

# HW #4

# Probabilistic Parsing

- Goals:
  - Learn about PCFGs
  - Implement PCKY
  - Analyze Parsing Evaluation
  - Assess improvements to PCFG Parsing

# Tasks

## 1. Train a PCFG

1. Estimate rule probabilities from treebank
2. Treebank is already in CNF
3. More ATIS data from Penn Treebank

## 2. Build CKY Parser

1. Modify (your) existing CKY implementation

# Tasks

## 3. Evaluation

1. Evaluate your parser using standard metric
2. We will provide **evalb** program and gold standard

## 4. Improvement

1. Improve your parser in some way:
  1. Coverage
  2. Accuracy
  3. Speed
2. Evaluate new parser

# Improvement Possibilities

- Coverage:
  - Some test sentences won't parse as is!
    - Lexical gaps (aka out-of-vocabulary [OOV] tokens)
      - ...remember to model the probabilities, too
- Better context modeling
  - e.g. — Parent Annotation
- Better Efficiency
  - e.g. — Heuristic Filtering, Beam Search
- No “cheating” improvements:
  - improvement can't change training by looking at test data

# evalb

- evalb available in `dropbox/21-22/571/hw4/tools`
- `evalb [...] <gold-file> <test-file>`
- `evalb --help` for more info
- NB: specify **full/absolute path** to evalb when invoking in your scripts