

These slides
are taken
from

Forest-Based Translation Rule Extraction



Haitao Mi

Institute of Computing Technology



中科院计算所
INSTITUTE OF
COMPUTING
TECHNOLOGY

Liang Huang

University of Pennsylvania



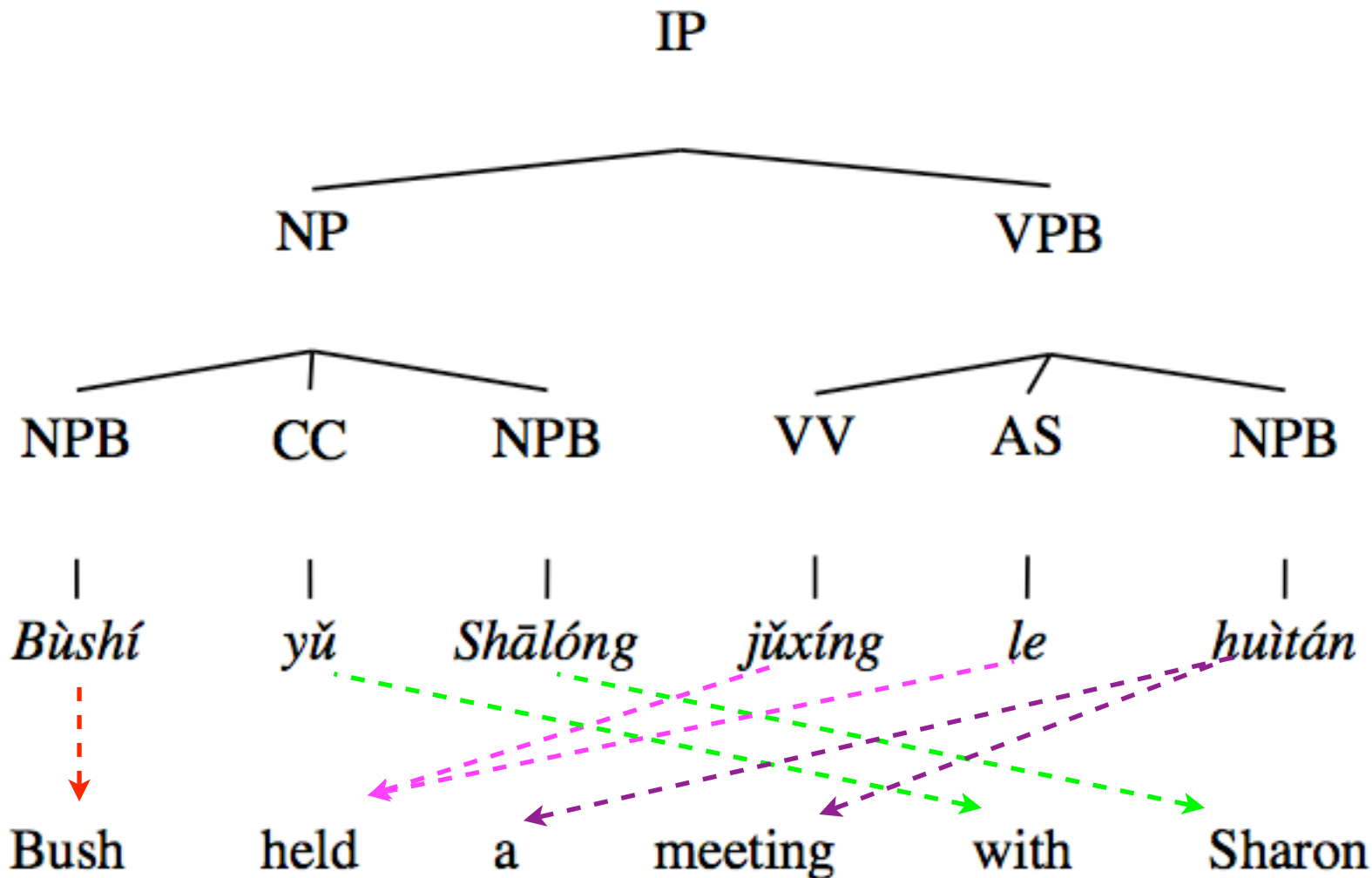
Penn
UNIVERSITY of PENNSYLVANIA

EMNLP 2008 talk, Honolulu, Hawaii, October 2008

This is NOT
the complete
presentation

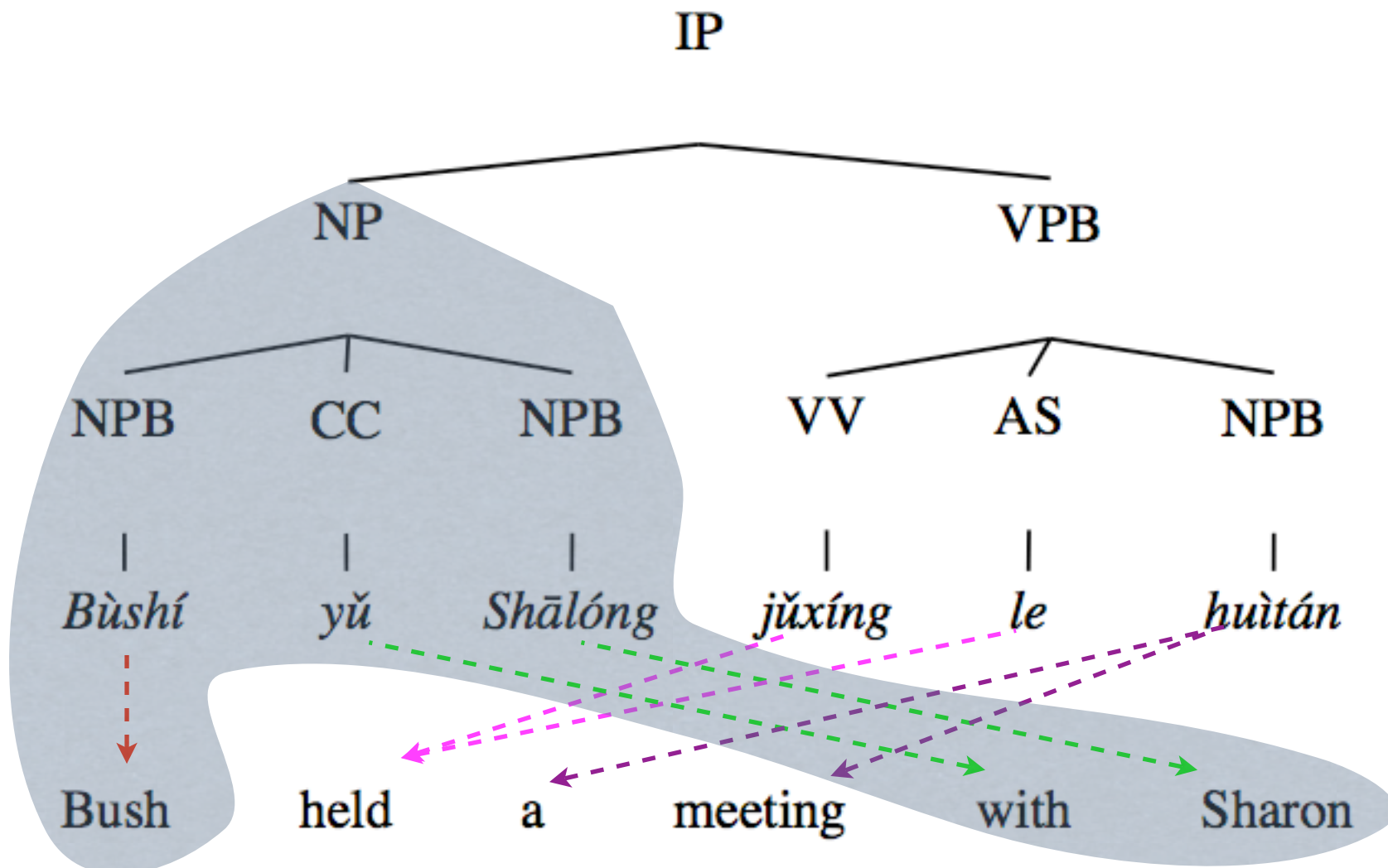
Where are the rules from?

- source parse tree, target sentence, and alignment
- intuition: **contiguous span**



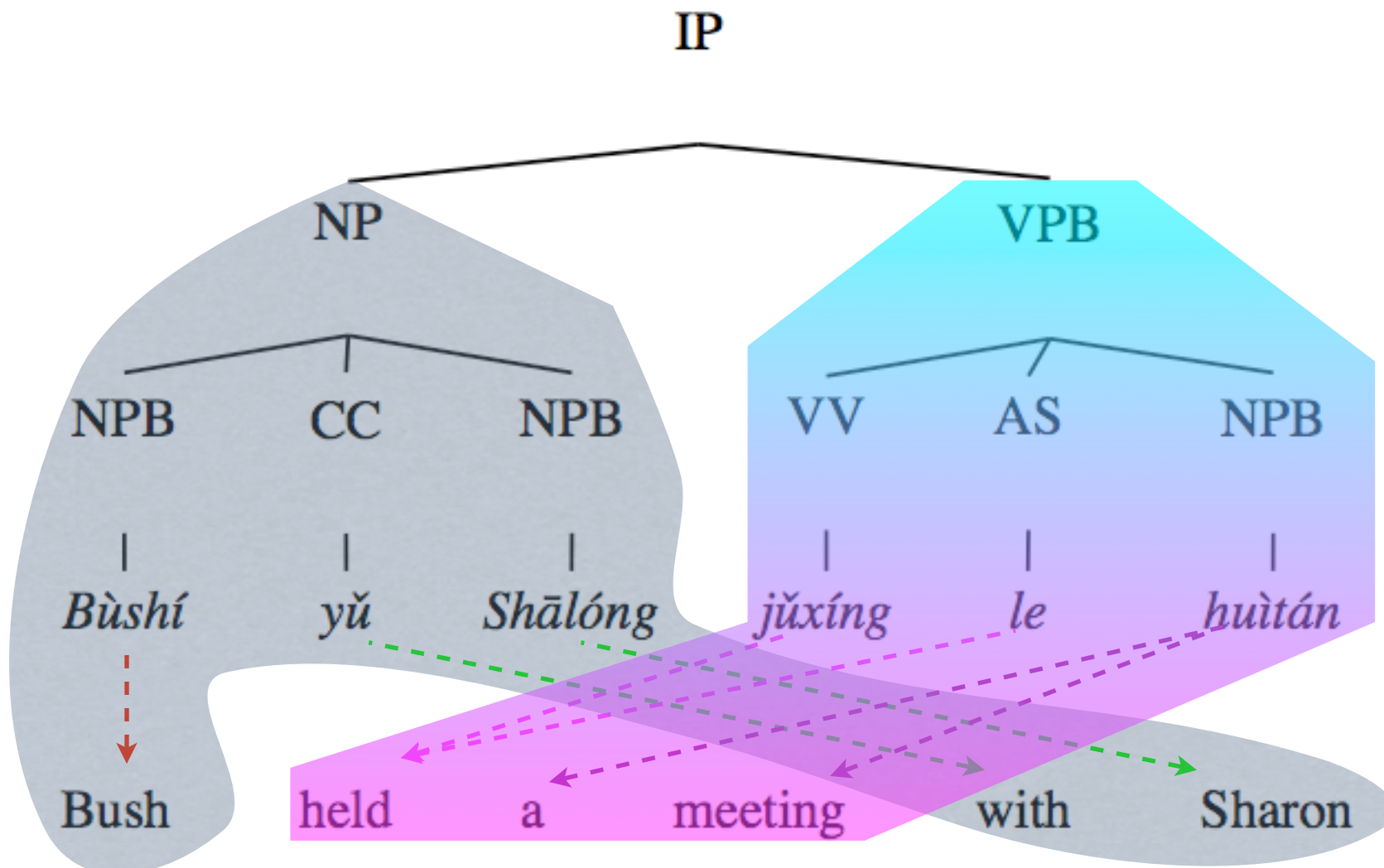
Where are the rules from?

- source parse tree, target sentence, and alignment
- intuition: **contiguous span**



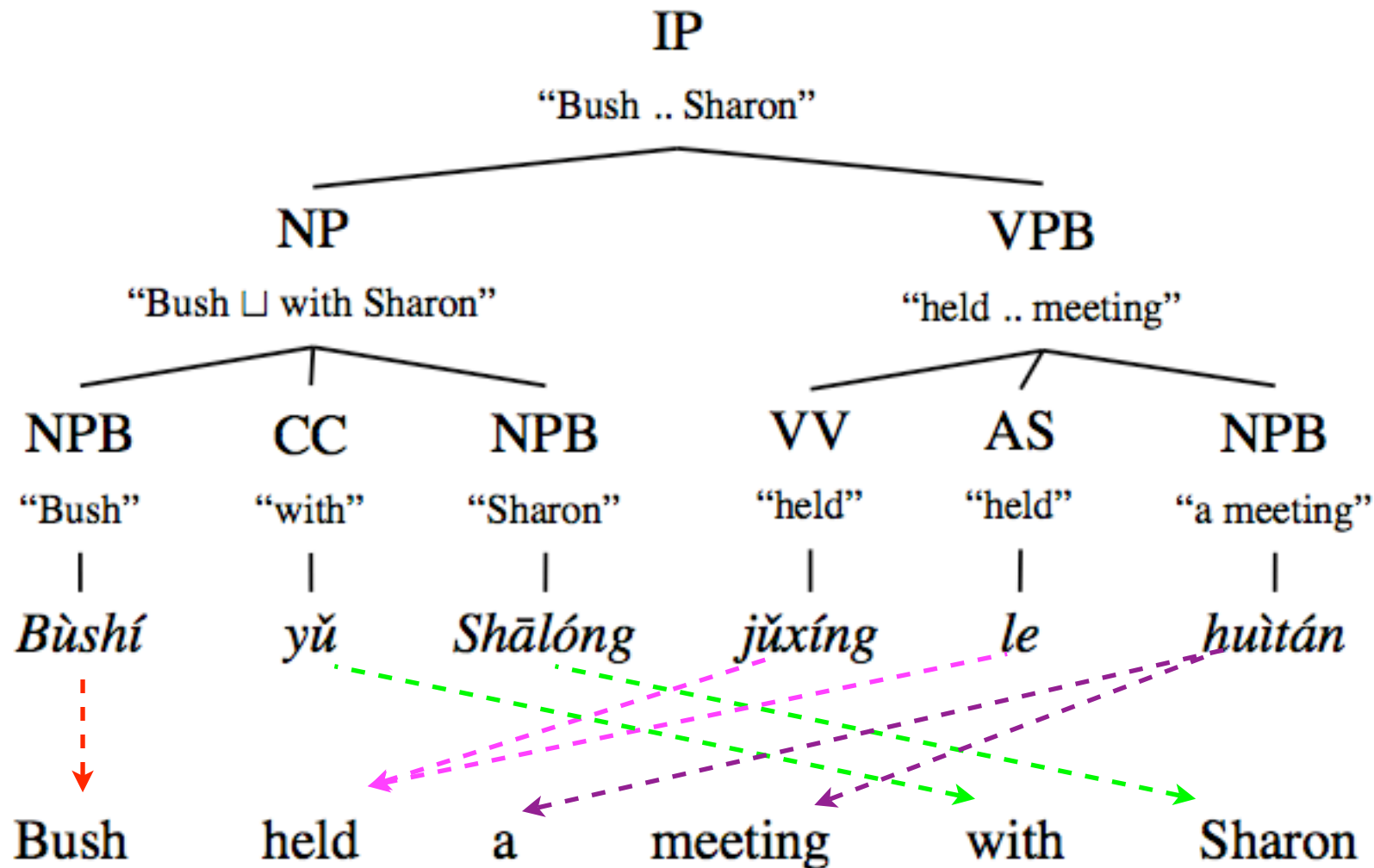
Where are the rules from?

