CS626: Speech, NLP and the Web

Going deeper into Deep Parsing with Constituency,

Dependency Parsing introduced

Pushpak Bhattacharyya

Computer Science and Engineering Department

IIT Bombay

Week of 21st September, 2020

Agenda for the week

- Deeper look into parsing
  - Constituency
  - Dependency
- Developing Probabilistic parsing
- Introduce Neural Parsing

Example

“The cameraman shot the batsman when he was near the minister.”

S→NP VP: apply on the core "The cameraman shot the batsman"

- NP
  - DT
    - "the"
  - NN
    - "cameraman"
- VP
  - VBD
    - "shot"
  - NP
    - DT
      - » the
    - NN
      - » batsman

"…when he was near the minister"

NP ("The cameraman")

- NP→ DT NN
- "The cameraman"
  - DT
    - The
  - NN
    - caemraman

VP (""shot the batsman when he was near the minister")

- VP→ VBD NP SBAR
- VBD
  - Shot
- NP
  - The batsman
- SBAR

…

SBAR→ WHADVP S

- WHAADVP
  - "when"
- NP
  - "he"

S →NP VP

- VP→ VBD PP
  - VBD
    - "was"
  - PP
    - P
      - "near"
    - NP→ DT NN ("the minister")

If cameraman was near the minister…

- S
  - NP
  - VP
    - VBD
      - "shot"
    - NP
      - "the batsman"
    - SBAR
      - "when he was near the minister"

If the batsman was near the minister…

- S
  - NP
  - VP
    - VBD
      - “shot”
    - NP
      - …
- NP
  - DT
    - » “the”
  - NN
    - » “the batsman”
  - SBAR
    - » “when he was near the minister”

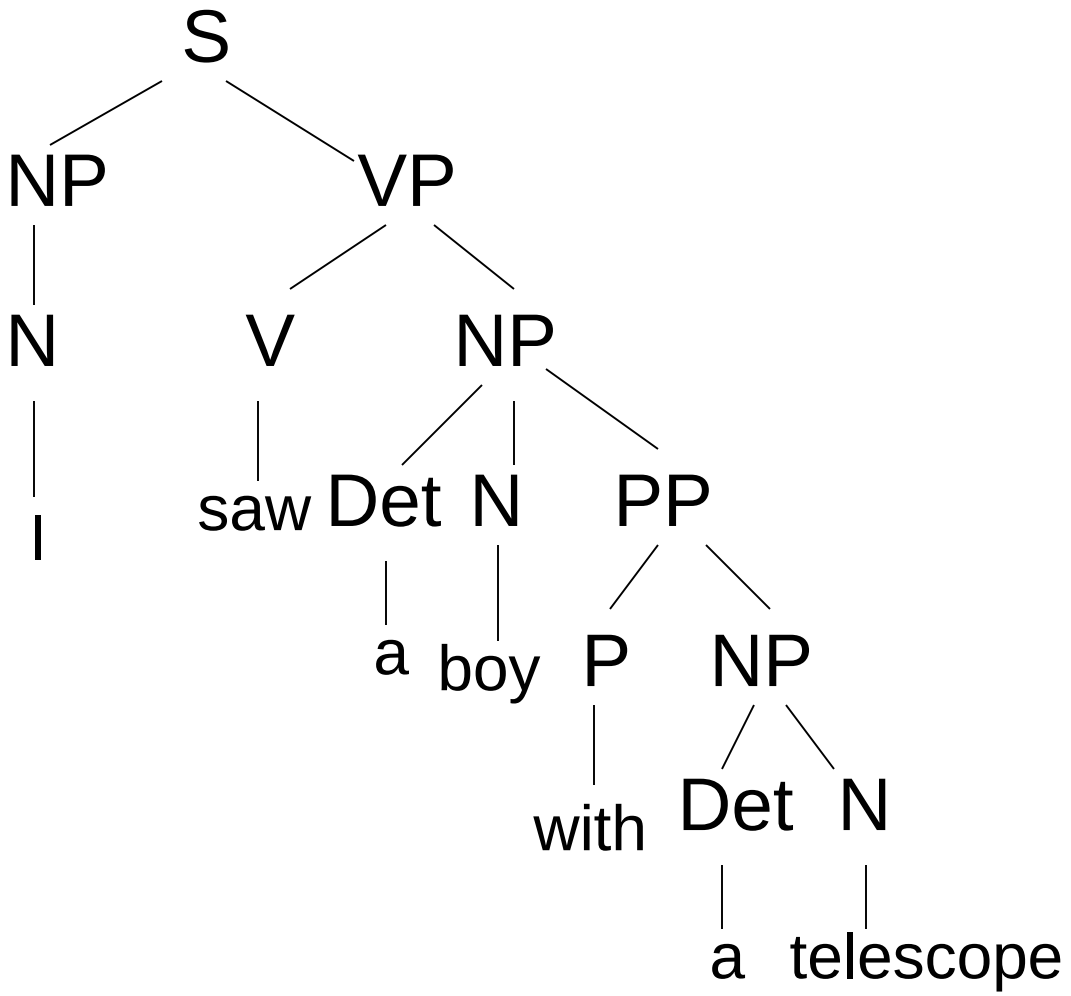
Coreference resolution

- Coreference resolution concerns
  - Finding different linguistic expressions that refers to same entity
- Eg:
  - Binding a pronoun with corresponding noun: **Anaphora Resolution**
- The cameraman shot the batsman when he was near the minister.
  - Ambiguity:
    - “He” refers to “The cameraman”, or
    - “He” refers to ”batsman”

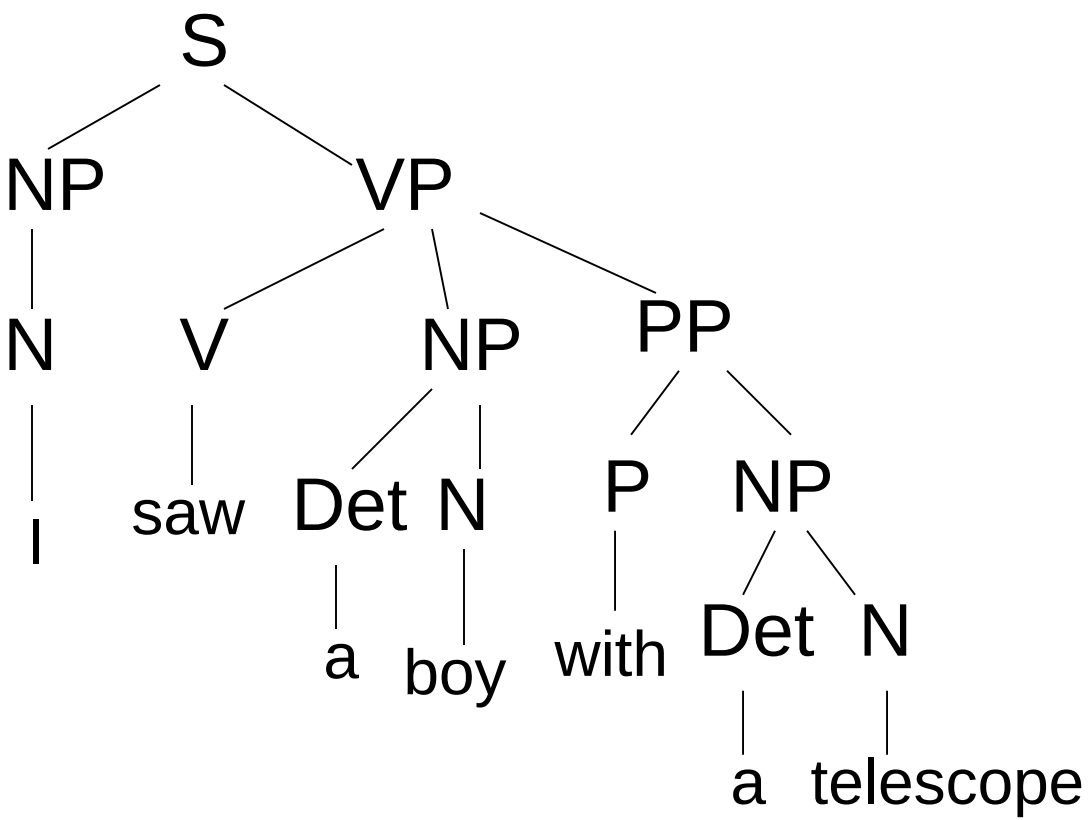
Types of adverbs

Category of adverb	Meaning	Examples
Adverbs of manner	In what manner, something is being done	quickly, softly
Adverbs of time	When the verb took place	recently, weekly
Adverbs of place	Where the verb took place	walking near the house, here, there
Adverbs of frequency	How often the verb occurs	How often the verb occurs
Adverbs of degree	Intensity of the verb	almost

Parsing challenge- PP ambiguity: “I saw the boy with a telescope”



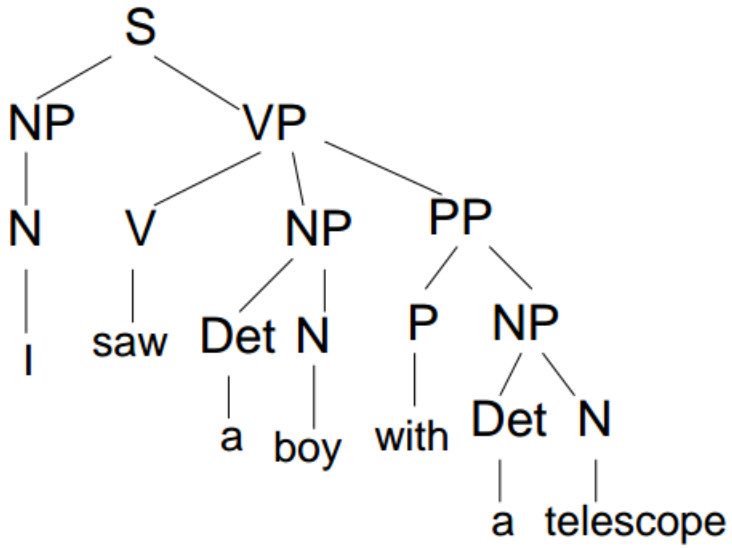
Constituency Parse Tree -2



Resolving PP attachment ambiguity

- The attachment of PP (Prepositional phrase) is determined based on the rule of proximity
- Rule of proximity
  - The lesser path length will represent the attachment point
  - Path length: the number of edges between two nodes

