



UPPSALA
UNIVERSITET

Treebank Grammars and Parser Evaluation

Syntactic analysis/parsing

2018-04-17

Ali Basirat

Department of Linguistics and Philology

Partly based on slides from Marco Kuhlmann





UPPSALA
UNIVERSITET

Recap: Probabilistic parsing



Probabilistic context-free grammars

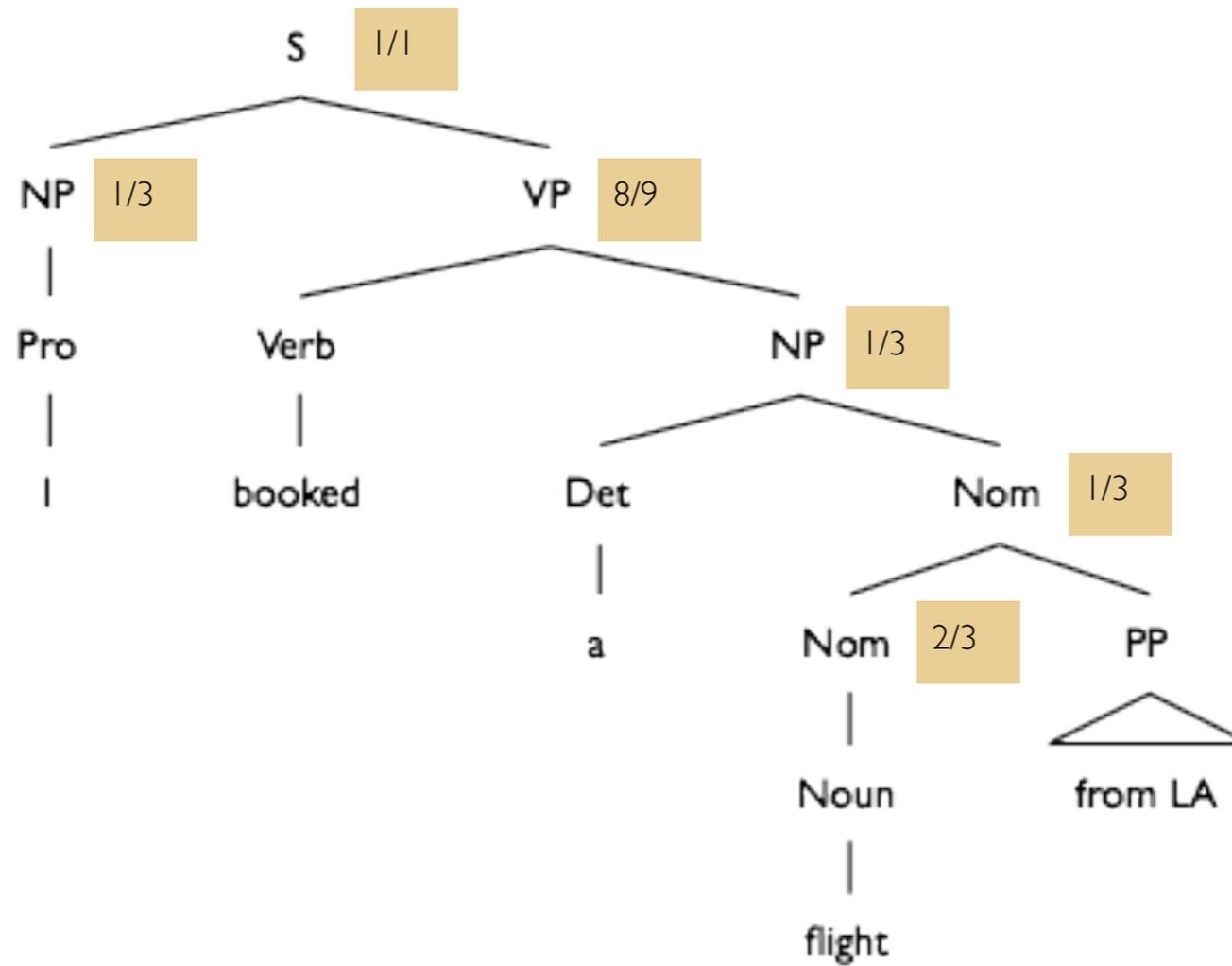
A **probabilistic context-free grammar (PCFG)**

