3rd International Conference on
Spoken Language Processing (ICSLP 94)
Yokohama, Japan
September 18-22, 1994

BSLP BASED LANGUAGE GRAMMARS FOR CHILD SPEECH

P.V.S. Rao and Nandini Bondale

Tata Institute of Fundamental Research, Colaba, Bombay-400 005, India.

ABSTRACT

We report in this paper our investigations aimed at evolving an approach to develop a grammar model for the speech of a child acquiring English as the first language. We use the occurrence of words in similar syntacto-semantic environment as the criterion for classifying words into equivalence classes. Consequently, words in the same class are interchangeable in a phrase or larger context. We identify phrases on the basis of n-gram frequencies of 'class exemplars' in the corpus and their meaningfulness. We group them into a small number of interchangeable phrase types. It turns out that the corpus of 723 sentences consists of 161 distinct phrase type sequence patterns and that of these, just 70 patterns have a frequency more than one and account for as many as 632 sentences in the original corpus. We use an algorithmic procedure to build a 'State Transition Network' which accounts for the sentence patterns. The STN consists of 11 states and generates most of the sentences in the corpus. This indicates that the approach is effective for modeling the grammar of the child.

1. INTRODUCTION

This paper describes the use of a Blank Slate Language Processor (BSLP) [1] developed by us for use in a speech recognition application [2], for modeling the language capability of a child during the language acquisition phase. The fact that the BSLP does not make any a priori assumptions about the grammar of the language makes it possible to use it in a wide variety of situations where restricted grammars and vocabularies are used. The child uses a small (even if growing) vocabulary and has rudimentary grammatical competence. His speech is aimed at facilitating his rather limited requirements (e.g. food, attention, comfort etc). His language thus bears many similarities with those used in restricted task environments by man-machine interaction systems. We have used a modified version of our BSLP to evolve a model for a grammar of spontaneous speech of a two-and-a-half-year old child acquiring English as a first language.

2. THE BLANK SLATE LANGUAGE PROCESSOR

The BSLP was developed as a post processor for an acoustic level recognizer used in a speech recognition system developed by us. It was intended to improve the performance of the speech recognition system for a railway reservation enquiry application in the Hindi language using higher level language properties.

The main feature of the BSLP is that it does not make any a priori assumptions regarding the grammar of the language but starts as if on a blank slate and acquires knowledge regarding the language structure during an interactive training phase. This knowledge was stored by the system in the form of a phrase-oriented STN. It accepted several alternatives for each word as provided by the acoustic level recognizer along with the associated confidence levels. Using these it generated a 'most likely sentence hypothesis', consistent with the constraints imposed by syntax and semantics. Within limits, it also corrected word recognition errors made by the acoustic level recognizer.

3. GRAMMAR FOR CHILD SPEECH

In his early years, every normal child spontaneously constructs a grammar for the language he acquires, without any formal training. At any given time, the model of the grammar constructed by the child gets reflected in his speech. While modeling the grammar, our aim is to evolve an idiosyncratic model for the child, rather than depend on a so called 'standard grammar' for a natural language which conforms with the linguistic competence of a hypothetical native speaker belonging to a homogeneous linguistic community (Neither exist in reality.)[3]. Our approach is therefore to derive the grammar model for an individual, based on his utterances. We use the BSLP technique for accomplishing this. In the following sections we describe the steps used in generating a corpus driven model of a grammar for the speech of a child acquiring language. Consequently, in contrast to existing literature [4,5], our emphasis in this presentation is on the techniques that have been developed, rather than on the model itself or on the developmental stages of language acquisition by the child,

4. THE APPROACH

The approach we follow here is broadly indicated in *fig. 1*. The detailed description of each of the steps indicated by the blocks is described below.

4.1 The Speech Corpus

Spontaneous speech of a two and a half year old child has been recorded in the natural environment of his home, during his interaction with his parents in various routine everyday situations. It was collected in several sessions, spanning over five weeks. A total of 450 minutes of recording has been transcribed manually. This transcription amounts to 723 Sentences (including 31 single word utterances). The 2733 word corpus uses a vocabulary of 460 distinct words.

