Tree-based and Forest-based Translation

Yang Liu

Institute of Computing Technology
Chinese Academy of Sciences

Liang Huang

Information Sciences Institute
University of Southern California

Outline

- Part 1: Tree-based Translation
  - Overview and Motivation
  - Tree-to-String Model and Decoding
  - Tree-to-String Rule Extraction
  - Language Model-Integrated Decoding: Cube Pruning
- Part 2: Forest-based Translation
  - Packed Forest
  - Forest-based Decoding
  - Forest-based Rule Extraction
- Part 3: Extensions
  - Tree-to-Tree Translation
  - Tree Sequence-based Translation
  - Joint Parsing and Translation
- Part 4: Conclusion

Natural Languages are Different

I love you

Я л ю б л ю т е б я

我爱你

당신을 사랑합니다

Eu te amo

Je t'aime

אני אוהב אותך

من شما را دوست دارم

Tôi yêu bạn

Ich liebe dich

Te quiero

Miluji tě

Ti amo

ผมรักคุณ

わ た し は 、 あ な た を 愛 し て い ま す

Ik hou van je

Jag älskar dig

By Google Translate

Translation is Hard!

connocting poopie

HELP ONESELF
TERMINATING MACHINE

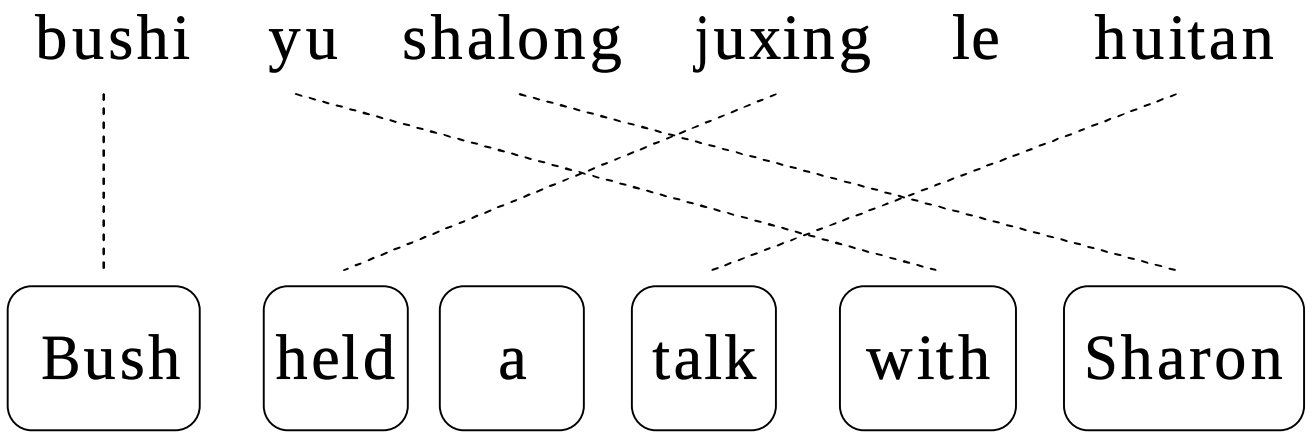
Machine Translation

布什 与 沙龙 举行 了 会谈

bushi yu shalong juxing le huitan

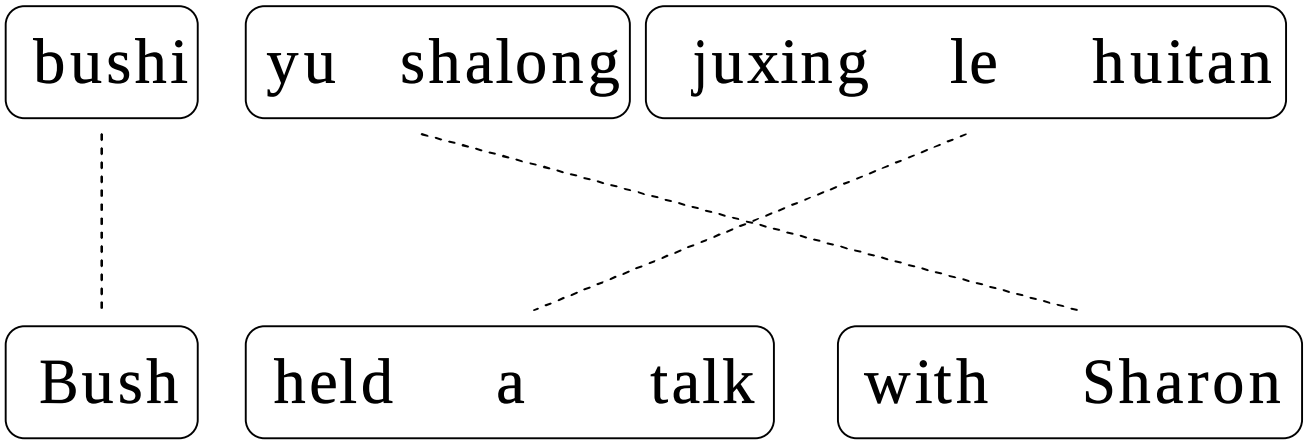
Bush held a talk with Sharon

Word-based MT



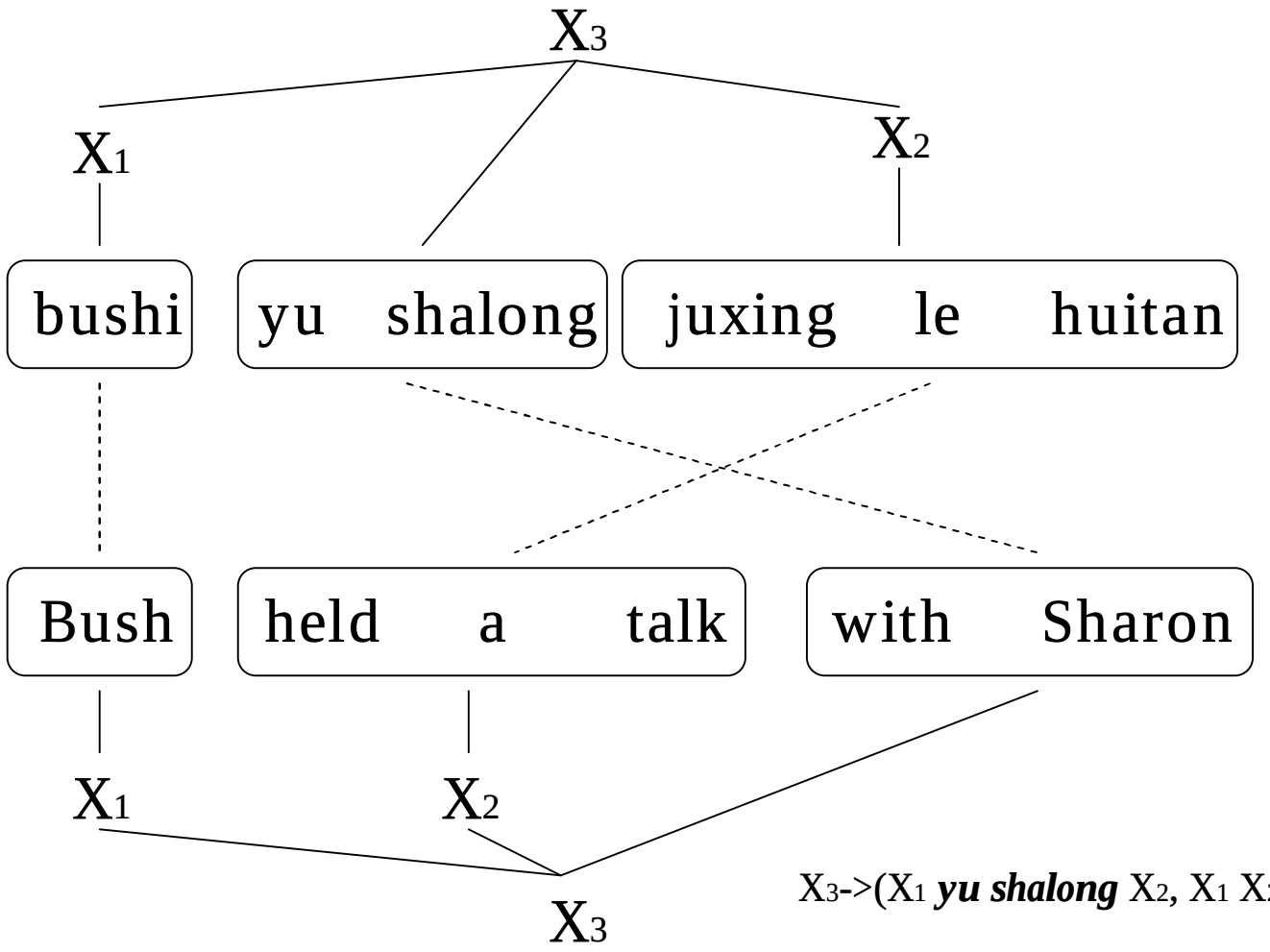
(Brown et al., 1993)

Phrase-based MT



(Koehn et al., 2003; Och and Ney, 2004)

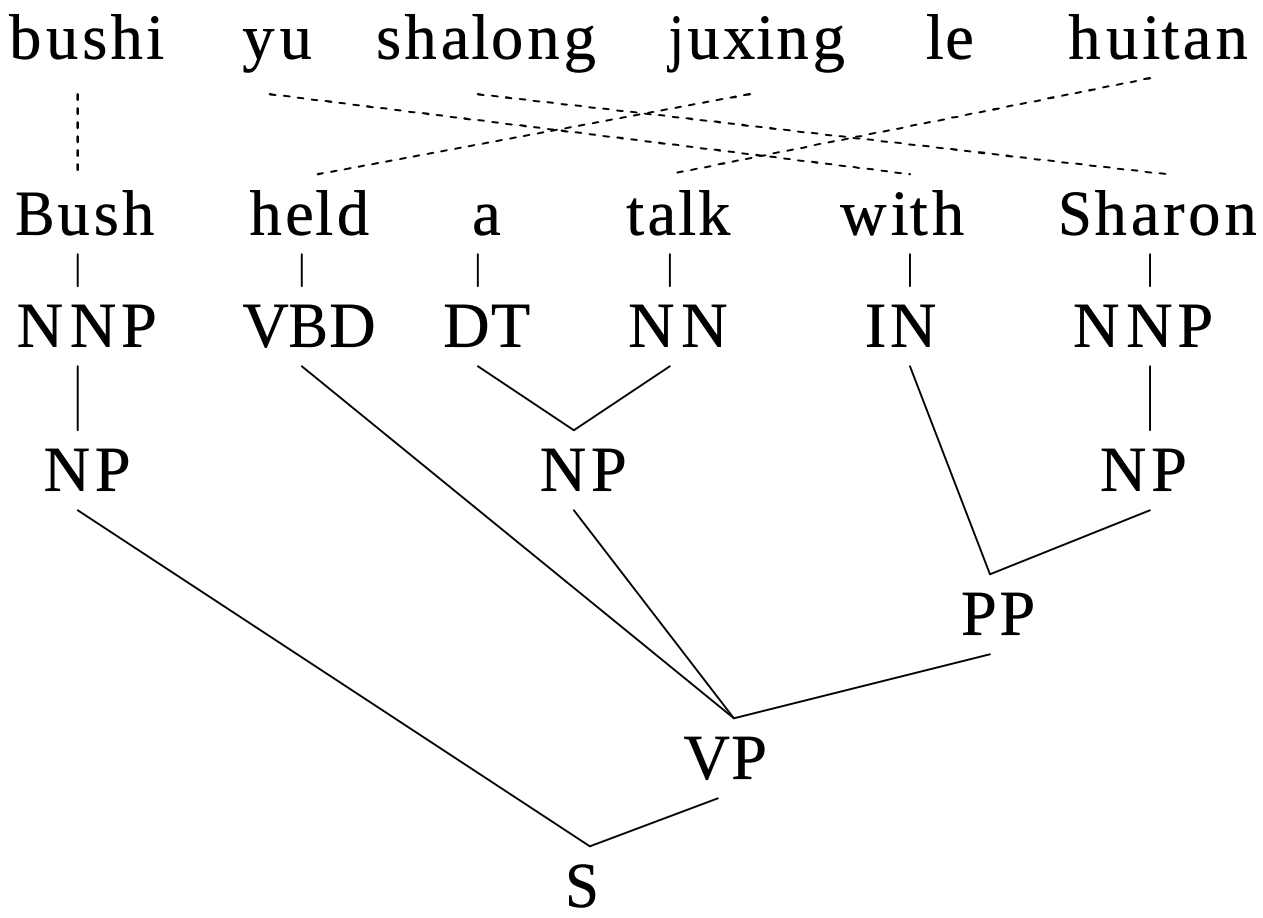
Hierarchical Phrase-based MT



X_3->(X_1 ***yu shalong*** X_2, X_1 X_2 ***with Sharon***)

(Chiang, 2005; Chiang, 2007)

Syntax-based MT



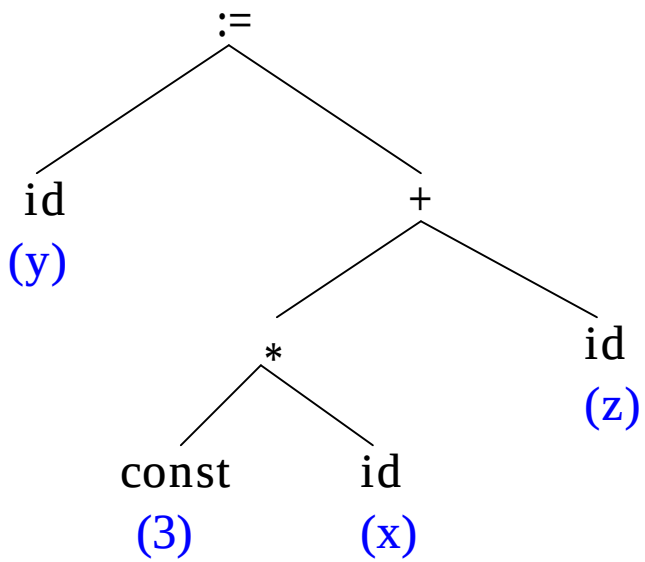
(Yamda and Knight, 2001; Galley et al., 2006; Shen et al., 2008)

Motivation

- Human Translation
  - Understand the source sentence
  - Generate the target sentence
- Compiling
  - Parse input program into a syntax tree
  - Generate code in machine language

Syntax-Directed Translation for Compiling

- Input: ***y*:=3*x+z**
- Parsing:



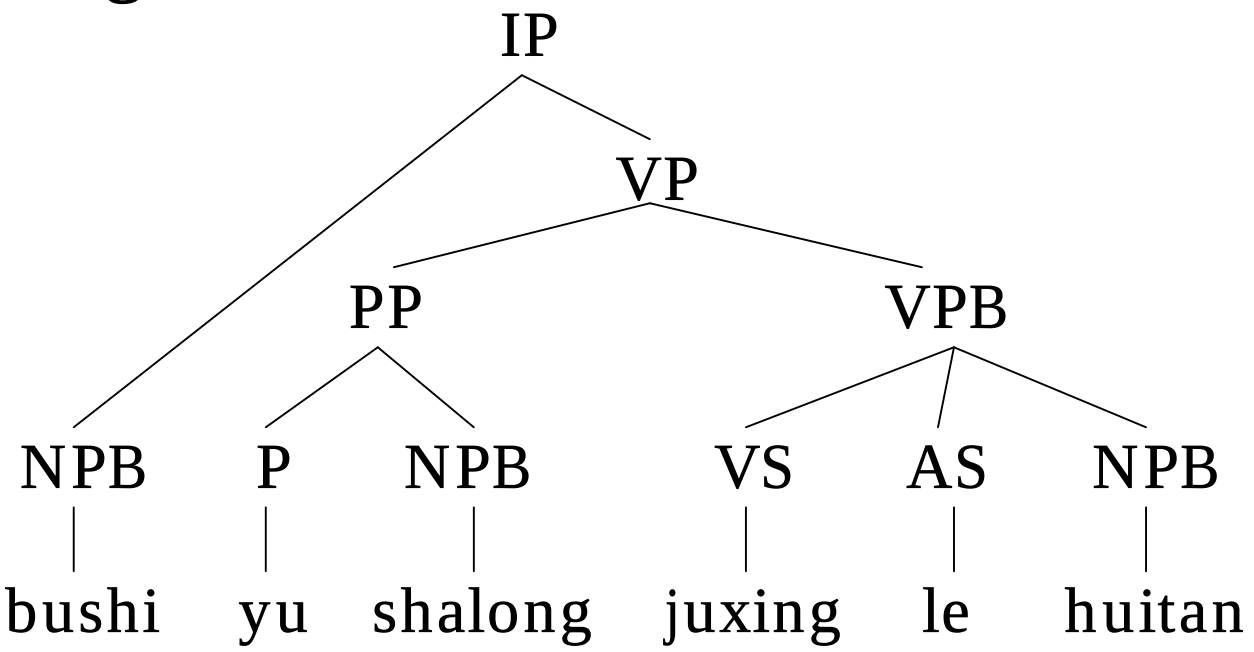
(Irons, 1961; Lewis and Stearns, 1968; Aho and Ullman., 1972)

Motivation

- Human Translation
  - Understand the source sentence
  - Generate the target sentence
- Compiling
  - Parse input program into a syntax tree
  - Generate code in machine language
- Machine Translation
  - Parse the source sentence into a tree
  - Recursively transfer the tree into the target language

Syntax-Directed Translation for MT

- Input: ***bushi yu shalong juxing le huitan***
- Parsing:



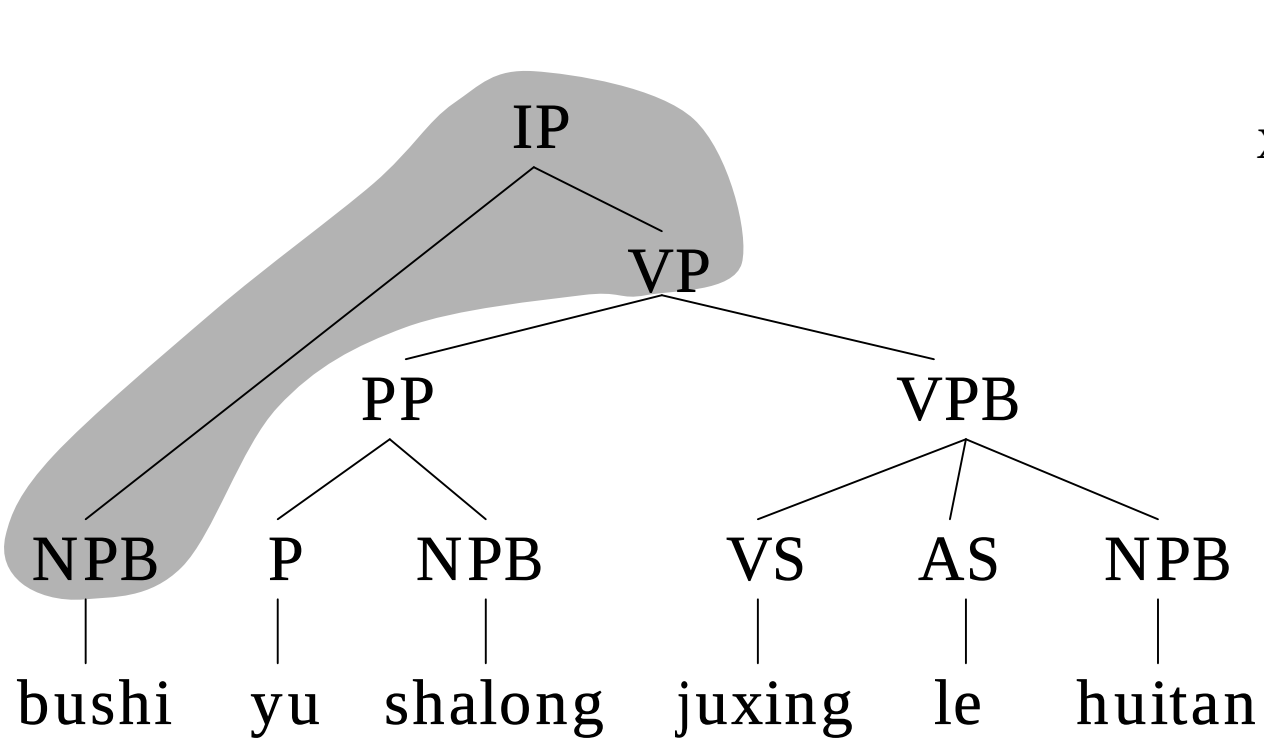
(Huang et al., 2006)

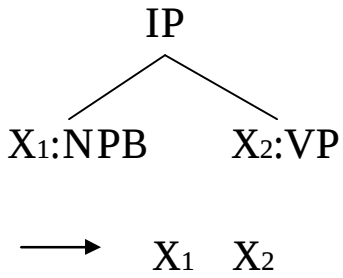
Outline

- Part 1: Tree-based Translation
  - *Overview and Motivation*
  - Tree-to-String Model and Decoding
  - Tree-to-String Rule Extraction
  - Language Model-Integrated Decoding: Cube Pruning
- Part 2: Forest-based Translation
  - Packed Forest
  - Forest-based Decoding
  - Forest-based Rule Extraction
- Part 3: Extensions
  - Tree-to-Tree Translation
  - Tree Sequence-based Translation
  - Joint Parsing and Translation
- Part 4: Conclusion

Tree-to-String Translation

- Recursive rewrite by pattern-matching

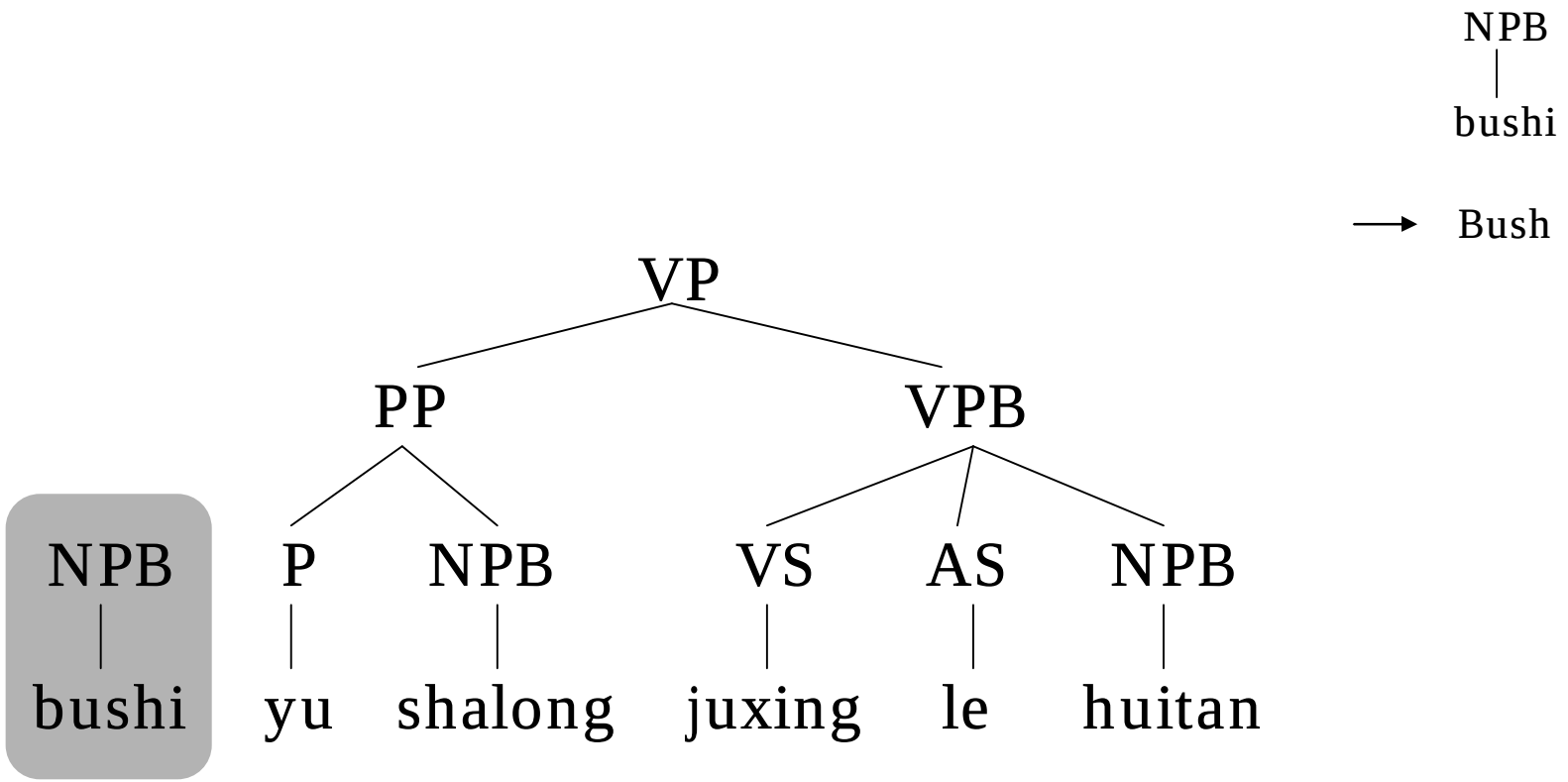




(Liu et al., 2006; Huang et al., 2006)

Tree-to-String Translation

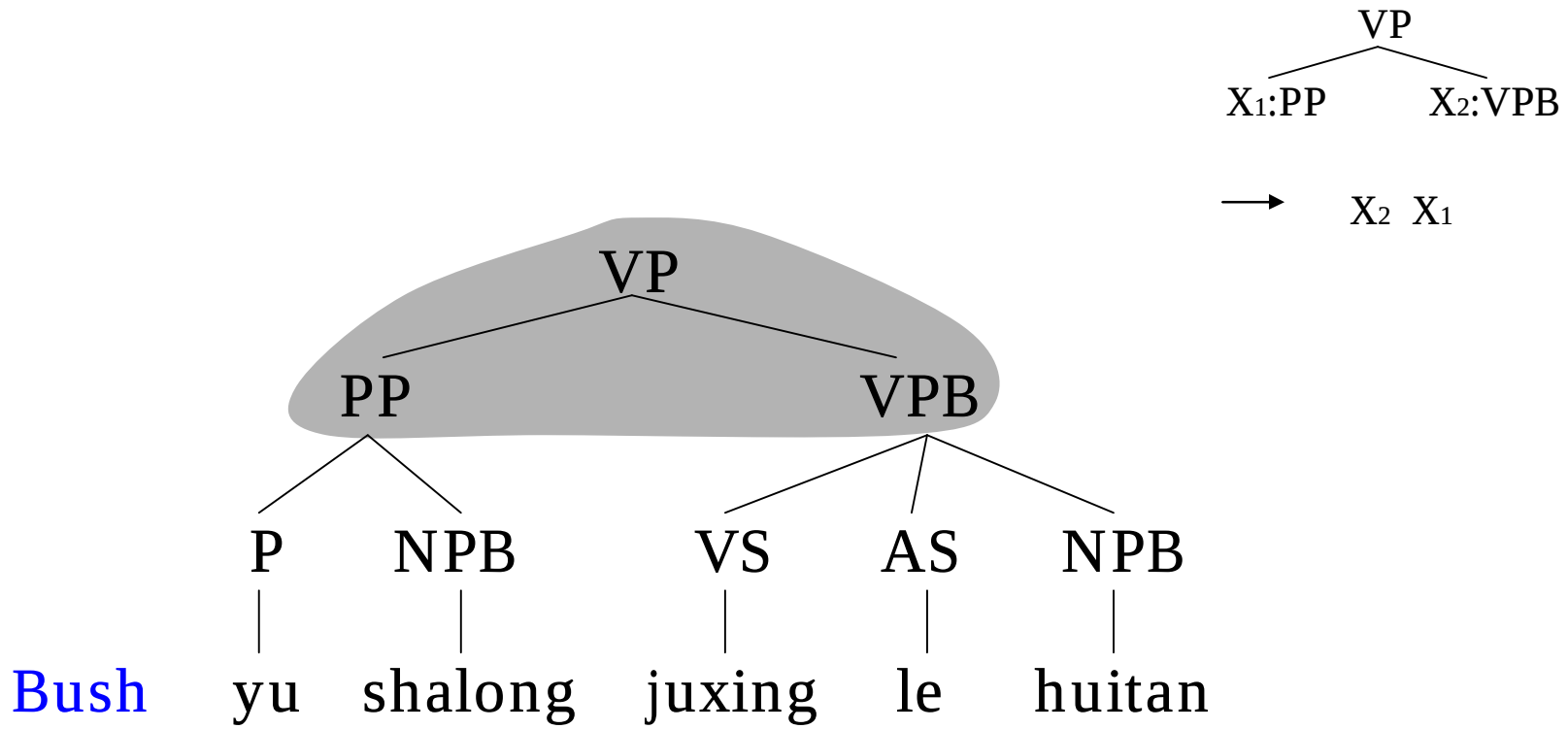
- Recursive rewrite by pattern-matching



(Liu et al., 2006; Huang et al., 2006)

Tree-to-String Translation

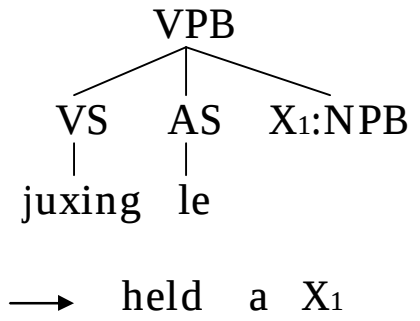
- Recursive rewrite by pattern-matching

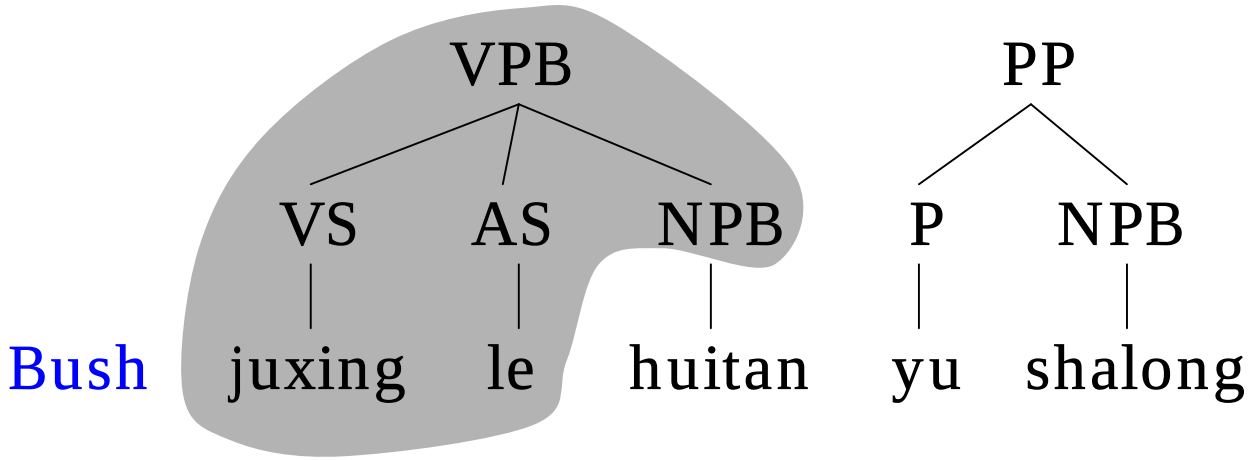


(Liu et al., 2006; Huang et al., 2006)

Tree-to-String Translation

- Recursive rewrite by pattern-matching

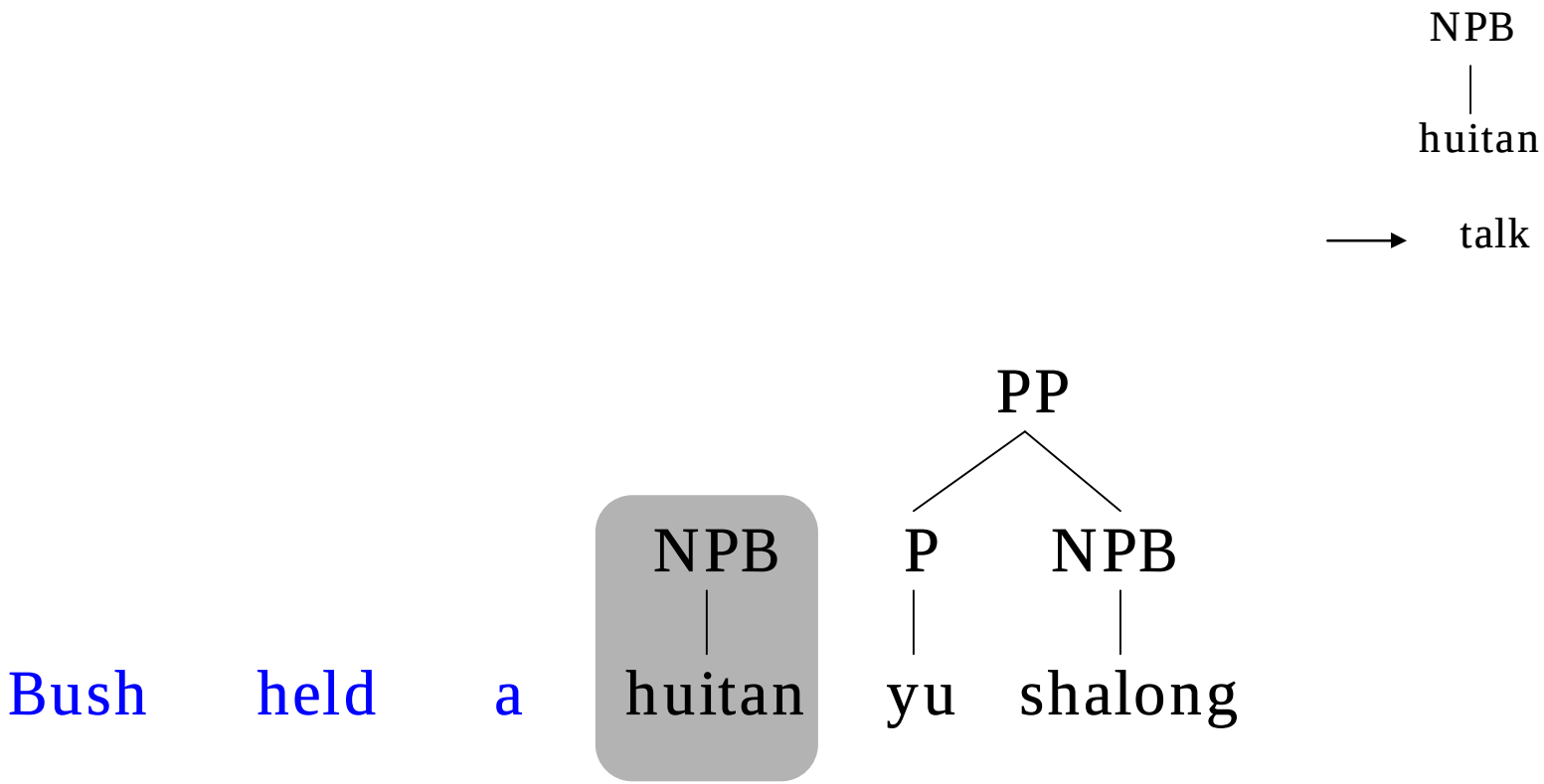




(Liu et al., 2006; Huang et al., 2006)

Tree-to-String Translation

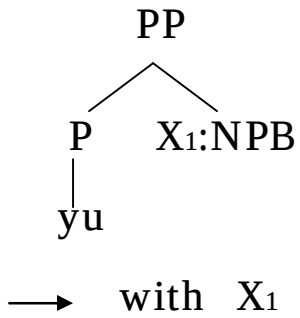
- Recursive rewrite by pattern-matching



(Liu et al., 2006; Huang et al., 2006)

