Why Do Children Say "Goed"?
A Computer Model of Child Generation

Mallory Selfridge
Department of EE and CS
University of Connecticut, Storrs, CT

1.0 Introduction

An important question in modelling child language generation is why children say regular forms of irregular words, such as "goed", (Clark and Clark, 1978) during development, although they never hear them. Three other general characteristics of children's generation also require explanation. First, Benedict's (1976) work suggests clearly that comprehension of various aspects of language precede the generation of those aspects. Second, the length of the utterances children say becomes generally longer as development proceeds (Bloom, 1973). Third, Wetstone and Friedlander (1973) suggest that first children say things in the wrong order, and then say things in the correct order.

In order to address these issues, this paper explores the hypothesis that learning to talk is driven by learning to understand. This hypothesis begins by assuming that the principal effect of learning to understand is the development of the lexicon as additional words are learned and their "definitions" are refined and modified. It further assumes that the language generation process is not learned, but is an innate part of a child's cognitive repertoire. Finally, it states that the ability to generate grows as the lexicon develops during the development of comprehension. The hypothesis predicts that a computer model which incorporated it would display the characteristics described above.

CHILD (Selfridge, 1980) was initially developed to model a subset of the development of comprehension about a limited domain in a child between the ages of one and two years, using context-based inference and learning rules to build a dictionary accessable to a conceptual analyzer (Birnbaum and Selfridge, 1979). New words were added to the dictionary, the meanings of ones it had learned were refined, and syntactic information on how to combine word meanings were stored under appropriate words. Meaning was modelled using Conceptual Dependency (Schank, 1973), while syntactic knowledge was represented using Sequential Structure (Selfridge, 1980) which encodes the utterance position of the filler of a slot of a word meaning. Thus CHILD's learning was based upon meaning, with syntactic knowledge indexed upon this meaning.

CHILD has now been equipped with a generator (Cullingford, Krueger, and Selfridge, 1981), which has access to the dictionary built up during comprehension learning. CHILD now learns to generate in the same limited domain, and offers explanations for the psychological phenomena described. In particular, although CHILD does not yet say "goed", it provides a precise account of how it could be given experiences which would lead it to say "goed".

2.0 Learning to Generate

Children speak in many different situations. CHILD only generates language in one of these, that in which a child describes something observed (Halliday, 1975). The user teaches CHILD to understand by providing utterances in situations in which CHILD can infer their meaning. To demonstrate the development of generation, the user provides CHILD with a Conceptual Dependency concept, simulating visual input. When given such a concept, CHILD attempts to generate it.

The following sequence of utterances interspersed with transitions summarizes the development of CHILD's generation capacity, and is drawn from a full run of CHILD during which it both learns to comprehend and generate. The statements referring to CHILD "learning" meaning and syntax summarize the learning of comprehension described in Selfridge (1980). In order to show development, CHILD has been supplied with the same concept to generate repeatedly: the concept for "the parent puts the ball on the table", (PTRANS ACTOR (PARENT1) OBJECT (BALL1) TO (TOP VAL (TABLE1))).

CHILD knows no words

CHILD says "um"

CHILD learns meaning of "ball"

CHILD says "ball"

CHILD learns meaning of "put"

CHILD says "ball put"

CHILD learns meaning of "table"

CHILD says "table ball put"

CHILD learns syntax of "put"

CHILD says "put ball table"

CHILD learns meaning of "on"

CHILD says "put ball table on"

CHILD learns syntax of "on"

CHILD says "put ball on table"

This progression shows that CHILD's generation does correspond to the data described earlier. First, CHILD certainly does learn to generate after it learns to understand. Second, the length of its utterances does indeed grow. Third, the ability to generate structurally correct utterances follows the ability to generate incorrect utterances.

3.0 Why Would CHILD Say "Goed"?

There are several properties of CHILD which would lead it to say "goed". First, its dictionary is ordered and the first word found during lookup which matches the concept being generated is used. Second, words are not forgotton and removed from the dictionary, but they may be "covered-up" by words learned later. Third, both new words and words whose meaning has been modified are placed at the top of the dictionary where they will be found first during lookup. The following developmental progression accounts for why children say "goed" according to the CHILD model. Each of the proposed learning events can be modelled by CHILD's existing learning rules.

The first stage is that in which children learn "go" and "went" as meaning PTRANS with the TIME slot containing PRESENT and PAST respectivly. The order in the dictionary is, top to bottom, "went" and "go". "Went" is on the top because it is presumably learned later than "go." At this stage the child will use these words correctly, since when he wants to generate a PTRANS with TIME as PAST he will lookup "went" directly, and when he wants to generate PTRANS with TIME as PRESENT, he will find "go" directly.

The second stage occurs when he learns "ed" as a separate lexical item, whose meaning is the filler PAST. This is learned during comprehension in the

same way other words are learned, and as a result "ed" is placed on the dictionary, whose order is now "ed", "went", "go". Although "ed" is now in the dictionary, the child desiring to generate PTRANS with TIME either PAST or PRESENT will still use "go" and "went", since dictionary lookup matches on the root concept, and the meaning of "ed" doesn't match PTRANS.