- source parse tree, target sentence, and alignment
- intuition: **contiguous span**



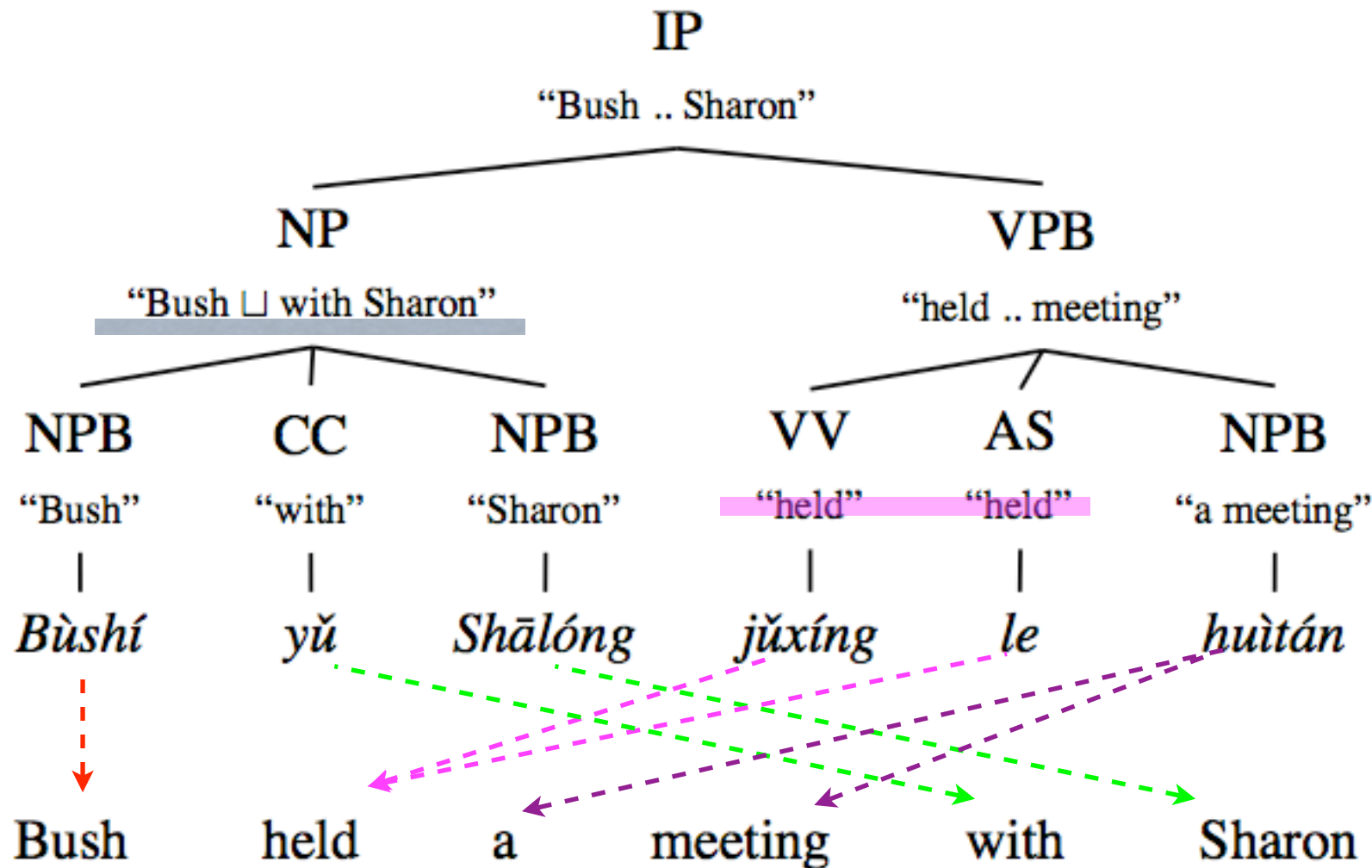
Where are the rules from?

- source parse tree, target sentence, and alignment
- compute target spans



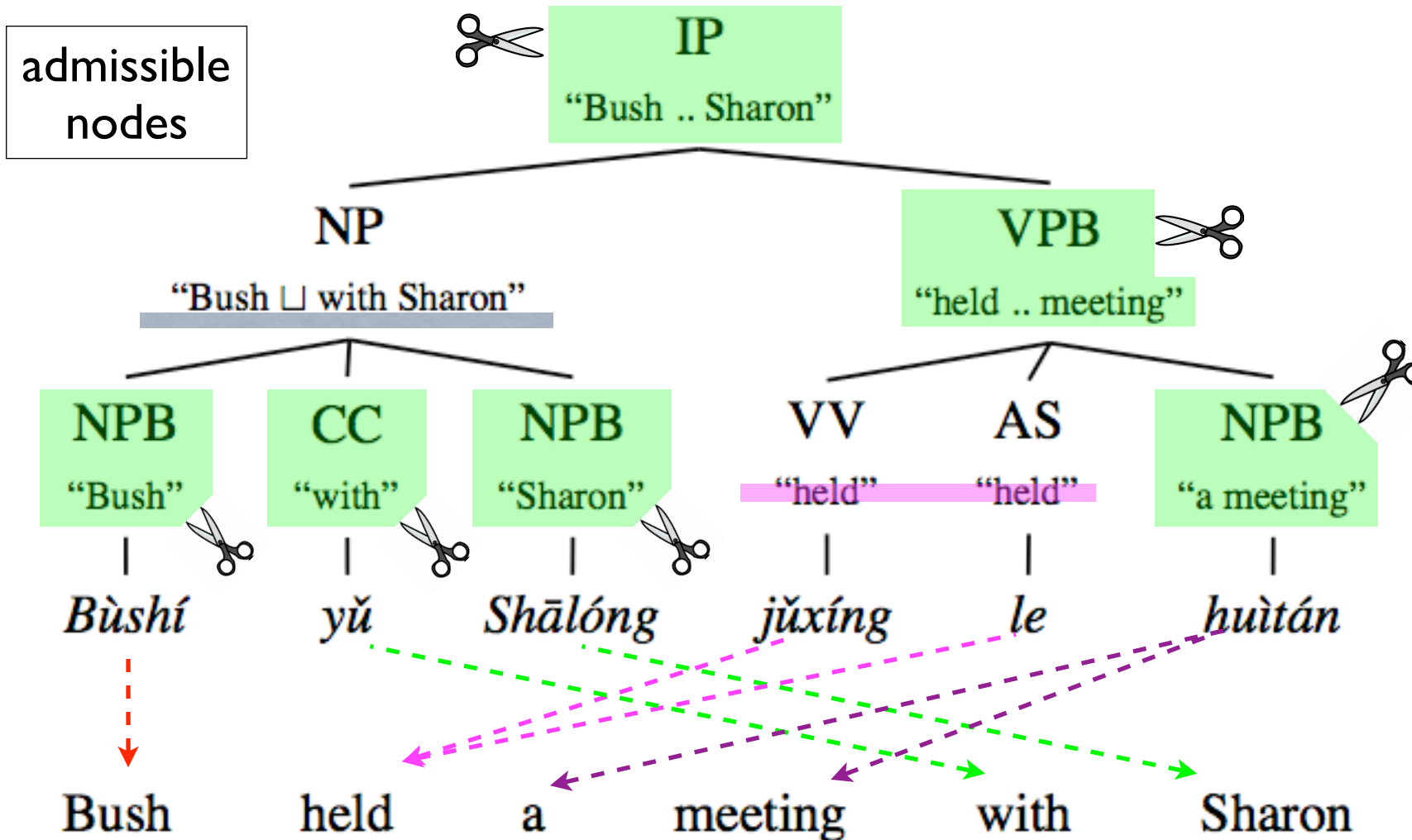
Where are the rules from?

- source parse tree, target sentence, and alignment
- well-formed fragment: contiguous and faithful target-span



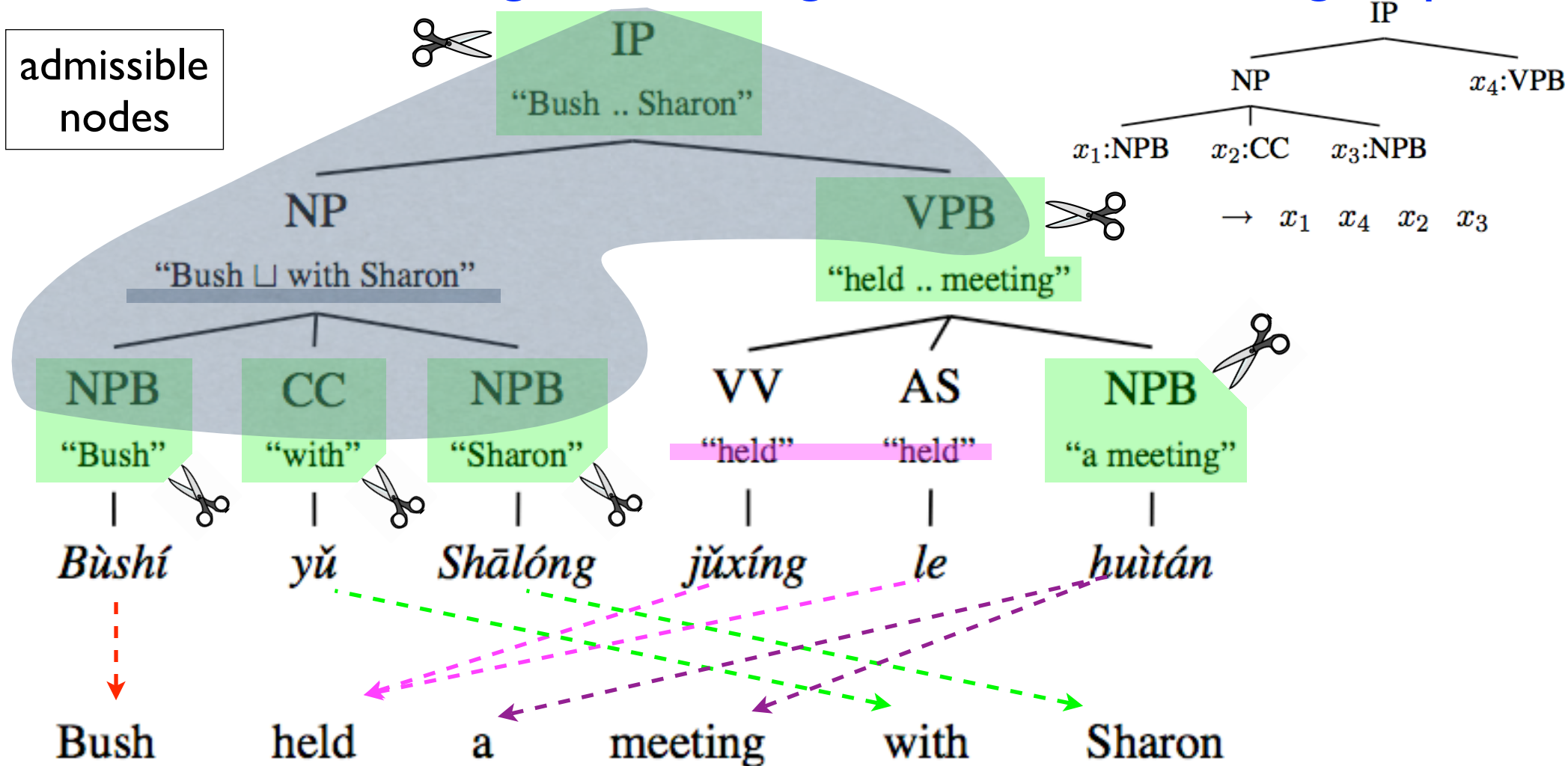
Where are the rules from?

- source parse tree, target sentence, and alignment
- well-formed fragment: contiguous and faithful target-span



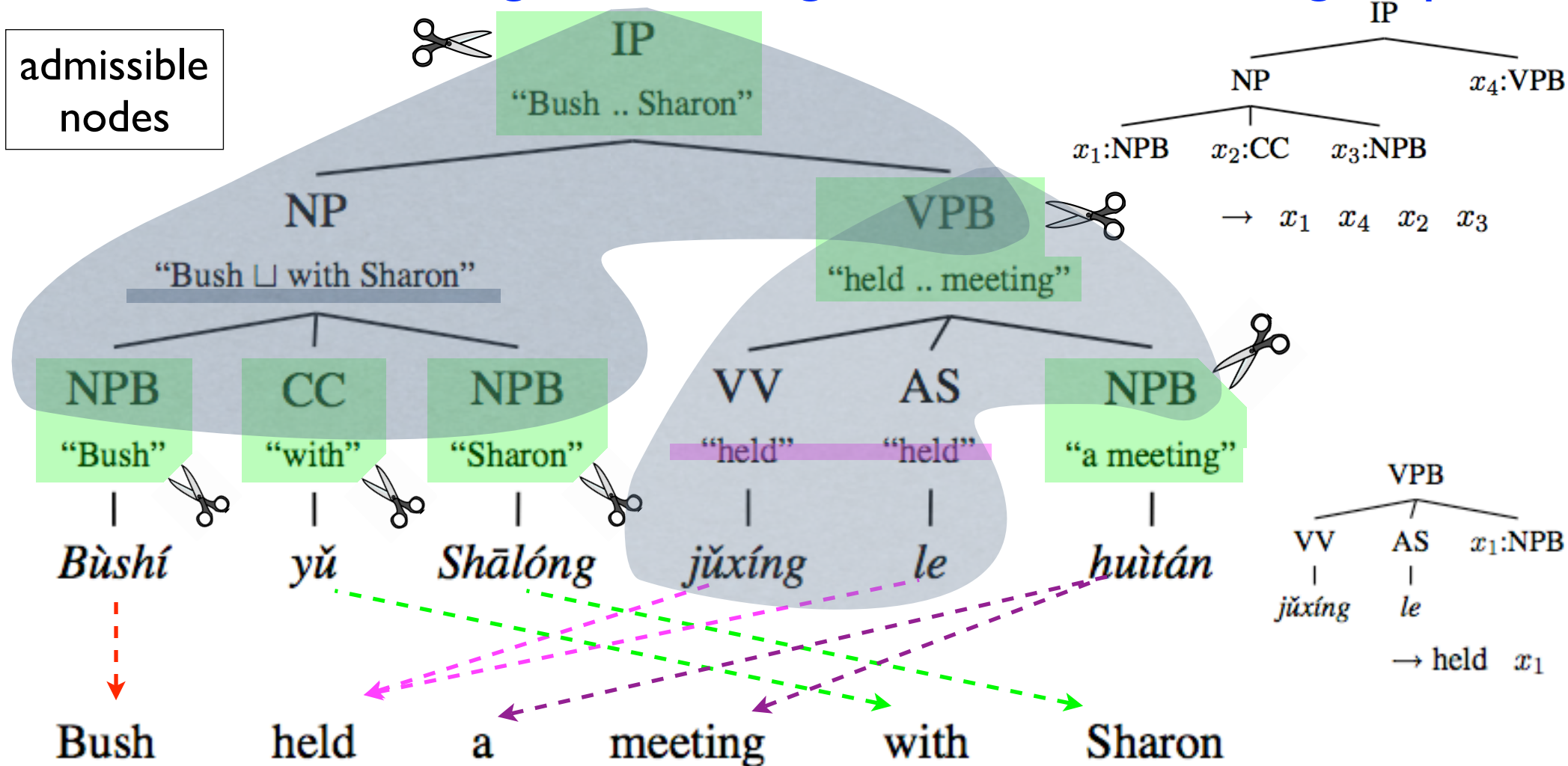
Where are the rules from?

- source parse tree, target sentence, and alignment
- well-formed fragment: contiguous and faithful target-span



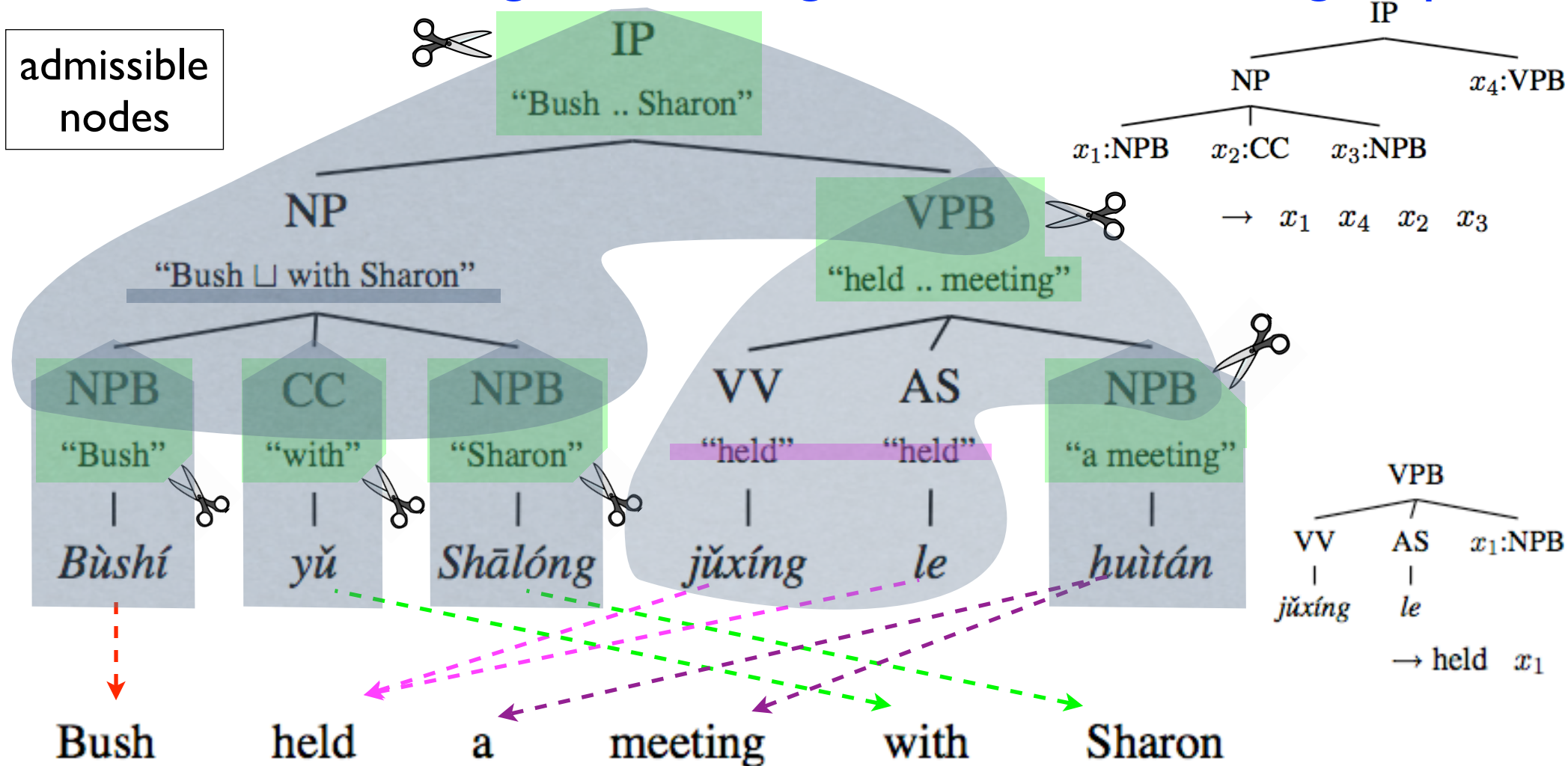
Where are the rules from?

