INFERNO.

El mezzo del camin di nostra uita
Mi ritrouai per una selua oscura;
Che la diritta uia era smarrita:
Et quanto a dir qual era, è cosa dura
Esta selua seluaggia et aspra et forte;
Che nel pensier rinuoua la paura.

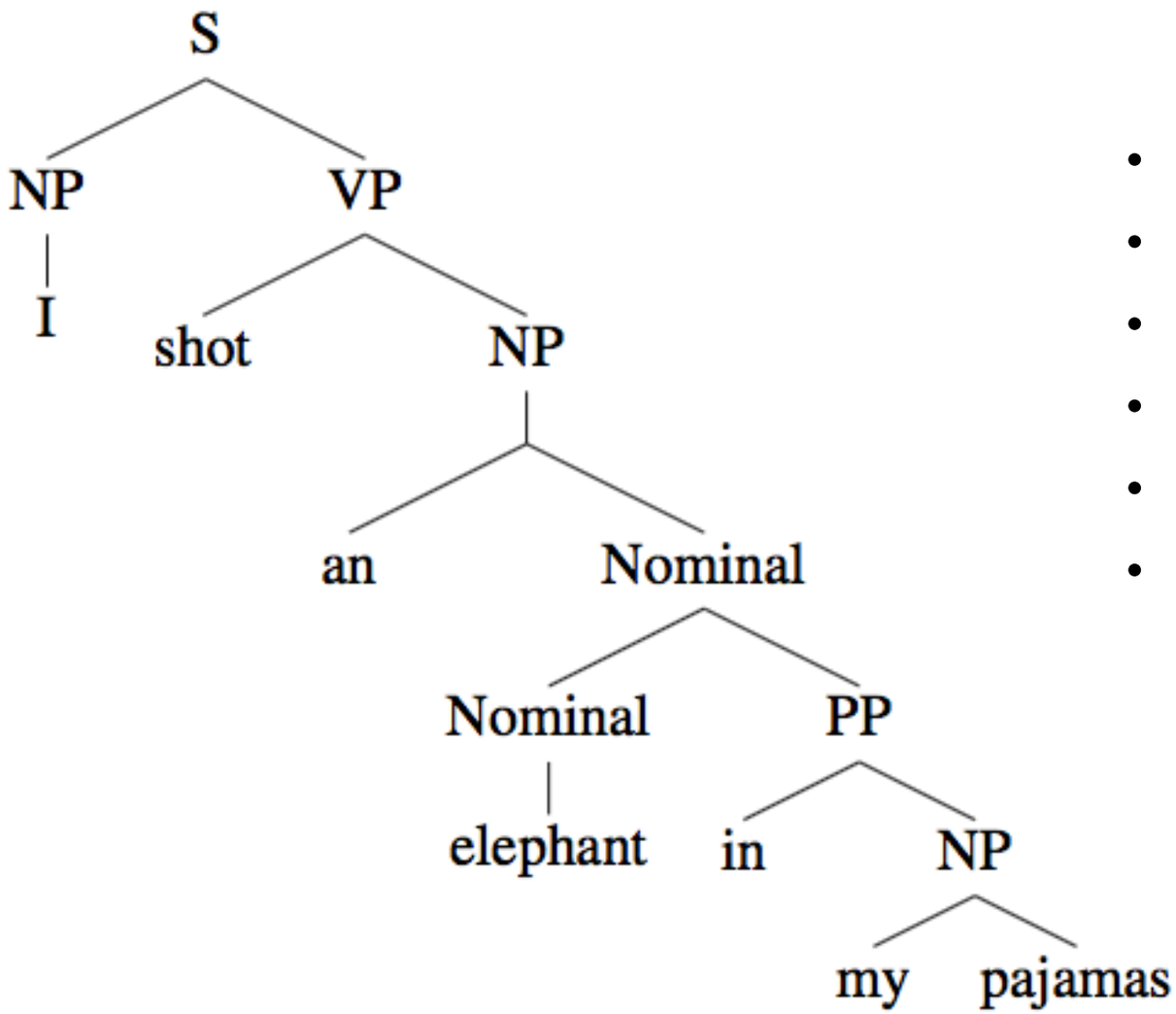
# Natural Language Processing

Info 159/259

Lecture 14: Phrase-structure parsing (Oct 9, 2018)

David Bamman, UC Berkeley – Guest Lecturer David Gaddy

# Constituents




*Every* internal node is a phrase

- my pajamas
- in my pajamas
- elephant in my pajamas
- an elephant in my pajamas
- shot an elephant in my pajamas
- I shot an elephant in my pajamas

Each phrase could be replaced by another of the same type of constituent

# Context-free grammar

A context-free grammar defines how symbols in a language combine to form valid structures

| | | |
|---|---|---|
| NP | → | Det Nominal |
| NP | → | ProperNoun |
| Nominal | → | Noun \| Nominal Noun |
| Det | → | a \| the |
| Noun | → | flight |

non-terminals

lexicon/terminals

# Context-free grammar

| | | |
|---|---|---|
| *N* | Finite set of non-terminal symbols | NP, VP, S |
| Σ | Finite alphabet of terminal symbols | the, dog, a |
| *R* | Set of production rules, each<br>$A \rightarrow \beta$<br>$\beta \in (\Sigma, N)$ | NP → DT JJ NN<br>Noun → dog |
| *S* | Start symbol | |

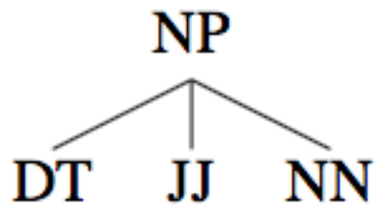
# Chomsky Normal Form (CNF)

| | | |
|---|---|---|
| *N* | Finite set of non-terminal symbols | NP, VP, S |
| Σ | Finite alphabet of terminal symbols | the, dog, a |
| *R* | Set of production rules, each<br>A → β<br>β = single terminal (from Σ) or two non-terminals (from *N*) | S → NP VP<br>Noun → dog |
| *S* | Start symbol | |

# Chomsky Normal Form (CNF)

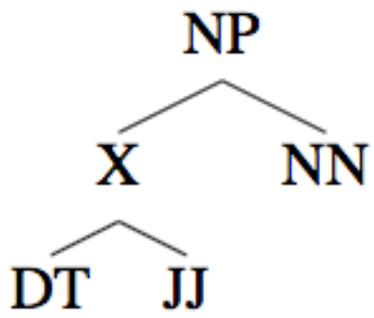
- Any CFG can be converted into weakly equivalent CNF (recognizing the same set of sentences as existing in the grammar but differing in their derivation).

NP → DT JJ NN




NP → X NN

X → DT JJ



| | | |
|---|---|---|
| S | → | NP VP |
| VP | → | VBD NP |
| VP | → | VP PP |
| Nominal | → | Nominal PP |
| Nominal | → | NN |
| Nominal | → | NNS |
| Nominal | → | PRP |
| PP | → | IN NP |
| NP | → | DT NN |
| NP | → | Nominal |
| NP | → | PRP$ Nominal |

| | | |
|---|---|---|
| VBD | → | shot |
| DT | → | an \| my |
| NN | → | elephant |
| NNS | → | pajamas |
| PRP | → | I |
| PRP$ | → | my |
| IN | → | in |

I shot an elephant in my pajamas

| | | |
|---|---|---|
| S | → | NP VP |
| VP | → | VBD NP |
| VP | → | VP PP |
| Nominal | → | Nominal PP |
| Nominal | → | pajamas \| elephant \| I |
| PP | → | IN NP |
| NP | → | DT NN |
| NP | → | pajamas \| elephant \| I |
| NP | → | PRP$ Nominal |

| | | |
|---|---|---|
| VBD | → | shot |
| DT | → | an \| my |
| PRP | → | I |
| PRP$ | → | my |
| IN | → | in |

I shot an elephant in my pajamas

# CKY

- Cocke-Kasami-Younger algorithm (also CYK) for parsing from a grammar expressed in CNF.
  - Kasami (1965)
  - Younger (1967)
  - Cocke and Schwartz (1970)

- Bottom-up dynamic programming

# CKY

- The subproblems for our dynamic program are all possible spans of contiguous words in the sentence.

I
shot
an
elephant
in
my
pajamas
0
1
2
3
4
5
6
7

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | | | | | | |
| | VBD [1,2] | | | | | |
| | | DT [2,3] | | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Each cell i,j keeps track of all phrase types that can be formed from *all* words from position i through position j

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | | | | | | |
| | VBD [1,2] | | | | | |
| | | DT [2,3] | | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

What phrases can be formed from “shot an elephant in”

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | | | | | | |
| | VBD [1,2] | | | | | |
| | | DT [2,3] | | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

What phrases can be formed from “I shot an elephant in my pajamas”

# CKY

- In CNF, each non-terminal generates two non-terminals

$$S \rightarrow NP\ VP$$

[S [NP I] [VP shot an elephant in my pajamas] ]

- The recursive step of our dynamic program just needs to select a split point and rule

- If the parent spans tokens i-j, then there is a k where the children span i-k and k-j.

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | | | | | | |
| | VBD [1,2] | | | | | |
| | | DT [2,3] | | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Does any rule generate PRP VBD?

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ∅ | | | | | |
| | VBD [1,2] | | | | | |
| | | DT [2,3] | | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Does any rule generate VBD DT?

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ∅ | | | | | |
| | VBD [1,2] | ∅ | | | | |
| | | DT [2,3] | | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Two possible places look for that split k

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | Ø | | | | | |
| | VBD [1,2] | Ø | | | | |
| | | DT [2,3] | | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Two possible places look for that split k

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | Ø | | | | | |
| | VBD [1,2] | Ø | | | | |
| | | DT [2,3] | | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Two possible places look for that split k

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | | | | |
| | VBD [1,2] | ø | | | | |
| | | DT [2,3] | | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Does any rule generate DT NN?

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ∅ | ∅ | | | | |
| | VBD [1,2] | ∅ | | | | |
| | | DT [2,3] | NP [2,4] | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Two possible places look for that split k

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ∅ | ∅ | | | | |
| | VBD [1,2] | ∅ | | | | |
| | | DT [2,3] | NP [2,4] | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Two possible places look for that split k

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ∅ | ∅ | | | | |
| | VBD [1,2] | ∅ | VP [1,4] | | | |
| | | DT [2,3] | NP [2,4] | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Three possible places look for that split k

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ∅ | ∅ | | | | |
| | VBD [1,2] | ∅ | VP [1,4] | | | |
| | | DT [2,3] | NP [2,4] | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Three possible places look for that split k

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ∅ | ∅ | | | | |
| | VBD [1,2] | ∅ | VP [1,4] | | | |
| | | DT [2,3] | NP [2,4] | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Three possible places look for that split k

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ∅ | ∅ | | | | |
| | VBD [1,2] | ∅ | VP [1,4] | | | |
| | | DT [2,3] | NP [2,4] | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

Three possible places look for that split k

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | | | |
| | VBD [1,2] | ø | VP [1,4] | | | |
| | | DT [2,3] | NP [2,4] | | | |
| | | | NP, NN [3,4] | | | |
| | | | | IN [4,5] | | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | |
| | | | NP, NN [3,4] | ø | ø | |
| | | | | IN [4,5] | ø | |
| | | | | | PRP$ [5,6] | |
| | | | | | | NNS [6,7] |

*elephant in
*an elephant in
*shot an elephant in
*I shot an elephant in

*in my
*elephant in my
*an elephant in my
*shot an elephant in my
*I shot an elephant in my

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | |
| | | | NP, NN [3,4] | ø | ø | |
| | | | | IN [4,5] | ø | |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |

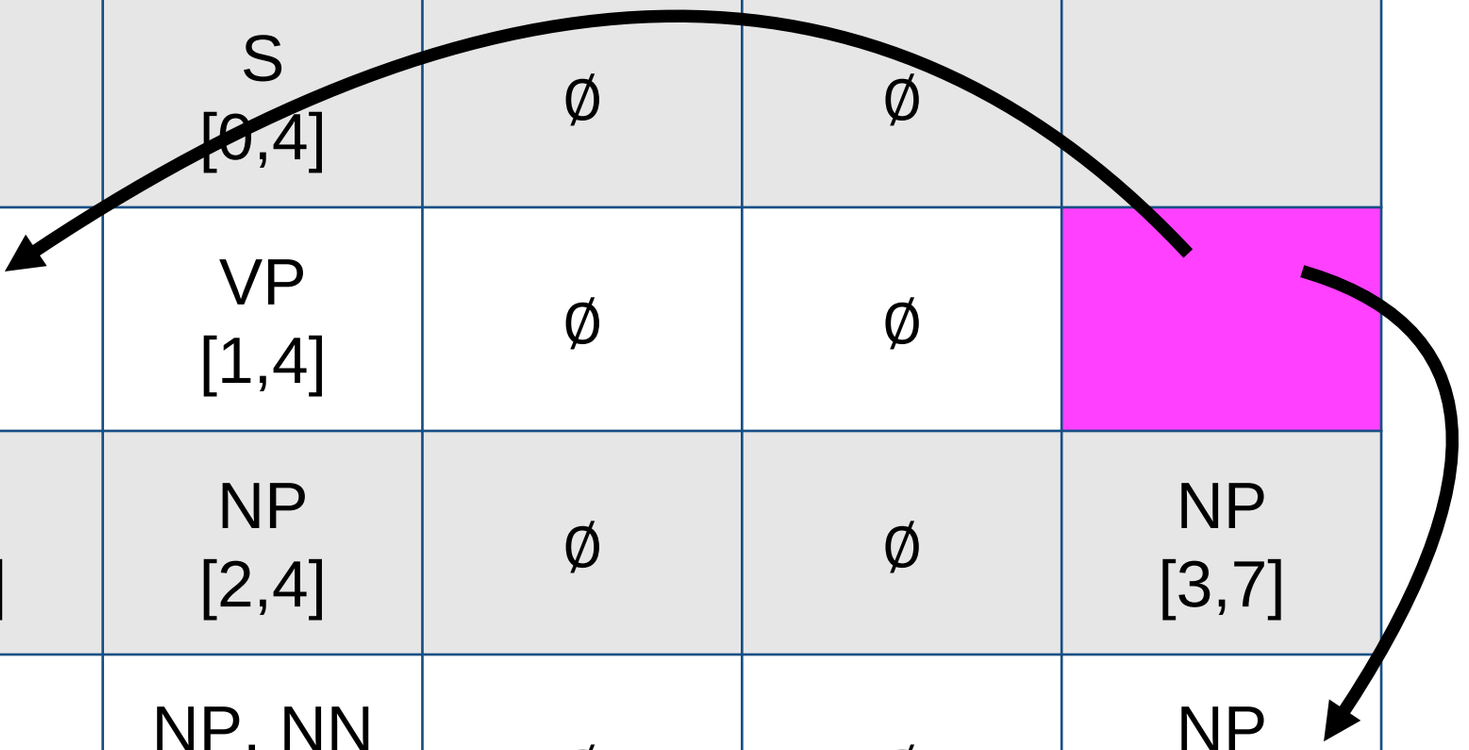
| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | |
| | | | NP, NN [3,4] | ø | ø | |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | |
| | | | NP, NN [3,4] | ø | ø | NP [3,7] |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |

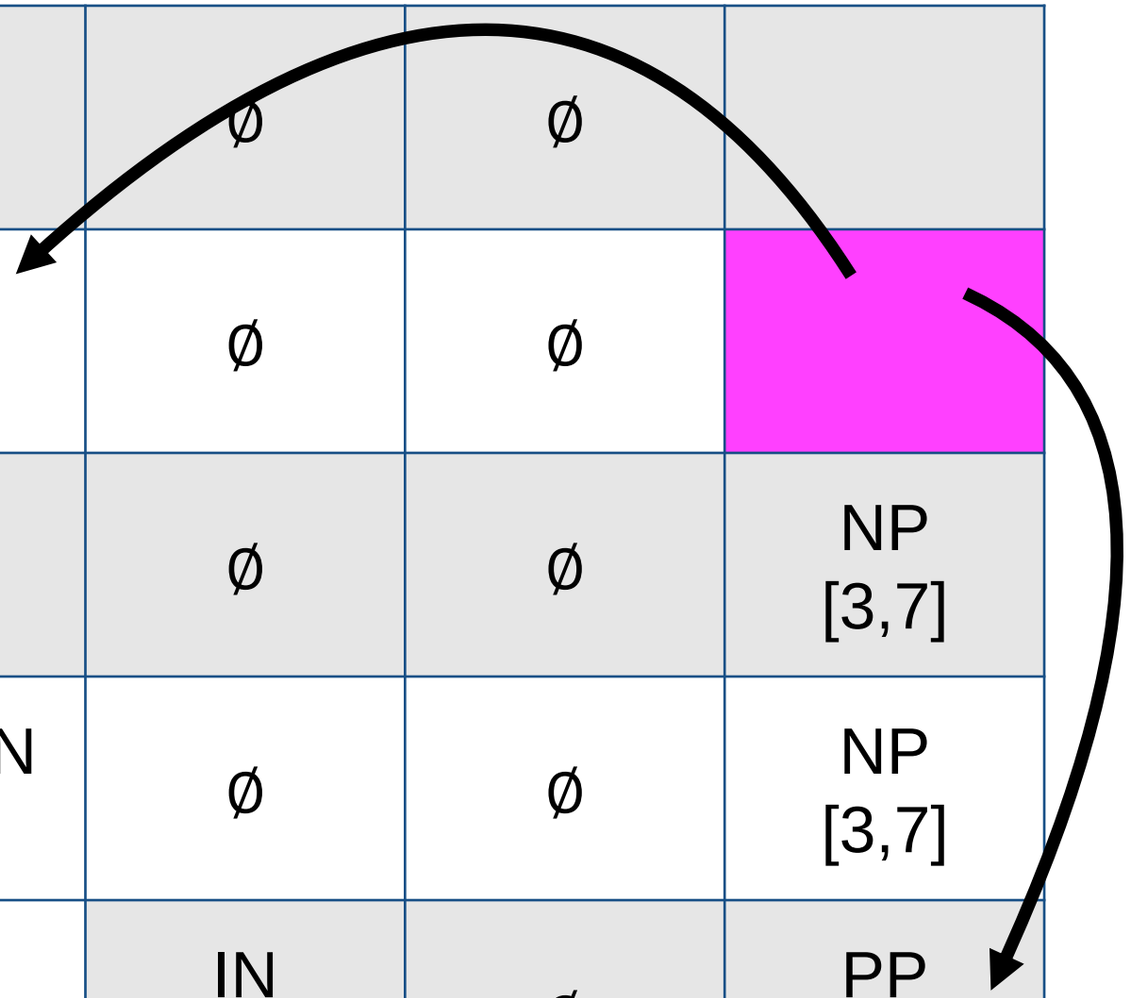
| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | NP [3,7] |
| | | | NP, NN [3,4] | ø | ø | NP [3,7] |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | NP [3,7] |
| | | | NP, NN [3,4] | ø | ø | NP [3,7] |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |

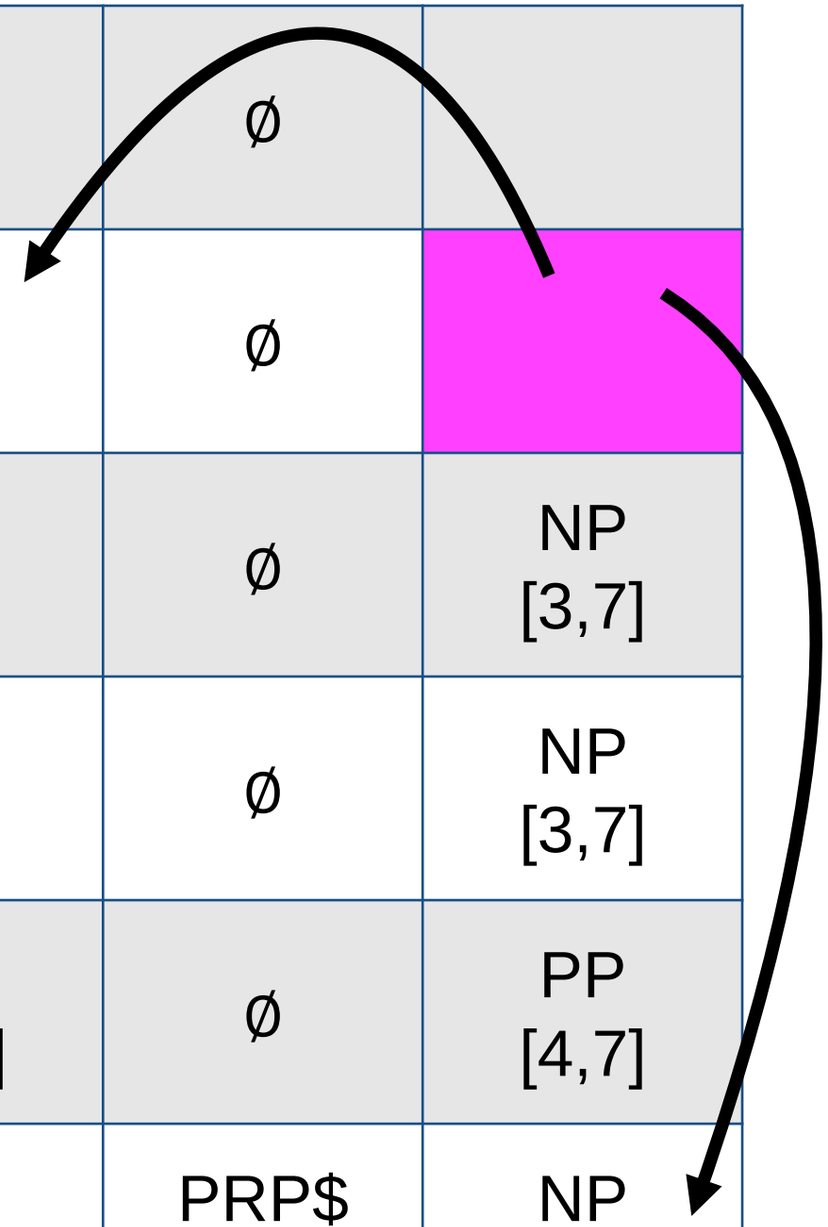
| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | NP [3,7] |
| | | | NP, NN [3,4] | ø | ø | NP [3,7] |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |



| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | NP [3,7] |
| | | | NP, NN [3,4] | ø | ø | NP [3,7] |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |



| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | NP [3,7] |
| | | | NP, NN [3,4] | ø | ø | NP [3,7] |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |



| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | NP [3,7] |
| | | | NP, NN [3,4] | ø | ø | NP [3,7] |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | |
| | | DT [2,3] | NP [2,4] | ø | ø | NP [3,7] |
| | | | NP, NN [3,4] | ø | ø | NP [3,7] |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | $VP_1$, $VP_2$ [1,7] |
| | | DT [2,3] | NP [2,4] | ø | ø | NP [2,7] |
| | | NP, NN [3,4] | ø | ø | NP [3,7] | |
| | | | IN [4,5] | ø | PP [4,7] | |
| | | | | PRP$ [5,6] | NP [5,7] | |
| | | | | | NNS [6,7] | |

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ∅ | ∅ | S [0,4] | ∅ | ∅ | |
| | VBD [1,2] | ∅ | VP [1,4] | ∅ | ∅ | $VP_1$, $VP_2$ [1,7] |
| | | DT [2,3] | NP [2,4] | ∅ | ∅ | NP [2,7] |
| | | | NP, NN [3,4] | ∅ | ∅ | NP [3,7] |
| | | | | IN [4,5] | ∅ | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |

Possibilities:

$S_1 \rightarrow NP\ VP_1$
$S_2 \rightarrow NP\ VP_2$
$? \rightarrow S\ PP$
$? \rightarrow PRP\ VP_1$
$? \rightarrow PRP\ VP_2$

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|

| | | | | | | |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | $S_1$, $S_2$ [0,7] |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | $VP_1$, $VP_2$ [1,7] |
| | | DT [2,3] | NP [2,4] | ø | ø | NP [2,7] |
| | | | NP, NN [3,4] | ø | ø | NP [3,7] |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |

Success! We've recognized a total of two valid parses

# CKY algorithm

```
function CKY-PARSE(words, grammar) returns table

  for j ← from 1 to LENGTH(words) do
      for all {A | A → words[j] ∈ grammar}
              table[j − 1, j] ← table[j − 1, j] ∪ A
      for i ← from j − 2 downto 0 do
          for k ← i + 1 to j − 1 do
              for all {A | A → BC ∈ grammar and B ∈ table[i, k] and C ∈ table[k, j]}
                      table[i,j] ← table[i,j] ∪ A
```

**Figure 12.5** The CKY algorithm.

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| NP, PRP [0,1] | ø | ø | S [0,4] | ø | ø | $S_1$, $S_2$ [0,7] |
| | VBD [1,2] | ø | VP [1,4] | ø | ø | $VP_1$, $VP_2$ [1,7] |
| | | DT [2,3] | NP [2,4] | ø | ø | NP [2,7] |
| | | | NP, NN [3,4] | ø | ø | NP [3,7] |
| | | | | IN [4,5] | ø | PP [4,7] |
| | | | | | PRP$ [5,6] | NP [5,7] |
| | | | | | | NNS [6,7] |

Runtime complexity?