Tree-to-String Translation

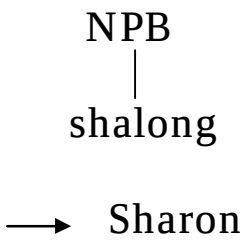
- Recursive rewrite by pattern-matching

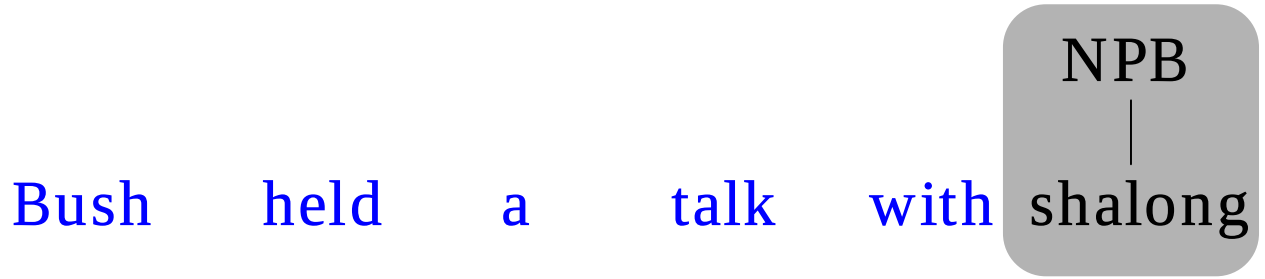


(Liu et al., 2006; Huang et al., 2006)

Tree-to-String Translation

- Recursive rewrite by pattern-matching





(Liu et al., 2006; Huang et al., 2006)

Tree-to-String Translation

- Recursive rewrite by pattern-matching

Tree-to-string translation

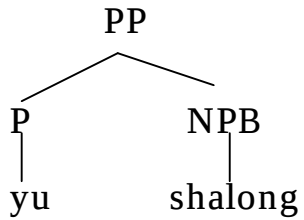
- Syntax-directed translation (e.g., Irons, 1961)
- Tree transducer (e.g., Knight and Graehl, 2005)
- Synchronous grammar (e.g., Eisner, 2003)
- …

Bush held a talk with Sharon

(Liu et al., 2006; Huang et al., 2006)

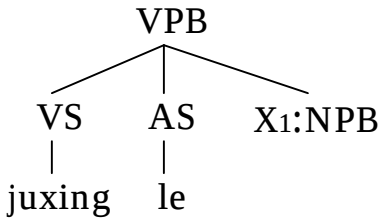
Expressive Power

phrasal translation



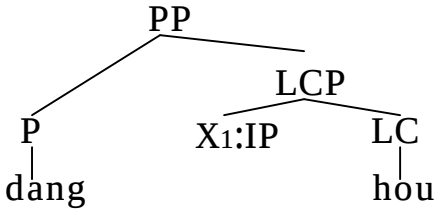
→ with Sharon

non-constituent phrase



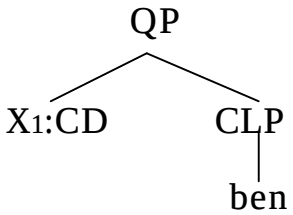
→ held a X1

non-contiguous phrase



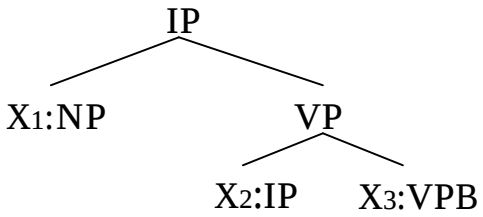
→ when X1

word omission



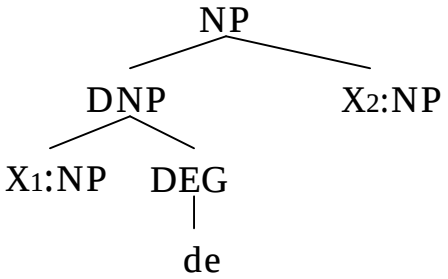
→ X1

multilevel re-ordering



→ X1 X3 X2

lexicalized re-ordering



→ X2 of X1

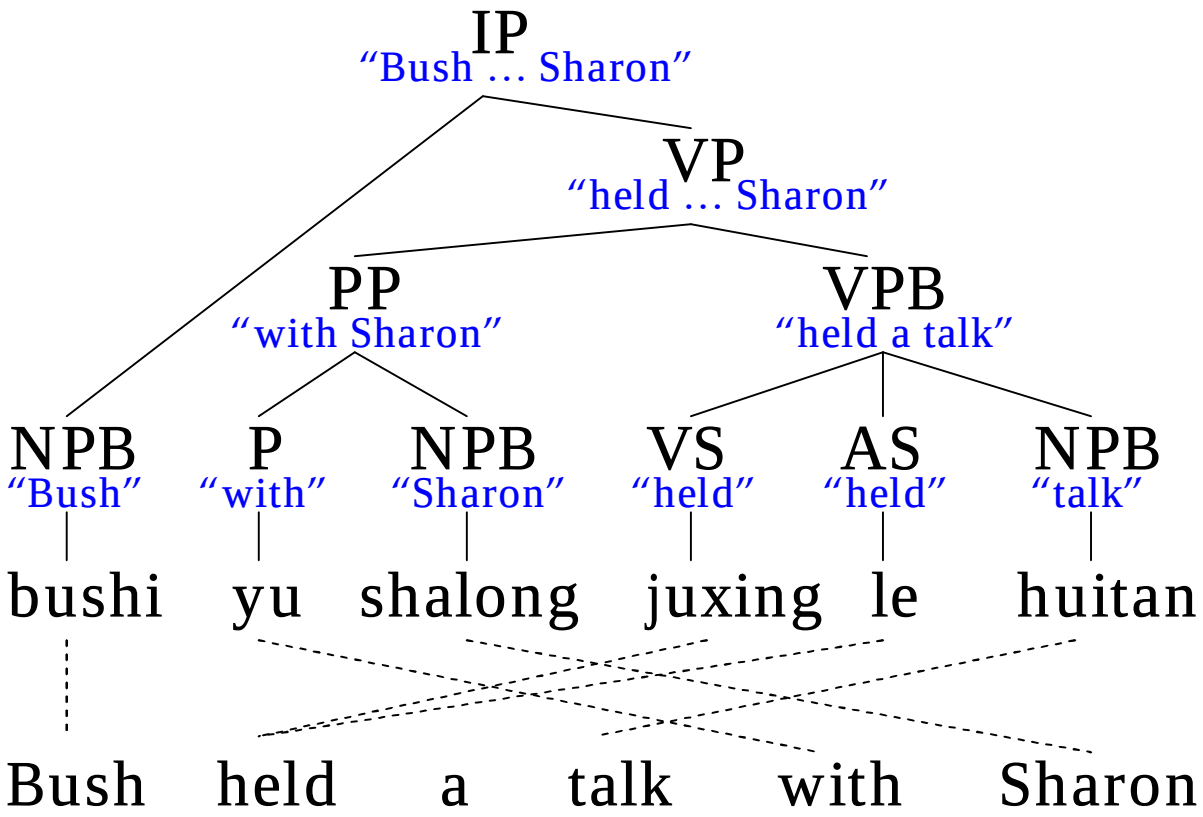
(Knight and Graehl, 2005)

Outline

- Part 1: Tree-based Translation
  - *Overview and Motivation*
  - *Tree-to-String Model and Decoding*
  - Tree-to-String Rule Extraction
  - Language Model-Integrated Decoding: Cube Pruning
- Part 2: Forest-based Translation
  - Packed Forest
  - Forest-based Decoding
  - Forest-based Rule Extraction
- Part 3: Extensions
  - Tree-to-Tree Translation
  - Tree Sequence-based Translation
  - Joint Parsing and Translation
- Part 4: Conclusion

Tree-to-String Rule Extraction

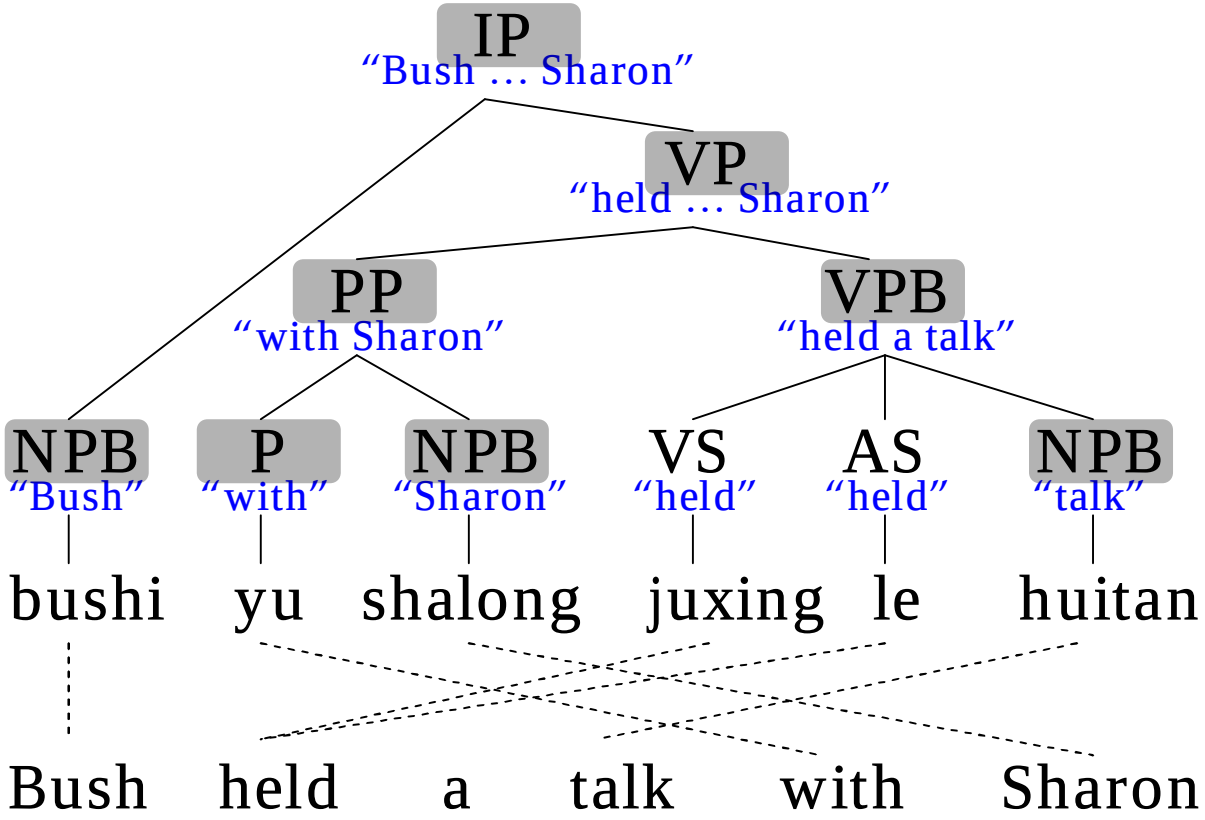
- Compute target spans



(Galley et al., 2004)

Tree-to-String Rule Extraction

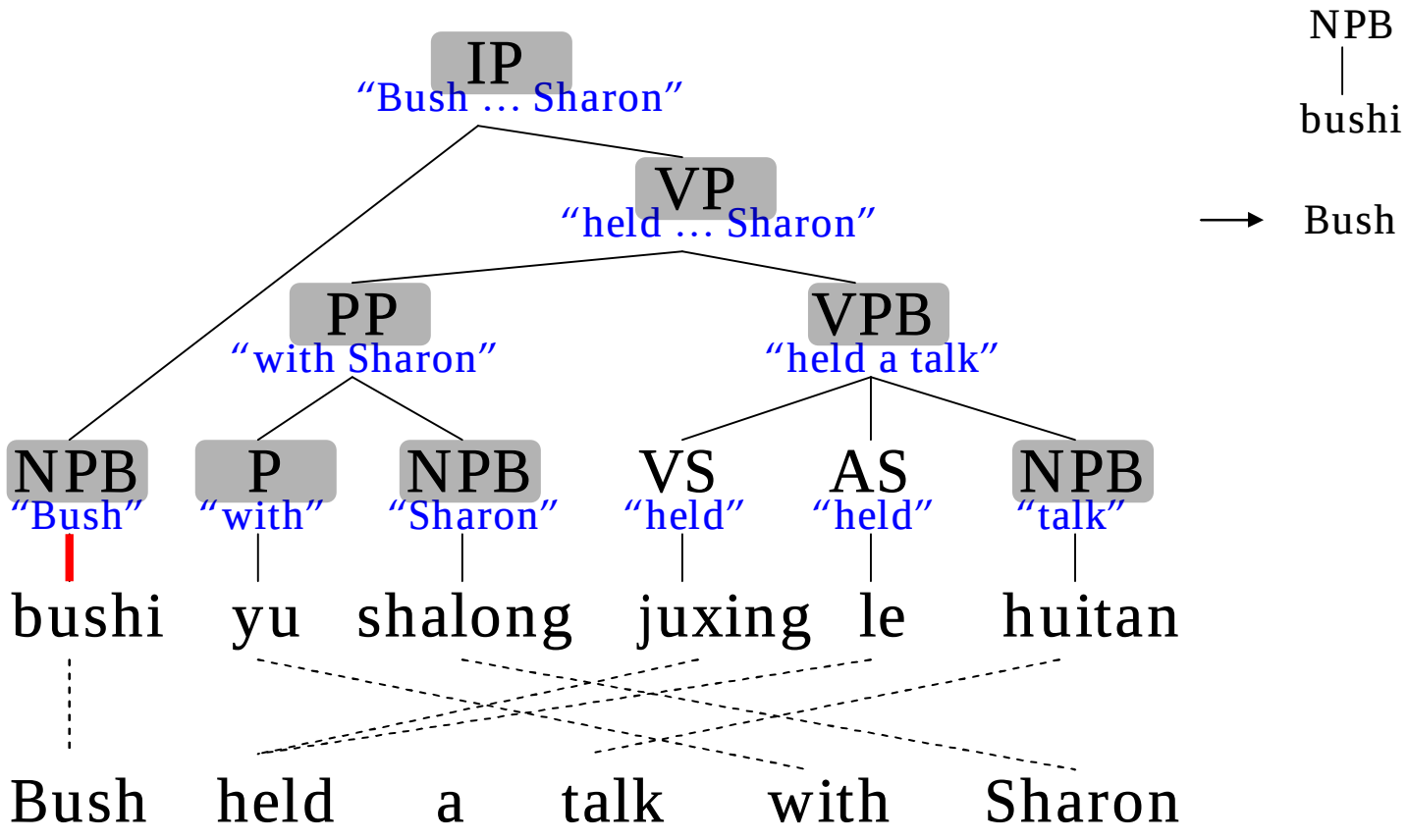
- Find admissible nodes



(Galley et al., 2004)

Tree-to-String Rule Extraction

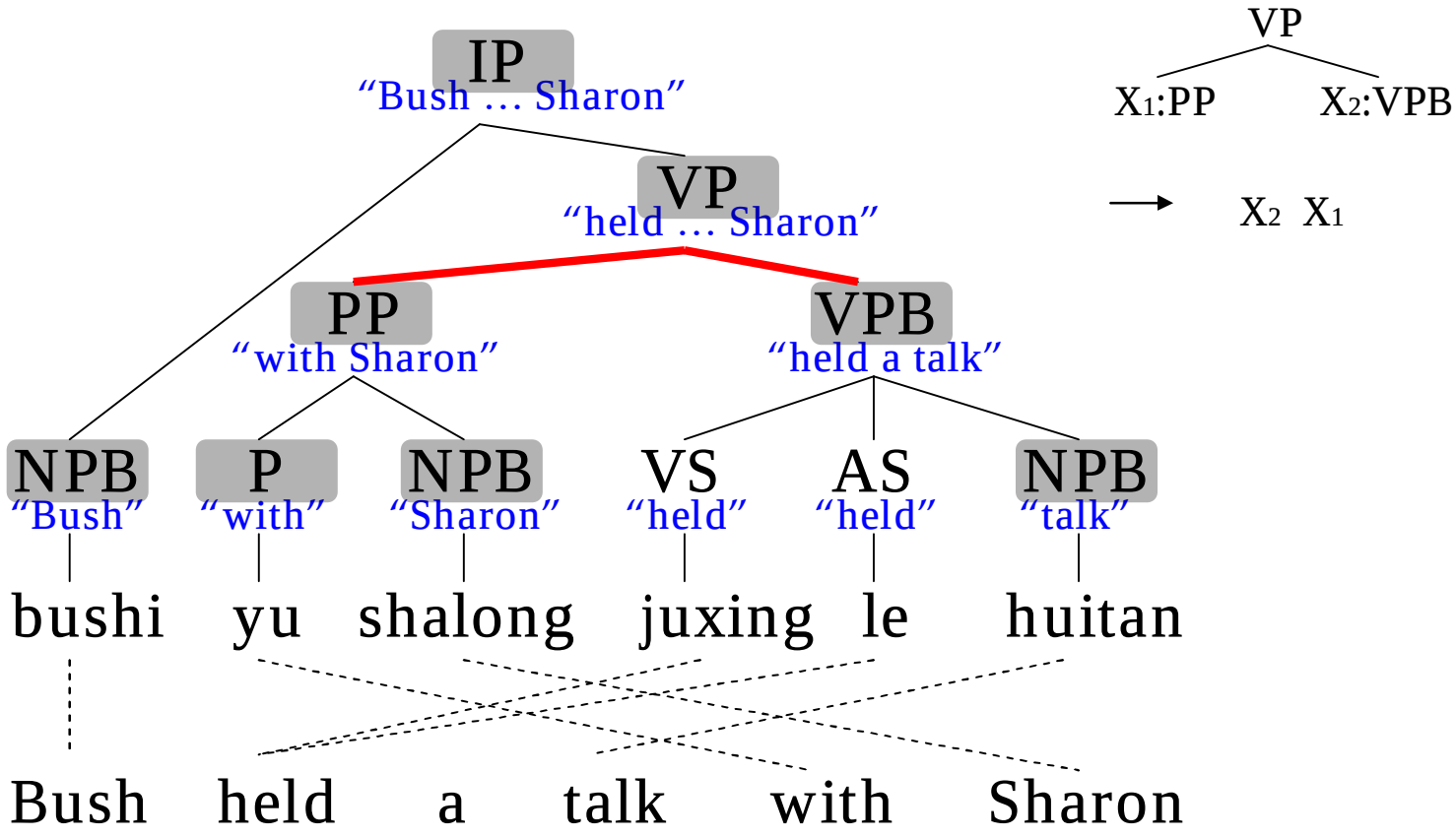
- Extract minimal rules



(Galley et al., 2004)

Tree-to-String Rule Extraction

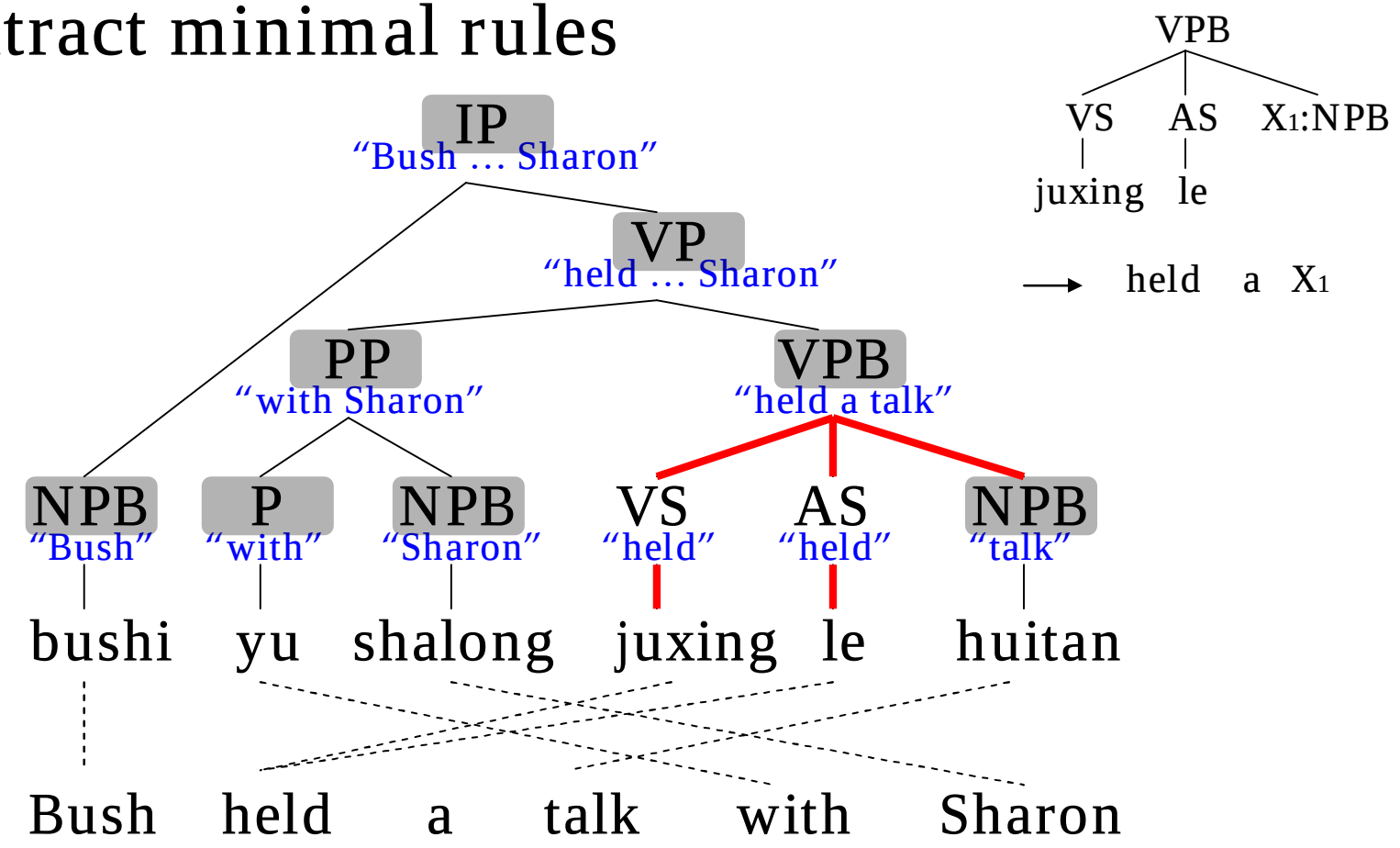
- Extract minimal rules



(Galley et al., 2004)

Tree-to-String Rule Extraction

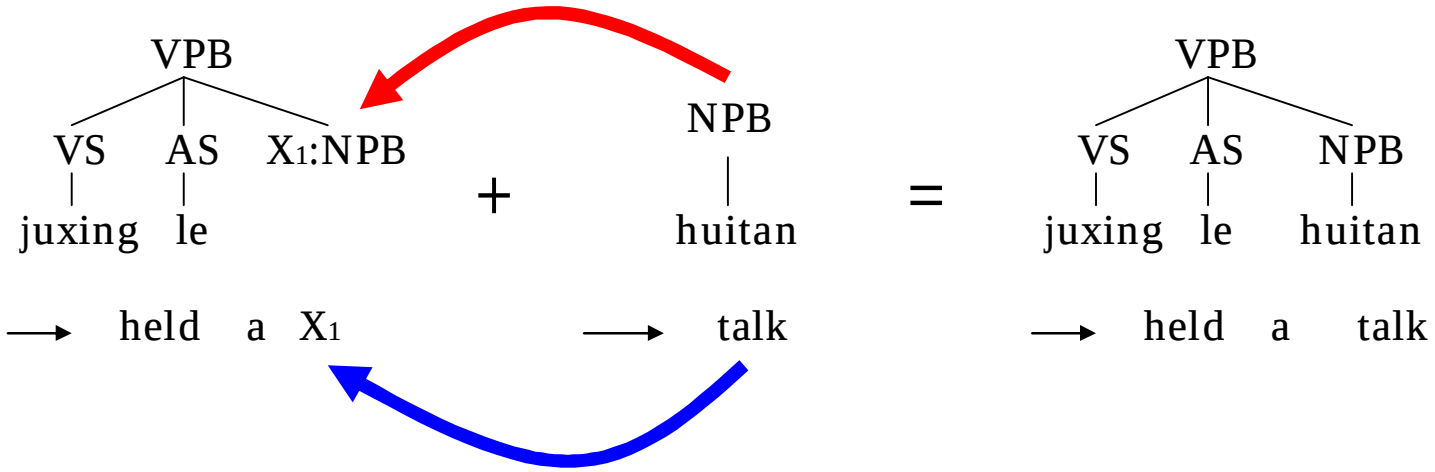
- Extract minimal rules



(Galley et al., 2004)

Tree-to-String Rule Extraction

- Get composed rules



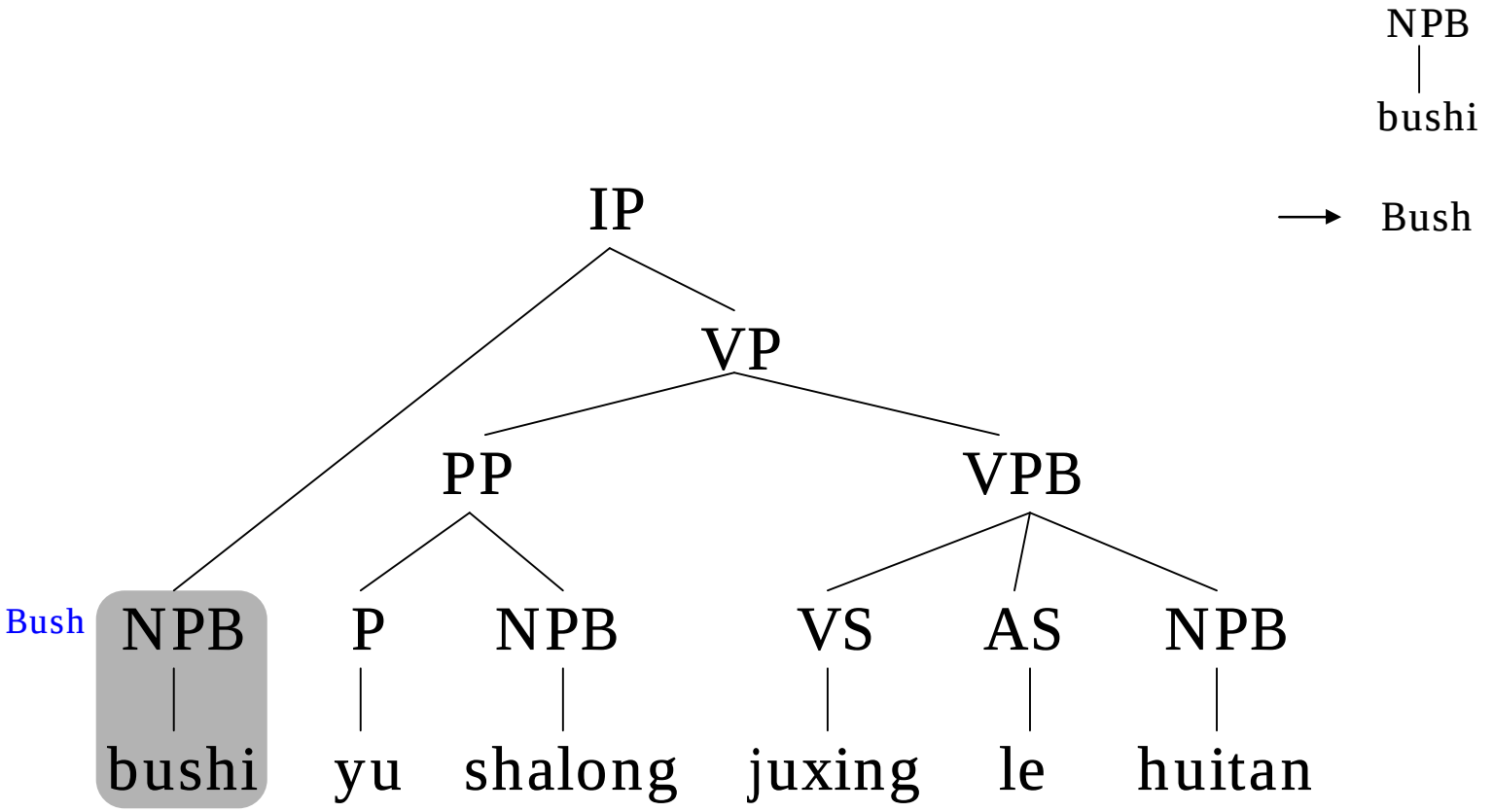
tree substitution

(Galley et al., 2006)

Outline

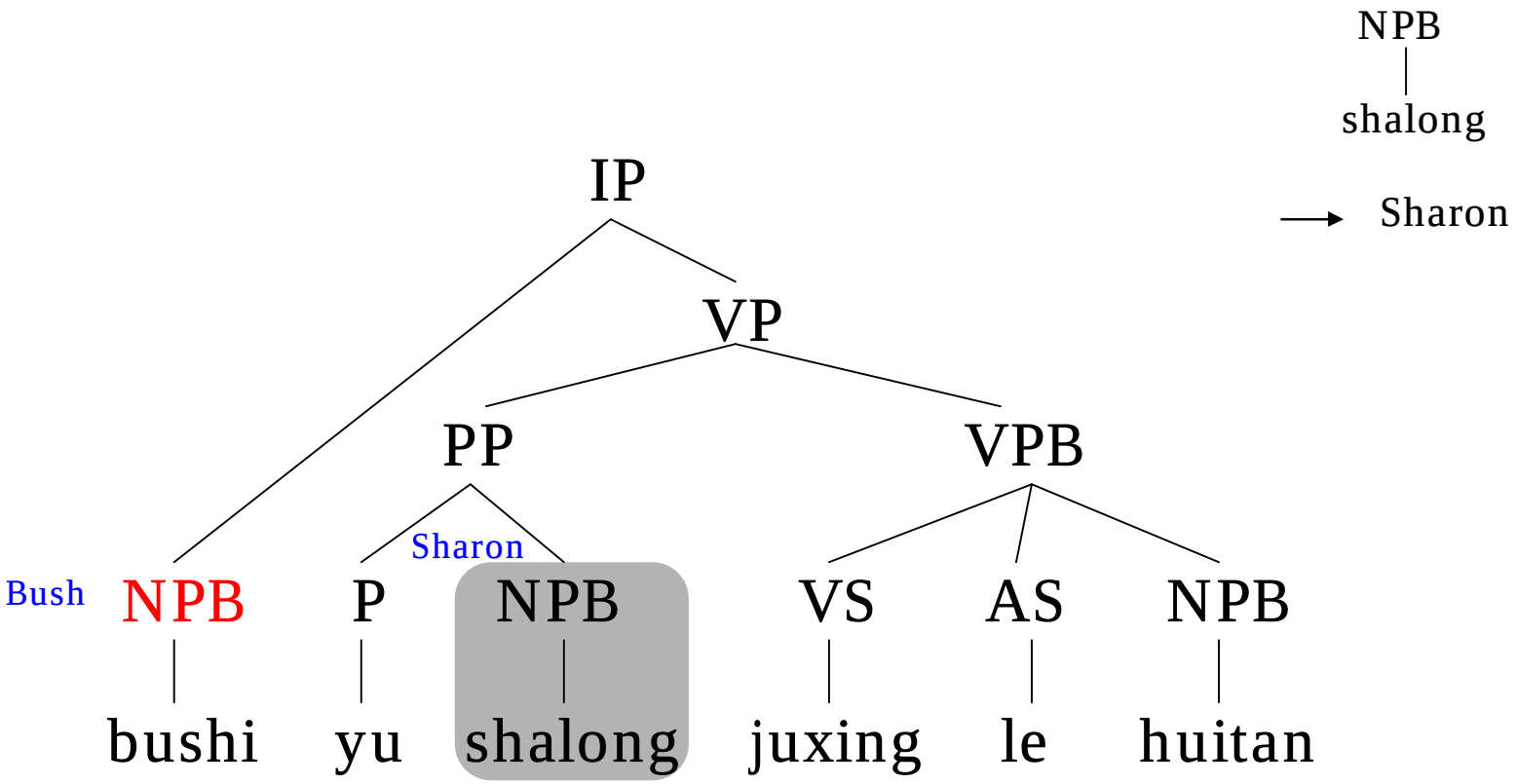
- Part 1: Tree-based Translation
  - *Overview and Motivation*
  - *Tree-to-String Model and Decoding*
  - *Tree-to-String Rule Extraction*
  - Language Model-Integrated Decoding: Cube Pruning
- Part 2: Forest-based Translation
  - Packed Forest
  - Forest-based Decoding
  - Forest-based Rule Extraction
- Part 3: Extensions
  - Tree-to-Tree Translation
  - Tree Sequence-based Translation
  - Joint Parsing and Translation
- Part 4: Conclusion

Bottom-up Decoding



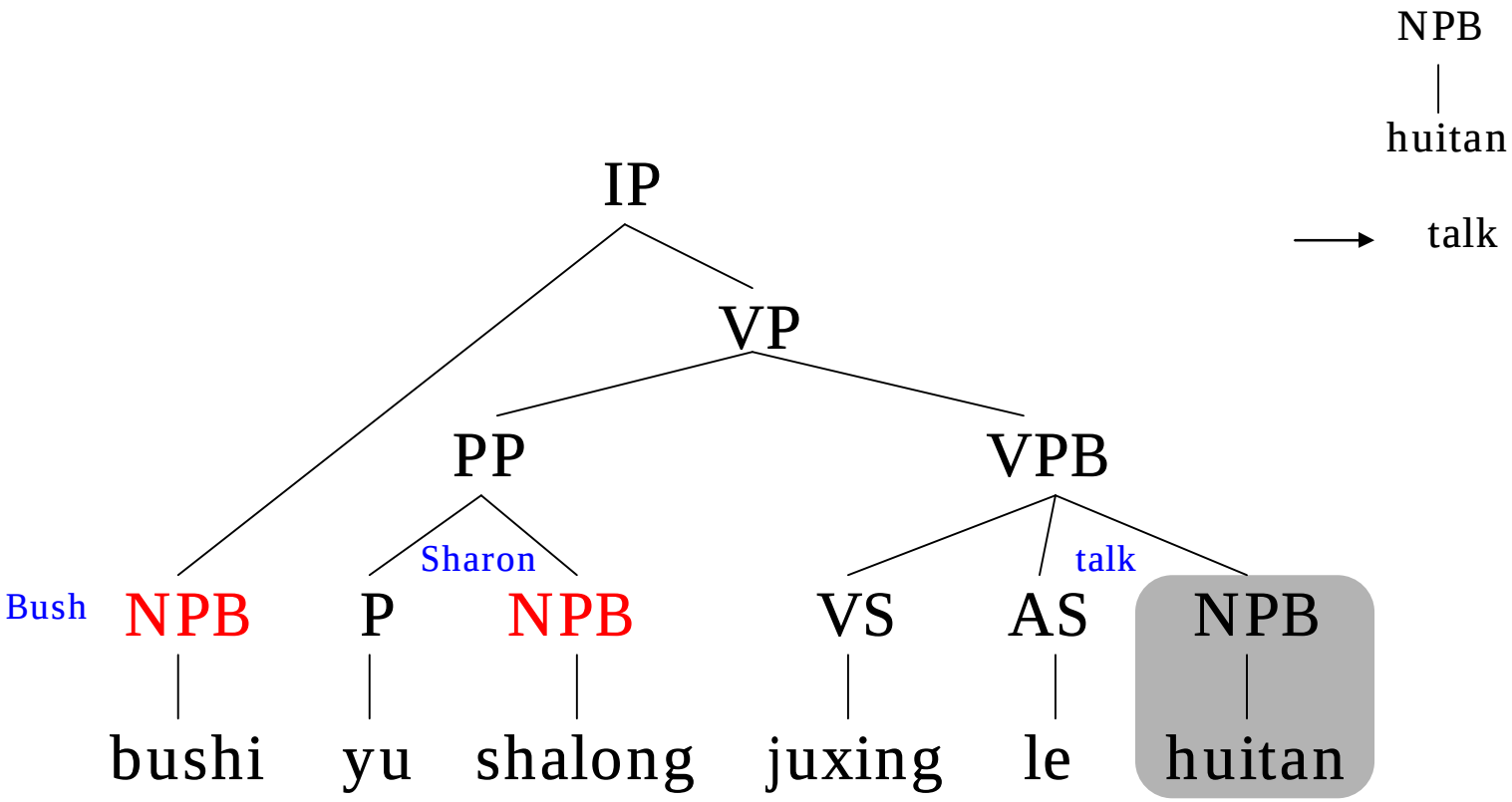
(Liu et al., 2006; Huang et al., 2006)

