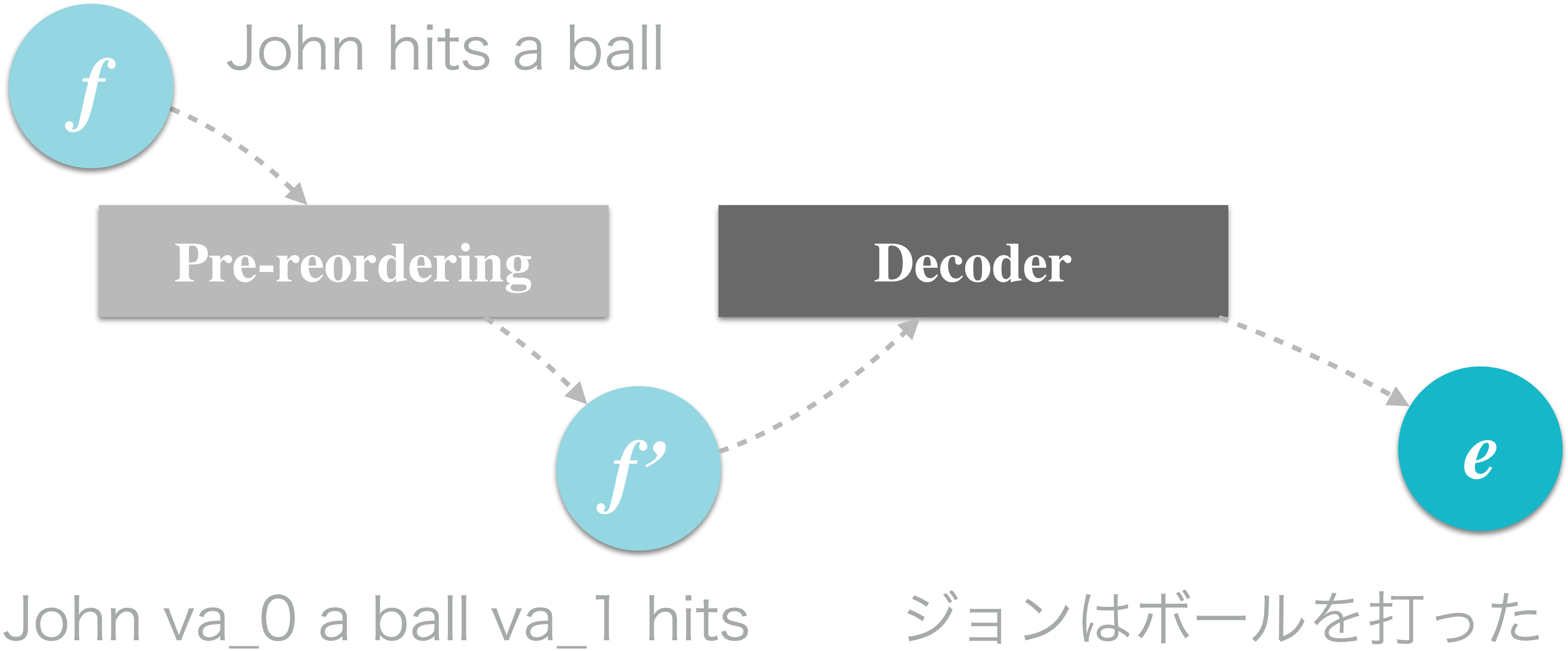
Weblio Pre-reordering SMT System

Zhongyuan Zhu @ WAT2014

Overview of pre-reordering systems

- Reorder input text before translation



John hits a ball

f

Pre-reordering

Decoder

f'

