A comparison of Double-DOP and DOP*

Reconsidering non-trivial DOP estimators

Benno Kruit Sara Veldhoen

Supervised by:
Andreas van Cranenburg Khalil Sima'an

University of Amsterdam (UvA)

Project AI, January 2014

Data Oriented Parsing
Introduction to DOP
Bias and Consistency
Double-DOP and DOP*: a comparison
Introduction to Double-DOP and DOP*
Comparison
Experiments
Results
Parsing Performance
Analyzing grammars
Summary

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
- Parsing Performance
- Analyzing grammars

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
- Parsing Performance
- Analyzing grammars

Parsing

- input: sentence

  John Loves Mary

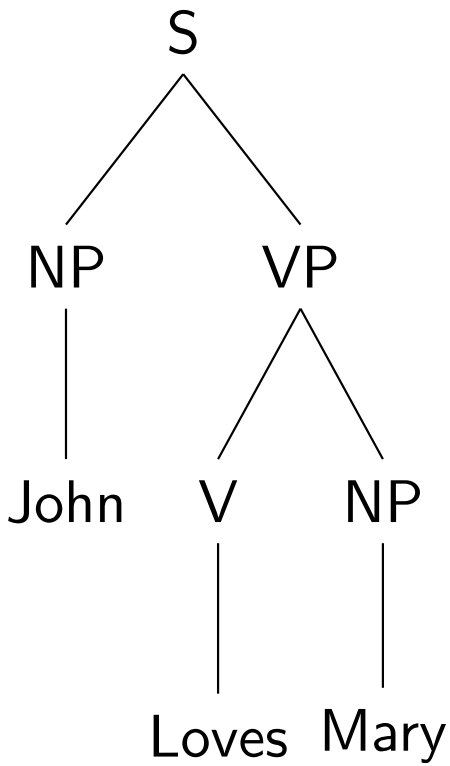
- output: constituent tree

Parsing

- input: sentence

  John Loves Mary

- output: constituent tree



Grammar

A grammar describes:

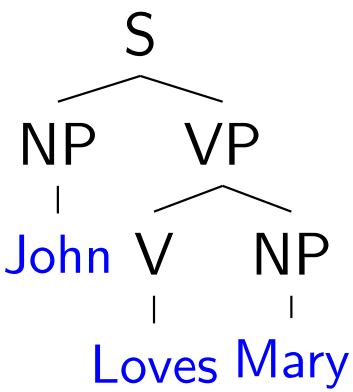
- how trees can be built
  - CFG's - elementary rules
  - TSG's - larger units: *fragments*
- how likely constructions are: *probabilistic* grammars
  - PCFG's - independence
  - PTSG's - derivations

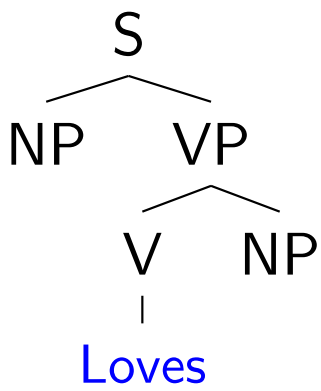
Grammar: CFG rules

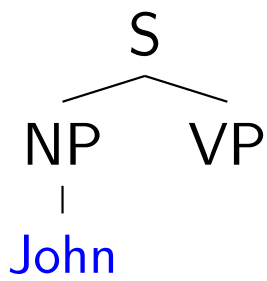
$S \rightarrow NP\ VP$
$VP \rightarrow V\ NP$
$NP \rightarrow John$
$NP \rightarrow Mary$
$V \rightarrow loves$

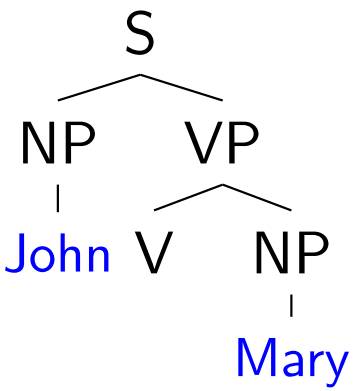
Data Oriented Parsing
Introduction to DOP
Bias and Consistency
Double-DOP and DOP*: a comparison
Introduction to Double-DOP and DOP*
Comparison
Experiments
Results
Parsing Performance
Analyzing grammars
Summary