Bottom-up Decoding



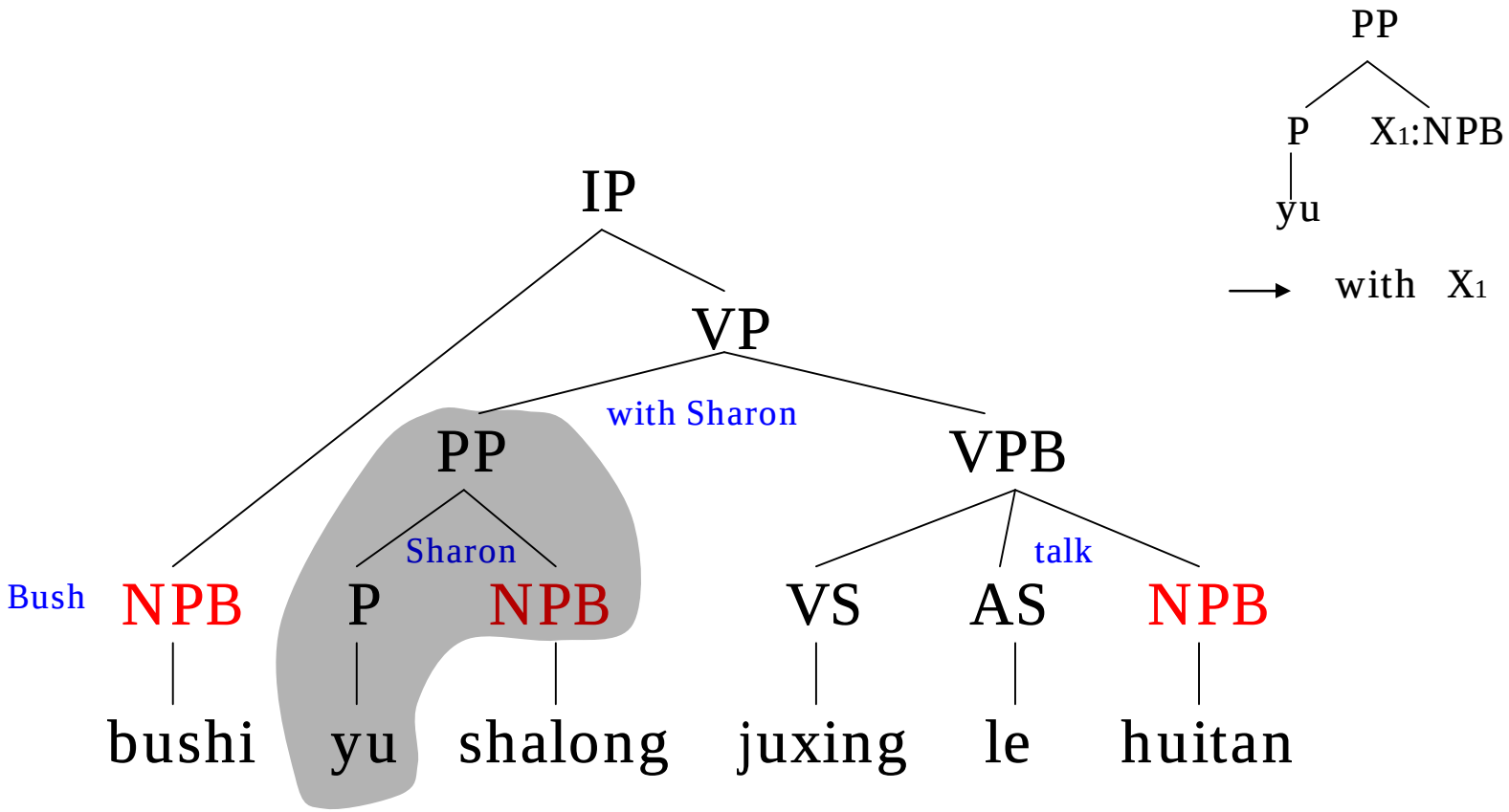
(Liu et al., 2006; Huang et al., 2006)

Bottom-up Decoding



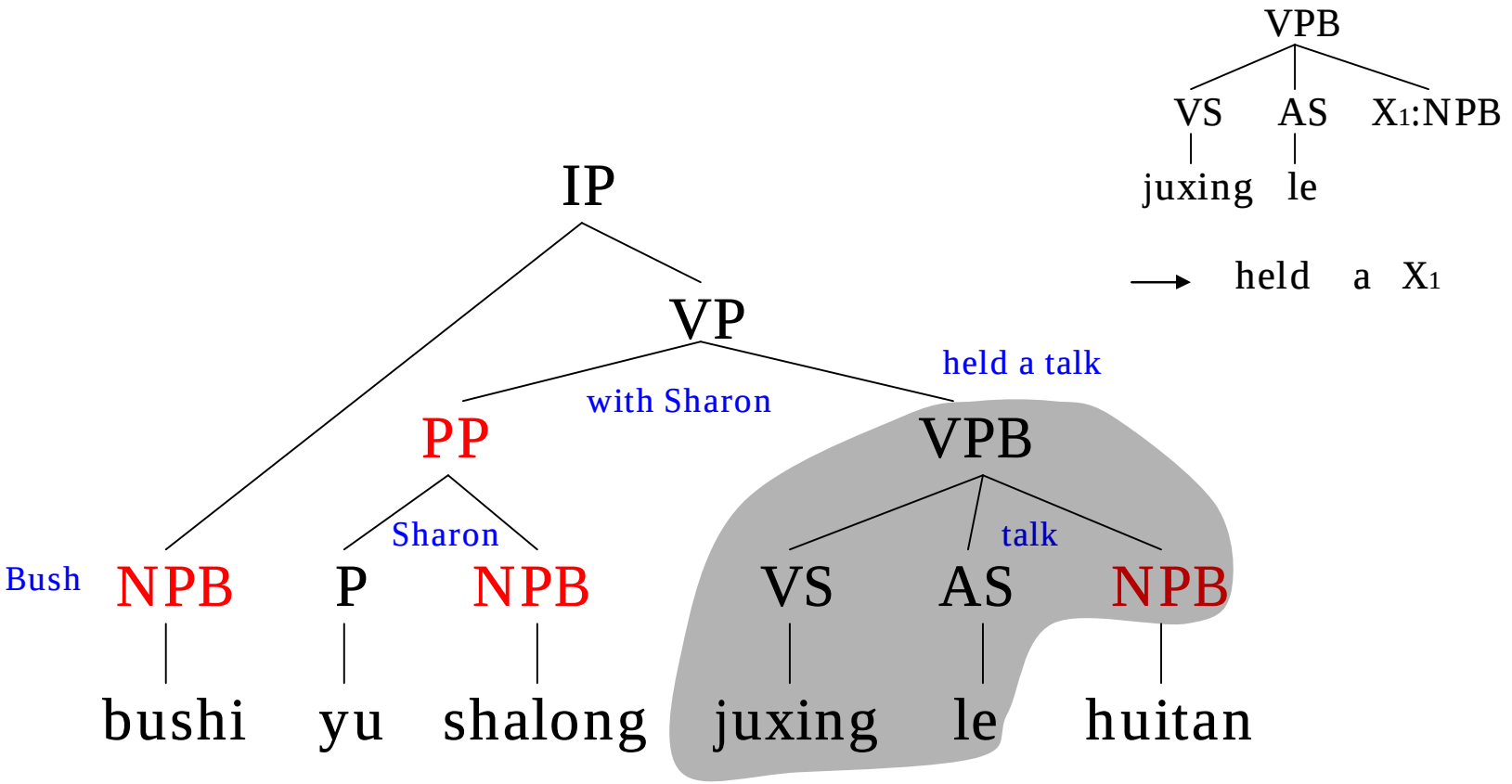
(Liu et al., 2006; Huang et al., 2006)

Bottom-up Decoding



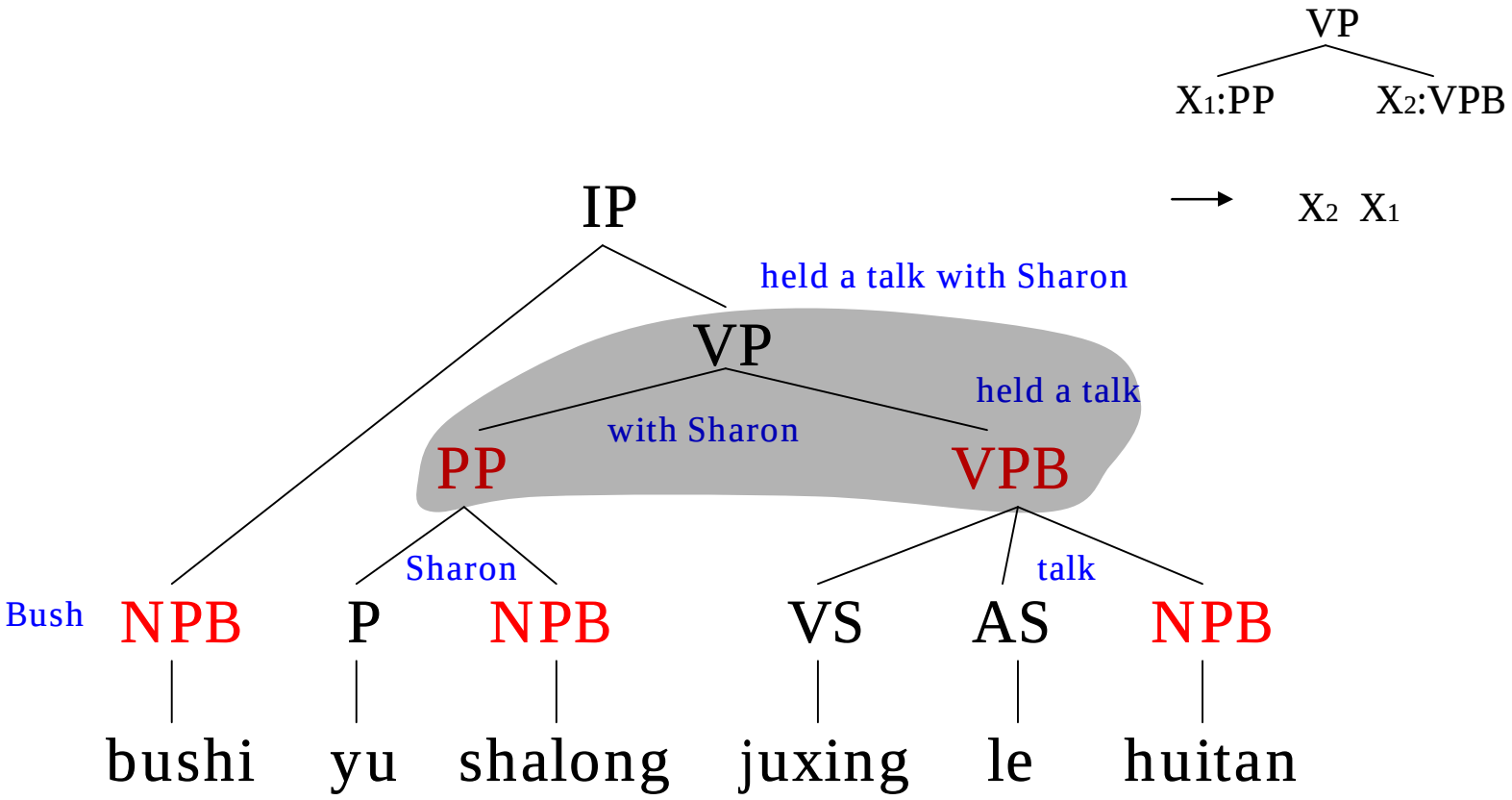
(Liu et al., 2006; Huang et al., 2006)

Bottom-up Decoding



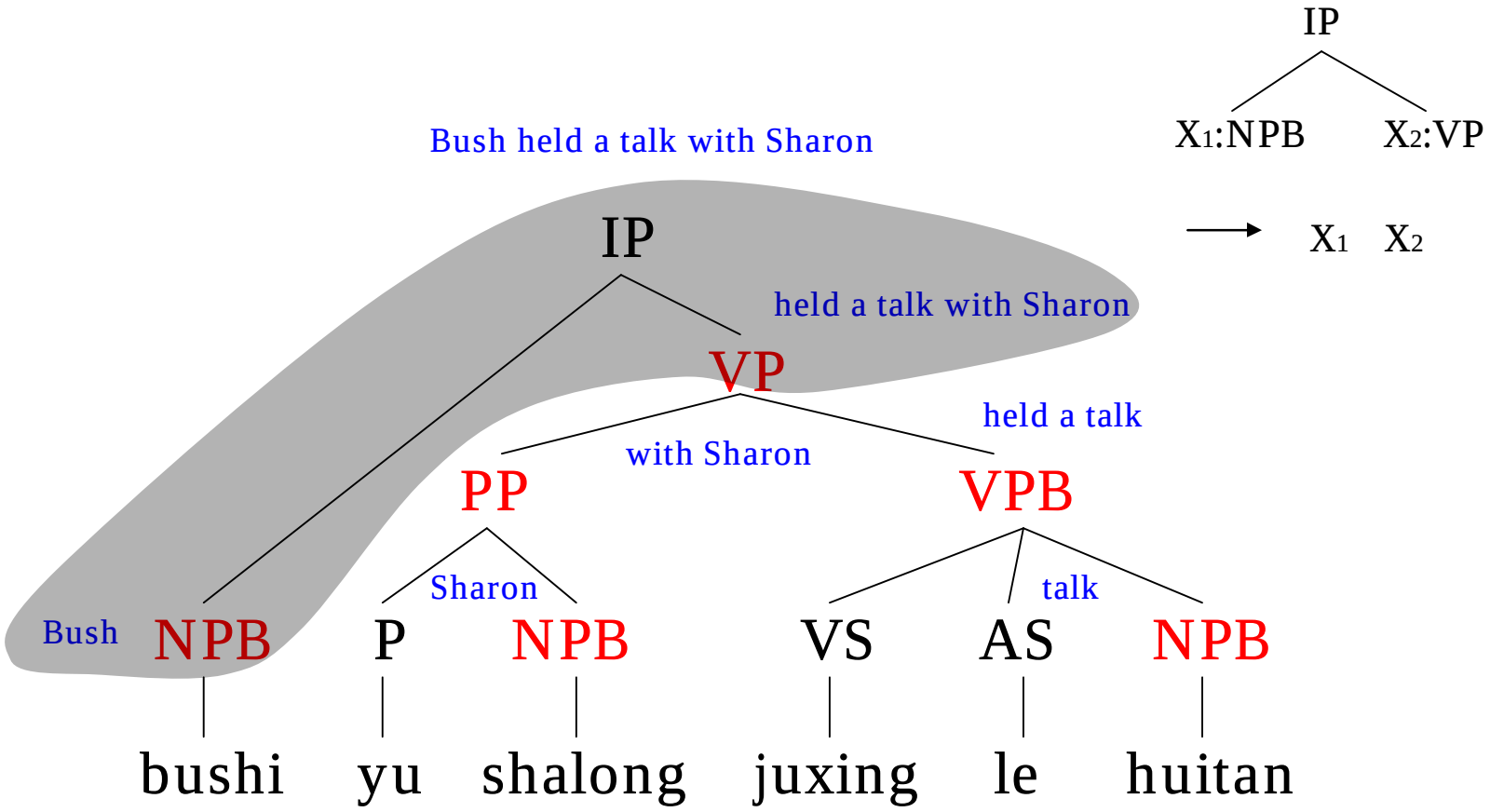
(Liu et al., 2006; Huang et al., 2006)

Bottom-up Decoding



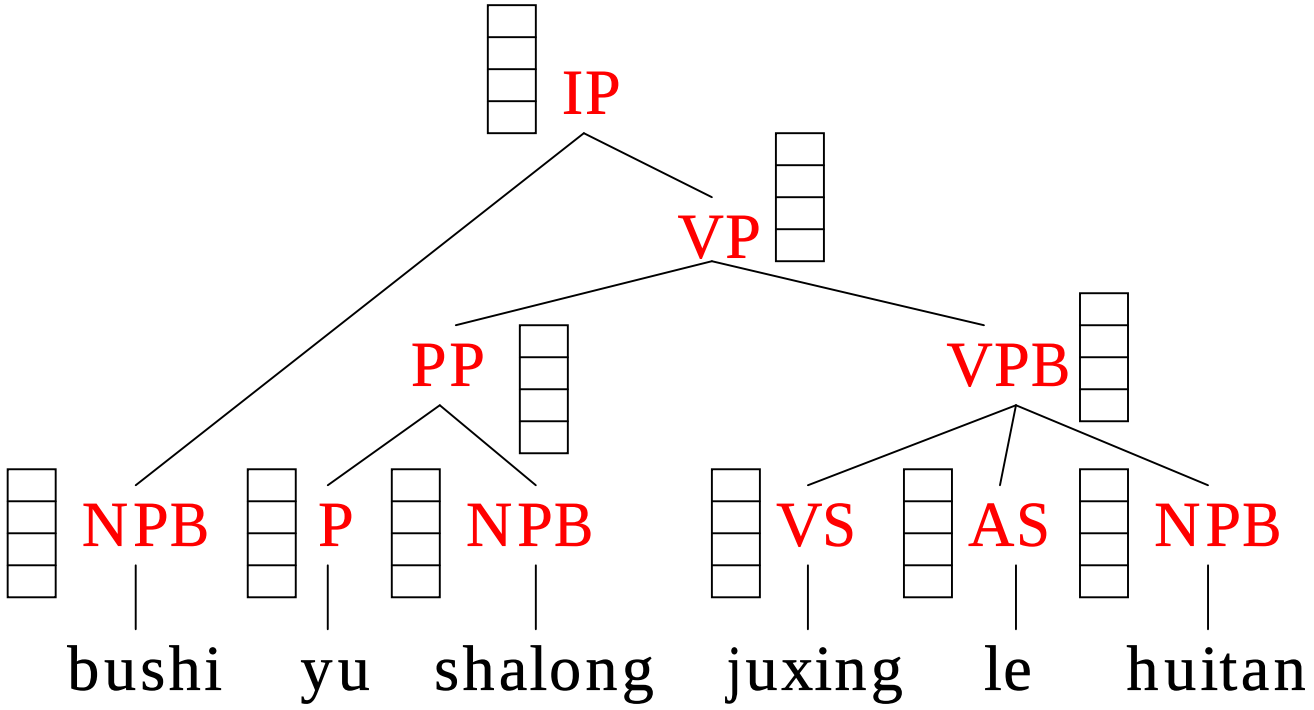
(Liu et al., 2006; Huang et al., 2006)

Bottom-up Decoding



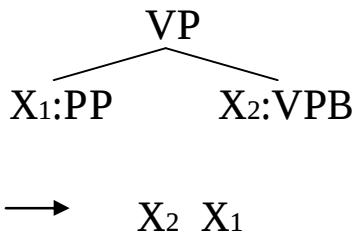
(Liu et al., 2006; Huang et al., 2006)

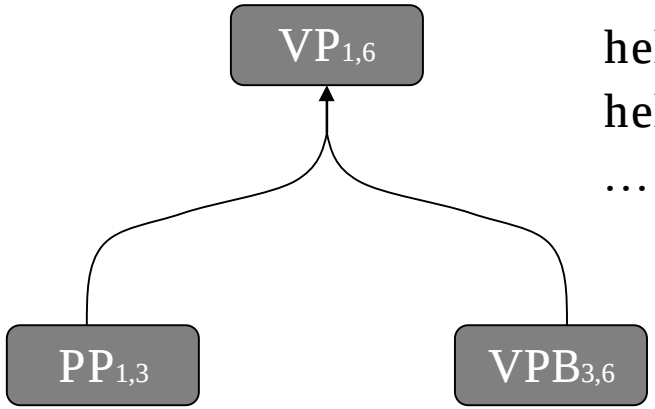
Beam Search



(Liu et al., 2006; Huang et al., 2006)

Exhaustive Search





held a talk with Sharon
held a talk and Sharon
held talks with Sharon
held talks and Sharon
…

with Sharon
and Sharon
Sharon with
Sharon and

held a talk
held talks
hold a talk
hold talks

Update Bigram LM Probability

p1=p(“with”) * p(“Sharon”|”with”) p2=p(“held”) * p(“a”|”held”)*p(“talk”|”a”)

with Sharon

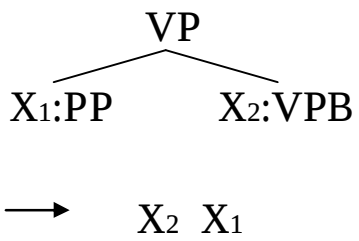
held a talk

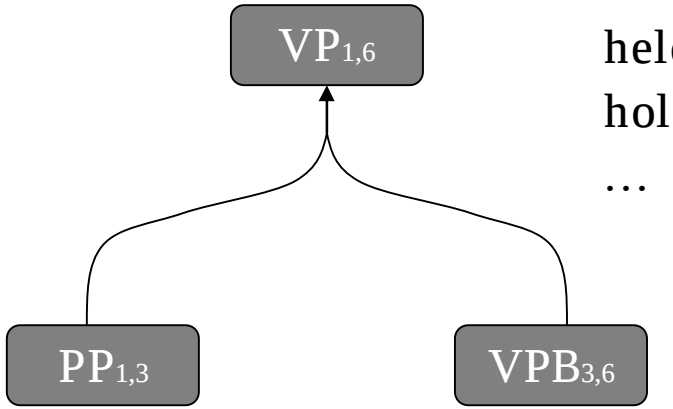
Only boundary words are used to update LM probability!

with Sharon held a talk p1*p2*p(“held”|”Sharon”)/p(“held”)

held a talk with Sharon p1*p2*p(“with”|”talk”)/p(“with”)

Exhaustive Search with a Bigram Language Model



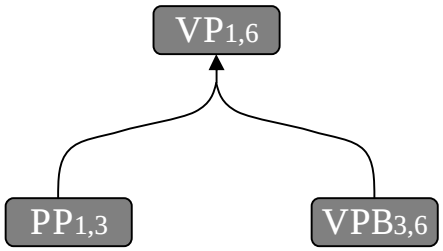


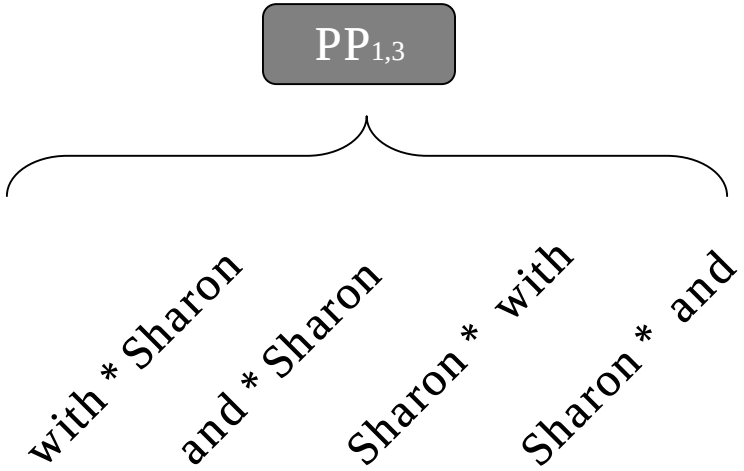
held * Sharon
hold * Sharon
held * with
hold * with
…

with * Sharon
and * Sharon
Sharon * with
Sharon * and

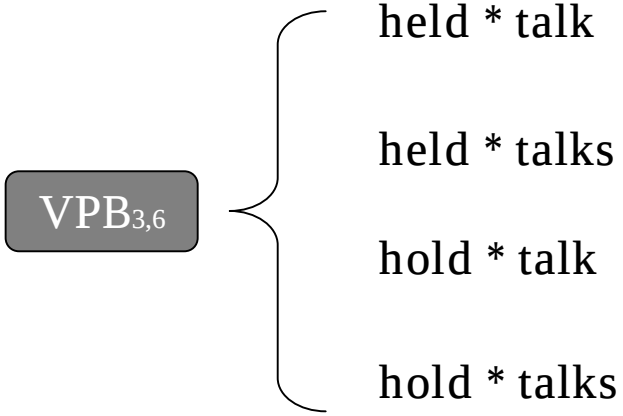
held * talk
held * talks
hold * talk
hold * talks

Monotonicity





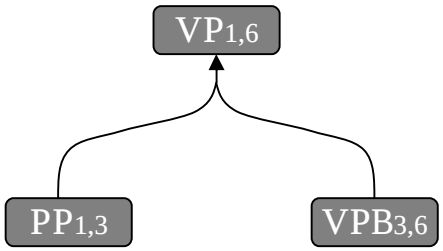
monotonic



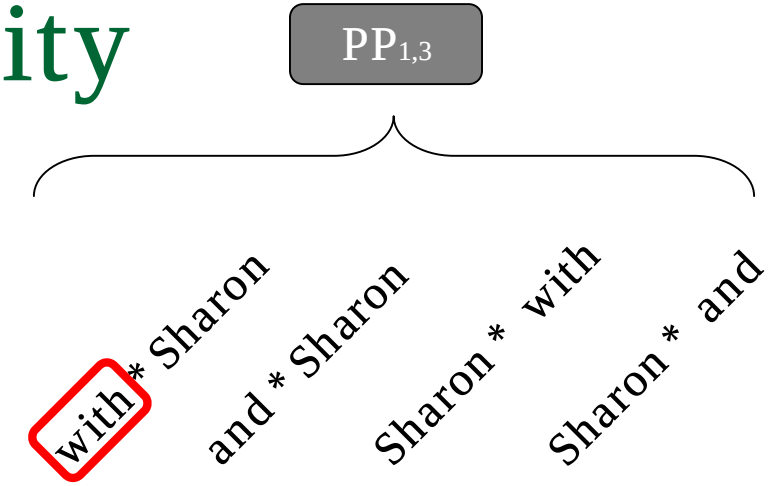
	1.0	3.0	4.0	6.5
1.0	2.0	4.0	5.0	7.5
1.1	2.1	4.1	5.1	7.6
2.0	3.0	5.0	6.0	8.5
3.5	4.5	6.5	7.5	10.0

(Huang and Chiang, 2005, 2007; Chiang, 2007)

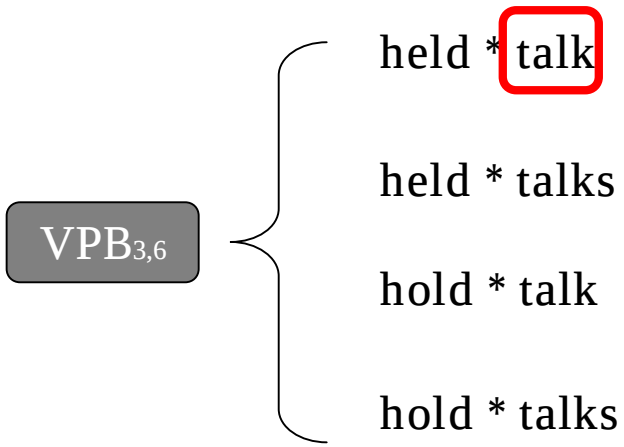
Non-Monotonicity



log(p(***with*** | ***talk***))-
log(p(***with***))



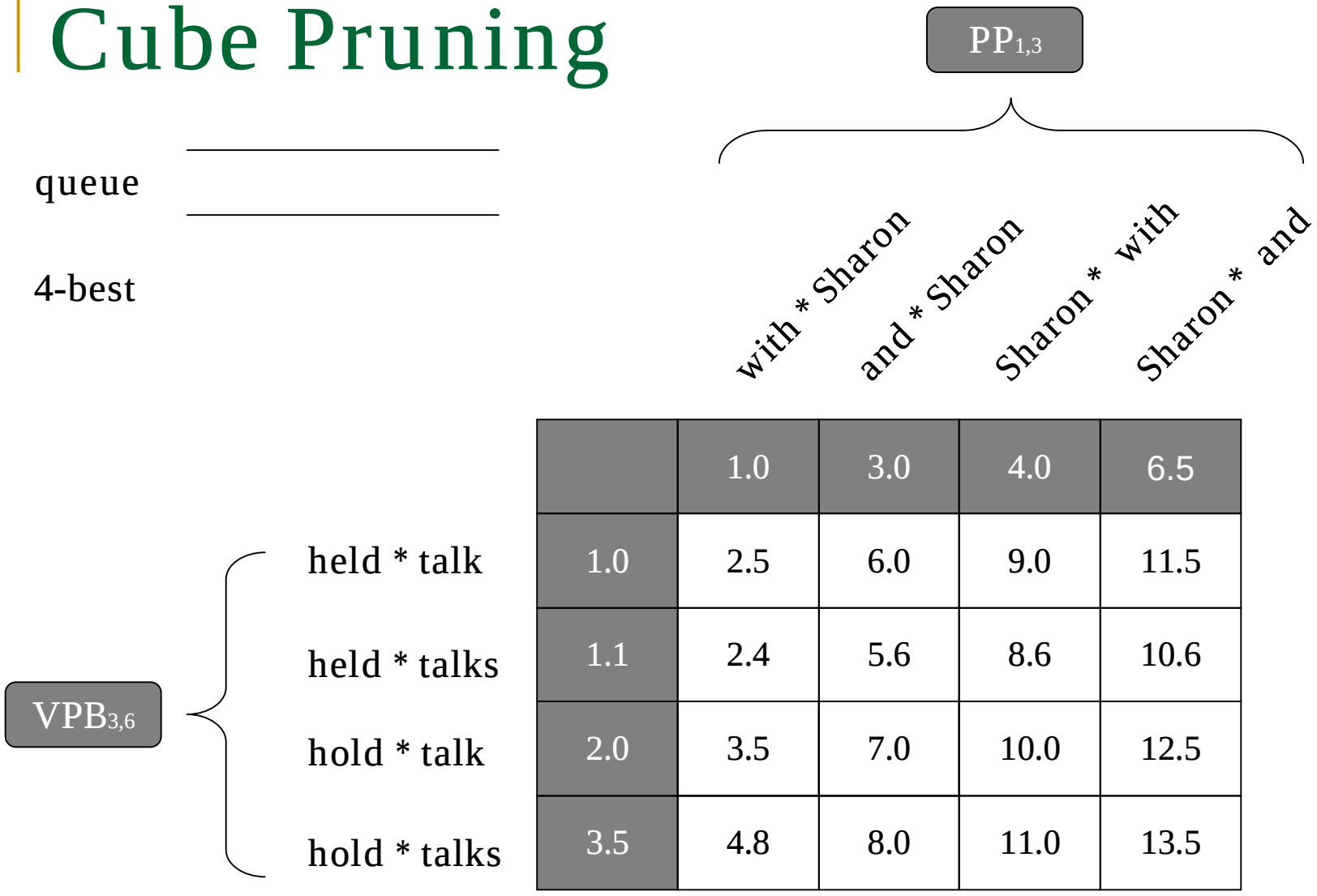
LM introduces non-monotonicity



	1.0	3.0	4.0	6.5
1.0	2.0 + 0.5	4.0 + 2.0	5.0 + 4.0	7.5 + 4.0
1.1	2.1 + 0.3	4.1 + 1.5	5.1 + 3.5	7.6 + 3.0
2.0	3.0 + 0.5	5.0 + 2.0	6.0 + 4.0	8.5 + 4.0
3.5	4.5 + 0.3	6.5 + 1.5	7.5 + 3.5	10 + 3.5

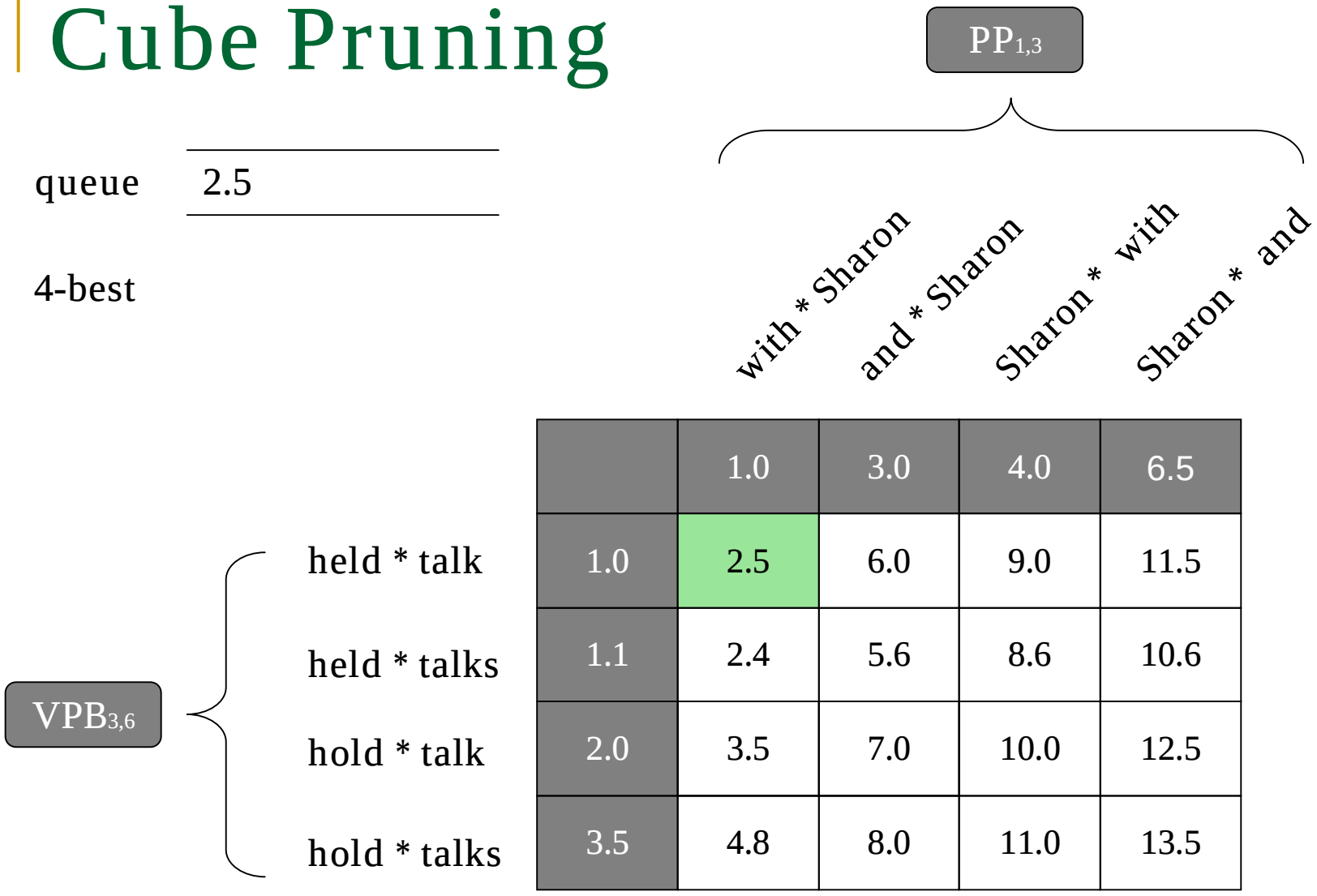
(Huang and Chiang, 2005, 2007; Chiang, 2007)

Cube Pruning



(Huang and Chiang, 2005, 2007; Chiang, 2007)

