STOCHASTIC CONTEXT-FREE GRAMMARS FOR ISLAND-DRIVEN PROBABILISTIC PARSING

Anna Corazza(*), Renato De Mori(**), Roberto Gretter(*), Giorgio Satta(*)

(*) Istituto per la Ricerca Scientifica e Tecnologica,
38050 Povo di Trento (Italy)
(**) School of Computer Science, McGill University,
3480 University str, Montreal, Quebec, Canada, H3A2A7

ABSTRACT

In automatic speech recognition the use of language models improves performance. Stochastic language models fit rather well the uncertainty created by the acoustic pattern matching. These models are used to score *theories* corresponding to partial interpretations of sentences. Algorithms have been developed to compute probabilities for theories that grow in a strictly left-to-right fashion. In this paper we consider new relations to compute probabilities of partial interpretations of sentences. We introduce theories containing a gap corresponding to an uninterpreted signal segment. Algorithms can be easily obtained from these relations. Computational complexity of these algorithms is also derived.

1 INTRODUCTION

The aim of Automatic Speech Understanding (ASU) is to process an uttered sentence, determining an optimal word sequence along with its interpretation. The success of such a process depends on the formal system we use to model natural language. There is strong evidence that stochastic regular grammars (for example Markov Models) do not capture the large-scale structure of natural language. In very recent years, there has been a growing interest toward more powerful stochastic rewriting systems, like stochastic context-free grammars (SCFG's; see among the others [Wright and Wrigley 89], [Lari and Young 90], [Jelinek *et al.* 90] and [Jelinek and Lafferty 90]). Stochastic grammars fit naturally the uncertainty created by the (pattern matching) acoustic search process; moreover SCFG's give syntactic prediction capabilities that are stronger than the Markov Models. Further motivations for this approach are reported in [Lari and Young 90] and [Jelinek *et al.* 90].

In ASU we are interested in generating partial interpretations of a spoken sentence called *theories*. We score them in terms of their likelihood $L(A, th) = O(\Pr(A \mid th)\Pr(th))$,[1] where $\Pr(A \mid th)$ is the probability that theory th derives the acoustic signal segment A and $\Pr(th)$ is the probability of the obtained theory. The most popular parsers used in Automatic Speech Recognition (ASR) generate and expand theories starting from the left and then proceeding rightward. In this case, the best theories already obtained can drive the analysis of the right portion of the input, restricting the class of possible next preterminals in order to maximize the probabilities of the new extended theories. For ASU, especially for dialogue systems, it may be useful to consider parsers that are "island-driven". These parsers focus on *islands*, that is words of particular semantic relevance which have been previously hypothesized with high acoustic evidence. Then they proceed outward, working in both directions. Island-driven approaches have been proposed and defended in [Woods 81] and [Giachin and Rullent 89]; in [Stock *et al.* 89] the predictive power of bidirectional parsing is also discussed. None of the parsers proposed in these works uses a stochastic grammar.

In this paper we consider the problem of scoring partial theories in the island-driven approach. An important quantity is $\Pr(th)$, i.e. the probability that a SCFG generates sequences of words

[1] We write $f(x) = O(g(x))$ whenever there exist constants c, $\overline{x} > 0$ such that $f(x) > c\ g(x)$ for every $x > \overline{x}$.

(islands) separated by *gaps*. The gaps are portions of the acoustic signal that are still uninterpreted in the context of *th*. We develop a theoretical framework to compute $\Pr(th)$ in the case th contains islands and gaps.

2 NOTATION AND DEFINITIONS

In this section definitions related to Stochastic Context Free Grammars (SCFGs) are introduced, along with the notation that will be used throughout this paper.