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



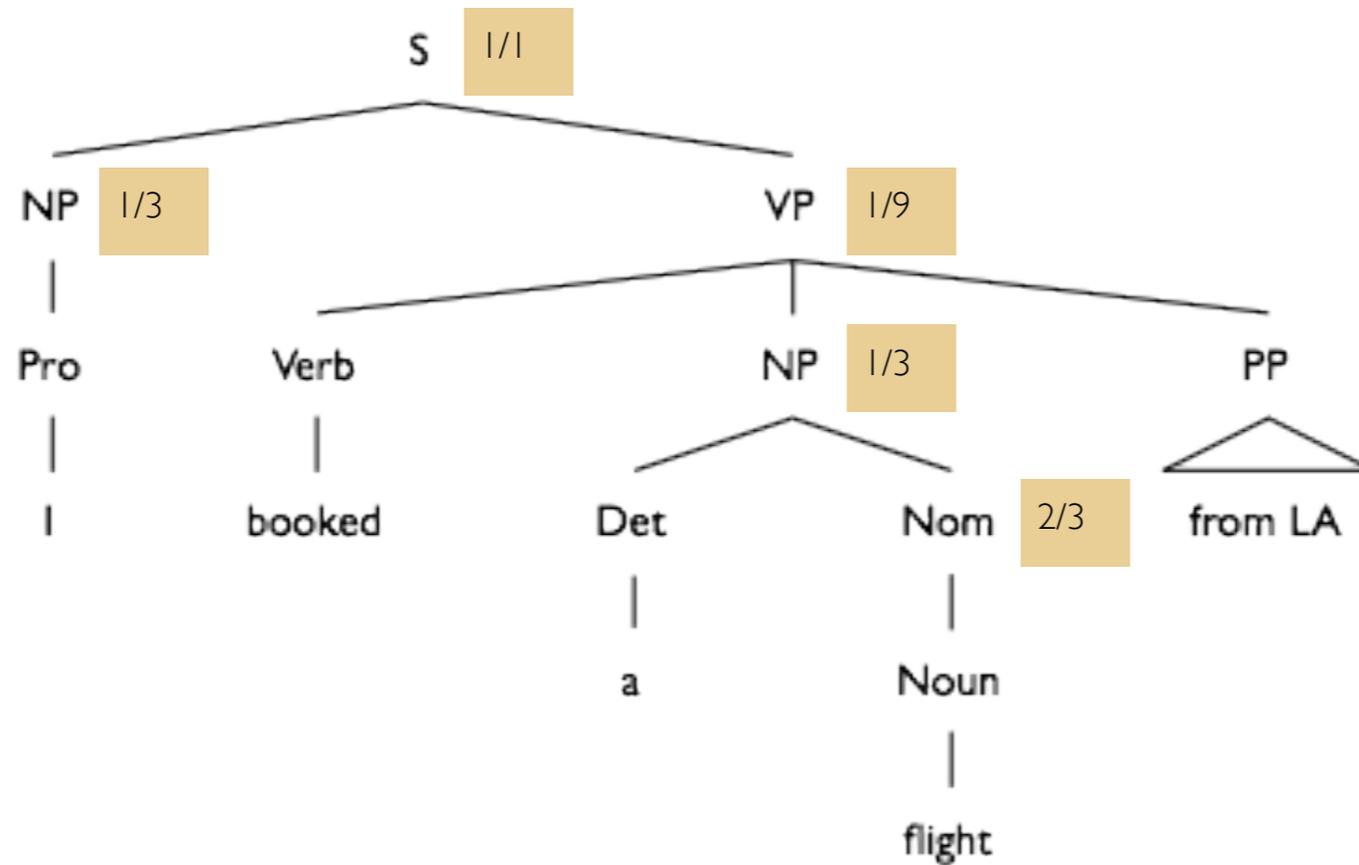
Probability of a parse tree



Probability: 16/729



Probability of a parse tree



Probability: 6/729



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
```



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
$S \rightarrow NP\text{-}SBJ VP .$	Pierre Vinken ... Nov. 29.
$NP\text{-}SBJ \rightarrow NP , ADJP ,$	Pierre Vinken, 61 years old,
$VP \rightarrow MD VP$	will join the board ...
$NP \rightarrow DT NN$	the board



The Penn Treebank

S → NP-SBJ VP . ((S

NP → NNP NNP

NP → CD NNS

(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 ,

ADJP → NP JJ



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

Treebank grammars

Problems with PCFG

- Poor independence assumption
- Lack of lexical conditioning



Poor Independence Assumption

- The probability of particular rule in a tree is independent of the rest of the tree
- But, the choice of node expansion depends on the rest of the tree
- Example: an NP in a subject position tends to be a pronoun, but in an object position tends to be a pronominal such as a proper noun
- We need to augment PCFG to model structural dependencies



Parent annotation

- To split non-terminals such as NP into multiple non-terminals such as NP^S and NP^{VP}
- It increases the size of the grammar
- It reduces the size of the training data for each grammar rule
- It can lead to overfitting
- The first paper in literature review introduces an split and merge algorithm