- source parse tree, target sentence, and alignment
- well-formed fragment: contiguous and faithful target-span

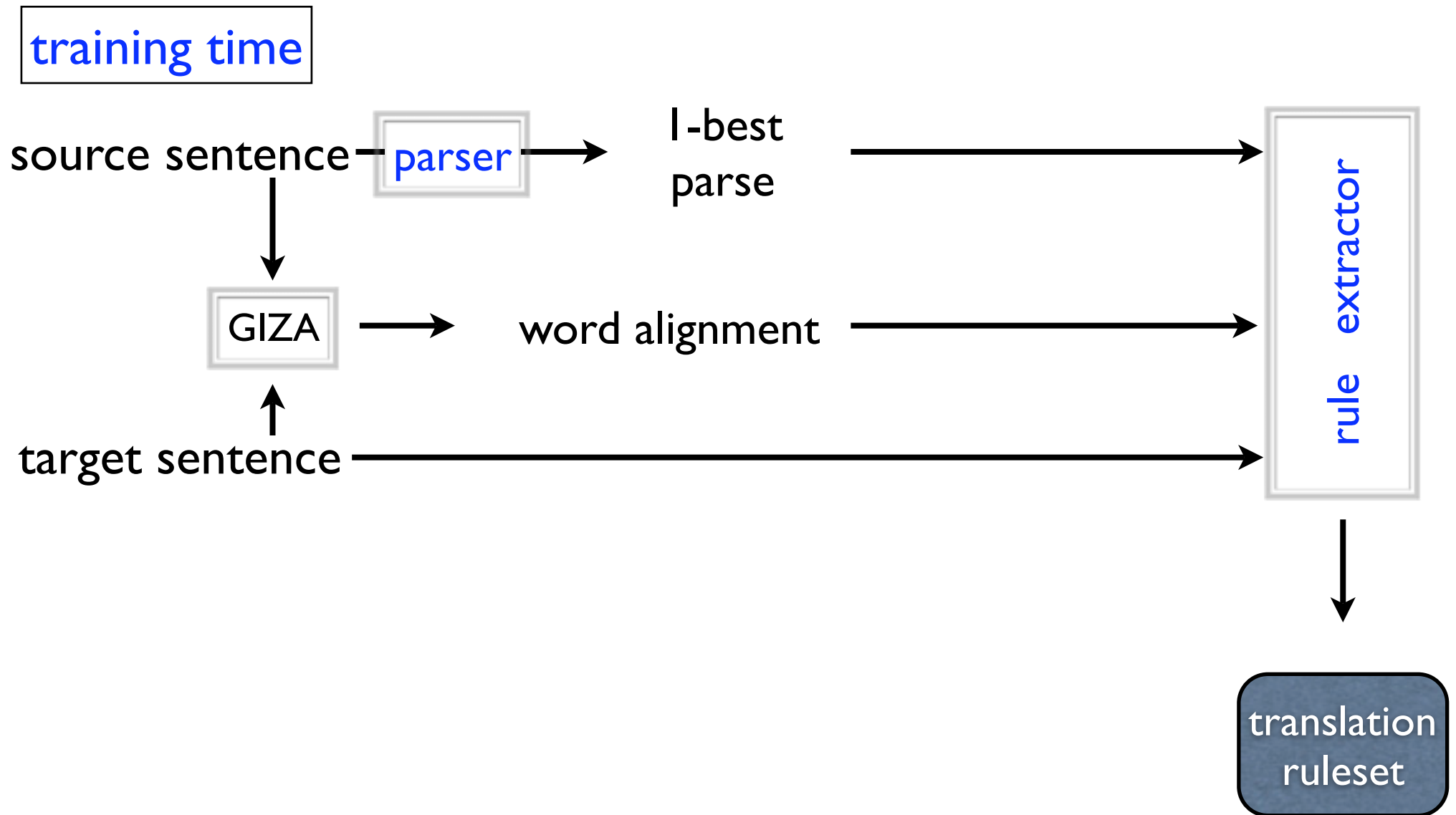


Where are the rules from?

- source parse tree, target sentence, and alignment
- well-formed fragment: contiguous and faithful target-span

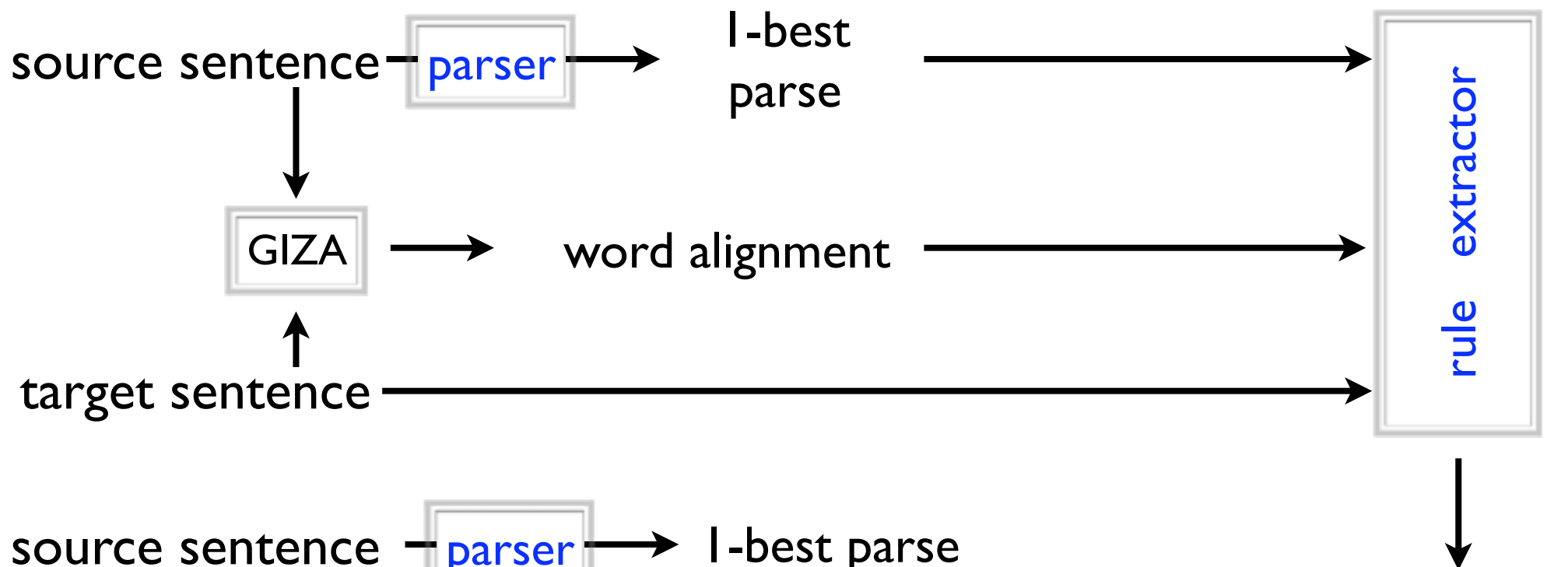


The Baseline Pipeline

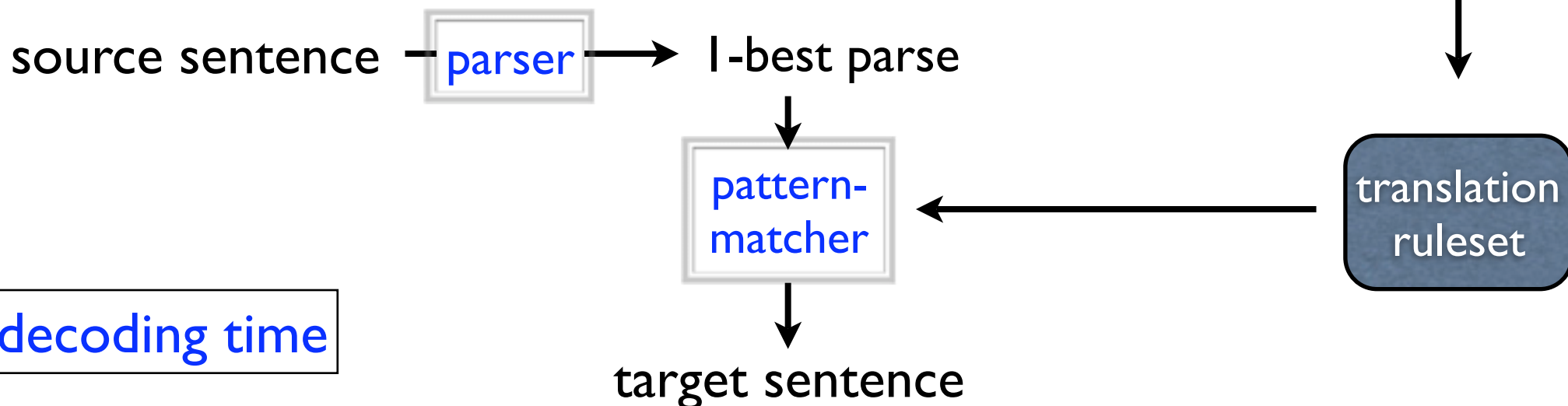


The Baseline Pipeline

training time

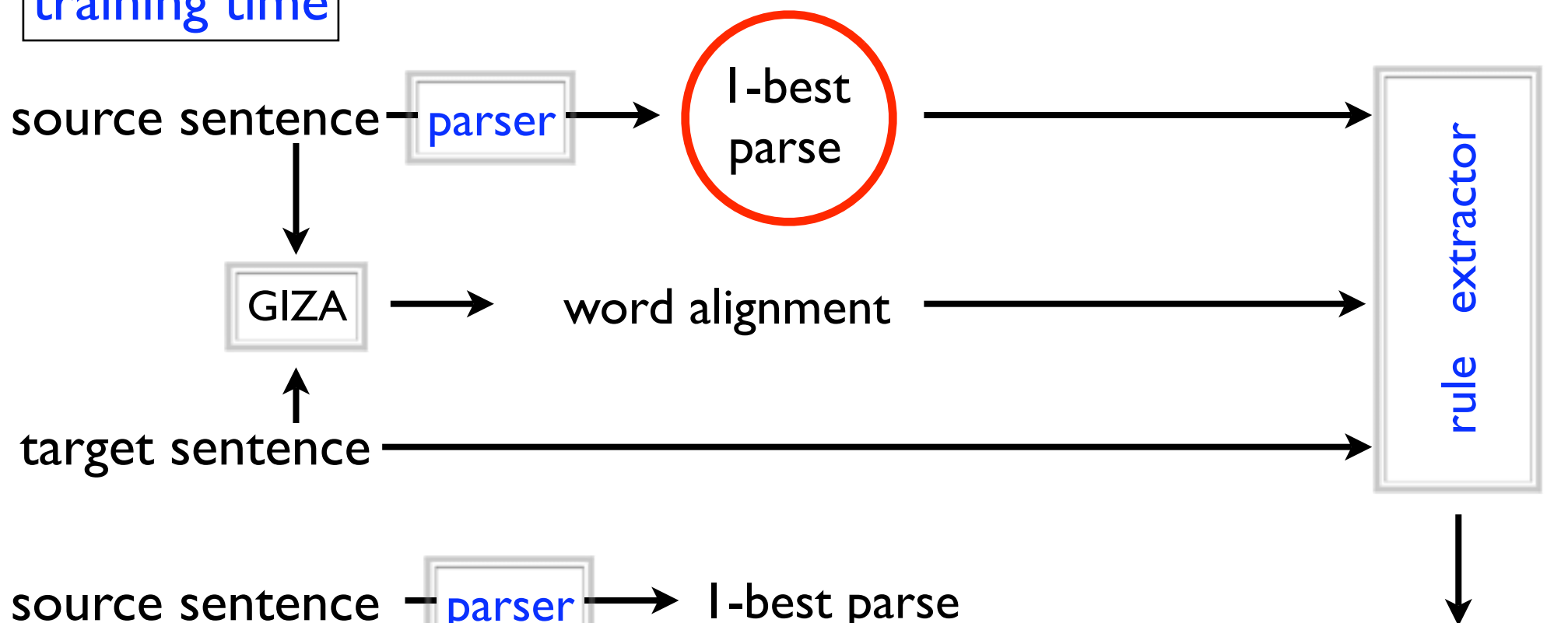


decoding time

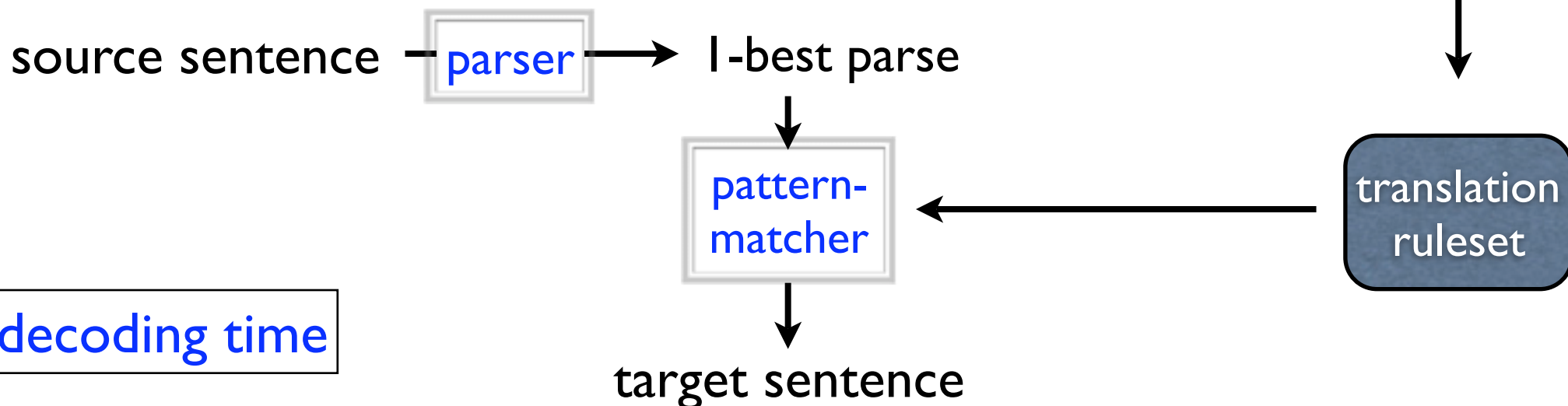


The Baseline Pipeline

training time



decoding time

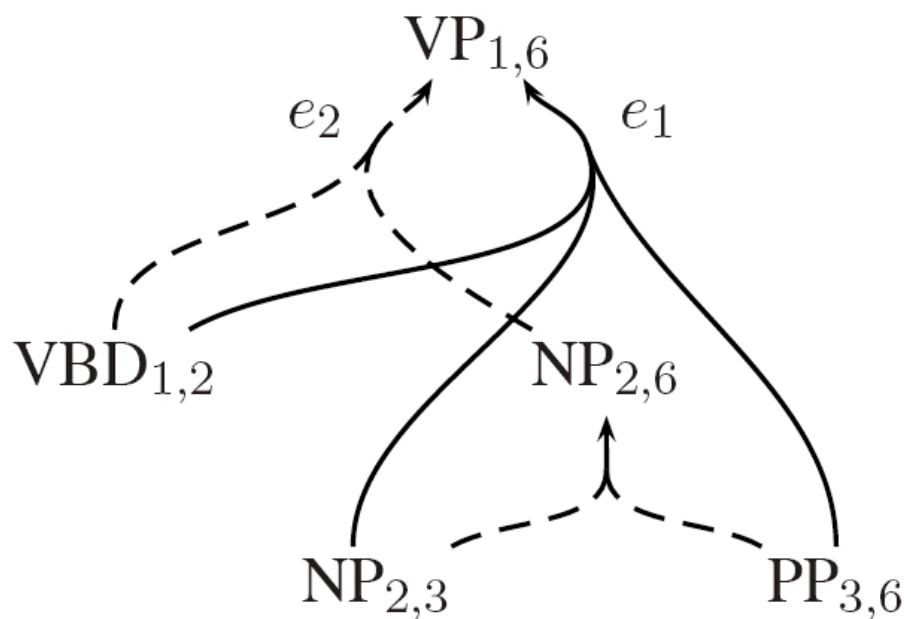


Outline

- Background: Tree-based Translation and Rule Extraction
- Forest-based Rule Extraction
 - Background: Parse Forest
 - Forest-based Extraction
 - Inside-Outside Forest Pruning
 - Fractional Rule Counts
- Related Work
- Experiments

