UPPSALA
UNIVERSITET

Treebank Grammars and Parser Evaluation

Syntactic analysis/parsing

2017-11-16

Sara Stymne
Department of Linguistics and Philology

Based on slides from Marco Kuhlmann

UPPSALA
UNIVERSITET

Recap: Probabilistic parsing

UPPSALA UNIVERSITET

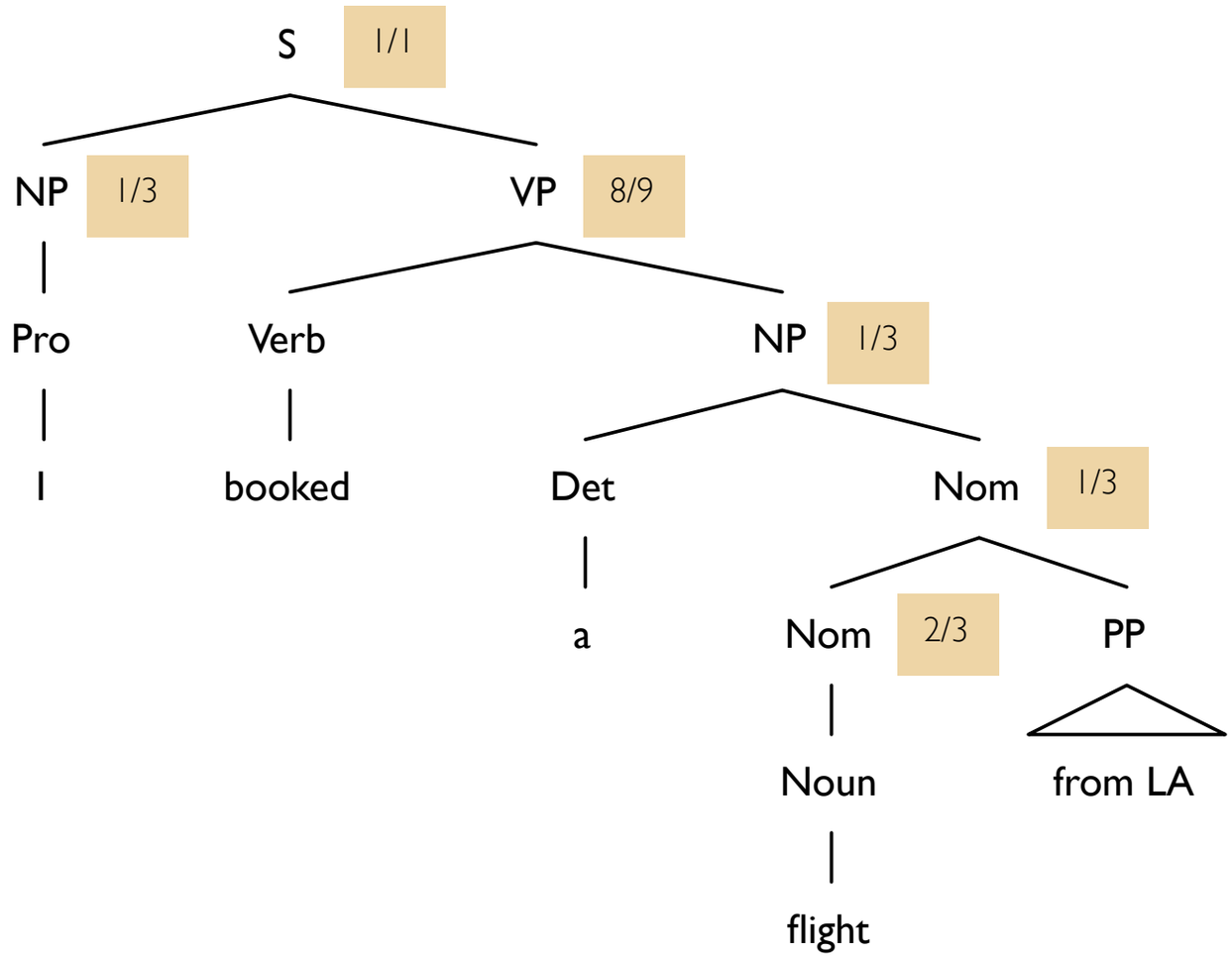
Probabilistic context-free grammars

A probabilistic context-free grammar (PCFG) is a context-free grammar where

- each rule *r* has been assigned a probability $p(r)$ between 0 and 1
- the probabilities of rules with the same left-hand side sum up to 1