At the third stage, the child learn that the meaning of "go" does not include the PRESENT filler of the time slot, perhaps because the child is learning about "go" and the future tense. He must also learn that the filler of the TIME slot of "go" must follow "go", perhaps because he learns to understand "go ing". Learning a modified meaning of "go" results in "go" being placed on the front of the dictionary, and its order is now "go", "ed", "went". At this stage, the child desiring to generate PTRANS with TIME as PRESENT will find the meaning of "go" matching this PTRANS, and will then search the dictionary to generate the PRESENT sub-concept. Since there is nothing there, he will use "go" alone. However, to generate PTRANS with TIME as PAST, he will again find "go" matching the PTRANS, and will search the dictionary to express the PAST subconcept. This time, however, he will find "ed", and will thus generate "goed".

The fourth stage occurs when the child hears "went" again, perhaps in ordinary discourse or as a parental correction to "goed". This correction results in "went" being placed in the front of the dictionary, whose order is now "went", "go", "ed". At this point, the child will use "go" for PTRANS with TIME as PRESENT, "ed" is available for expressing the PAST time for ordinary action words, but he will cease using "goed", since he finds "went" expresses PTRANS with TIME as PAST directy.

4.0 Conclusions

CHILD offers an explanation for the psychological data described earlier. Generation follows comprehension because generation cannot occur until comprehension learning adds words to the dictionary. Length of utterances increases because the number of words available to express a concept increases during comprehension learning. Utterances are generated with incorrect structure before they are generated with proper structure because syntax is indexed under word meanings, word meanings are learned before their syntax, and word meanings without syntax are available to the generator before word meanings with syntax. In particular, this paper proposes a specific explanation of why children say "goed": because generation is driven by understanding, because the dictionary is ordered and the first word found matching the concept to be generated is used, and because new words and words with refined meanings are placed on the top of the dictionary.

This account of why children say "goed" makes four specific predictions. First, since "went" is still in the dictionary even though the child says "goed", the child will still understand the utterances containing "went" during the stage at which he is generating "goed". Second, since "went" was relearned as a result of an experience specifically with "went", this model predicts that a child will continue to say "goed" at least until he has heard "went" again; that is, after he begins to use "goed", he will never say "went" before he hears "went" again. Third, since corrections or relearning are specific to individual words, the transition from the third to fourth stages for various irregular words will occur individually. That is, if the child is saying "goed" and "runned", learning "went" again will not result in the child also saying "ran". Rather, a specific experience with "ran" is needed. Fourth, this model predicts that a child will never say "goed" until after he has learned to understand "go" in a tense other than the present, because this experience teaches him that the TIME slot under "go" is empty, as is needed to say "goed".

Acknowledgements

Dr. Roger Schank's assistance in this work was invaluable. Dr. Richard Cullingford has made many valuable suggestions, and Dr. Katherine Nelson has provided many insights into problems of modelling child language learning. Larry Birnbaum, Marie Bienkowski and Jamie Callan have contributed both ideas and criticisms.

Bibliography

Benedict, H. (1976). Language Comprehension in 10 to 16 Month-old Infants. (thesis) Department of Psychology, Yale University, New Haven, Connecticut.

Birnbaum, L., and Selfridge, M. (1981). in Inside Artificial Intelligence: Five Programs plus Miniatures. Schank R. and Riesbeck C.K. (eds.), Lawrence Earlbaum Associates, Hillsdale, NJ.

Bloom, L., 1973. One Word at a Time. Mouton, The Hague.

Bruner, J.S., Goodnow, J. J., and Austin, G.A., (1956). A Study of Thinking. John Wiley and Sons, New York

Clark, E.V. and Clark, H.H., (1978). Psychology and Language, Harcourt Brace Jovanovich.

Cullingford, R.E., Krueger, M.W., and Selfridge, M. (1981) A Picture and a Thousand Words: Automated Explanations as a Component of a Computer-Aided Design System. Forthcoming.

Halliday, M.A.K., (1975). Learning How to Mean: Explorations in the Development of Language Edward Arnold, London.

Schank, R. C., (1973). Identification of Conceptualizations Underlying Natural Language. In R. C. Schank and K. M. Colby (eds.) Computer Models of Thought and Language W.H. Freeman and Co., San Fransisco.

Selfridge, M. (1980). A Process Model of Language Acquisition. Computer Science Technical Report 172, Yale University, New Haven, Ct.

Westone, H. and Friedlander, (1973). The Effect of Word Order on Young Children's Responses to Simple Questions and Commands. Child Dev. 44:734-740

Winston, P. (1975). Learning Structural Descriptions from Examples. In P.H. Winston (ed.) The Psychology of Computer Vision. McGraw-Hill, New York.