Packed Forest

- a compact representation of many parses
- by sharing common sub-derivations
- polynomial-space encoding of exponentially large set

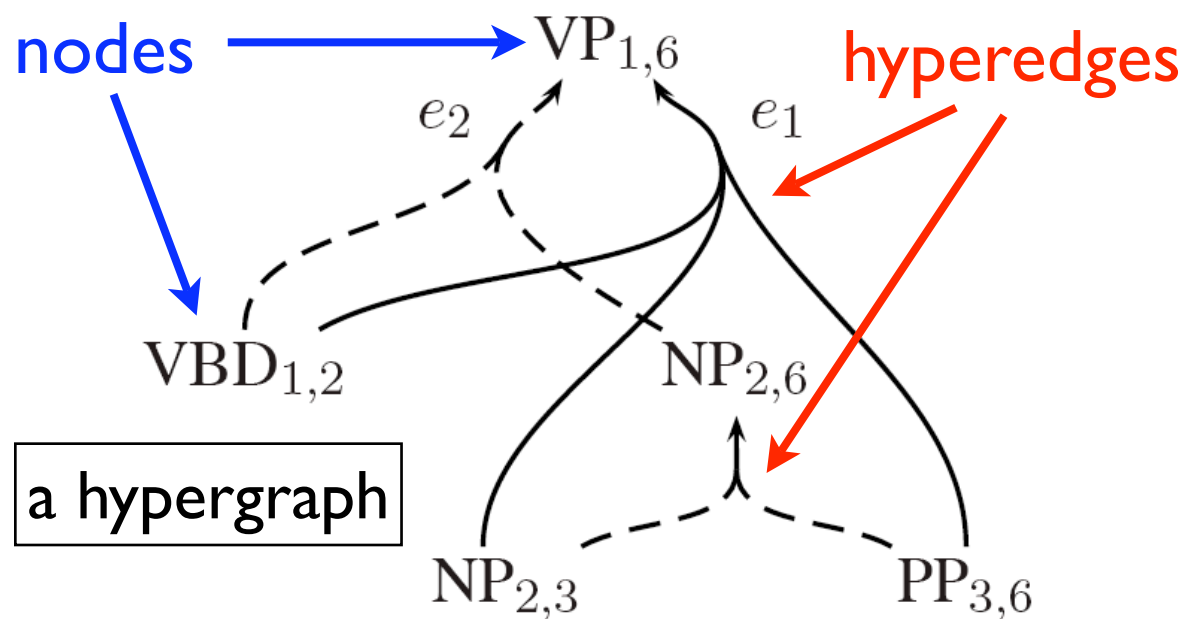


$$e_1 \quad \frac{VBD_{1,2} \quad NP_{2,3} \quad PP_{3,6}}{VP_{1,6}}$$

0 I 1 saw 2 him 3 with 4 a 5 mirror 6

Packed Forest

- a compact representation of many parses
- by sharing common sub-derivations
- polynomial-space encoding of exponentially large set

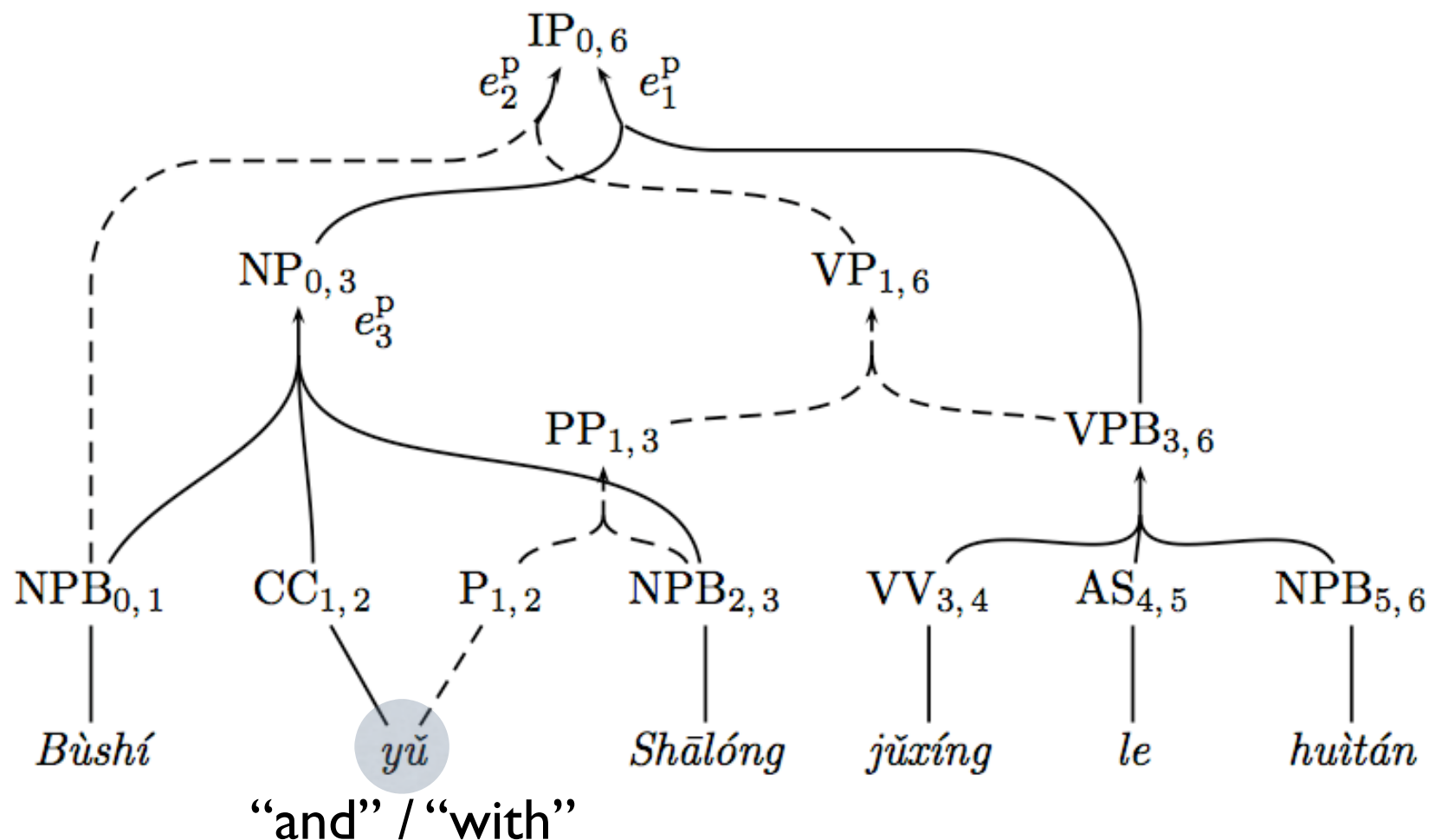


$$e_1 \frac{VBD_{1,2} \quad NP_{2,3} \quad PP_{3,6}}{VP_{1,6}}$$

0 I 1 saw 2 him 3 with 4 a 5 mirror 6

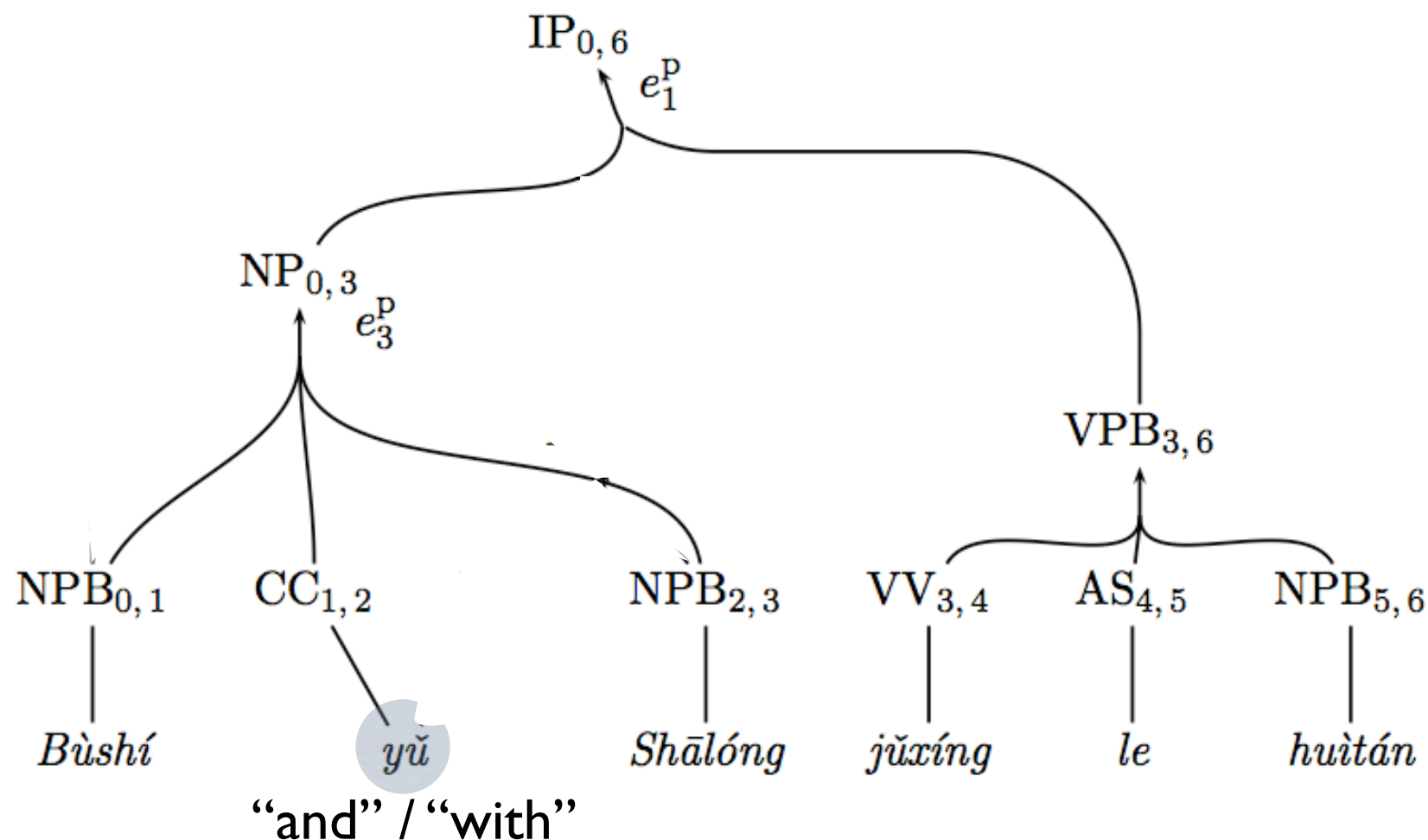
Chinese Forest

- parse the input into a forest instead of I-best tree
- Chinese *yu* can be either a CC (“and”) or P (“with”)



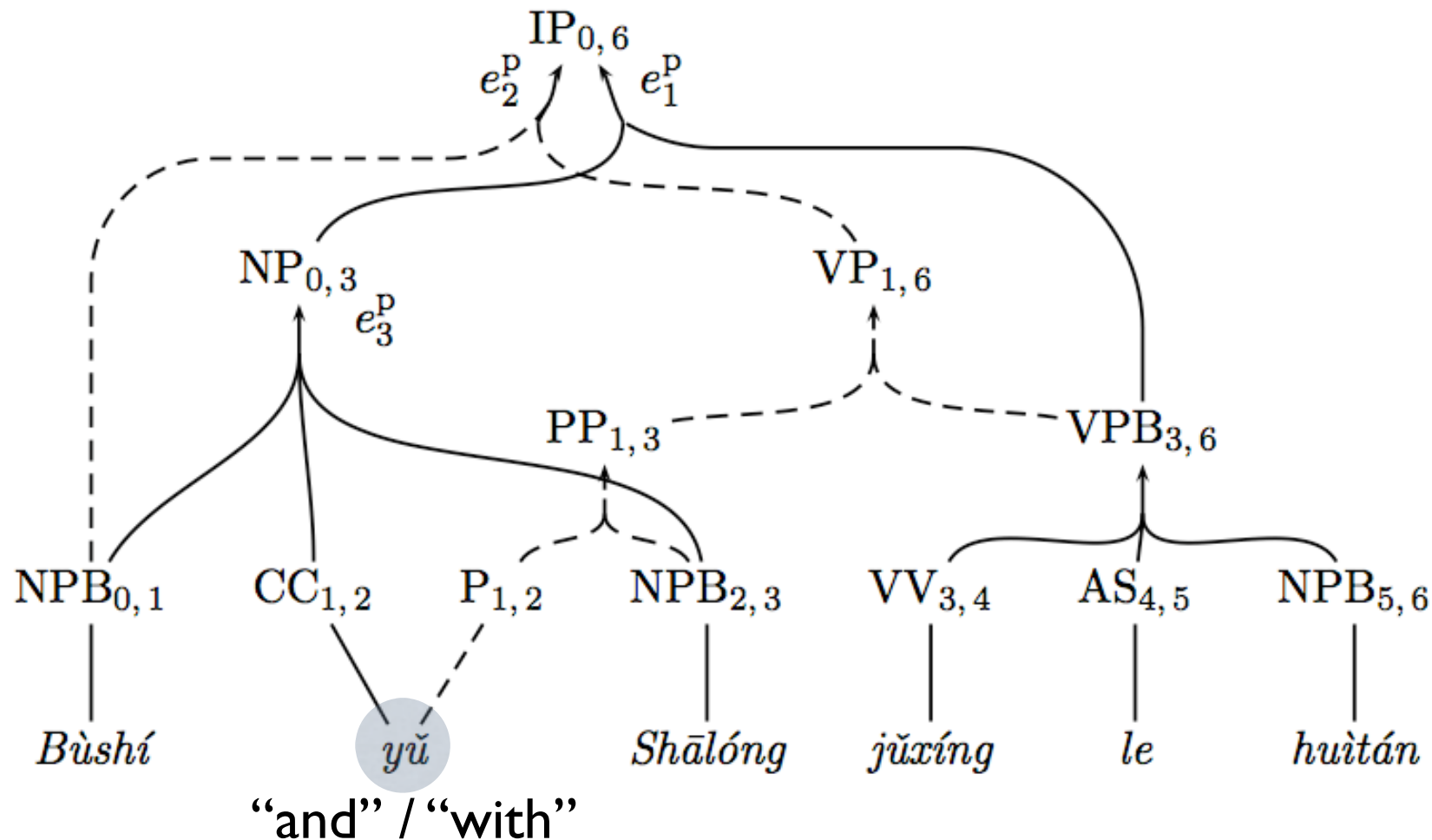
Chinese Forest

- parse the input into a forest instead of I-best tree
- Chinese *yu* can be either a CC (“and”) or P (“with”)