# CFG

- This use of CKY allows us to:
    - check whether a sentence in grammatical in the language defined by the CFG
        - enumerate all possible parses for a sentence
- But it doesn’t tell us on its one which of those possible parses is most likely.

Under a full grammar, how many valid parses does an average-length sentence have?

a) 1-10

b) 10-100

c) 100-1000

d) 1000+

- Answer: **1000+!**

- Of course, most of these aren’t parses humans would even think of and are completely non-sensical

- Giving someone 1000 different possible parses usually isn’t very helpful to them – what we really want is the single most likely parse

# PCFG

- Probabilistic context-free grammar: each production is also associated with a probability.
- This lets us calculate the probability of a parse for a given sentence; for a given parse tree T for sentence S comprised of n rules from R (each A → β):

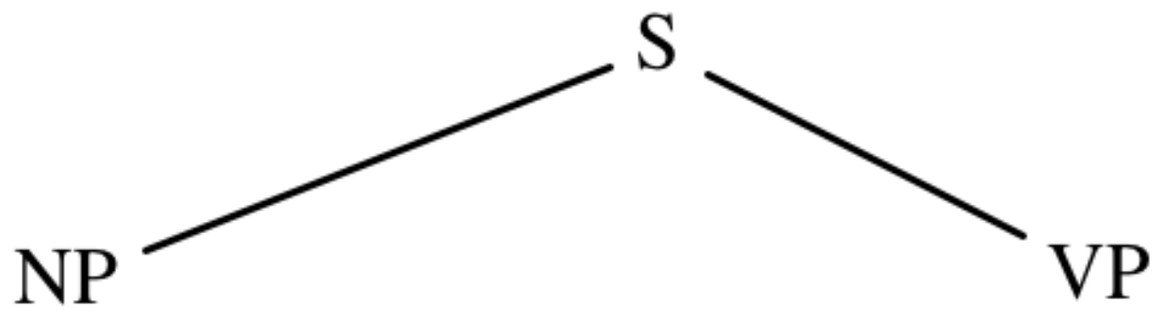
$$P(T, S) = \prod_{i}^{n} P(\beta \mid A)$$

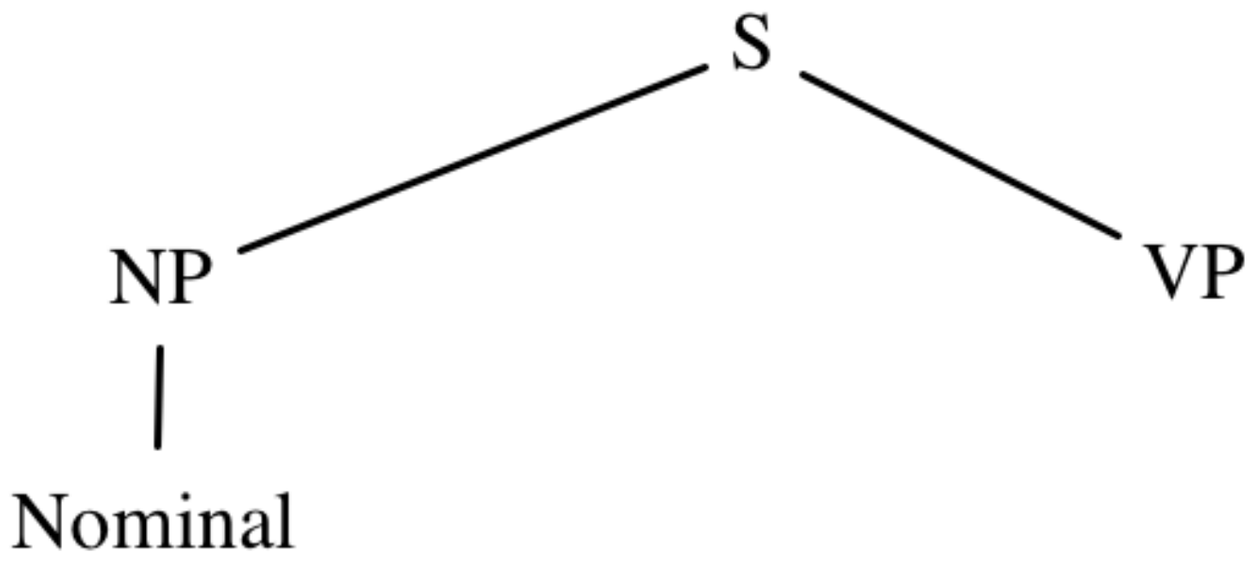
# PCFG

| | | |
|---|---|---|
| *N* | Finite set of non-terminal symbols | NP, VP, S |
| Σ | Finite alphabet of terminal symbols | the, dog, a |
| *R* | Set of production rules, each<br>$A \rightarrow \beta$ [p]<br>$p = P(\beta \mid A)$ | S → NP VP<br>Noun → dog |
| *S* | Start symbol | |

