UPPSALA
UNIVERSITET

Treebanks, evaluation
Discussion

Syntactic analysis/parsing

2024

Sara Stymne
Department of Linguistics and Philology

Based on slides by Marco Kuhlmann

UPPSALA
UNIVERSITET

Treebank grammars

Reading rules off the trees

Given a treebank, we can construct a grammar by reading rules off the phrase structure trees.

Sample grammar rule	Span
S → NP-SBJ VP .	Pierre Vinken … Nov. 29.
NP-SBJ → NP , ADJP ,	Pierre Vinken, 61 years old,
VP → MD VP	will join the board …
NP → DT NN	the board

Properties of treebank grammars

- Treebank grammars are typically rather flat.
  Annotators tend to avoid deeply nested structures.
- Grammar transformations.
  In order to be useful in practice, treebank grammars need to be transformed in various ways.
- Treebank grammars are large.
  The vanilla PTB grammar has 29,846 rules.

Estimating rule probabilities

- The simplest way to obtain rule probabilities is relative frequency estimation.
  - Step 1: Count the number of occurrences of each rule in the treebank.
  - Step 2: Divide this number by the total number of rule occurrences for the same left-hand side.
  - The grammar that you use in the assignment is produced in this way.

UPPSALA
UNIVERSITET

Parser evaluation

Evaluation measure

- Precision:
  Out of all brackets found by the parser,
  how many are also present in the gold standard?

- Recall:
  Out of all brackets in the gold standard,
  how many are also found by the parser?

- F1-score:
  harmonic mean between precision and recall:
  2 × precision × recall / (precision + recall)

Evaluation and transformation

- It is good practice to always re-transform the grammar if it has been transformed, for instance into CNF
- In assignment 1 you will do your evaluation on the parse trees in CNF
  - It affects the scores, so they are not comparable to scores on the original treebank
  - This is not really good practice
  - But, it simplifies the assignment!

UPPSALA
UNIVERSITET

Assignment 1

Parser and grammar

- `def CKY(pcfg, norm_words):`
  - `pcfg` is a class
  - Access its variables:
  - `pcfg.N`
  - `pcfg.q1[X,norm]`
  - …

UPPSALA
UNIVERSITET

Practicalities

UPPSALA UNIVERSITET

Literature seminar 1, Feb 7

- *Recurrent neural network grammars*, Dyer, Kuncoro, Ballesteros, and Smith
- Detailed instructions on the course web page
  - Read the article carefully
  - Work through the given questions
  - Be prepared to discuss the article and questions
- Make an effort to try to understand the paper!
  - But fine if you do not understand everything, especially the details about neural networks and maths
- **The seminar will help in understanding the paper!**

UPPSALA
UNIVERSITET

Literature seminars

- The seminar is obligatory and part of the examination
- Will be held on campus
- If you do not attend, have not prepared, or do not take part in the discussion: written report instead
- Groups and times will be available on the web page soon
- 45 minutes per group

UPPSALA UNIVERSITET

Coming sessions

- Wednesday, Jan 31, **8–10**:
  - Lecture on transition-based dependency parsing
    - Recorded videos + exercise available
- Monday Feb. 5:
  - Supervision in Chomsky
- Wednesday Feb 7:
  - Literature seminar 1: 9–12 (groups TBA)