



Insertion Operator for Bayesian Tree Substitution Grammars

NTT Communication Science Laboratories

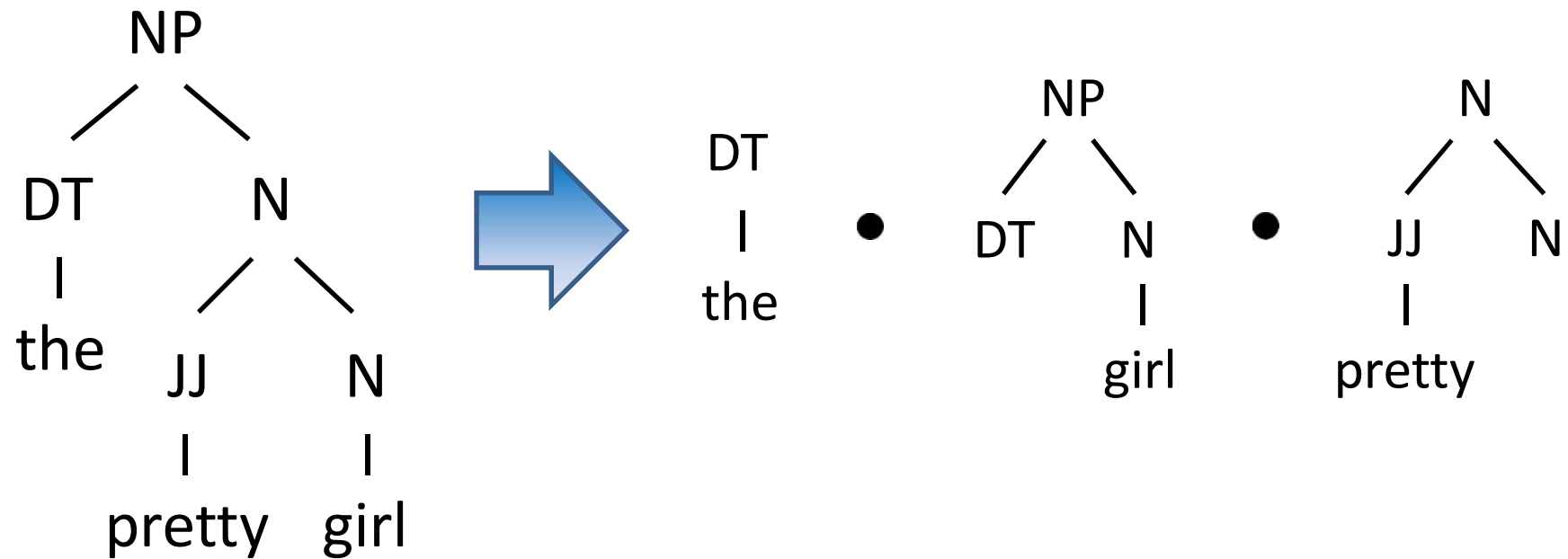
*[Hiroyuki Shindo](#) Akinori Fujino Masaaki Nagata



Task

Grammar induction from Treebank

Derivation



Previous Work

- Context free grammar (CFG)
 - Tree substitution grammar (TSG)
 - Generalization of CFG
 - No annotated derivations
 - ➡ Bayesian learning [Cohn et al. 09, Post et al. 09]
 - Tree Insertion grammar (TIG)
 - Tree Adjoining grammar (TAG)
 - No annotated derivations
 - ➡ Linguistic heuristics [Chang 03, Chen et al. 06]
 - ➡ Bayesian learning?
- Substitution operator
- Substitution operator
- Insertion operator