Resolving PP attachment ambiguity: Example



- I saw a boy with a telescope
- PP attachment ambiguity
  - with a telescope
    - boy (noun)
    - saw (verb)
- Path length between boy and PP = 4
- Path length between saw and PP = 3
- PP “with a telescope” is attached to “saw”
- Meaning: I used the telescope to see a boy

“Attachment” is the crux of the matter

- PP attachment
  - “I saw the boy with a telescope”
    - PP- “with the telescope”
- Clause attachment
  - “The cameraman shot the batsman when he was near the minister”
    - Clause- “when he was near the minister”

Difference between SBAR and S

- Contribution of Generative Grammar (Noam Chomsky)
- SBAR is used in complex sentences
  - Complex sentences: have clauses
  - Should start with “wh”-word
- S is used for “normal” sentences

“The cameraman shot the batsman when he was near the minister”

- S:
  - “The cameraman shot the batsman when he was near the minister”
- SBAR
  - “when he was near the minister”
- S
  - “he was near the minister”

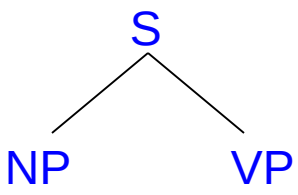
“I know the boy who lives in Delhi”

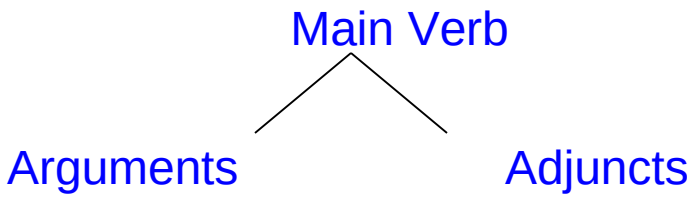
- S
  - NP: “I”
  - VP: “know the boy…”
    - VB: “know”
    - NP: “the boy”
      - WHNP
      - S
        - …
  - S
    - VP
      - VBZ: “lives”
        - PP
          - P: “in”
          - NNP: “Delhi”

Types of sentences

- Simple
  - “The cameraman shot the batsman”
- Complex
  - “The cameraman shot the batsman when he was near the minister”
- Compound
  - “The cameraman shot the batsman and he was happy”

Two kinds of parse representations: Constituency Vs. Dependency





Example: raw sentence

The strongest rain shut down the financial hub of Mumbai

(from: Stanford parser https://nlp.stanford.edu/software/lex-parser.shtml)

Example: POS Tagged sentence

The/DT strongest/JJS rain/NN shut/VBD down/RP the/DT financial/JJ hub/NN of/IN Mumbai/NNP

Constituency parse

```
(S
  (NP
      (DT The)
       (JJS strongest)
         (NN rain))
   )

(VP
    …
```

```
(VP
  (VP
     (VBD shut)
     (PRT (RP down))
  (NP
     (NP
       (DT the) (JJ financial)
       (NN hub))
     (PP (IN of)
          (NP (NNP Mumbai)))))
```

Dependency Parse

root(ROOT-0, shut-4)

nsubj(shut-4, rain-3)

prt(shut-4, down-5)

det(rain-3, the-1)

amod(rain-3, strongest-2)

dobj(shut-4, hub-8)

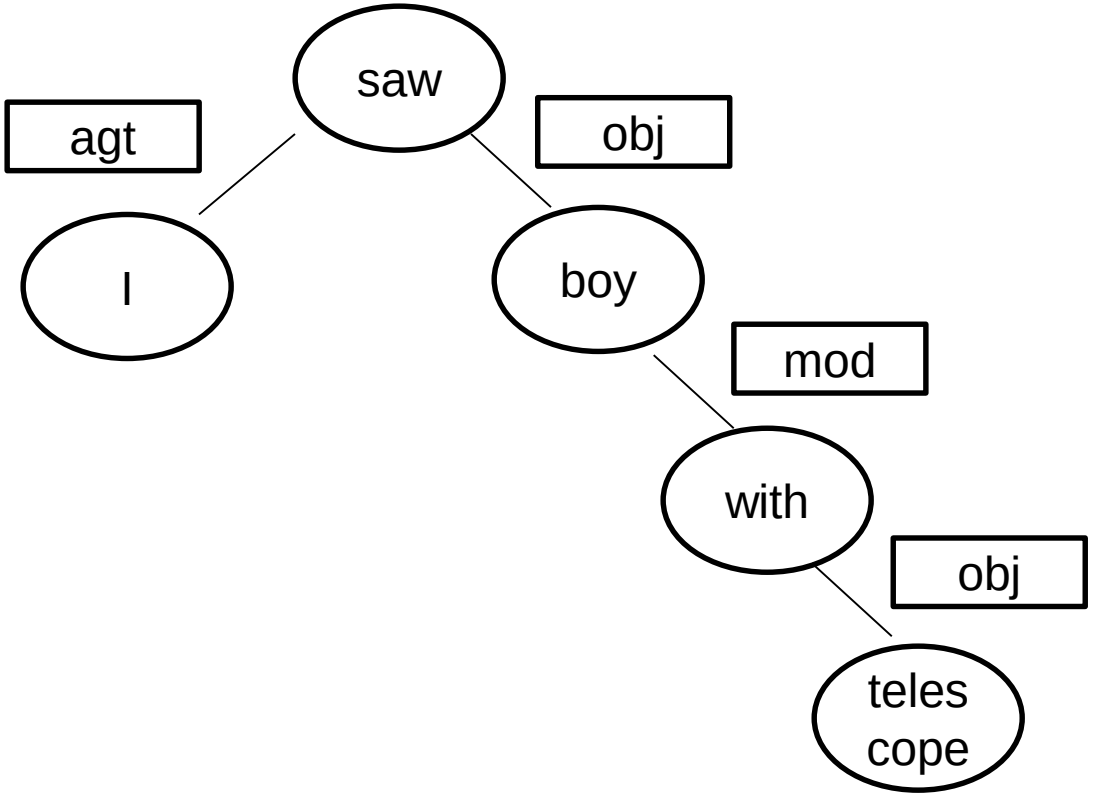
det(hub-8, the-6)

amod(hub-8, financial-7)

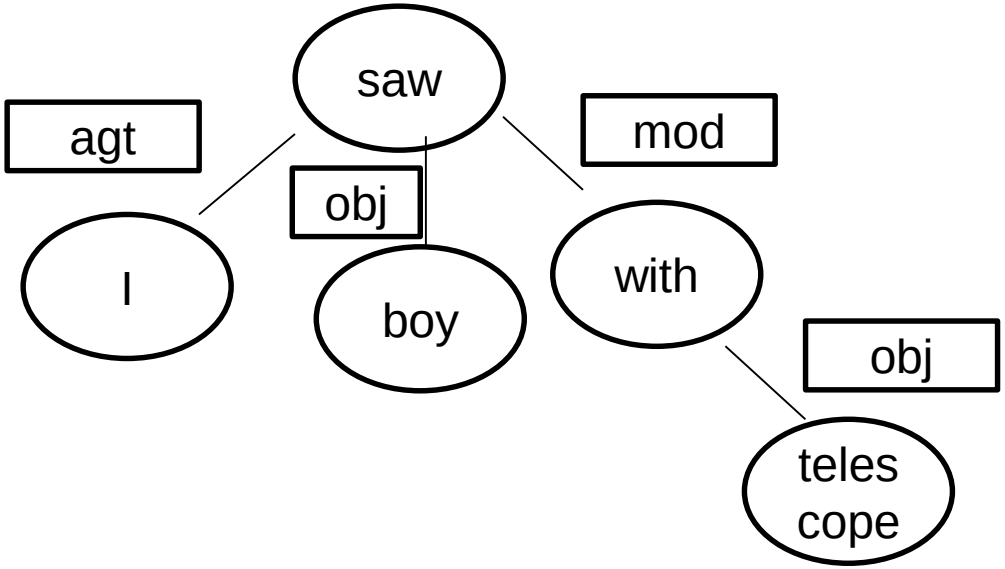
prep(hub-8, of-9)

pobj(of-9, Mumbai-10)