Cube Pruning

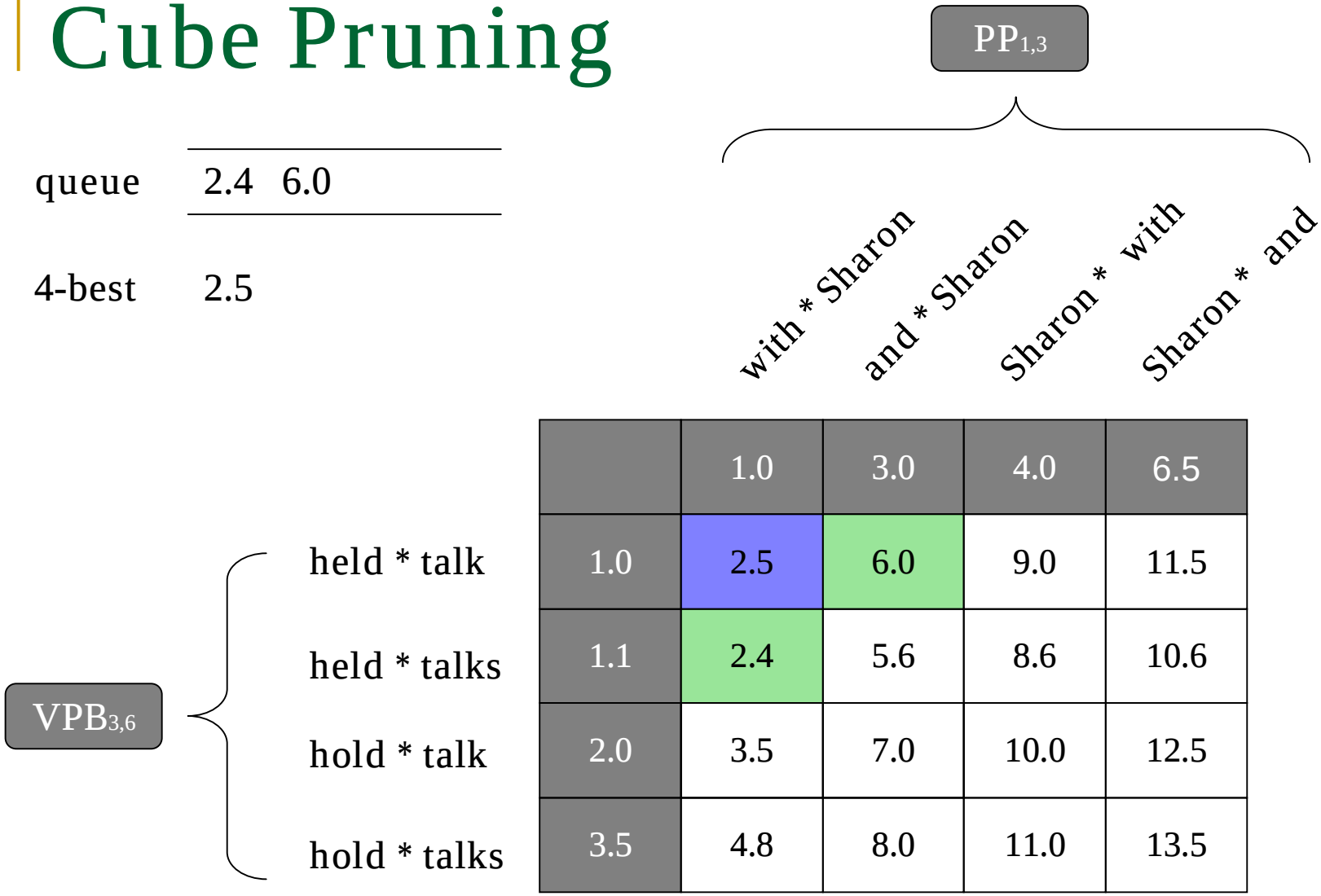


(Huang and Chiang, 2005, 2007; Chiang, 2007)

Cube Pruning

queue 2.4 6.0

4-best 2.5

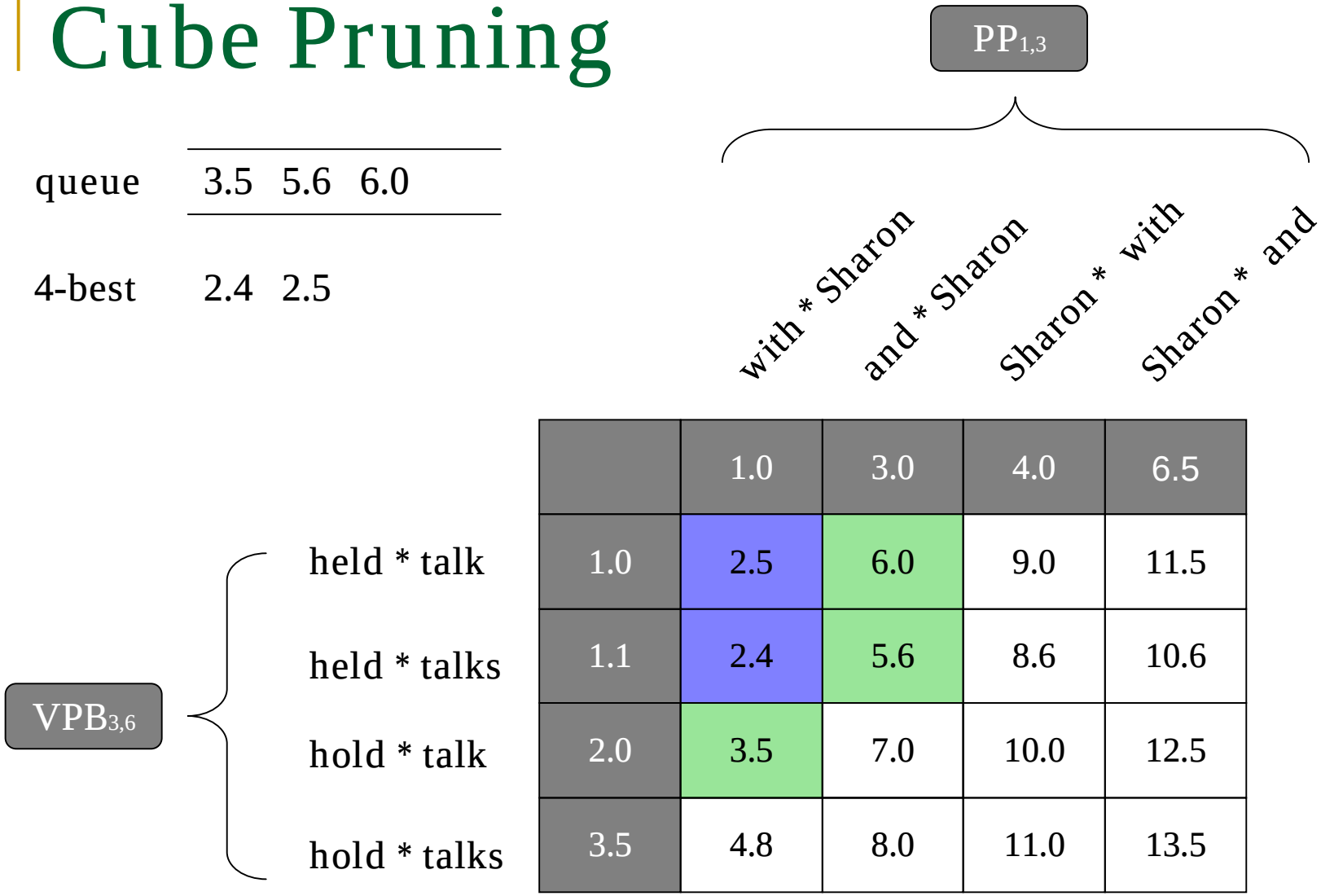


(Huang and Chiang, 2005, 2007; Chiang, 2007)

Cube Pruning

queue 3.5 5.6 6.0

4-best 2.4 2.5

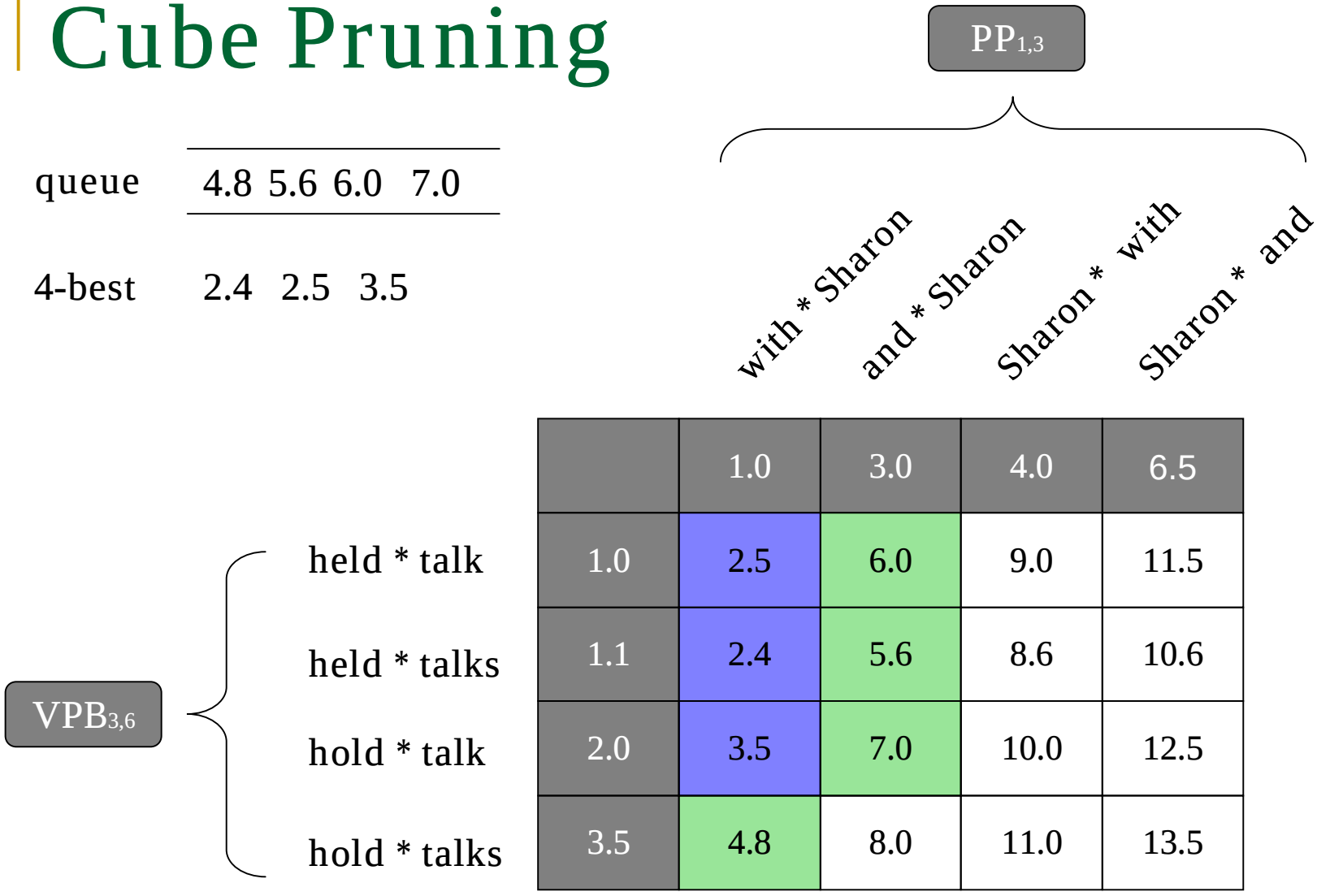


(Huang and Chiang, 2005, 2007; Chiang, 2007)

Cube Pruning

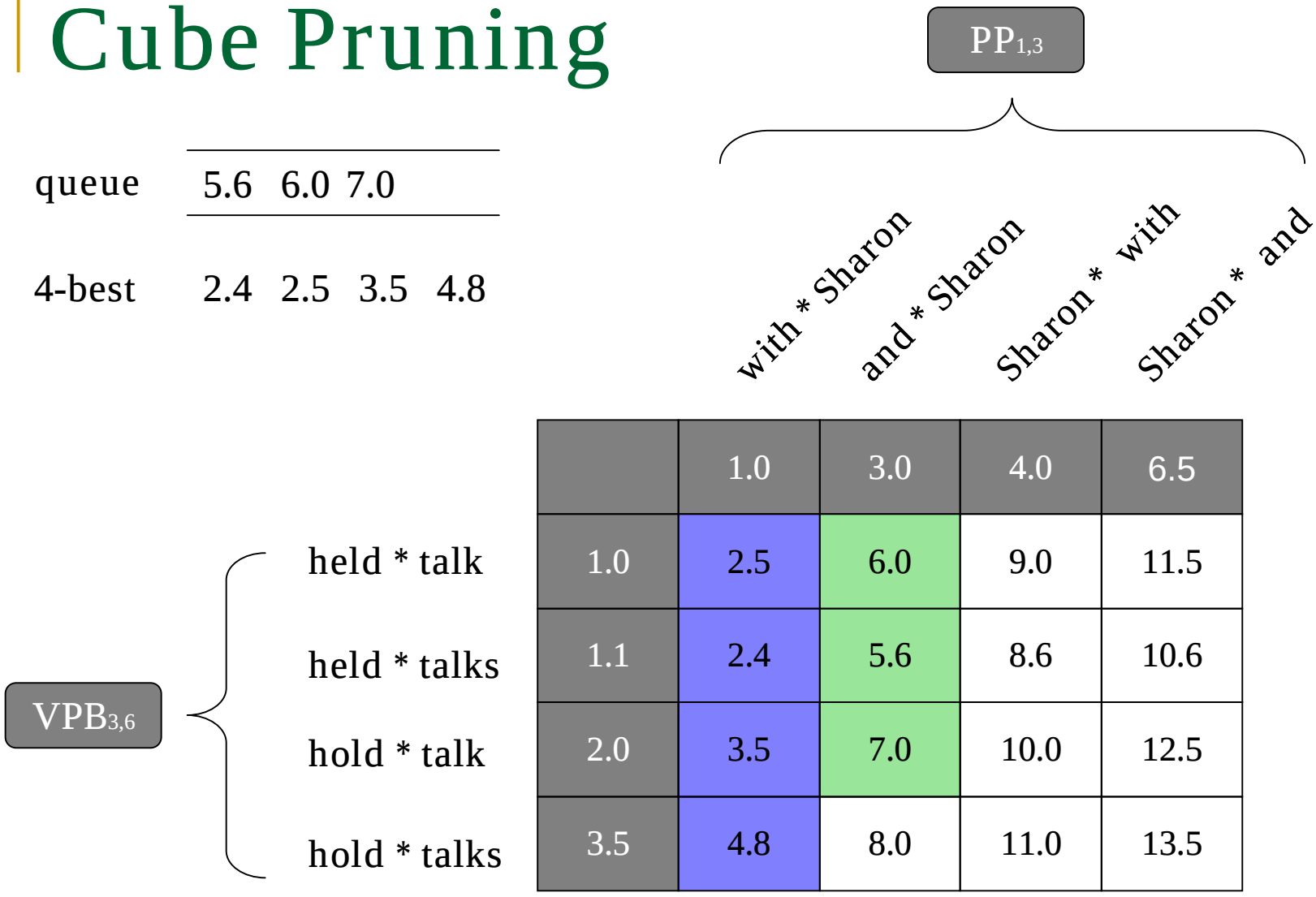
queue 4.8 5.6 6.0 7.0

4-best 2.4 2.5 3.5



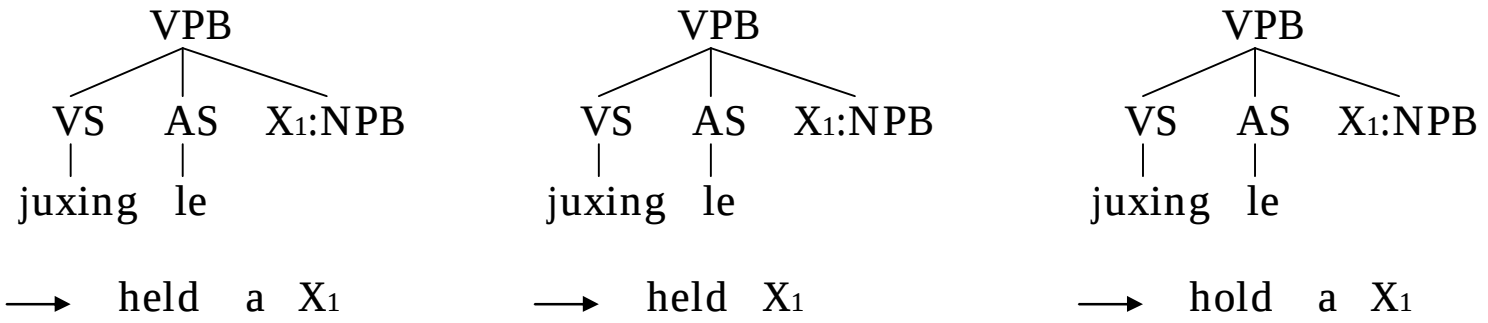
(Huang and Chiang, 2005, 2007; Chiang, 2007)

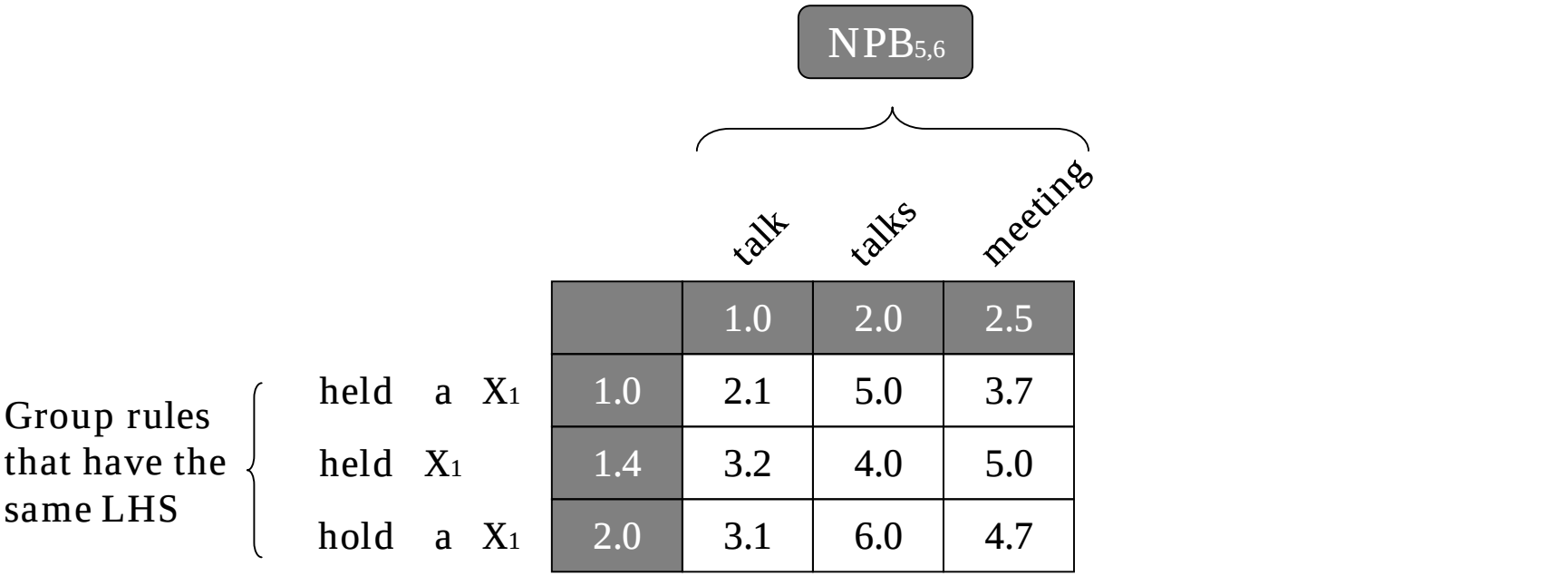
Cube Pruning



(Huang and Chiang, 2005, 2007; Chiang, 2007)

Cube Pruning within Rule Group

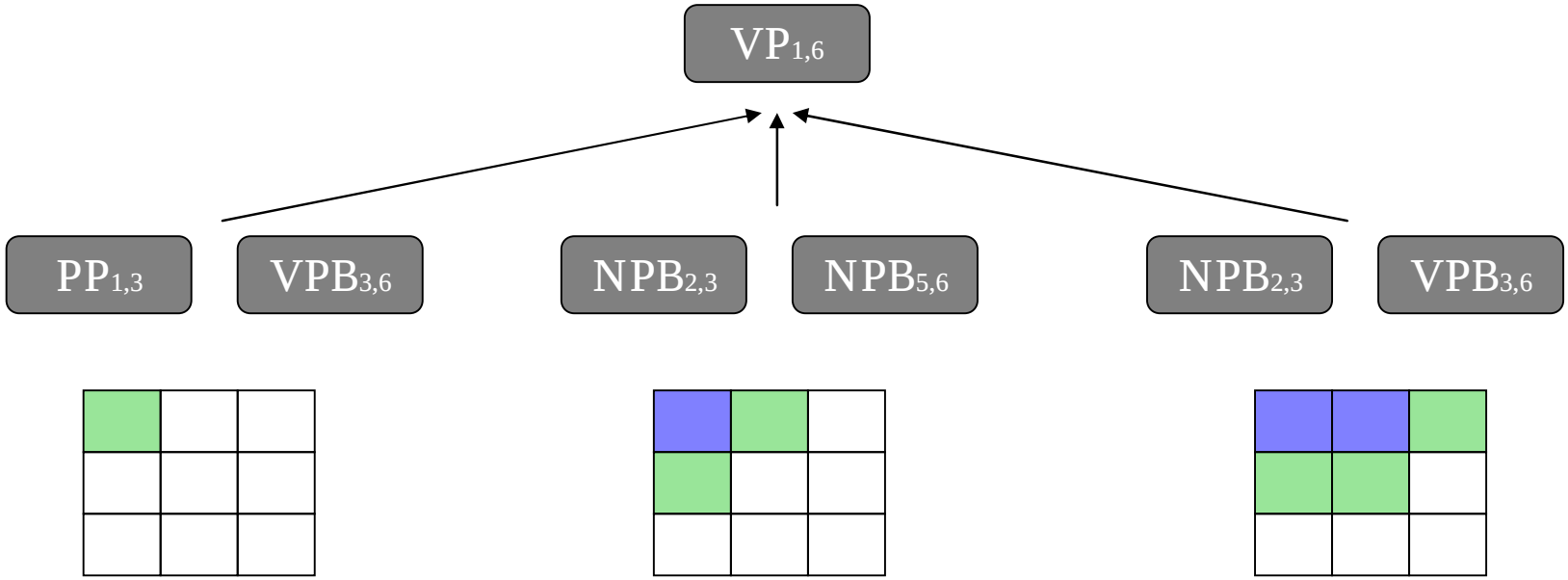




	1.0	2.0	2.5
1.0	2.1	5.0	3.7
1.4	3.2	4.0	5.0
2.0	3.1	6.0	4.7

(Huang and Chiang, 2005, 2007; Chiang, 2007)

Cube Pruning within Node



process all rules simultaneously!

significant savings of computation

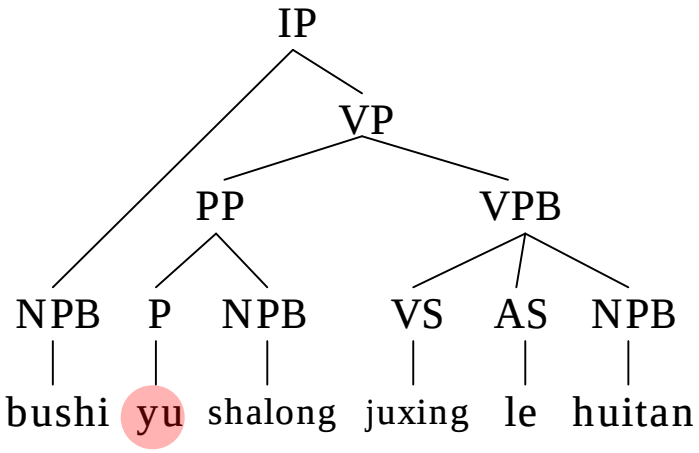
(Huang and Chiang, 2005, 2007; Chiang, 2007)

Outline

- *Part 1: Tree-based Translation*
  - *Overview and Motivation*
  - *Tree-to-String Model and Decoding*
  - *Tree-to-String Rule Extraction*
  - *Language Model-Integrated Decoding: Cube Pruning*
- Part 2: Forest-based Translation
  - Packed Forest
  - Forest-based Decoding
  - Forest-based Rule Extraction
- Part 3: Extensions
  - Tree-to-Tree Translation
  - Tree Sequence-based Translation
  - Joint Parsing and Translation
- Part 4: Conclusion

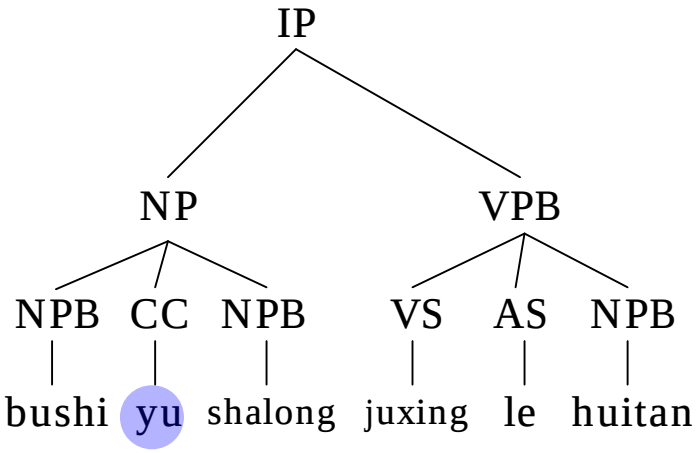
Syntactic Ambiguity

It is important to choose a correct tree for producing a good translation!



with

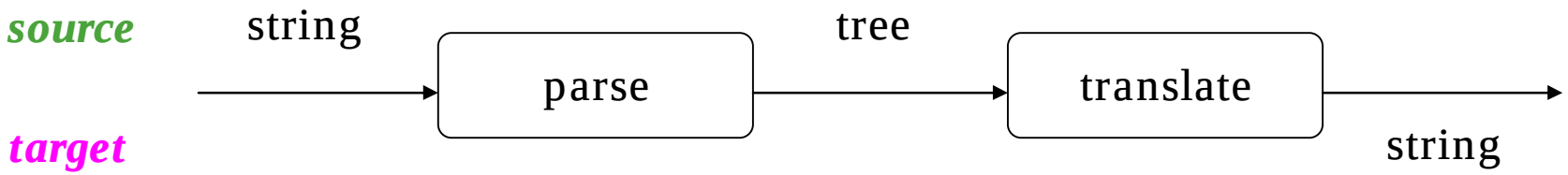
``Bush held a talk with Sharon''



and

``Bush and Sharon held a talk''

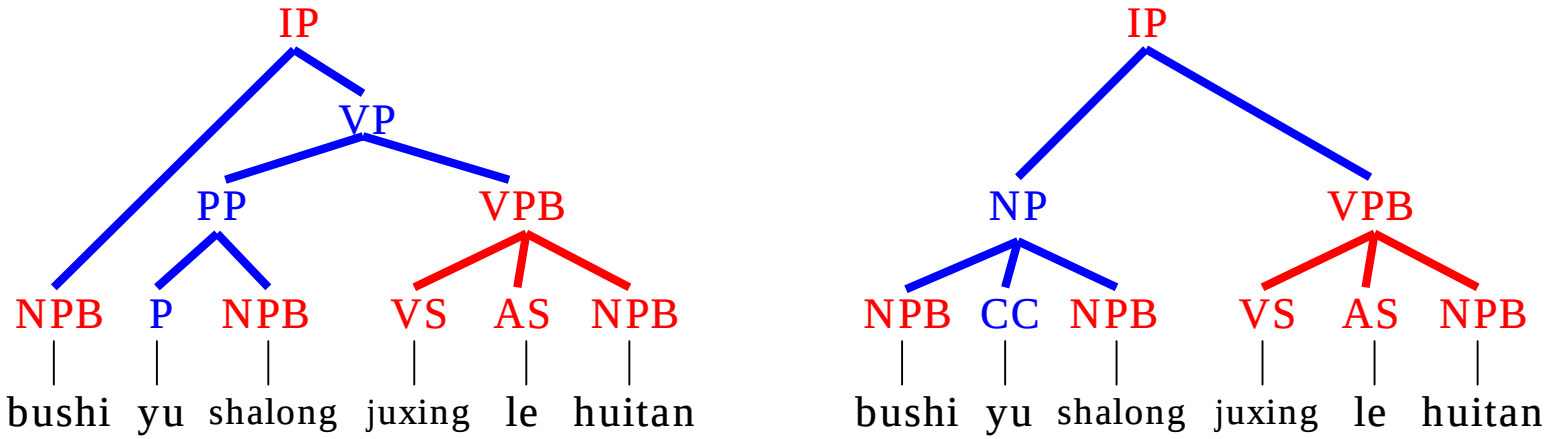
Parsing Mistake Propagation



parsing mistakes potentially introduce translation mistakes!

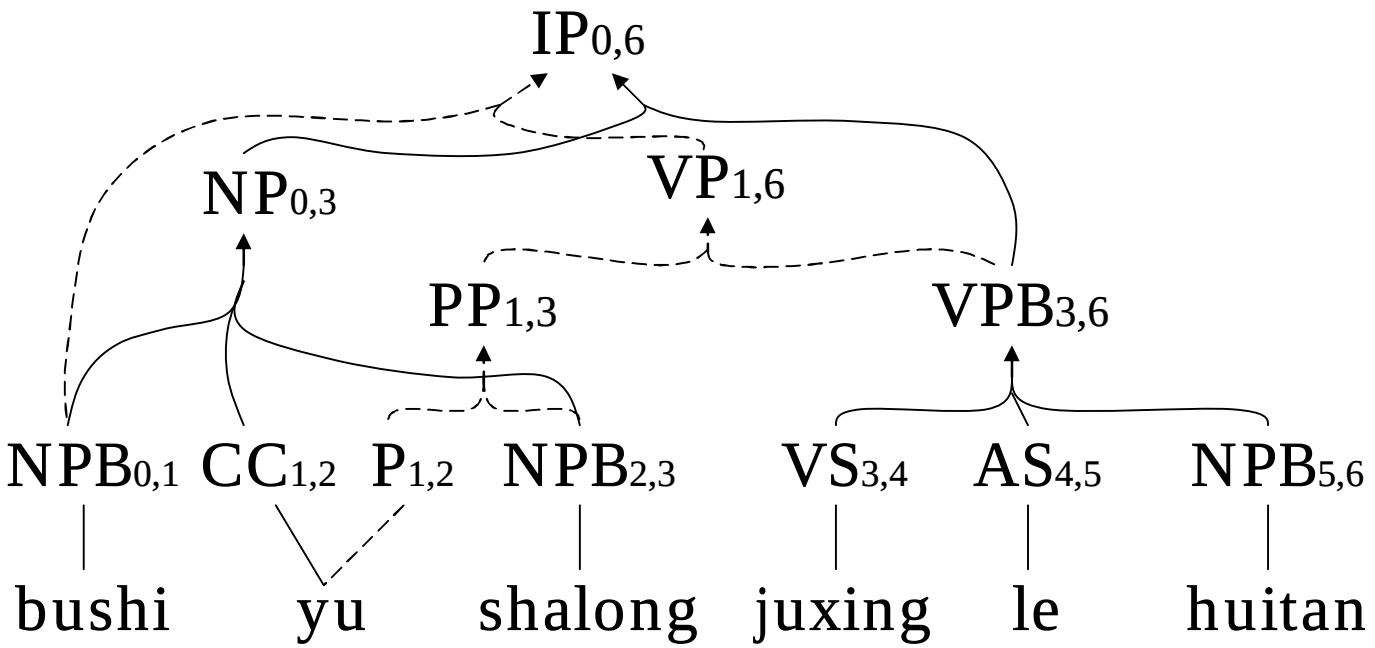
(Quirk and Corston-Oliver, 2006)

1-best Trees => ***n***-best Trees?



Very few variations among the ***n***-best trees!