e

John va_0 a ball va_1 hits

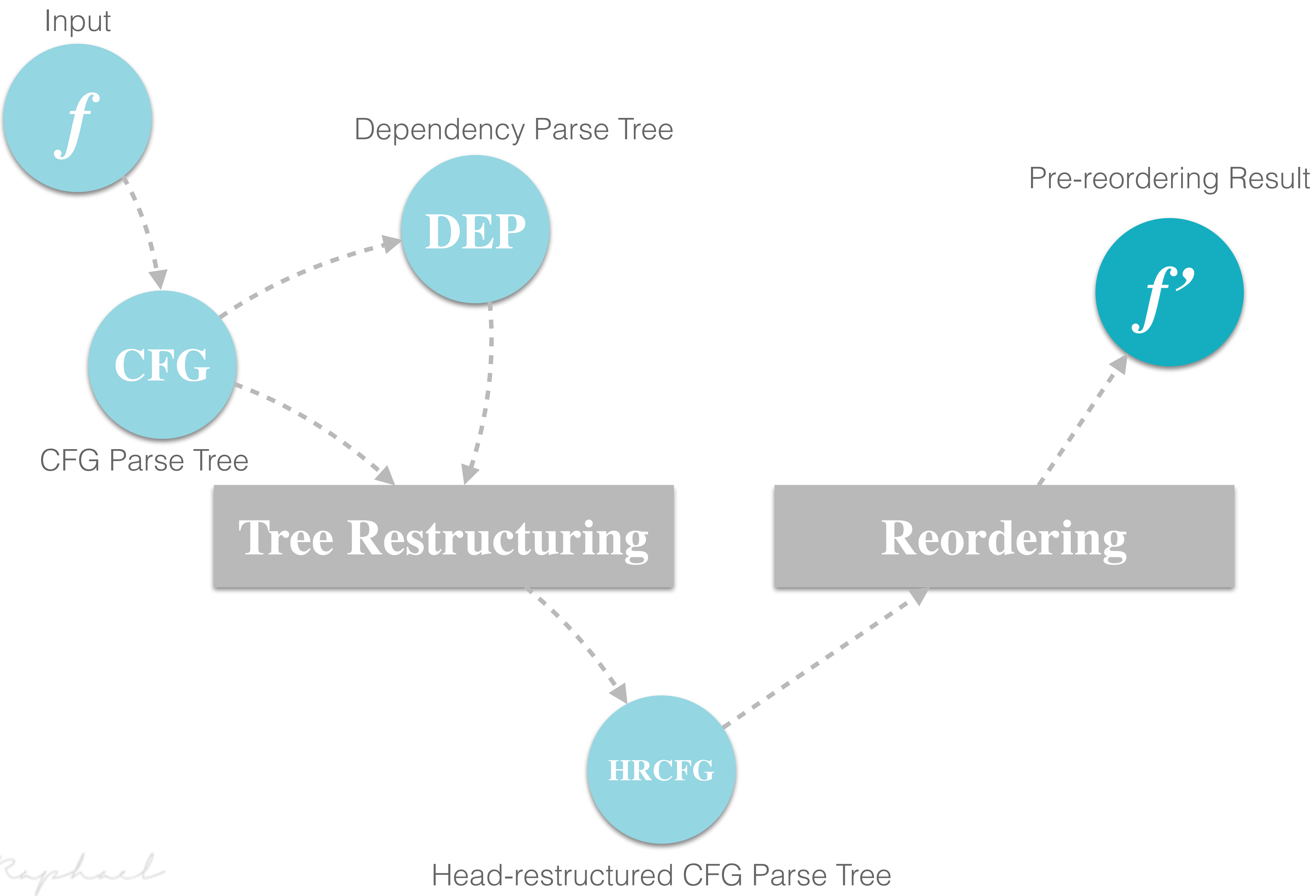
ジョンはボールを打った

Approaches of pre-reordering

- Syntactic pre-reordering with parse trees
  - Rule-based
    - Head-finalization (Isozaki et al., 2010)
  - Supervised learning with word alignments
    - Automatically learning Rewrite Patterns (Xia and McCord, 2004)
- Syntactic pre-reordering without parse tree
  - LADER (Neubig et al., 2012)

Pre-reordering model in our system

Overview of our pre-reordering system



Input

f

Dependency Parse Tree

DEP

CFG

CFG Parse Tree

Tree Restructuring

Reordering

Pre-reordering Result

f'

HRCFG

Head-restructured CFG Parse Tree

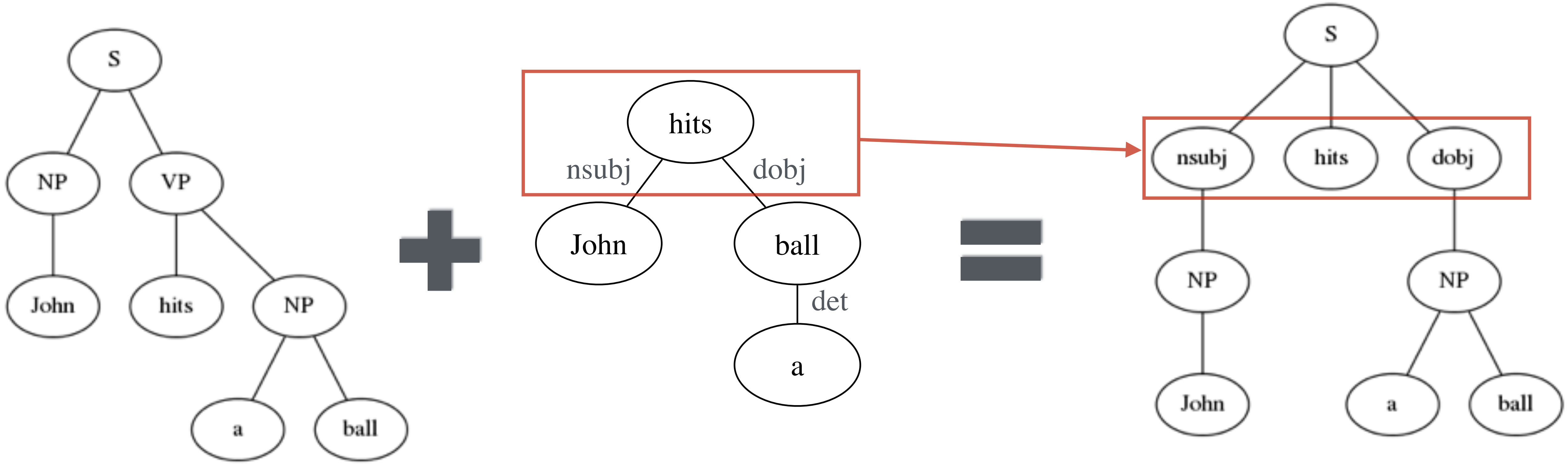
Head-restructured CFG Parse Tree

- Problem of CFG parse tree
  - Hard to capture long-distance reordering patterns
- Problem of Dependency parse tree
  - Fully lexicalized parse tree leads to a sparse reordering model

Head-restructured CFG Parse Tree

- Our approach
  - Restructure a CFG parse tree to inject head information into it