“I saw the boy with a telescope”: Dependency Parse Tree - 1



Dependency Parse Tree - 2



Probabilistic parsing

Example of Sentence labeling: Parsing

$[_{S1}[_{S}[_{S}[_{VP}[_{VB}$Come$][_{NP}[_{NNP}$July$]]]]]$

$[_{,}$,$]$

$[_{CC}$ and$]$

$[_{S}$ $[_{NP}$ $[_{DT}$ the$]$ $[_{JJ}$ IIT$]$ $[_{NN}$ campus$]]$

$[_{VP}$ $[_{AUX}$ is$]$

$[_{ADJP}$ $[_{JJ}$ abuzz$]$

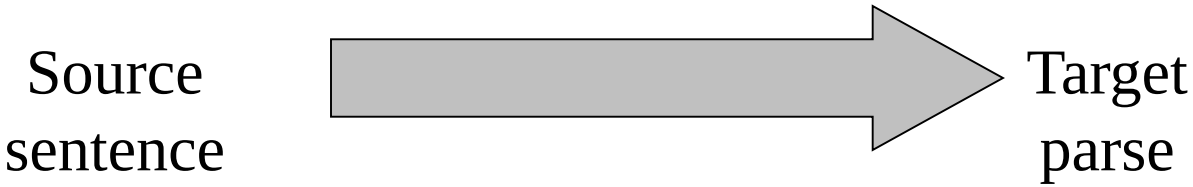
$[_{PP}[_{IN}$ with$]$

$[_{NP}[_{ADJP}$ $[_{JJ}$ new$]$ $[_{CC}$ and$]$ $[$ $_{VBG}$ returning$]]$

$[_{NNS}$ students$]]]]]]$

$[_{.}$.$]]]$

Noisy Channel Modeling



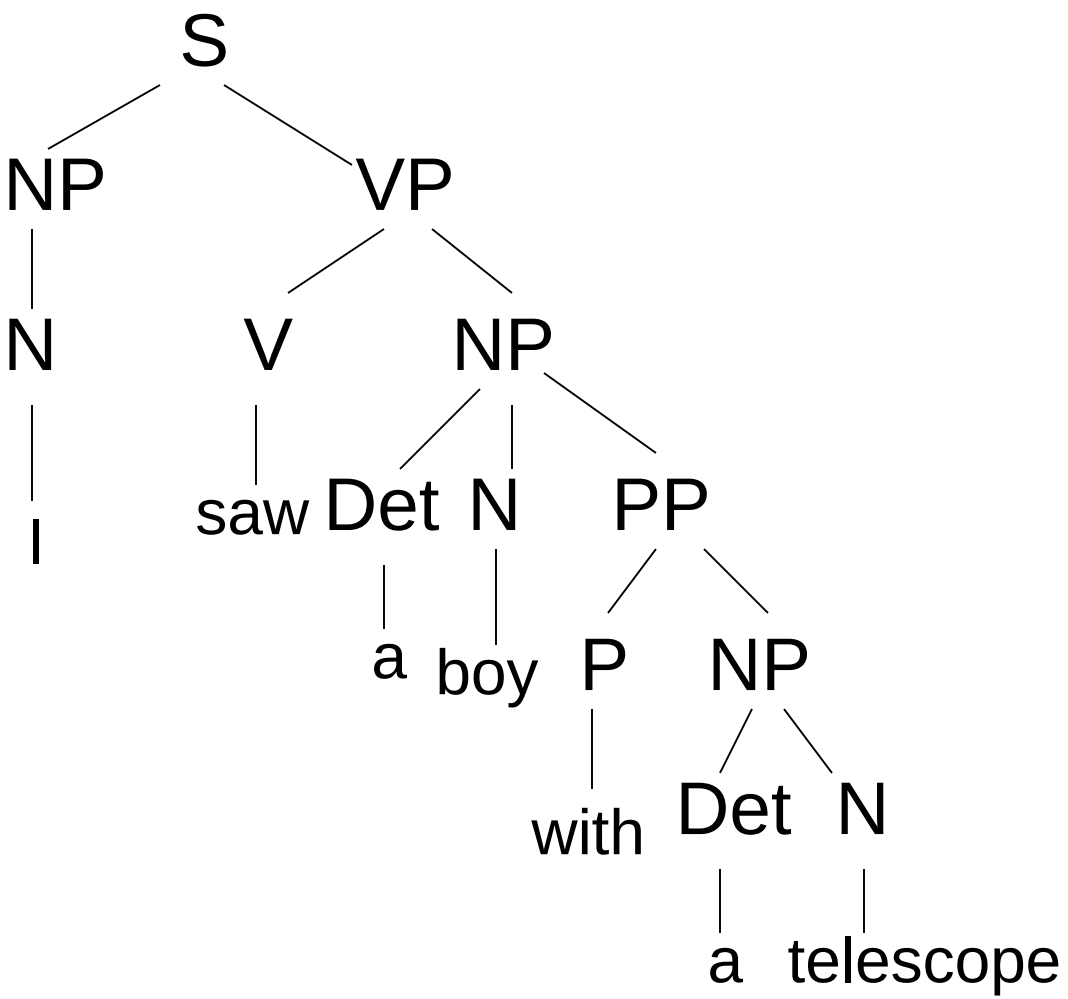
$$T^* = \underset{T}{\operatorname{argmax}}\,[P(T|S)]$$

$$= \underset{T}{\operatorname{argmax}}\,[P(T).P(S|T)]$$

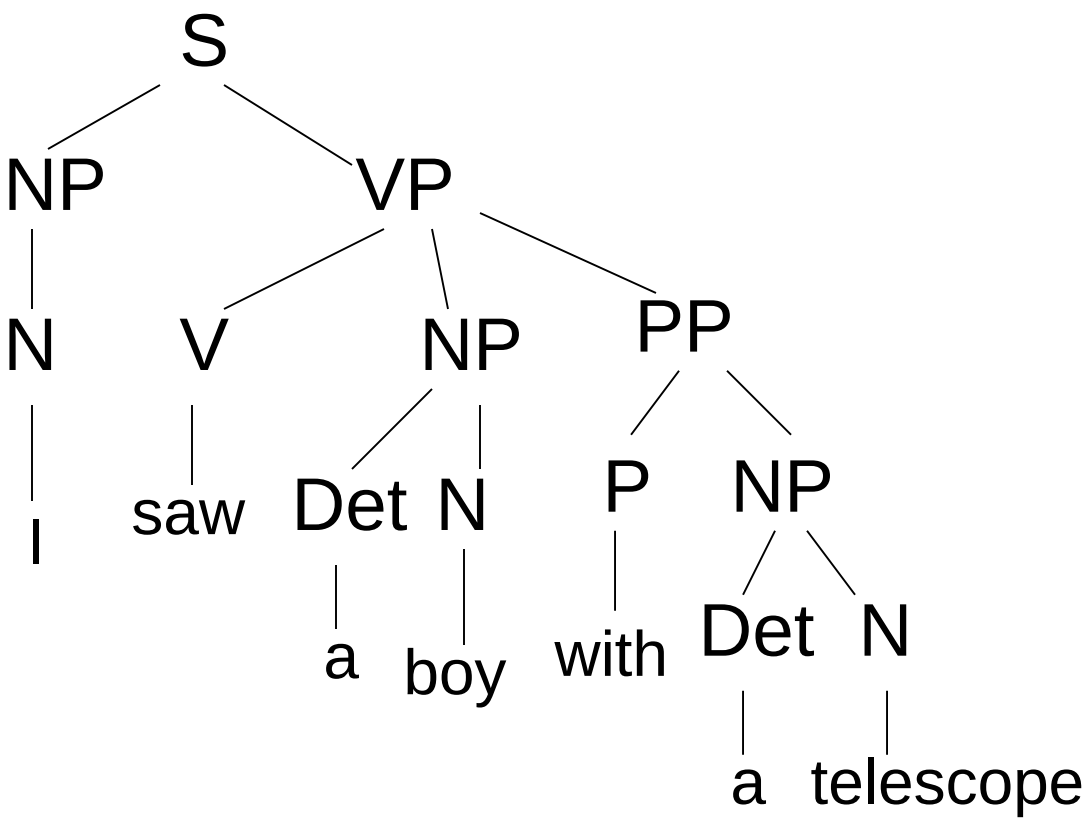
$$= \underset{T}{\operatorname{argmax}}\,[P(T)]$$

since given the parse the sentence is completely determined and $P(S|T)=1$

I saw a boy with a telescope: Tree - 1



Constituency Parse Tree -2



Formal Definition of PCFG

- A PCFG consists of
  - A set of terminals $\{w_k\}$, k = 1,….,V

    $\{w_k\}$ = { child, teddy, bear, played…}
  - A set of non-terminals $\{N^i\}$, i = 1,…,n

    $\{N_i\}$ = { NP, VP, DT…}
  - A designated start symbol N^1
  - A set of rules $\{N^i \rightarrow \zeta^j\}$, where ζ^j is a sequence of terminals & non-terminals

    NP $\rightarrow$ DT NN
  - A corresponding set of rule probabilities

Rule Probabilities

- Rule probabilities are such that

$$\forall i \quad \sum_{\mathrm{i}} \mathrm{P}(\mathrm{N}^i \rightarrow \zeta^j) = 1$$

E.g., P(NP $\rightarrow$ DT NN) = 0.2

P(NP $\rightarrow$ NN) = 0.5

P(NP $\rightarrow$ NP PP) = 0.3

- P(NP $\rightarrow$ DT NN) = 0.2
  - Means 20 % of the training data parses use the rule NP $\rightarrow$ DT NN