Lack of lexical conditioning

- Words play a role in the pre-terminal rules of PCFG
- They can be useful in the other places in the grammar
- We need to augment PCFG to deal with lexical dependency statistics
- Probabilistic Lexicalized CFGs – a seminar on 29th of April



UPPSALA
UNIVERSITET

Parser evaluation



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/dev*, 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**.



signature!



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:
$$2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$$



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



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)



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



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

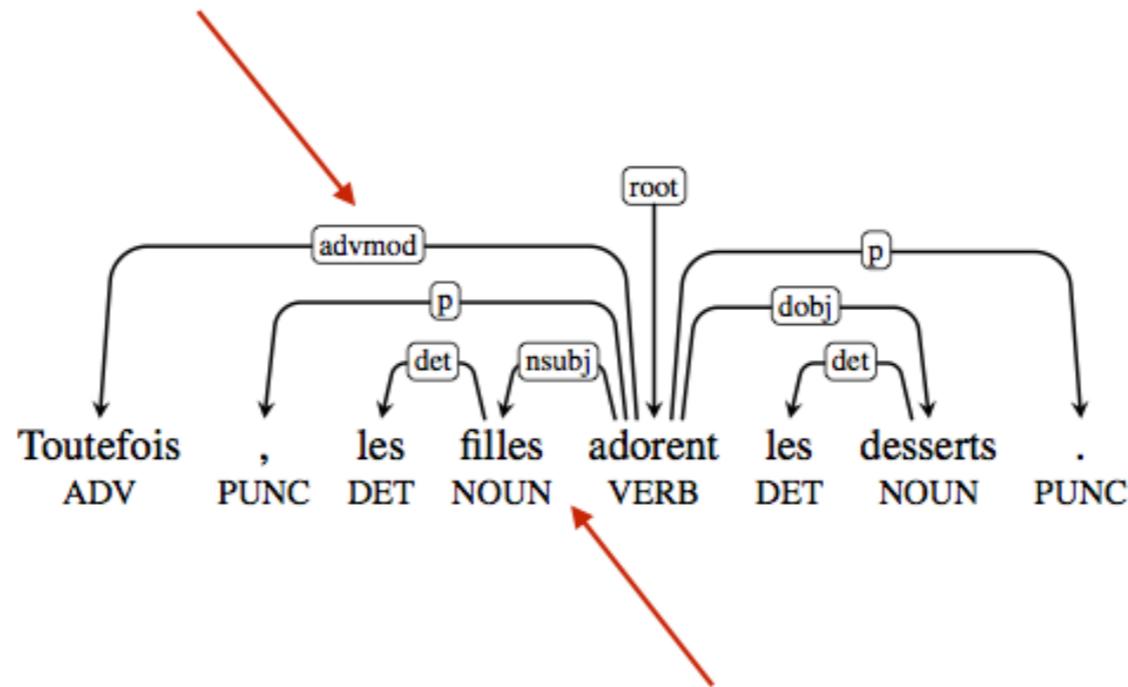


Parser evaluation

UPPSALA
UNIVERSITET

**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 2.1: 60+ languages, 100+ treebanks

More to come in next release!



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



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



Universal dependency parsing results

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.



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 10.4.1-2, 12.3, 12.8, 13.3



Next

- The Earley algorithm
- Seminar: Lexicalized grammars – Group 3
- Start reading the first paper of literature review
- Work on assignment 1 (and 2)
 - Important to get started, think of your overall workload!
 - Contact me if you need help!