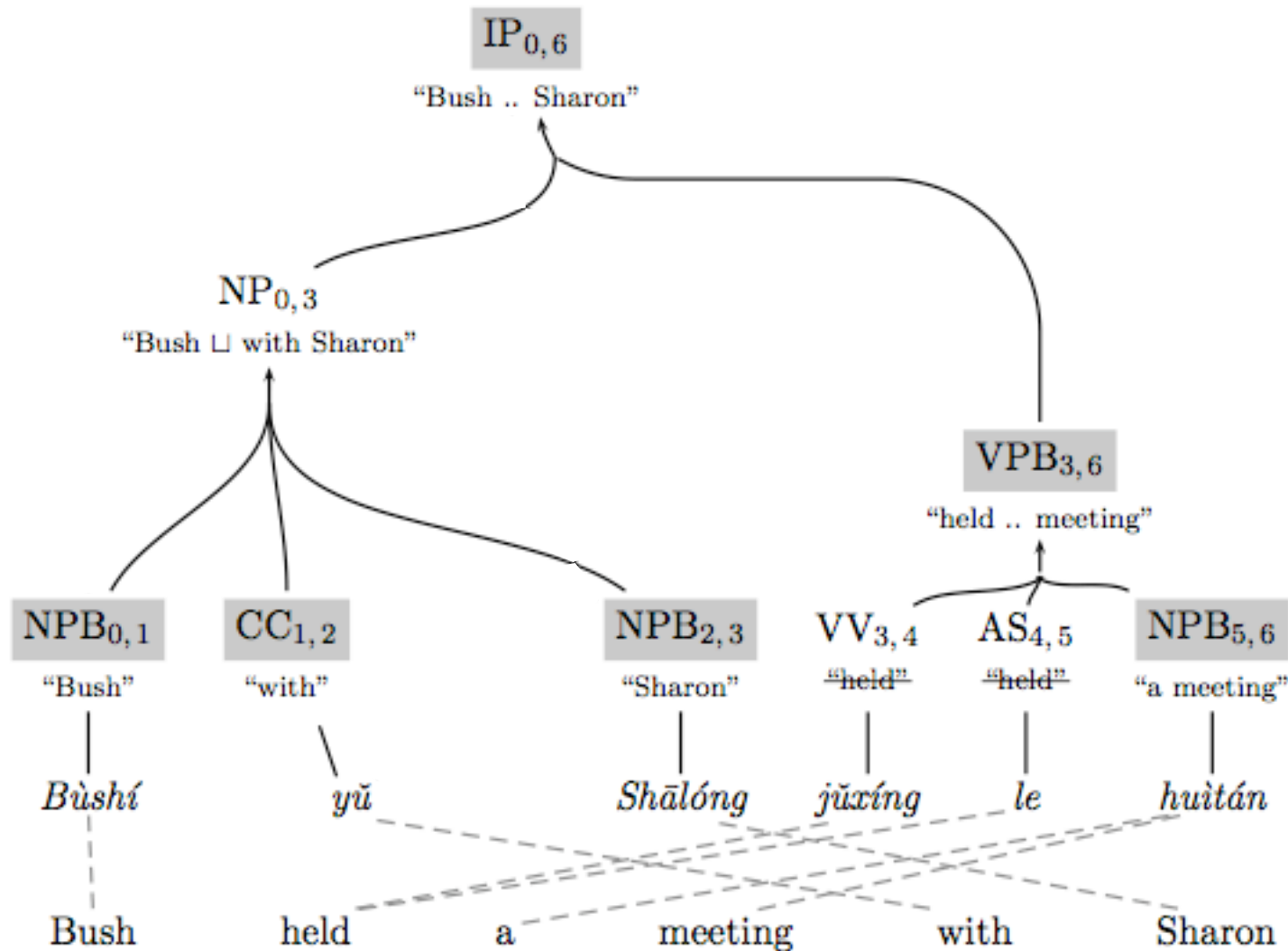
Chinese Forest

- parse the input into a forest instead of I-best tree
- Chinese *yu* can be either a CC (“and”) or P (“with”)



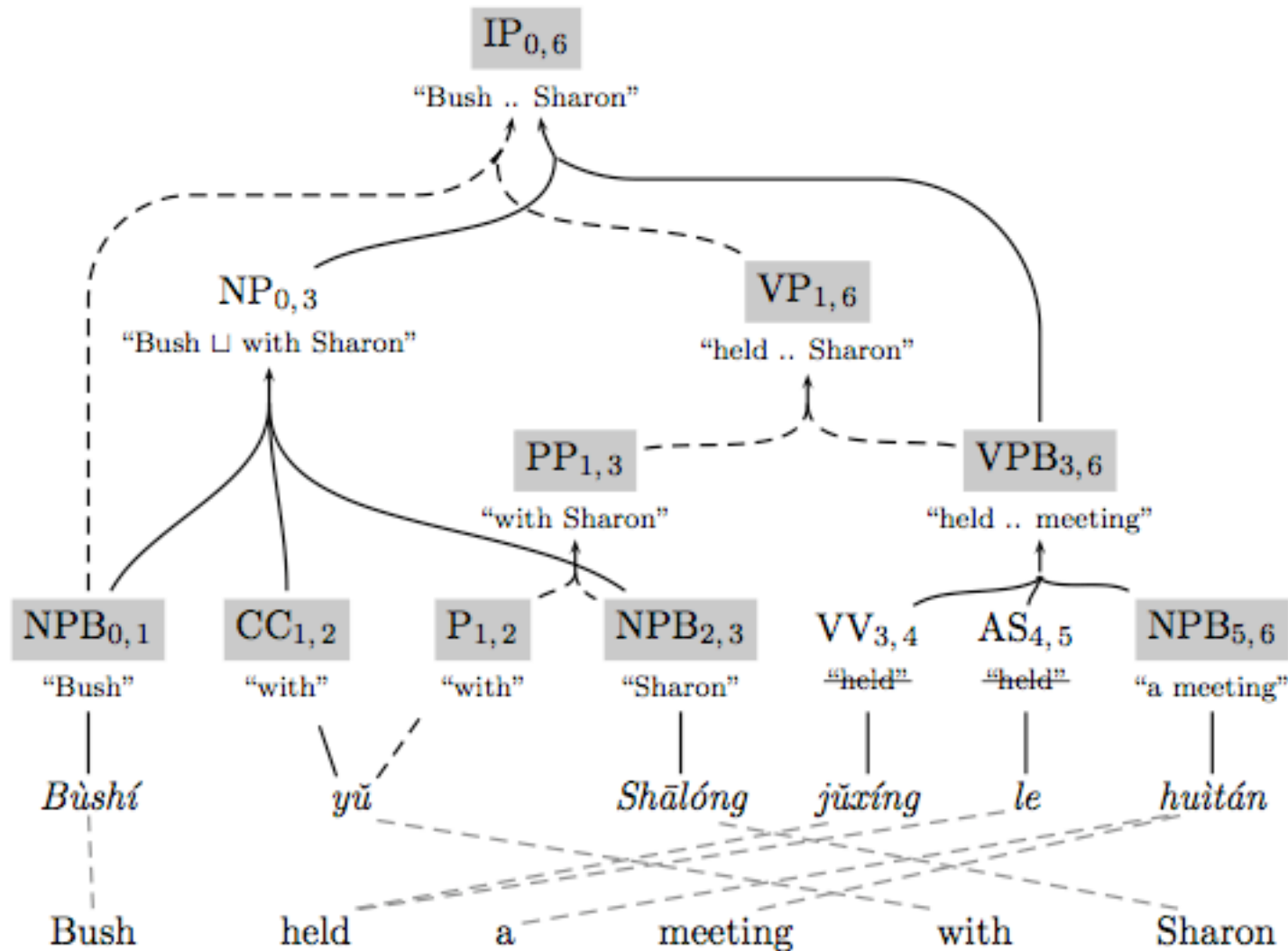
Forest-based Rule Extraction

- same at “where to cut”; different at “how to cut”



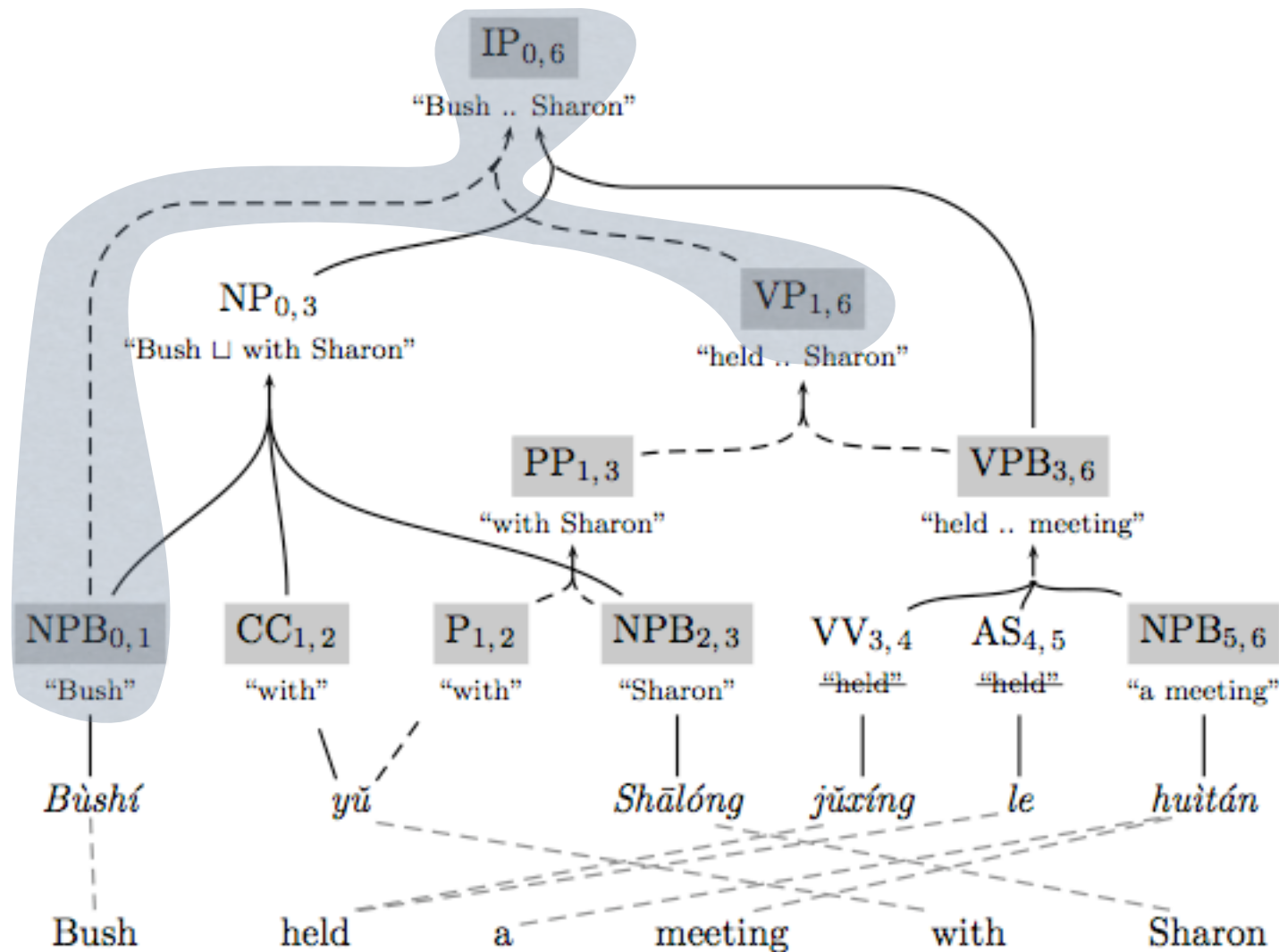
Forest-based Rule Extraction

- same at “where to cut”; different at “**how** to cut”



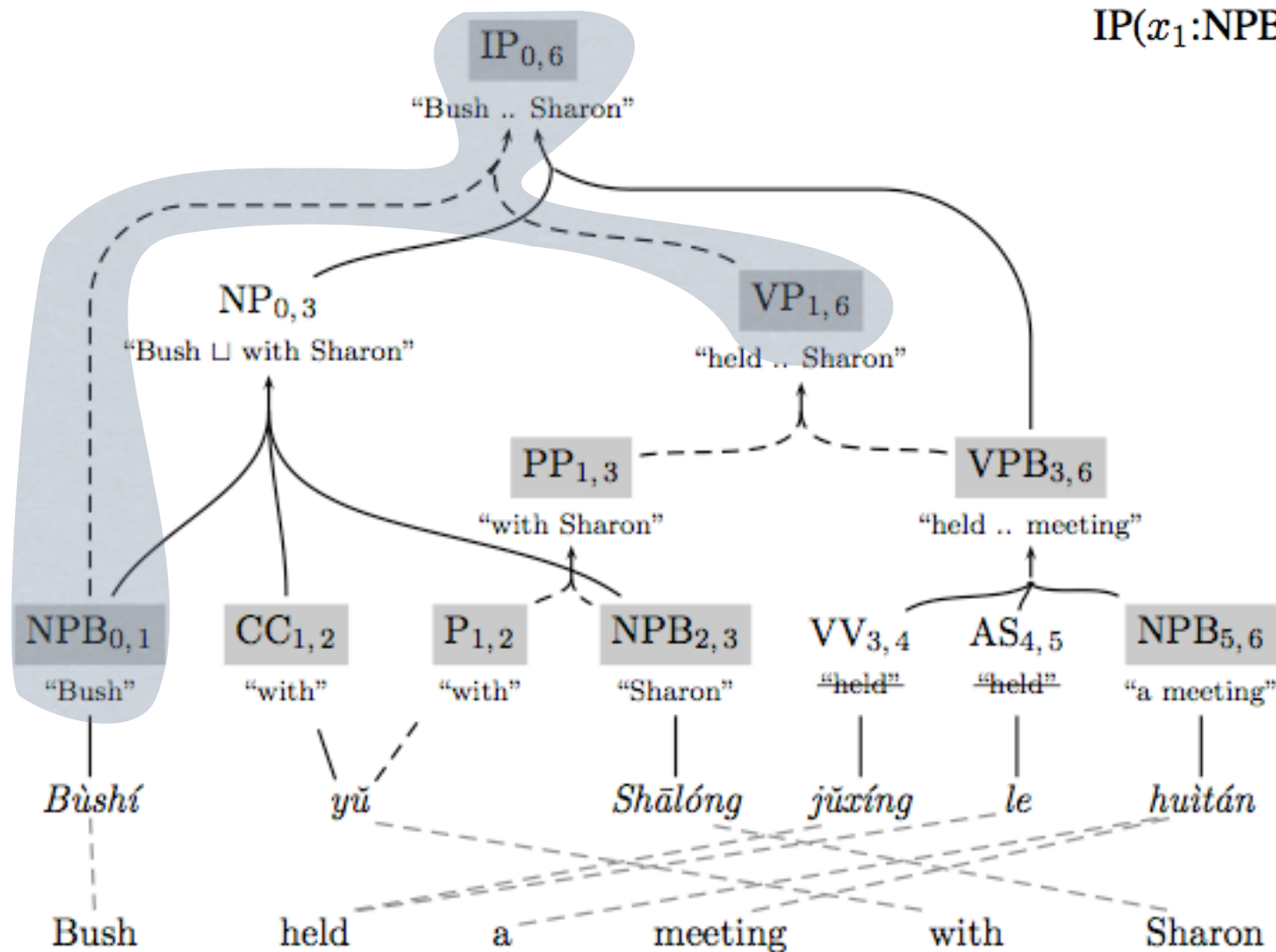
Forest-based Rule Extraction

- same at “where to cut”; different at “**how** to cut”