Head word is always lexicalized



CFG Parse Tree

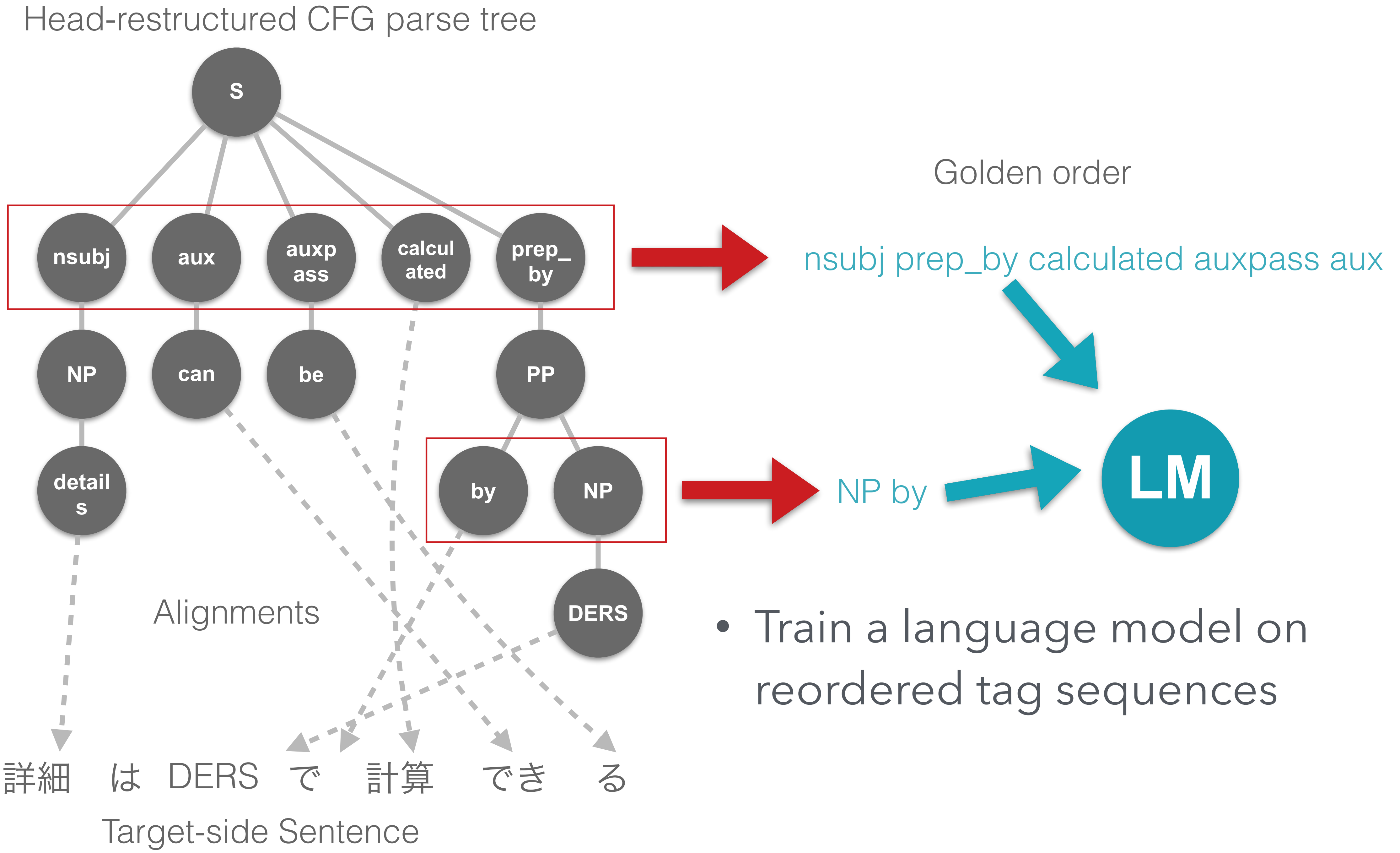
Dependency Parse Tree

Head-restructured CFG Parse Tree (HRCFG)

Raphael

Learning reordering model based on LM

- Extract tag sequences in golden order



nsubj prep_by calculated auxpass aux

NP by

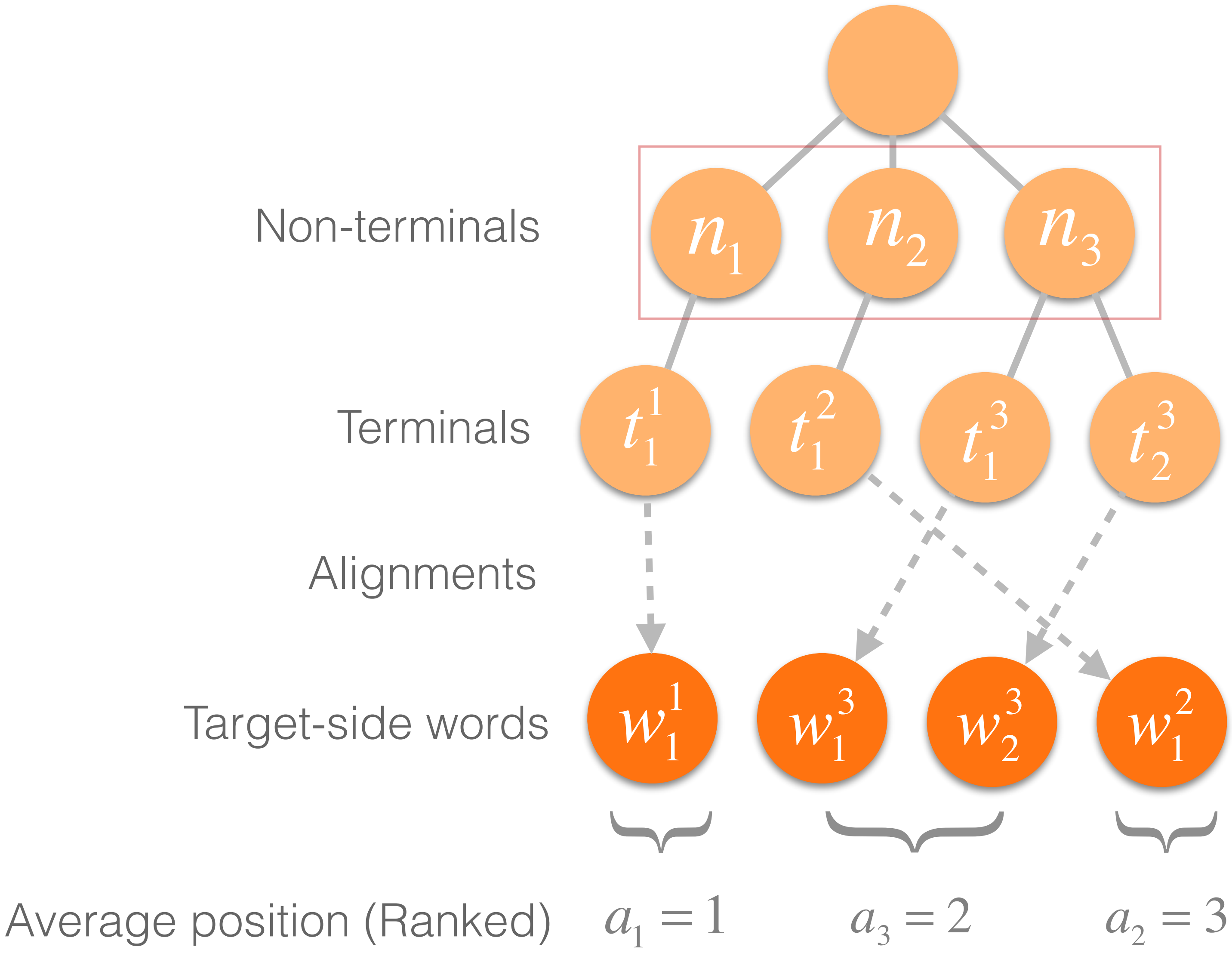
- Train a language model on reordered tag sequences

Raphael

Finding golden order with word alignments

- Given a bilingual sentence pair, source-side parse tree and word alignments,

  the golden order of a node layer is defined as



For nodes $(n_1, n_2, \ldots, n_k)$

Initial order:

$$o_0 = (1, 2, \ldots, k)$$

Golden order:

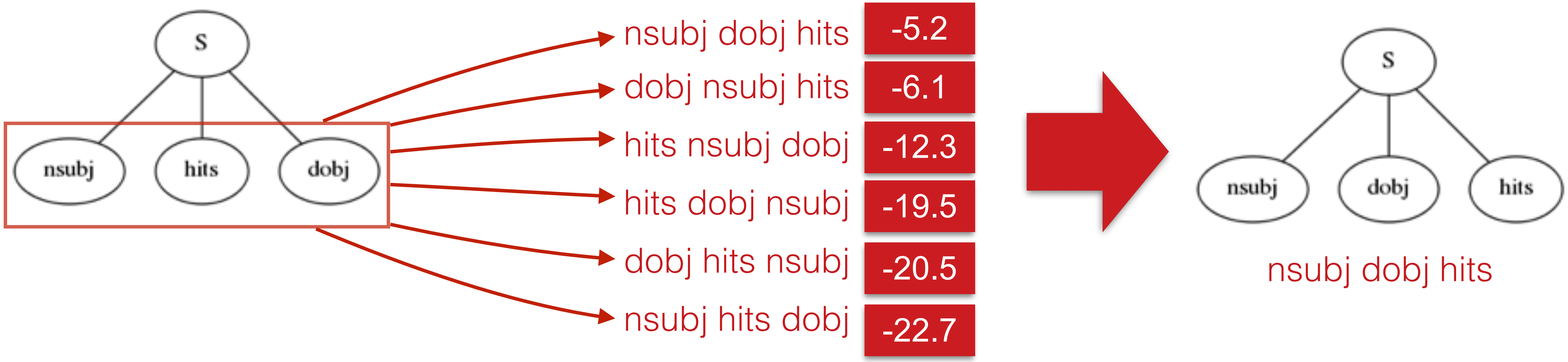
$$\hat{o} = (a_1, a_2, \ldots, a_k)$$

Raphael

Reordering a input parse tree

1.List all possible orders for a treelet

3. Select the best order to adjust the treelet

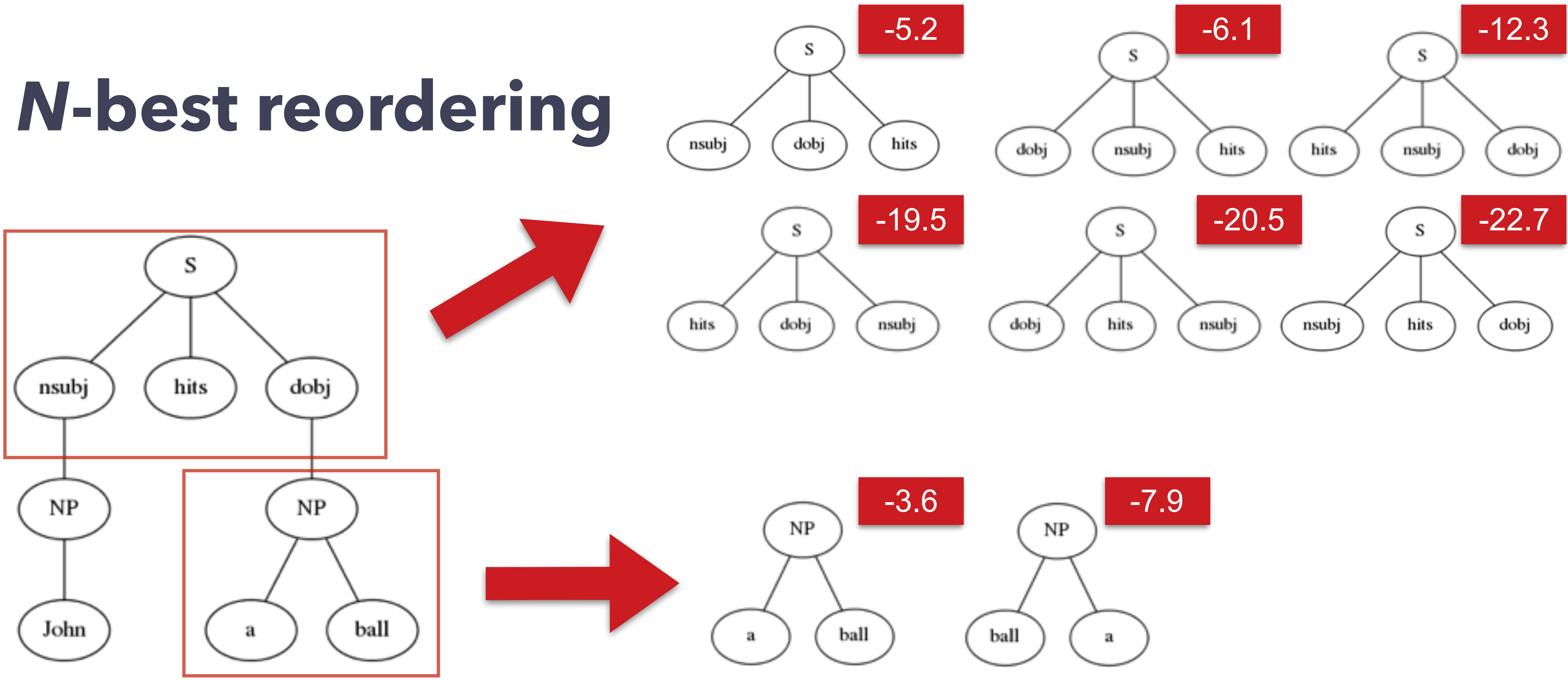


nsubj dobj hits	-5.2
dobj nsubj hits	-6.1
hits nsubj dobj	-12.3
hits dobj nsubj	-19.5
dobj hits nsubj	-20.5
nsubj hits dobj	-22.7

nsubj dobj hits

2.Score them with language model

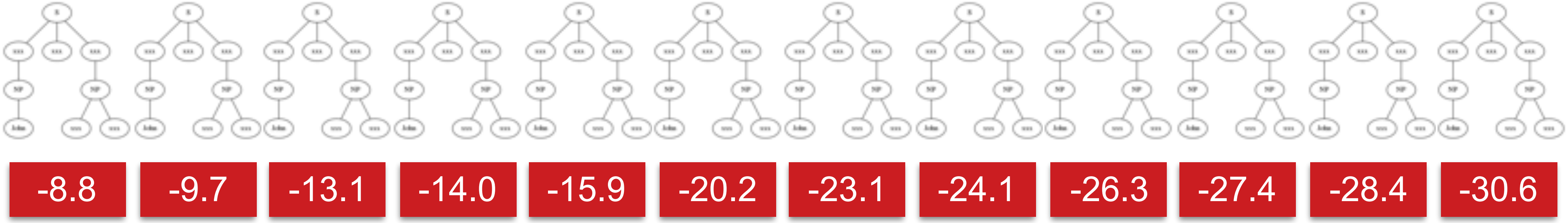
N-best reordering



Reordered treelets with LM scores

All 12 possible combinations here

Selected *N*-best results by accumulated scores (Cube Pruning is applied in the practice)



