Doubling DOP*

Kruit, Veldhoen

Data Oriented
Parsing
Introduction to DOP
Bias and Consistency

Double-DOP and
DOP*: a
comparison
Introduction to
Double-DOP and
DOP*
Comparison
Experiments

Results
Analyzing grammars
Parsing Performance

Summary

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

Doubling DOP*

Kruit, Veldhoen

Data Oriented
Parsing
Introduction to DOP
Bias and Consistency

Double-DOP and
DOP*: a
comparison
Introduction to
Double-DOP and
DOP*
Comparison
Experiments

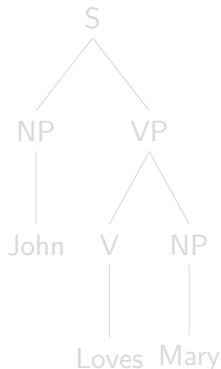Results
Analyzing grammars
Parsing Performance

Summary

Outline

Doubling DOP*

Kruit, Veldhoen

Data Oriented
Parsing
Introduction to DOP
Bias and Consistency

Double-DOP and
DOP*: a
comparison
Introduction to
Double-DOP and
DOP*
Comparison
Experiments

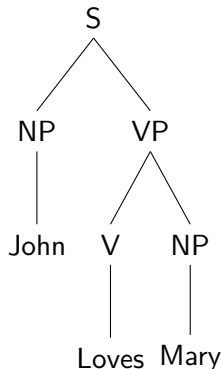Results
Analyzing grammars
Parsing Performance

Summary

Parsing

- input: sentence

    John Loves Mary

- output: constituent tree

Kruit, Veldhoen

Parsing

Doubling DOP*

Kruit, Veldhoen

Data Oriented
Parsing
Introduction to DOP
Bias and Consistency

Double-DOP and
DOP*: a
comparison
Introduction to
Double-DOP and
DOP*
Comparison
Experiments

Results
Analyzing grammars
Parsing Performance

Summary

- ▶ 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
  - ▶ TSG's - derivations

Outline

Doubling DOP*

Kruit, Veldhoen

Data Oriented
Parsing
Introduction to DOP
Bias and Consistency

Double-DOP and
DOP*: a
comparison
Introduction to
Double-DOP and
DOP*
Comparison
Experiments

Results
Analyzing grammars
Parsing Performance

Summary

Title

Outline

Doubling DOP*

Kruit, Veldhoen

Data Oriented
Parsing
Introduction to DOP
Bias and Consistency

Double-DOP and
DOP*: a
comparison
Introduction to
Double-DOP and
DOP*
Comparison
Experiments

Results
Analyzing grammars
Parsing Performance

Summary

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

Doubling DOP*

Kruit, Veldhoen

Data Oriented
Parsing
Introduction to DOP
Bias and Consistency

Double-DOP and
DOP*: a
comparison
Introduction to
Double-DOP and
DOP*
Comparison
Experiments

Results
Analyzing grammars
Parsing Performance

Summary

Comparison

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

Outline

Doubling DOP*

Kruit, Veldhoen

Data Oriented
Parsing
Introduction to DOP
Bias and Consistency

Double-DOP and
DOP*: a
comparison
Introduction to
Double-DOP and
DOP*
Comparison
Experiments

Results
Analyzing grammars
Parsing Performance

Summary

Outline

Doubling DOP*

Kruit, Veldhoen

Data Oriented
Parsing
Introduction to DOP
Bias and Consistency

Double-DOP and
DOP*: a
comparison
Introduction to
Double-DOP and
DOP*
Comparison
Experiments

Results
Analyzing grammars
Parsing Performance

Summary

Maximal overlap $\leftrightarrow$ shortest derivation

Split ↔ one vs. the rest

Outline

F1 scores

Kruit, Veldhoen

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.