UPPSALA UNIVERSITET

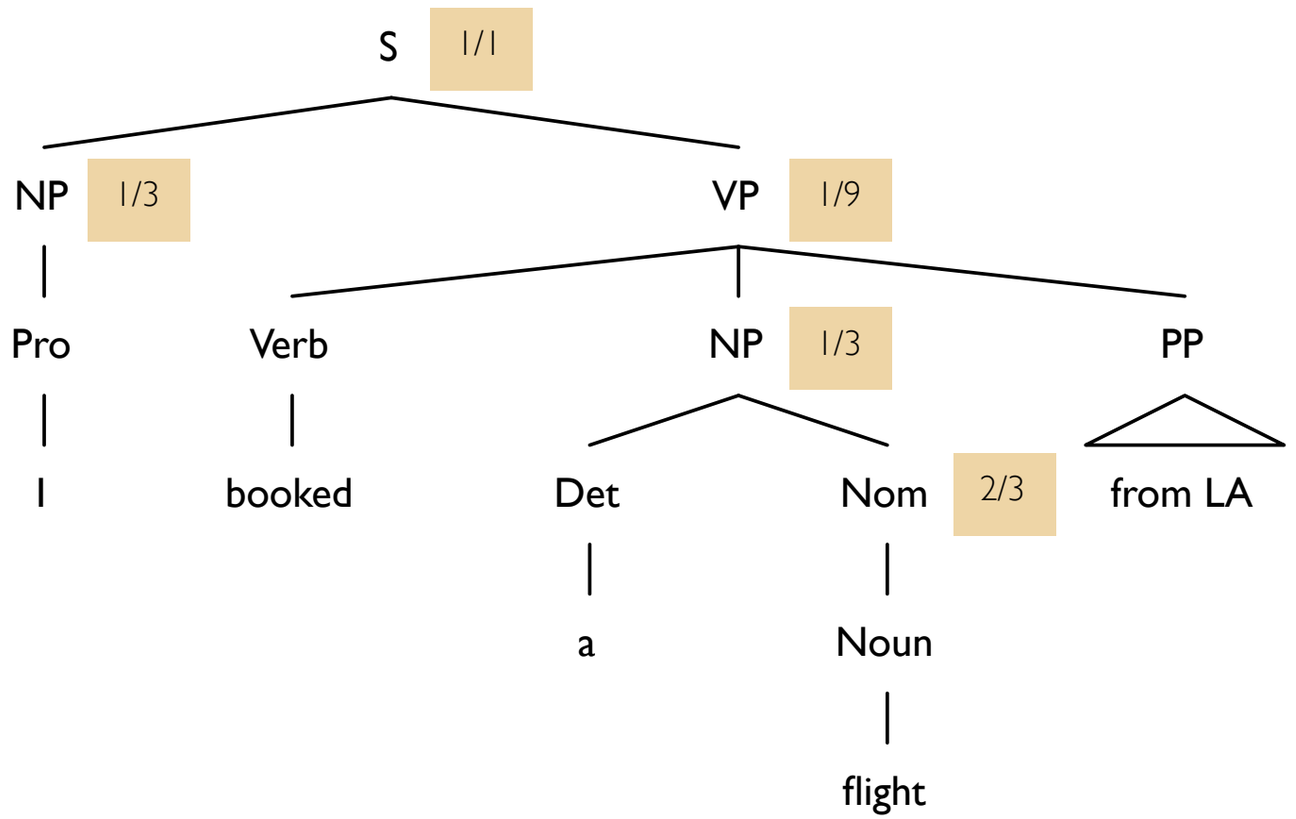
Probability of a parse tree



Probability: 16/729

UPPSALA UNIVERSITET

Probability of a parse tree



Probability: 6/729

UPPSALA UNIVERSITET

Computing the most probable tree

```
for each max from 2 to n
  for each min from max - 2 down to 0
    for each syntactic category C
      double best = undefined
      for each binary rule C -> C₁ C₂
        for each mid from min + 1 to max - 1
          double t₁ = chart[min][mid][C₁]
          double t₂ = chart[mid][max][C₂]
          double candidate = t₁ * t₂ * p(C -> C₁ C₂)
          if candidate > best then
            best = candidate
      chart[min][max][C] = best
```

UPPSALA UNIVERSITET

Backpointers

```
if candidate > best then
  best = candidate
  // We found a better tree; update the backpointer!
  backpointer = (C -> C₁ C₂, min, mid, max)
...
chart[min][max][C] = best
backpointerChart[min][max][C] = backpointer
```

UPPSALA
UNIVERSITET

Treebank grammars

Treebanks

- Treebanks are corpora in which each sentence has been annotated with a syntactic analysis.
- The annotation process requires detailed guidelines and measures for quality control.
- Producing a high-quality treebank
  is both time-consuming and expensive.