Experiments

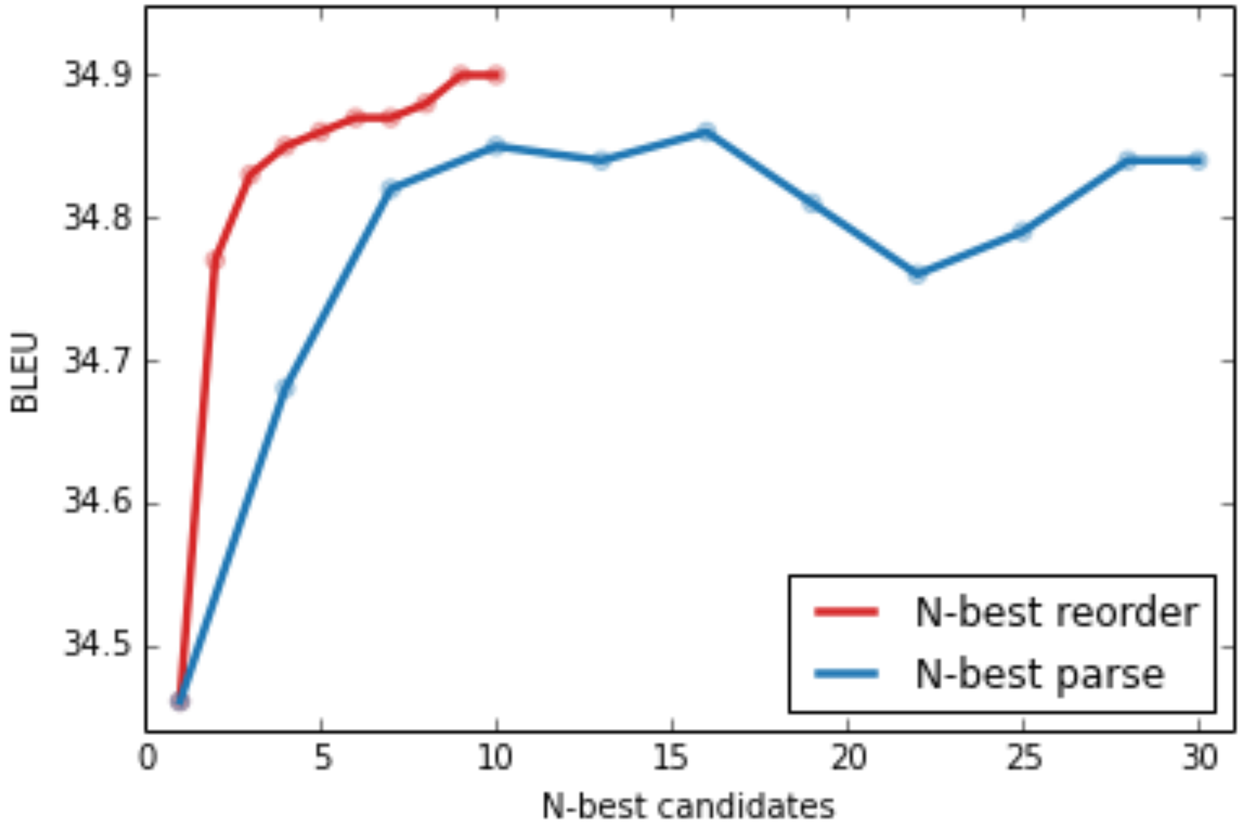
In-house experiments

	BLEU	RIBES
1-best parse + 1 best reorder	34.46	0.7817
N-best parse + 1 best reorder	34.80	0.7851
1-best parse + *N*-best reorder	34.90	0.7857
N-best parse + *N*- best reorder	**35.10**	**0.7887**

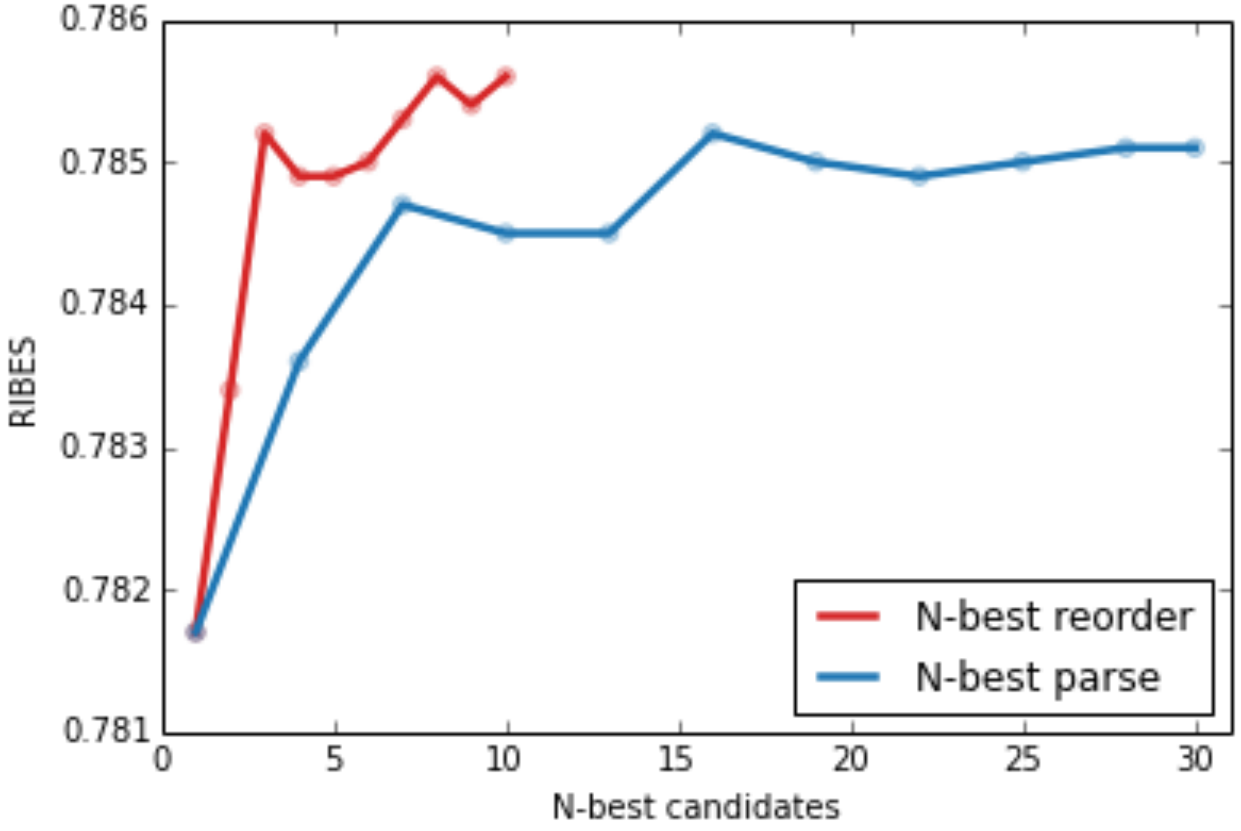
- For "*N*-best reorder", 10 candidate reordering results are considered.
- For "*N*-best parse", 30 candidate parse trees are considered.
- We select the final translation by the sum of translation score (given by decoder) and the score of pre-reordering.