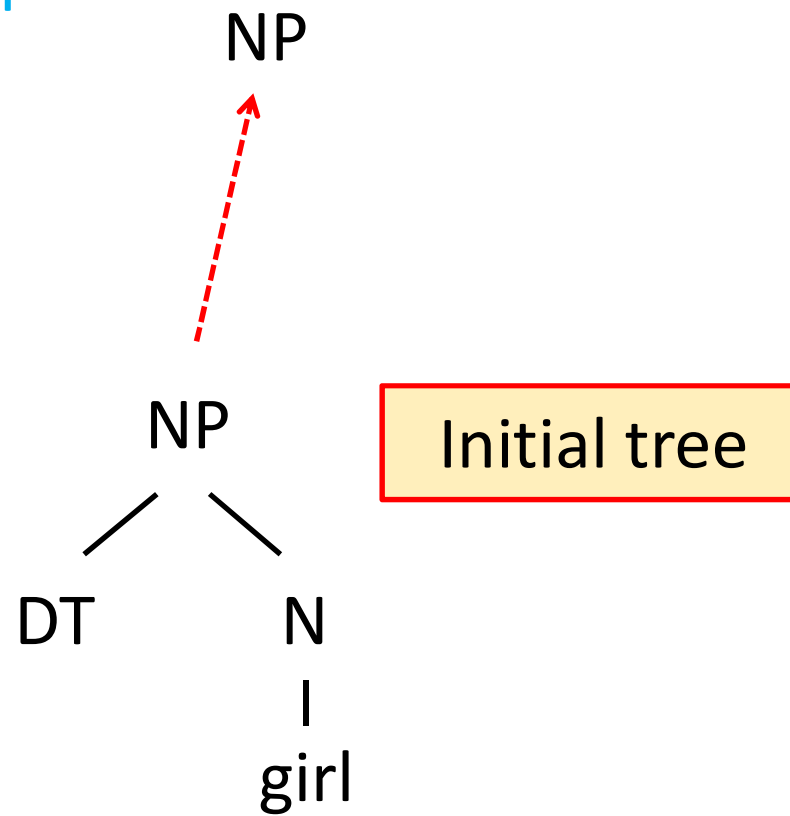
TSG

Substitution operator

NP

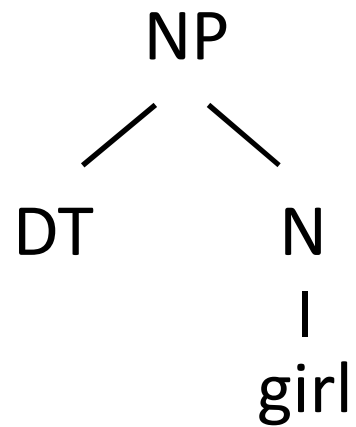
TSG

Substitution operator



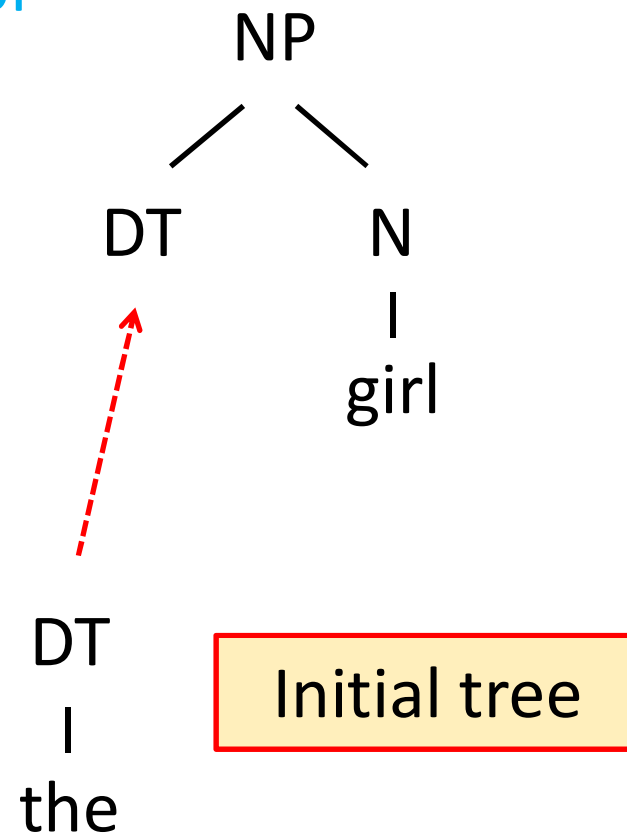
TSG

Substitution operator



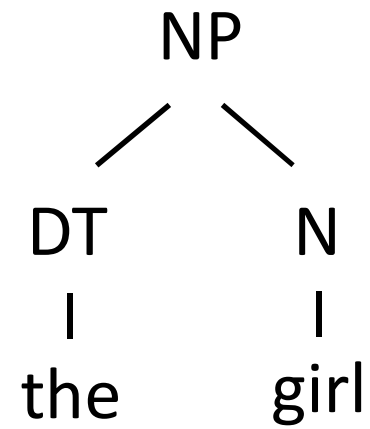
TSG

Substitution operator



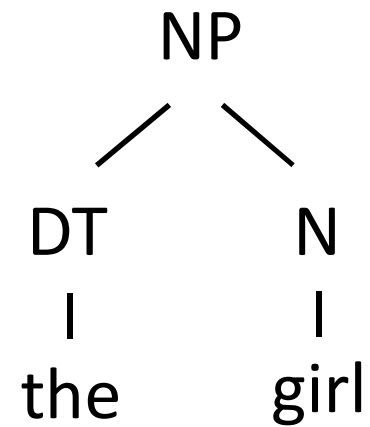
TSG

Substitution operator



