Kruit, Veldhoen

Data Oriented
Parsing

Introduction to DOP Bias and Consistency

Double-DOP and DOP\*: a

Introduction to Double-DOP ar DOP\*

Comparison Experiments

Result

Analyzing grammars Parsing Performance

Summary

Doubling DOP\*

A comparison of Double-DOP and DOP\*

Benno Kruit Sara Veldhoen

Supervised by:
Andreas van Cranenburg Khalil Sima'an

University of Amsterdam (UvA)

Project AI, January 2014

## Outline

# 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

### Doubling DOP\*

Kruit, Veldhoen

#### Data Oriented Parsing

Introduction to DOP Bias and Consistency

# Double-DOP and DOP\*: a

Introduction

Double-DOP and DOP\*
Comparison

Experiment

#### Result

Analyzing grammars
Parsing Performance

ummarv

Introduction to DOP Bias and Consistency

Double-DOP and DOP\*: a

Introduction Double-DOP

DOP\*
Comparison

Experimen

Result

Analyzing grammars

Summarv

# Data Oriented Parsing Introduction to DOP

Bias and Consistency

Double-DOP and DOP\*: a comparison

Introduction to Double-DOP and DOP

Comparisor

Experiments

## Results

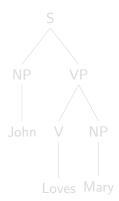
Analyzing grammars
Parsing Performance

# **Parsing**

► input: sentence

John Loves Mary

output: constituent tree



Doubling DOP\*

Kruit, Veldhoen

Data Oriented Parsing

Introduction to DOP Bias and Consistency

Double-DOP and DOP\*: a

comparison

Double-DOP DOP\*

Comparison Experiments

Resul

Analyzing grammars
Parsing Performance

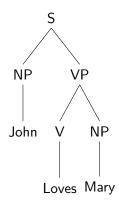
Summary

# **Parsing**

input: sentence

John Loves Mary

output: constituent tree



Doubling DOP\*

Kruit, Veldhoen

Data Oriented Parsing

Introduction to DOP Bias and Consistency

Double-DOP and DOP\*: a

comparison

Double-DOP DOP\*

Comparison Experiments

Resul

Analyzing grammars
Parsing Performance

Summar

#### Data Oriented Parsing

Introduction to DOP Bias and Consistency

# Double-DOP and DOP\*: a

Introduction

Double-DOP and DOP\*
Comparison

Experiment:

#### Result

Analyzing grammars

ummarv

## 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

### Data Oriented Parsing

Introduction to DOP Bias and Consistency

# Double-DOP and DOP\*: a

comparison

Double-DOP DOP\* Comparison

Comparison Experiments

#### Results

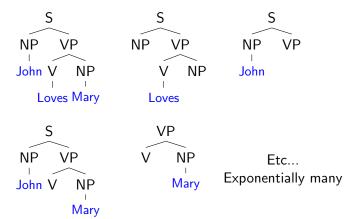
Analyzing grammars

Summary

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

V→ loves

# Grammar: Tree fragments



Doubling DOP\*

Kruit, Veldhoen

Data Oriented Parsing

Introduction to DOP Bias and Consistency

Double-DOP and

comparison

Double-DOP as DOP\* Comparison

omparison xperiments

Result

Analyzing grammars Parsing Performance

ummary

Double-DOP and

Introduction Double-DOP DOP\*

Comparison

Experimer

Resul

Analyzing grammars

Summary

## Data Oriented Parsing

Introduction to DOP

Bias and Consistency

Double-DOP and DOP\*: a comparison

Introduction to Double-DOP and DOP

Comparisor

Experiments

## Results

Analyzing grammars
Parsing Performance

Double-DOP and DOP\*: a

Introduction to
Double-DOP and

Comparison

. .

Result

Analyzing grammars

Summary

- Assumption
  - Language is an infinite parse tree distribution
  - ► Treebank is a finite sample
- Estimate the true distribution
- Expected estimation should improve when the treebank grows → expected loss should decline
- ▶ Consistency: Expected loss becomes 0 when the sample size approaches  $\infty$

Double-DOP and

- Assumption
  - An estimator should approach any distribution
  - Even finite distributions!
- ▶ If there's a distribution that doesn't match its expected estimate, the estimator is biased.
- What about unseen data?
- Bias is good

Introduction to DOP Bias and Consistency

Double-DOP and DOP\*: a

comparison
Introduction to
Double-DOP and

DODP\*
Comparison

Exper.

Resul

Analyzing grammars

Summary

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

## Double-DOP

Extraction: Maximal Overlap

Estimation: relative frequency

► Coverage: PCFG rules

### Doubling DOP\*

Kruit, Veldhoen

#### Data Oriented Parsing

Bias and Consistency

# Double-DOP and DOP\*: a

comparison
Introduction to
Double-DQP and

DOP\*
Comparison

Comparison Experiments

#### Resul

Analyzing grammars

Summary

Double-DOP and DOP\*: a comparison

Introduction to Double-DOP and DOP\*

Comparison Experiments

Results

Analyzing grammars

ummary

► Held-out estimation - HC and EC

Extraction: Shortest derivations

► Estimation: relative frequency in shortest derivations

Coverage: smoothing PCFG rules with probability p<sub>unkn</sub>

Double-DOP and

Double-DOP and

Comparison

Comparison

Double-DOP and DOP\*: a comparison

# Comparison

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

Doubling DOP\*

Kruit, Veldhoen

Data Oriented

Double-DOP and

Comparison

Introduction to DOP Bias and Consistency

Double-DOP and DOP\*: a

Introduction Double-DOP DOP\*

Experiments

Resul

Analyzing grammars

Summary

Double-DOP and DOP\*: a comparison

Introduction to Double-DOP and DOP\*

Experiments

## Results

Analyzing grammars
Parsing Performance

Introduction to DOP Bias and Consistency

Double-DOP and DOP\*: a

Introduction Double-DOP DOP\*

Comparison Experiments

Result

Analyzing grammars Parsing Performance

Summary

Introduction to DOP
Bias and Consistency

Double-DOP and DOP\*: a comparison Introduction to Double-DOP and Double-DOP

Comparison

Experiments

## Results

Analyzing grammars

Parsing Performance

# Maximal overlap ↔ shortest derivation

Doubling DOP\*

Kruit, Veldhoen

# Double-DOP and

Analyzing grammars

# Split $\leftrightarrow$ one vs. the rest

 $\mathsf{Doubling}\;\mathsf{DOP}^*$ 

 $Kruit,\ Veldhoen$ 

Data Oriented Parsing

Bias and Consistency

Double-DOP and

comparison Introduction

DOP\*\*
Comparison

Recult

Analyzing grammars

Summar

Parsing Performance

Results

Parsing Performance

## F1 scores

## $\mathsf{Doubling}\;\mathsf{DOP}^*$

 $Kruit,\ Veldhoen$ 

#### Data Oriented Parsing

Bias and Consistency

# Double-DOP and DOP\*: a

Introduction t

DOP\*
Comparison

Comparison Experiments

#### Result

Analyzing grammars Parsing Performance

Summary

# Double-DOP and

Double-DOP and

Summary

Outlook

- ▶ Why not?
- Is it just grammar size?

So shortest derivations are not as useful as they seem

And a split moves weight to large fragments

Performance is not related to consistency