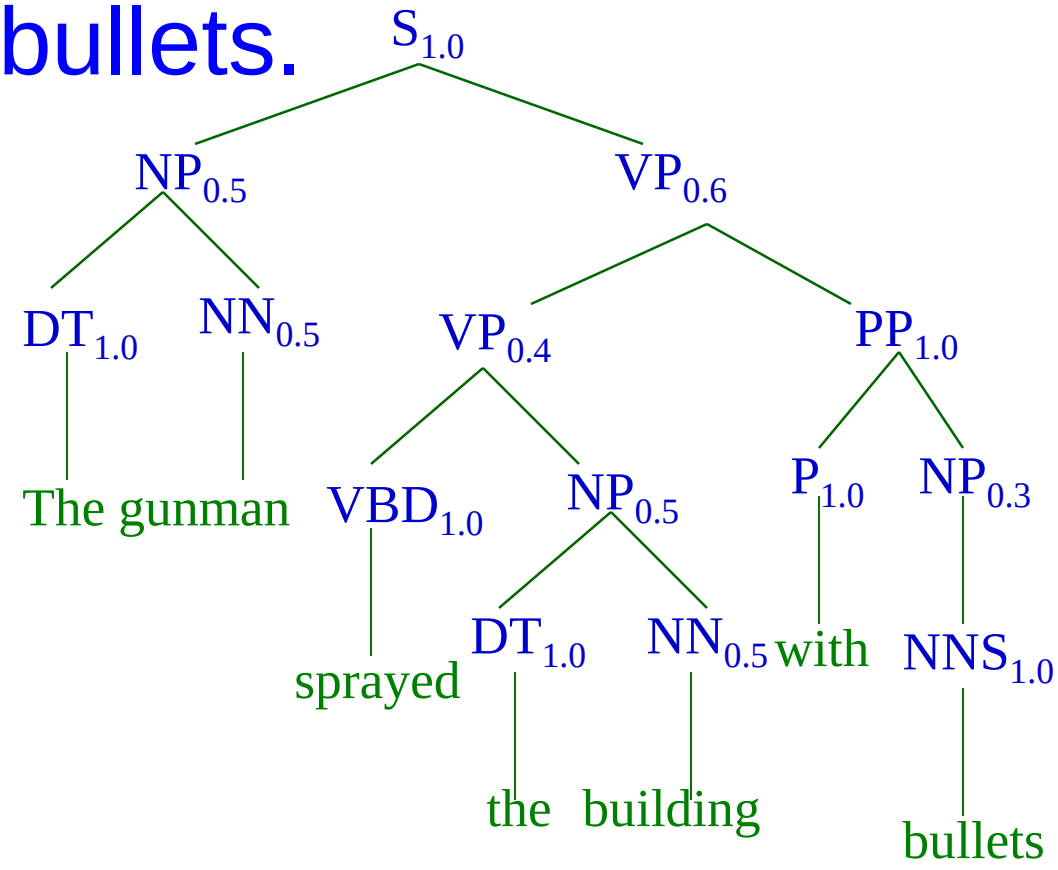
Probabilistic Context Free Grammars

- S → NP VP 1.0
- NP → DT NN 0.5
- NP → NNS 0.3
- NP → NP PP 0.2
- PP → P NP 1.0
- VP → VP PP 0.6
- VP → VBD NP 0.4

- DT → the 1.0
- NN → gunman 0.5
- NN → building 0.5
- VBD → sprayed 1.0
- NNS → bullets 1.0

Example Parse t_1`

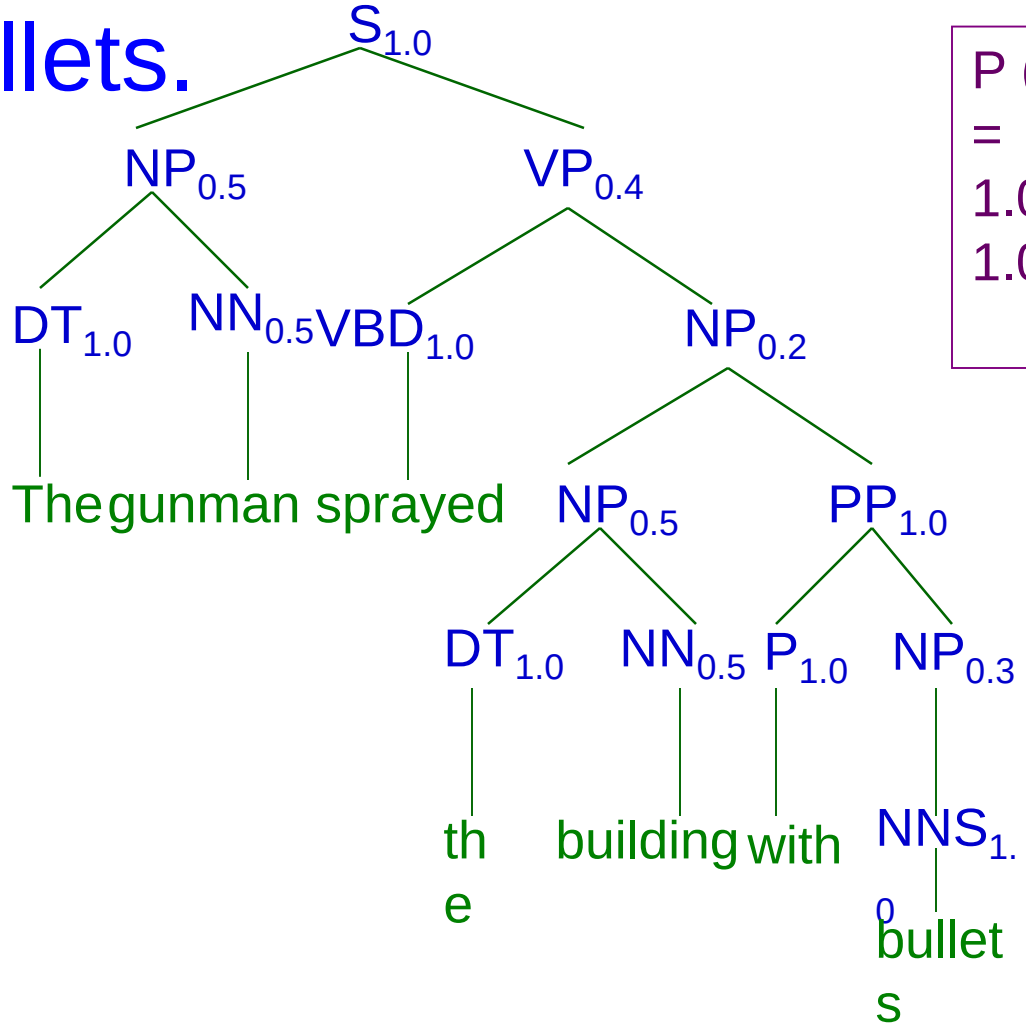
- The gunman sprayed the building with bullets.



$P(t_1) = 1.0 * 0.5 * 1.0 * 0.5 * 0.6 * 0.4 * 1.0 * 0.5 * 1.0 * 0.5 * 1.0 * 1.0 * 0.3 * 1.0 = 0.00225$

Another Parse t_2

- The gunman sprayed the building with bullets.

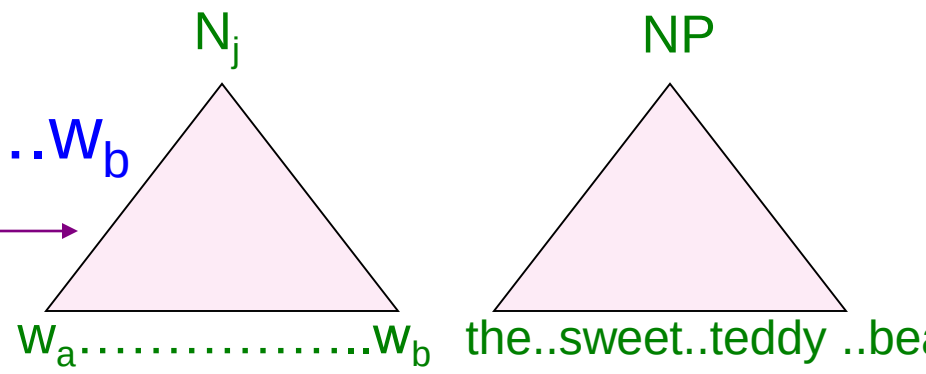


$P(t_2)$
$= 1.0 * 0.5 * 1.0 * 0.5 * 0.4 * 1.0 * 0.2 * 0.5 * 1.0 * 0.5 * 1.0 * 1.0 * 0.3 * 1.0$
$= 0.0015$

Probability of a sentence

- Notation :
  - w_{ab} – subsequence $w_a \ldots w_b$
  - N_j dominates $w_a \ldots w_b$

    or yield(N_j) = $w_a \ldots w_b$



- Probability of a sentence = $P(w_{1m})$

$$P(w_{1m}) = \sum_t P(w_{1m}, t)$$

→ Where t is a parse tree of the sentence

$$= \sum_t P(t) P(w_{1m} \mid t)$$

$$= \sum_{t: yield(t) = w_{1m}} P(t) \qquad \because P(w_{1m} \mid t) = 1$$

If t is a parse tree for the sentence w_{1m}, this will be 1 !!

Assumptions of the PCFG model

- Place invariance :

  P(NP $\rightarrow$ DT NN) is same in locations 1 and 2

- Context-free :

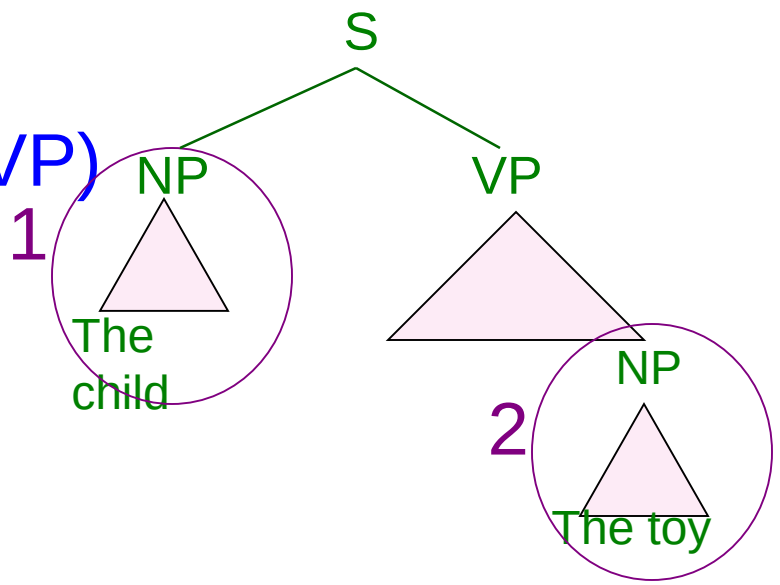
  P(NP $\rightarrow$ DT NN | anything outside “The child”)

  = P(NP $\rightarrow$ DT NN)

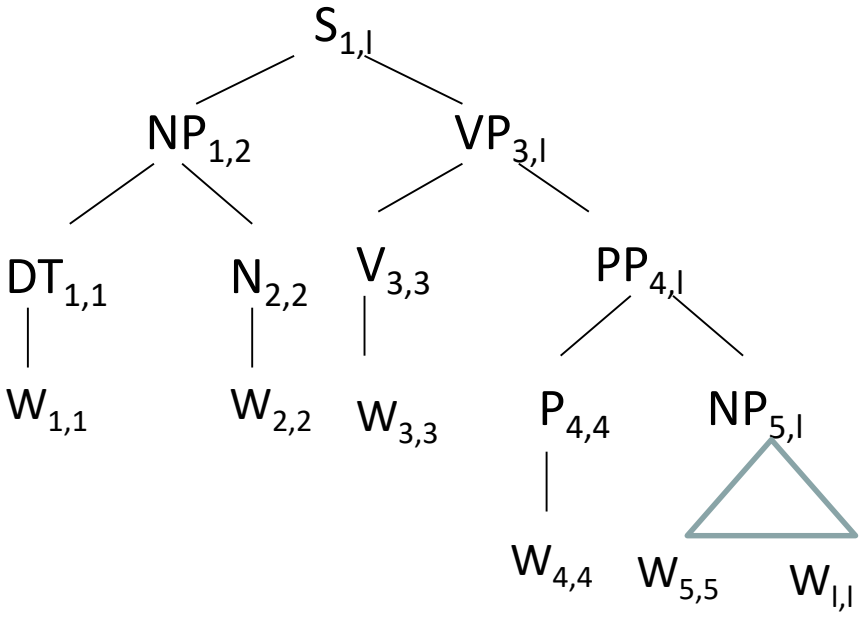
- Ancestor free : At 2,

  P(NP $\rightarrow$ DT NN|its ancestor is VP)

  = P(NP $\rightarrow$DT NN)



Probability of a parse tree (cont.)