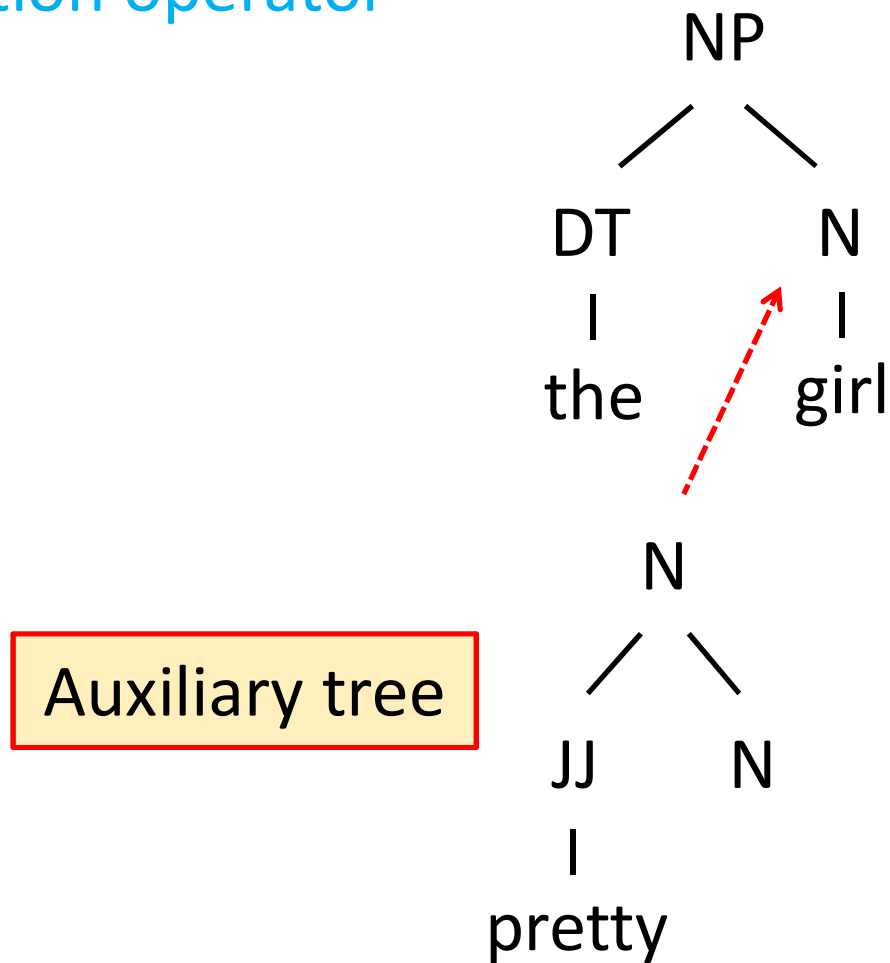
TSG + Insertion

Insertion operator



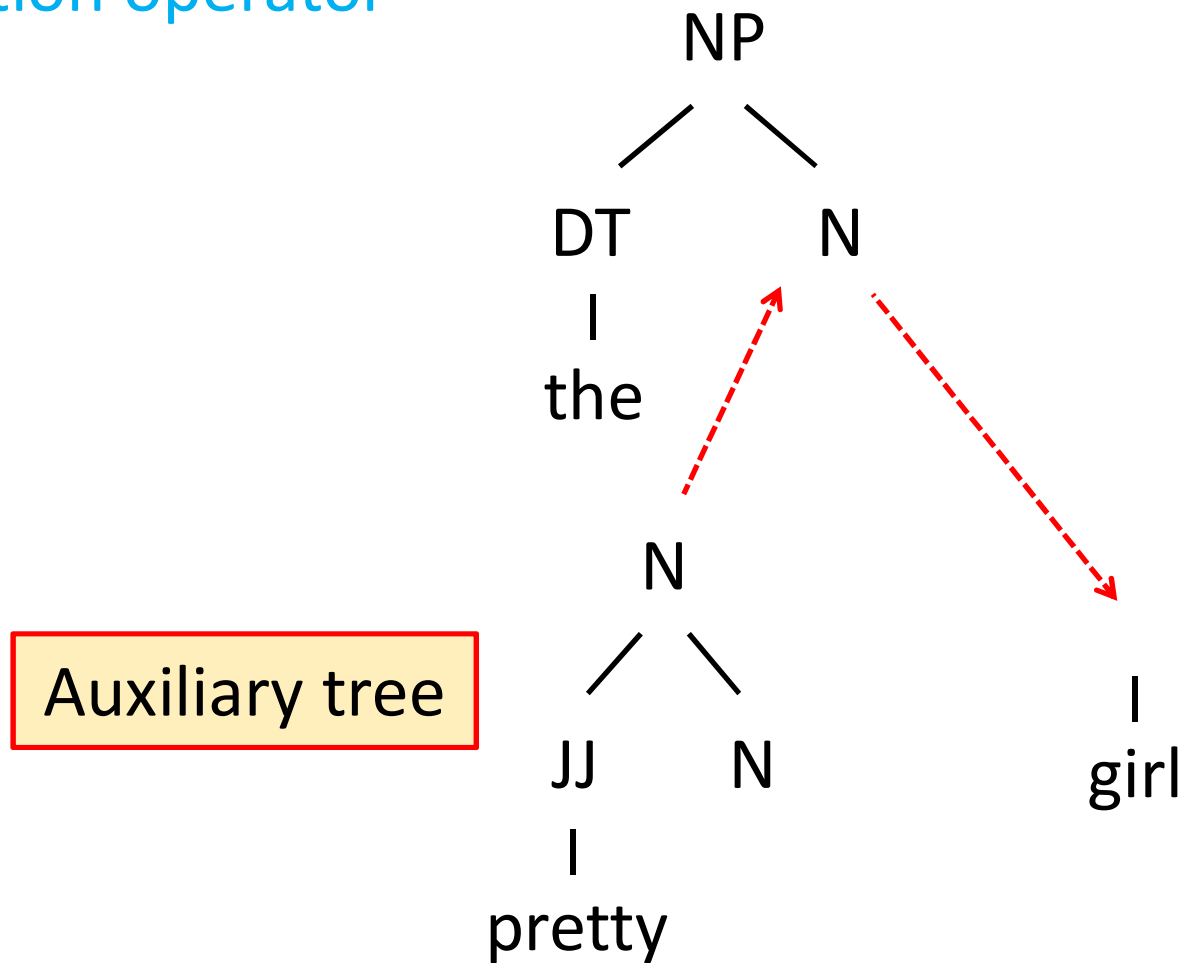
TSG + Insertion

Insertion operator



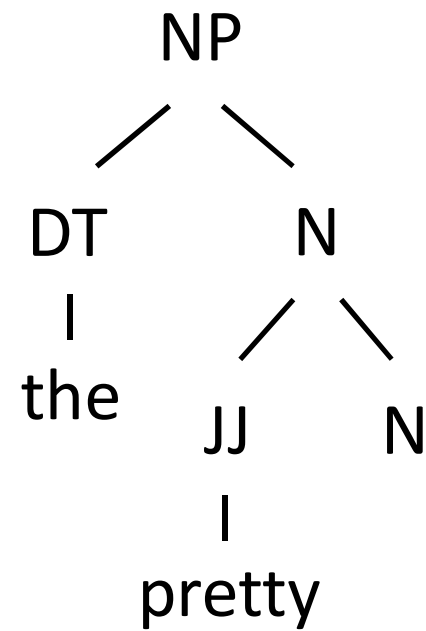
TSG + Insertion

Insertion operator



TSG + Insertion

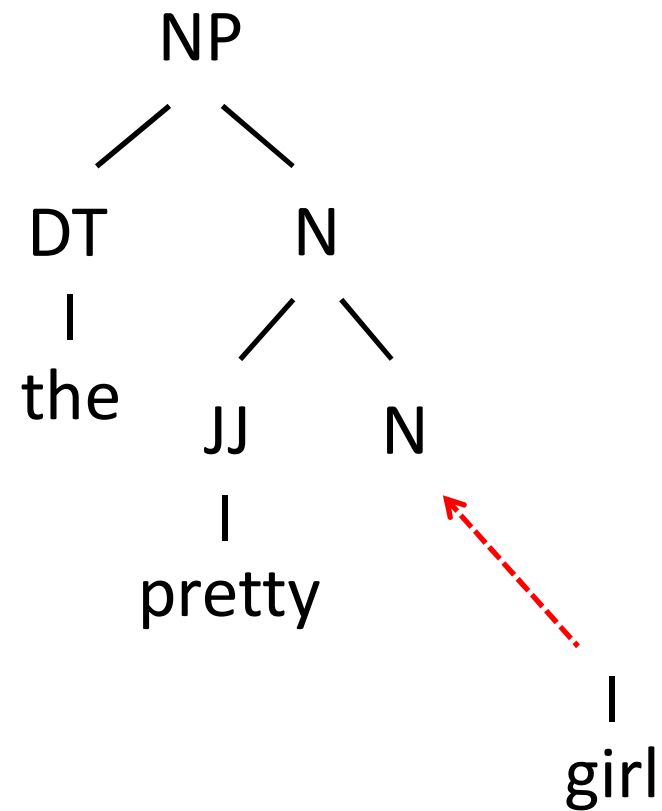
Insertion operator



|
girl

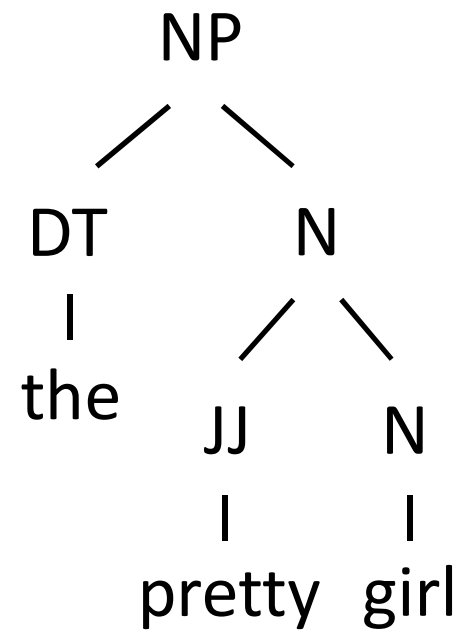
TSG + Insertion