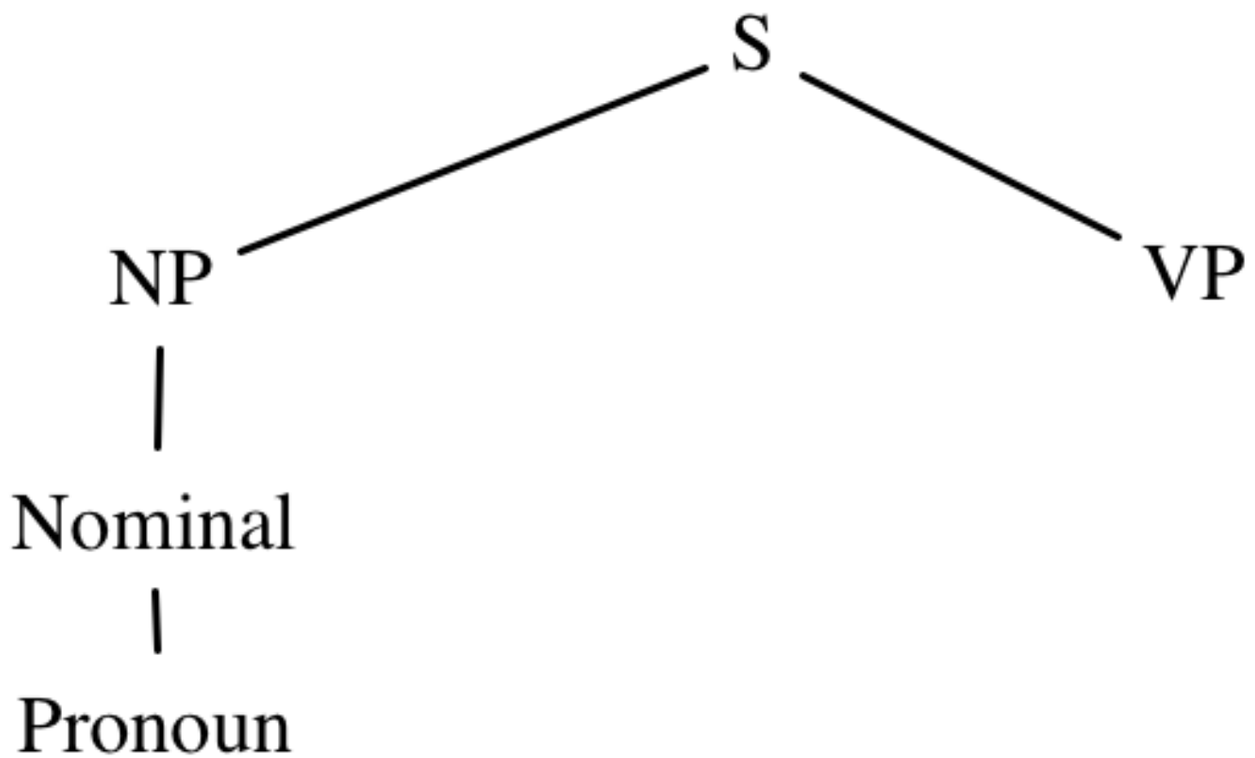
S

$P(\text{NP VP} \mid \text{S})$

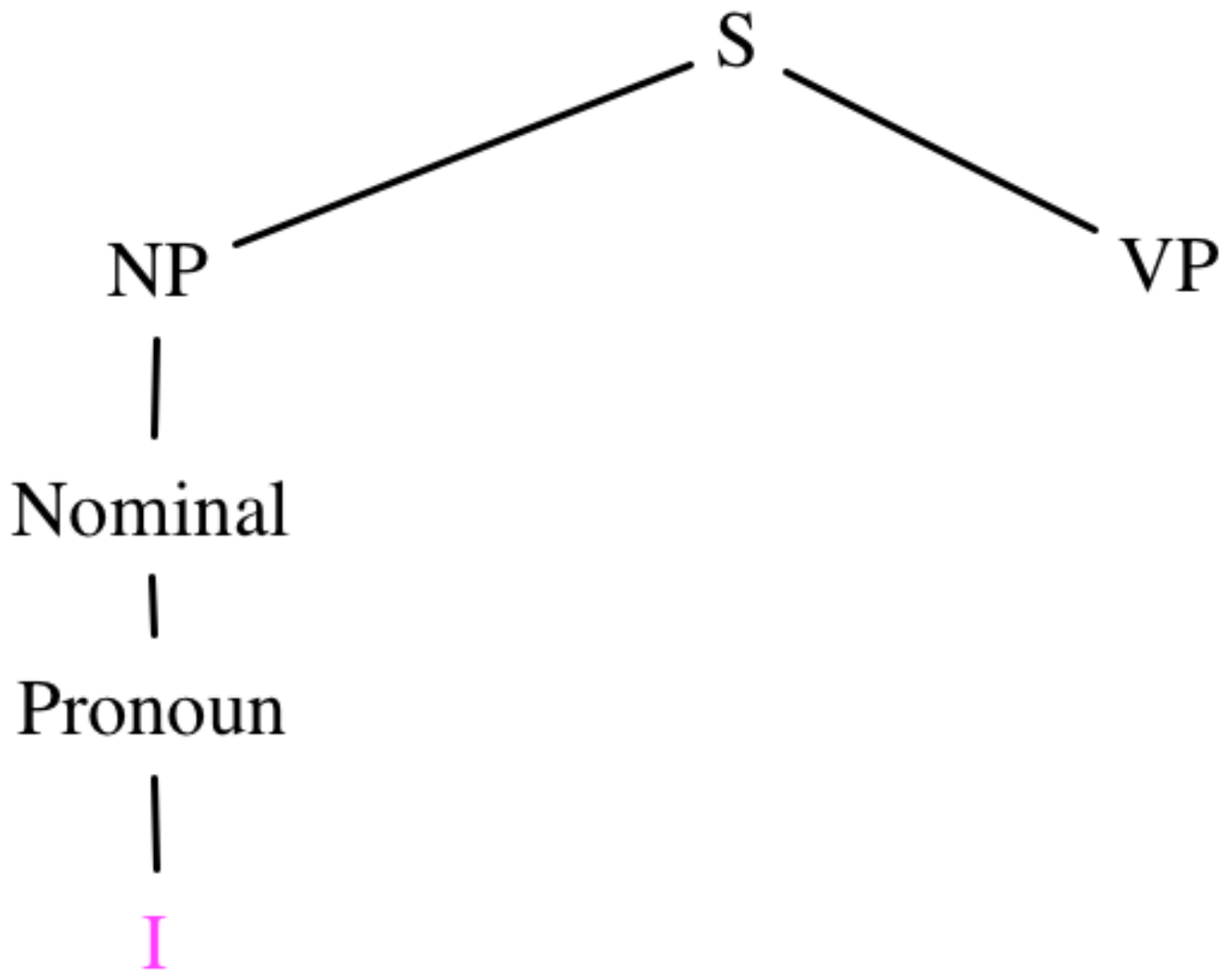




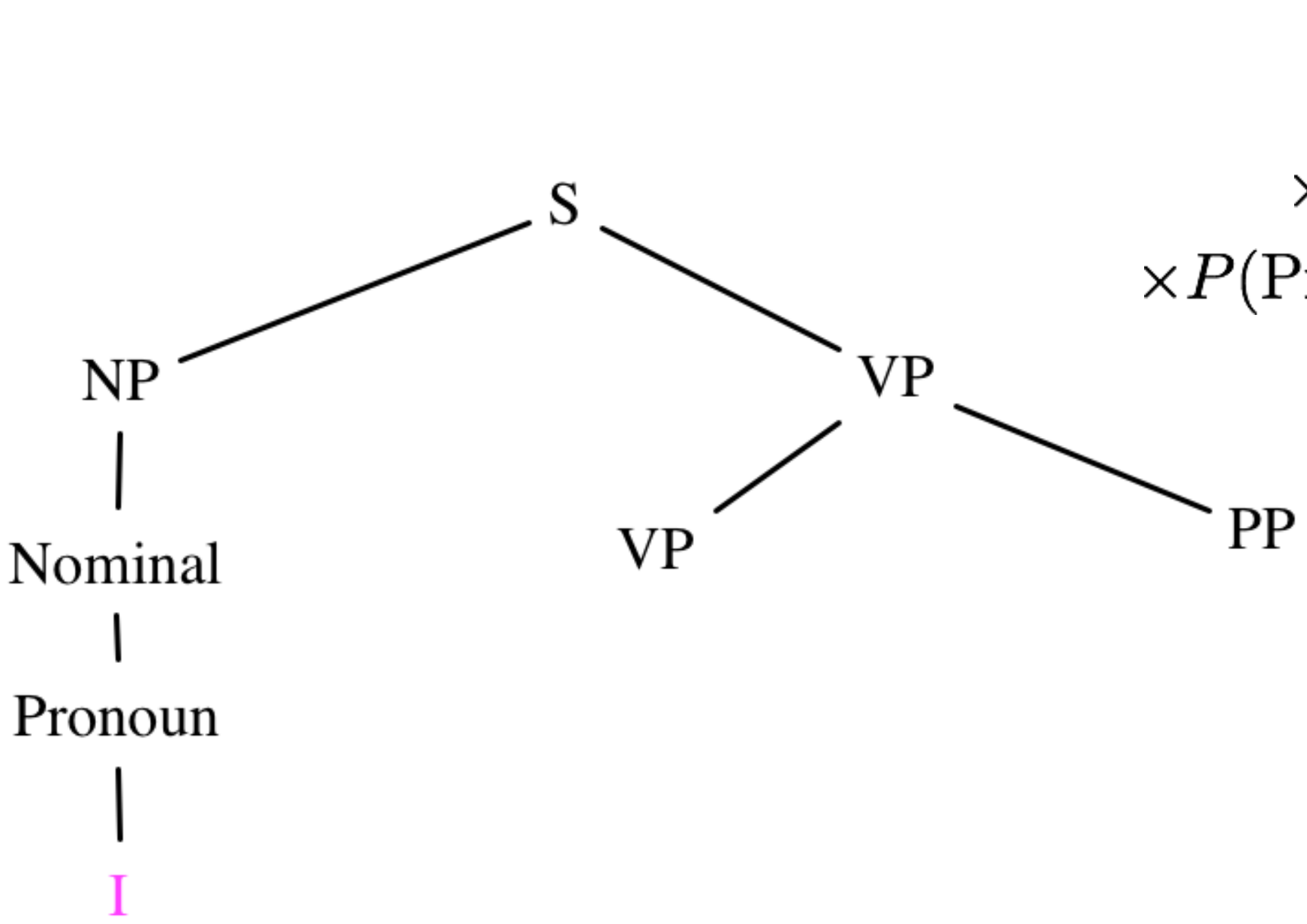

$$P(\text{NP VP} \mid \text{S})$$
$$\times P(\text{Nominal} \mid \text{NP})$$



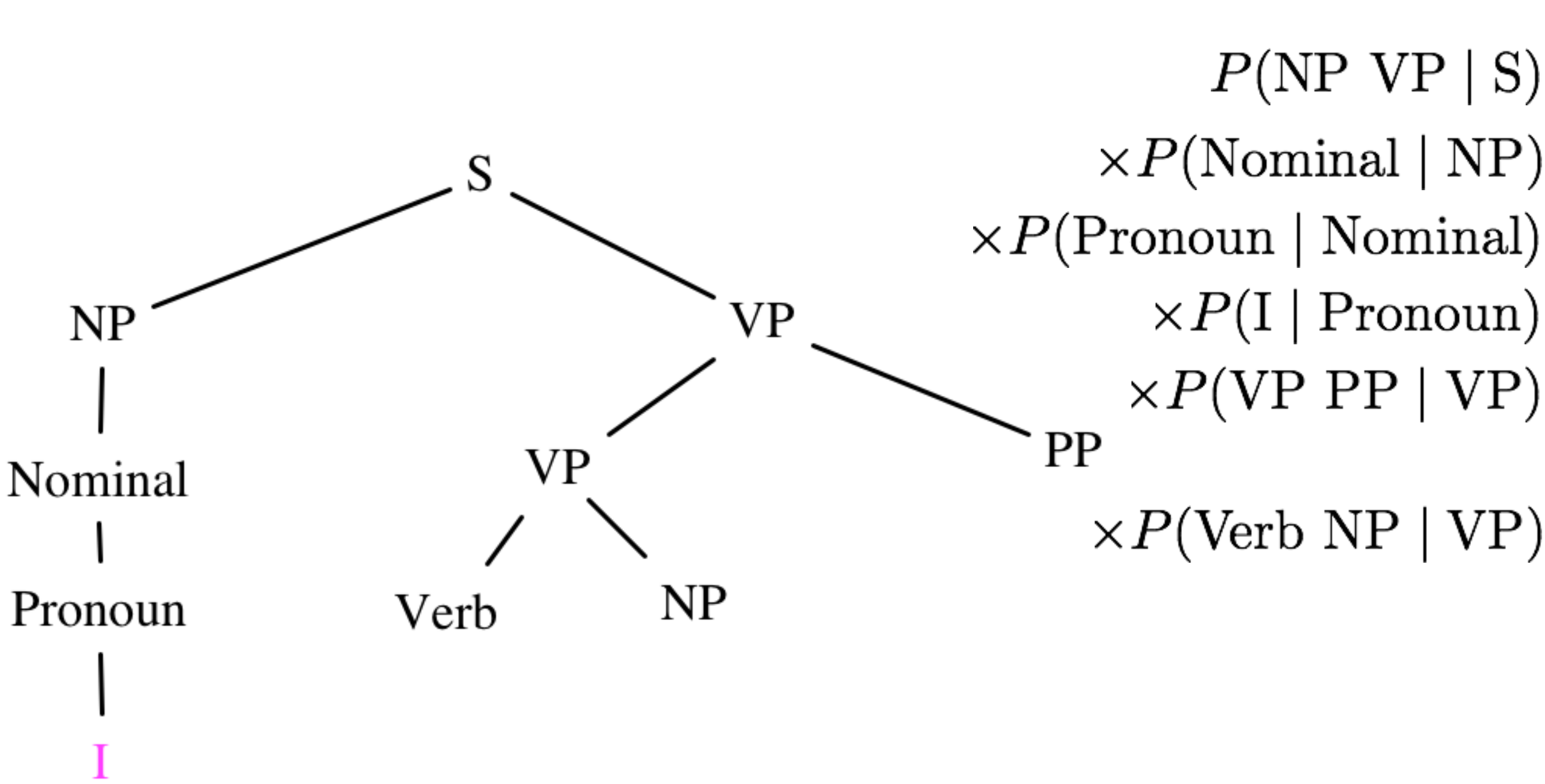

$$P(\text{NP VP} \mid \text{S})$$
$$\times P(\text{Nominal} \mid \text{NP})$$
$$\times P(\text{Pronoun} \mid \text{Nominal})$$



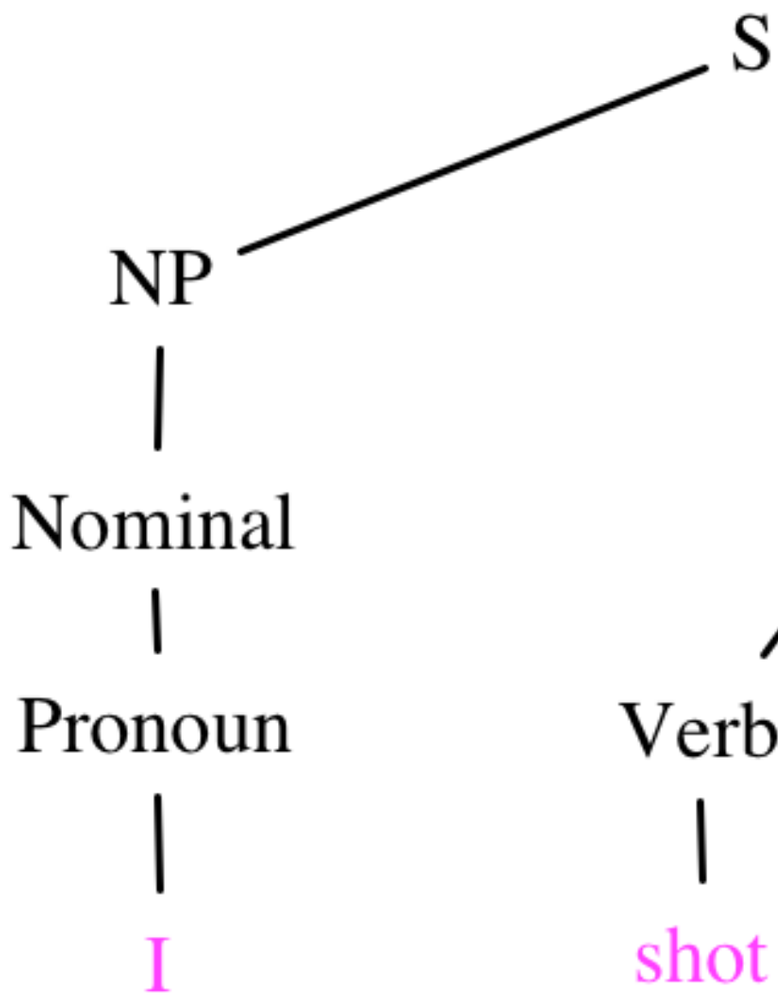

$$P(\text{NP VP} \mid \text{S})$$
$$\times P(\text{Nominal} \mid \text{NP})$$
$$\times P(\text{Pronoun} \mid \text{Nominal})$$
$$\times P(\text{I} \mid \text{Pronoun})$$