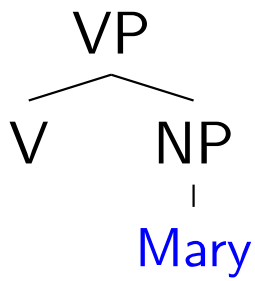
Grammar: Tree fragments







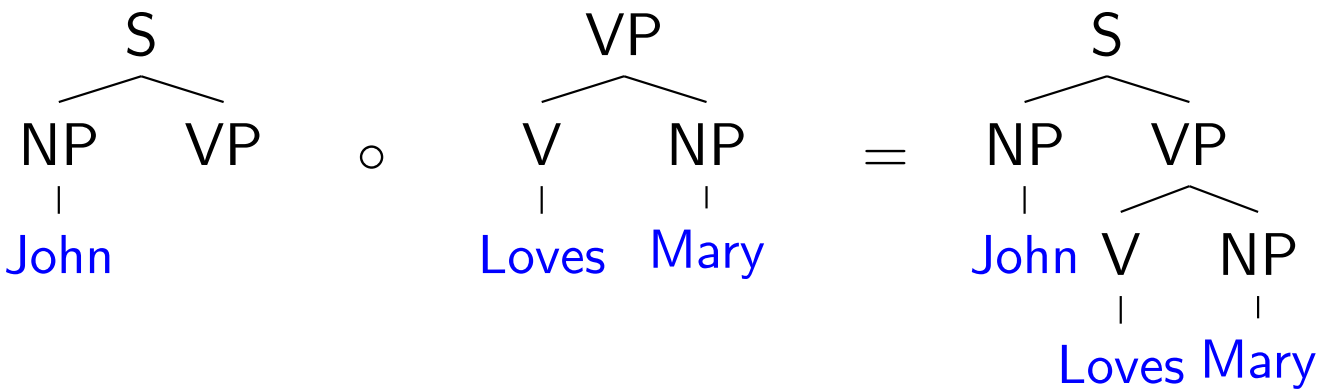




Etc...

Exponentially many

Grammar: Tree fragments



$$0.1 \times 0.3 = 0.03$$

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
- Parsing Performance
- Analyzing grammars

Consistency

- 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 $\rightarrow$ expected *loss* should decline
- **Consistency**: Expected loss becomes 0 when the sample size approaches ∞

Data Oriented Parsing
Introduction to DOP
Bias and Consistency
Double-DOP and DOP*: a comparison
Introduction to Double-DOP and DOP*
Comparison
Experiments
Results
Parsing Performance
Analyzing grammars
Summary

Bias

- 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**

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
- Parsing Performance
- Analyzing grammars

Data Oriented Parsing
Introduction to DOP
Bias and Consistency

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

Results
Parsing Performance
Analyzing grammars

Summary

Double-DOP

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

Data Oriented Parsing
Introduction to DOP
Bias and Consistency
Double-DOP and DOP*: a comparison
Introduction to Double-DOP and DOP*
Comparison
Experiments
Results
Parsing Performance
Analyzing grammars
Summary

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

Data Oriented Parsing
- Introduction to DOP
- Bias and Consistency

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

Results
- Parsing Performance
- Analyzing grammars

Comparison

- Shortest derivations or Maximal overlap
- Split or full estimation
- Consistency

Data Oriented Parsing
Introduction to DOP
Bias and Consistency
Double-DOP and DOP*: a comparison
Introduction to Double-DOP and DOP*
Comparison
Experiments
Results
Parsing Performance
Analyzing grammars
Summary