An SCFG is defined as a quadruple $G_s = (\mathrm{N}, \Sigma, P, S)$, where N is a finite set of *nonterminal* symbols, Σ is a finite set of *terminal* symbols disjoint from N, P is a finite set of *productions* of the form $H \rightarrow \alpha$, $H \in \mathrm{N}$, $\alpha \in (\Sigma \cup \mathrm{N})^*$, and $S \in \mathrm{N}$ is a special symbol called *start symbol*. Each production is associated with a probability, indicated with $\Pr(H \rightarrow \alpha)$. The grammar G_s is *proper* if the following relation holds:

$$\sum_{\alpha \in (\Sigma \cup N)^*} \Pr(H \rightarrow \alpha) = 1, \quad H \in \mathrm{N}. \tag{1}$$

An SCFG G_s is in *Chomsky Normal Form* (CNF) if all productions in G_s are in one of the following forms:

$$H \rightarrow FG \quad H \rightarrow w, \quad H, F, G \in \mathrm{N}, w \in \Sigma. \tag{2}$$

For reasons discussed in [Jelinek *et al.* 90] it is useful to have the SCFG in CNF; in the following we will always refer to SCFGs in CNF.

The derivation of a string by the grammar G_s is usually represented as a parse (or derivation) tree, whose nodes indicate the productions employed in the derivation itself. It is also possible to associate with each derivation tree the probability that it was generated by the grammar G_s. This probability is the product of the probabilities of all the rules employed in the derivation.

Given a string $z \in \Sigma^*$, the notation $H < z >$, $H \in \mathrm{N}$, indicates the set of all trees with root H generated by G_s and spanning z. Therefore $\Pr(H < z >)$ is the sum of the probabilities of these subtrees, i.e. the probability that the string z has been generated by G_s starting from symbol H. We assume that the grammar G_s is *consistent* [Gonzales and Thomason 78]. This means that the following condition holds:[2]

$$\sum_{z \in \Sigma^*} \Pr(S < z >) = 1. \tag{3}$$

From this hypothesis it follows that a similar condition holds for all nonterminals.

A possible application of an island driven parser to a task of ASU is the following. On the basis of a previously obtained theory (partial interpretation) $u = w_i \dots w_{i+p}$ and of some non-syntactic knowledge, predictions can be made for words not necessarily adjacent to u. This introduces a gap within the theory that represents a not yet recognized part of the input sentence. Then further syntactical and acoustical analyses will try to fill in the gap. The gap will be then filled by further syntactical and acoustical analysis. Therefore we will deal with theories that can be represented as follows:

$$\begin{array}{ll} th: & w_i \dots w_{i+p} x_1 \dots x_m w_j \dots w_{j+q} y_1 \dots y_k \dots \\ \text{or} & u x^{(m)} v y^{(*)} \end{array} \tag{4}$$

where $w_i \dots w_{i+p} = u$ and $w_j \dots w_{j+q} = v$ indicate strings of already recognized terminals ($i, j > 0, p, q \geq 0, j > i+p$) while $x_1 \dots x_m = x^{(m)}$, $m \geq 0$ and $y_1 \dots y_k \dots = y^{(*)}$ stand for gaps with specified length m ($x^{(m)}$) or (finite) unspecified length ($x^{(*)}$). We will also indicate a gap with x meaning that either $x = x^{(m)}$ or $x = x^{(*)}$. In our notation, i and j are position indices, p and q are shift indices, m indicates a (known) gap length and k, h are used as running indices. Finally, Σ^* represents the set of all strings of finite length over Σ, while $\Sigma^m, m \geq 0$ is the set of all strings in Σ^* of length m.

[2] The normalization property expressed in (1) above guarantees that the probabilities of all (finite and infinite) derivations sum to one, but the language generated by the grammar only corresponds to the subset of the finite derivations, whose probability can be less than one.

We studied both the cases in which gap x has specified or unspecified length (see [Corazza *et al.* 90]). In practical cases, it is possible to estimate from the acoustic signal the probability distribution of the number of words filling the gap. Since this makes more significant the case in which the gap length is specified, in this work we will focus our attention on theories of the form $x = ux^{(m)}vy^{(*)}$.

3 PARTIAL DERIVATION TREE PROBABILITIES

For the calculation of the probability $\Pr(S < uxvy^{(*)} >)$, called *prefix-string-with-gap probability*, we use some quantities already introduced by other authors, like the *inside probability* $\Pr(H < u >)$ [Baker 79], [Lari and Young 90], [Jelinek *et al.* 90] or the *prefix-string probability* $\Pr(H < ux >)$ [Jelinek and Lafferty 90]. In [Jelinek and Lafferty 90] an algorithm is proposed for the computation of the latter probability in the case of unspecified gap length ($\Pr(H < ux^{(*)} >)$). We sketch here a similar algorithm for the cases in which the gap length equals m.