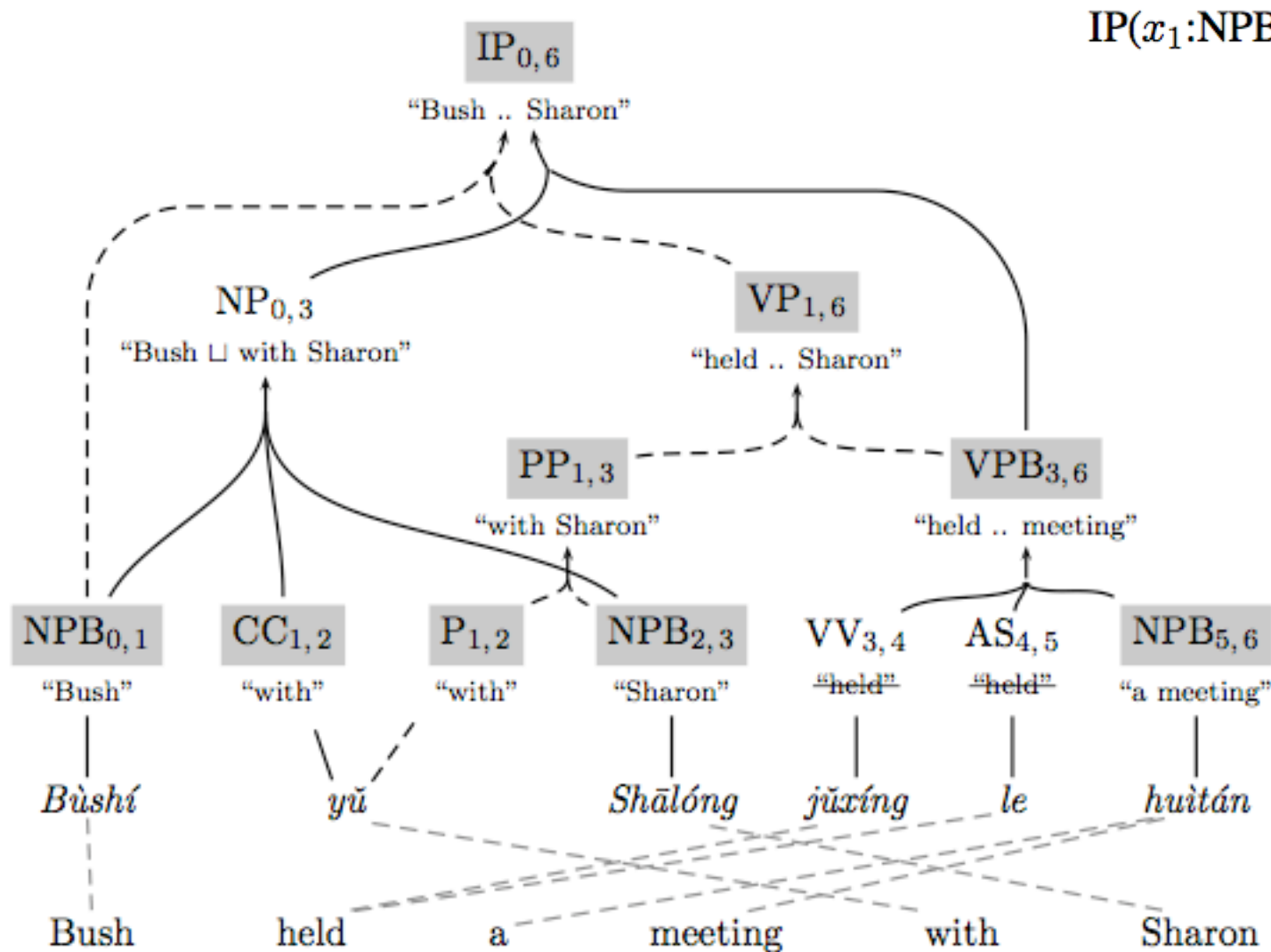
Forest-based Rule Extraction

- same at “where to cut”; different at “**how** to cut”



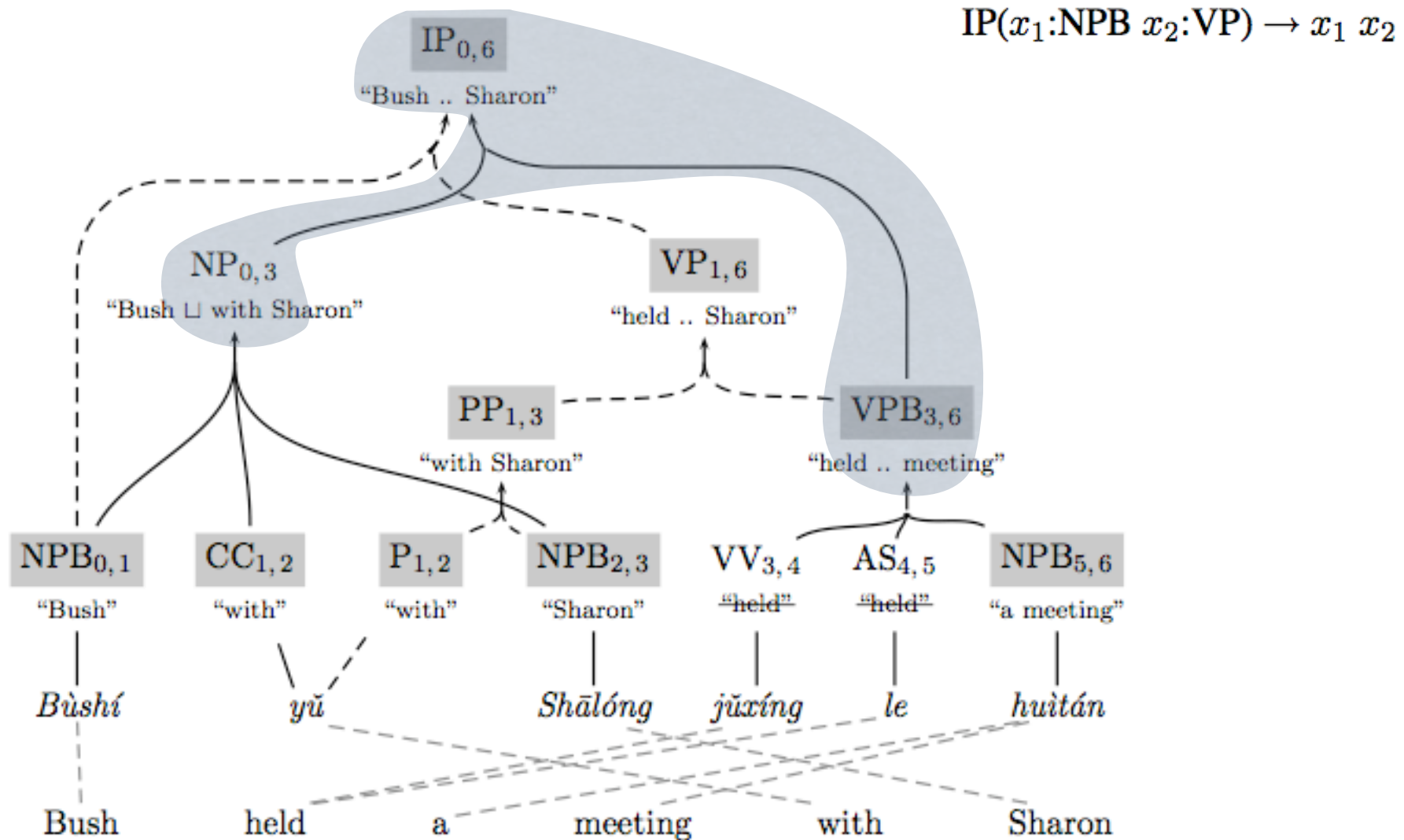
Forest-based Rule Extraction

- same at “where to cut”; different at “how to cut”



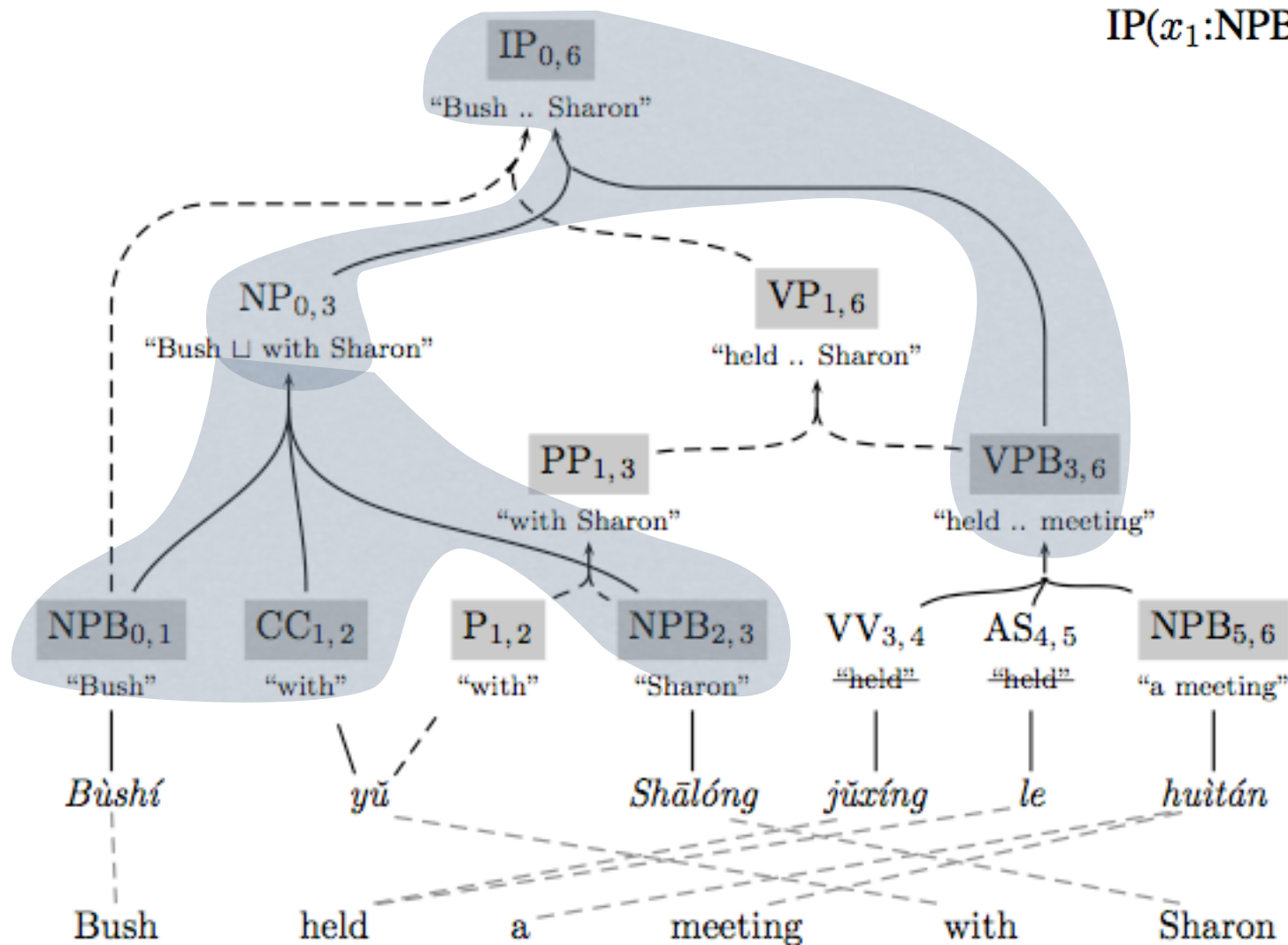
Forest-based Rule Extraction

- same at “where to cut”; different at “how to cut”



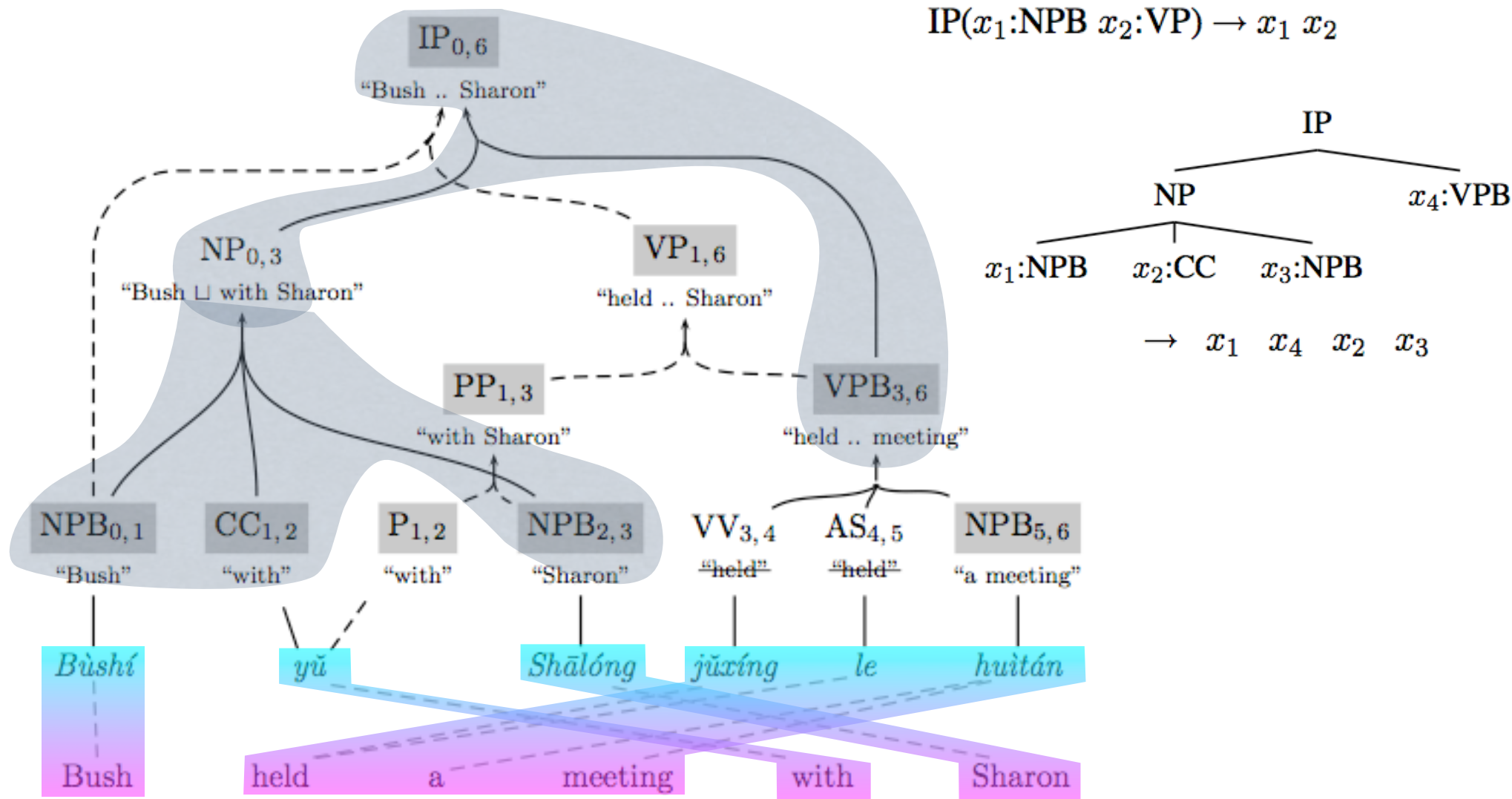
Forest-based Rule Extraction

- same at “where to cut”; different at “how to cut”



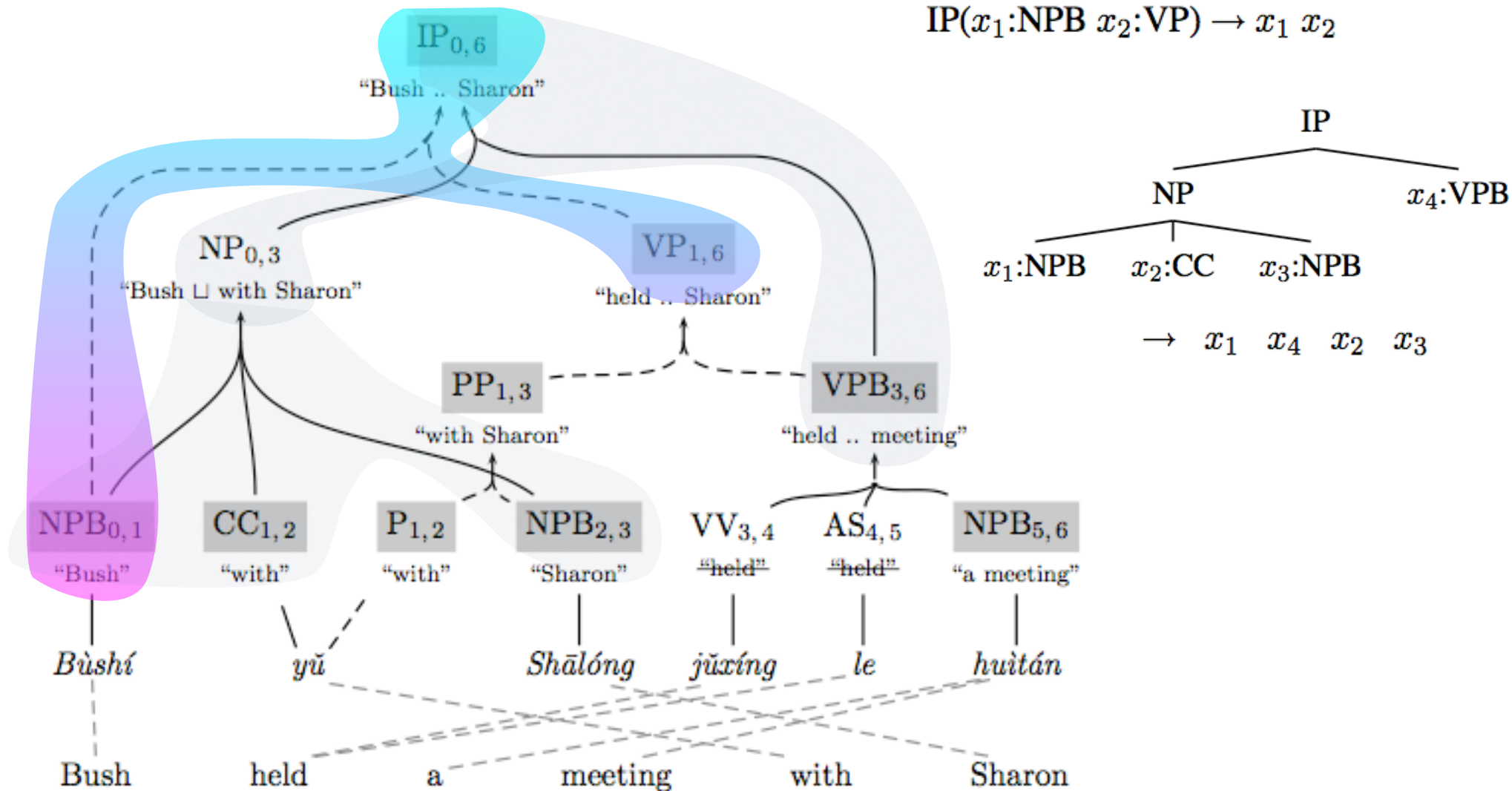
Forest-based Rule Extraction

- same at “where to cut”; different at “how to cut”