Insertion operator



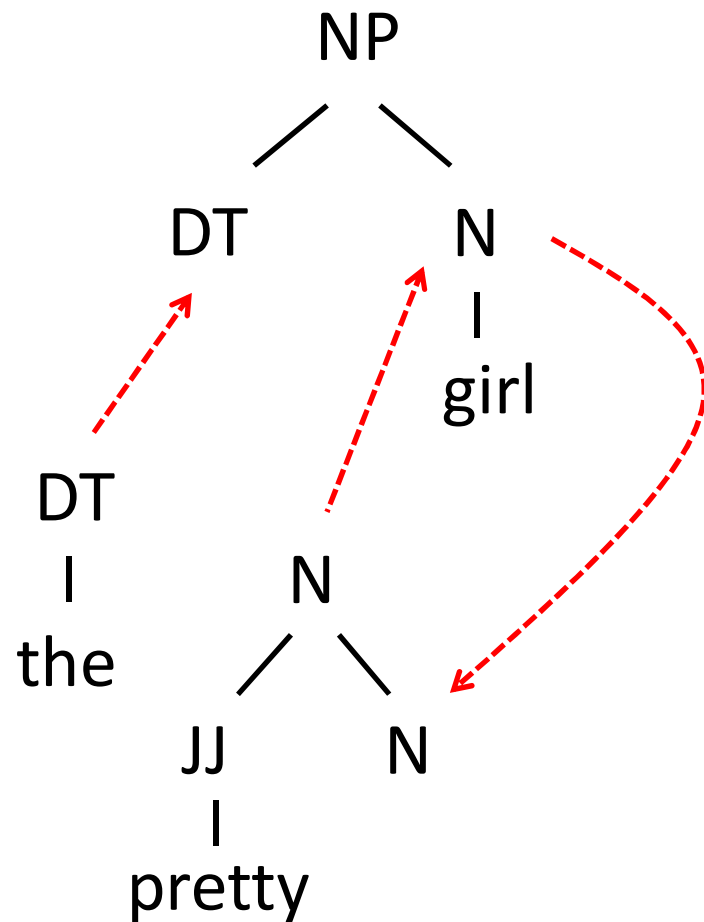
TSG + Insertion

Insertion operator

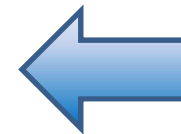
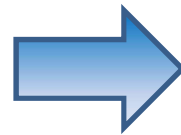


TSG + Insertion

Derivation

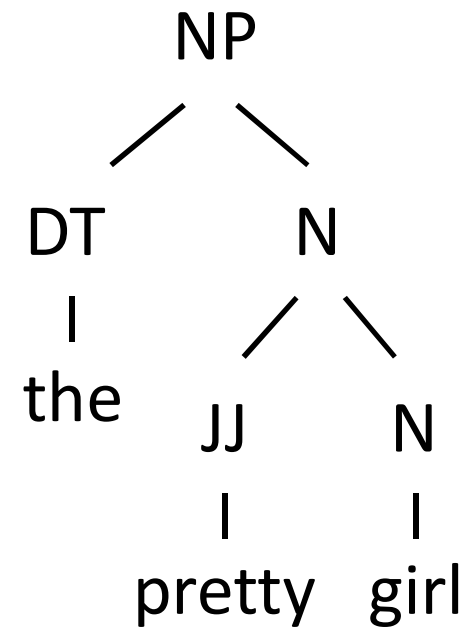


Generation



Inference

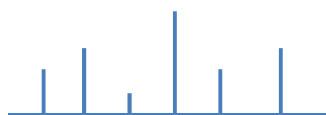
Observed



Probability Model of TSG + Insertion

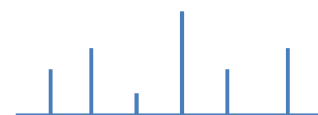
- Probability distribution over **initial** and **auxiliary** trees
- **Pitman-Yor process** prior [Pitman and Yor 97]

Initial trees



$$G_X \sim \text{PYP}(d_X, \theta_X, P_0)$$
$$e|X \sim G_X$$

Auxiliary trees



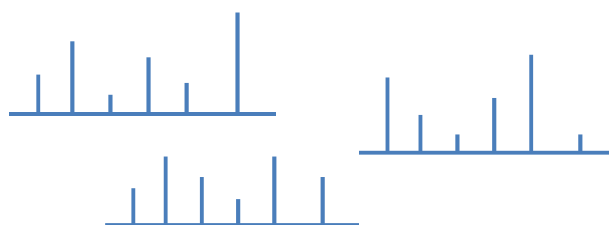
$$H_X \sim \text{PYP}(d'_X, \theta'_X, P_0)$$
$$f|X \sim H_X$$

