Elements of Language Processing and Learning
Lab assignment
Stage 1: Computing the Probability of a Tree
Report

Benno Kruit, 10576223
Sara Veldhoen, 10545298

November 14, 2013

The objective is to compute the probability of a parse tree, e.g. a tree in the test set.

In general, the probability of a tree is the multiplication of the probability of the production, with the probabilities of the children. We actually compute the log probability to avoid arithmetic underflow due to the multiplication of small probabilities. Therefore, in the algorithm we perform an addition.

First step: annotate, call logScoreHelper The function logScoreHelper essentially has two cases:

- Basis: preterminal node. If the tree is a preterminal with one child, a leaf with a terminal (a word), the lexicon is called to compute the probability of this production. The probability of the children is not
- Recursive: internal node. If the tree is an inner node with trees as children, the grammar is called to compute the probability of the production. The probability of the children is computed via a recursive call to logScoreHelper. A distinction is made between internal nodes with one or two children, thus calling the grammar for unary or binary rules respectively.

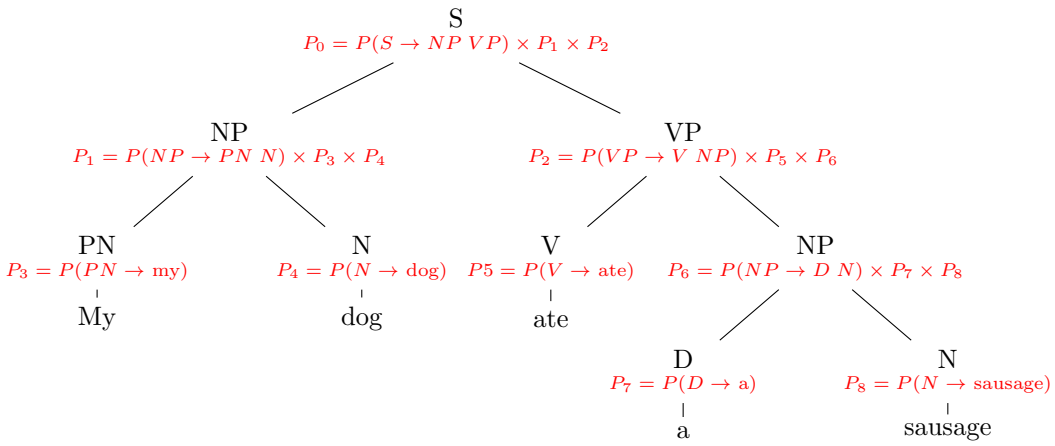

Figure 1: The probability of a tree is a multiplication of the probability of the production with the probability of its children