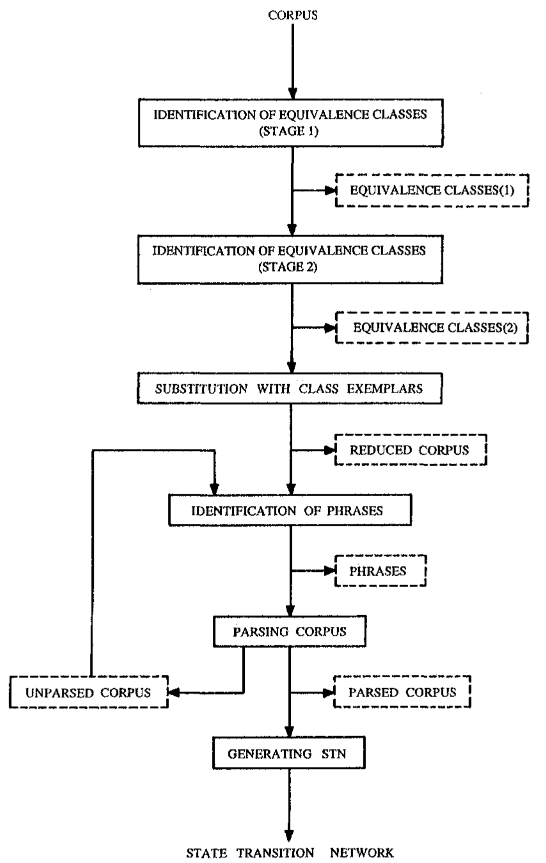

Fig. 1 - Generation of Corpus-driven grammar model for Child Speech using BSLP.

4.2 Syntactic Tagging

The first step in organizing the corpus for further processing is the categorization of the vocabulary comprising the corpus into equivalence classes. Each equivalence class would represent a 'syntactic category' and, possibly, a 'semantic subcategory'. Members of an equivalence class would be mutually interchangeable in a phrase frame or in larger contexts. The criterion used for classification of words is their occurrence in similar syntacto-semantic environments. This environment is defined in terms of the immediate neighbors of the word on either side. We do this classification in two stages.

4.2.1 Grouping (Stage 1) In the first stage of classification, we define context in terms of (30) frequently occurring unigrams, bigrams and trigrams. For each word in the corpus, we determine the conditional probabilities (CP's) of its occurrence to the left or to the right of each of these n-gram strings. (There will be two CP's per word per n-gram). We then map the words into a multidimensional space where each dimension defines the one of the CP's defined. We then cluster the words in this space. Ideally, each cluster so formed would cover an equivalence class; however, due to the limited corpus and the vagaries of grammar, some members of each class would be missed and some interlopers would get into each cluster. The final decision regarding classification of each word in a cluster is taken interactively. The results of this classification are summarized in Table 1.

TABLE 1
EQUIVALENCE CLASSES AT STAGE 1

Class 1 - (Nouns and Pronouns)
elephant(8), I(202), it(132), you(48), me(18), tea(17), we(15), milk(12), thing(11), animal(6), puzzle(5), noise(5), nail(5), ball(5)

Class 2 - (Verb, Auxiliary verb)
show(2), want(94), will(60), come(28), take(20), keep(20), sit(18), do(18), see(913), are(13), can(12), put(11), give(11), fall(11), drink(11)

Class 3 - (Adjective, Possessive pronoun)
nice(7), this(132), the(87), that(65), my(23), some(16), all(12), two(9), more(9), hundred(9), a(8).

Class 4 - (Preposition)
for(13), in(30), of(22), on(21), from(9), like(9), with(7)

The numbers in parentheses indicate frequencies of occurrence in the corpus.

4.2.2 Grouping (Stage 2) The second stage is aimed at identifying and capturing words which were missed during the first stage of classification. Contextual similarity with known members of an equivalence class would be a good criterion for inclusion in that class. To facilitate evaluation of this similarity, we arbitrarily choose Class Exemplars (CE's) to represent each class (the underlined words in Table 1) and 'reduce' the original corpus by substituting each classified word by the appropriate CE which, for all purposes, functions as a syntactic tag for that word. Statistics regarding occurrence of the CE in a given context indicate the overall behavior of the class as a whole. Hence, candidate words can be included in an equivalence class on the basis of similarity with the CE.

We use immediate context, i.e. occurrence as the central word of a trigram frame as the criterion for classification. A word is added to an equivalence class if it occurs within the same set of trigram frames as the class exemplar. The words which get added in this manner to the earlier equivalence classes are listed in Table 2.

TABLE 2
EQUIVALENCE CLASSES AT STAGE 2

Class no.	Words at stage 1	New words added
1	38	84
2	53	24
3	21	11
4	7	4

4.3 Identification of Phrases

In this context, 'phrases' are frequently occurring word strings which convey information on their own; e.g *the nice elephant, under the table* etc. The procedure consists in selecting word strings from frequently occurring bigrams and trigrams of CE's on the basis of meaningfulness. For example, we choose the italicized strings as phrases from the following frequently occurring n-grams.

TABLE 3
A PHRASE SET

elephant for
show for
will show
elephant is
not show
for elephant
will elephant
don't show
elephant not

The phrases so identified are used to parse the corpus. 'Unparsed' parts of the corpus (sub-strings of the sentences that are not covered by the phrase set used) are then processed using the same procedure iteratively to add more phrases to the currently existing phrase set, as necessary. Two iterations are adequate to parse practically all the sentences of the corpus. These phrases can be grouped into a small number of 'phrase types', such that phrases in each group are interchangeable within a sentence frame.