N-best reordering & *N*-best parse tree inputs

- Incorporating multiple reordering results and parse trees benefits automatic scores.



BLEU

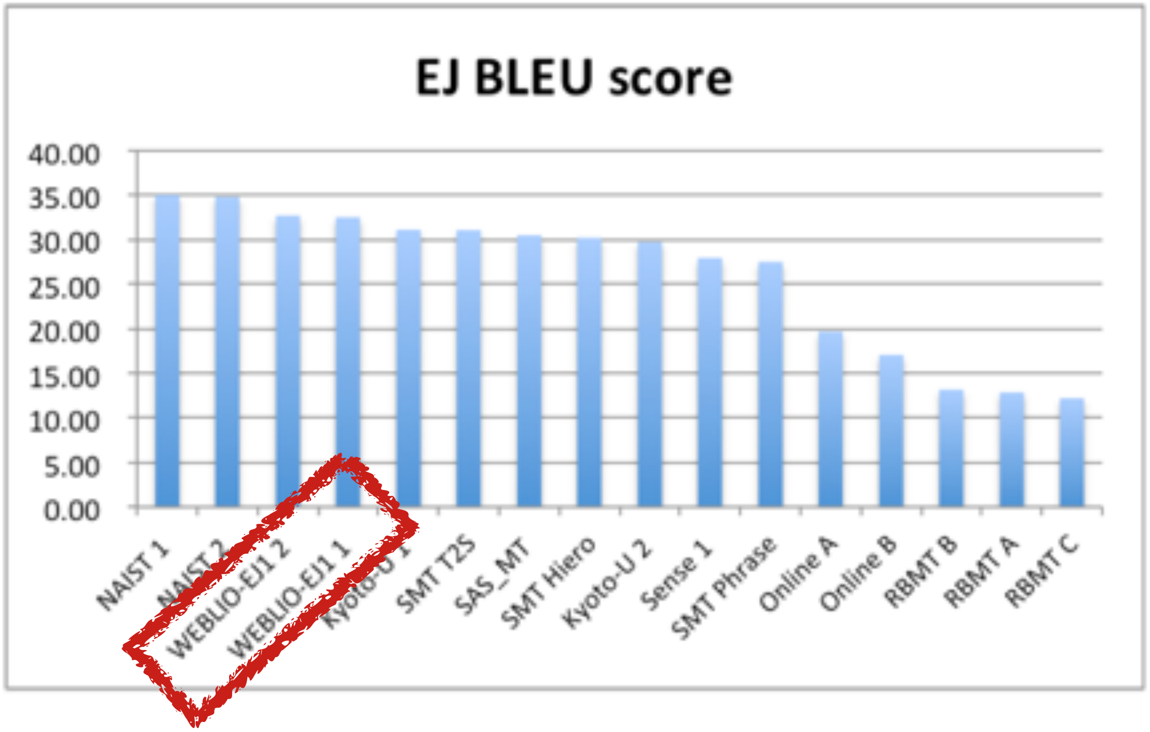


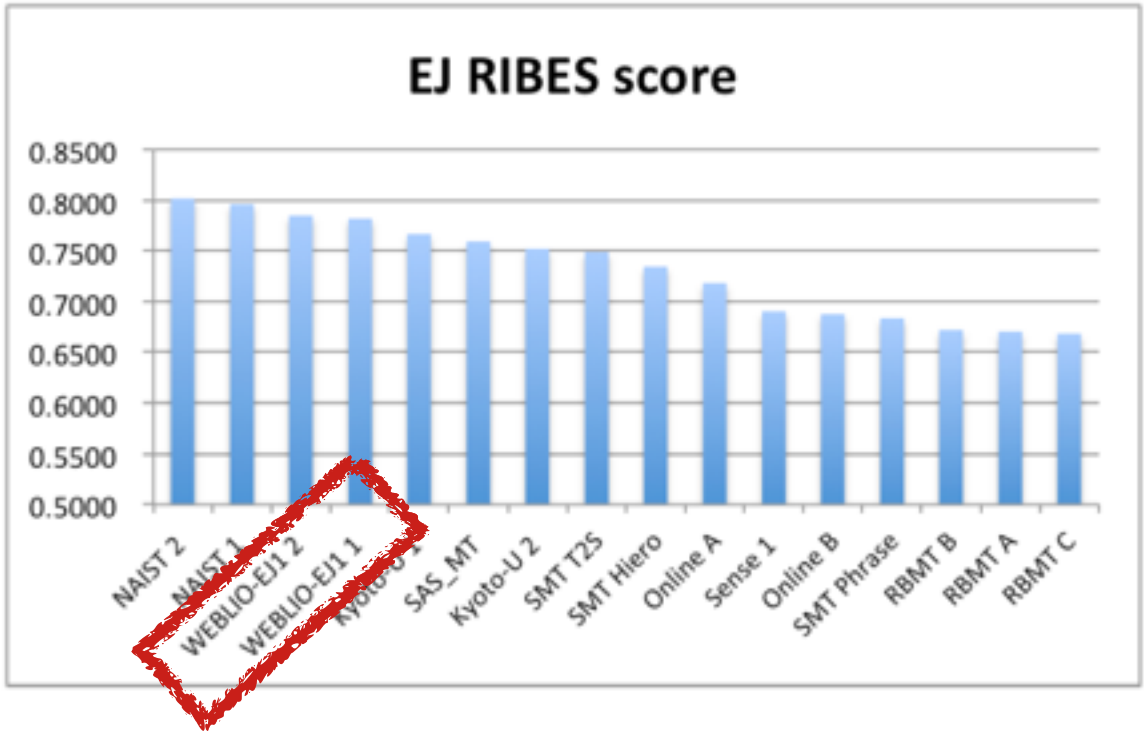