Packed Forest

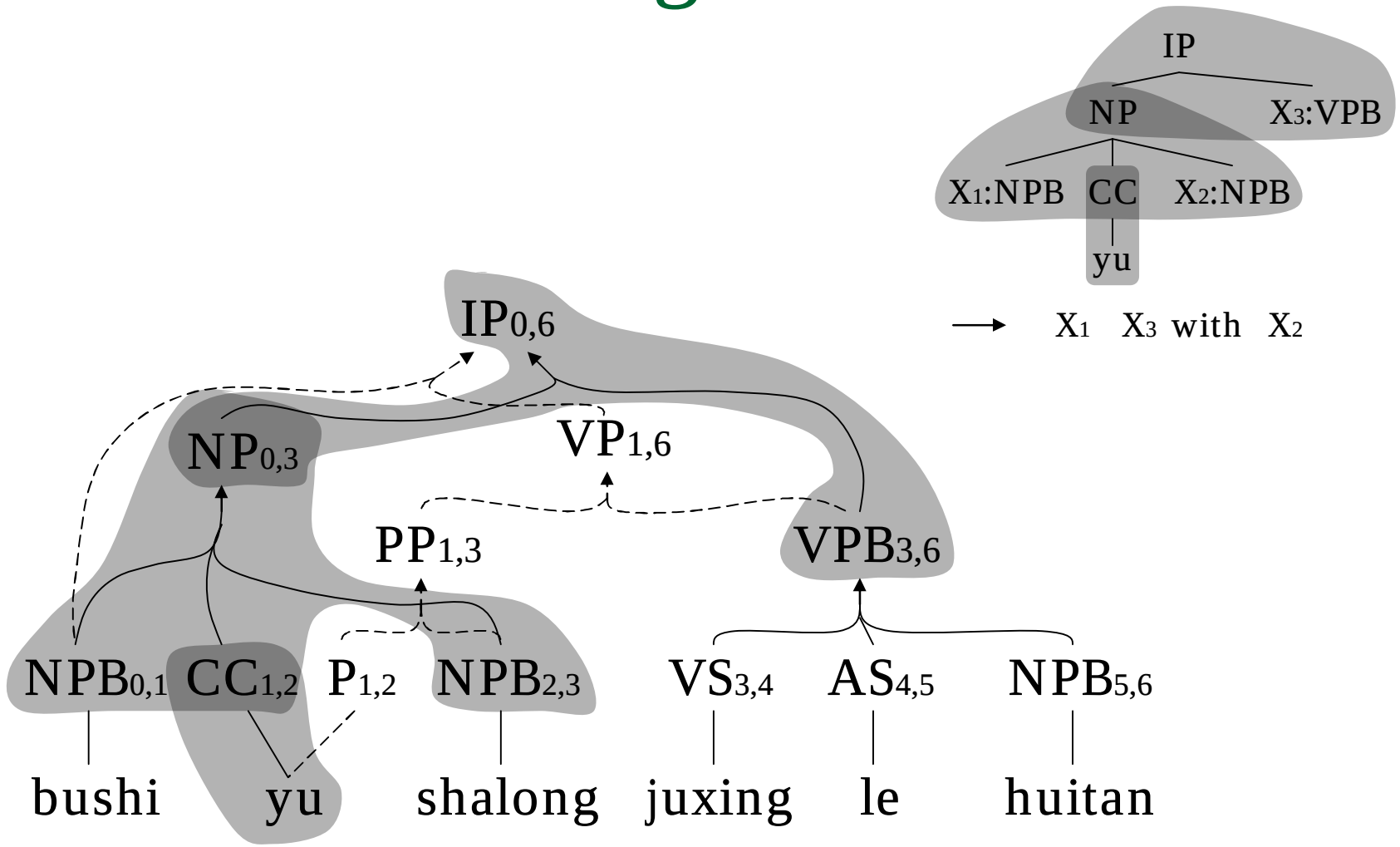


(Billot and Lang, 1989; Klein and Manning, 2001; Huang and Chiang, 2005)

Outline

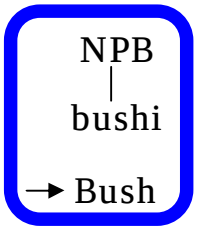
- *Part 1: Tree-based Translation*
  - *Overview and Motivation*
  - *Tree-to-String Model and Decoding*
  - *Tree-to-String Rule Extraction*
  - *Language Model-Integrated Decoding: Cube Pruning*
- Part 2: Forest-based Translation
  - *Packed Forest*
  - Forest-based Decoding
  - Forest-based Rule Extraction
- Part 3: Extensions
  - Tree-to-Tree Translation
  - Tree Sequence-based Translation
  - Joint Parsing and Translation
- Part 4: Conclusion

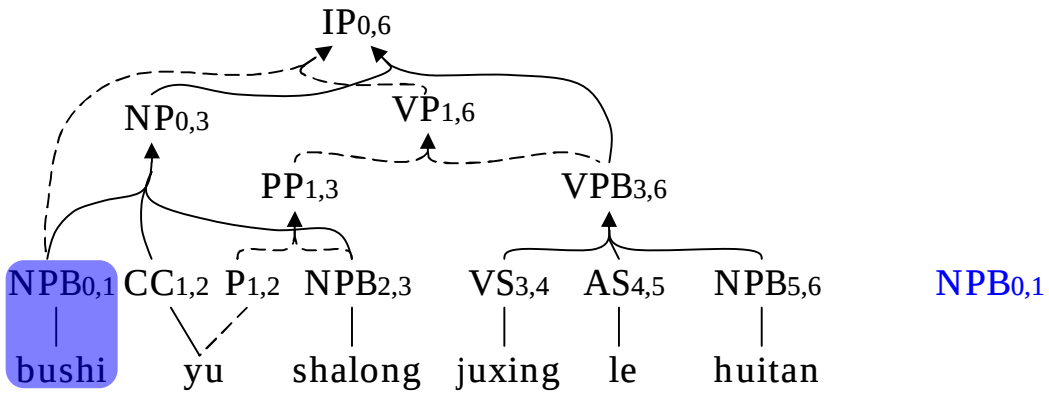
Pattern Matching on Forest



(Mi et al., 2008)

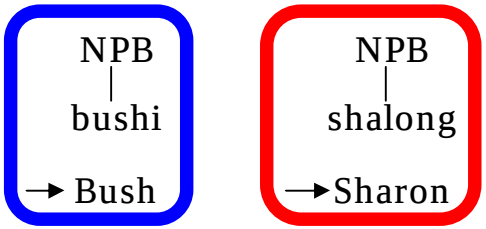
Translation Forest

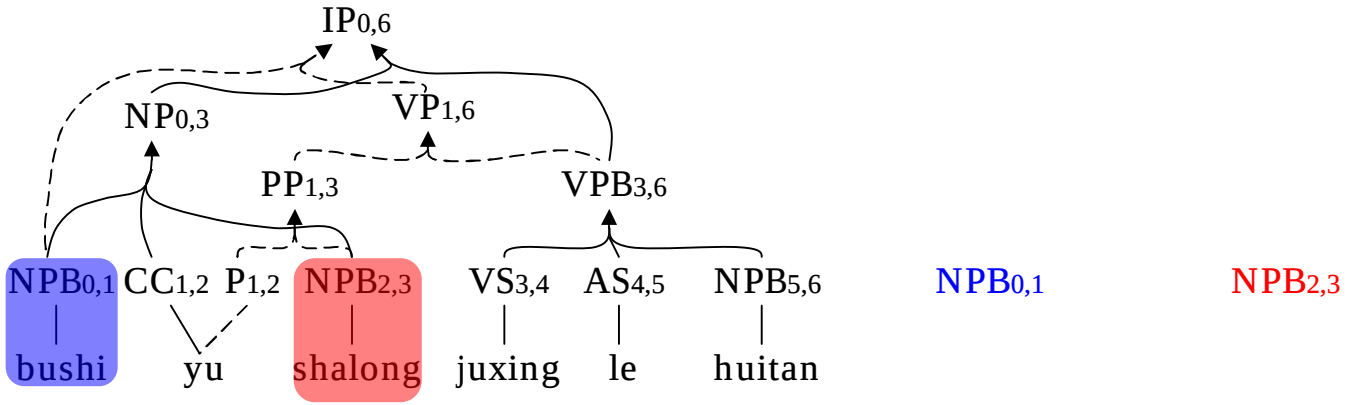




(Mi et al., 2008)

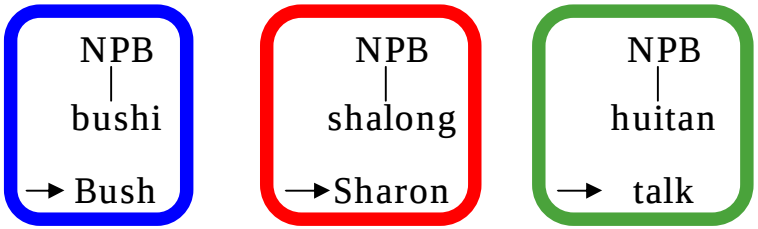
Translation Forest

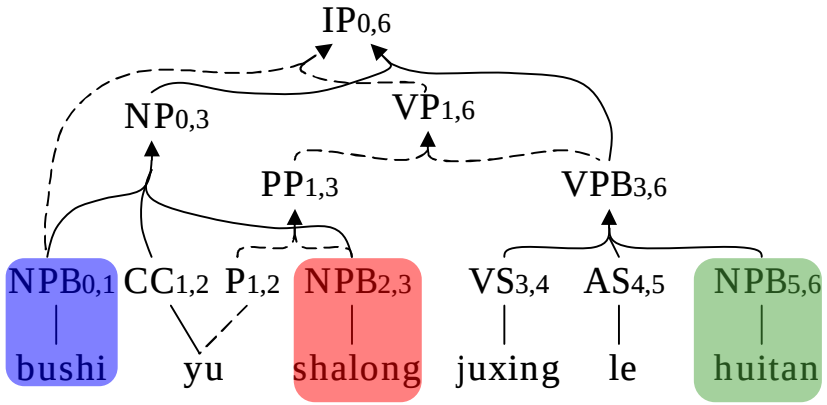




(Mi et al., 2008)

Translation Forest

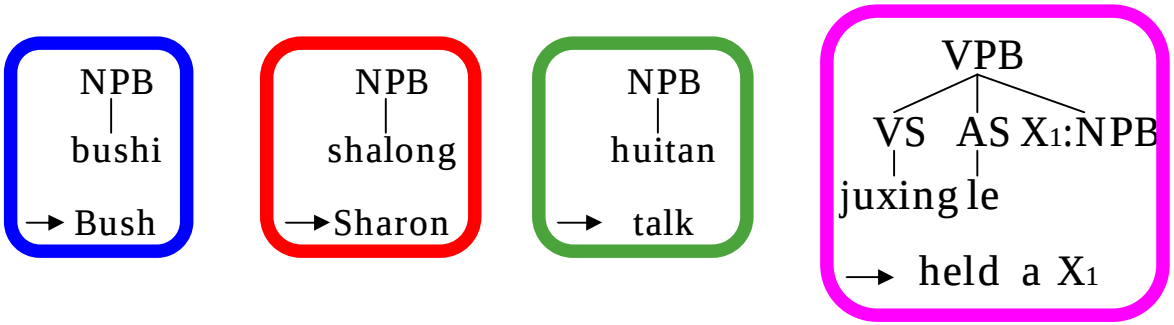


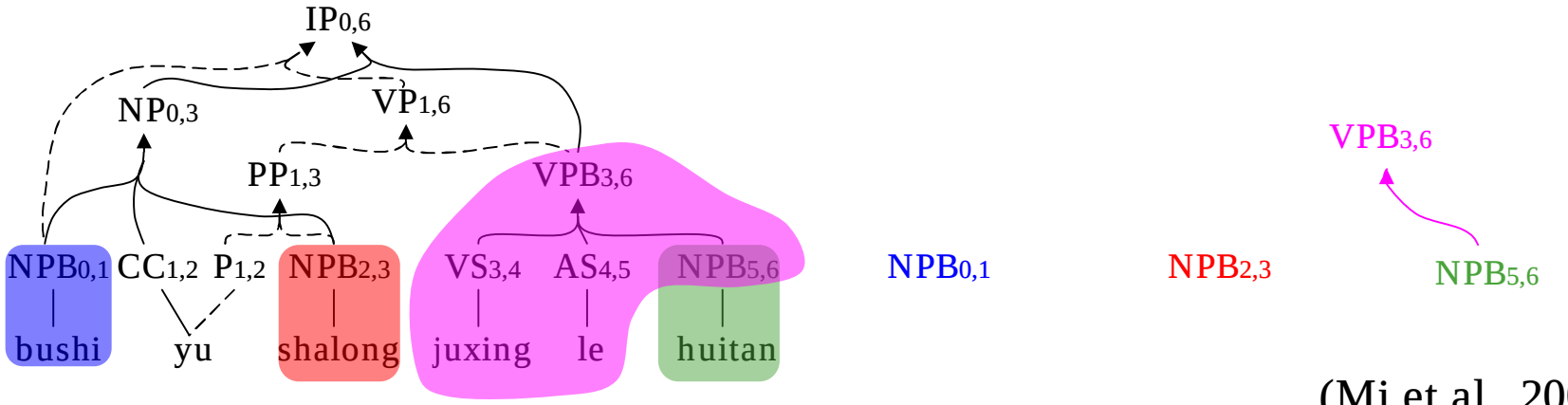


NPB0,1 NPB2,3 NPB5,6

(Mi et al., 2008)

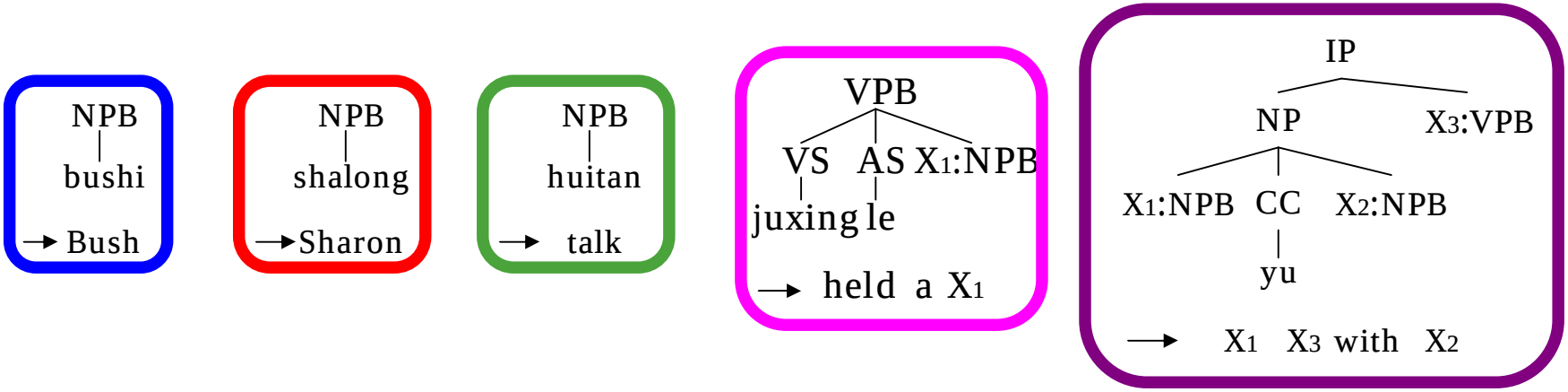
Translation Forest

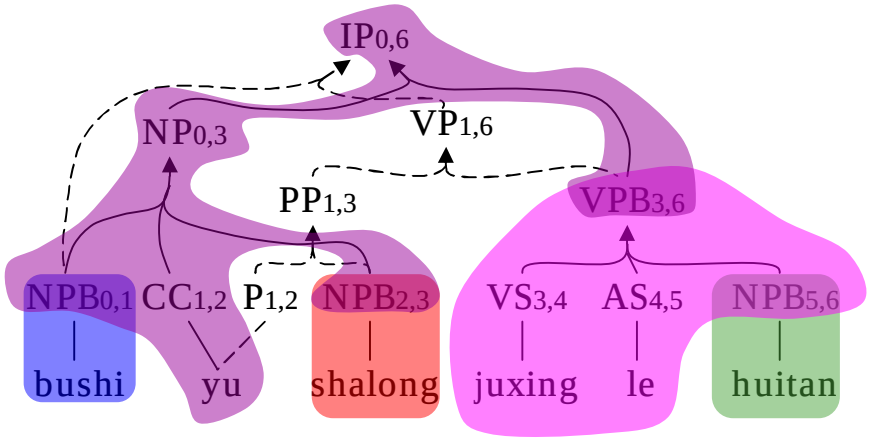


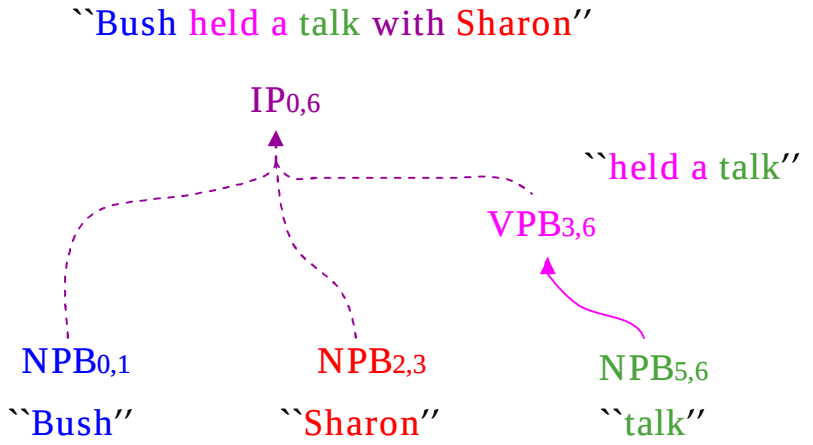


(Mi et al., 2008)

Translation Forest

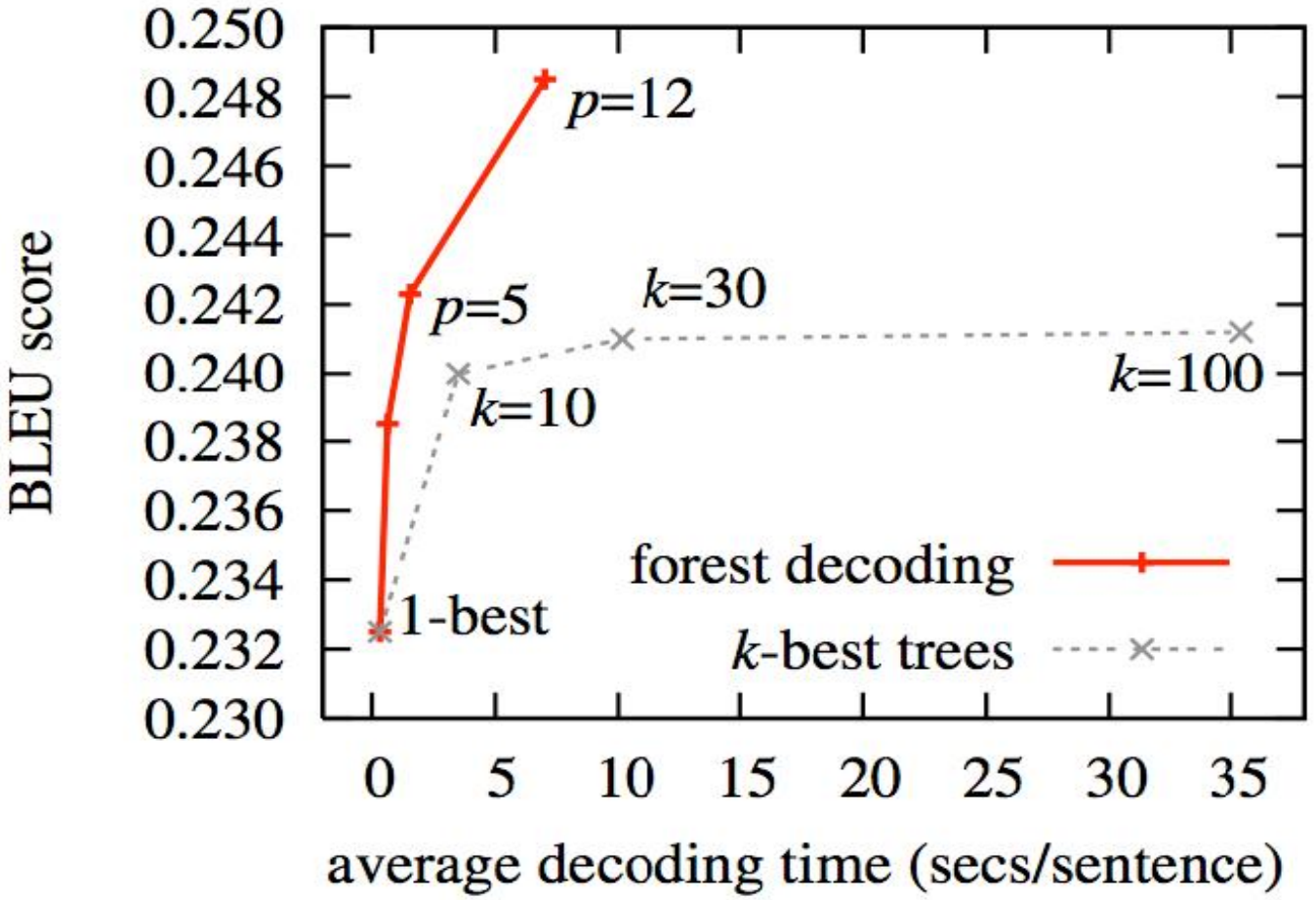






(Mi et al., 2008)

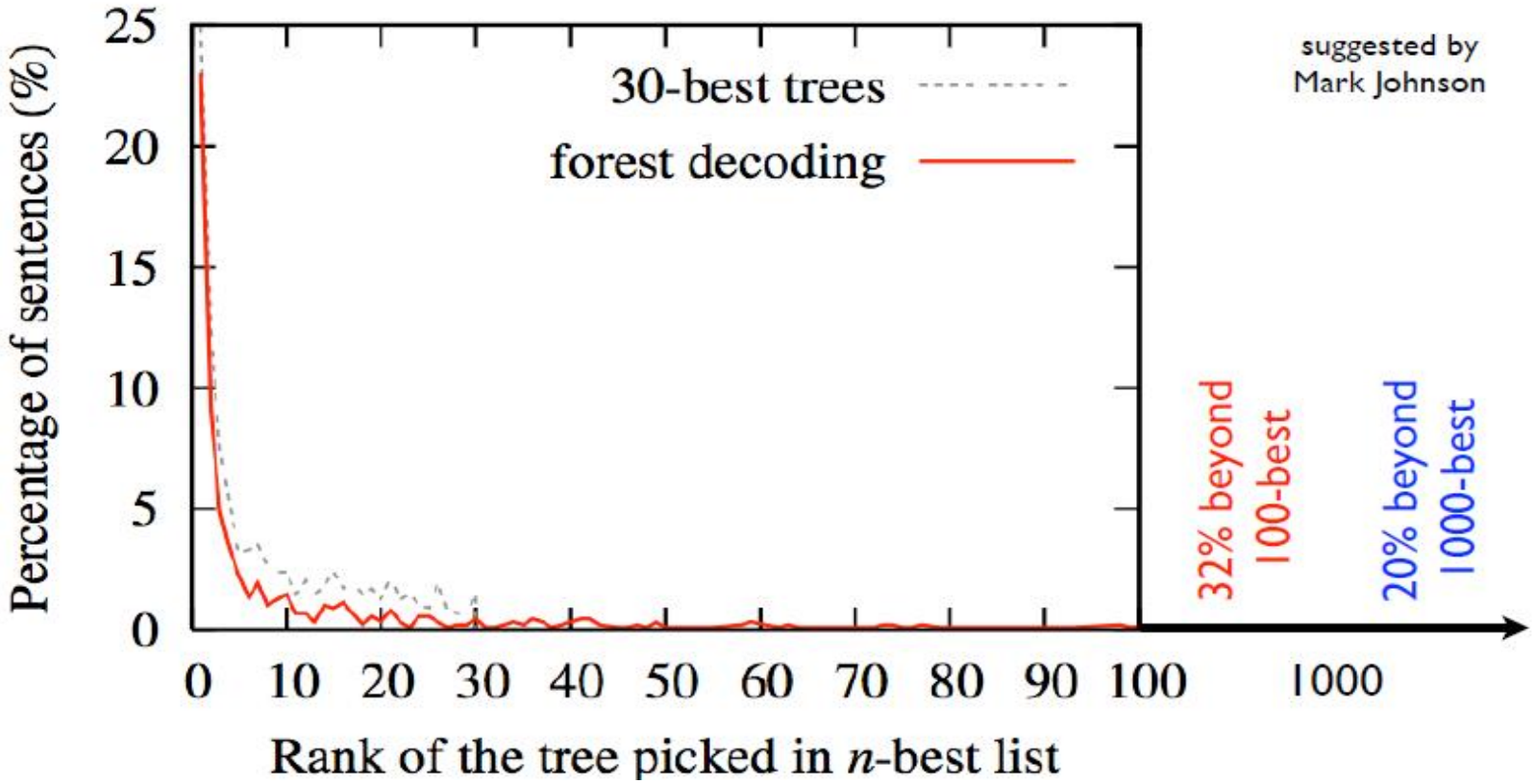
N-best Trees Vs. Forest



(Mi et al., 2008)

Forest as Virtual ∞-best list

- How often is the *i*th-best tree picked by the decoder?



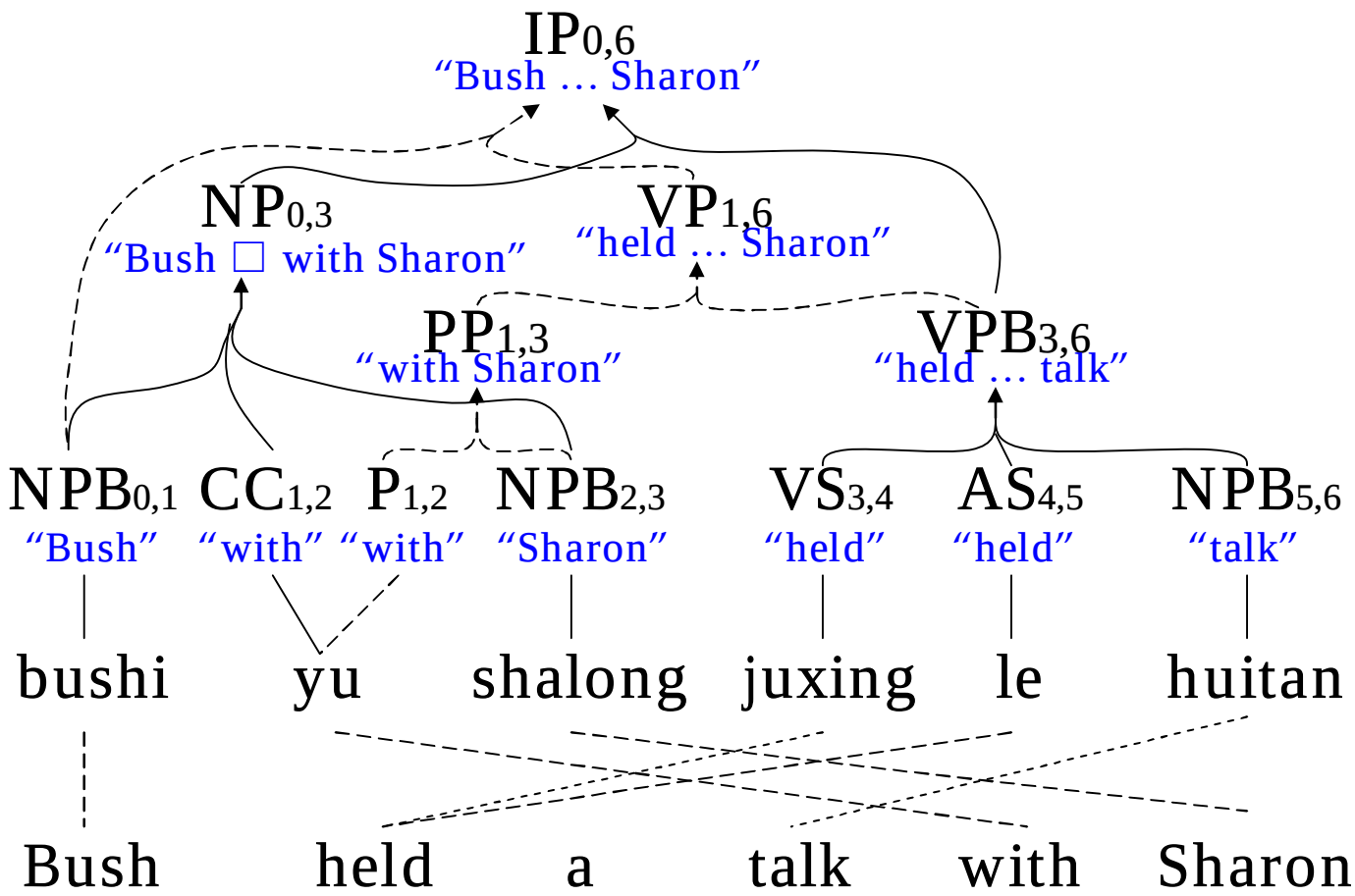
(Mi et al., 2008)

Outline

- *Part 1: Tree-based Translation*
  - *Overview and Motivation*
  - *Tree-to-String Model and Decoding*
  - *Tree-to-String Rule Extraction*
  - *Language Model-Integrated Decoding: Cube Pruning*
- Part 2: Forest-based Translation
  - *Packed Forest*
  - *Forest-based Decoding*
  - Forest-based Rule Extraction
- Part 3: Extensions
  - Tree-to-Tree Translation
  - Tree Sequence-based Translation
  - Joint Parsing and Translation
- Part 4: Conclusion

Forest-based Rule Extraction

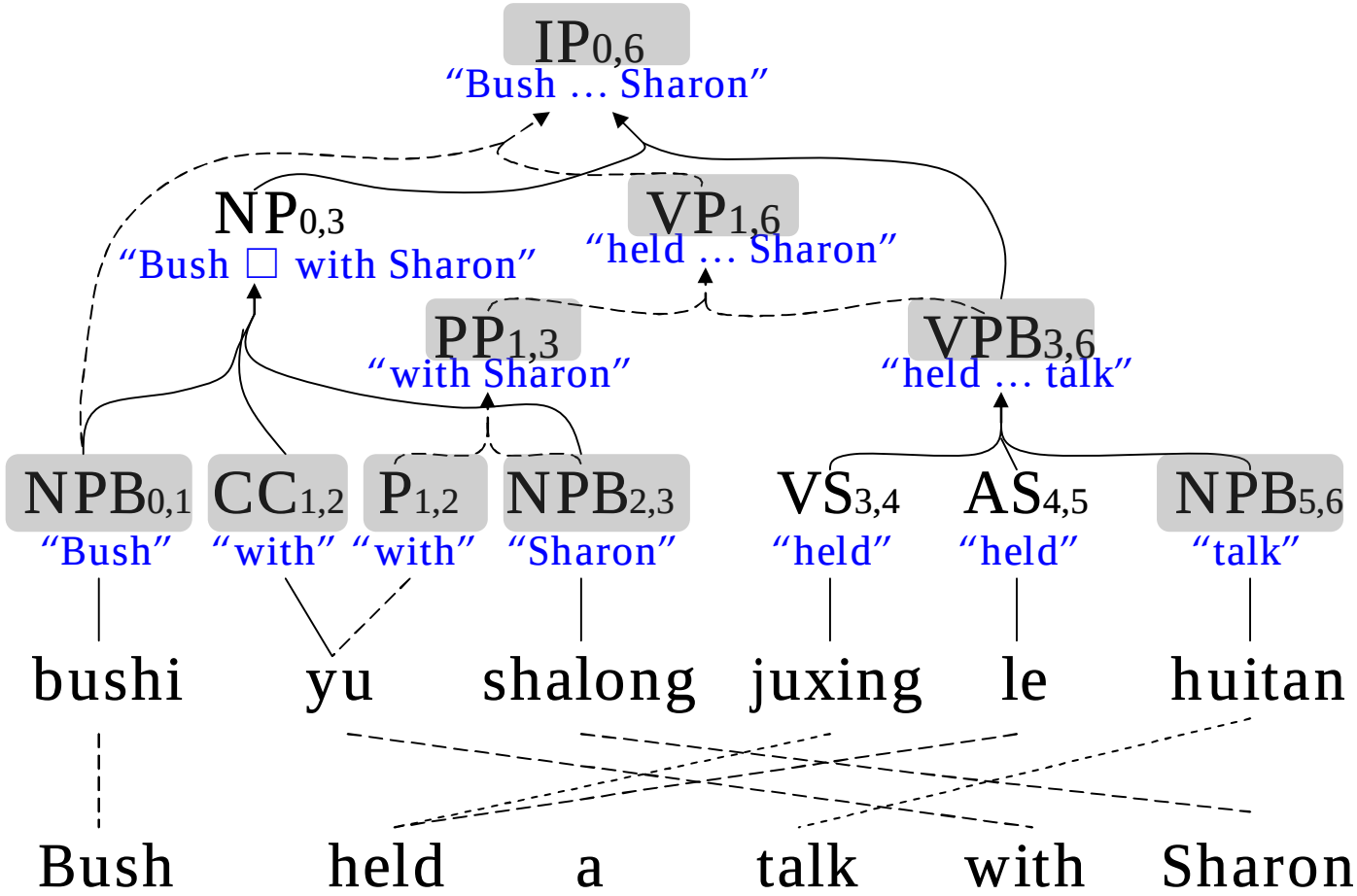
- Compute target spans



(Mi and Huang, 2008)

Forest-based Rule Extraction

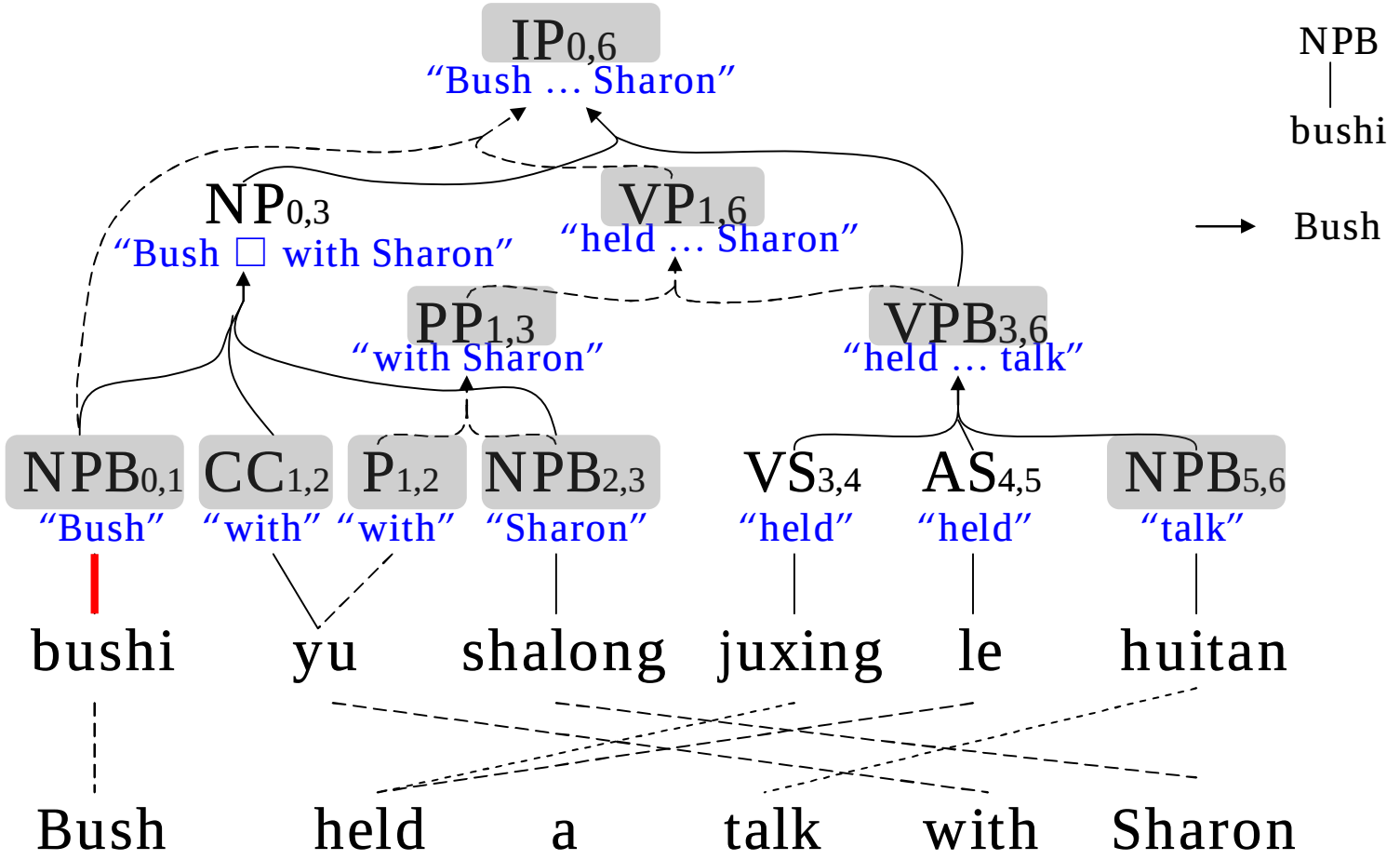
- Compute admissible nodes



(Mi and Huang, 2008)

Forest-based Rule Extraction

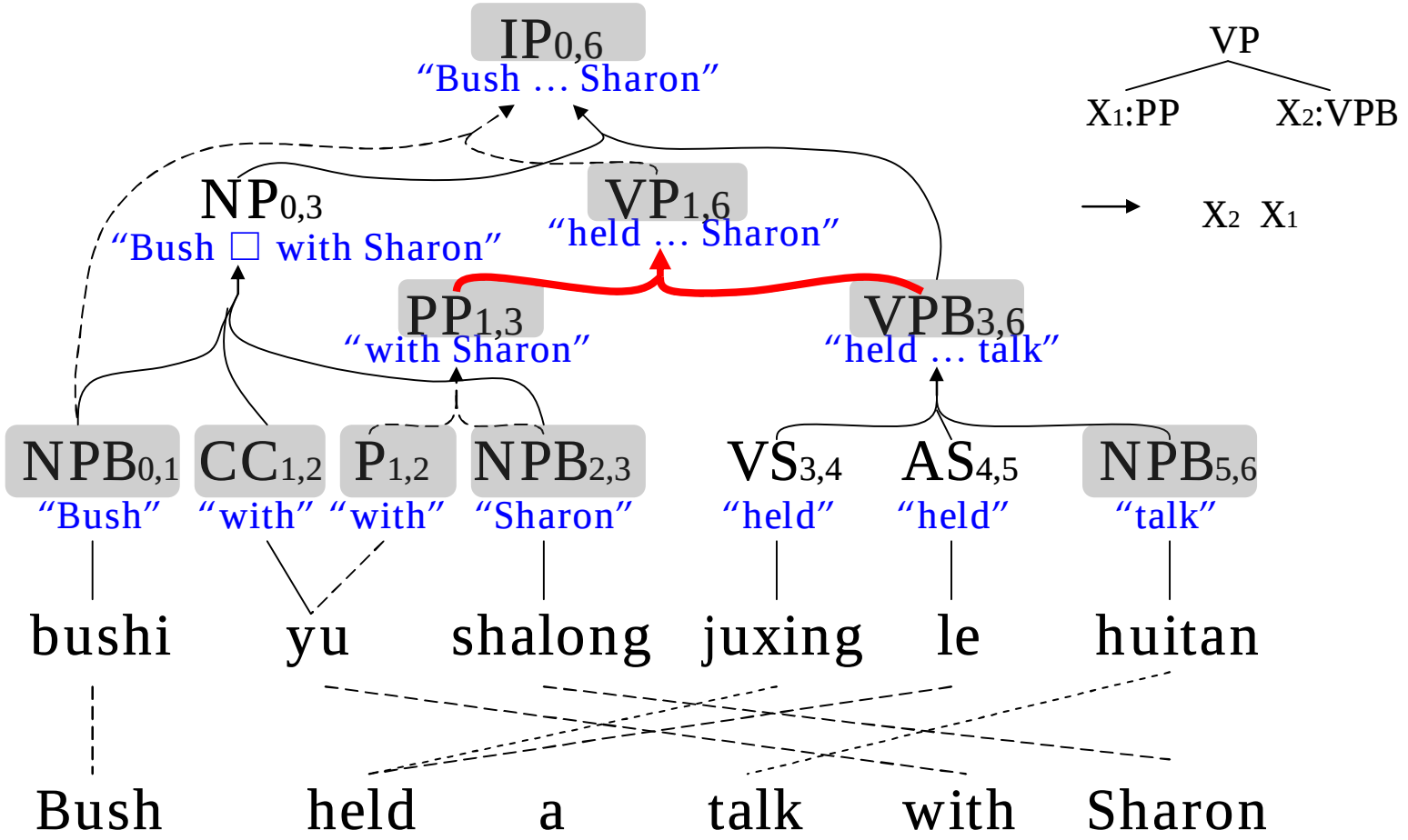
- Extract minimal rules



(Mi and Huang, 2008)