3.1 Prefix-string and Suffix-string probabilities

In the case of a known length gap $x^{(m)}$, a prefix-string probability $\Pr(H < ux^{(m)} >)$ can be computed on the basis of the following relation. Since G_s is in Chomsky Normal Form, if $|ux^{(m)}| > 1$ then H must directly derive two nonterminals G_1 and G_2. According to the way the string $ux^{(m)}$ can be divided into two parts spanned by G_1 and G_2 respectively, one can distinguish two different situations: in the first one, G_1 spans just a proper prefix of u and G_2 spans the remaining part of u and the gap; in the second one, G_1 entirely spans u plus a possible prefix of the gap. Based on these cases, the following relation can be established:

$$\Pr(H < ux^{(m)} >) = \sum_{G_1 G_2} \Pr(H \rightarrow G_1 G_2)[$$
$$\sum_{k=0}^{p-1} \Pr(G_1 < w_i \dots w_{i+k} >) \times$$
$$\times Pr(G_2 < w_{i+k+1} \dots w_{i+p} x^{(m)} >) +$$
$$+ \sum_{k=0}^{m-1} \Pr(G_1 < ux_1^{(k)} >) \Pr(G_2 < x_2^{(m-k)} >)]. \quad (5)$$

Note that gap $x^{(m)}$ has been split into two shorter gaps $x_1^{(k)}$ and $x_2^{(m-k)}$. By a recursive application of (5), prefix-string probabilities can be computed using both the following initial condition:[3]

$$\Pr(H < w_i x^{(0)} >) = \Pr(H \rightarrow w_i) \quad (6)$$

and the *gap probabilities* $\Pr(H < x^{(m)} >)$, which are the sum of the probabilities of all trees with root H and yield of length m. Gap probabilities can be recursively computed as follows:

$$\Pr(H < x^{(m)} >) = \sum_{G_1, G_2 \in N} \Pr(H \rightarrow G_1 G_2) \times$$
$$\times \sum_{j=1}^{m-1} \Pr(G_1 < x^{(j)} >) \Pr(G_2 < x^{(m-j)} >), \; m > 1. \quad (7)$$

$$\Pr(H < x^{(1)} >) = \sum_{w \in \Sigma} \Pr(H \rightarrow w); \quad (8)$$

In a similar way we can define $\Pr(< xv >)$ as the *suffix-string probability*; its computation can be easily obtained from expressions that are symmetrical with respect to the ones employed for the prefix-string probability. Details are not pursued here.

We introduce now two probabilities that will be useful in calculating the prefix-string-with-gap probability: the *gap-in-string probability* $\Pr(H < uxv >)$ and the *island probability* $\Pr(H < xvy^{(*)} >)$.

3.2 Gap-in-string probabilities

For the gap-in-string probability computation we can distinguish three independent and mutually

[3] By convention, $x^{(0)}$ is the null string ε, i.e. the string whose length is zero.

exclusive cases, according to the position of the boundary between the two parts of string uxv spanned by the two children G_1 and G_2 of H. The first word of the string spanned by G_2 can belong to the initial string $u = w_i \dots w_{i+p}$, to the gap x or to the final string $v = w_j \dots w_{j+q}$.

In the case of known length gap one gets:

$$
\begin{aligned}
&\Pr(H < w_i \dots w_{i+p} x^{(m)} w_j \dots w_{j+q} >) = \\
&= \sum_{G_1 G_2} \Pr(H \rightarrow G_1 G_2)[\\
&\qquad \sum_{k=0}^{p-1} \Pr(G_1 < w_i \dots w_{i+k} >) \times \\
&\qquad\qquad \times \Pr(G_2 < w_{i+k+1} \dots w_{i+p} x^{(m)} v >) + \\
&+ \sum_{k=0}^{m} \Pr(G_1 < u x_1^{(k)} >) \Pr(G_2 < x_2^{(m-k)} v >) + \\
&+ \sum_{k=0}^{q-1} \Pr(G_1 < u x^{(m)} w_j \dots w_{j+k} >) \times \\
&\qquad\qquad \times \Pr(G_2 < w_{j+k+1} \dots w_{j+q} >)]. \qquad (9)
\end{aligned}
$$

