# Doubling DOP*

## A comparison of Double-DOP and DOP*

Benno Kruit  Sara Veldhoen

Supervised by:
Andras van Cranenburg  Khalil Sima'an

University of Amsterdam (UvA)

Project AI, January 2014

# Outline

# Outline

# Parsing

- input: sentence

  *John Loves Mary*

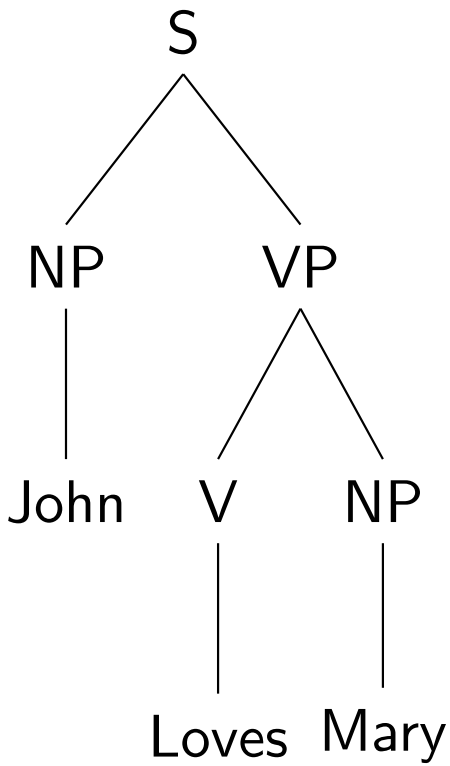
- output: constituent tree

# Parsing

- input: sentence

  *John Loves Mary*

- output: constituent tree



# Grammar

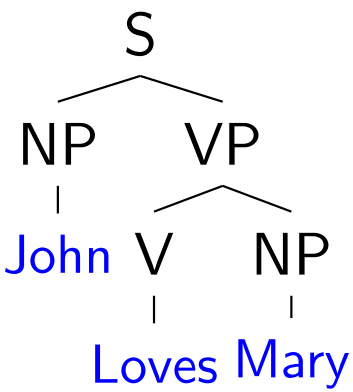
A grammar describes:

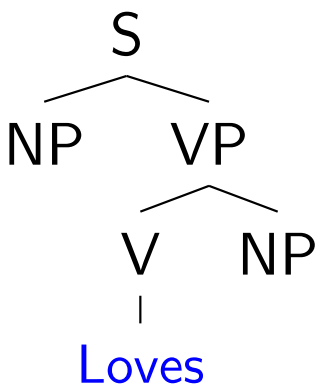
- how trees can be built
  - CFG's - elementary rules
  - TSG's - larger units: *fragments*
- how likely constructions are: *probabilistic* grammars
  - PCFG's - independence
  - PTSG's - derivations

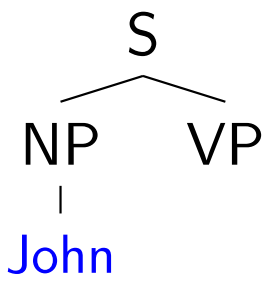
# Grammar: CFG rules

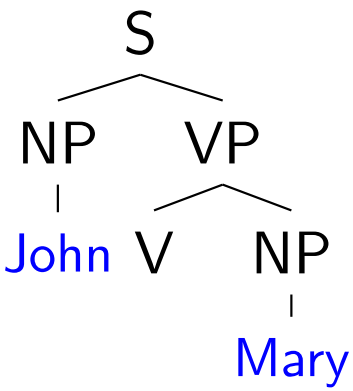
$S \rightarrow NP\ VP$
$VP \rightarrow V\ NP$
$NP \rightarrow John$
$NP \rightarrow Mary$
$V \rightarrow loves$

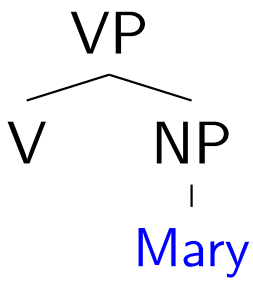
# Grammar: Tree fragments




Etc...

Exponentially many

# Outline

# Title

# Outline

# Double-DOP

- Extraction: Maximal Overlap
- Estimation: relative frequency
- Coverage: PCFG rules

# DOP*

- Held-out estimation - *HC* and *EC*
- Extraction: Shortest derivations
- Estimation: relative frequency *in shortest derivations*
- Coverage: smoothing PCFG rules with probability $p_{unkn}$

# Outline

# Comparison

- Shortest derivations or Maximal overlap
- Held-out estimation or one vs. the rest

# Outline

# Outline

# Maximal overlap $\leftrightarrow$ shortest derivation

# Split $\leftrightarrow$ one vs. the rest

# Outline

# F1 scores

# Summary

- The first main message of your talk in one or two lines.
- The second main message of your talk in one or two lines.
- Perhaps a third message, but not more than that.

- Outlook
  - Something you haven't solved.
  - Something else you haven't solved.