Forest-based Rule Extraction

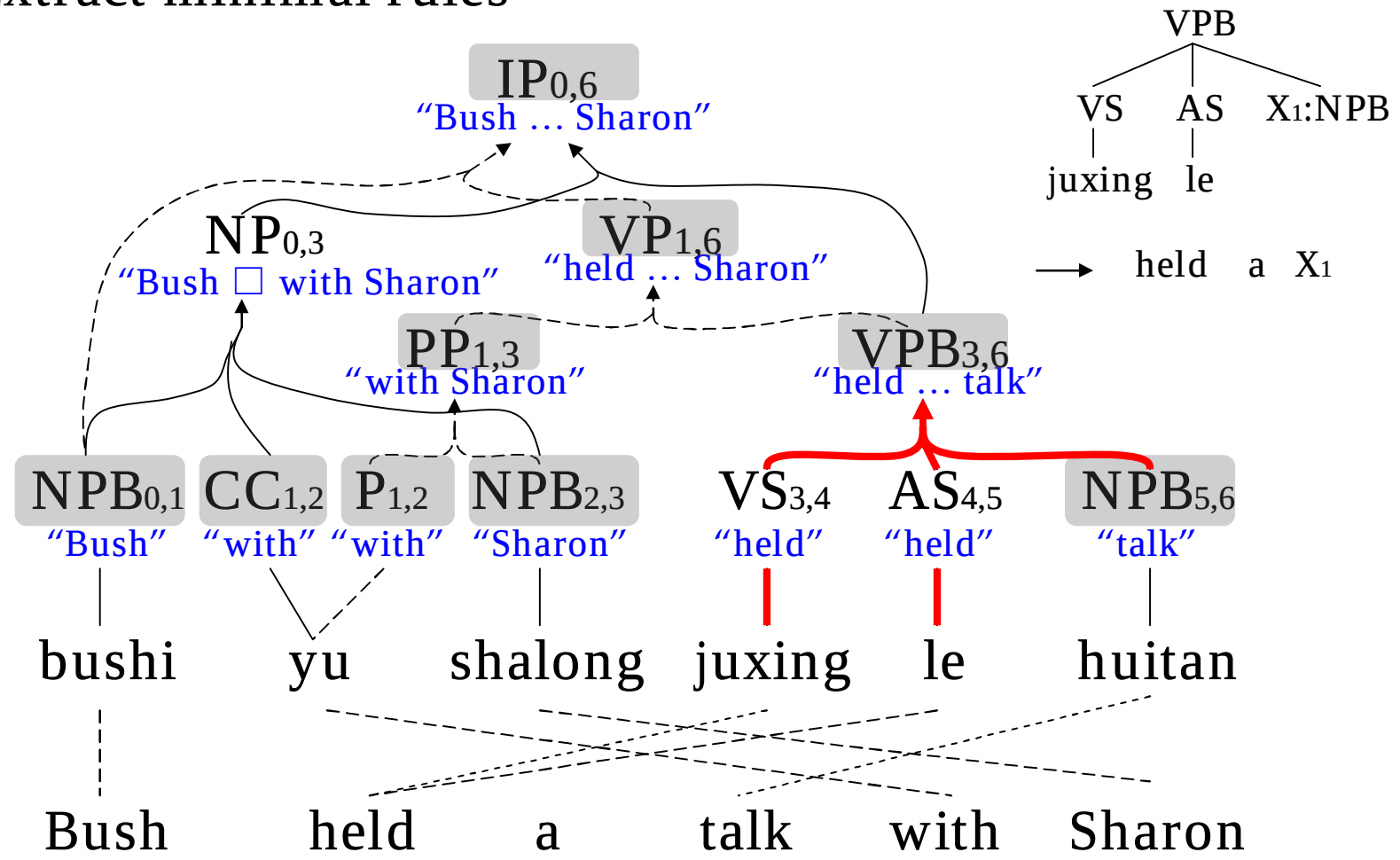
- Extract minimal rules



(Mi and Huang, 2008)

Forest-based Rule Extraction

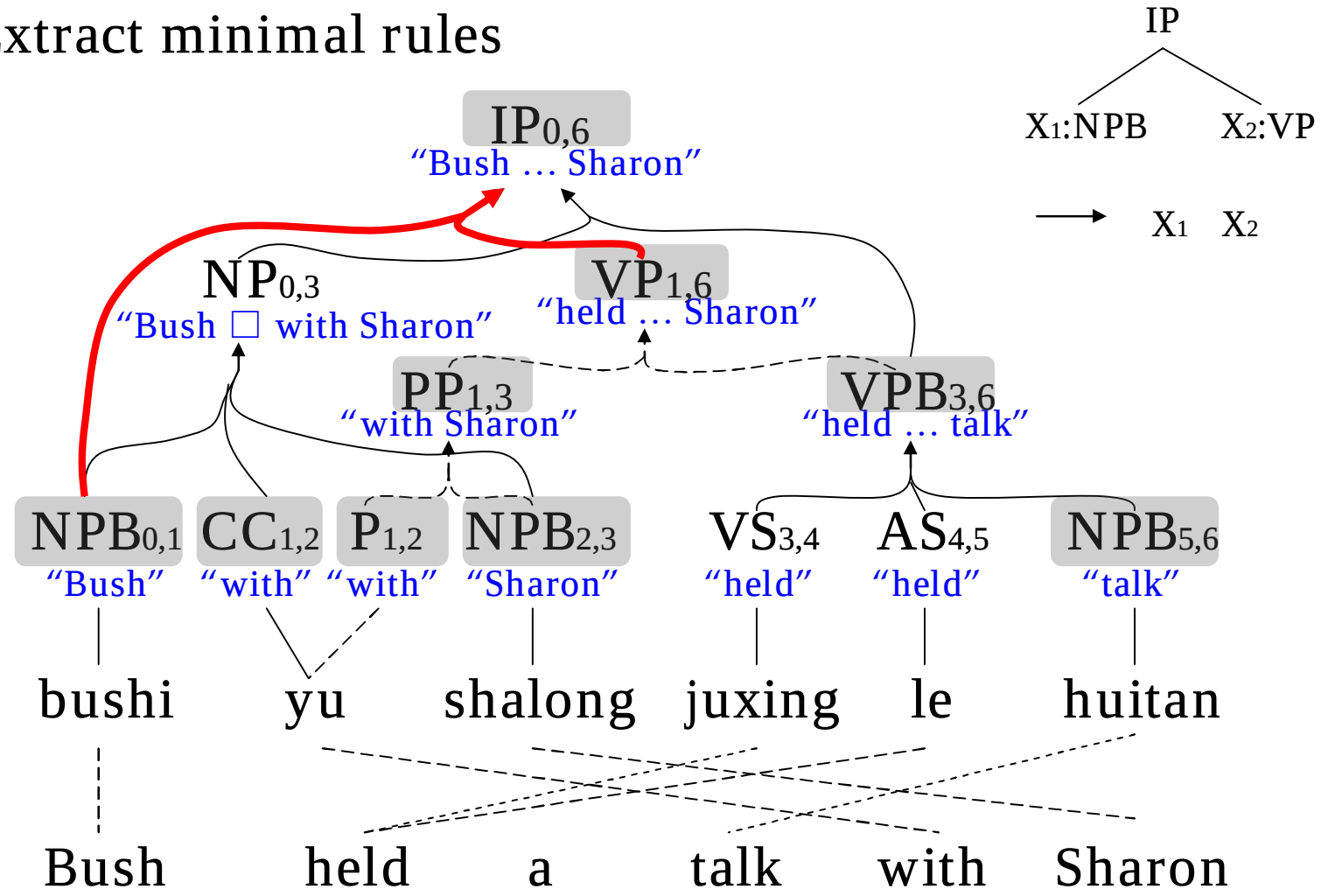
- Extract minimal rules



(Mi and Huang, 2008)

Forest-based Rule Extraction

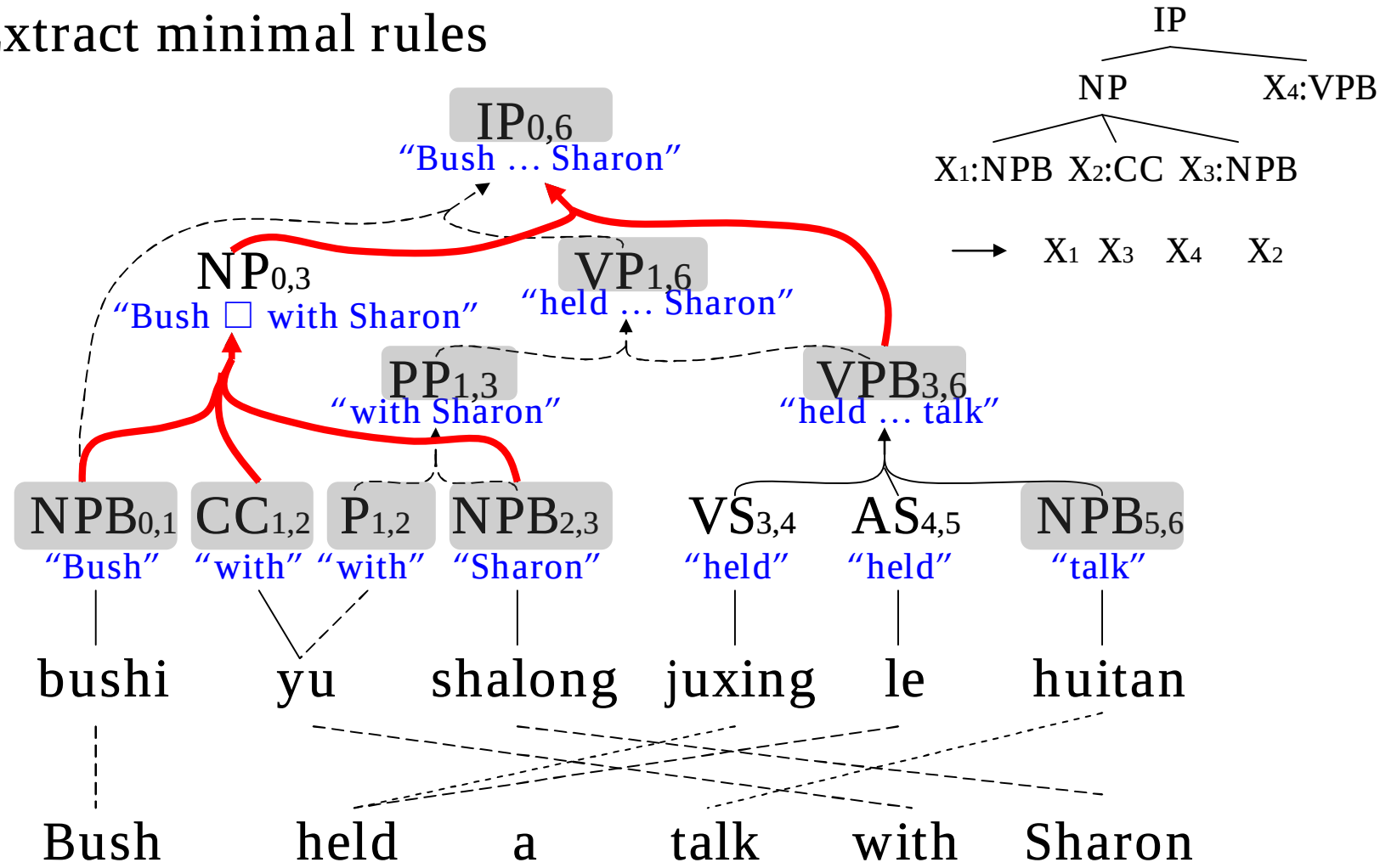
- Extract minimal rules



(Mi and Huang, 2008)

Forest-based Rule Extraction

- Extract minimal rules



(Mi and Huang, 2008)

Rule Probabilities and Rule Count

$$P(r \mid lhs(r)) = \frac{c(r)}{\sum_{r':lhs(r')=lhs(r)} c(r')}$$

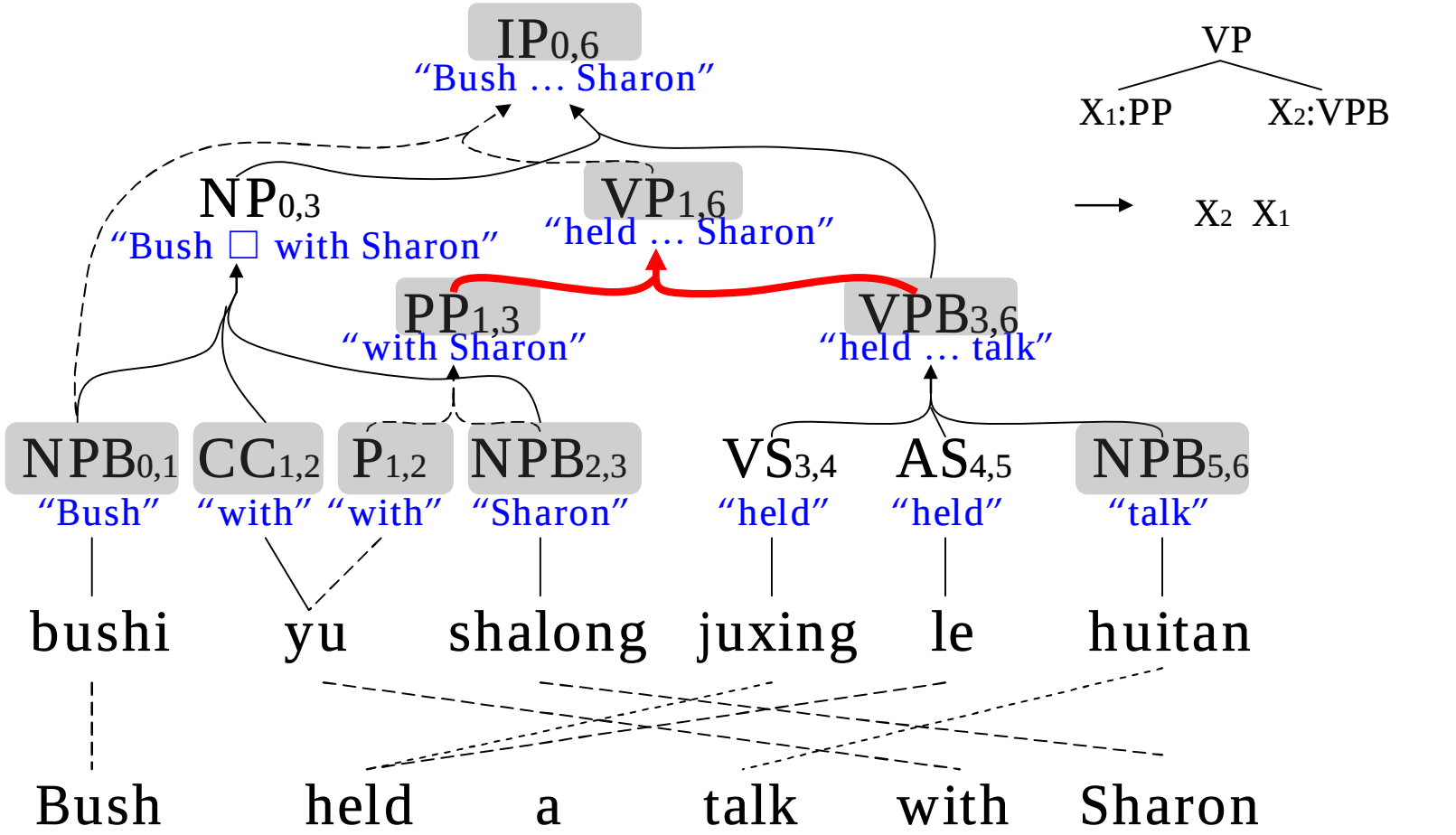
$$P(r \mid rhs(r)) = \frac{c(r)}{\sum_{r':rhs(r')=rhs(r)} c(r')}$$

$$P(r \mid root(lhs(r))) = \frac{c(r)}{\sum_{r':root(lhs(r'))=root(lhs(r))} c(r')}$$

How often does a rule occur in training examples?

Fractional Count

Q: What 's the count of this rule on this training example?

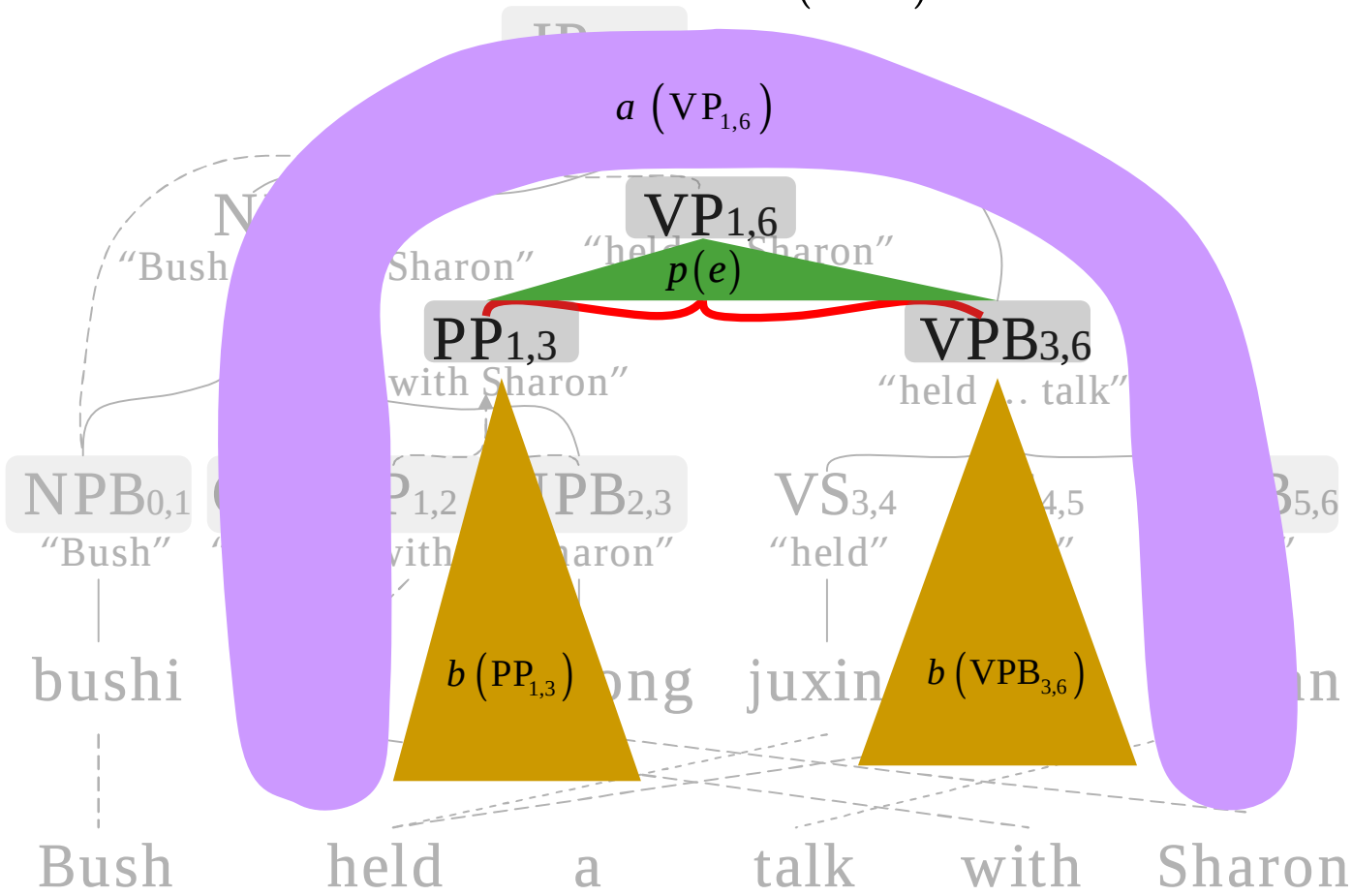


(Mi and Huang, 2008)

Fractional Count

$$ab(\{e\}) = a(\mathrm{VP}_{1,6}) \times p(e) \times b(\mathrm{PP}_{1,3}) \times b(\mathrm{VPB}_{3,6})$$

$$c(r) = \frac{ab(lhs(r))}{ab(\mathrm{TOP})}$$



(Mi and Huang, 2008)

Results

decoding

rule extraction

	1-best tree	forest
1-best tree	0.2560	0.2674
forest	0.2679	0.2816

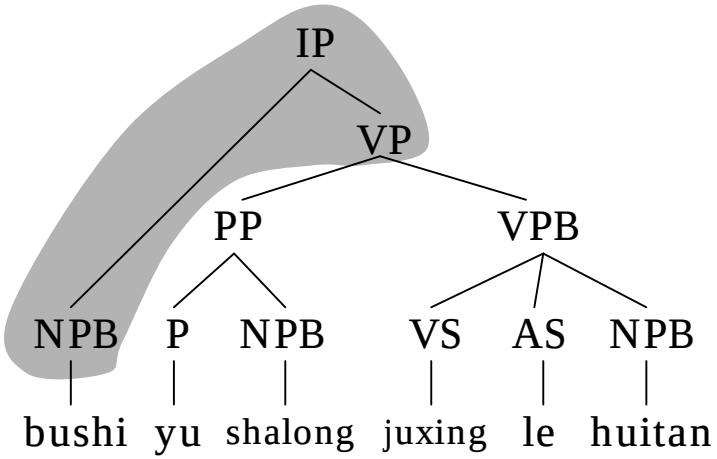
(Mi and Huang, 2008)

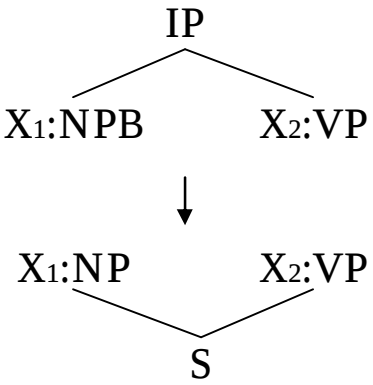
Outline

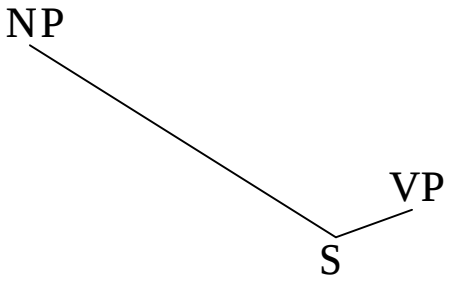
- *Part 1: Tree-based Translation*
  - *Overview and Motivation*
  - *Tree-to-String Model and Decoding*
  - *Tree-to-String Rule Extraction*
  - *Language Model-Integrated Decoding: Cube Pruning*
- *Part 2: Forest-based Translation*
  - *Packed Forest*
  - *Forest-based Decoding*
  - *Forest-based Rule Extraction*
- Part 3: Extensions
  - Tree-to-Tree Translation
  - Tree Sequence-based Translation
  - Joint Parsing and Translation
- Part 4: Conclusion

Tree-to-Tree Translation

- Recursive rewrite by pattern-matching



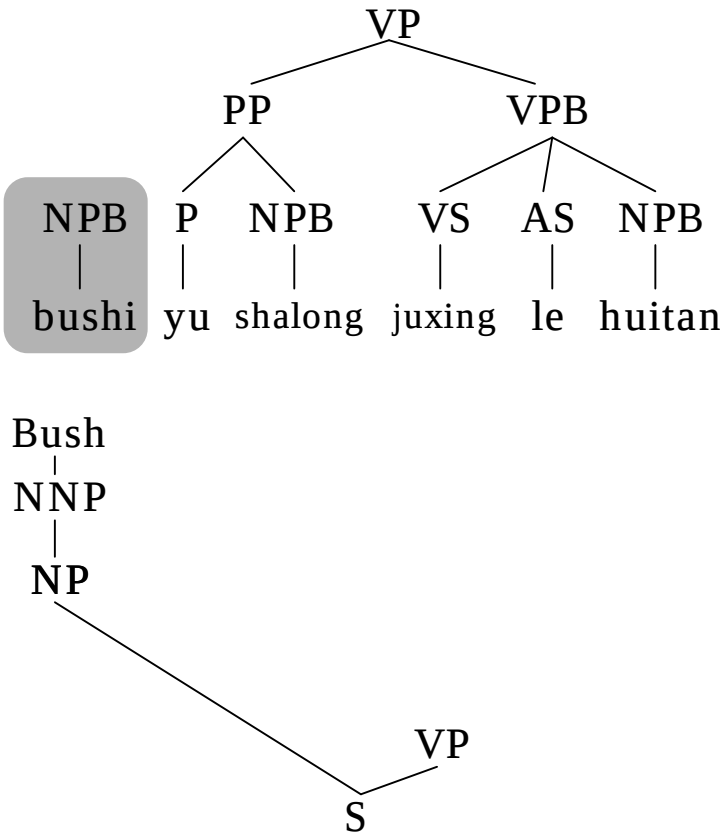


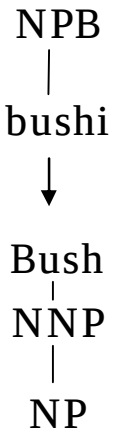


(Eisner 2003, Zhang, 2007)

Tree-to-Tree Translation

- Recursive rewrite by pattern-matching

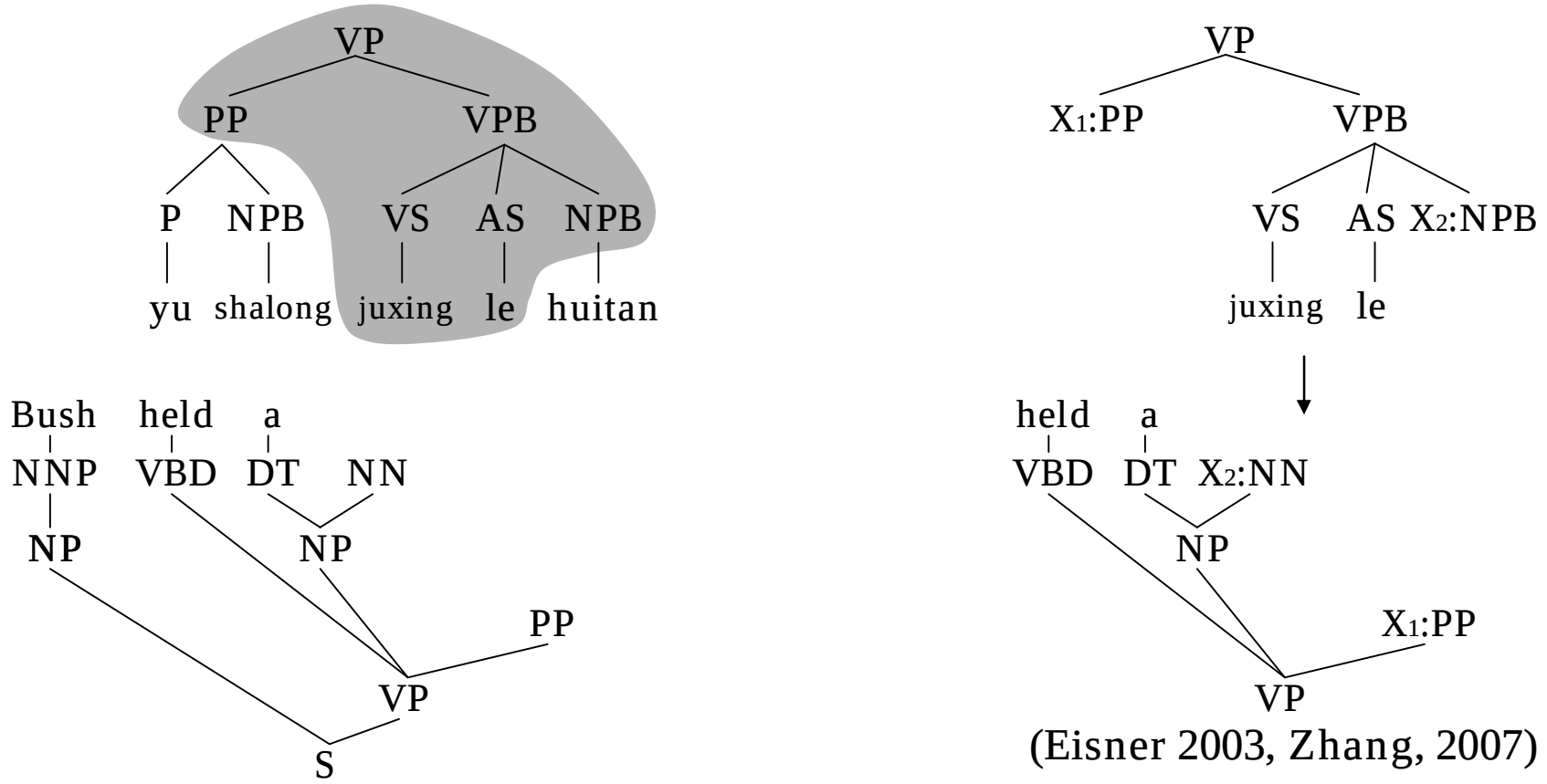




(Eisner 2003, Zhang, 2007)

Tree-to-Tree Translation

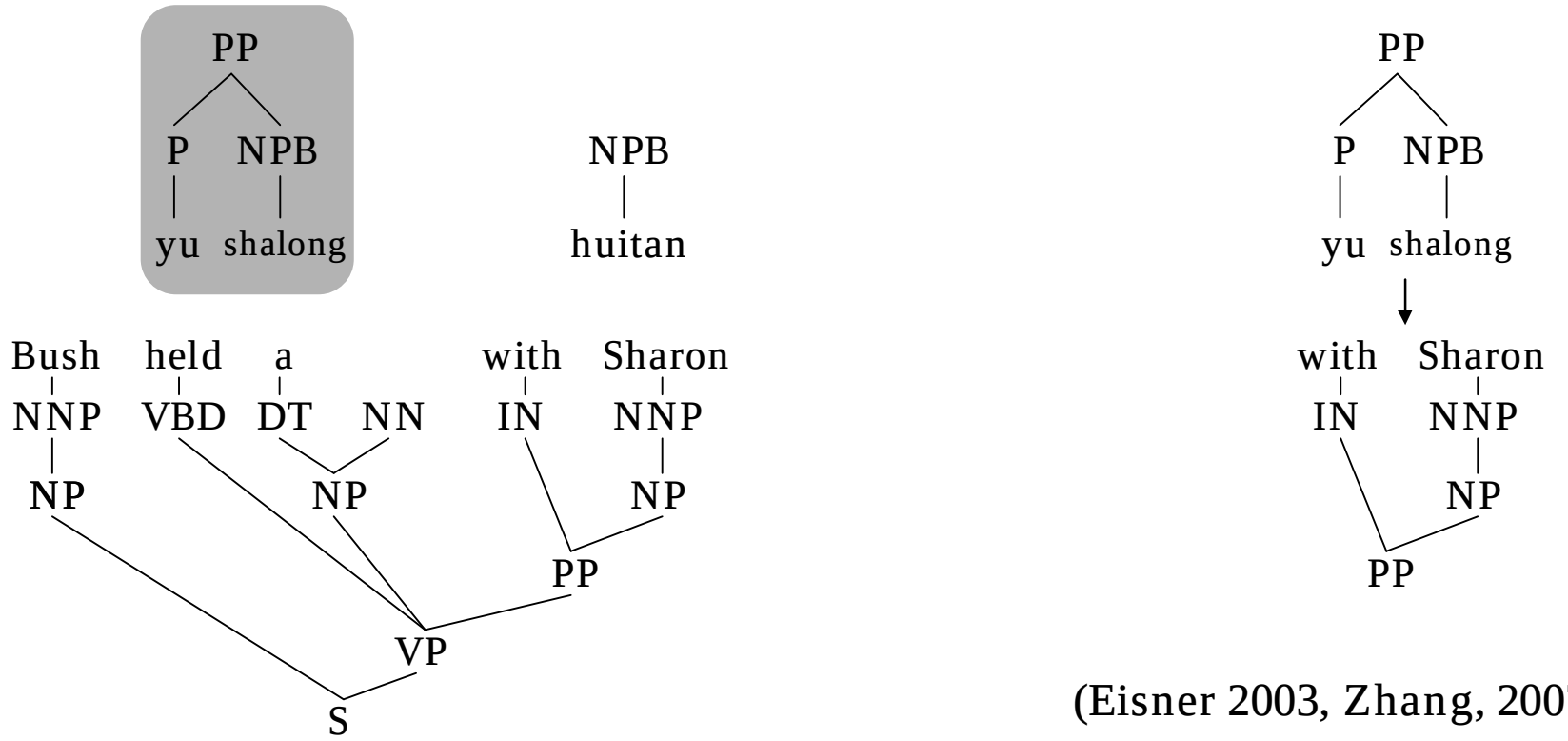
- Recursive rewrite by pattern-matching



(Eisner 2003, Zhang, 2007)

Tree-to-Tree Translation

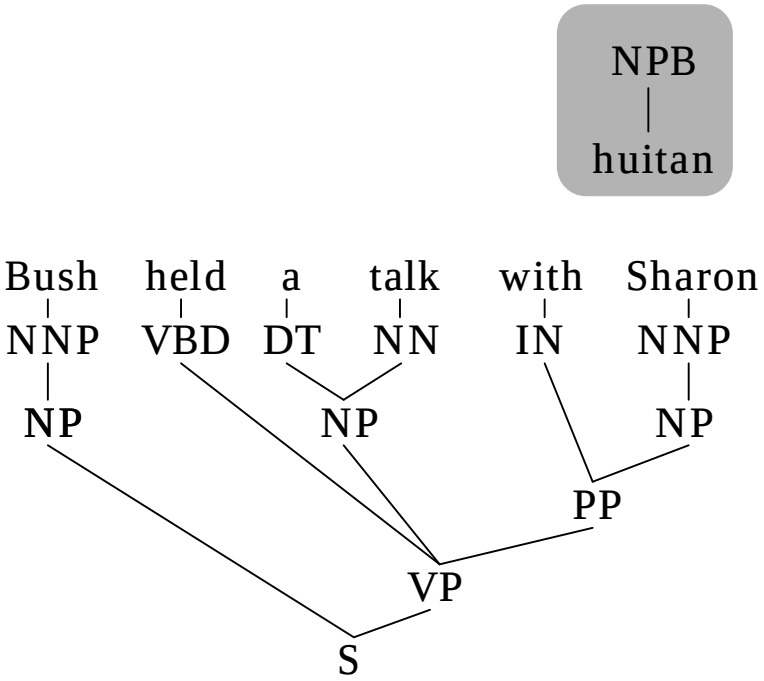
- Recursive rewrite by pattern-matching

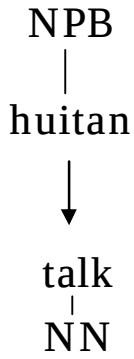


(Eisner 2003, Zhang, 2007)