The Penn Treebank

- One of the most widely known treebanks is the Penn TreeBank (PTB).
- The PTB was compiled at the University of Pennsylvania; the latest release was in 1999.
- Most well known is the Wall Street Journal section of the Penn Treebank.
- This section contains 1 million tokens from the Wall Street Journal (1987–1989).

The Penn Treebank

```
( (S
    (NP-SBJ
      (NP (NNP Pierre) (NNP Vinken) )
      (, ,)
      (ADJP
        (NP (CD 61) (NNS years) )
        (JJ old) )
      (, ,) )
  (VP (MD will)
    (VP (VB join)
      (NP (DT the) (NN board) )
      (PP-CLR (IN as)
        (NP (DT a) (JJ nonexecutive) (NN director) ))
      (NP-TMP (NNP Nov.) (CD 29) )))
  (. .) ))
```

PTB bracket labels

Word	Description
NNP	Proper noun
CD	Cardinal number
NNS	Noun, plural
JJ	Adjective
MD	Modal
VB	Verb, base form
DT	Determiner
NN	Noun, singular
IN	Preposition
…	…

Phrase	Description
S	Declarative clause
NP	Noun phrase
ADJP	Adjective phrase
VP	Verb phrase
PP	Prepositional
ADVP	Adverb phrase
RRC	Reduced relative
WHNP	*Wh*-noun phrase
NAC	Not a constituent
…	…

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

UPPSALA UNIVERSITET

The Penn Treebank

```
( (S
    (NP-SBJ
      (NP (NNP Pierre) (NNP Vinken) )
      (, ,)
      (ADJP
        (NP (CD 61) (NNS years) )
        (JJ old) )
      (, ,) )
    (VP (MD will)
      (VP (VB join)
        (NP (DT the) (NN board) )
        (PP-CLR (IN as)
          (NP (DT a) (JJ nonexecutive) (NN director) ))
        (NP-TMP (NNP Nov.) (CD 29) )))
    (. .) ))
```

UPPSALA UNIVERSITET

Treebank grammars

The Penn Treebank

```
( (S
    (NP-SBJ
      (NP (NNP Pierre) (NNP Vinken) )
      (, ,)
      (ADJP
        (NP (CD 61) (NNS years) )
        (JJ old) )
      (, ,) )
  (VP (MD will)
    (VP (VB join)
      (NP (DT the) (NN board) )
      (PP-CLR (IN as)
        (NP (DT a) (JJ nonexecutive) (NN director) ))
      (NP-TMP (NNP Nov.) (CD 29) )))
  (. .) ))
```

S → NP-SBJ VP .

The Penn Treebank

```
( (S
    (NP-SBJ
      (NP (NNP Pierre) (NNP Vinken) )
      (, ,)
      (ADJP
        (NP (CD 61) (NNS years) )
        (JJ old) )
      (, ,) )
    (VP (MD will)
      (VP (VB join)
        (NP (DT the) (NN board) )
        (PP-CLR (IN as)
          (NP (DT a) (JJ nonexecutive) (NN director) ))
        (NP-TMP (NNP Nov.) (CD 29) )))
    (. .) ))
```

NP-SBJ → NP , ADJP ,

The Penn Treebank

```
( (S
    (NP-SBJ
      (NP (NNP Pierre) (NNP Vinken) )
      (, ,)
      (ADJP
        (NP (CD 61) (NNS years) )
        (JJ old) )
      (, ,) )
    (VP (MD will)
      (VP (VB join)
        (NP (DT the) (NN board) )
        (PP-CLR (IN as)
          (NP (DT a) (JJ nonexecutive) (NN director) ))
        (NP-TMP (NNP Nov.) (CD 29) )))
    (. .) ))
```

ADJP → NP JJ

The Penn Treebank

```
( (S
    (NP-SBJ
      (NP (NNP Pierre) (NNP Vinken) )
      (, ,)
      (ADJP
        (NP (CD 61) (NNS years) )
        (JJ old) )
      (, ,) )
    (VP (MD will)
      (VP (VB join)
        (NP (DT the) (NN board) )
        (PP-CLR (IN as)
          (NP (DT a) (JJ nonexecutive) (NN director) ))
        (NP-TMP (NNP Nov.) (CD 29) )))
    (. .) ))
```