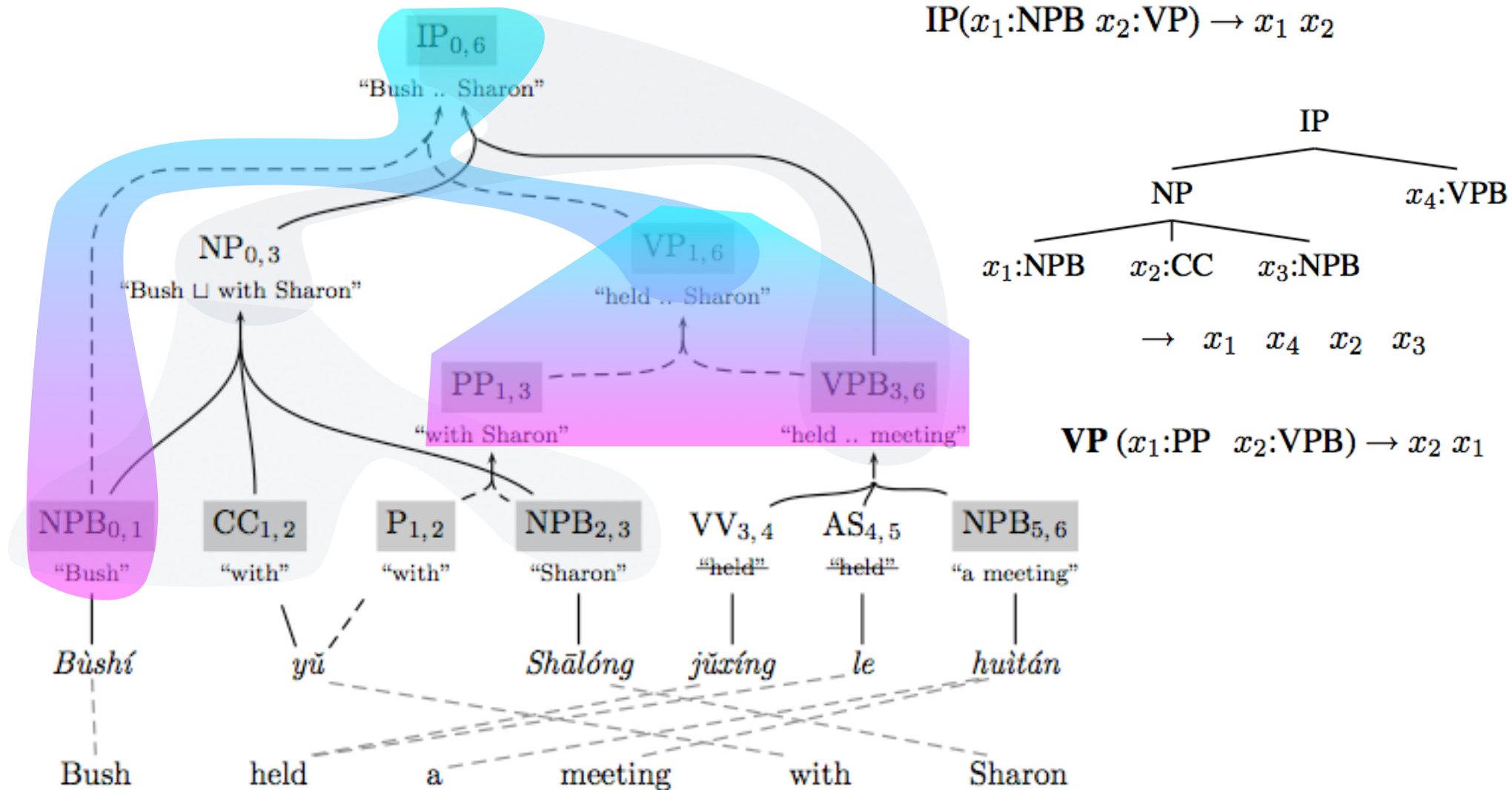
Forest-based Rule Extraction

- forest can extract smaller chunks of rules



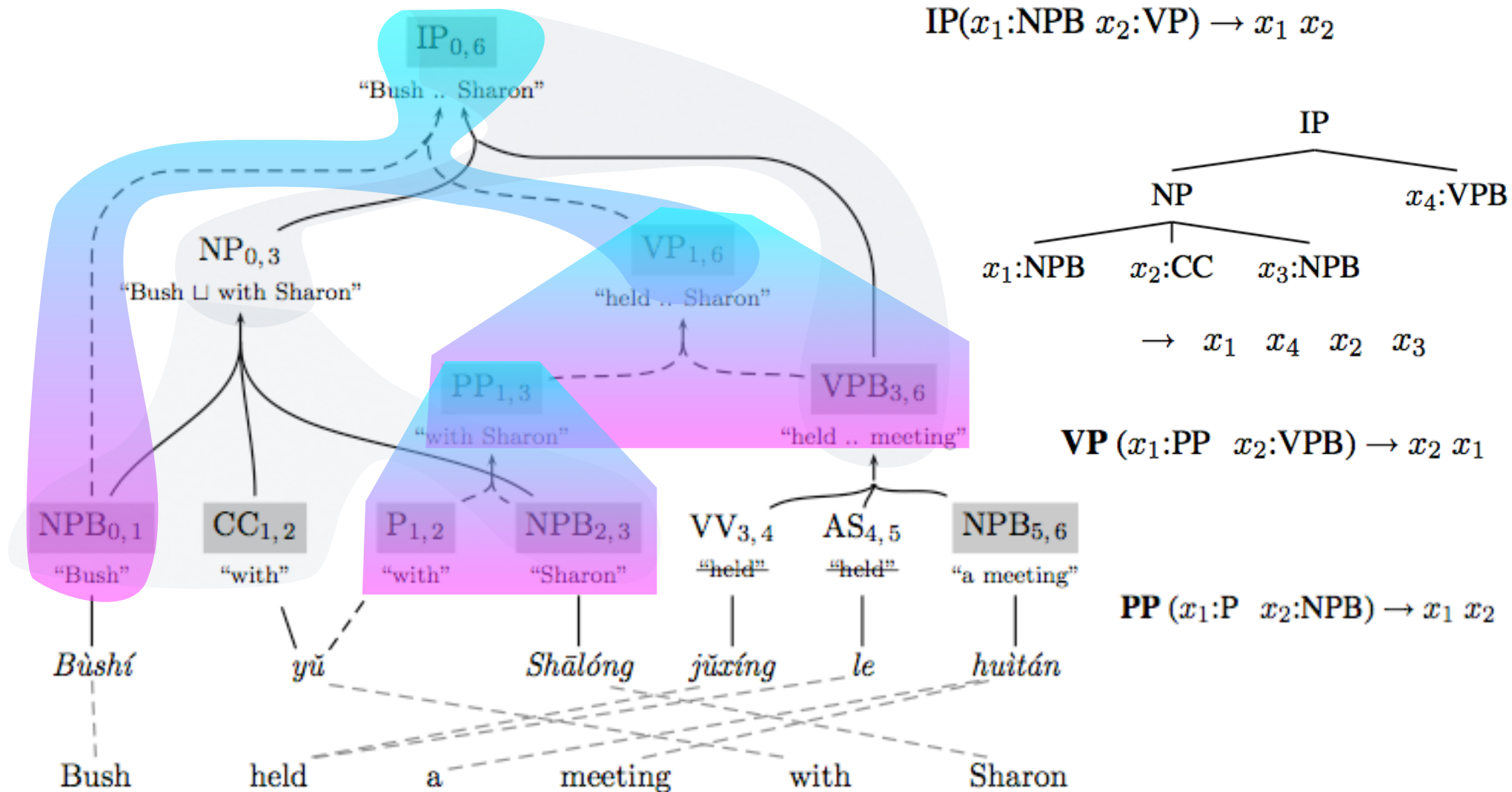
Forest-based Rule Extraction

- forest can extract smaller chunks of rules

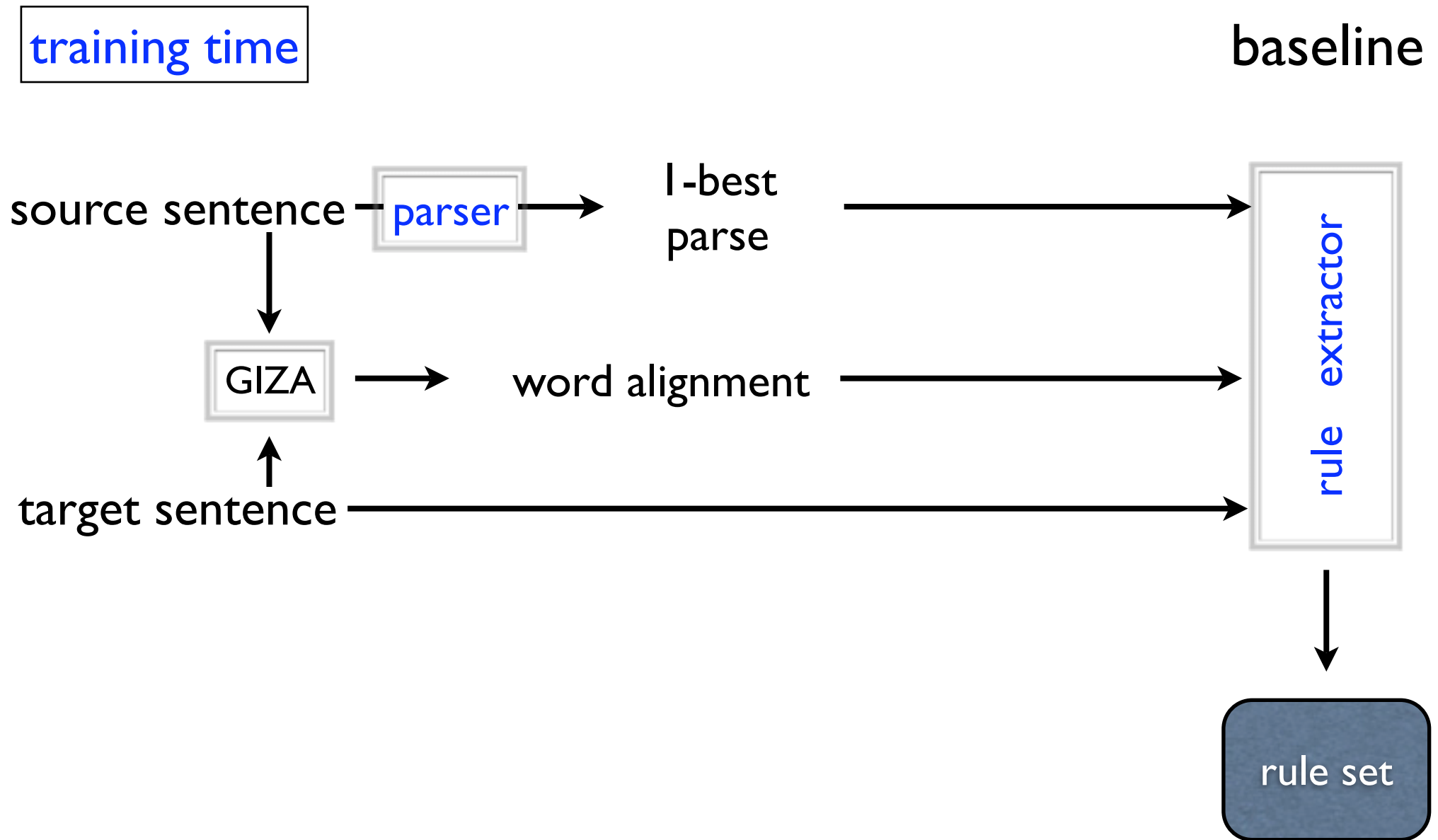


Forest-based Rule Extraction

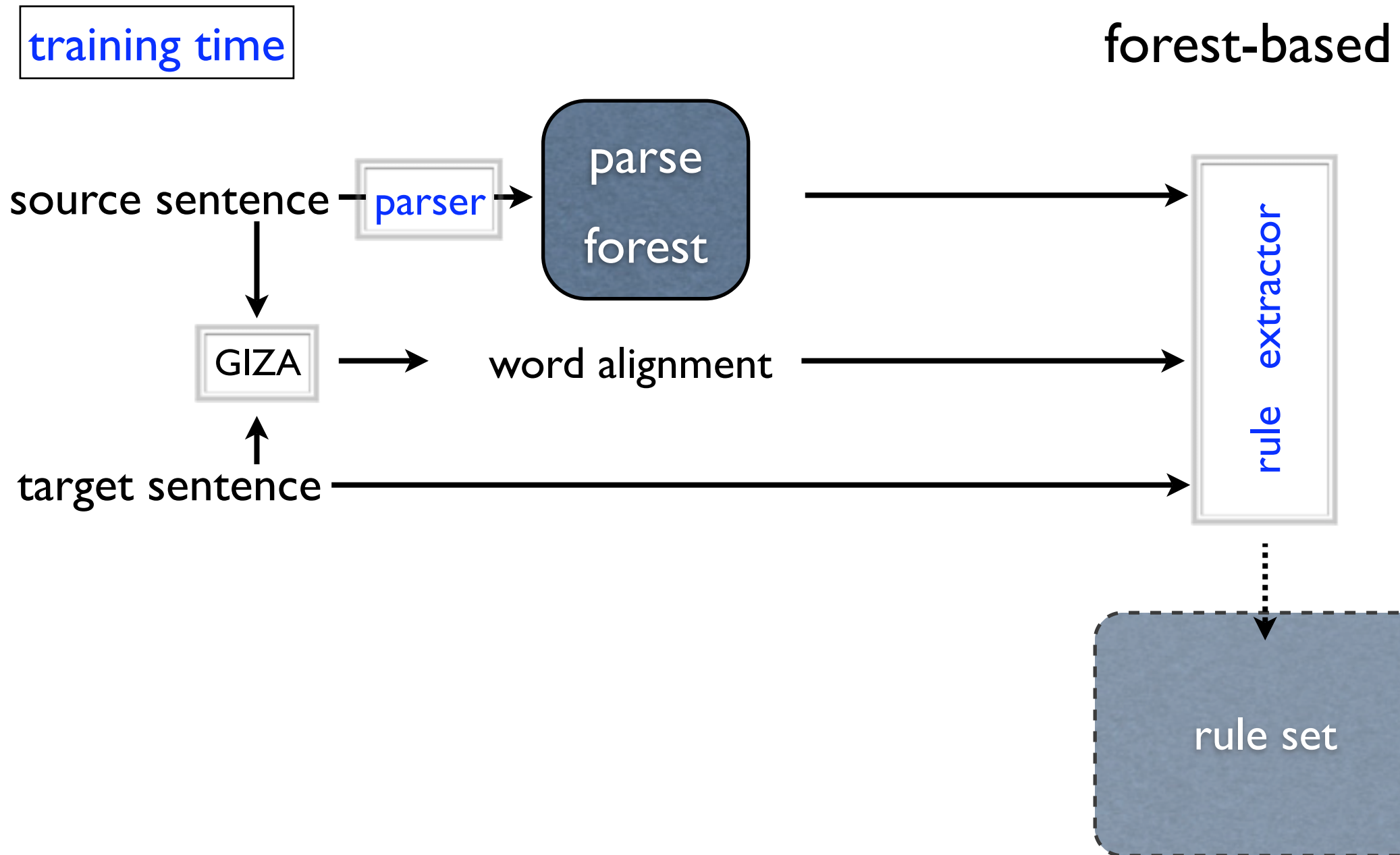
- forest can extract smaller chunks of rules



