Do you want Mommy to take it and put it up?
Taked up and put up

Remember we took the basket to her?
Tooked basket

If you fall down you ll break the horn
Fell down and broke horn

Figure 4

The resulting behavior of the model is summarized in Figure 3. Since the choice of past-tense form is entirely dependent upon the input to the model, no conclusions can be drawn about the the specific past-tense forms which are learned, but depending upon the input data, the model may (1) keep an erroneous -ed ending, may (2) proceed through a period of instability in which the output vacillates between an erroneous -ed ending and the correct irregular form, or (3) the model may discard the erroneous form and replace it by the irregular form. All these forms of behavior are exhibited in the sample output from a series of computer runs which is presented for illustration in Figure 4. To make apparent the changing preferences of the model, only past-tense forms are shown in the output exhibited here. A portion of the internal representation of the model is graphically represented in Figure 5, the information given the model for this illustration is presented in Figure 6, and a summary of the confidence factors attached to the different past-tense forms is summarized in Figure 7.



cl = class
pt = past-tense form

Figure 5
A Schematic Representation of a Portion of the Internal Representation of Our Model

Since developing the capacity to learn the past tense forms of verbs, we have successfully applied the same strategies to the learning of plural forms of nouns which name physical objects. We believe that this ability to generalize the processes employed by the model lends support to the validity of the processes.

Lexicon

away, bank, basket, bat, book, break, broke, chair, coat, come, Daddy, Dale, do, door, down, fall, fell, game, glove, go, hang, here, horn, in, Mommy, on, one, open, other, out, outside, over, paper, pick, pocketbook, put, racket, see, sit, take, that, time, together, took, try, up, use, Ursula, want, watch, what, when, whistle, you

Relation list

(possn-reln $owner), (breaking-reln break),(broking-reln broke), (coming-reln come), (doing-reln do), (down-reln down), (falling-reln fall), (felling-reln fell), (going-reln go), (hanging-reln hang), (here-reln here), (in-reln in), (on-reln on), (opening-reln open), (out-reln out), (outside-reln outside), (over-reln over), (picking-reln pick), (putting-reln put), (seeing-reln see), (sitting-reln sit), (that-reln that), (taking-reln take), (tooking-reln took), (trying-reln try), (up-reln up), (using-reln use), (wanting-reln want), (watching-reln watch), (what-reln what), (when-reln when)

Feature list

person:Dale, Daddy, Mommy, Ursula

place: here, outside

Slot-filling features

possn-reln $owner person

here-reln $slot1:pointing-obj

Closed class

and

Figure 6

Information Given the Model for this run

	First Run	Second Run	Final Run
breaked	.8	.8	.8
broke	.955	.99 *	.995 *
broked	.957 *	.957	.957
failed	.909 *?	.909	.909
fell	.75	.958 *	.981 *
felled	.909 *?	.957	.956
opened	.875	.941	.961
picked	.955	.983	.992
put	.944	.984	995 *
puted	.967 *	.99 *	.993
taked	.96 *	.986 *	.987
took	.875	.958	.979
tooked	.941	.976	.993 *

* wins

Figure 7

Confidence Factors Attached to Past-Tense Forms

CONCLUSION

We would emphasize that the purpose of this illustration is not to make any specific claims about the learning of past-tense forms in English, but rather to illustrate that a paradigm such as ours may be sensitive to the input data and my exhibit varied behavior. The use of dynamic rule schemas and confidence factors has been used to model the phenomena of generalization, over-generalization, and subsequent correction of over-generalized forms.

This discussion has been offered as an example of the kind of answers that computational models can proffer for consideration and experimentation. Other issues which the same model explores are (1) what variation occurs in the model as specific constraints are built-in or omitted, (2) how can the use of input filters focus on different aspects of the input data over time, (3) how does variation in meaning representation and sets of semantic features affect the learning process. We believe that the development of models such as ours will have a large impact on future work in language acquisition.

REFERENCES

Arbib, M. A., Conklin, J. E., and Hill, J. (1986) From Schema Theory to Language, Oxford University Press, New York, NY.

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

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

Bybee, J., & Slobin, D. (1982) "Rules and Schemas in the Development and Use of the English Past Tense", Language, vol. 58, 2, pp. 265-289.

deVilliers, J. (1985) "Learning How to Use Verbs: Lexical Coding and the Influence of the Input", Journal of Child Language, 12, pp. 587-595.

deVilliers, J., & deVilliers, P. (1979) Early Language, Harvard University Press, Cambridge, MA.

Hill, J. (1985) "Using a Computational Model of Language Acquisition to Address Questions in Linguistic Inquiry", Proceedings of the Seventh Annual Conference of the Cognitive Science Society, Irvine, CA.

________ (1984) "Combining Two Term Relations" Evidence in Support of Flat Structure," Journal of Child Language, 11,pp. 673-678.

________ (1983) "A Computational Model of Language Acquisition in the Two-Year-Old," Cognition and Brain Theory, 6(3),pp. 287-317.

________ (1982) "A Computational Model of Language Acquisition in the Two-year-old," Ph.D. Dissertation, University of Massachusetts at Amherst, reproduced by the Indiana University Linguistics Club, Bloomington IN, February 1983.

Hill, J., & Arbib, M. A. (1984) "Schemas, Computation, and Language Acquisition," Human development, 27,pp. 282-296.

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

Kuczaj, S.A. II, (1982) "On the Nature of Syntactic Development", in S. Kuczaj, ed, Language Development: Volume 1: Syntax and Semantics, Lawrence Erlbaum Associates, Hillsdale, N.J.

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

MacWhinney, B., & Snow, C. (1985) "The Child Data Exchange System", _Journal of Child Language_, 12, pp. 271-296.

Maratsos, M., & Chalkley, M. (1980) "The Internal Language of Children's Syntax: The Ontogenesis and Representation of Syntactic Categories", In K. Nelson, ed., _Children's Language: Vol.2_, Gardner Press, N.Y.

Rumelhart, D., & McClelland, J. (1987, to appear) "Learning the Past Tenses of English Verbs: Implicit Rules or Parallel Distributive Processes?", In B. MacWhinney, ed., _Mechanisms of Language Acquisition_, Lawrence Erlbaum Associates, Hillsdale, NJ.