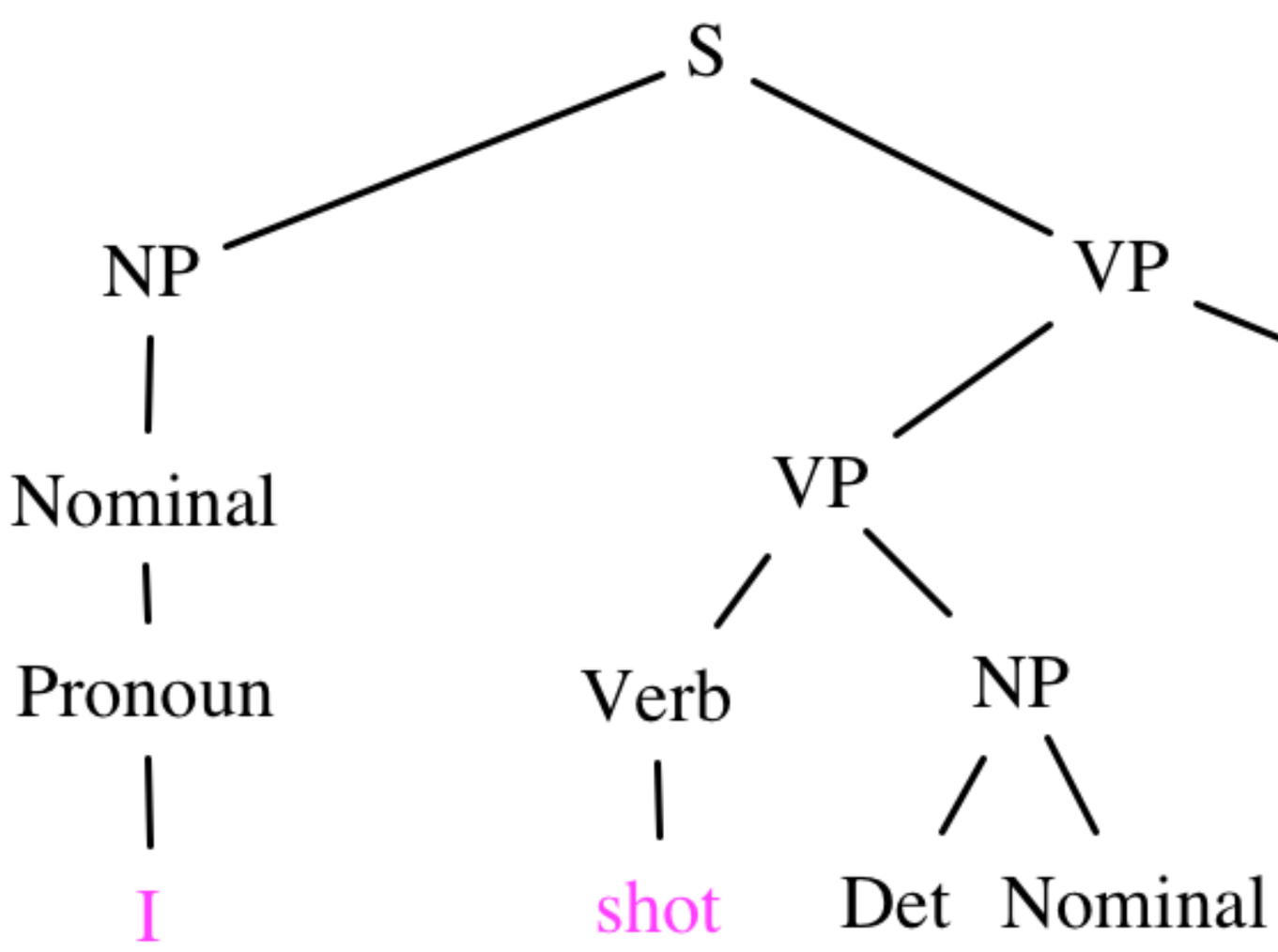

$$P(\text{NP VP} \mid \text{S})$$
$$\times P(\text{Nominal} \mid \text{NP})$$
$$\times P(\text{Pronoun} \mid \text{Nominal})$$
$$\times P(\text{I} \mid \text{Pronoun})$$
$$\times P(\text{VP PP} \mid \text{VP})$$



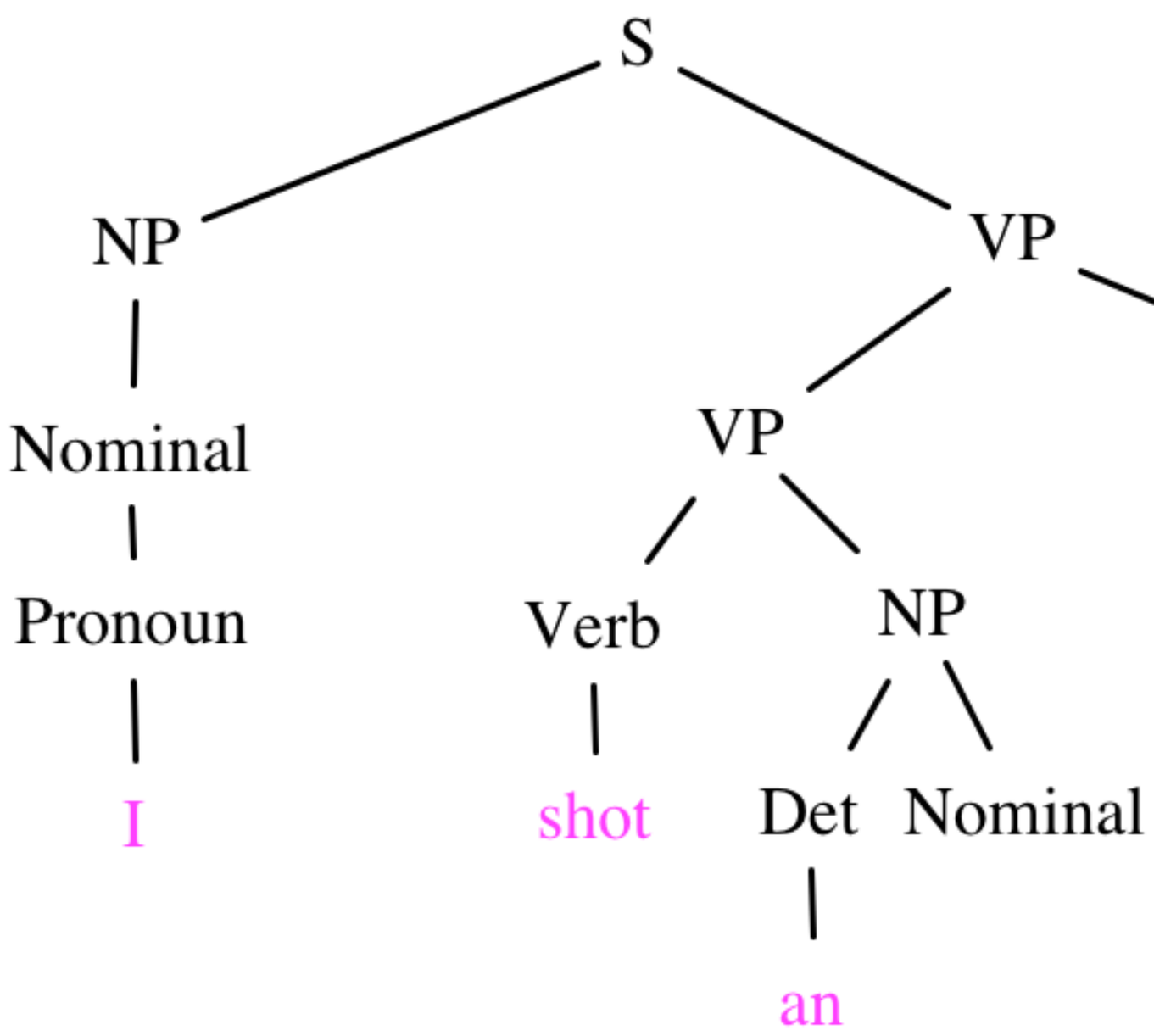

$$P(\text{NP VP} \mid \text{S})$$
$$\times P(\text{Nominal} \mid \text{NP})$$
$$\times P(\text{Pronoun} \mid \text{Nominal})$$
$$\times P(\text{I} \mid \text{Pronoun})$$
$$\times P(\text{VP PP} \mid \text{VP})$$
$$\times P(\text{Verb NP} \mid \text{VP})$$



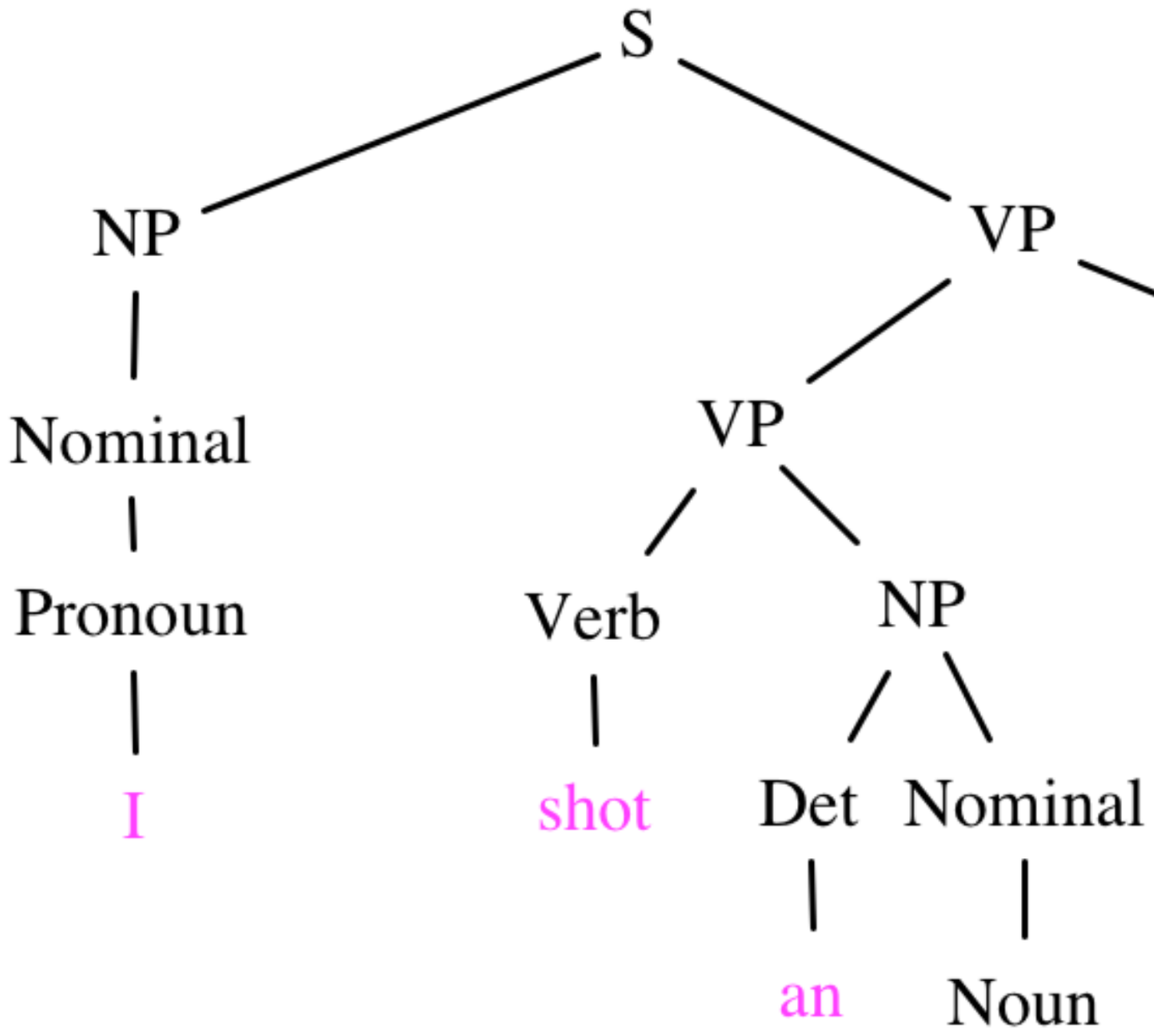

$$P(\text{NP VP} \mid \text{S})$$
$$\times P(\text{Nominal} \mid \text{NP})$$
$$\times P(\text{Pronoun} \mid \text{Nominal})$$
$$\times P(\text{I} \mid \text{Pronoun})$$
$$\times P(\text{VP PP} \mid \text{VP})$$
$$\times P(\text{Verb NP} \mid \text{VP})$$
$$\times P(\text{shot} \mid \text{Verb})$$



