Traitement automatique du langage
TP 4 — PCFGs and Parsing

Asheesh Gulati
Exercises prepared by Yves Scherrer
01.11.2018

Submit by 14.11.2018 midnight.

Implement a probabilistic CKY parser

Implement the CKY parsing algorithm in Python using the pseudo-code given in the course material. Given a grammar and a sentence, the program should be able to determine the most probable syntactic analysis and to output the associated parse tree. We provide you with a toy grammar (*toygrammar.json* on Chamilo) and a few example sentences to test your algorithm.

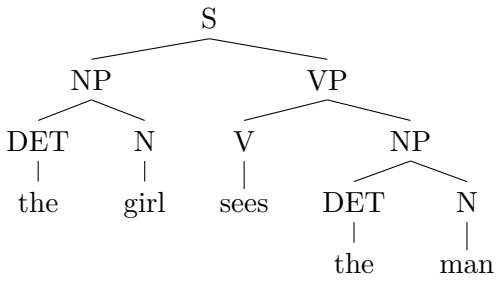
Some indications about the grammar:

- The grammar is stored as a JSON file. It can be loaded into a Python dictionary using *json.load()*.
- I contains the initial symbol, NTR contains a list of non-terminal rules, and TR contains a list of terminal rules.
- Each rule is represented as a list: the first element is the left hand side, the last element contains the rule probability, and the elements in between (2 in case of non-terminal rules, 1 in case of terminal rules) constitute the right hand side of the rule.

Some indications about the parsing algorithm:

- You can safely assume that the grammar is in Chomsky Normal Form.
- It might be useful to store the applied terminal rules as *backpointers* to recover the best parse tree.
- It might be useful to store the rules as dictionaries of dictionaries, using the left hand side as the first key, the right hand side as the second key, and the probability as the value.

- Your parser should gracefully handle sentences that cannot be parsed, e.g. because of missing rules.
- Your parser should display the parse tree and its probability. The simplest way of displaying a parse tree is using nested brackets. For example, the tree:



  can be represented as:

  (S (NP (DET the) (N girl)) (VP (V sees) (NP (DET the) (N man))))

Here are a few example sentences with their expected outputs using the toy grammar:

- *the man the girl*

  No analysis.

- *the girl sees the telescope*

  (S (NP (DET the) (N girl)) (VP (V sees) (NP (DET the) (N telescope))))0.001728

- *the girl sees the telephone*

  No analysis (the word *telephone* is not in the grammar).

- *the telescope watches the man with the girl*

  (S (NP (DET the) (N telescope)) (VP (V watches) (NPPP (NP (DET the) (N man)) (PP (P with) (NP (DET the) (N girl)))))) 0.0005376

- *the man sees the man with the telescope*

  (S (NP (DET the) (N man)) (VP (V sees) (NPPP (NP (DET the) (N man)) (PP (P with) (NP (DET the) (N telescope)))))) 0.001344