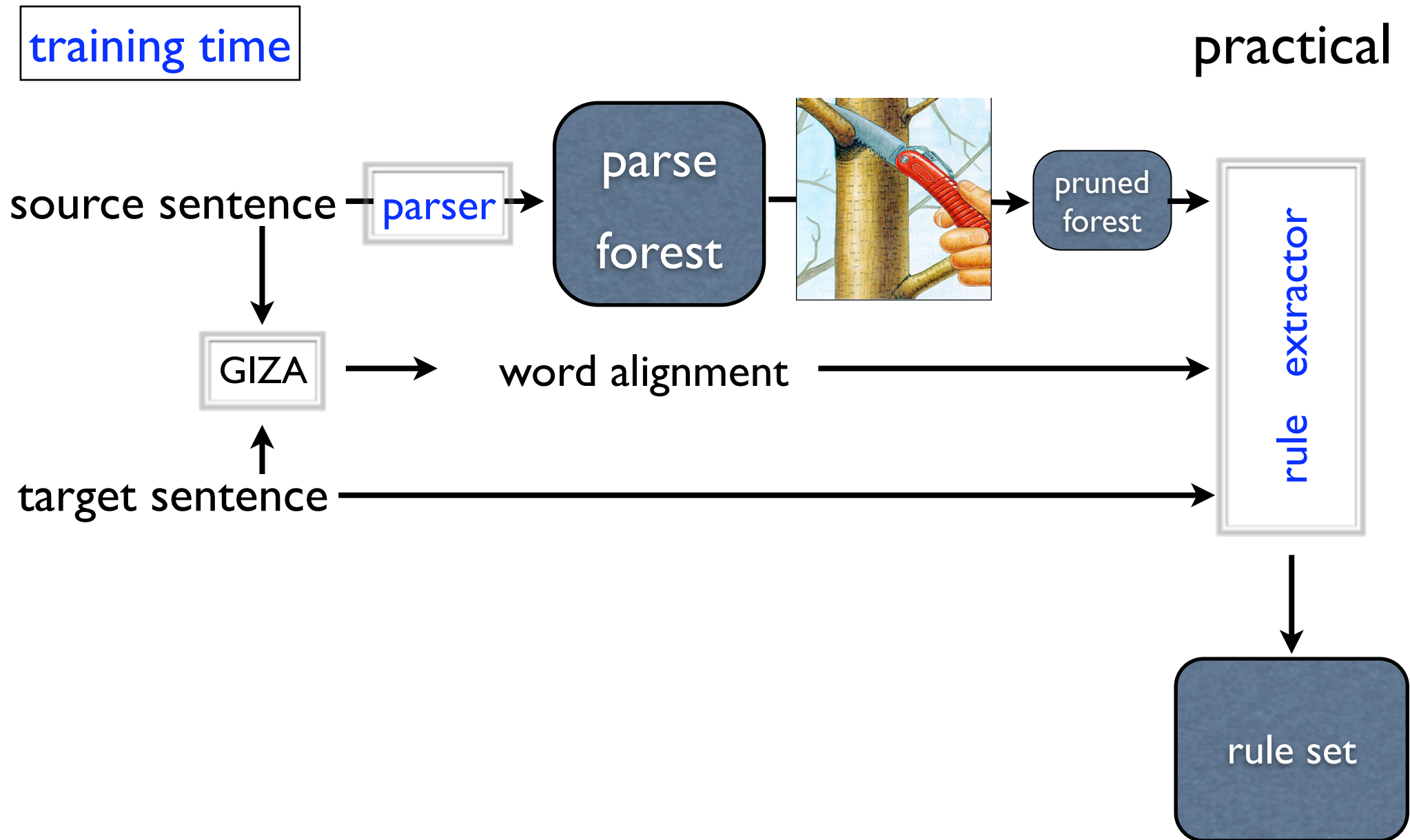
Rule Extraction Pipeline



Rule Extraction Pipeline



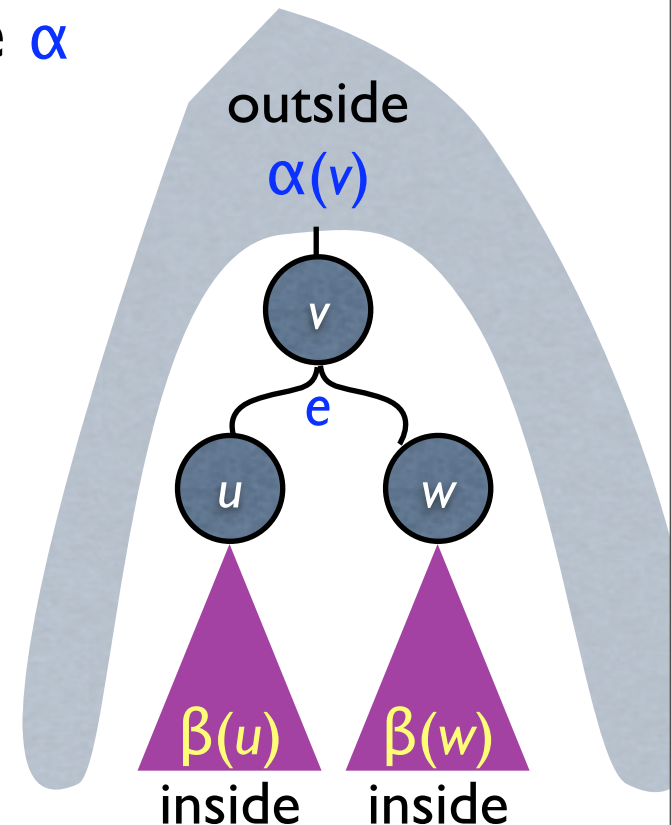
Rule Extraction Pipeline



Inside-Outside Forest Pruning

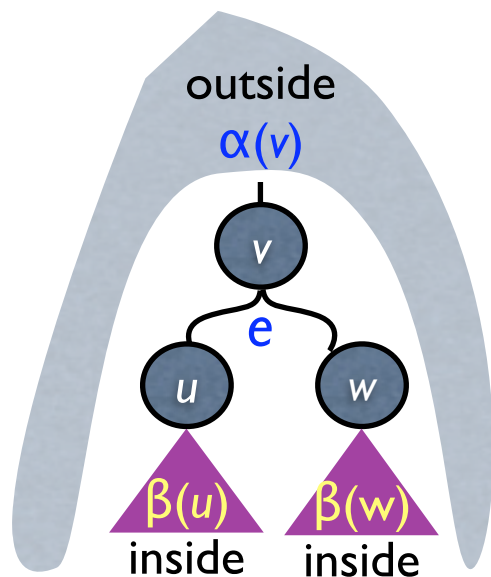
- prune *unpromising* hyperedges
 - cost of best derivation that traverses e
- inside-outside, (max) marginal probs
 - first compute Viterbi inside β , outside α
- merit $\alpha\beta(e) = \alpha(v) \cdot p(e) \cdot \beta(u) \beta(w)$
 - similar to “expected count” in EM
- prune away a hyperedge e if
$$\alpha\beta(e) / \beta(\text{TOP}) > p$$
for some threshold p

(amount of deviation from 1-best derivation)



Fractional Rule Counts

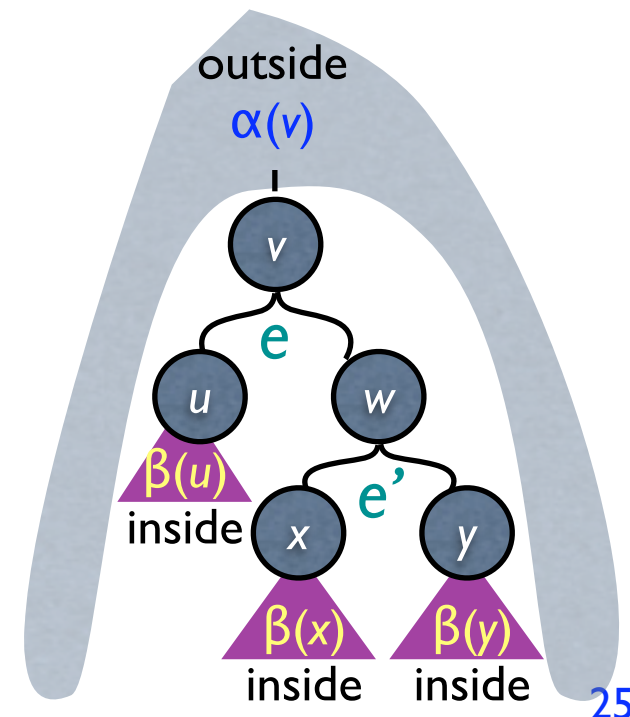
- tree-based: every rule extracted gets a unit count
- forest-based: should penalize rules extracted from non 1-best parses
- each rule gets a **fractional count** based on parse hyperedges
- same idea as forest pruning: inside-outside merit



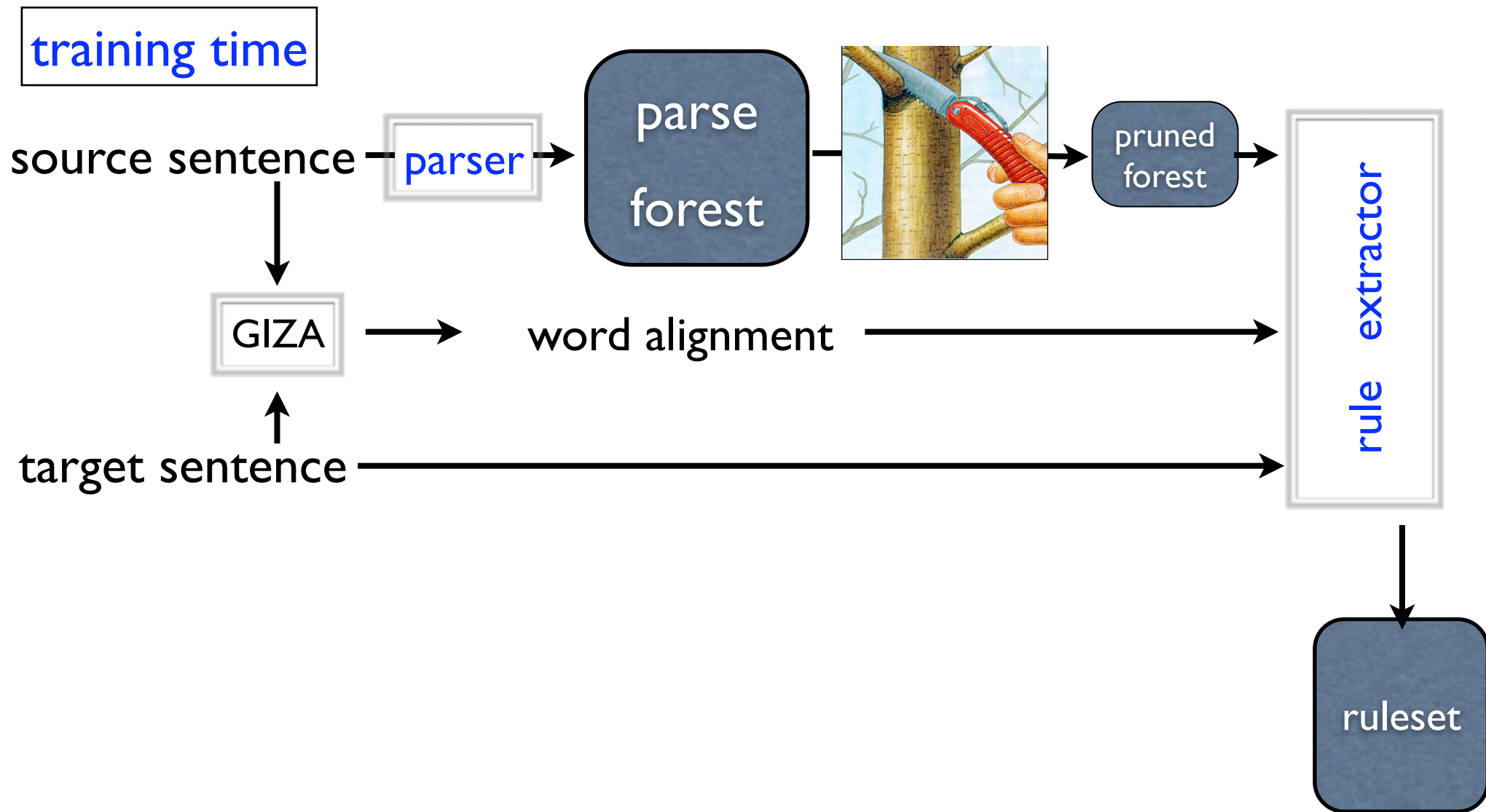
$$\alpha\beta(r) = \alpha\beta(\{e, e'\}) = \alpha(v)$$

- $p(e) p(e')$
- $\beta(u)\beta(x)\beta(y)$

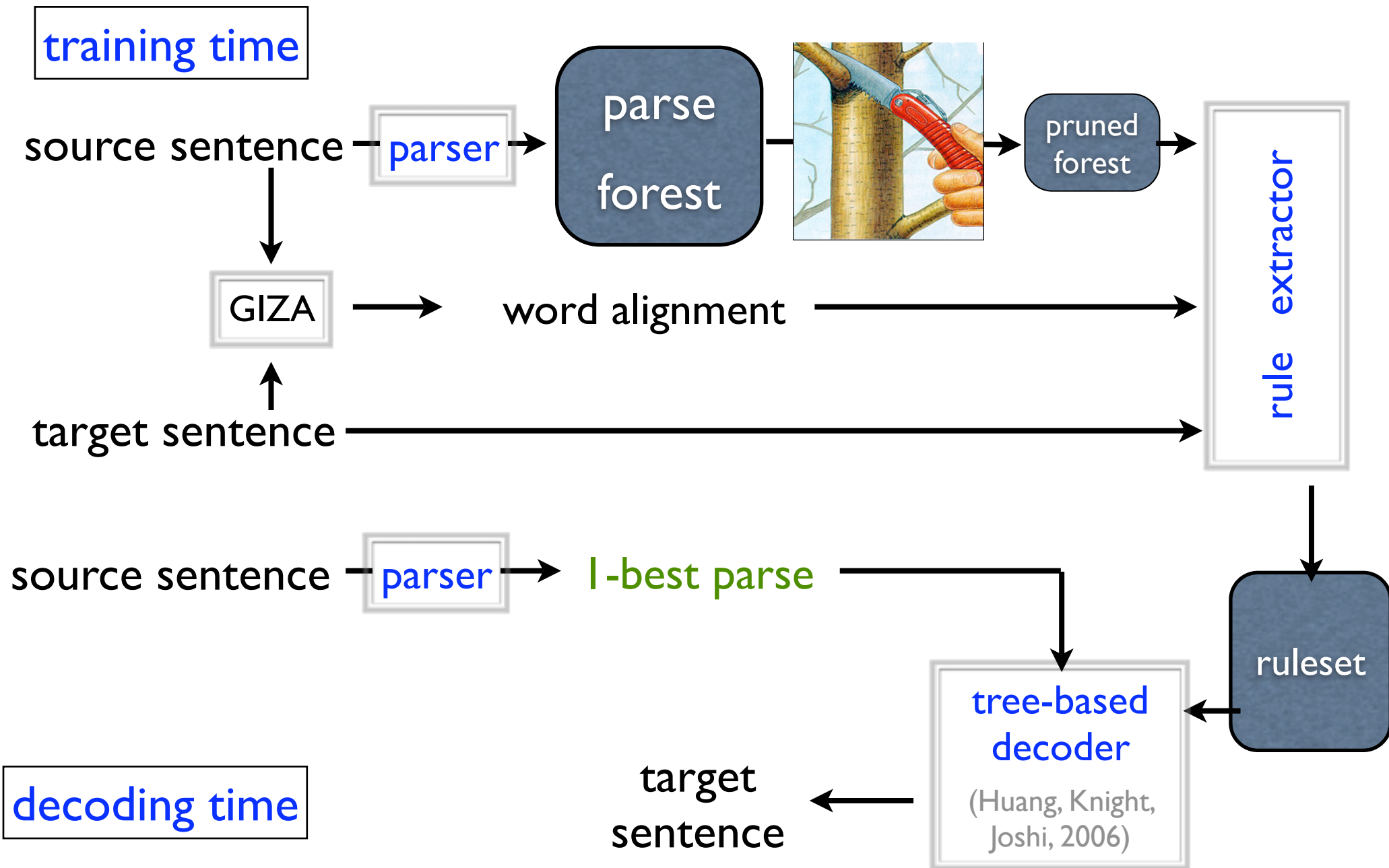
$$\text{count}(r) = \alpha\beta(r) / \beta(\text{TOP})$$



The Whole Forest Pipeline



The Whole Forest Pipeline



The Whole Forest² Pipeline