$P(t|s) = P(t \mid S_{1,l})$

$= P(NP_{1,2}, DT_{1,1}, w_{1,1}$
$N_{2,2}, w_{2,2}$
$VP_{3,l}, V_{3,3}, w_{3,3}$
$PP_{4,l}, P_{4,4}, w_{4,4}\ NP_{5,l}, w_{5\ldots l} \mid S_{1,l})$

$= P(NP_{1,2}, VP_{3,l} \mid S_{1,l}) * P(DT_{1,1}, N_{2,2} \mid NP_{1,2}) *$
$P(w_{1,1} \mid DT_{1,1}) * P(w_{2,2} \mid N_{2,2}) * P(V_{3,3}, PP_{4,l} \mid VP_{3,l}) *$
$P(w_{3,3} \mid V_{3,3}) * P(P_{4,4}, NP_{5,l} \mid PP_{4,l}) * P(w_{4,4} \mid P_{4,4}) *$
$P(w_{5\ldots l} \mid NP_{5,l})$

(Using Chain Rule, Context Freeness and Ancestor Freeness)

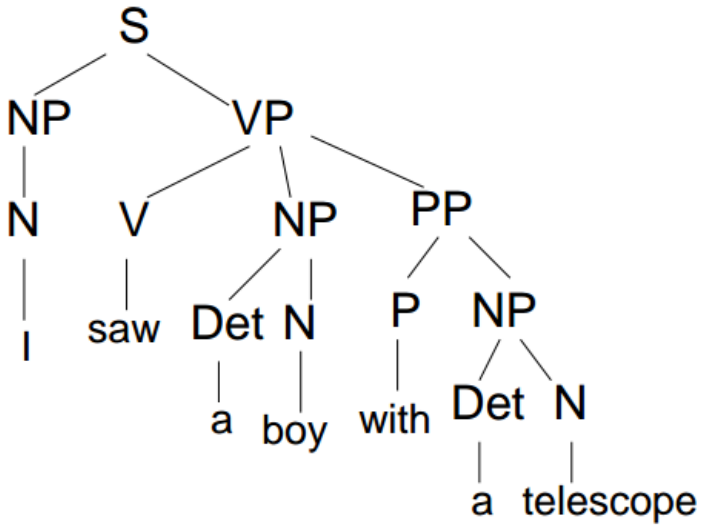
Why probability in Parsing

Domination

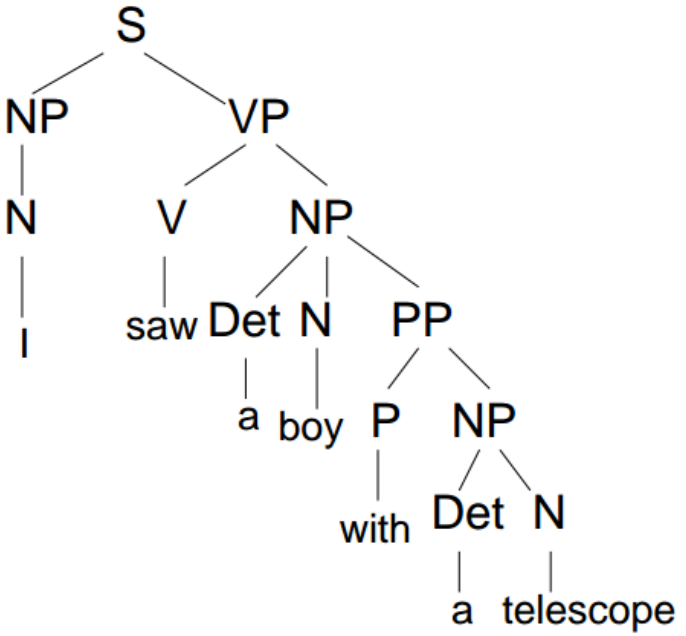
- A sentence is dominated by the symbol S through domination of segments by phrases
- Examples
  - The capital of a country dominates the whole country.
  - The capital of a state dominates the whole state.
  - The district headquarter dominates the district.
  - IIT Bombay is dominated by the administration of IIT Bombay.
  - Administration dominates Heads of Depts
  - The department is dominated by head of the department.

Ambiguity in determining domination

I saw a boy with a telescope.



- “saw” dominated by VP
- “a boy” dominated by NP
- “with a telescope” dominated by PP
- Yield of first NP is “a telescope”

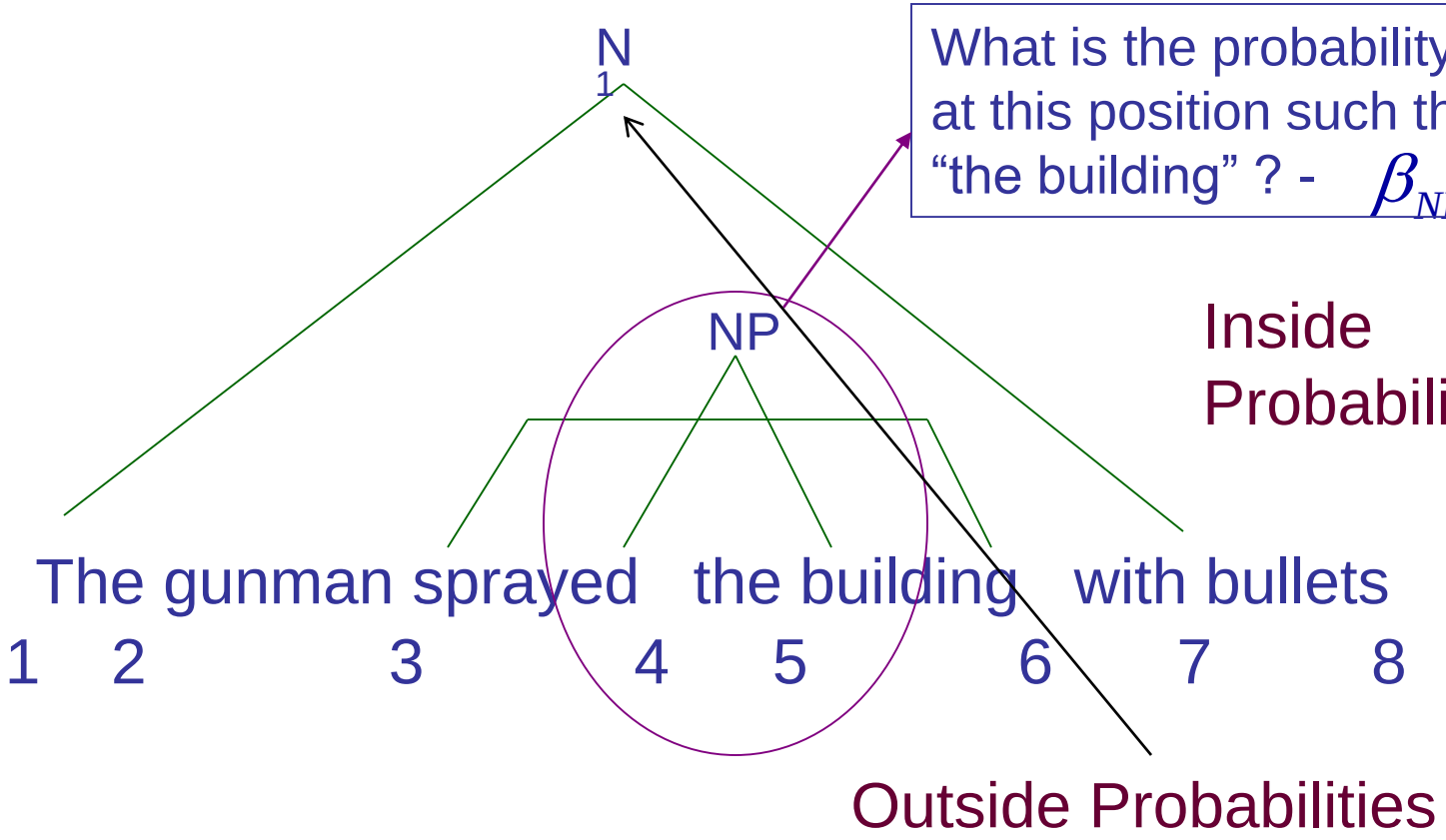


- “saw” dominated by VP
- “with a telescope” dominated by PP
- “a boy with a telescope” dominated by NP
- Yield of NP is a “a boy with a telescope”

Need of Probabilistic Parsing

- Main Intuition
  - Resolving the uncertainty
    - which non-terminal dominates how much territory in the sentence.
- The ambiguity in determining
  - The yield of NP
  - Will the NP dominate “a boy” or “a boy with a telescope”

Interesting Probabilities



What is the probability of having a NP at this position such that it will derive "the building" ? - $\beta_{NP}(4,5)$

Inside Probabilities

Outside Probabilities

What is the probability of starting from N^1 and deriving "The gunman sprayed", a NP and "with bullets" ? - $\alpha_{NP}(4,5)$

Parse tree for the given sentence using probabilistic CYK parsing

0 **The** 1 **gunman** 2 **sprayed** 3 **the** 4 **building** 5 **with** 6 **bullets** 7

- Two parse trees are possible because the sentence has attachment ambiguity .