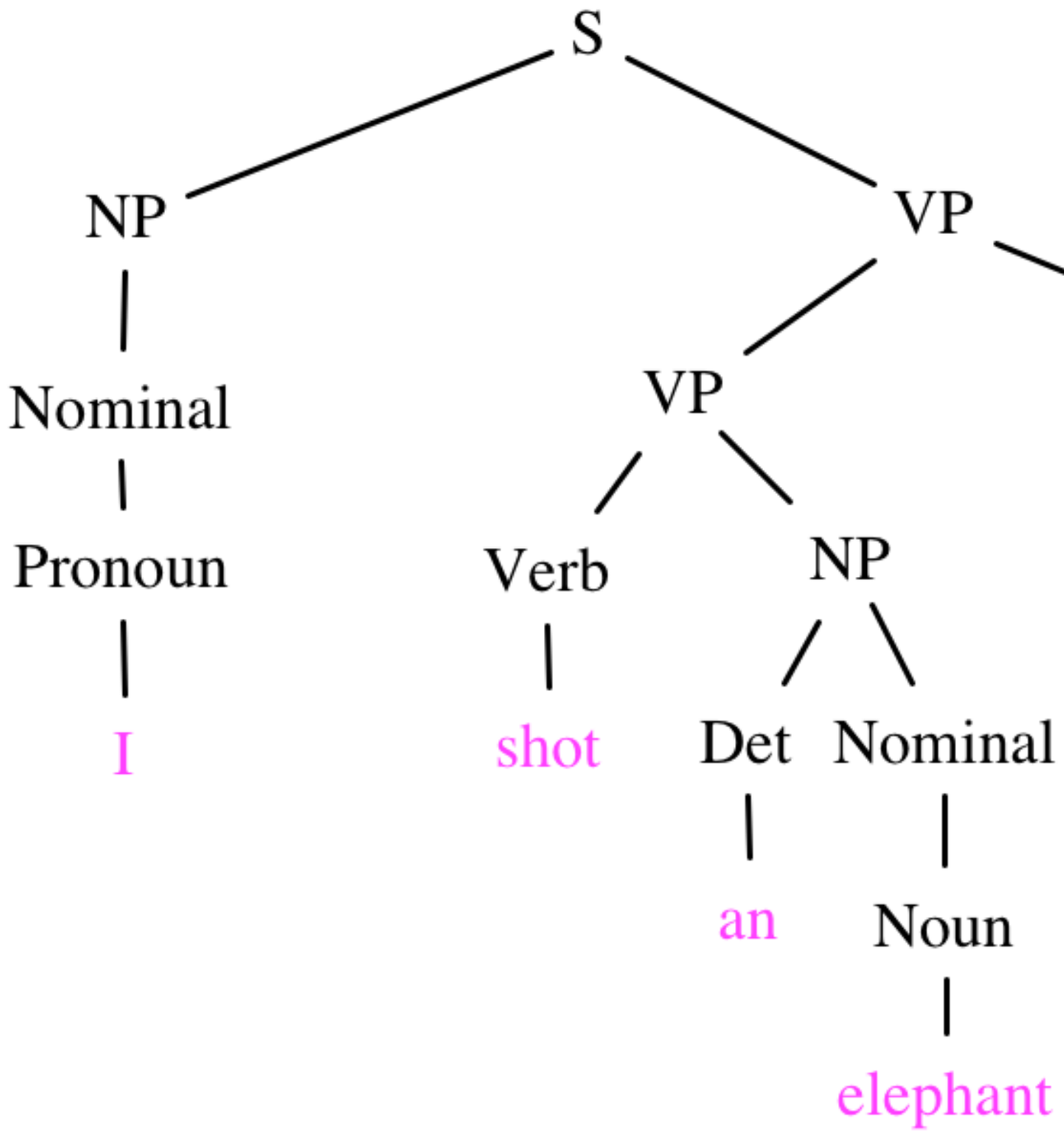

$$
\begin{aligned}
&P(\text{NP VP} \mid \text{S}) \\
\times &P(\text{Nominal} \mid \text{NP}) \\
\times &P(\text{Pronoun} \mid \text{Nominal}) \\
\times &P(\text{I} \mid \text{Pronoun}) \\
\times &P(\text{VP PP} \mid \text{VP}) \\
\times &P(\text{Verb NP} \mid \text{VP}) \\
\times &P(\text{shot} \mid \text{Verb}) \\
\times &P(\text{Det Nominal} \mid \text{NP})
\end{aligned}
$$



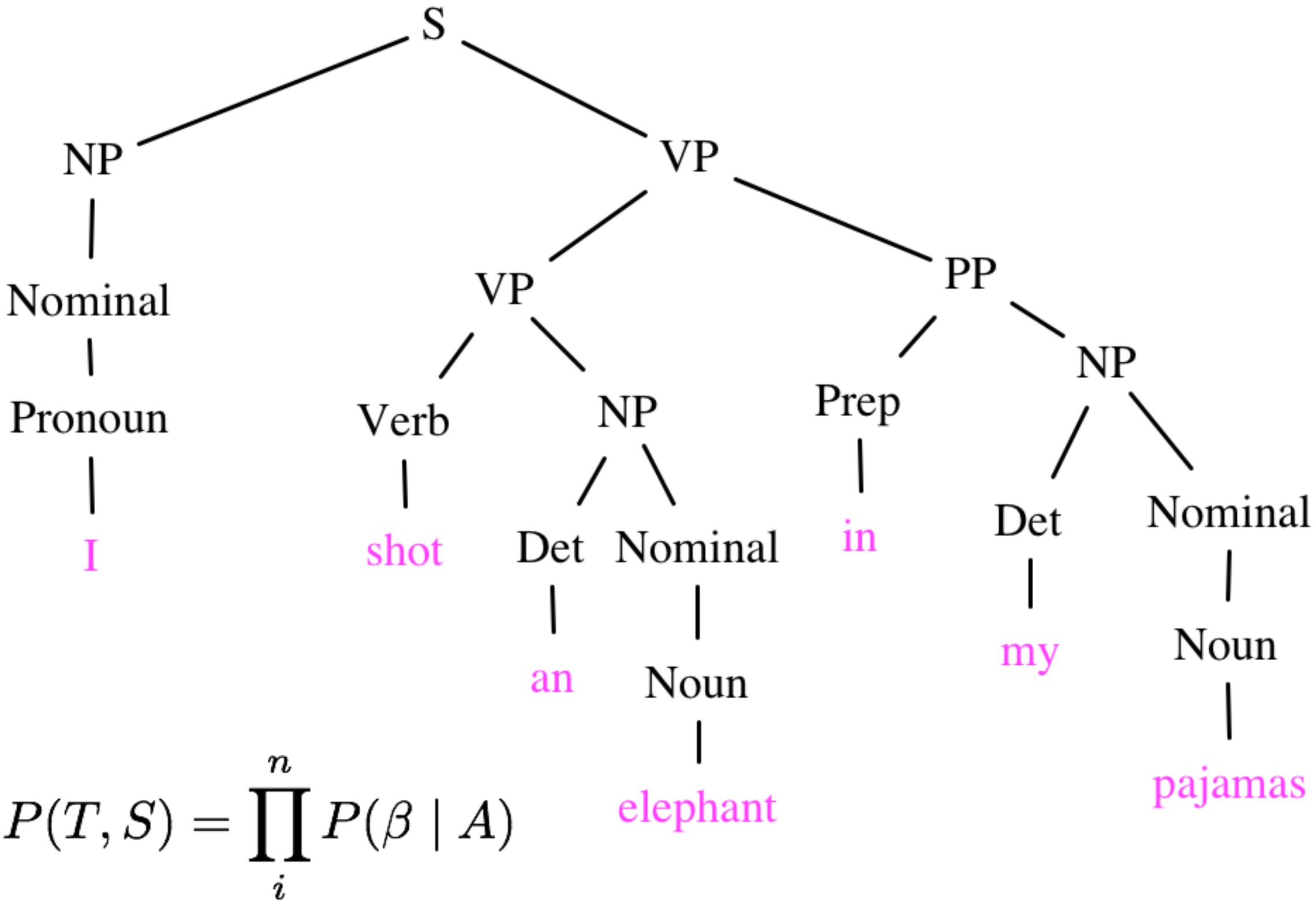

$$P(\text{NP VP} \mid \text{S})$$
$$\times P(\text{Nominal} \mid \text{NP})$$
$$\times P(\text{Pronoun} \mid \text{Nominal})$$
$$\times P(\text{I} \mid \text{Pronoun})$$
$$\times P(\text{VP PP} \mid \text{VP})$$
$$\times P(\text{Verb NP} \mid \text{VP})$$
$$\times P(\text{shot} \mid \text{Verb})$$
$$\times P(\text{Det Nominal} \mid \text{NP})$$
$$\times P(\text{an} \mid \text{Det})$$




$$
\begin{aligned}
&P(\text{NP VP} \mid \text{S}) \\
\times &P(\text{Nominal} \mid \text{NP}) \\
\times &P(\text{Pronoun} \mid \text{Nominal}) \\
\times &P(\text{I} \mid \text{Pronoun}) \\
\times &P(\text{VP PP} \mid \text{VP}) \\
\times &P(\text{Verb NP} \mid \text{VP}) \\
\times &P(\text{shot} \mid \text{Verb}) \\
\times &P(\text{Det Nominal} \mid \text{NP}) \\
\times &P(\text{an} \mid \text{Det}) \\
\times &P(\text{Noun} \mid \text{Nominal})
\end{aligned}
$$




$$
\begin{aligned}
&P(\text{NP VP} \mid \text{S}) \\
\times &P(\text{Nominal} \mid \text{NP}) \\
\times &P(\text{Pronoun} \mid \text{Nominal}) \\
\times &P(\text{I} \mid \text{Pronoun}) \\
\times &P(\text{VP PP} \mid \text{VP}) \\
\times &P(\text{Verb NP} \mid \text{VP}) \\
\times &P(\text{shot} \mid \text{Verb}) \\
\times &P(\text{Det Nominal} \mid \text{NP}) \\
\times &P(\text{an} \mid \text{Det}) \\
\times &P(\text{Noun} \mid \text{Nominal}) \\
\times &P(\text{elephant} \mid \text{Noun})
\end{aligned}
$$




$$P(T,S) = \prod_{i}^{n} P(\beta \mid A)$$

# CKY for PCFGs

- A PCFG gives us a mechanism for assigning scores (here, probabilities) to different parses for the same sentence.

- But we often care about is finding the single best parse with the highest probability.

- In general, when people talk about CKY, they mean the PCFG version, since the CFG version isn't very useful.

# CKY for PCFGs

- We calculate the max probability parse using CKY by storing the probability of each phrase within each cell as we build it up.
- In particular, we fill the cell for span i-j and label A with the maximum over splits and rules of

$$table(i, j, A) = P(A \rightarrow BC) \times table(i, k, B) \times table(k, j, C)$$

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| PRP:0.04 [0,1] | | | | | | |
| | VBD:0.04 [1,2] | | | | | |
| | | DT:0.05 [2,3] | | | | |
| | | | NN:0.03 [3,4] | | | |
| | | | | IN:0.10 [4,5] | | |
| | | | | | PRP$:0.12 [5,6] | |
| | | | | | | NNS:0.01 [6,7] |

Probaiblity of a terminal (word) given its tag

$$P(A \rightarrow \beta)$$

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| PRP:0.04 [0,1] | ∅ | ∅ | | | | |
| | VBD:0.04 [1,2] | ∅ | | | | |
| | | DT:0.05 [2,3] | NP:0.00015 [2,4] | | | |
| | | | NN:0.03 [3,4] | | | |
| | | | | IN:0.10 [4,5] | | |
| | | | | | PRP$:0.12 [5,6] | |
| | | | | | | NNS:0.01 [6,7] |

$$table(2, 4, NP) = P(\text{NP} \rightarrow \text{DT NN}) \times table(2, 3, DT) \times table(3, 4, NN)$$

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| PRP:0.04 [0,1] | ø | ø | | | | |
| | VBD:0.04 [1,2] | ø | VP:0.0000006 [1,4] | | | |
| | | DT:0.05 [2,3] | NP:0.00015 [2,4] | | | |
| | | | NN:0.03 [3,4] | | | |
| | | | | IN:0.10 [4,5] | | |
| | | | | | PRP$:0.12 [5,6] | |
| | | | | | | NNS:0.01 [6,7] |

We just calculated this value and can use it now

$$table(1,4,VP) = P(\text{VP} \rightarrow \text{VBD NP}) \times table(1,2,VBD) \times table(2,4,NP)$$

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| PRP:0.04 [0,1] | ø | ø | S: 0.0000000048 [0,4] | | | |
| | VBD:0.04 [1,2] | ø | VP:0.0000006 [1,4] | | | |
| | | DT:0.05 [2,3] | NP:0.00015 [2,4] | | | |
| | | | NN:0.03 [3,4] | | | |
| | | | | IN:0.10 [4,5] | | |
| | | | | | PRP$:0.12 [5,6] | |
| | | | | | | NNS:0.01 [6,7] |

We just calculated this value and can use it now

$$table(0,4,S) = P(\text{S} \rightarrow \text{NP VP}) \times table(0,1,NP) \times table(1,4,VP)$$