Tree-to-Tree Translation

- Recursive rewrite by pattern-matching

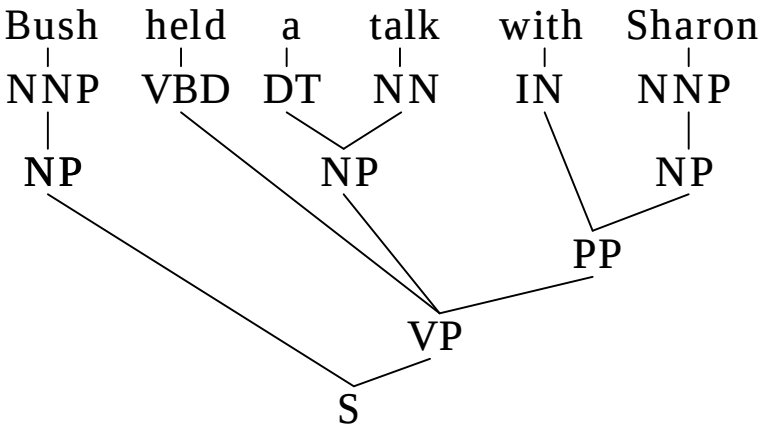




(Eisner 2003, Zhang, 2007)

Tree-to-Tree Translation

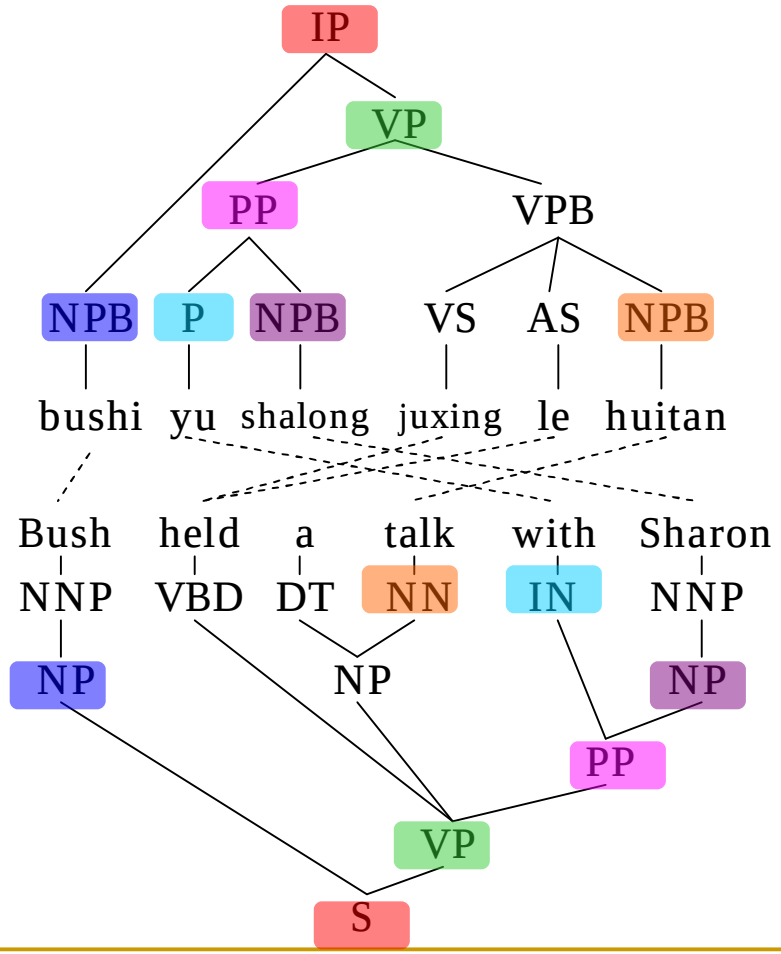
- Recursive rewrite by pattern-matching



(Eisner 2003, Zhang, 2007)

Tree-to-Tree Rule Extraction

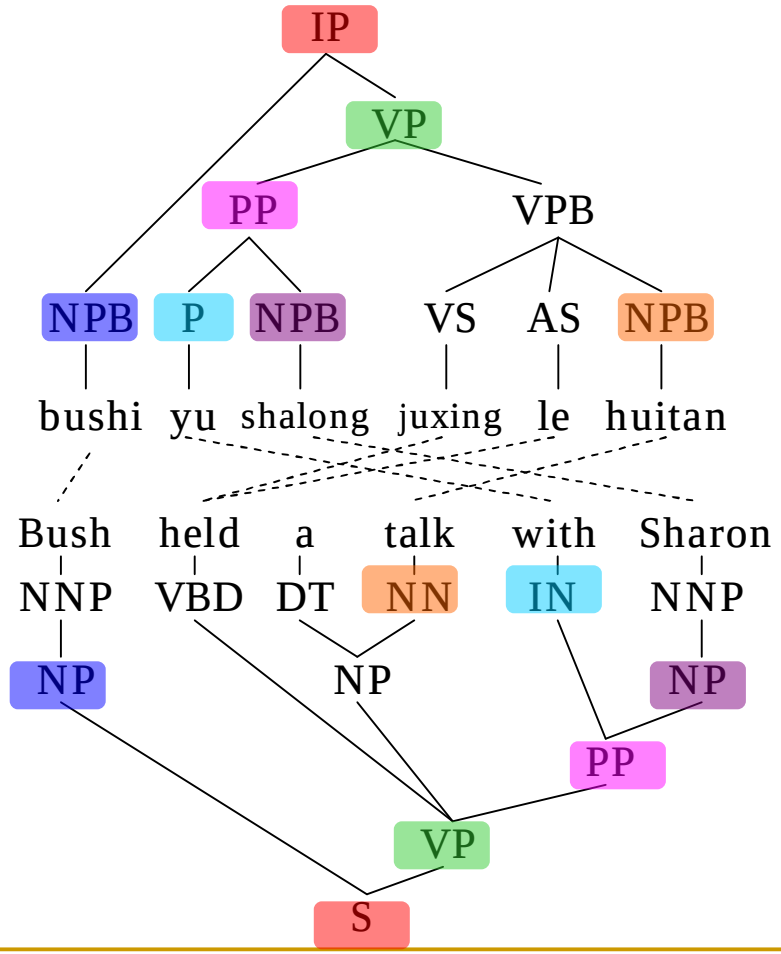
- Find admissible node pairs

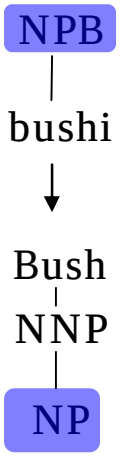


(Zhang, 2007, Liu et al., 2009a)

Tree-to-Tree Rule Extraction

- Extract minimal rules

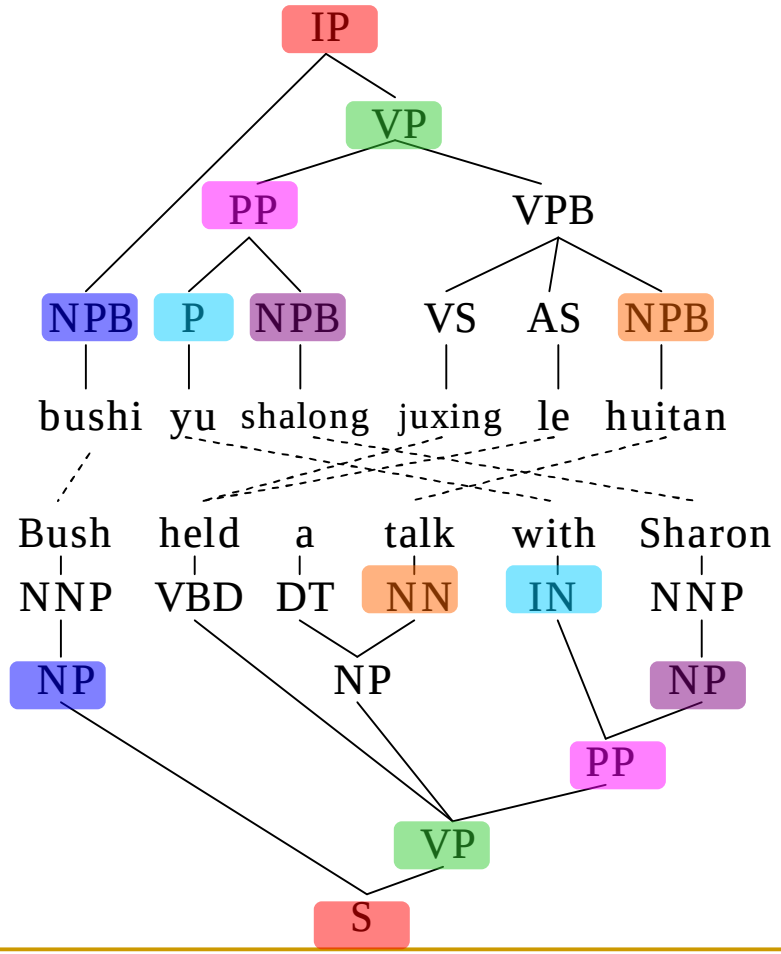


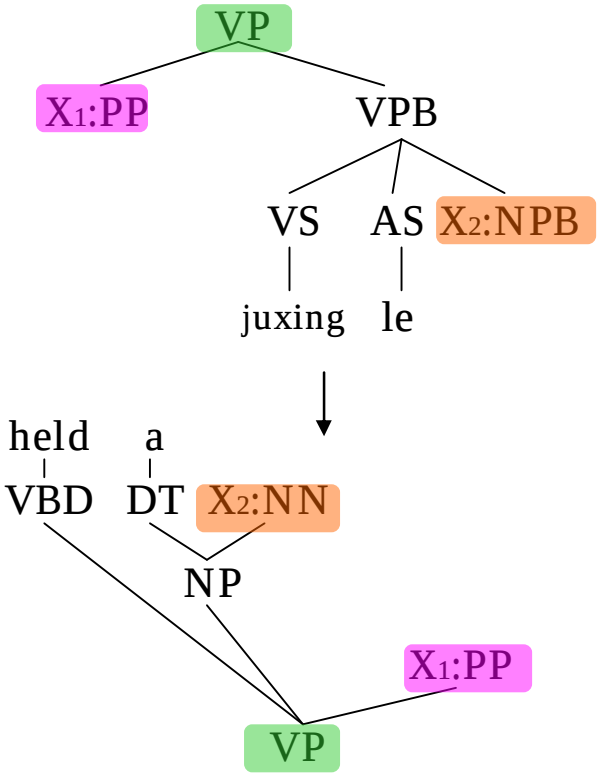


(Zhang, 2007, Liu et al., 2009a)

Tree-to-Tree Rule Extraction

- Extract minimal rules

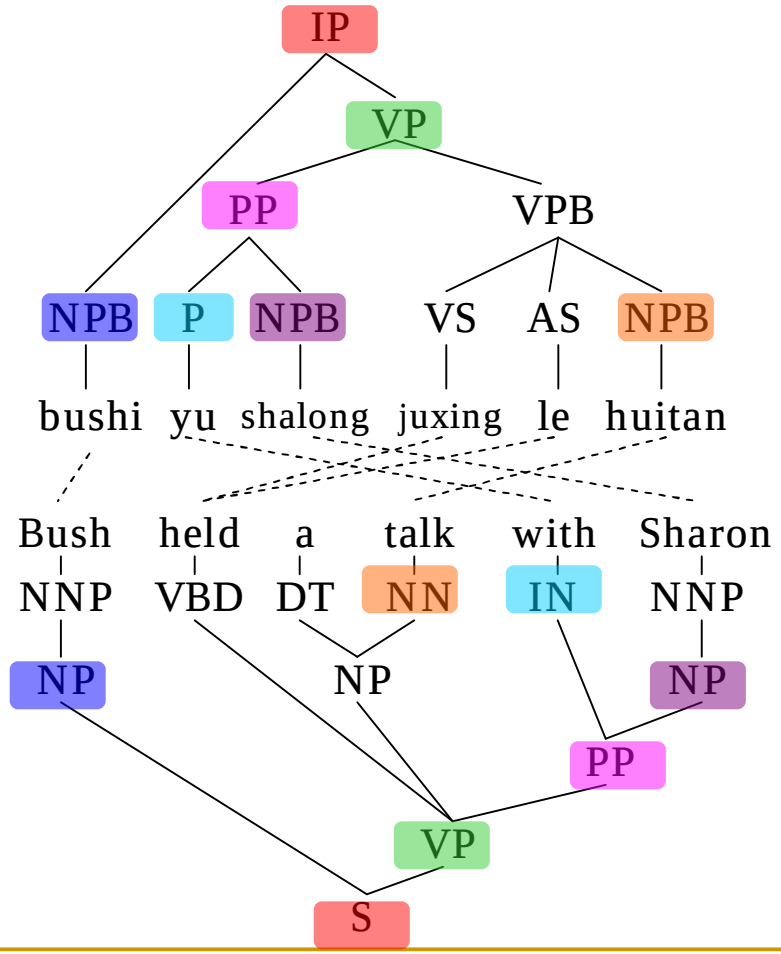


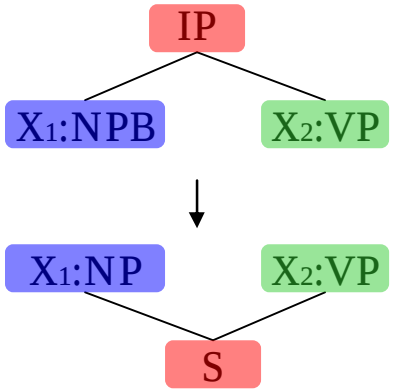


(Zhang, 2007, Liu et al., 2009a)

Tree-to-Tree Rule Extraction

- Extract minimal rules

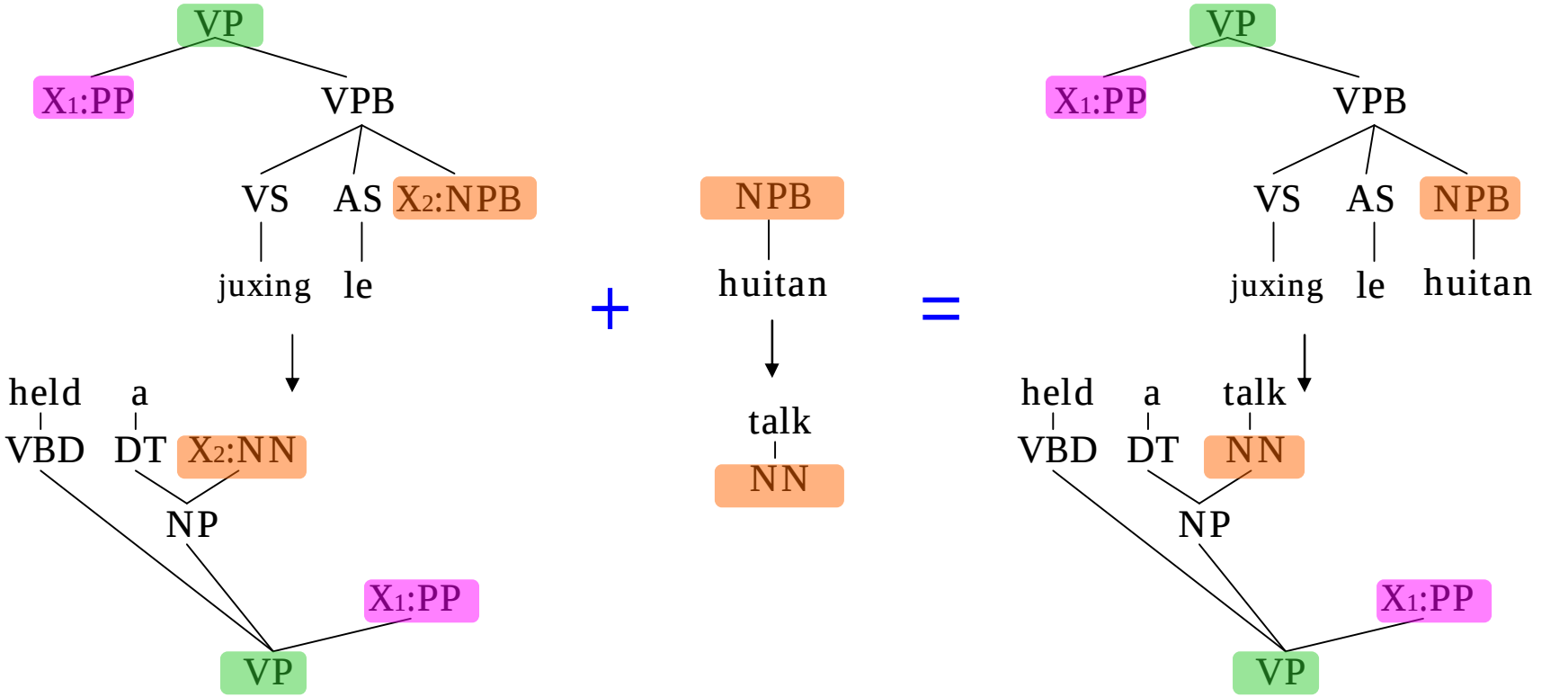




(Zhang, 2007, Liu et al., 2009a)

Tree-to-Tree Rule Extraction

- Get composed rules



(Zhang, 2007, Liu et al., 2009a)

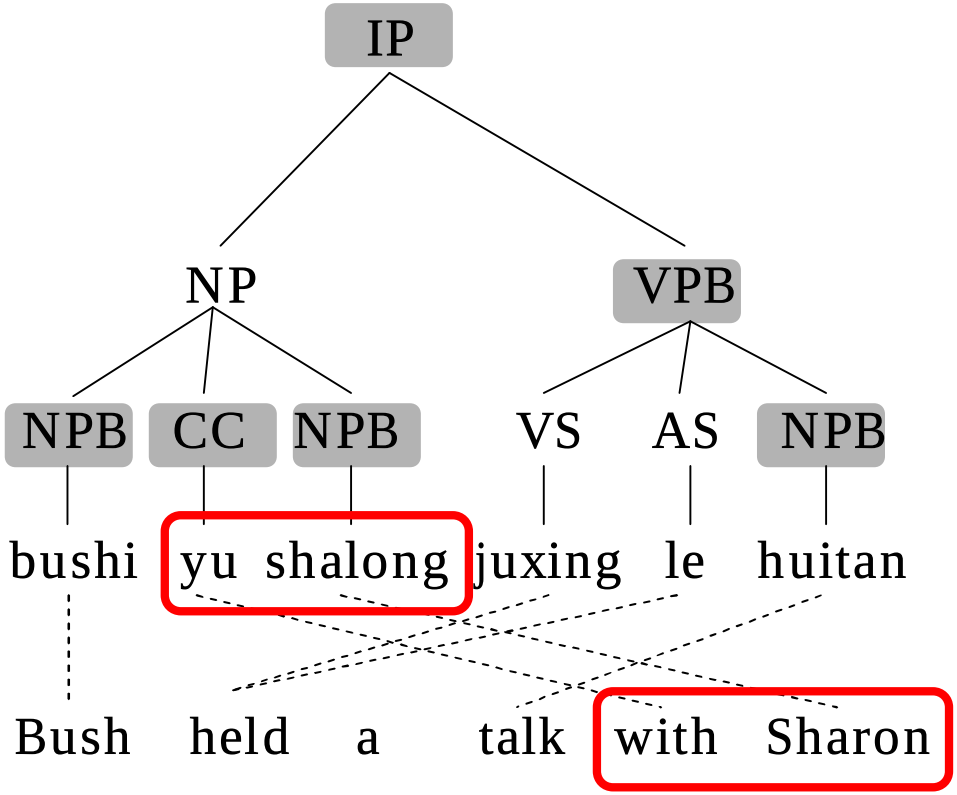
Challenges

- Tree-to-tree translation is over-constrained
  - Poorest rule coverage
  - Suffers from parsing mistake propagation on both sides
- Recent advances
  - Use tree sequence (Zhang et al., 2008)
  - Use packed forest (Liu et al., 2009a)
  - Fuzzy extraction and decoding (Chiang, 2010)

Outline

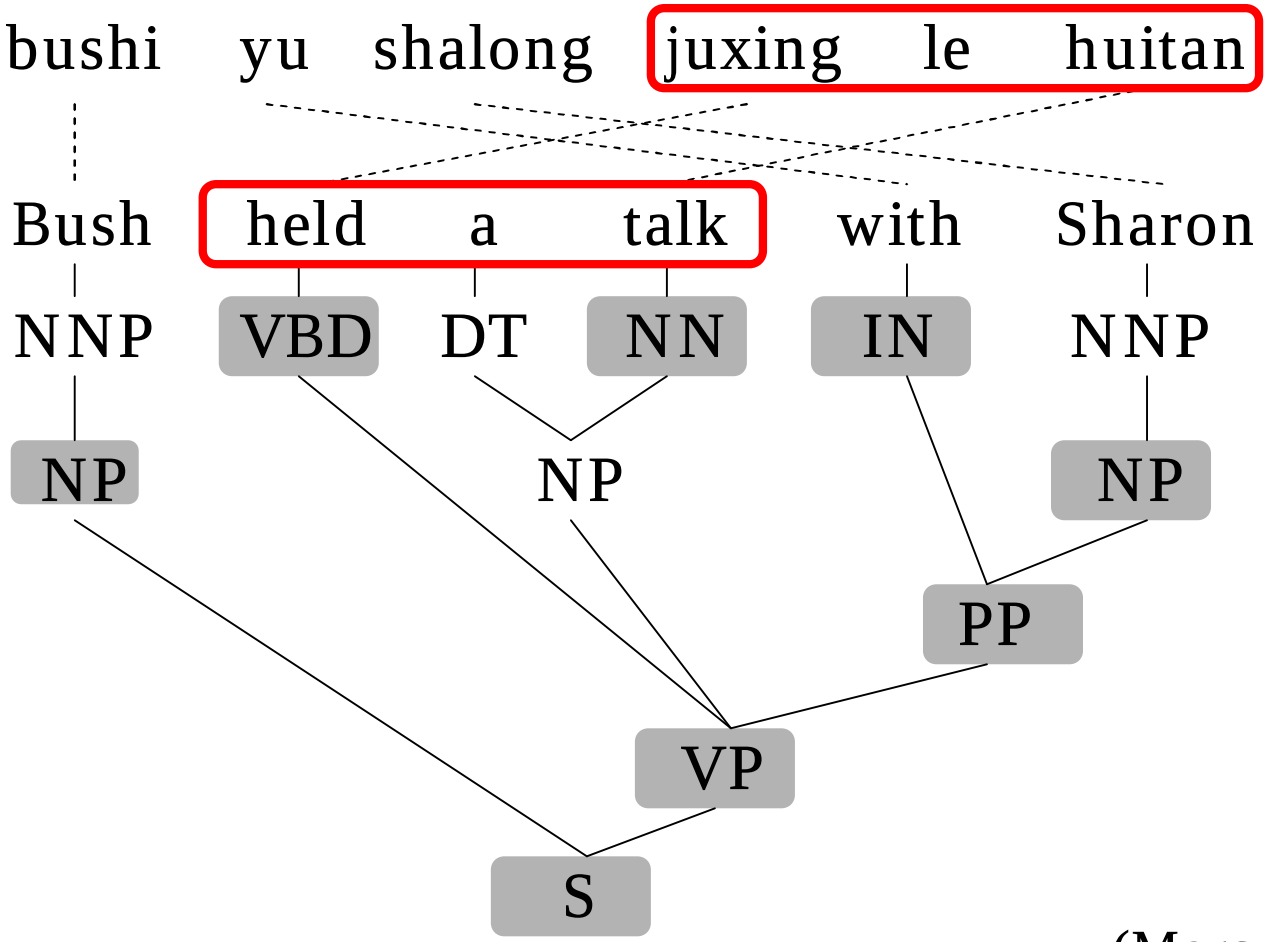
- *Part 1: Tree-based Translation*
  - *Overview and Motivation*
  - *Tree-to-String Model and Decoding*
  - *Tree-to-String Rule Extraction*
  - *Language Model-Integrated Decoding: Cube Pruning*
- *Part 2: Forest-based Translation*
  - *Packed Forest*
  - *Forest-based Decoding*
  - *Forest-based Rule Extraction*
- Part 3: Extensions
  - *Tree-to-Tree Translation*
  - Tree Sequence-based Translation
  - Joint Parsing and Translation
- Part 4: Conclusion

Non-Constituent Phrase Pairs



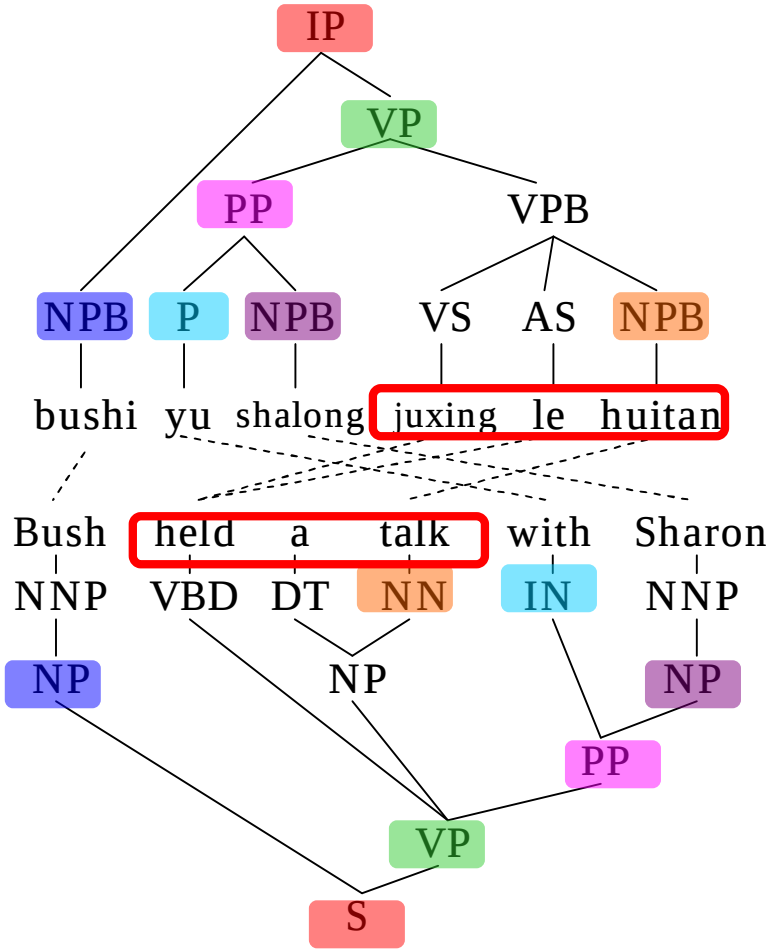
(Marcu et al., 2006)

Non-Constituent Phrase Pairs



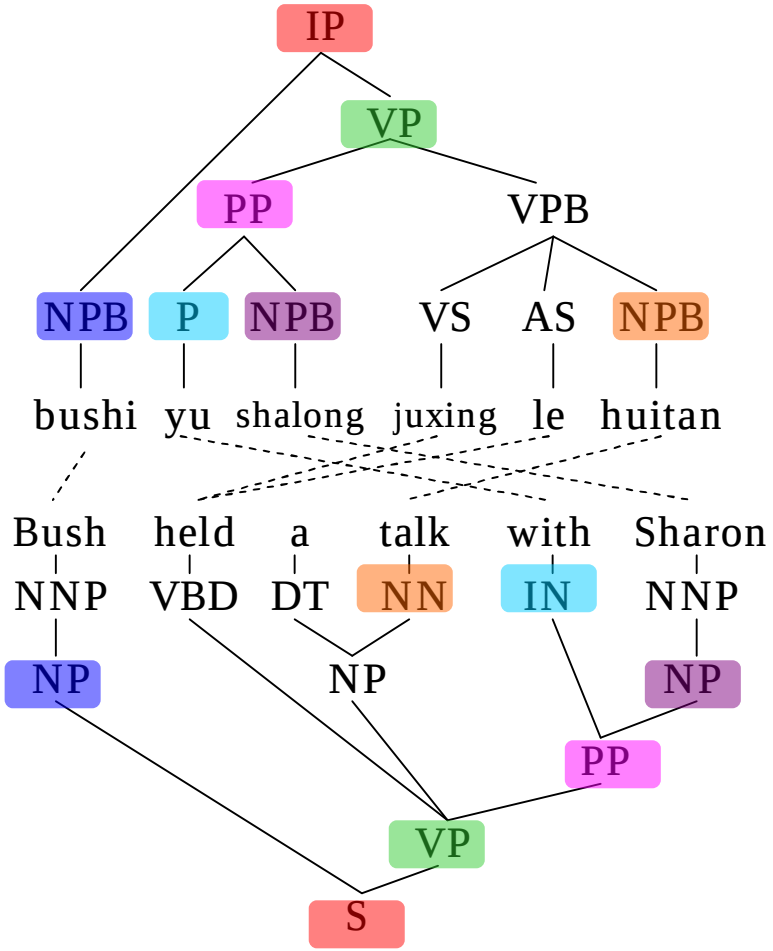
(Marcu et al., 2006)

Non-Constituent Phrase Pairs



(Marcu et al., 2006)

Rule Coverage



phrase pair	s2s	t2s	s2t	t2t
(bushi, Bush)	√	√	√	√
(yu, with)	√	√	√	√
(shalong, Sharon)	√	√	√	√
(huitan, talk)	√	√	√	√
(yu shalong, with Sharon)	√	√	√	√
(juxing le, held)	√	×	√	×
(juxing … huitan, held … talk)	√	√	×	×
(yu … huitan, held … Sharon)	√	√	√	√
(bushi … huitan, Bush … Sharon)	√	√	√	√
	100%	89%	89%	78%

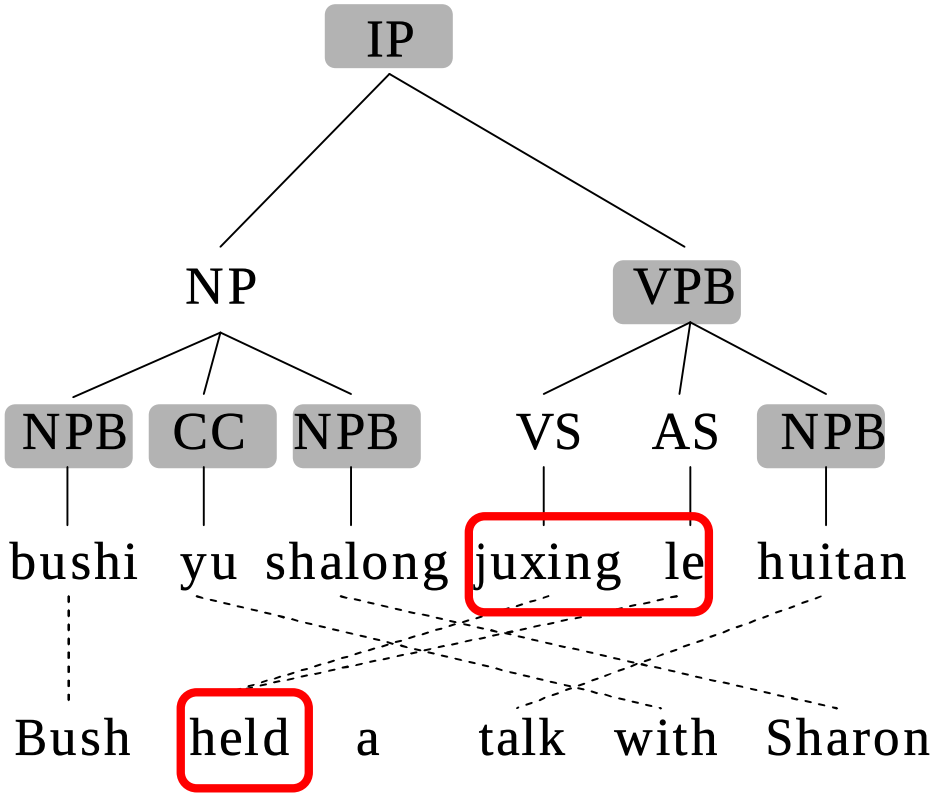
Rule Coverage

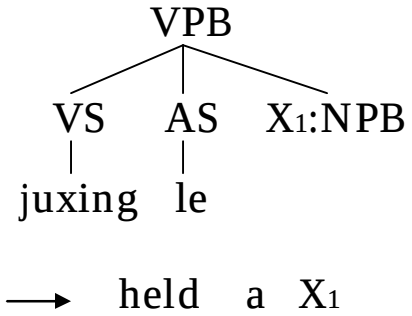
model	human	automatic
string-to-string	100%	100%
tree-to-string	78%	75%
string-to-tree	76%	72%
tree-to-tree	68%	60%

Results from (Chiang, 2010)

Solutions

- Extend to larger rules

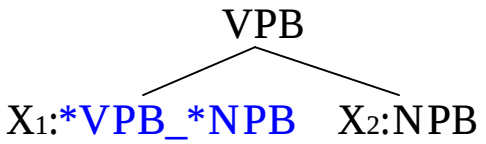


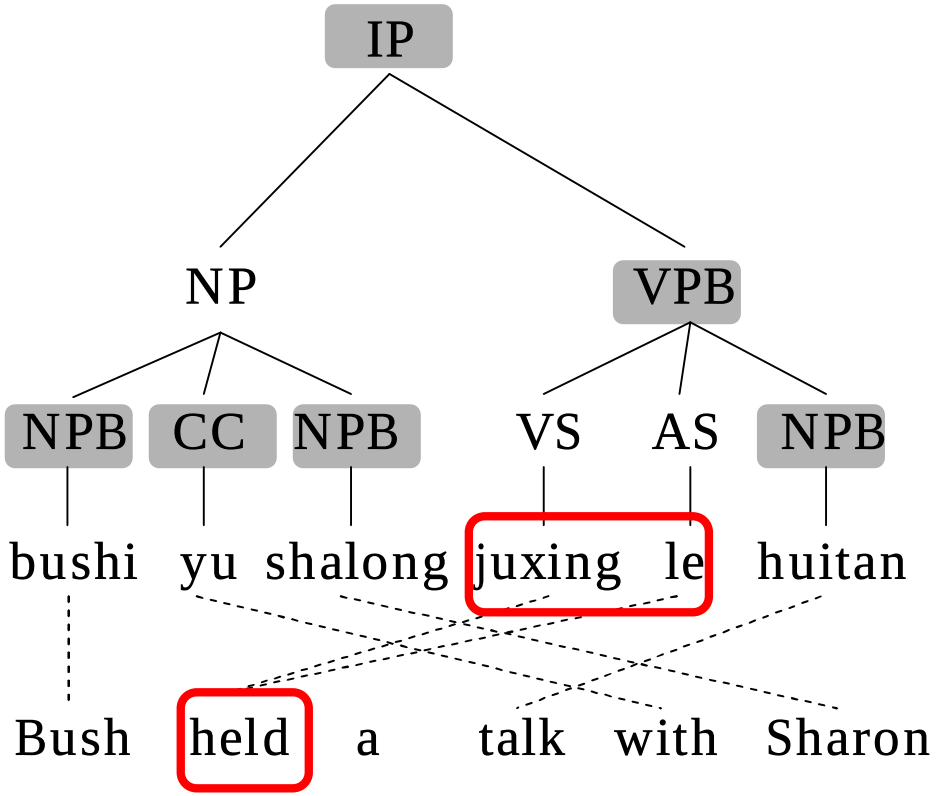


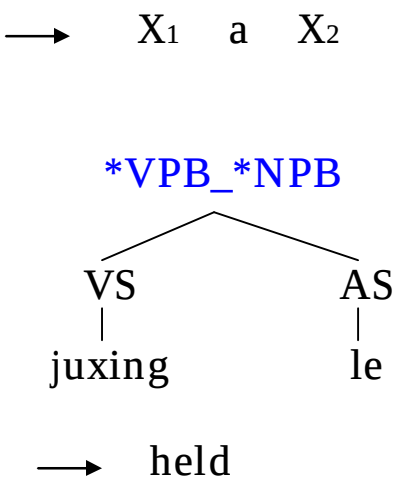
(Galley et al., 2006)