Example

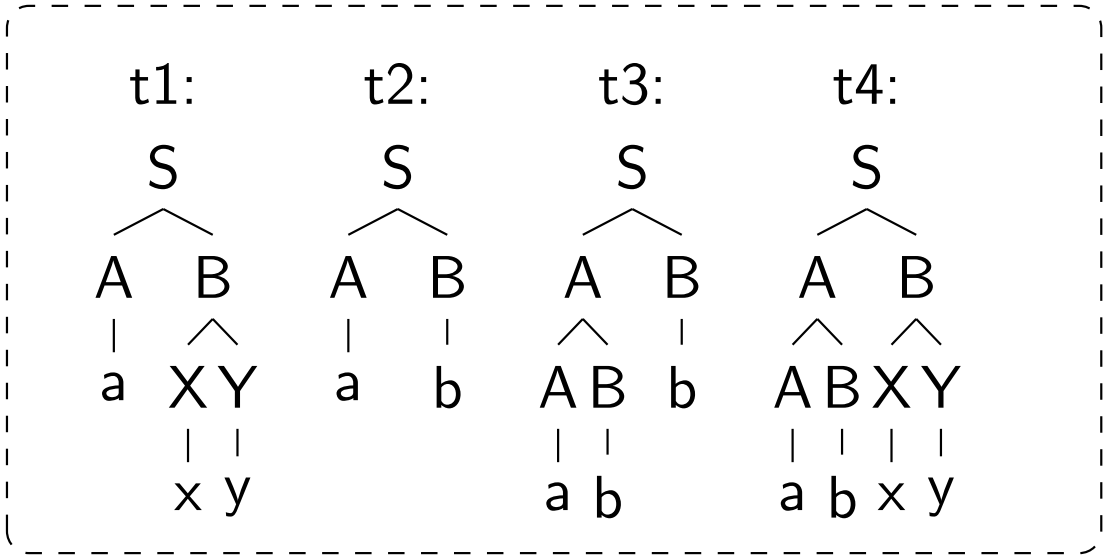


Figure : A toy treebank

Example

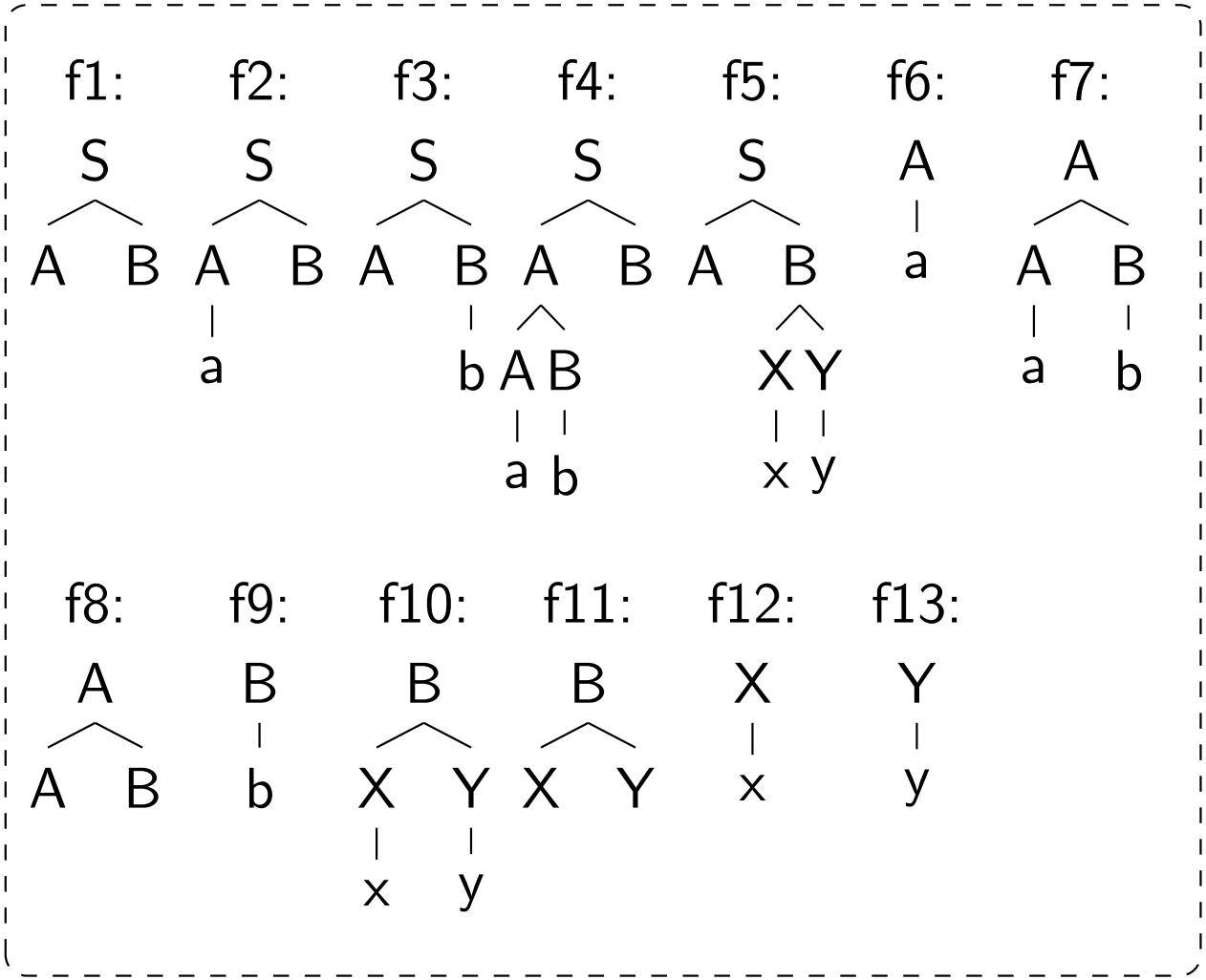


Figure : Some extracted fragments

Example

	Maximal overlap	weight	Shortest deriv.[1]	weight
f1	(t1,t3),(t2,t4)	4/12	-	0
f2	(t1,t2)	2/12	1b, 2a	1/4
f3	(t2,t3)	2/12	2b, 3b	1/4
f4	(t3,t4)	2/12	3a, 4b	1/4
f5	(t1,t4)	2/12	1a, 4a	1/4
f6	(t1,t3),(t1,t4), (t2,t3),(t2,t4)	4/6	1a, 2b	1/2
f7	-	0	3b, 4a	1/2
f8	CFG rule	2/6	-	0
f9	(t2,t3),(t2,t4), (t3,t4)	4/6	2a, 3a	1/2
f10	-	0	1b, 4b	1/2
f11	CFG rule	2/6	-	0
f12	CFG rule	2/2	-	0
f13	CFG rule	2/2	-	0

Table : Weight assignment of MO and SD, full estimation

[1] $t_1 = f_5 \circ f_6(\text{a})$ or $f_2 \circ f_{10}(\text{b})$, $t_2 = f_2 \circ f_8(\text{a})$ or $f_3 \circ f_6(\text{b})$, $t_3 = f_4 \circ f_8(\text{a})$ or $f_3 \circ f_7(\text{b})$, $t_4 = f_5 \circ f_7(\text{a})$ or $f_4 \circ f_9(\text{b})$

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
- Parsing Performance
- Analyzing grammars

Experiments

Estimation and Parsing with the `Disco-Dop` framework.
Three grammars:

- Maximal Overlap Full (Double-DOP)
- Maximal Overlap Split
- Shortest Derivation Split (DOP*)

Data Oriented Parsing
Introduction to DOP
Bias and Consistency
Double-DOP and DOP*: a comparison
Introduction to Double-DOP and DOP*
Comparison
Experiments
Results
Parsing Performance
Analyzing grammars
Summary

Data

Wall Street Journal (WSJ) section of the Penn Treebank
Preprocessing:

- Removing functions
- Binarizing by Markovization (h=1 v=1)

Data Oriented Parsing
Introduction to DOP
Bias and Consistency
Double-DOP and DOP*: a comparison
Introduction to Double-DOP and DOP*
Comparison
Experiments
Results
Parsing Performance
Analyzing grammars
Summary

Algorithm

- Estimation
  - Full: Maximal Overlap
  - Split: 10 random folds, interpolating results
    - Maximal Overlap
    - Shortest Derivation
  - Smoothing
- Parsing
  - Input: sentences with sentences with a POS-tag attached to each word
  - Output: Parsing accuracy scores

Data Oriented Parsing
Introduction to DOP
Bias and Consistency
Double-DOP and DOP*: a comparison
Introduction to Double-DOP and DOP*
Comparison
Experiments
Results
Parsing Performance
Analyzing grammars
Summary