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| PRP:0.04 [0,1] | ø | ø | S: 0.0000000048 [0,4] | | | |
| | VBD:0.04 [1,2] | ø | VP:0.0000006 [1,4] | | | |
| | | DT:0.05 [2,3] | NP:0.00015 [2,4] | | | |
| | | | NN:0.03 [3,4] | | | |
| | | | | IN:0.10 [4,5] | | |
| | | | | | PRP$:0.12 [5,6] | |
| | | | | | | NNS:0.01 [6,7] |

Note these values are getting very small! Better to add in log space

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| PRP: -3.21 [0,1] | ø | ø | S: -19.2 [0,4] | | | |
| | VBD: -3.21 [1,2] | ø | VP: -14.3 [1,4] | | | |
| | | DT: -3.0 [2,3] | NP: -8.8 [2,4] | | | |
| | | | NN: -3.5 [3,4] | | | |
| | | | | IN: -2.3 [4,5] | | |
| | | | | | PRP$: -2.12 [5,6] | |
| | | | | | | NNS: -4.6 [6,7] |

Note these values are getting very small! Better to add in log space

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| PRP: -3.21 [0,1] | ø | ø | S: -19.2 [0,4] | ø | ø | |
| | VBD: -3.21 [1,2] | ø | VP: -14.3 [1,4] | ø | ø | $VP_1$, $VP_2$ [1,7] |
| | | DT: -3.0 [2,3] | NP: -8.8 [2,4] | ø | ø | NP: -24.7 [2,7] |
| | | | NN: -3.5 [3,4] | ø | ø | NP: -19.4 [3,7] |
| | | | | IN: -2.3 [4,5] | ø | PP: -13.6 [4,7] |
| | | | | | PRP$: -2.12 [5,6] | NP: -9.0 [5,7] |
| | | | | | | NNS: -4.6 [6,7] |

For any phrase type spanning [i,j], we only need to keep the max probability given the assumptions of a PCFG

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| PRP: -3.21 [0,1] | ø | ø | S: -19.2 [0,4] | ø | ø | |
| | VBD: -3.21 [1,2] | ø | VP: -14.3 [1,4] | ø | ø | VP: -30.2 [1,7] |
| | | DT: -3.0 [2,3] | NP: -8.8 [2,4] | ø | ø | NP: -24.7 [2,7] |
| | | | NN: -3.5 [3,4] | ø | ø | NP: -19.4 [3,7] |
| | | | | IN: -2.3 [4,5] | ø | PP: -13.6 [4,7] |
| | | | | | PRP$: -2.12 [5,6] | NP: -9.0 [5,7] |
| | | | | | | NNS: -4.6 [6,7] |

For any phrase type spanning [i,j], we only need to keep the max probability given the assumptions of a PCFG

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| PRP: -3.21 [0,1] | ø | ø | S: -19.2 [0,4] | ø | ø | S: -35.7 [0,7] |
| | VBD: -3.21 [1,2] | ø | VP: -14.3 [1,4] | ø | ø | VP: -30.2 [1,7] |
| | | DT: -3.0 [2,3] | NP: -8.8 [2,4] | ø | ø | NP: -24.7 [2,7] |
| | | | NN: -3.5 [3,4] | ø | ø | NP: -19.4 [3,7] |
| | | | | IN: -2.3 [4,5] | ø | PP: -13.6 [4,7] |
| | | | | | PRP$: -2.12 [5,6] | NP: -9.0 [5,7] |
| | | | | | | NNS: -4.6 [6,7] |

For any phrase type spanning [i,j], we only need to keep the max probability given the assumptions of a PCFG

| I | shot | an | elephant | in | my | pajamas |
|---|---|---|---|---|---|---|
| PRP: -3.21 [0,1] | ∅ | ∅ | S: -19.2 [0,4] | ∅ | ∅ | S: -35.7 [0,7] |
| | VBD: -3.21 [1,2] | ∅ | VP: -14.3 [1,4] | ∅ | ∅ | VP: -30.2 [1,7] |
| | | DT: -3.0 [2,3] | NP: -8.8 [2,4] | ∅ | ∅ | NP: -24.7 [2,7] |
| | | | NN: -3.5 [3,4] | ∅ | ∅ | NP: -19.4 [3,7] |
| | | | | IN: -2.3 [4,5] | ∅ | PP: -13.6 [4,7] |
| | | | | | PRP$: -2.12 [5,6] | NP: -9.0 [5,7] |
| | | | | | | NNS: -4.6 [6,7] |

As in Viterbi, backpointers let us keep track on the path through the chart that leads to the best derivation

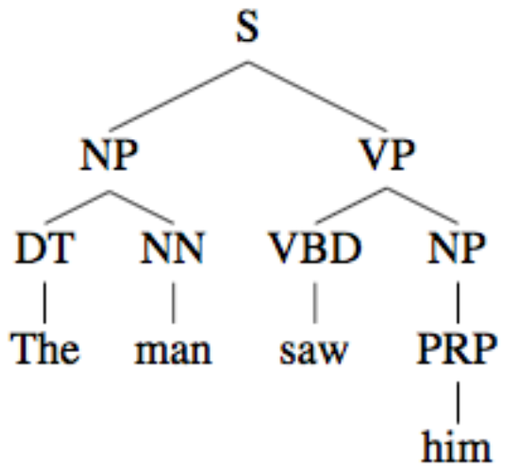
# PCFG

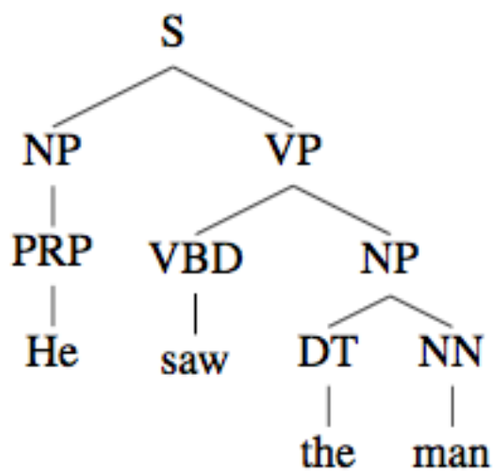
- Probabilistic context-free grammar: each production is also associated with a probability.
- This lets us calculate the probability of a parse for a given sentence; for a given parse tree T for sentence S comprised of n rules from R (each A → β):

$$P(T, S) = \prod_{i}^{n} P(\beta \mid A)$$

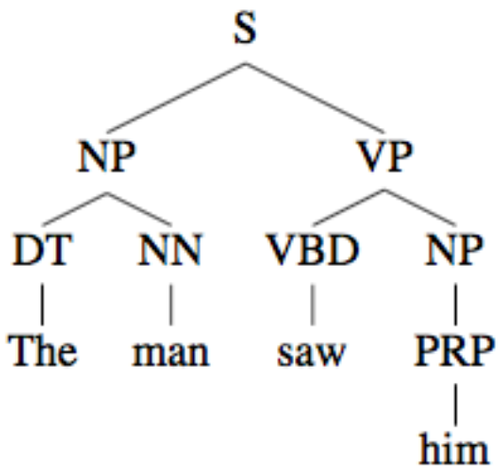
# Problems with PCFGs

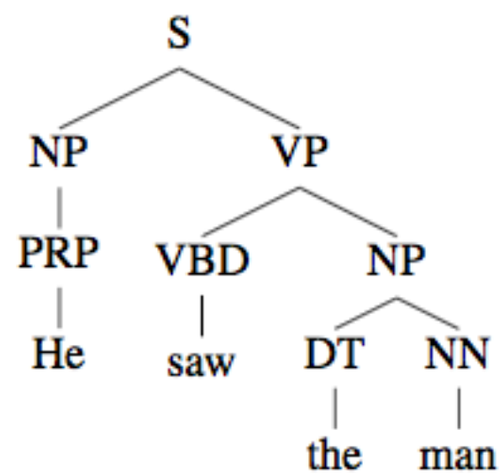
- Strong independence assumptions
  - Each production (e.g., NP → DT NN) is independent of the rest of tree.
  - In real use, productions are strongly dependent on their place in the tree.



# Problems with PCFGs




| | Pronoun | Non-Pronoun |
|---|---|---|
| Subject | 91% | 9% |
| Object | 34% | 66% |

SLP3; Francis et al. 1999

# Problems with PCFGs

| | Pronoun | Non-Pronoun |
|---|---|---|
| Subject | 91% | 9% |
| Object | 34% | 66% |

Maximum likelihood estimates from Switchboard:

- P(NP → DT NN) = 0.28
- P(NP → PRP) = 0.25

SLP3; Francis et al. 1999

# Splitting non-terminals

Rather than having a single rule for each non-terminal $P(NP \rightarrow DT\ NN)$, we can condition on some context (Johnson 1998)

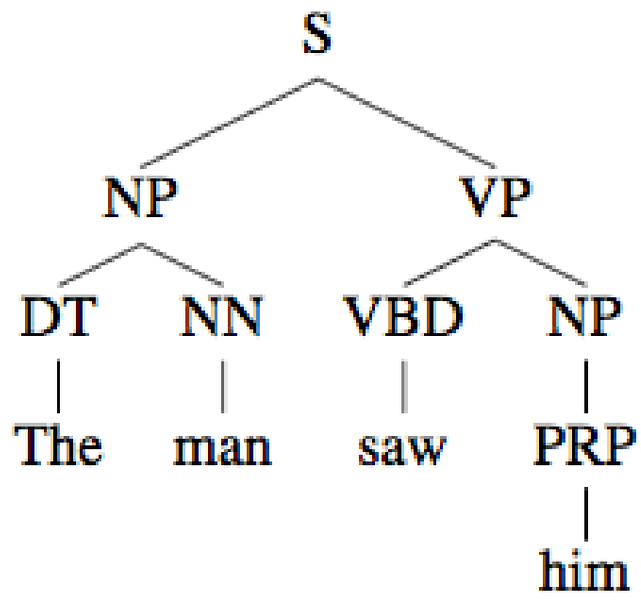
- $P_{subject}(NP \rightarrow DT\ NN)$
- $P_{object}(NP \rightarrow DT\ NN)$

# Splitting non-terminals

- Subjects/objects are structural relations in phrase structure trees

- Subject = NP child of S
- Object = NP child of VP child of S

(For some subjects/objects; other rules for embedded structures)

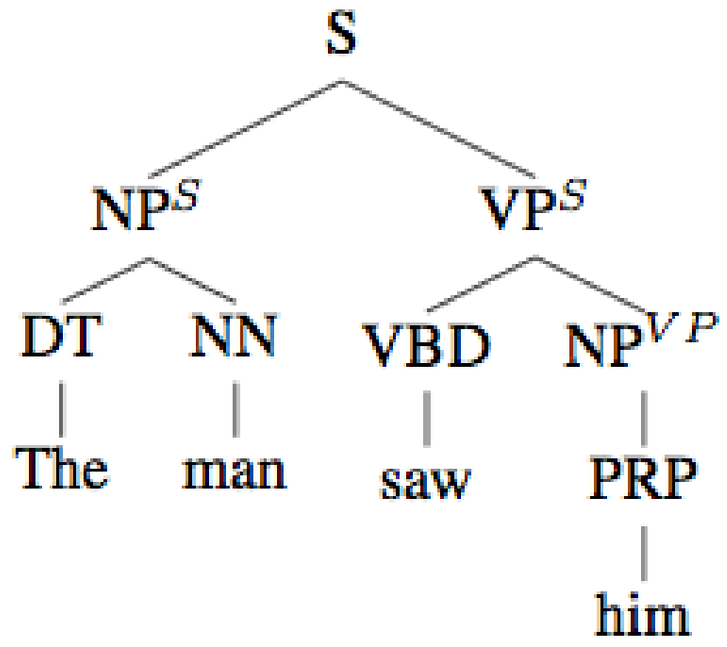


# Parent annotation

We can encode context in a general way by annotating each node in a tree with its parent

This lets us learn different probabilities for:

- $NP^S$ (subjects)
- $NP^{VP}$ (objects)



