# Statistical Agenda Parsing

*Robert J. Bobrow*
BBN Systems and Technologies
10 Moulton Street.
Cambridge, MA 02138

## ABSTRACT

This paper presents the results of converting a standard Graham/Harrison/Ruzzo (GHR) parser for a unification grammar into an agenda-driven parsing system. The agenda is controlled by statistical measures of grammar-rule likelihood obtained from a training set.

The techniques in the agenda parser lead to substantial reductions in chart size and parse time, and can be applied to any chart-based parsing algorithm without hand-tuning.

## INTRODUCTION

In a Graham/Harrison/Ruzzo (GHR) parser, the chart is used to maintain a record of syntactic constituents that have been found (terms) and grammatical rules that have been partially matched (dotted rules). Parsing strategies such as GHR, CKY and other algorithms can be viewed as methodical ways of filling the chart which guarantee to explore all possible extensions of dotted rules by terms.

An agenda is an alternative chart-filling algorithm with the goal of finding some term covering the entire input without necessarily filling in all of the chart. If terms can be ranked by "goodness" and the grammar can produce multiple analyses of a given string, then one goal for an agenda is to produce the "best" parse first.

The alternative goal we have chosen for DELPHI is to use the agenda mechanism to reduce the search necessary to produce ACCEPTABLE (see below) parses. This results in sparsely populated charts, approaching the extreme (and probably unattainable) goal of deterministic parsing, in which the only terms and dotted rules entered into the chart are those which appear as parts of the final parse.

The techniques involved in statistical agenda parsing allow "low probability" rules to be added to a grammar without significant cost in terms of either erroneous parses or increased parse time. These low probability rules greatly increase the coverage and robustness of the system by accounting for unusual or marginal constructions.

## DELPHI AGENDA PARSING

Most techniques for search space reduction involve careful tuning of the grammar or the parsing mechanism. This is very labor intensive and can place limits on the grammatical coverage of the system (Abney 1990). Our approach is to use an automated statistical technique for ranking rules based on their use in parsing a training set with the same grammar (under the control of an all-paths GHR parser without human supervision).

This approach also allows us to include grammatical rules that are of use only rarely, or in specialized domains, and to learn how applicable they are to a body of sentences. To take into account general linguistic tendencies, we augment the statistical ranking by a small number of general agenda ordering strategies.

The DELPHI agenda mechanism is based on three "schedulable" action types:

1. the insertion of a term into the chart,
2. the insertion of a dotted rule into the chart, and
3. the (conditional) "pair extension" of a dotted rule by a term.

In principle one would like to order those actions in terms of the probability that they lead to a final parse. The initial implementation of the agenda mechanism uses an approximation to this ordering.

## USE OF STATISTICAL MEASURES

There are two types of measures that one might estimate to help the agenda parsing mechanism. They are (1) category expansion probabilities and (2) rule success probabilities.

### Category Expansion Probabilities

Category expansion probabilities are perhaps the more obvious of the two measures. The goal is to determine the probability that a given syntactic category (e.g., NP) is expanded by a given grammar rule in a valid parse.

These probabilities allow one to estimate the probability that a given tree is the expansion of a given category. Bayes' rule may be used to calculate the relative probabilities of various parse trees for a specified input string.

### Rule Success Probabilities

Using rule success probabilities, the goal is to determine the probability that a term inserted into the chart by a particular rule will be part of a final parse.

## Training

In order to train the agenda mechanism, a set of sentences is parsed using the all-paths GHR parser and their charts are analyzed.

For each rule (R) in the grammar we determine three numbers:

1. NT(R), the number of terms in the charts based on that rule.
2. NDR(R), the number of dotted rules initiated in the chart based on that rule.
3. NGT(R), the number of "good terms" based on that rule, ones that are constituents of an ACCEPTABLE parse (i.e., ones leading to executable database commands for ATIS).

For each category C in the grammar, we calculate one number:

4. NGT(C), the number of terms with that category which are constituents in an acceptable parse.

The ratio NGT(R)/NT(R) is an estimate of the probability that a term based on R will appear in the final parse, and NGT(R)/NDR(R) is an estimate of the probability that the initiation of a dotted rule based on R will lead to a good term. (Note that in DELPHI, each word sense is treated as if it were a separate grammar rule, and so this mechanism takes into account the relative likelihood of various word senses in the training set.)

If C(R) is the category produced by the rule R, then the category expansion probability of R is NGT(R)/NGT(C(R)).