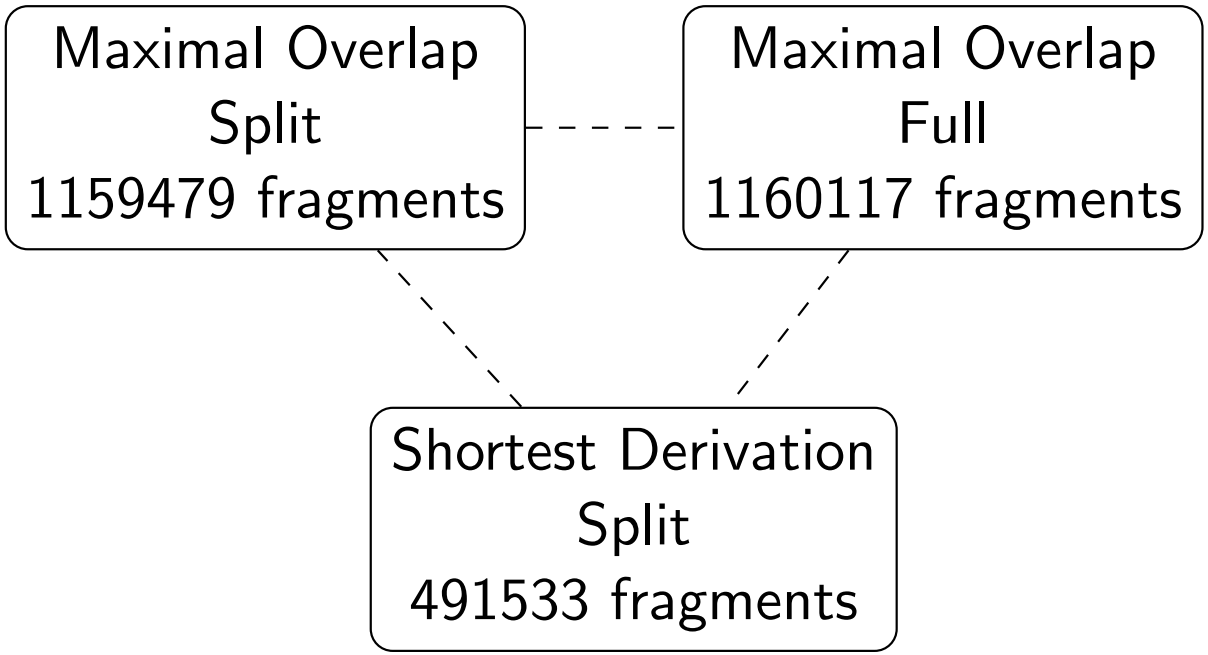


Figure : The grammars and their size

$p_{unkn} = 1.41 \times 10^{-3}$

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
- Parsing Performance
- Analyzing grammars

Scores

	Maximal Overlap Full	Maximal Overlap Split	Shortest Derivation Split
labeled recall	86.17	85.11	79.20
labeled precision	86.05	85.50	79.32
labeled f-measure	86.11	85.31	79.26
exact match	28.32	25.87	16.52

