

# Elements of Language Processing and Learning

Project 1: Statistical Parsing, Steps 1 & 2

25 November, 2012

By:

Georgios Methenitis, Stud.Id:10407537

Marios Tzakris, Stud.Id:10407537

## 1 Introduction

In this project we had to implement a prototype statistical parser. Having the given treebanks we had to parse and extract the rules using this statistical parser (Step 1), resulting to a probabilistic context free grammar (PCFg). Then, given this extracted grammar, we implement the Cocke-Younger-Kasami (CYK) algorithm, in order to parse some test sentences generating a parse-forest for each one of them. In Section 2, we are going to describe the implementation of our parser. In Section 3, we are going to describe the implementation of the CYK algorithm as well as, our assumptions in respect to the unknown words, and a brief explanation of every step of this algorithm through pseudo-code. Section 4, serves as an epilogue to the first two steps of this project.

## 2 Parser

Each rule is represented by the non-terminal left-hand side node, and the right-side node or nodes. a typical example of binary rule is this:  $X \rightarrow \alpha \beta$ . In general in our treebanks there were mostly binary rules but in some cases we had to handle and unary rules which have this form:  $X \rightarrow \alpha$ .

## 3 CYK Algorithm

agargbarebeatb

## 4 Conclusion

In this first two steps, we have seen....