- Total 16 multiplications are required to make both the parse trees using probabilistic CYK.

- Number of multiplications is less in comparison to a probabilistic parsing which prepares the two parse trees independently with 28 multiplication.

	The 1	gunman 2	Sprayed 3	the 4	Building 5	with 6	Bullets 7
0	β_{DT} (0-1) =1.0	β_{NP} (0-2) =0.25					β_{S}(0-7) =0.006
1		β_{NN} (1-2) =0.5					
2			β_{VBD}(2-3) =1.0		β_{VP} (2-5) =0.1		β_{VP}(2-7) =0.024
3				β_{DT}(3-4) =1.0	β_{NP} *(3-5) =0.25*		β_{NP}(3-7) =0.015
4					β_{NN} (4-5) =0.5		
5						β_{P}(5-6) =1.0	β_{PP}(5-7) =0.3
6							$\beta_{NP/NNS}$(6-7) =1.0

Calculation of values for each non terminal occuring in the CYK table

$\beta_{DT}(0\text{-}1) = 1.0$ (From Grammar rules)

$\beta_{NN}(1\text{-}2) = 0.5$ (From Grammar rules)

$\beta_{NP}(0\text{-}2) = P(\text{the gunman} \mid NP_{0\text{-}2}, G)$

$= P(NP \rightarrow DT\ NN) * \beta_{DT}(0\text{-}1) * \beta_{NN}(1\text{-}2)$

$= 0.5 * 1.0 * 0.5$

$= 0.25$

$\beta_{VBD}(2\text{-}3) = 1.0$ (From Grammar rules)

$\beta_{DT}(3\text{-}4) = 1.0$ (From Grammar rules)

$\beta_{NN}(4\text{-}5) = 0.5$ (From Grammar rules)

$\beta_{NP}(3\text{-}5) = P(\text{the building} \mid NP_{3\text{-}5}, G)$

$= P(NP \rightarrow DT\ NN) * \beta_{DT}(3\text{-}4) * \beta_{NN}(4\text{-}5)$

$= 0.5 * 1.0 * 0.5$

$= 0.25$

$\beta_{VP}(2\text{-}5) = P(VP \rightarrow VBD\ NP) * \beta_{VBD}(2\text{-}3) * \beta_{NN}(3\text{-}5)$

$= 0.4 * 1 * 0.25$

$= 0.1$

$\beta_{P}(5\text{-}6) = 1.0$ (From Grammar rules)

$\beta_{NP/NNS}(6\text{-}7) = 1.0$ (From Grammar rules)

$\beta_{PP}(5\text{-}7) = P(PP \rightarrow P\ NP) * \beta_{P}(5\text{-}6) * \beta_{NP/NNS}(6\text{-}7)$

$= 1.0 * 1.0 * 0.3$

$= 0.3$

$\beta_{NP}(3\text{-}7) = P(NP \rightarrow NP\ PP) * \beta_{NP}(3\text{-}5) * \beta_{PP}(5\text{-}7)$

$= 0.2 * 0.25 * 0.3$

$= 0.015$

$\beta_{VP}(2\text{-}7) = (P(VP \rightarrow VBD\ NP) * \beta_{VBD}(2\text{-}3) * \beta_{NP}(3\text{-}7) + P(VP \rightarrow VP\ PP) * \beta_{VP}(2\text{-}5) * \beta_{PP}(5\text{-}7))$

$= 0.4 * 1 * 0.015 + 0.6 * 0.1 * 0.3$

$= 0.024$

$\boldsymbol{\beta_{S}(0\text{-}7) = P(S \rightarrow NP\ VP) * \beta_{NP}(0\text{-}2) * \beta_{VP}(2\text{-}7)}$

$\boldsymbol{= 1 * 0.25 * 0.024}$

$\boldsymbol{= 0.006}$

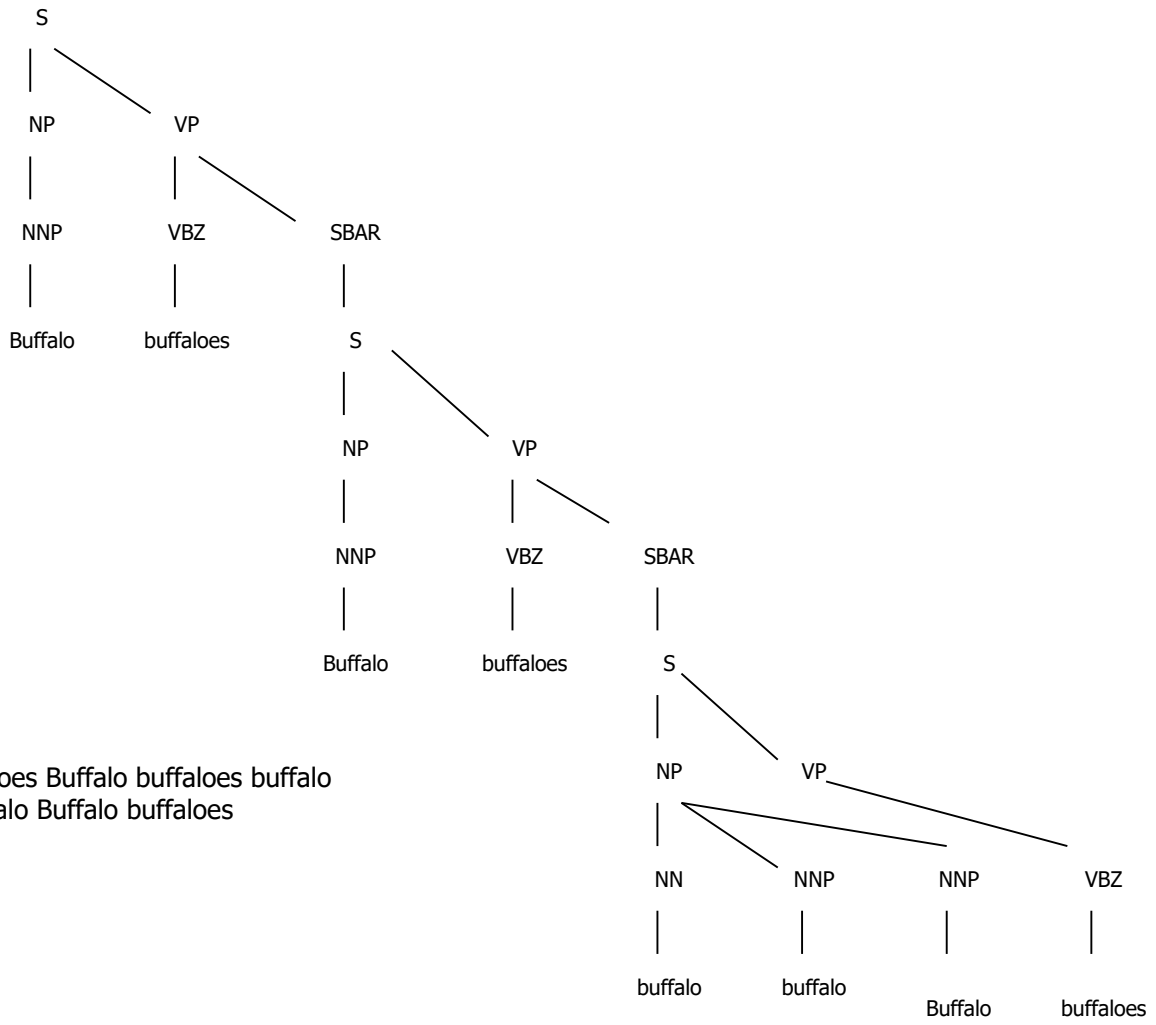
A very difficult parsing situation!

Repeated Word handling

Sentence on Buffaloes!