The inner summations in (9) contain products of already defined probabilities, along with terms that can be computed recursively with the following initial condition ($p = q = 0$):

$$
\begin{aligned}
&\Pr(H < w_i x^{(m)} w_j >) = \sum_{G_1, G_2} \Pr(H \rightarrow G_1 G_2) \times \\
&\times \sum_{k=0}^{m} \Pr(G_1 < w_i x_1^{(k)} >) \Pr(G_2 < x_2^{(m-k)} w_j >) \qquad (10)
\end{aligned}
$$

3.3 Island probabilities

As for the gap-in-string case, the island probability computation involves three cases, depending on the position of the first word of the string spanned by G_2 with respect to the island $v = w_j \dots w_{j+q}$. The three sets of strings generated in the three cases above are probabilistically independent, but not disjoint in the case of unspecified length gap. Due to this fact, in such a case one must also consider the probability products, then obtaining a quadratic system of equations. On the other hand, the following relation is obtained for the case of m-length gap:

$$
\begin{aligned}
&\Pr(H < x^{(m)} w_j \dots w_{j+q} y^{(*)} >) = \sum_{G_1, G_2} \Pr(H \rightarrow G_1 G_2)[\\
&\qquad \sum_{k=1}^{m} \Pr(G_1 < x_1^{(k)} >) \times \\
&\qquad\qquad \times \Pr(G_2 < x_2^{(m-k)} w_j \dots w_{j+q} y^{(*)} >) + \\
&\qquad + \sum_{k=0}^{q-1} \Pr(G_1 < x^{(m)} w_j \dots w_{j+k} >) \times \\
&\qquad\qquad \times \Pr(G_2 < w_{j+k+1} \dots w_{j+q} y^{(*)} >) + \\
&\qquad + \Pr(G_1 < x^{(m)} w_j \dots w_{j+q} y_1^{(*)} >) \times \\
&\qquad\qquad \times \Pr(G_2 < y_2^{(*)} >)] \qquad (11)
\end{aligned}
$$

where the term $\Pr(G_2 < y_2^{(*)} >)$ equals 1. Using the definition of $Q_L(H \Rightarrow G_1\ G_2)$ given in [Jelinek and Lafferty 90] one can solve the recursion in (11) in the same way the recursive equation for the prefix-string probability is solved there, obtaining:

$$
\begin{aligned}
&\Pr(H < x^{(m)} w_j \dots w_{j+q} y^{(*)} >) = \\
&= \sum_{G_1, G_2} Q_L(H \Rightarrow G_1 G_2) C_{xvy}(G_1, G_2) \qquad (12)
\end{aligned}
$$

in which:

$$
\begin{aligned}
&C_{xvy}(G_1, G_2) = \\
&= \sum_{k=1}^{m} \Pr(G_1 < x_1^{(k)} >) \times \\
&\qquad \times \Pr(G_2 < x_2^{(m-k)} w_j \dots w_{j+q} y^{(*)} >) + \\
&+ \sum_{k=0}^{q-1} \Pr(G_1 < x^{(m)} w_j \dots w_{j+k} >) \times \\
&\qquad \times \Pr(G_2 < w_{j+k+1} \dots w_{j+q} y^{(*)} >). \qquad (13)
\end{aligned}
$$

The term $C_{xvy}(G_1, G_2)$ contains a summation of products between gap probabilities and island probabilities over a left gap shorter than x, along with a summation of products between suffix-string probabilities (with known length gap) and prefix-string probabilities (with unspecified length gap). Equation (13) can be solved recursively, with the initial condition ($x^{(0)} = \varepsilon$):

$$C_{vy}(G_1, G_2) = \sum_{k=0}^{q-1} \Pr(G_1 < w_j \dots w_{j+k} >) \times$$
$$\times \Pr(G_2 < w_{j+k+1} \dots w_{j+q} y^{(*)} >). \qquad (14)$$