Probability Model of TSG + Insertion

- Probability distribution over **initial** and **auxiliary** trees
- **Pitman-Yor process** prior [Pitman and Yor 97]

Base distribution

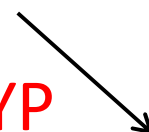
P_0



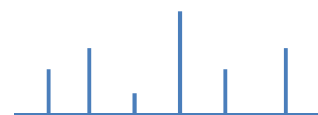
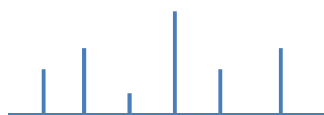
MLE of CFG



PYP



PYP



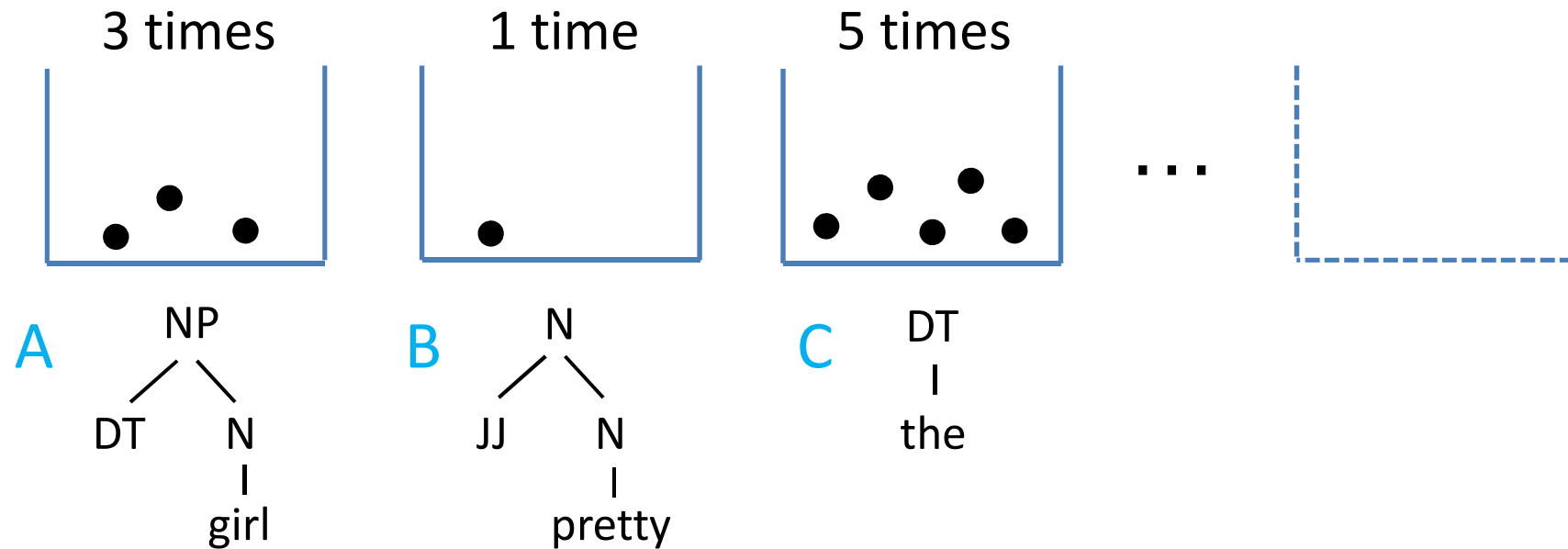
$$G_X \sim \text{PYP}(d_X, \theta_X, P_0)$$

$$e|X \sim G_X$$

$$H_X \sim \text{PYP}(d'_X, \theta'_X, P_0)$$

$$f|X \sim H_X$$

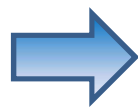
Pitman-Yor Process



- a) Draw subtree from *cache* with proportional to the frequency
- b) Draw new subtree (= create a box) from the base distribution



"Rich get richer" statistics



Encourage compact grammar

Inference

Blocked Metropolis-Hastings Sampler

[Johnson et al. 07, Cohn et al. 10]

For each sentence,

1. Calculate the inside probability.
2. Sample a derivation tree in a top-down manner.
3. Accept or reject the sample by MH test.