4.4 Automatic Generation of STN

Sentences in the corpus can be seen as ordered sequences of phrase types and hence, phrases. It is possible to construct a First Order Markovian State Transition Network which generates sentences in the corpus. The phrases in this net would form a partially ordered set. This ordering is determined on the basis of their relative positions in the sentences of the corpus. Loops can occur if the phrase ordering reverses ($p_i < p_j$ and $p_j < p_i$ i.e. p_i follows p_j and p_j follows p_i) either in the same sentence or in different sentences. In such cases, the same phrase is given different tags to avoid looping; e.g. the phrase sequence $p_i\ p_j\ p_i\ p_k\ p_i$ would generate loops unless each of the three p_i's are tagged differently, say as p_{i1}, p_{i2} and p_{i3}. This is accomplished as follows. Phrases (say $p_1, p_2, p_3, \ldots p_n$) are arranged in an *n* by *n* matrix. An element $c_{i,j}$ of the matrix would be 1 if $p_i<p_j$ in the phrase ordering where i and j range over 1 to n. Such a matrix would be a triangular matrix if the set is fully ordered and if the ordering of rows and columns in the matrix is the same as that of the phrases. The diagonal elements would all be zero because the condition $p_i<p_j$ can not be satisfied when i=j. The number of non-zero elements in each row determines the rank order of that phrase in the set.

In our case, the set would be partially ordered; there would be more than one row containing the same number of non-zero elements. A non-zero diagonal element indicates that the corresponding phrase precedes as well as follows some other phrase or phrases in the sentences in the corpus. This phrase is (notionally) split and treated as two or more distinct phrases by appropriate tagging. Each such splitting would change the matrix size and require repetition of the loop checking process.

The partial ordering partitions the phrases into a number of rank-ordered groups, where the phrases in each group are unordered. The STN can then be generated (see *Fig. 2*) in a straightforward manner, by treating the phrases as states and connecting these to allow for transitions between the states to account for sentences in the corpus.

5. EXPERIMENTAL RESULTS

All the steps represented by boxes in *Fig.1* have been tested out on the corpus. The results obtained show that each of these steps is effective in achieving the aim that is set out for them, as indicated in Tables 1, 2, and 3. The corpus of 723 sentences, reduces to 385 distinct sentence types after reduction (i.e. substitution of individual words by the respective class exemplars). After deletion of negative and vocative words, identifying strings of CE's as phrases and finally, substituting phrases by phrase type representatives, we are left with 161 distinct phrase type sequence patterns (i.e. sentence types). Of these, just 70 sentence types have a frequency more than one and account for as many as 632 sentences in the original corpus. The STN obtained is shown in *Fig. 2*. For simplicity, the version illustrated here covers only those sentence types which occur with frequencies greater than 5 (The maximum frequency is 71). The thick lines indicate transitions corresponding to sentence types with frequencies equal to or greater than 10 and single lines indicate transitions in the frequency range 5 to 10.

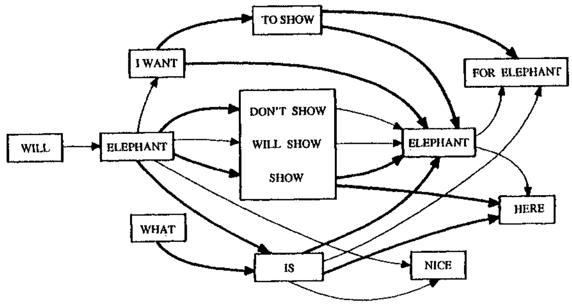

Fig. 2 - STN for the (reduced) corpus.

We have demonstrated the feasibility of adopting the Blank Slate Language Processor approach for analyzing and modeling a grammar for the spoken language of a two-and-a-half-year old child. While there is scope for further refinement and tuning of the techniques, the results indicated in this paper are promising enough to continue research in this direction.

6. ACKNOWLEDGEMENT

We are grateful to S. Deshpande, Anilkumar, U. Bhosale, and Ms. A. Kundu student trainees, for their help in implementing and testing the system, and to the Department of Electronics, Government of India and the United Nations Development Program for their financial support.

7. REFERENCES

[1] Rao P.V.S. and Nandini Bondale, "Blank Slate Language Processor for Speech Recognition", *Proc. ICSLP-92,* **Vol 1**, pp. 329-332, 1992.

[2] Rao P.V.S., "Voice: An integrated speech recognition system for the Hindi language", *Speech Commn.*, **13**, pp. 197-205, 1993.

[3] Klein W. and Dittmar N., "Developing grammars", Springer-Verlag, 1979.

[4] Menyuk P., "Sentences children use", MIT Press, 1969.

[5] Fletcher P. and Garman M. (ed), "Language acquisition", Cambridge Univ. Press, 1986.