3.4 Prefix-string-with-gap probabilities

An expression for the prefix-string-with-gap probability $\Pr(H < ux^{(m)}vy^{(*)} >)$ can now be obtained directly from the four cases where the boundary between the two children of H belongs to u, to the gap x, to the island v or to the final gap y:

$$\Pr(H < w_i \dots w_{i+p} x^{(m)} w_j \dots w_{j+q} y^{(*)} >) =$$
$$= \sum_{G_1, G_2} \Pr(H \rightarrow G_1 G_2)[$$
$$\sum_{k=0}^{p-1} \Pr(G_1 < w_i \dots w_{i+k} >) \times$$
$$\times \Pr(G_2 < w_{i+k+1} \dots w_{i+p} x^{(m)} v y^{(*)} >) +$$
$$+ \sum_{k=0}^{m} \Pr(G_1 < u x_1^{(k)} >) \Pr(G_2 < x_2^{(m-k)} v y^{(*)} >) +$$
$$+ \sum_{k=0}^{q-1} \Pr(G_1 < u x^{(m)} w_j \dots w_{j+k} >) \times$$
$$\times \Pr(G_2 < w_{j+k+1} \dots w_{j+q} y^{(*)} >) +$$
$$+ \Pr(G_1 < u x^{(m)} v y_1^{(*)} >) \Pr(G_2 < y_2^{(*)} >)]. \qquad (15)$$

Solving the recursion in (15) in the same way as for (11), one obtains:

$$\Pr(H < w_i \dots w_{i+p} x^{(m)} w_j \dots w_{j+q} y^{(*)} >) =$$
$$= \sum_{G_1, G_2} Q_L(H \Rightarrow G_1 G_2) D_{uxvy}(G_1, G_2) \qquad (16)$$

where:

$$D_{uxvy}(G_1, G_2) =$$
$$\sum_{k=0}^{p-1} \Pr(G_1 < w_i \dots w_{i+k} >) \times$$
$$\times \Pr(G_2 < w_{i+k+1} \dots w_{i+p} x^{(m)} v y^{(*)} >) +$$
$$+ \sum_{k=0}^{m} \Pr(G_1 < u x_1^{(k)} >) \Pr(G_2 < x_2^{(m-k)} v y^{(*)} >) +$$
$$+ \sum_{k=0}^{q-1} \Pr(G_1 < u x^{(m)} w_j \dots w_{j+k} >) \times$$
$$\times \Pr(G_2 < w_{j+k+1} \dots w_{j+q} y^{(*)} >). \qquad (17)$$

As for previous computations in this section, equation (17) consists of summations over products of already defined probabilities along with a recursive term $\Pr(G_2 < w_{i+k+1} \dots w_{i+p} x^{(m)} v y^{(*)} >)$ which can be computed starting with the initial condition $(p = 0)$:

$$D_{w_i xvy}(G_1, G_2) =$$
$$= \sum_{k=0}^{m} \Pr(G_1 < w_i x_1^{(k)} >) \Pr(G_2 < x_2^{(m-k)} v y^{(*)} >) +$$
$$+ \sum_{k=0}^{q-1} \Pr(G_1 < w_i x^{(m)} w_j \dots w_{j+k} >) \times$$
$$\times \Pr(G_2 < w_{j+k+1} \dots w_{j+q} y^{(*)} >). \qquad (18)$$

4 COMPLEXITY EVALUATION

Based on the relation presented in the last section, algorithms for the computation of the probabilities defined there can be developed strightforwardly. In the present section we discuss the computational complexity for the cases of major interest (details about the derivation of the complexity expressions are simple but tedious, and therefore will not be reported here). The assumed model of computation is the *Random Access Machine*, taken under the *uniform cost criterion* (see [Aho *et al.* 74]). We are mainly concerned here with worst-case time complexity results.

We will indicate with $|P|$ the size of set P, i.e. the number of productions in G_s. All the probabilities defined in Section 3 depend upon the grammar G_s, strings u and v and the lengths of gaps x and y. Table 1 summarizes worst-case time complexity for sets of these probabilities.