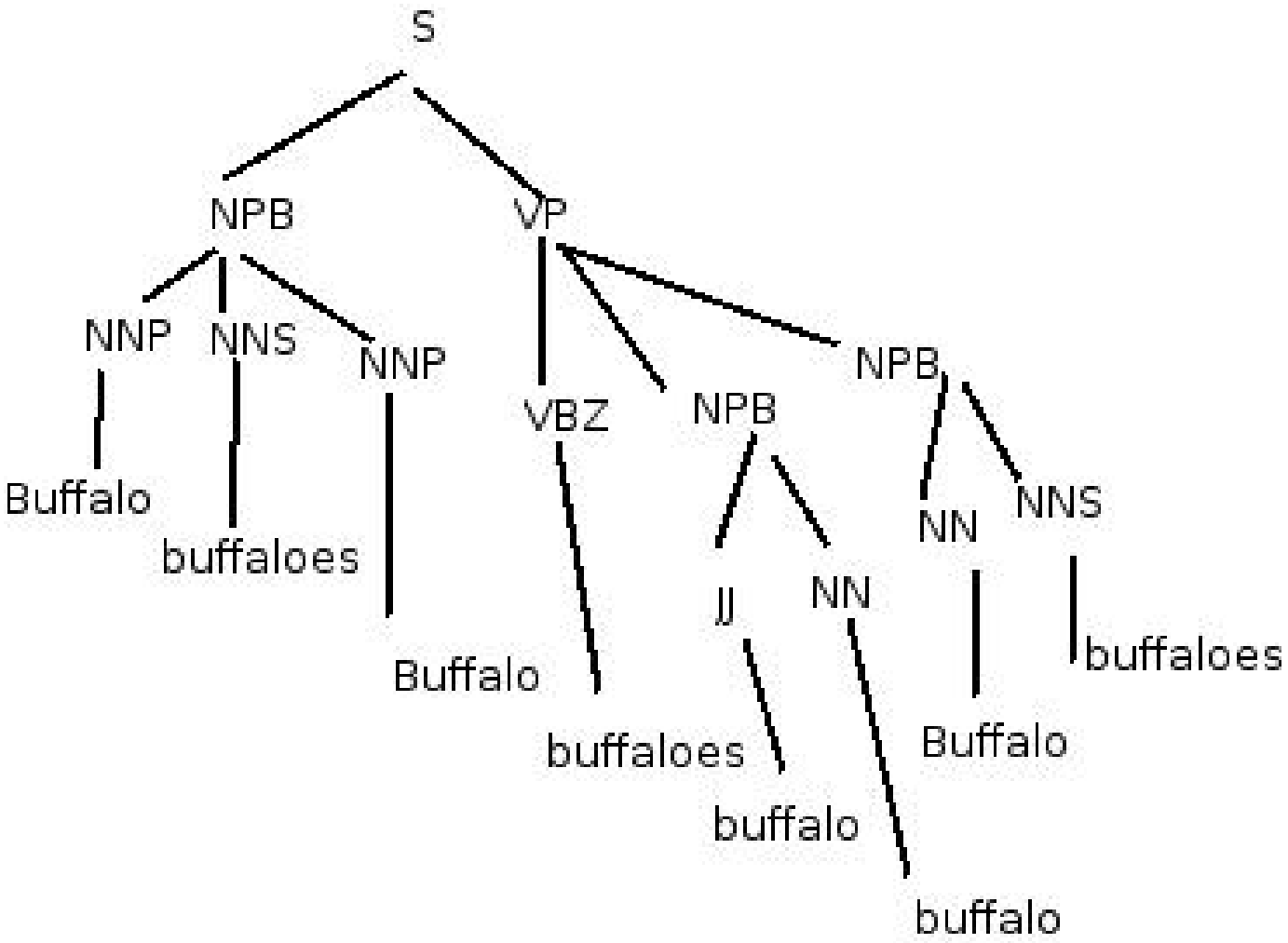
Buffaloe buffaloes Buffaloe buffaloes buffaloe buffaloe Buffaloe buffaloes

Charniak

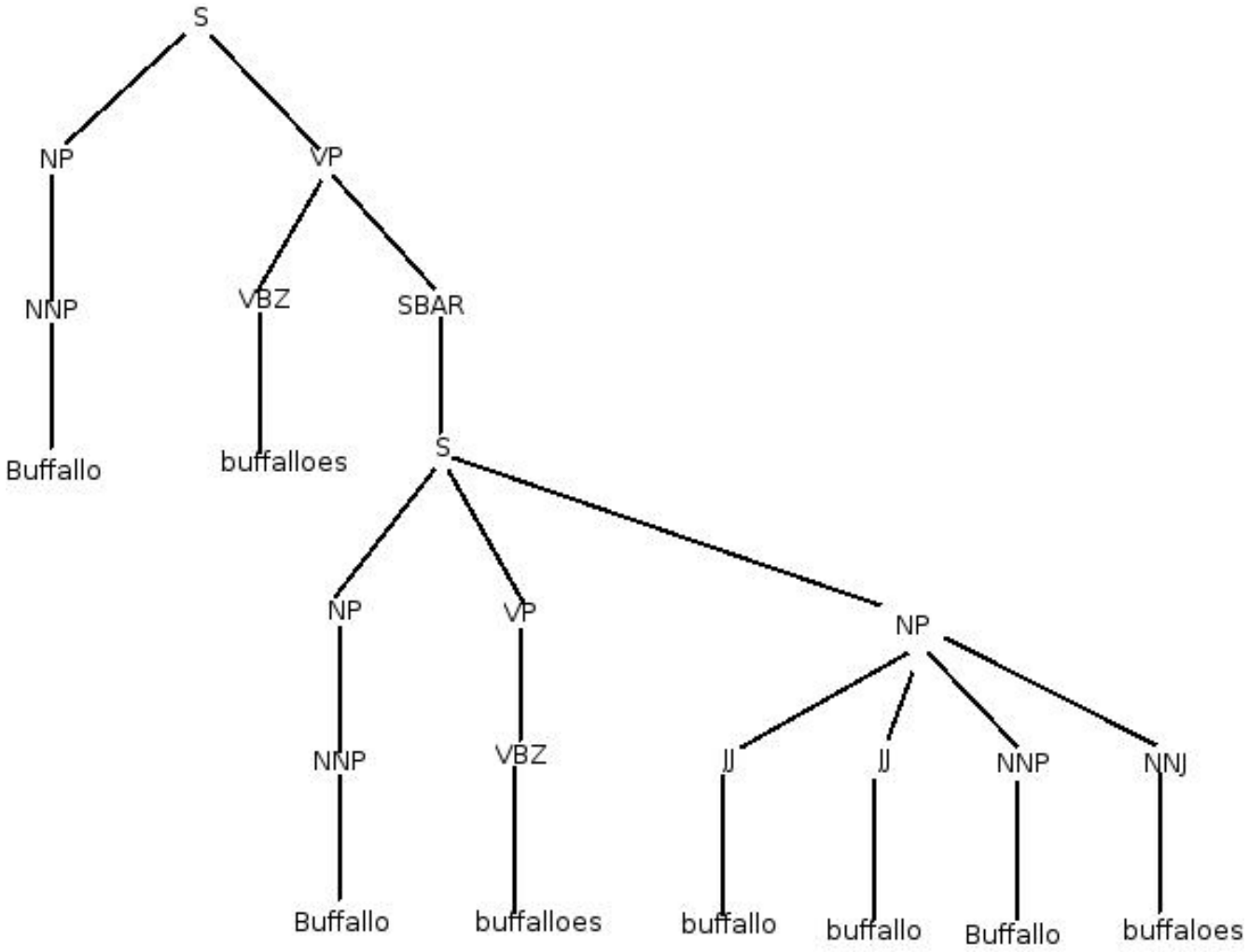


Buffalo buffaloes Buffalo buffaloes buffalo buffalo Buffalo buffaloes

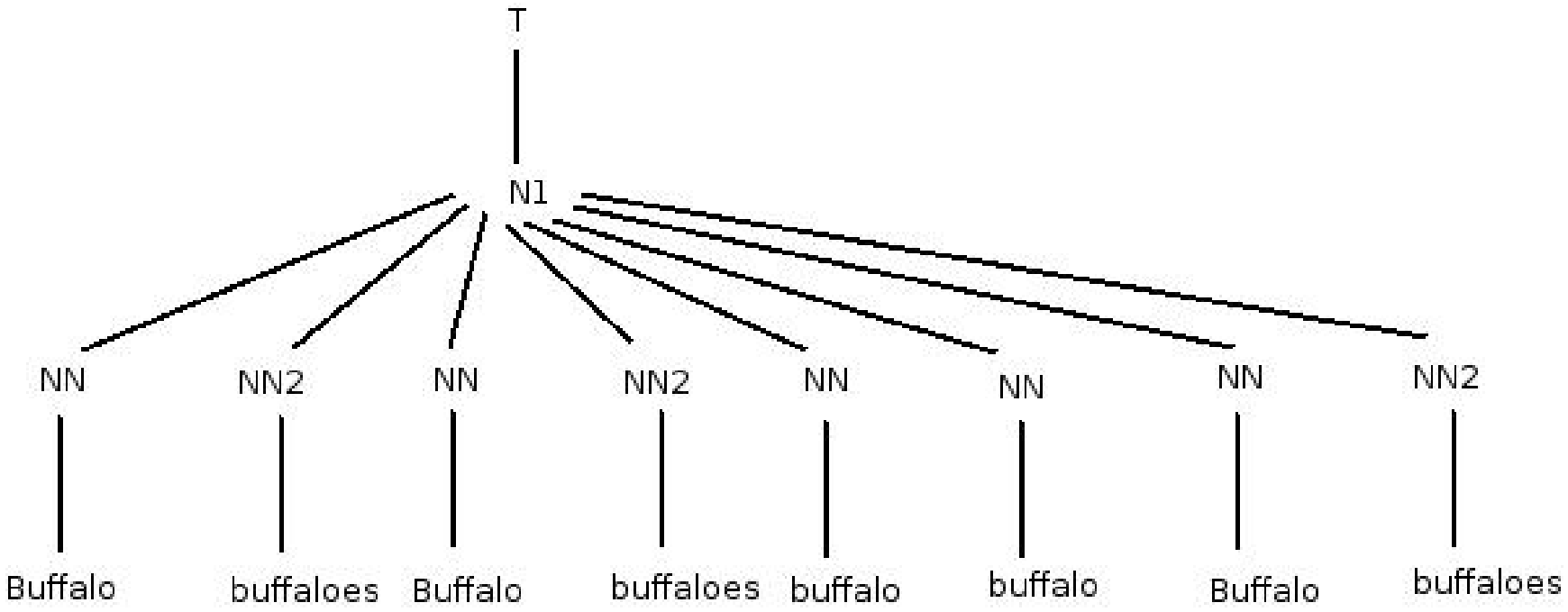
Collins


S
NPB
VP
NNP
NNS
NNP
VBZ
NPB
NPB
Buffalo
buffaloes
Buffalo
buffaloes
JJ
NN
NN
NNS
buffaloes
Buffalo
buffalo
buffalo