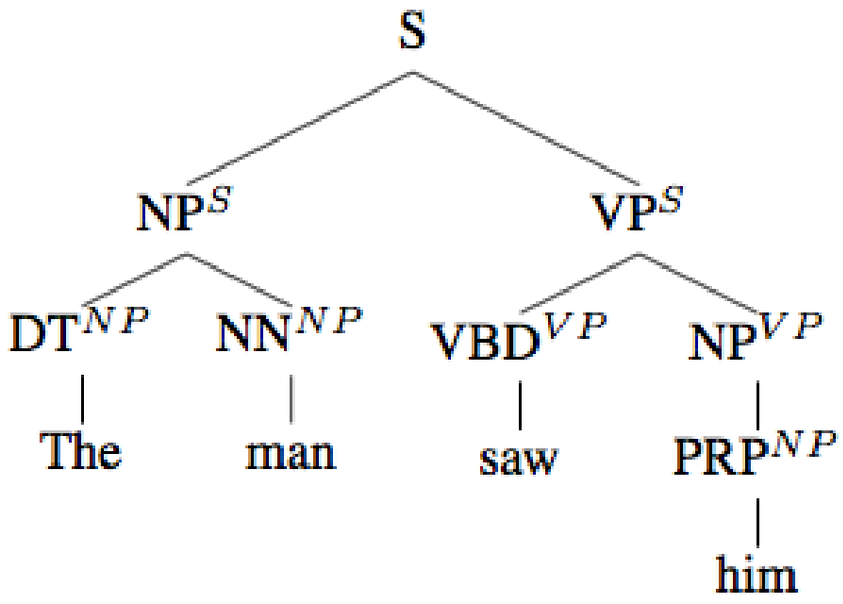
# Parent annotation

We can also split the pre-terminal POS tags too

(Klein and Manning 2003)

This lets us learn different probabilities

- $P(RB^{VP} \rightarrow not)$
- $P(RB^{NP} \rightarrow not)$



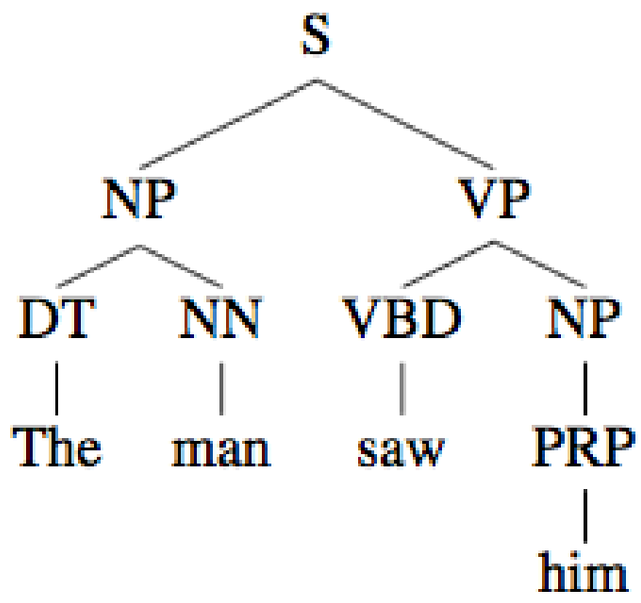
# Splitting non-terminals

- Dramatically increases the size of the grammar → less training data for each production

- Other approaches search for best splits that maximize the training data likelihood (Petrov et al 2006)

# Problems with PCFGs

- P(VBD → saw) — "saw" itself doesn't influence the structure above it except through that pre-terminal.

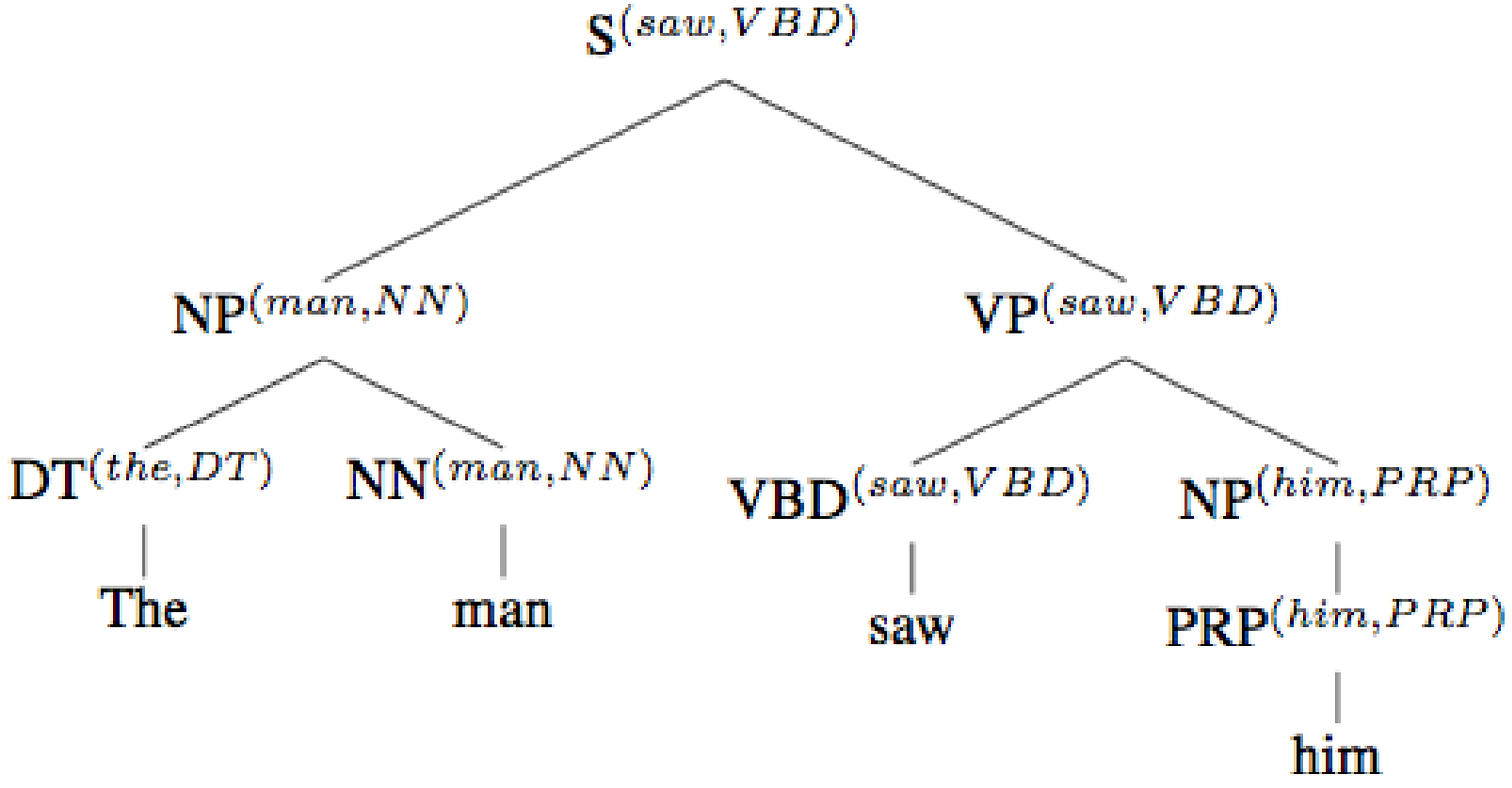



- What if the verb was "sneezed"?

- Lexicon information in a PCFG has little influence on the overall parse structure

# Lexicalized PCFGs

- Annotate each node with its head + POS tag of head



# Lexicalized PCFGs

- Annotate each node with its head + POS tag of head
- We can't have a rule for each fine-grained production — e.g. $P(S^{(saw, VBD)} \rightarrow NP^{(man, NN)}\ VP^{(saw, VBD)})$
- Different models make different independent assumptions to make this quantity tractable (Collins 1999, Charniak 1997)

# Performance Comparison

| | F1 |
|---|---|
| Vanilla PCFG | 72.6 |
| Manually splitting (Klein 2003) | 86.4 |
| Automatic splitting (Petrov 2006) | 89.6 |
| Lexicalized PCFG (Collins 2003) | 88.2 |

# Neural Span Parser

- Lots of different neural parsing methods, today going to talk about Stern et al. *A Minimal Span-Based Neural Constituency Parser*

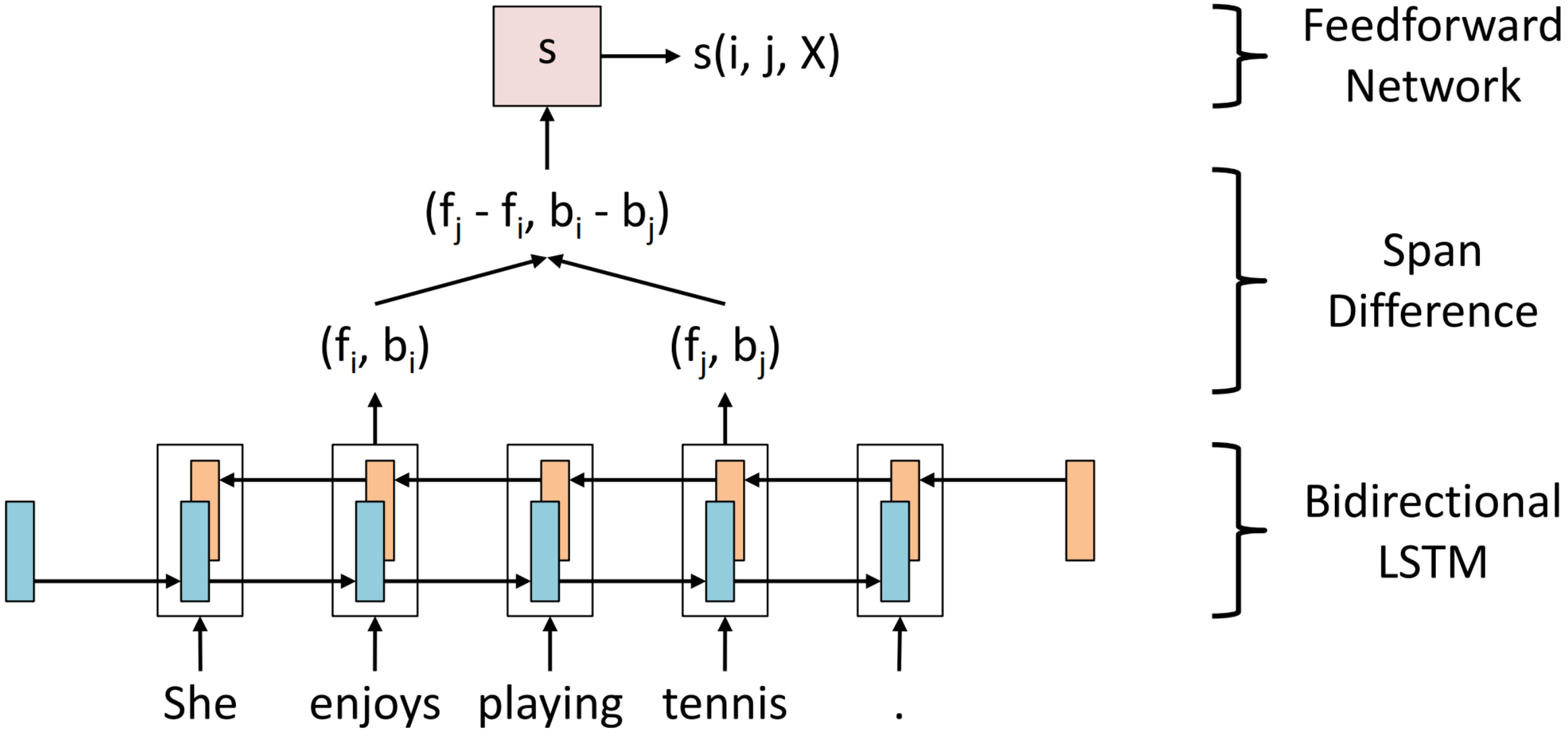
- This model uses a chart-based decoding scheme similar to CKY, but gets scores from the output of a neural network

# Neural Span Parser

- Instead of giving a score for each production rule, the neural span parser directly computes a score for each labeled span

- The dynamic program is only needed to ensure the output is a tree

# Neural Span Parser



# Performance Comparison

| | F1 |
|---|---|
| Automatic splitting (Petrov 2006) | 89.6 |
| Non-neural re-ranking parser (Charniak 2005) | 91.0 |
| Neural span parser (Stern 2017) | 91.7 |
| State-of-the-art neural (Kitaev 2018) | 93.6 |
| + extra unlabeled data (Kitaev 2018) | 95.1 |