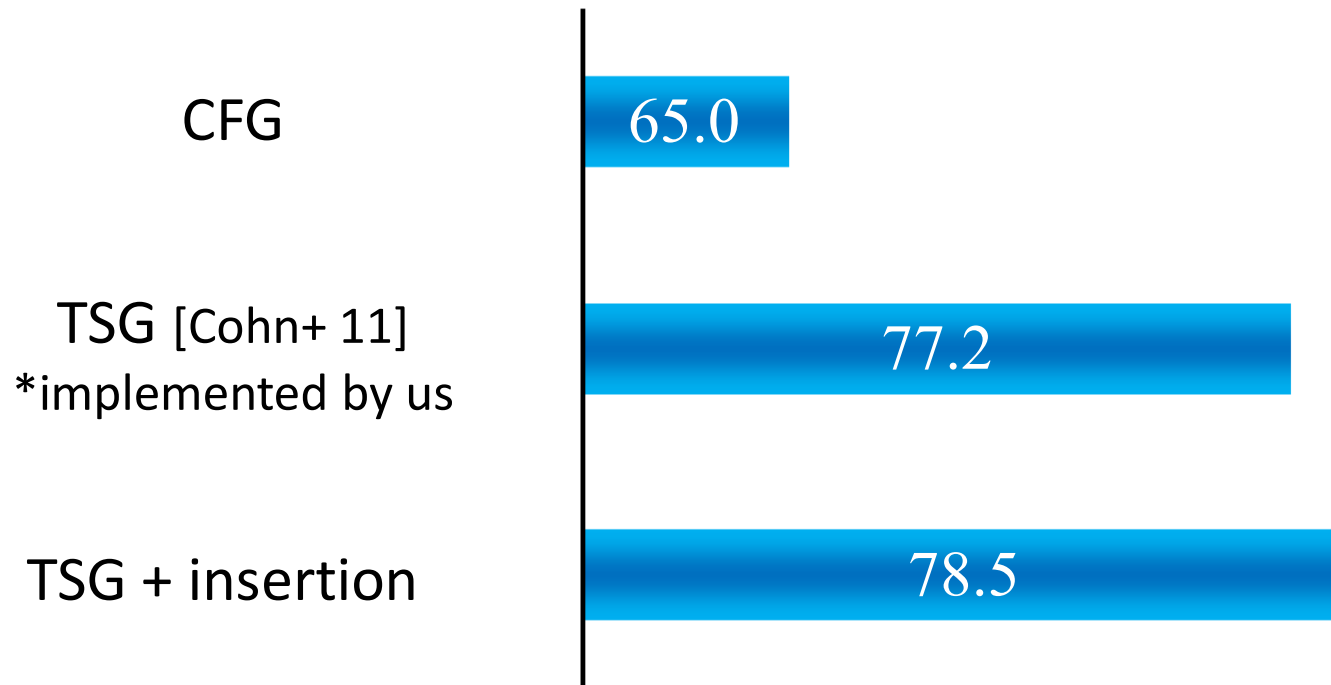
Experiment

Parsing performance

Experiment

Parsing performance (small dataset)