computed set	*time complexity*
island probabilities	
1. $\{\Pr(H < x^{(m)} w_j \dots w_{j+q} y^{(*)} >) \mid H \in \mathrm{N}\}$	$O(\lvert P\rvert \max\{q^3, m^2 q\})$
prefix-string-with-gap probabilities	
2. $\{\Pr(H < w_i \dots w_{i+p} x^{(m)} w_j \dots w_{j+q} y^{(*)} >) \mid H \in \mathrm{N}\}$	$O(\lvert P\rvert \max\{p^3, q^3, pm^2, qm^2\})$
one word extension for island probabilities	
3. $\{\Pr(H < x^{(m)} w_j \dots w_{j+q} a y^{(*)} >) \mid H \in \mathrm{N}\}$	$O(\lvert P\rvert \max\{q^2, m^2\})$
4. $\{\Pr(H < x^{(m-1)} a w_j \dots w_{j+q} y^{(*)} >) \mid H \in \mathrm{N}\}$	$O(\lvert P\rvert \max\{m^2 q, mq^2\})$
one word extension for prefix-string-with-gap probabilities	
5. $\{\Pr(H < w_i \dots w_{i+p} x^{(m)} w_j \dots w_{j+q} a y^{(*)} >) \mid H \in \mathrm{N}\}$	$O(\lvert P\rvert \max\{p^2, q^2, m^2, (m+q)p\})$
6. $\{\Pr(H < w_i \dots w_{i+p} x^{(m-1)} a w_j \dots w_{j+q} y^{(*)} >) \mid H \in \mathrm{N}\}$	$O(\lvert P\rvert \max\{p^2 q, pq^2, p^2 m, pm^2\})$

Table 1: *Worst-case time complexity for the computation of the probabilities of some sets of theories. Symbol $a \in \Sigma$ indicates a one word extension of a theory whose probability had already been computed.*

Both island and prefix-string-with-gap probabilities require cubic time computations (rows 1 and 2). Rows 3 to 6 account for cases in which one have to compute the probability of a theory that has been obtained from a previously analyzed theory by means of a single word extension. In these cases, using a dynamic technique, one can dispense from the computation of elements already involved in the calculation of the previous theory. One word extension on the side of the unknown length gap $y^{(*)}$ costs quadratic time both in the case of island and prefix-string-with-gap probabilities. The one word extension on the side of the known length gap $x^{(m)}$ costs cubic time. This asymmetry can be justified observing from (15) that the addition of a single word between a string and a bounded gap forces the reanalysis of a quadratic number of new subterms. Note that this is also true for well known dynamic methods for CFG recognition (e.g. the CYK algorithm [Younger 67]): one word change in the *middle part* of a string implies a cubic-time whole recomputation in the worst-case. In fact there is an interesting parallelism between those methods, the Inside algorithm and the methods discussed here (see [Corazza *et al.* 90] for a discussion).

5 DISCUSSION

A framework has been developed to score partial sentence interpretations in ASU systems. General motivations for modeling naturall anguage by SCFG's can be found in [Jelinek *et al.* 90], while the importance of scoring measures that are compatible with island-driven strategies has been already pointed out in [Woods 81]. In the present section we discuss major advantages of the studied approach and possible applications of the derived framework.

We are mainly interested in sentence interpretation systems. Even if semantical and pragmatical predictive models are not defined, we can rely on high-level heuristic information sources. This knowledge can be used to predict words on the base of previous partial interpretations. Predictions may be words not adjacent to the stimulating segments. These words can be recovered using word-spotting techniques.[4] Thus, the only way to employ the available heuristic information is to parse sentences in a discontinuous way. This means that the parser has first to find an island and then to fill the gap between the stimulating segment and the island itself. This technique produces partial analyses that are interleaved by gaps and that can be scored using our method.

[4] Word-spotting techniques allow one to find occurences of one (or more) given word in a speech signal. In these systems there is a trade off between "false alarms" and "missing words" that can be controlled by a threshold obtained from training speech.

The framework introduced in this paper can also be used to predict words adjacent to an already recognized string and to compute the probability that the first (last) word x_1 (x_m) of a gap is a certain symbol $a \in \Sigma$. This new word will extend the current theory. Words adjacent to an existing theory can be hypothesized by selecting the word(s) which maximize the prefix-string-with-gap probability of the theory augmented with it. Instead of computing these probabilities for all the elements in the dictionary, it is possible to restrict this expensive process to the preterminal symbols (as in [Jelinek and Lafferty 90]). The approach discussed so far should be compared with standard lattice parsing techniques, where no restriction is imposed by the parser on the word search space (see, for example [Chow and Roukos 89] and the discussion in [Moore *et al.* 89]).