NP → CD NNS

The Penn Treebank

```
( (S
    (NP-SBJ
      (NP (NNP Pierre) (NNP Vinken) )
      (, ,)
      (ADJP
        (NP (CD 61) (NNS years) )
        (JJ old) )
      (, ,) )
    (VP (MD will)
      (VP (VB join)
        (NP (DT the) (NN board) )
        (PP-CLR (IN as)
          (NP (DT a) (JJ nonexecutive) (NN director) ))
        (NP-TMP (NNP Nov.) (CD 29) )))
    (. .) ))
```

NP → NNP NNP

Coverage of treebank grammars

- A treebank grammar will account for all analyses in the treebank.
- It can also be used to derive sentences that were not observed in the treebank.

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

UPPSALA UNIVERSITET

Different types of evaluation

- Intrinsic versus extrinsic evaluation.
  Evaluate relative to some gold standard vs. evaluate in the context of some specific task
- Automatic versus manual evaluation.
  Evaluate relative to some predefined measure vs. evaluate by humans.

Standard evaluation in parsing

- Intrinsic and automatic
- Parsers based on treebank grammars are evaluated by comparing their output to some gold standard.
- For this purpose, the treebank is customarily split into three sections: *training*, *tuning*, and *testing*.
- The parser is developed on *training* and *tuning*; final performance is reported on *testing*.

Bracket score

- The standard measure to evaluate phrase structure parsers is bracket score.
- Bracket: [min, max, category]
- One compares the brackets found by the parser to the brackets in the gold standard tree.
- Performance is reported in terms of precision, recall, and F-score.

Bracket score

- The standard measure to evaluate phrase structure parsers is bracket score.
- Bracket: [min, max, category]

signature!

- One compares the brackets found by the parser to the brackets in the gold standard tree.
- Performance is reported in terms of precision, recall, and F-score.

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 2 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

More about treebanks

UPPSALA UNIVERSITET

Treebank types - examples

- Phrase-structure treebanks
  - Penn treebank (English, and Chinese, Arabic)
  - NEGRA (German)
- Dependency treebanks
  - Prague Dep. treebank (Czech, + other)
  - Danish Dep. treebank (Danish)
  - Converted phrase-structured treebanks (e.g. Penn)
- Other
  - CCGBank (CCG, English)
  - LinGO Redwoods (HPSG, English)

UPPSALA UNIVERSITET

Swedish Treebank

- Combination of two older treebanks which have been merged and harmonized:
  - SUC (Stockholm-Umeå Corpus)
  - Talbanken
- Size: ~350 000 tokens
- Phrase structure annotation with functional labels
- Converted to dependency annotation
- Some parts checked by humans, some annotated automatically

UPPSALA UNIVERSITET

Domains and languages

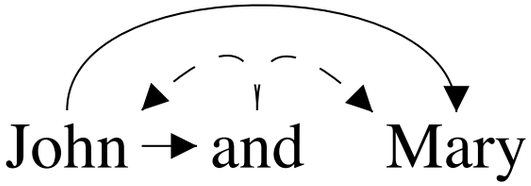
- Most of the parsing research was traditionally performed for English on the Wall Street Journal part of Penn Treebank
- Results for other English domains and for other languages are often worse than English WSJ
- Possible reasons
  - Parsing methods developed for English tends to work best for English (WSJ)
  - Language differences
  - Annotation differences
  - Treebank size and quality
  - ...

Treebank annotation issues

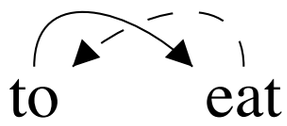
- Not only one possible annotation
- Important to have clear guidelines
- Quality control in the annotation project