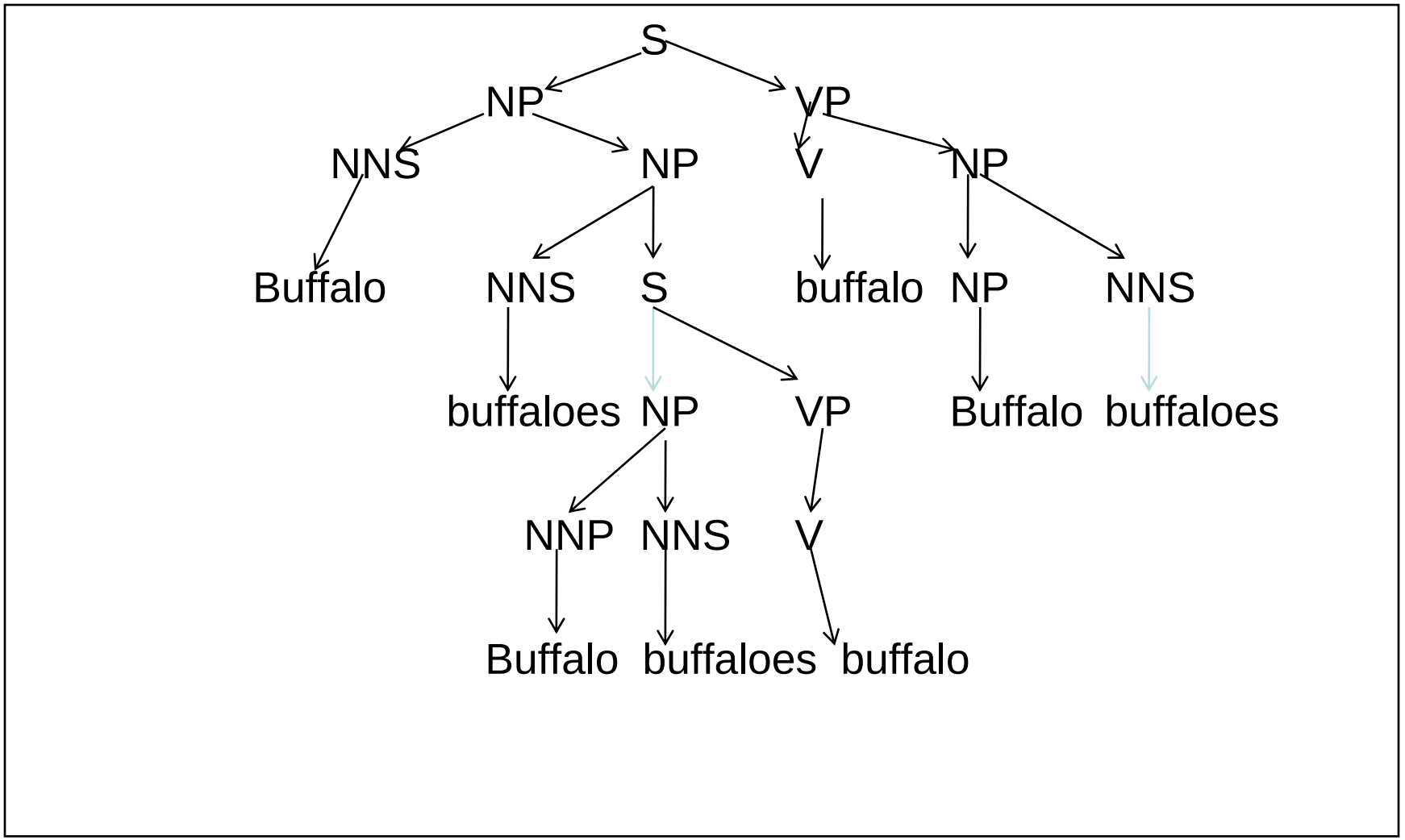
Stanford


S
NP
VP
NNP
VBZ
SBAR
Buffallo
buffalloes
S
NP
VP
NP
NNP
VBZ
JJ
JJ
NNP
NNJ
Buffallo
buffalloes
buffallo
buffallo
Buffallo
buffaloes

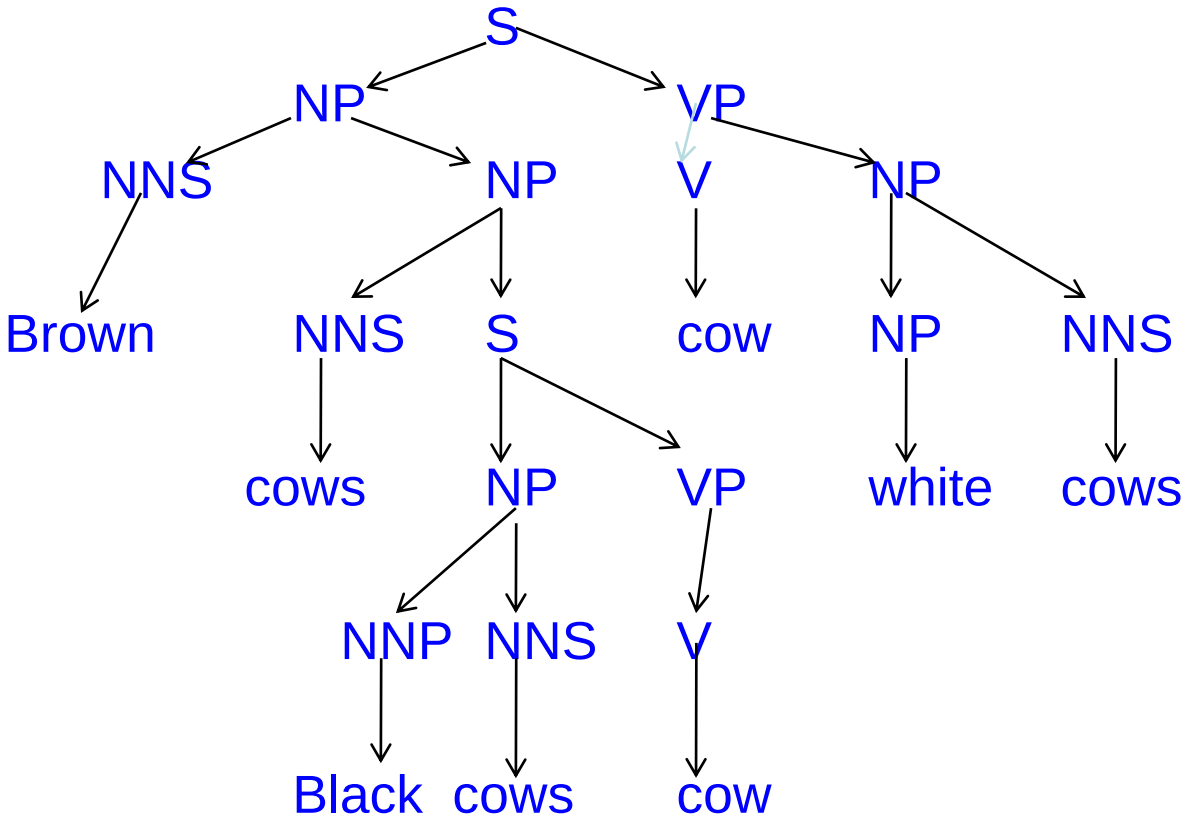
RASP


T
N1
NN
NN2
NN
NN2
NN
NN
NN
NN2
Buffalo
buffaloes
Buffalo
buffaloes
buffalo
buffalo
Buffalo
buffaloes

Correct parse



Another sentence of same structure

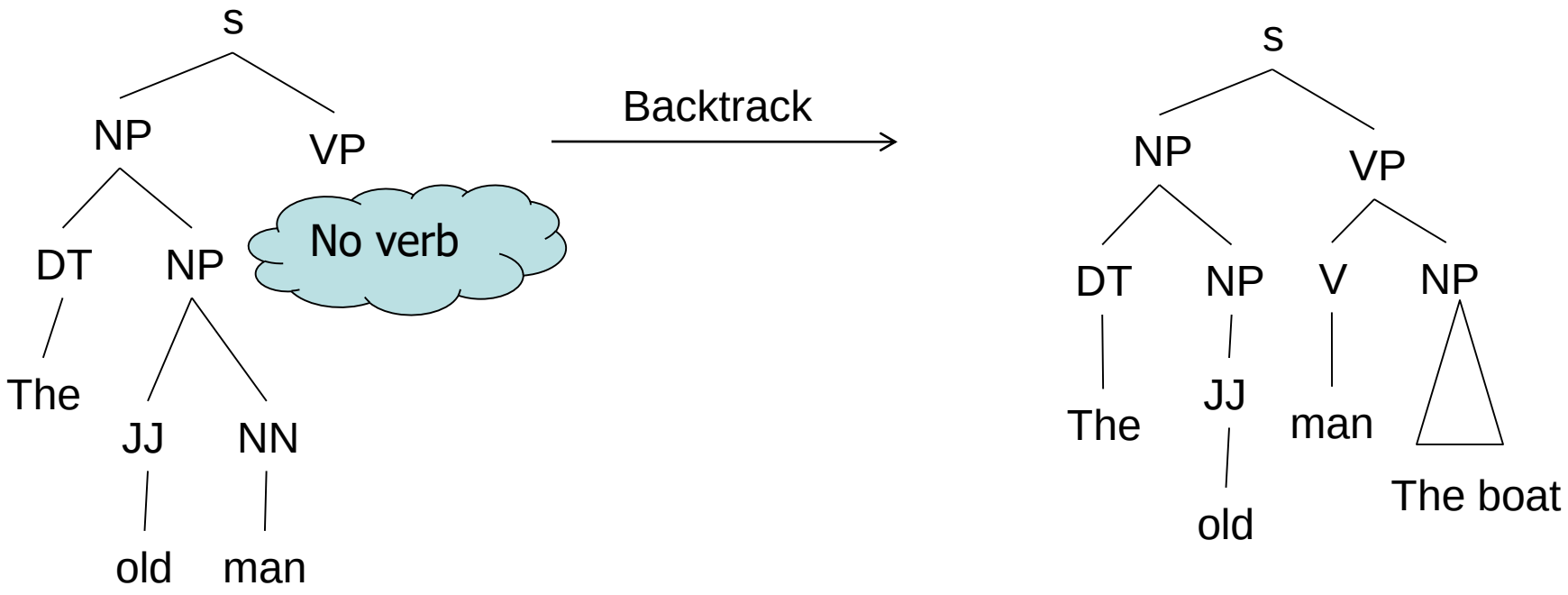


Observation

- Collins and Charniak come close to producing the correct parse.
- RASP tags all the words as nouns.

Another phenomenon: Garden pathing

e.g. The old man the boat.



Another example: The horse raced past the garden fell.