F1 score

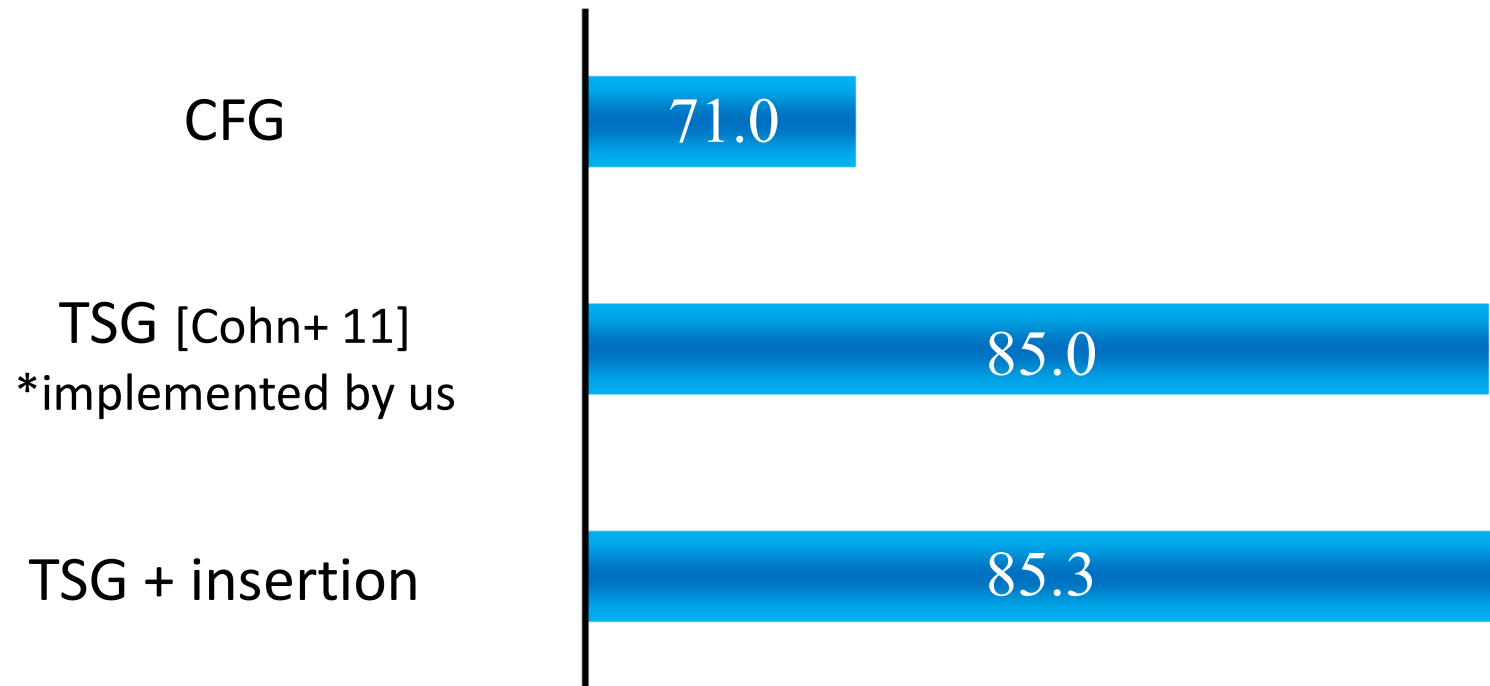


WSJ Penn Treebank (training: sec. 2, test: sec. 22)

Experiment

Parsing performance (full treebank dataset)

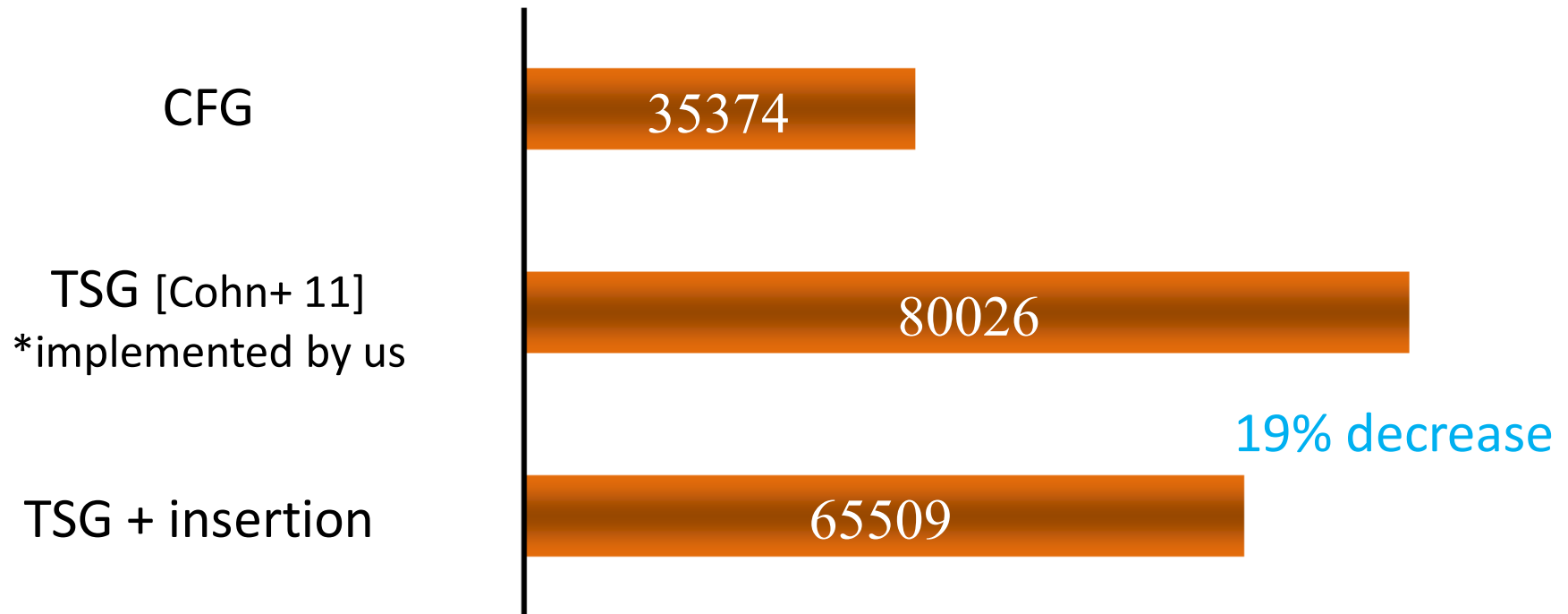
F1 score



WSJ Penn Treebank (training: sec. 2-21, test: sec. 23)

Experiment

The number of subtrees (initial and auxiliary trees)



WSJ Penn Treebank (training: sec. 2-21, test: sec. 23)

Experiment

Examples of auxiliary trees obtained by our model

- (NP (NP) (: -))
- (NP (NP) (ADVP (RB respectively)))
- (PP (PP) (, ,))
- (VP (VP) (RB then))
- (VP (VP) (RB not))
- (QP (QP) (IN of))
- (S (S) (: ;))

Summary

Insertion operator for Bayesian TSG model

Model:

- Distribution over initial and auxiliary trees
- Pitman-Yor process prior for compact grammar

Inference:

- Efficient MCMC sampler

Results:

- Outperforms CFG and TSG on small dataset
- Identical performance on large dataset,
making the grammar size 19% smaller than TSG



Thank You!