UPPSALA UNIVERSITET

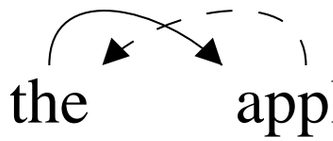
Dependency annotation options



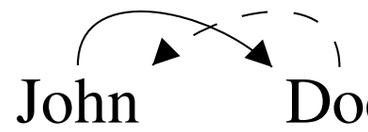
(a) Coordination



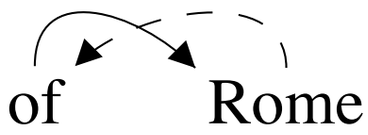
(b) Infinitive Verbs



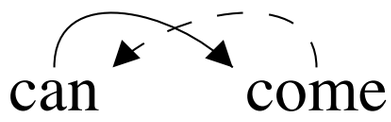
(c) Noun Phrases



(d) Noun Sequence



(e) Prepositional Phrases



(f) Verb Groups

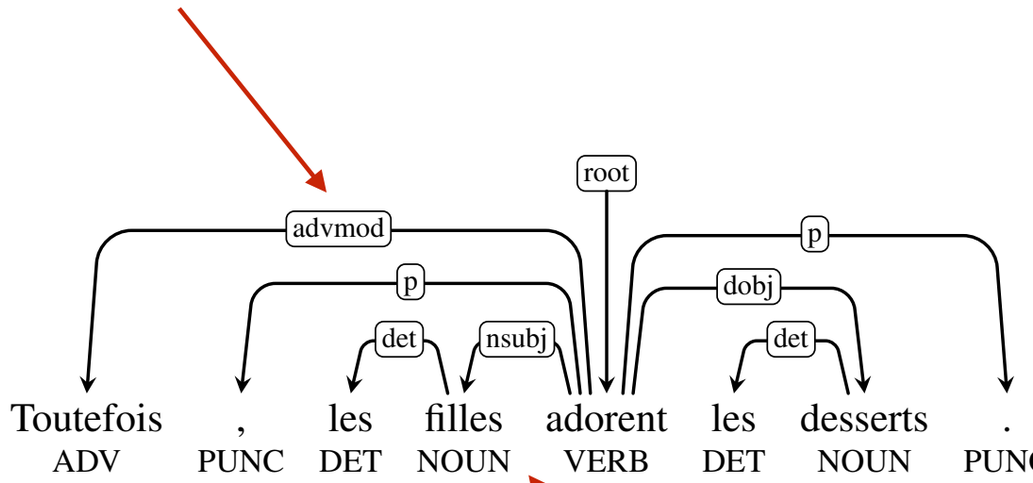
Schwartz et al. CoLING 2012.

UPPSALA UNIVERSITET

Univeral dependencies

Stanford dependencies (de Marneffe et al, 2006), adapted and harmonised for cross-lingual consistency

Version 1.1:
English
Finnish
French
German
Italian
Indonesian
Japanese
Korean
Portuguese
Spanish
Swedish
March 2014



Google part-of-speech tags (Petrov et al, 2012), fine-grained language specific tags if available

Version 1.0:
English
French
German
Korean
Spanish
Swedish
July 2013

from Joakim Nivre

Version 1.2: 33 languages, 37 treebanks
Version 2.0: >60 languages, >100 treebanks
Many more in next release!

Universal dependency principles

- Maximize parallelism
  - Don’t annotate the same thing in different ways
  - Don’t make different things look the same
- Don’t overdo it
  - Don’t annotate things that aren’t there
  - Languages select from a universal pool of categories
  - Allow language-specific extensions
- Use content words as heads

UPPSALA UNIVERSITET

Usefulness of consistent annotations

- Compare empirical results across languages
- Cross-lingual structure transfer
- Evaluate cross-lingual learning
- Build and maintain multilingual systems
- Make comparative linguistic studies
- Validate linguistic typology
- Make progress towards a universal parser

.

UPPSALA UNIVERSITET

Dependency parsing

- Dependency parsing has traditionally been evaluated for many languages:
- CoNLL 2006-2007 shared task
  - 10-13 languages
  - Different annotation schemes
- Universal dependencies
  - Many, and continually more, languages
  - Harmonized annotation

UPPSALA UNIVERSITET

Univeral dependency parsing results

Language	LAS, 2013	LAS, 2016
German	64.84	80.7
English	78.54	82.2
Swedish	70.90	85.9
Spanish	70.29	87.3
French	73.37	85.5
Korean	55.85	82.5

From
McDonald et al. ACL 2013.
Dozat et al., CoNLL 2017.

UPPSALA UNIVERSITET

Summary

- One can extract probabilistic context-free grammars from treebanks.
- Parsers can be evaluated by comparing their output against a gold standard.
- Reading: J&M 12.4, 14.3, 14.7

UPPSALA UNIVERSITET

Overview this week

- Lecture Tuesday: The Earley algorithm
- Lecture Thursday: advanced PCFG+supervision
- Start reading the seminar article
- Work on assignment 1 and 2
  - Important to get started, think of your overall workload!
  - Contact me if you need help!