Table : Results for 1229 sentences of length$\leq$ 40

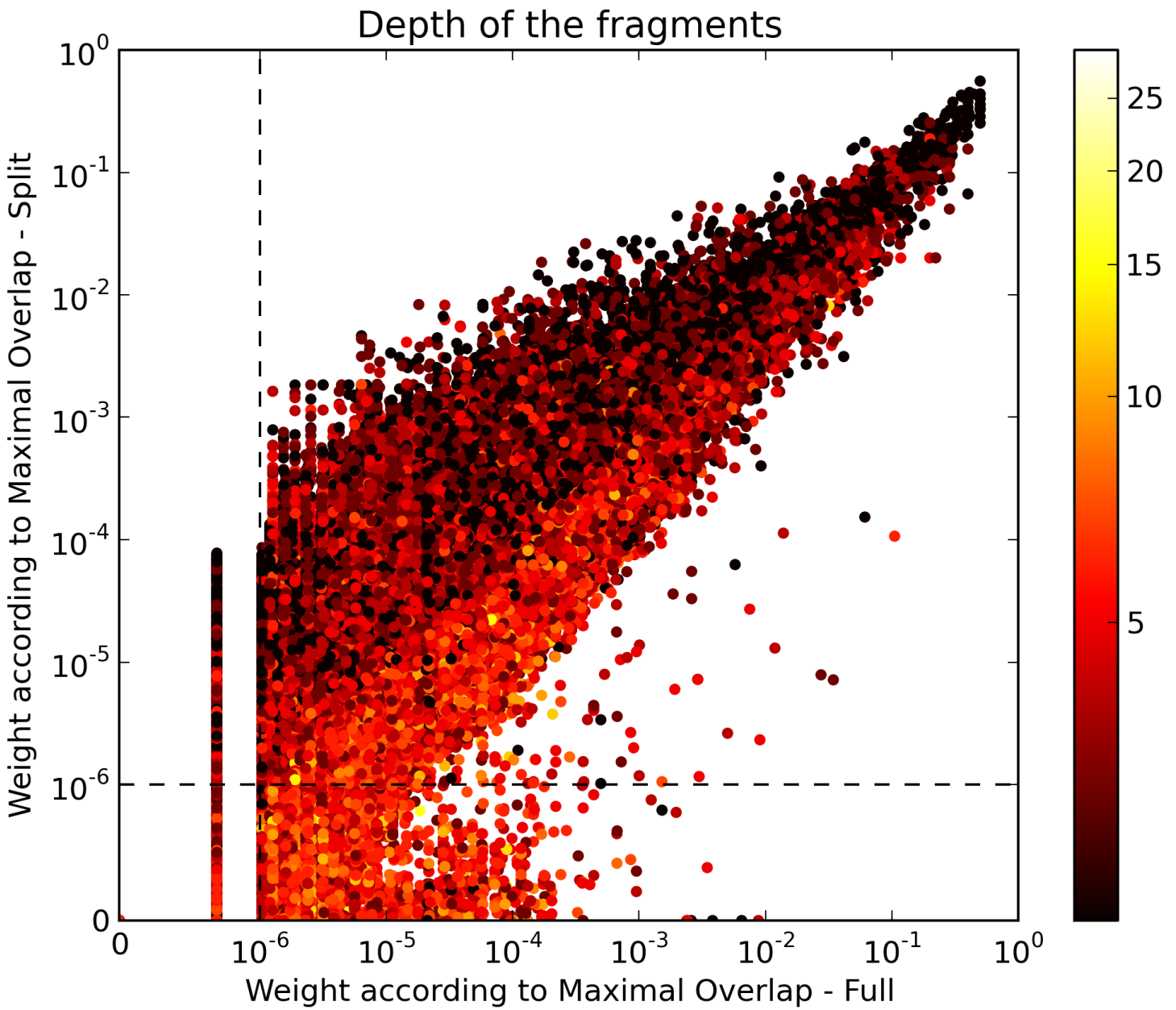
Outline

Data Oriented Parsing
- Introduction to DOP
- Bias and Consistency

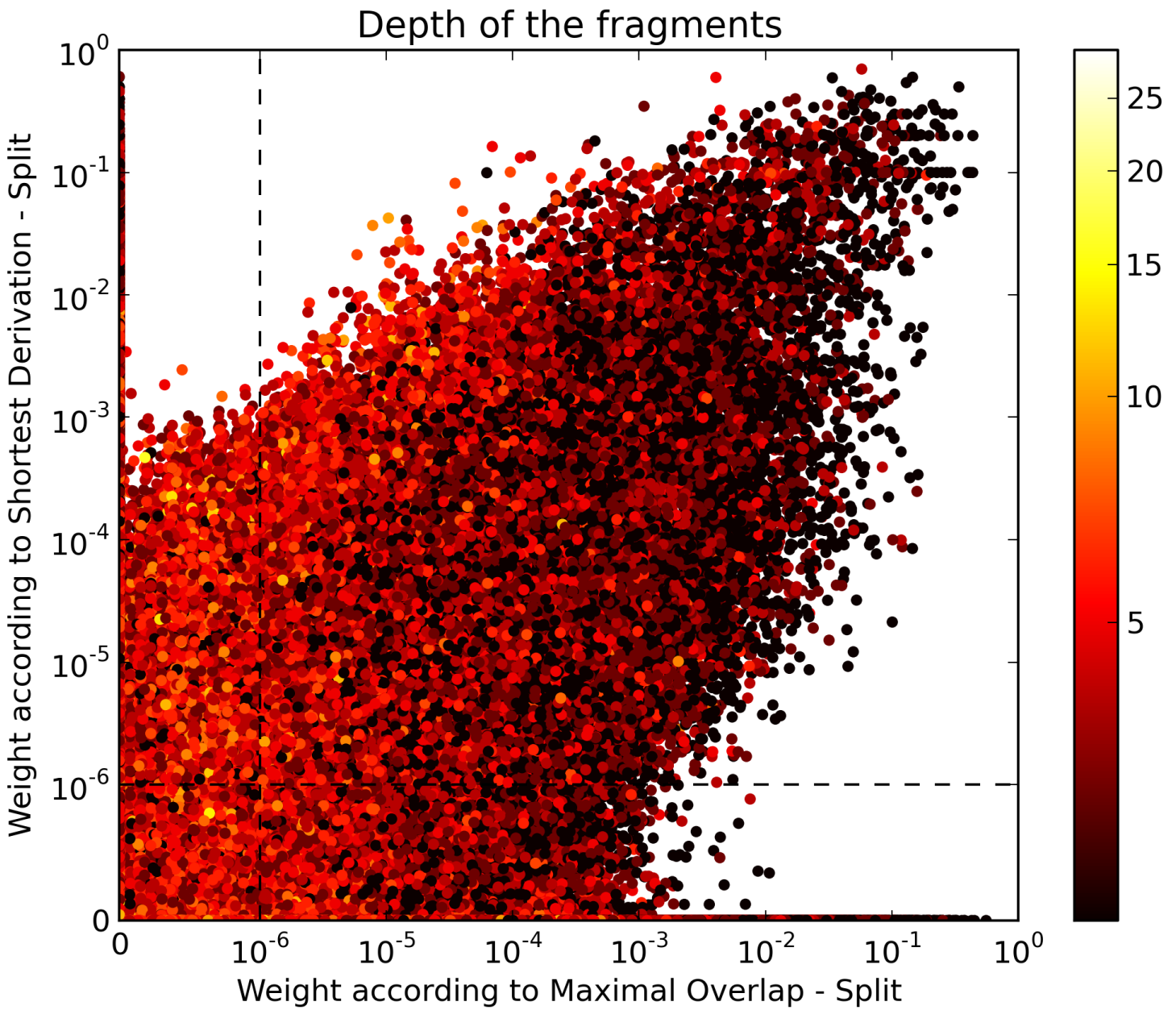
Double-DOP and DOP*: a comparison
- Introduction to Double-DOP and DOP*
- Comparison
- Experiments