RIBES

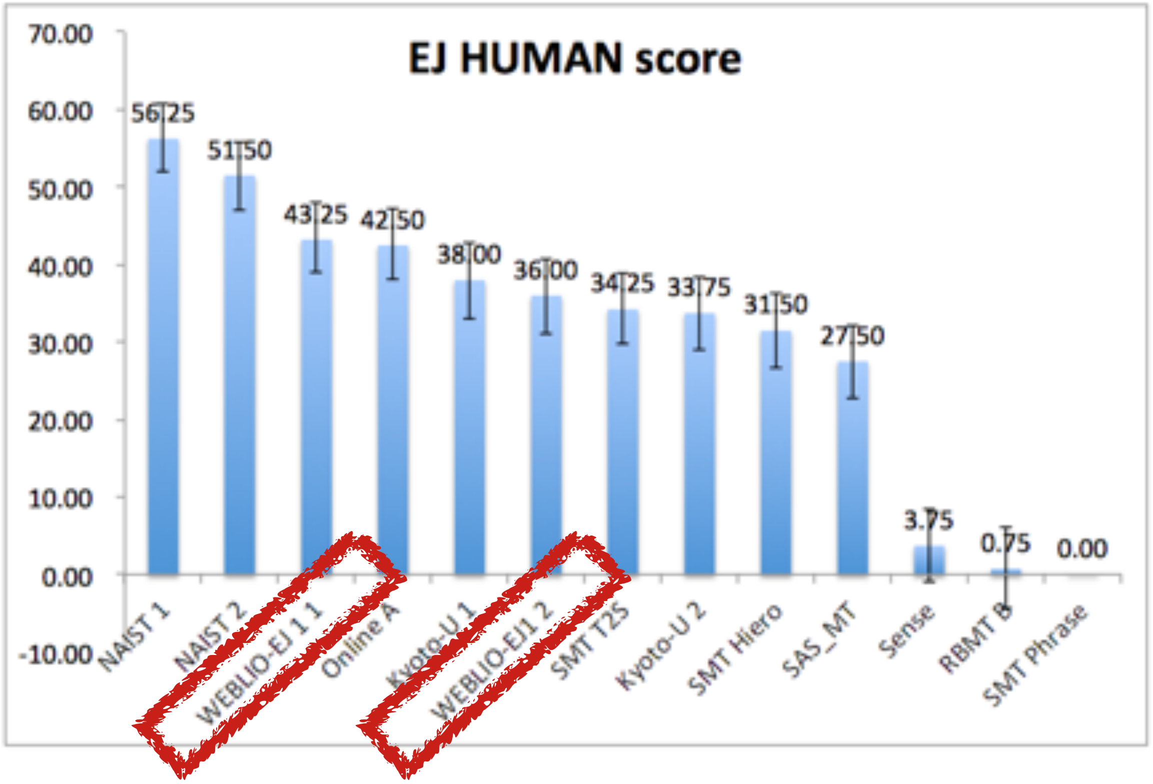
Official evaluation results

	BLEU	RIBES	HUMAN
N-best reorder	34.87	0.7869	+43.25
N-best reorder + N-best parse	35.04	0.7900	+36.00
BASELINE PBMT	29.80	0.6919	0.00