## Preliminary Results for Different Measures

Using rule success probabilities leads to substantial reduction (a factor of more than 3) in chart size. In general, one might expect that better estimates of such probabilities, based on category expansion probabilities in the tree below the term, would lead to improved results, even though these estimates require somewhat more computation than rule success probabilities alone.

We have compared the use of category expansion probabilities with the use of rule success probabilities in several variations of the agenda mechanism, and have found that rule success probabilities produce superior results, although the reasons for this are not entirely clear.

An experiment using category expansion probabilities alone led to larger charts than produced by the use of rule success probabilities in isolation. Combining category expansion probabilities with rule success probabilities appeared to be no better than just using the rule success probabilities.

## AGENDA STRUCTURES

The structure of the agenda mechanism appears to be as important as the statistical measures used to order agenda items. Experience with probabilistic agendas in speech processing would suggest an approach in which all information relevant to ordering is combined into a single numeric measure and used to order a single queue. In principal, this allows different measures to interact and for strength in one measure to make up for weakness in another.

We experimented with this approach in a system which had a single agenda in which all three of the schedulable action types described above were placed. The statistical measures described above were combined in a weighted fashion with priorities based on the size of the constituents, the position of the right hand end of the constituent and the action type. A number of experiments were run, giving different weightings to the different parameters, but all of these experiments led to charts that were 20% to 40% larger than the alternative structured agenda described below.

The structured-agenda approach involves the creation of a 2-dimensional array of agendas, as illustrated in figure 1.

| Action Type | Rightmost Endpoint | | | | |
|---|---|---|---|---|---|
| | N | N-1 | ... | 1 | 0 |
| Pairs | A1 | A4 | ... | | |
| Rules | A2 | A5 | | | |
| Terms | A3 | A6 | | | |

**Figure 1:** Agenda

Each cell of the array consists of a single type of action, e.g. term insertion, and all of the actions in the list Ai in a cell have the same rightmost end. Within the cell, the actions in the list Ai are ordered by probability estimates.

For each step, the first non-empty cell (starting with A1 and going in the order shown in figure 1) is chosen, and the first item on its agenda is run. This has the effect of reinforcing progress to the right through the input string, of choosing the most appropriate action for such motion at each step, and favoring close attachment of modifiers.

## DELPHI RESULTS

Measurements of chart-size and time reductions for BBN's DELPHI grammar running on the ATIS training and test sets indicate the improvements possible with several variations of the basic agenda mechanism. For example, using the structured agenda on 551 sentences of training data from June 1990, the chart size was reduced by a factor of 3.24, and the total processing time reduced by a factor of 1.82.

This result underestimates the improvement gained by agenda parsing, since somewhat more than 10% of the "sentences" in the training data were ill-formed according to our grammar (many were ill-formed according to any plausible grammar!). Since a properly operating agenda system will eventually produce the same chart that the GHR parser does, and since that entire chart must be searched before a string is determined to be unparseable, the performance of any agenda mechanism must reduce to that of the GHR parser for such inputs.

Another set of experiments was performed with a set of 539 "parseable" strings taken from the combination of the June 1990 and February 1991 ATIS training set. For this set the speedup was a factor of 3.8 and the chart size reduction was well over 3.5. (The hedge on chart size reduction is because data for the chart size of 5 sentences in the GHR parser was not obtained, the charts overflowed available memory . At this time the ratio of that chart size to the size of the agenda parser chart was over 30.)

The introduction of probabilistic agenda parsing, combined with the application of software engineering techniques, has sped up natural language analysis considerably. The average time for parsing, semantic interpretation and discourse processing (of a 551 sentence training corpus) in our DELPHI system was lowered to 1.43 seconds per sentence, with a median time of 0.99 seconds, on a Sun 4/280.

## REFERENCES


1. Abney, Steven P. "Rapid Incremental Parsing with Repair", *Proceedings of the Sixth Annual Conference of the UW Centre for the New Oxford English Dictionary and Text Research*,University of Waterloo, Waterloo, Ontario, Canada, October 28-30, 1990,.

2. Graham, Susan L., Michael A. Harrison, and Walter L. Ruzzo,"An Improved Context-Free Recognizer", *ACM Transactions on Programming Languages and Systems*, Vol. 2, Number 3,1980,.

3. Kasami, T. "An Efficient Recognition and Syntax Analysis Algorithm for Context-Free Languages", *Science Report AFCRL-65-758*, Air Force Cambridge Research Laboratory, 1966.

4. Bates, M., Boisen, S., Ingria, R. and Stallard, D. "BBN ATIS System Progress Report - June 1990", *Proceedings of the Speech and Natural Language Workshop* (June, 1990), Morgan Kaufmann Publishers, Inc., 1990.