Results
- Parsing Performance
- Analyzing grammars

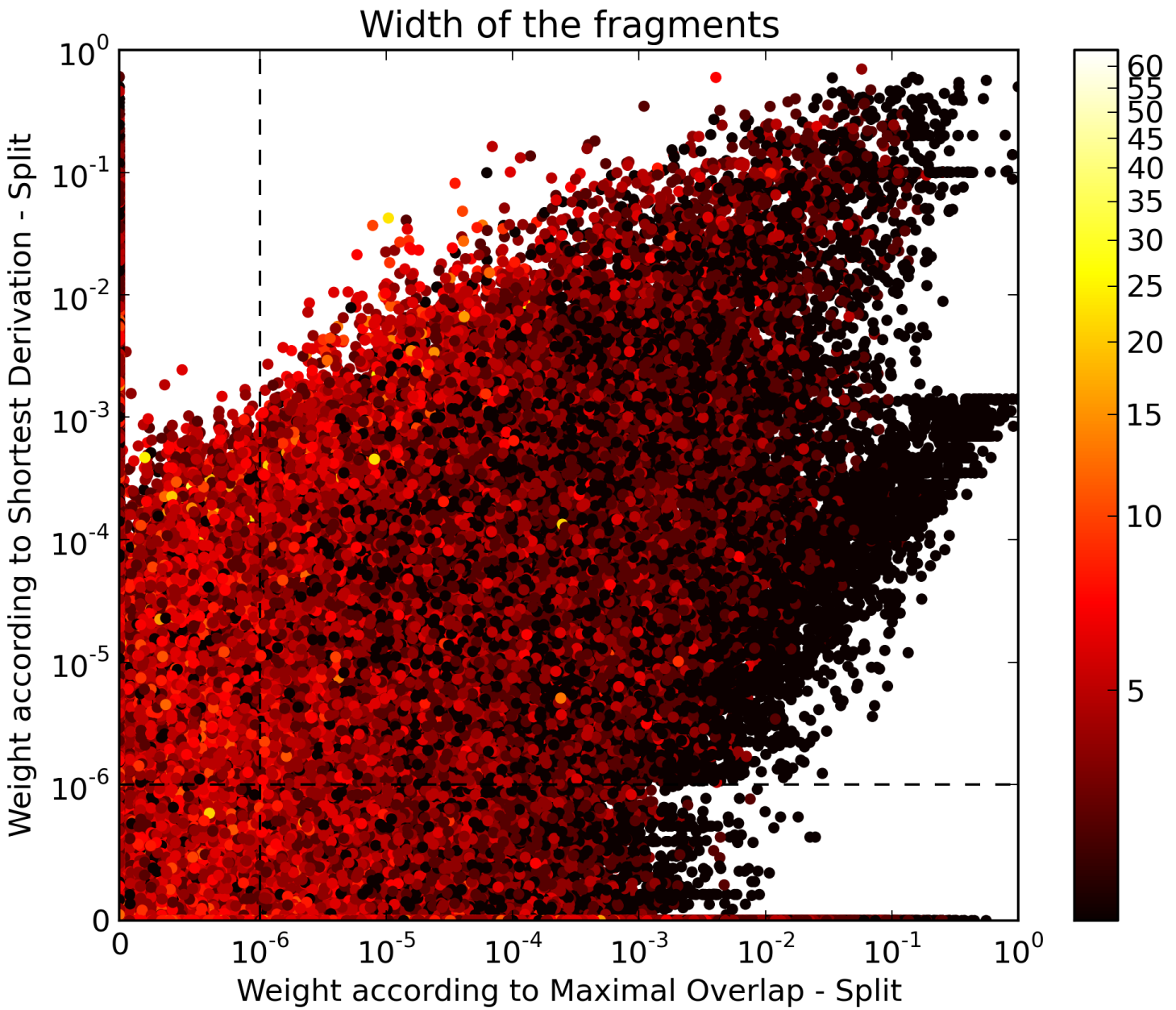
Split $\leftrightarrow$ Full



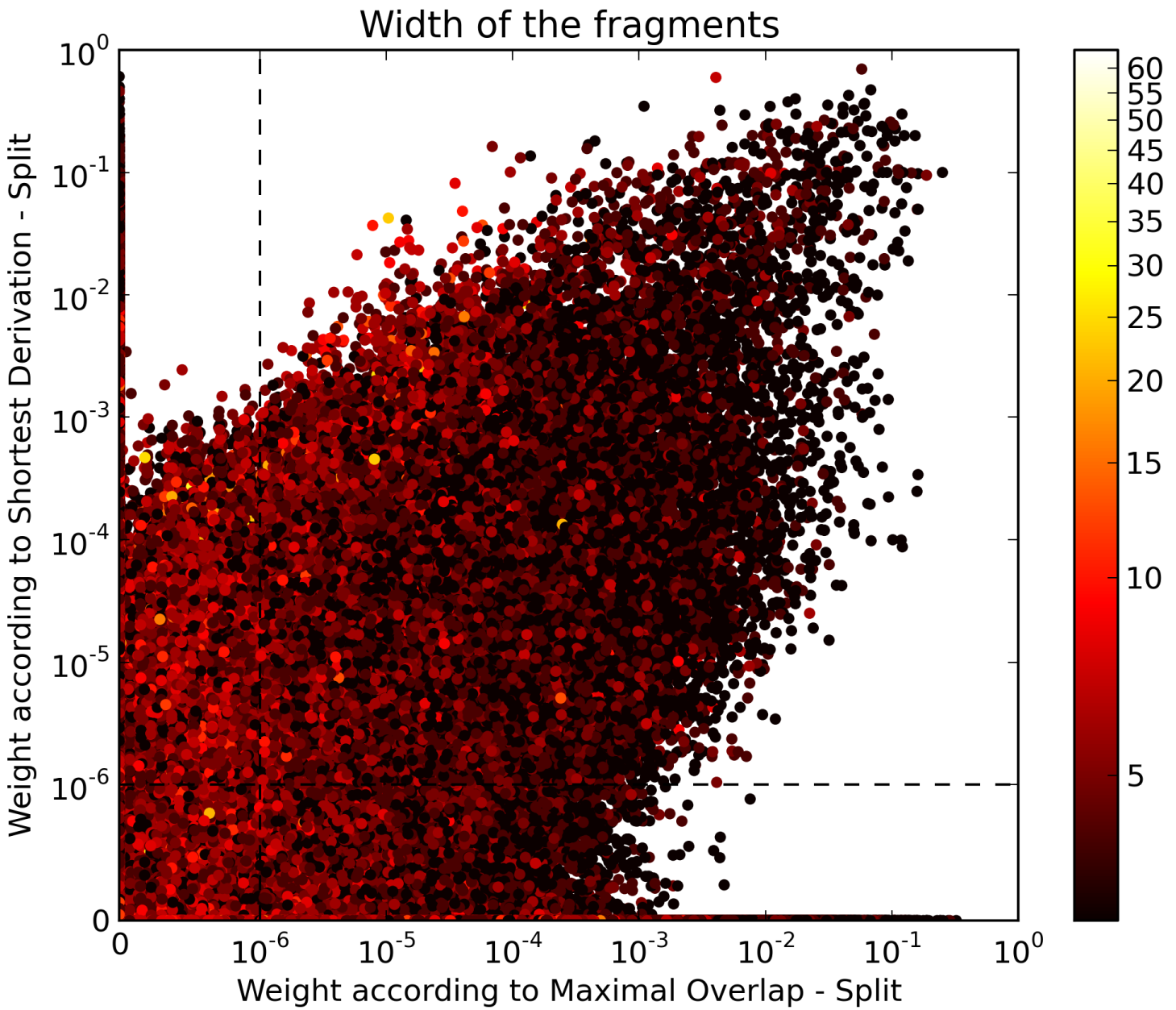
Maximal overlap $\leftrightarrow$ shortest derivation



Width


Width of the fragments
Weight according to Shortest Derivation - Split
Weight according to Maximal Overlap - Split
0
10^{-6}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
60
55
50
45
40
35
30
25
20
15
10
5

Width


Width of the fragments
Weight according to Shortest Derivation - Split
Weight according to Maximal Overlap - Split
0
10^{-6}
10^{-5}
10^{-4}
10^{-3}
10^{-2}
10^{-1}
10^{0}
60
55
50
45
40
35
30
25
20
15
10
5

Summary

- Shortest Derivation moves weight to larger fragments
- Split moves weight to smaller fragments
- Performance is not necessarily related to consistency: DOP* has bad parsing performance

- Outlook
  - Further analysis
  - Other estimators

Acknowledgments

- Andreas van Cranenburgh
- Khalil Sima'an