Official evaluation results



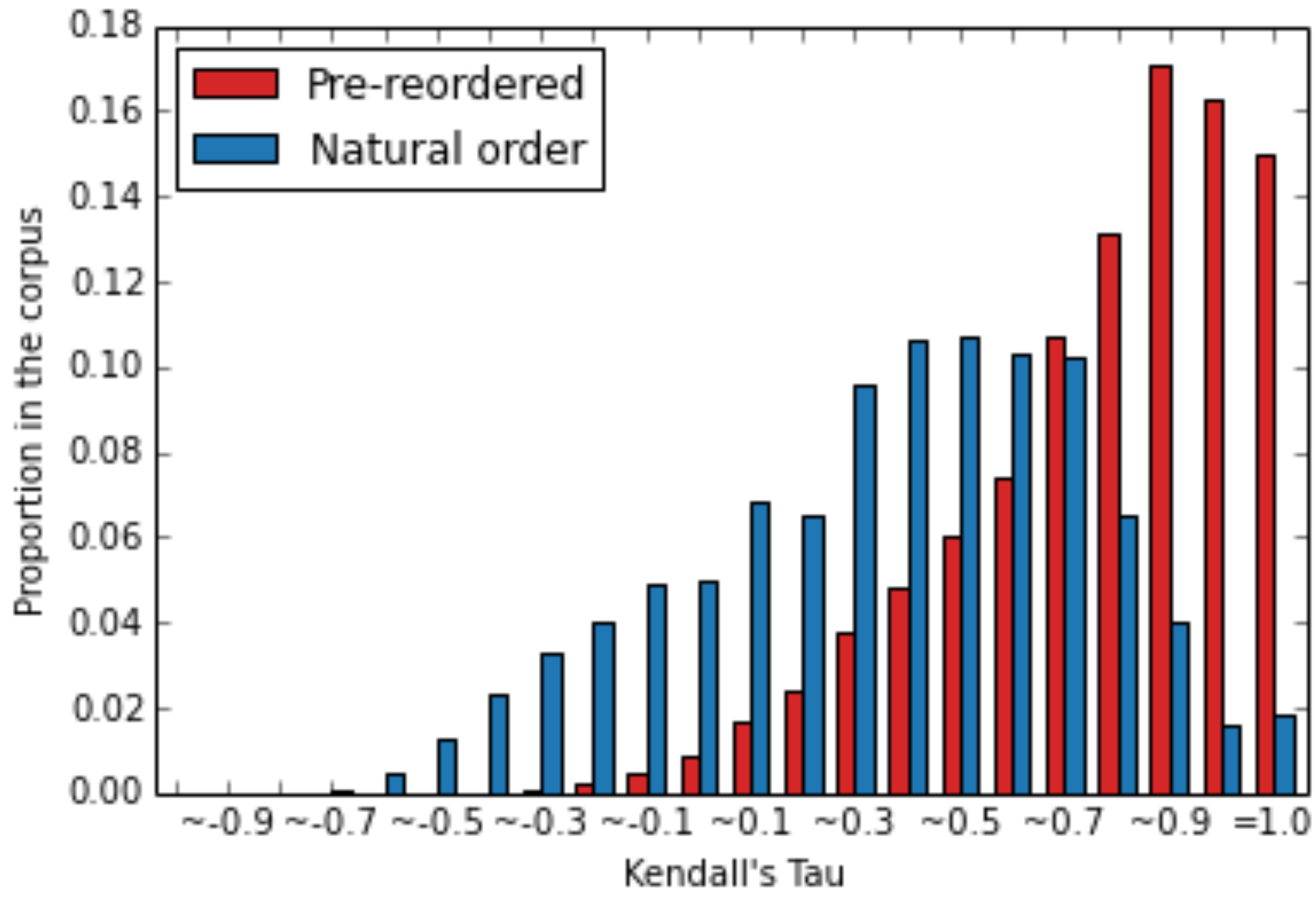




Raphael

Effect of pre-ordering

- Identical ordered sentences increases to 15%



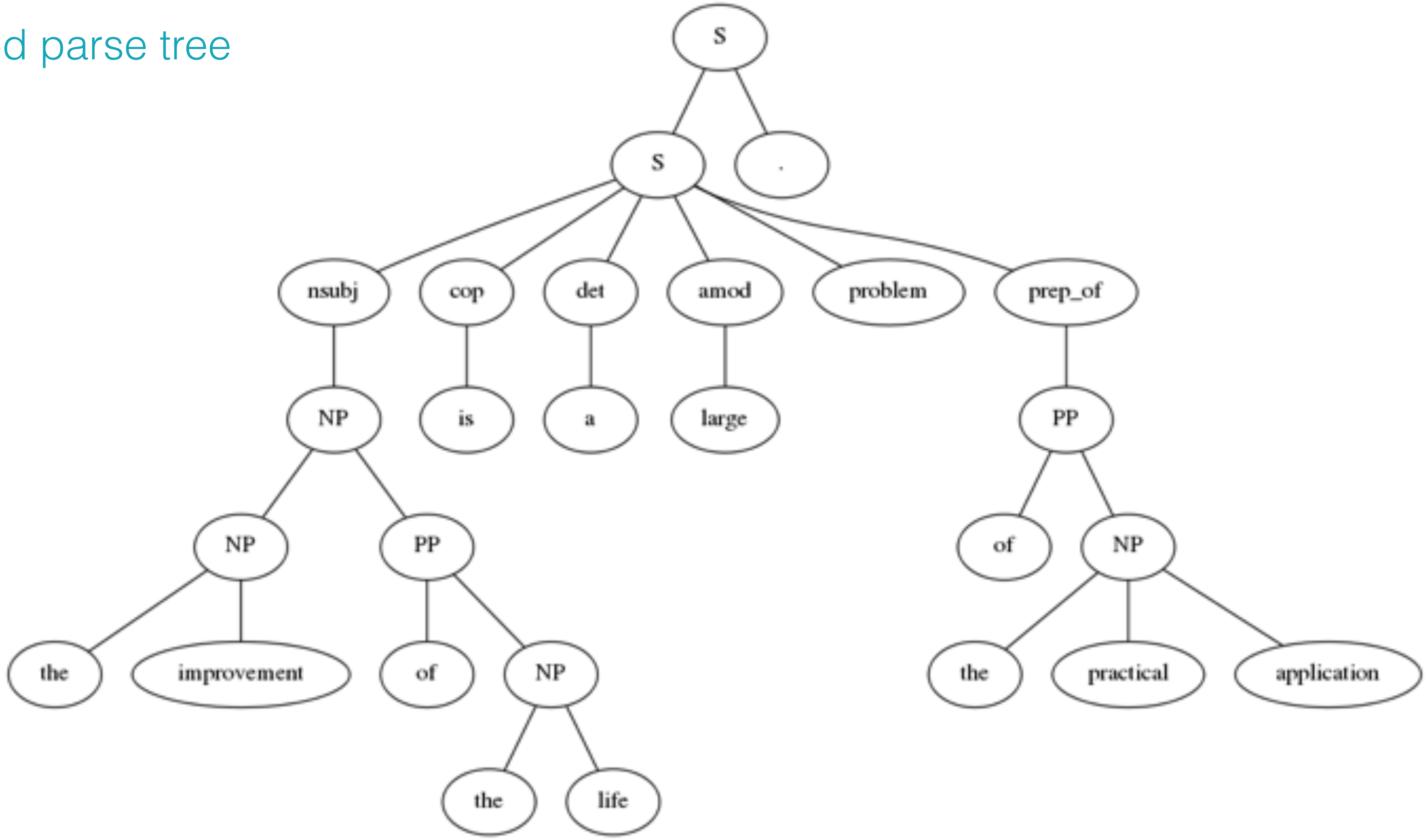
Closer in order

Example of pre-reordering

Original input

the improvement of the life is a large problem of the practical application.

Restructured parse tree



Reordered input

the life of the improvement va_nsubjpass the practical application of a large problem is .

Reference

寿命 の 向上 が 実用 化 の 大きな 課題 で あ る 。

Review

- Language model is just a quick solution to the reordering problem, sometimes it fails in simple cases.
  - Sparseness problem
- To gain more from forest input, it’s necessary to integrate it inside the pre-reordering model.

Online demonstrations

Head-restructured CFG parse tree

http://raphael.uaca.com/demos/hdtree

Pre-reordering

http://raphael.uaca.com/demos/raphreorder

Thanks.