Our framework accounts for bidirectional expansion of partial analyses; this improves the predictive capabilities of the system. In fact, bidirectional strategies can be used in restricting the syntactic search space for gaps surrounded by two partial analyses. This point has been discussed in [Stock *et al.* 89] for cases of one word length gaps. We propose a generalization to m-length gaps and to cases where partial analyses do not represent only complete parse trees but also partial derivation trees.

As a final remark, notice that the proposed framework requests the SCFG to be in Chomsky normal form. Although every SCFG G_s can be cast in CNF, such a process may result in quadratic size expansion of G_s, where the size of G_s is roughly proportional to the sum of the length of all productions in G_s. The proposed framework can be easily generalized to other kinds of bilinear forms with linear expansion in the size of G_s (for example the *canonical two form* [Harrison 78]). This consideration deserves particular attention because in natural language applications the size of the grammar is considerably larger than the input sentence length.

References

[Aho *et al.* 74] A.V.Aho, J.E.Hopcroft and J.D.Ullman: "The Design Analysis of Computer Algorithms", Addison-Wesley Publishing Company, Reading, MA, 1974.

[Baker 79] J.K.Baker: "Trainable Grammars for Speech Recognition", Proceedings of the Spring Conference of the Acoustical Society of America, 1979.

[Chow and Roukos 89] Y.L.Chow and S.Roukos: "Speech Understanding Using a Unification Grammar", Proceedings of the IEEE International Conference on Acoustic, Speech and Signal Processing, 1989, Glasgow, Scotland.

[Corazza *et al.* 90] A.Corazza, R.DeMori, R.Gretter and G.Satta: "Computation of Probabilities for an Island-Driven Parser" Technical Report SOCS 90-19, Mc Gill University, MONTREAL, Quebec, H3A 2A7 CANADA. Also as Technical Report TR9009-01, IRST, Trento, Italy, 1990.

[Giachin and Rullent 89] E.P.Giachin and C.Rullent: "A Parallel Parser for Spoken Natural Language", Proceedings of the Eleventh International Joint Conference on Artificial Intelligence, 1989, Detroit, Michigan USA, pp.1537-1542.

[Gonzales and Thomason 78] R.C.Gonzales and M.G.Thomason: "Syntactic Pattern Recognition", Addison-Wesley Publishing Company, Reading, MA, 1978.

[Harrison 78] M.A.Harrison: "Introduction to Formal Language Theory", Addison-Wesley Publishing Company, Reading, MA, 1978.

[Jelinek *et al.* 90] F.Jelinek J.D.Lafferty and R.L.Mercer: "Basic Method of Probabilistic Context Free Grammars", Internal Report, T.J.Watson Research Center, Yorktown Heights, NY 10598, 85 pages.

[Jelinek and Lafferty 90] F.Jelinek and J.D.Lafferty: "Computation of the Probability of Initial Substring Generation by Stochastic Context Free Grammars", Internal Report, Continuous Speech Recognition Group, IBM Research, T.J.Watson Research Center, Yorktown Heights, NY 10598, 10 pages.

[Lari and Young 90] K.Lari and S.J.Young: "The Estimation of Stochastic Context-Free Grammars using the Inside-Outside Algorithm" Computer Speech and Language, vol.4, n.1, 1990, pp.35-56

[Moore *et al.* 89] R.M.Moore F.Pereira and H.Murveit: "Integrating Speech and Natural Language Processing", Proceedings of the Speech and Natural Language Workshop, 1989, Philadelphia, Pennsylvania, pp.243-247.

[Stock *et al.* 89] O.Stock R.Falcone and P.Insinnamo: "Bidirectional Chart: A Potential Technique for Parsing Spoken Natural Language Sentences" Computer Speech and Language, vol.3, n.3, 1989, pp.219-237.

[Woods 81] W.A.Woods: "Optimal Search Strategies for Speech Understanding Control" Artificial Intelligence, vol.18, n.3, 1981, pp.295-326.

[Wright and Wrigley 89] J.H.Wright and E.N.Wrigley: "Probabilistic LR Parsing for Speech Recognition" International Workshop on Parsing Technologies, Pittsburg, PA, pp.105-114.

[Younger 67] D.H.Younger: "Recognition and Parsing of Context-Free Languages in Time n^3" Information and Control, vol. 10, 1967, pp.189-208.