Solutions

- Add pseudo nodes



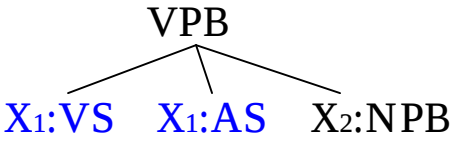


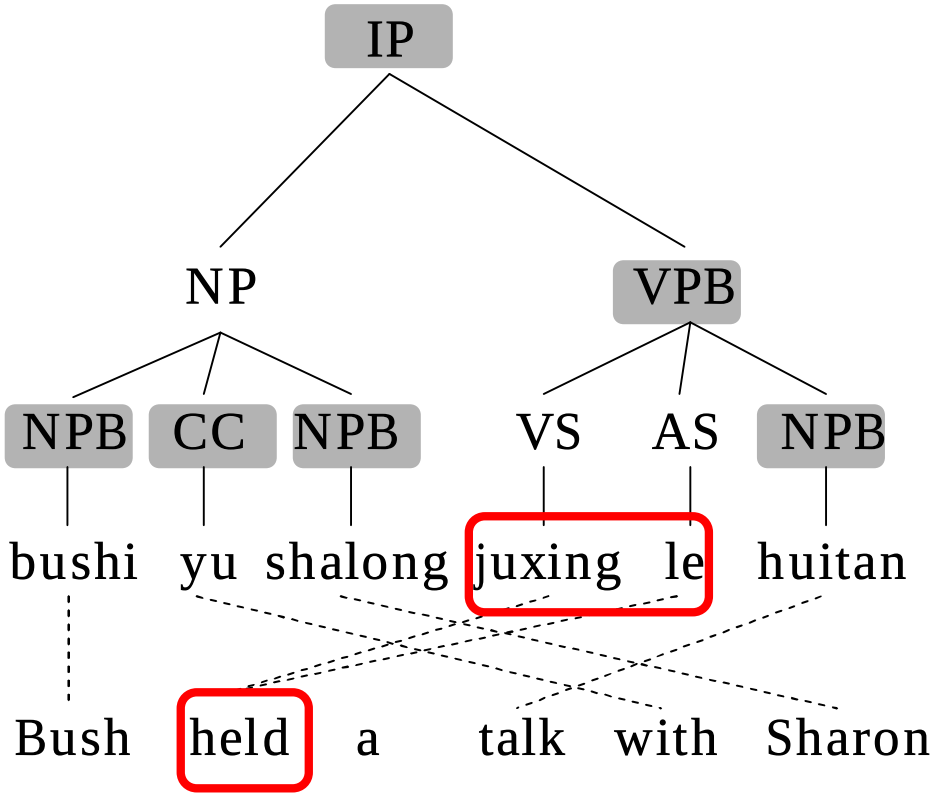


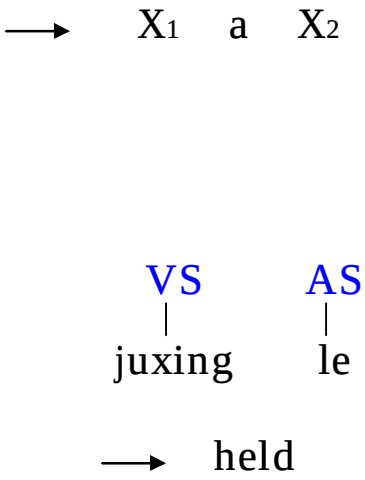
(Marcu et al., 2006)

Solutions

- Use tree sequences







(Liu et al., 2007; Zhang et al., 2008)

Tree-Sequence + Forest

system	input	rule	BLEU
Moses	string	string-to-string	25.7
tree-to-string	tree	tree-to-string	26.1
		tree-sequence-to-string	27.0
	forest	tree-to-string	27.7
		tree-sequence-to-string	28.8

Results from (Zhang et al., 2009)

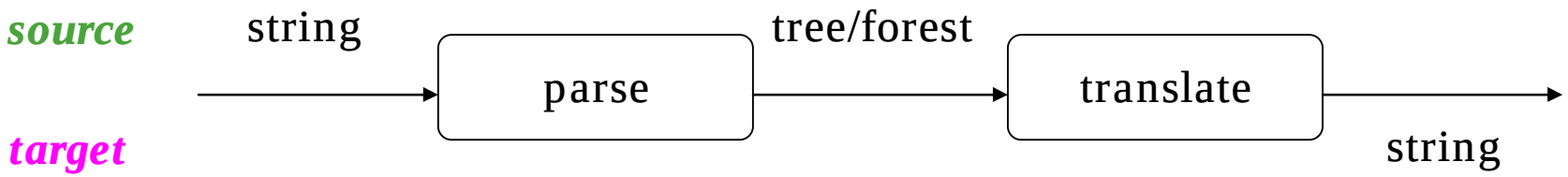
Other Solutions

- Re-structure syntax-trees (Wang et al., 2007)
- Offer more trees (Mi and Huang, 2008)
- Re-align syntax trees and strings (May and Knight, 2007)
- Well-formed dependency structures (Shen et al., 2008)
- Gibbs sampling (Cohn and Blunsom, 2009)
- Joint decoding (Liu et al., 2009b)

Outline

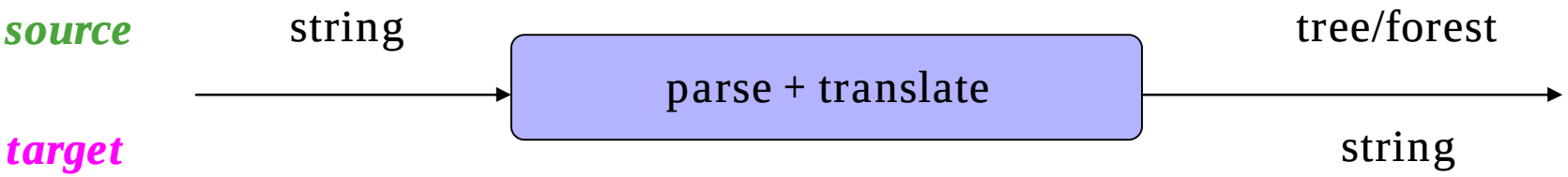
- *Part 1: Tree-based Translation*
  - *Overview and Motivation*
  - *Tree-to-String Model and Decoding*
  - *Tree-to-String Rule Extraction*
  - *Language Model-Integrated Decoding: Cube Pruning*
- *Part 2: Forest-based Translation*
  - *Packed Forest*
  - *Forest-based Decoding*
  - *Forest-based Rule Extraction*
- Part 3: Extensions
  - *Tree-to-Tree Translation*
  - *Tree Sequence-based Translation*
  - Joint Parsing and Translation
- Part 4: Conclusion

Separate Parsing and Translation



- Separate grammar for parsing and translation
- decoding is fast!

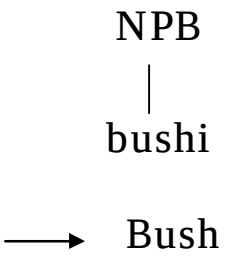
Joint Parsing and Translation

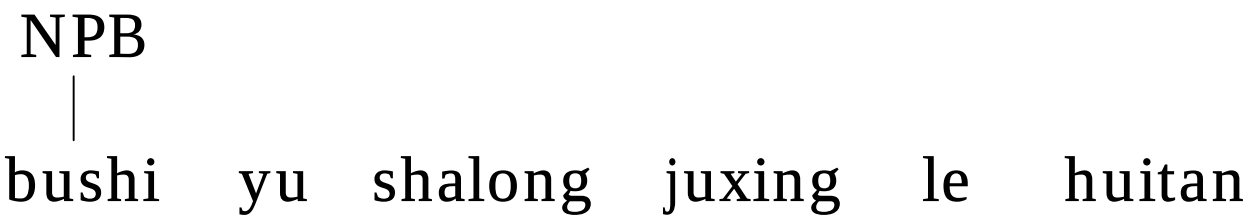


- Its search space is larger than tree/forest
- It is a translator as well as a parser
- Parsing interacts with translation

(Liu and Liu, 2010)

Tree-to-String Translation as Parsing

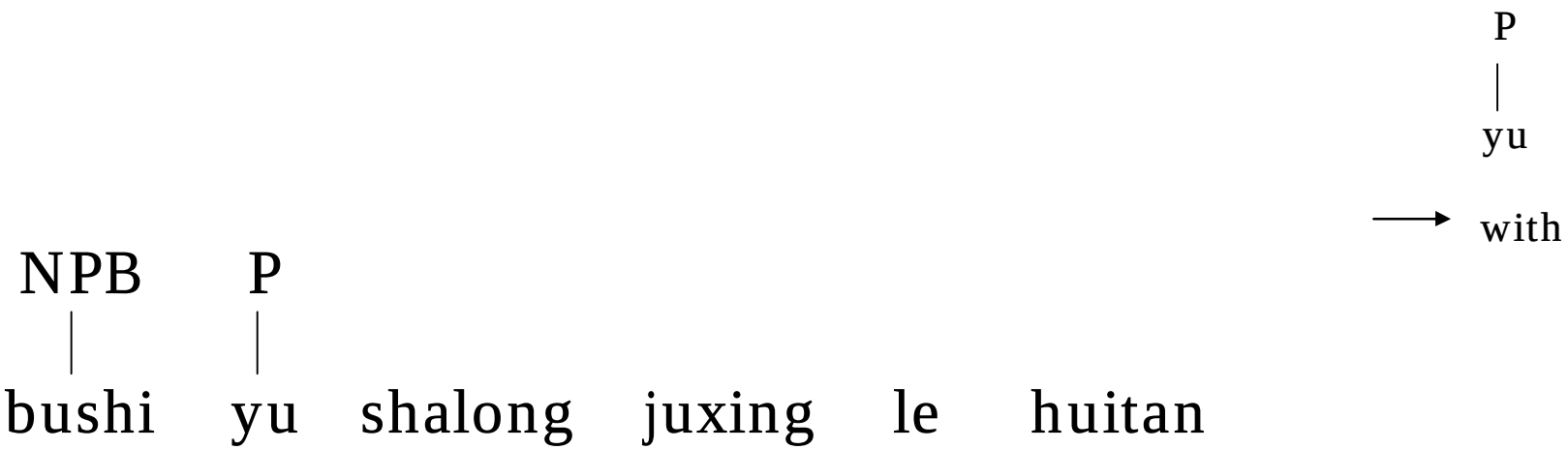




Bush

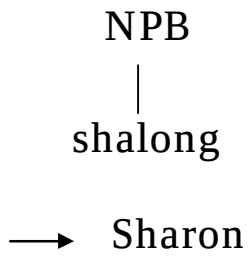
(Liu and Liu, 2010)

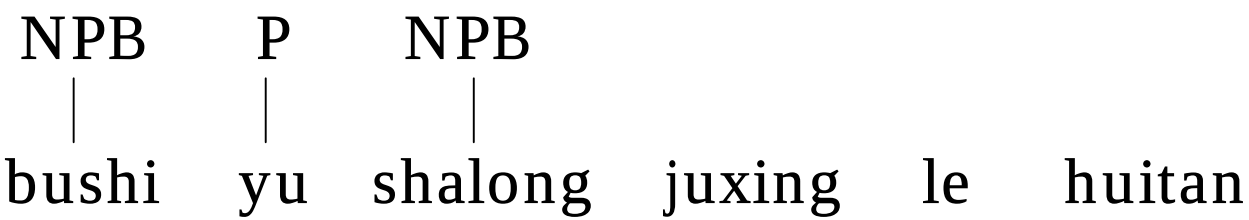
Tree-to-String Translation as Parsing



(Liu and Liu, 2010)

Tree-to-String Translation as Parsing

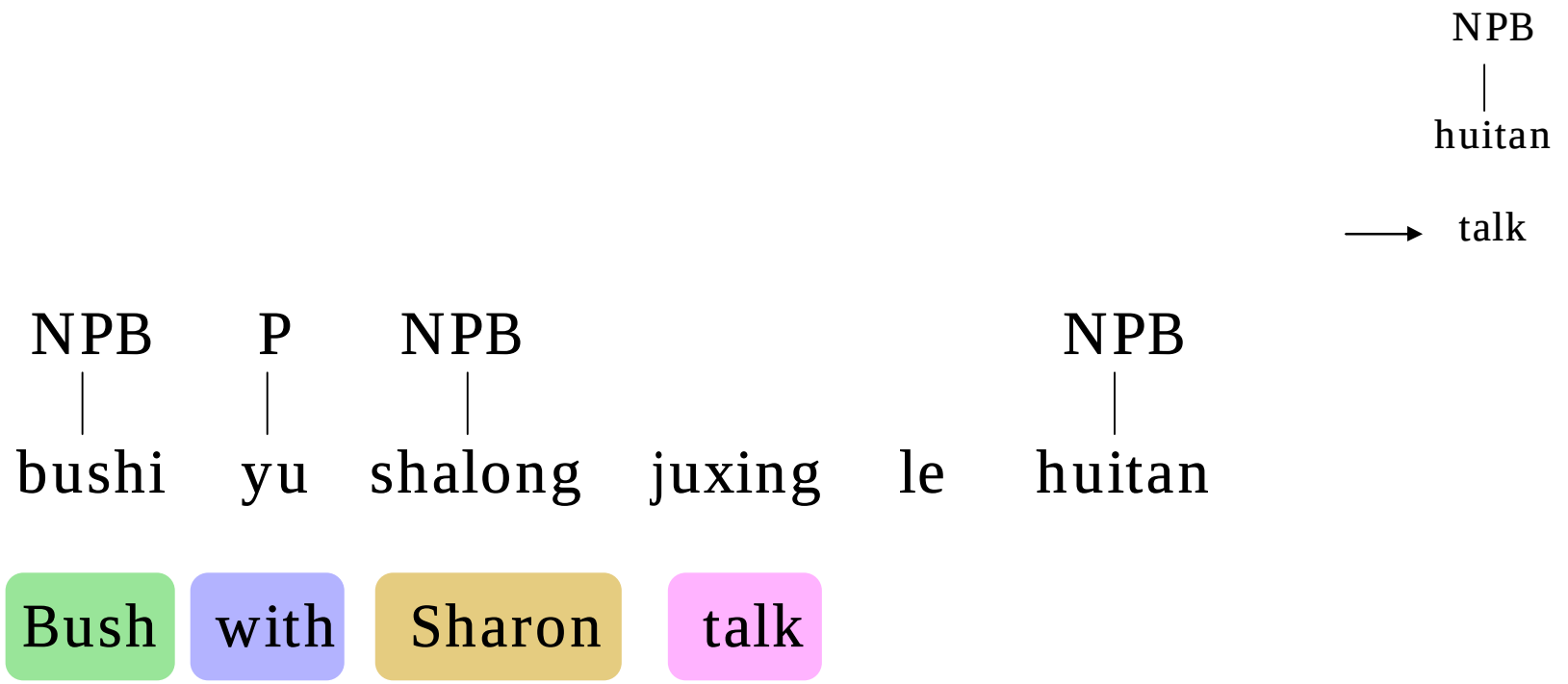




Bush with Sharon

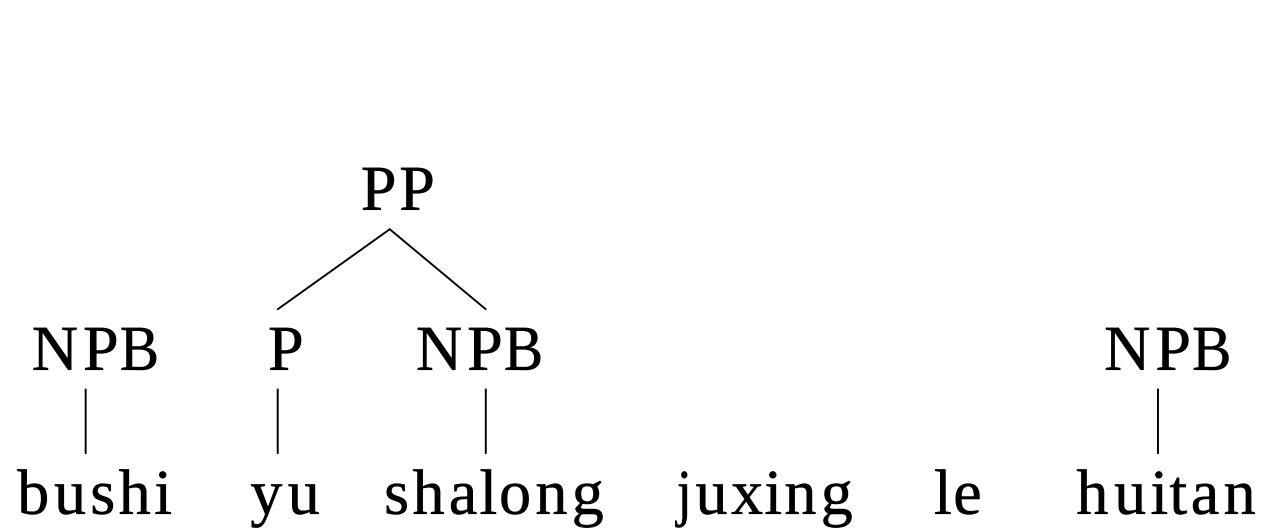
(Liu and Liu, 2010)

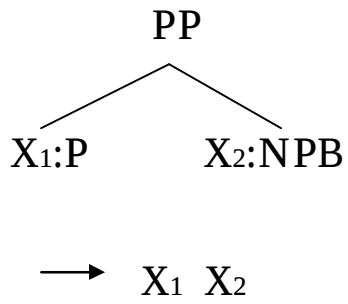
Tree-to-String Translation as Parsing



(Liu and Liu, 2010)

Tree-to-String Translation as Parsing

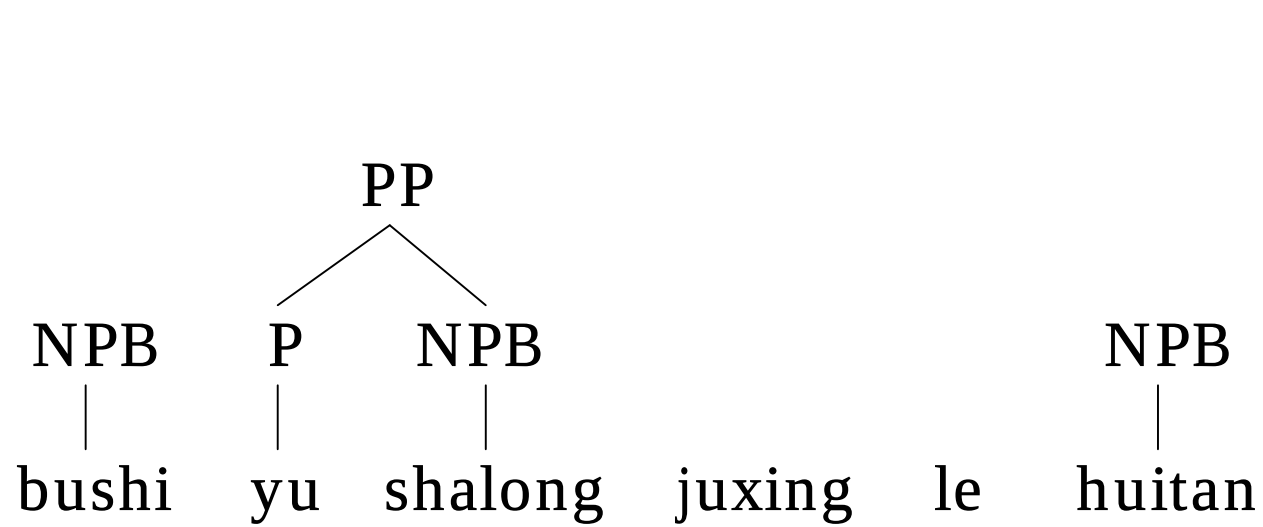


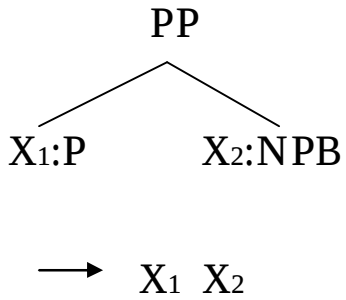


Bush with Sharon talk

(Liu and Liu, 2010)

Tree-to-String Translation as Parsing

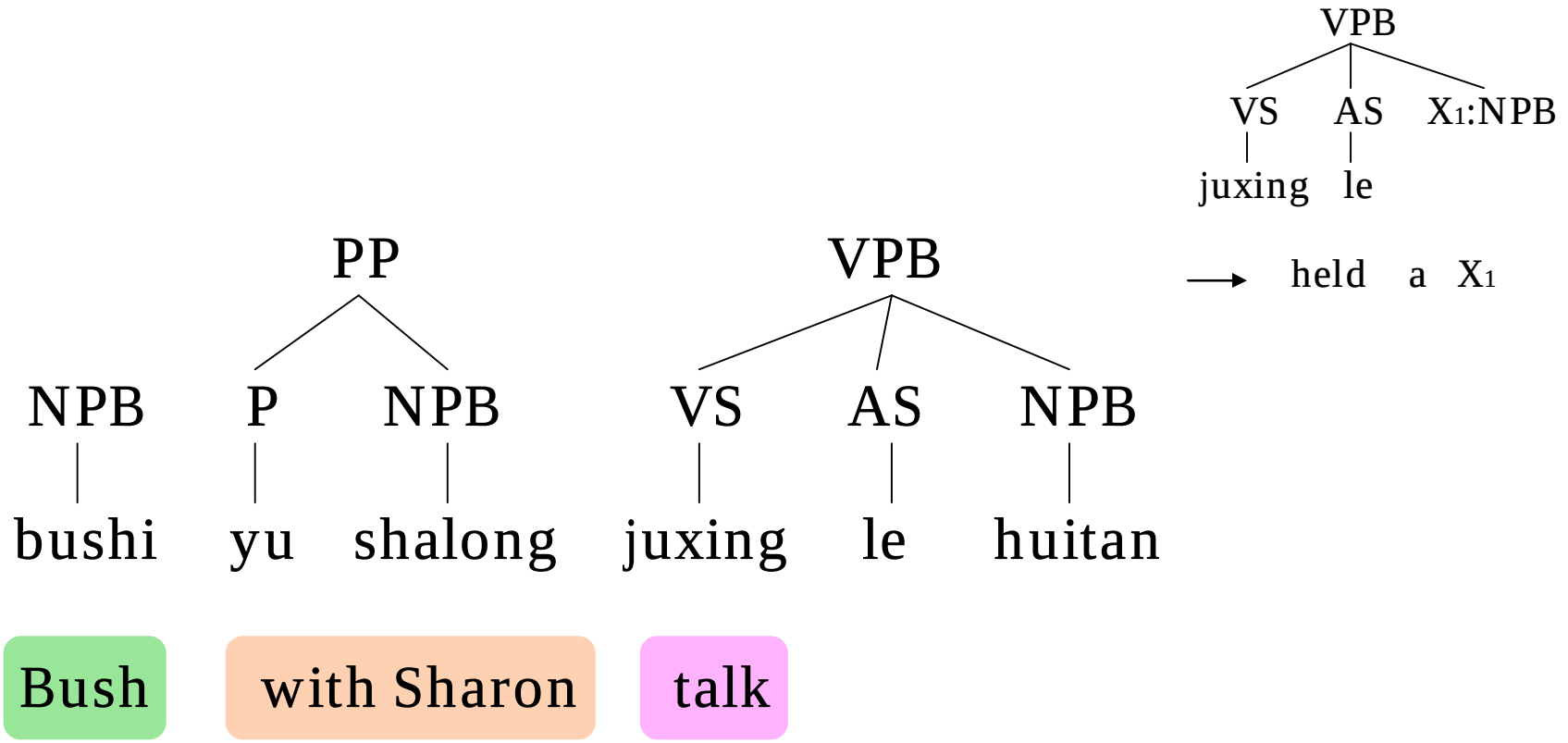




Bush with Sharon talk

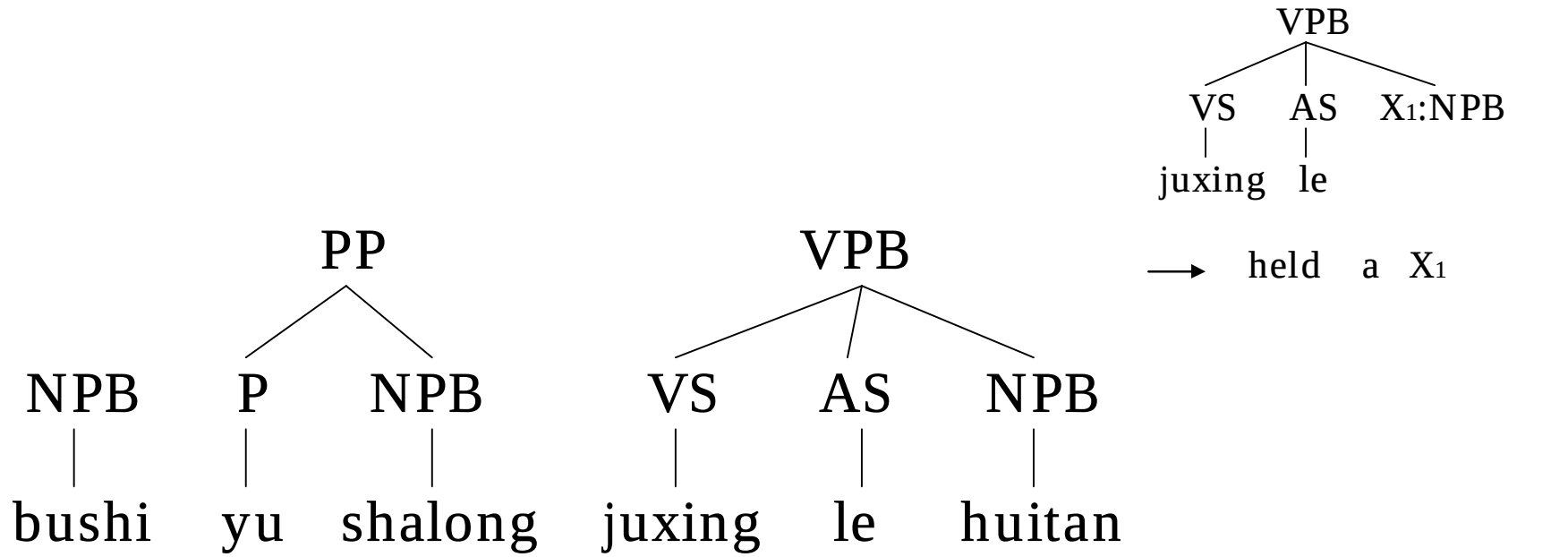
(Liu and Liu, 2010)

Tree-to-String Translation as Parsing



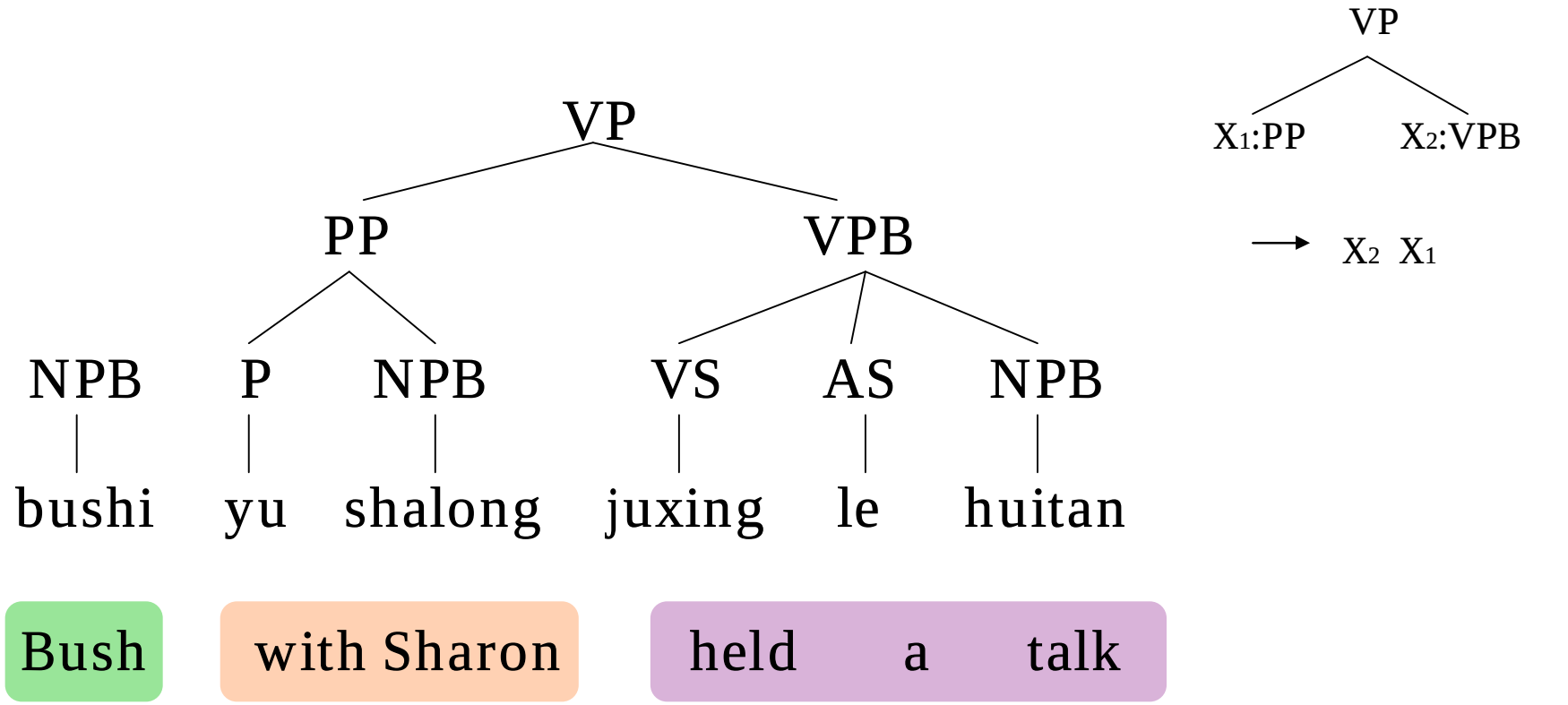
(Liu and Liu, 2010)

Tree-to-String Translation as Parsing



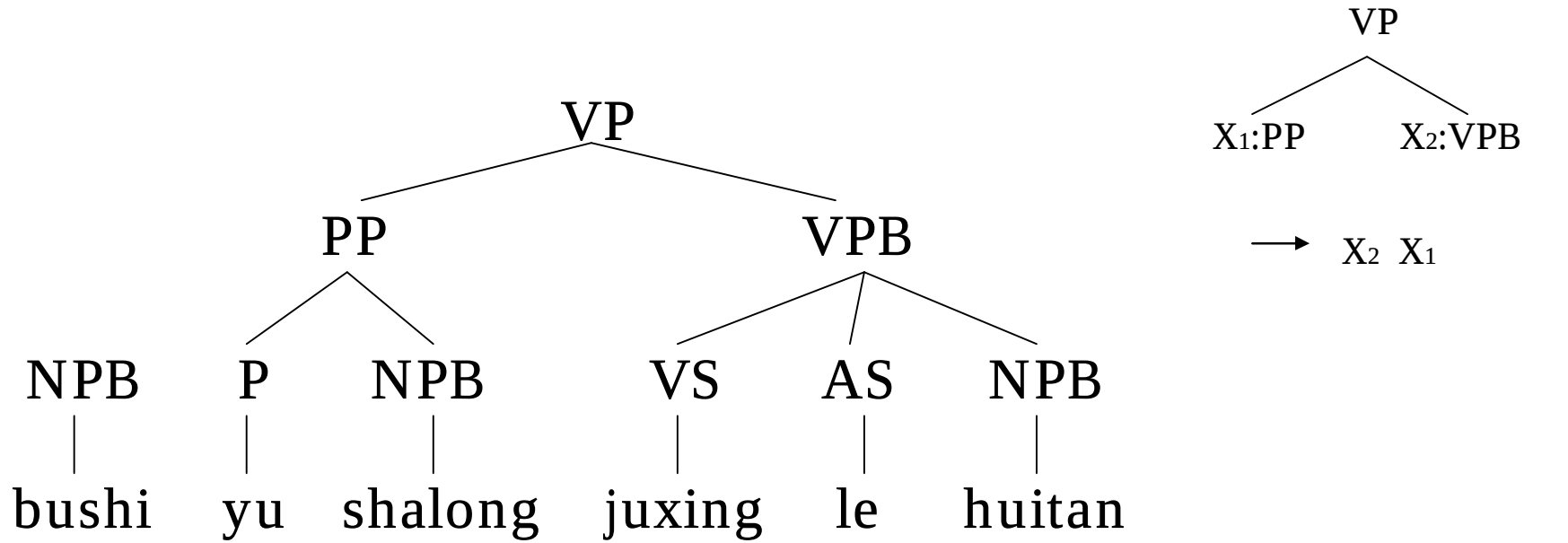
(Liu and Liu, 2010)

Tree-to-String Translation as Parsing



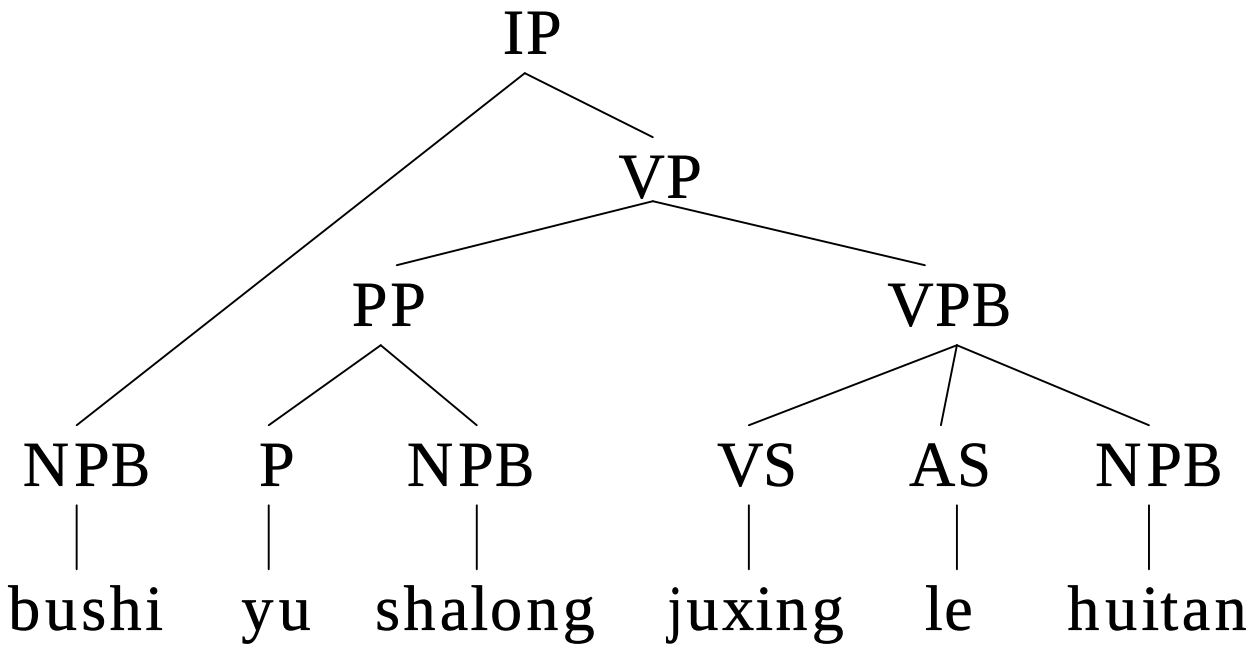
(Liu and Liu, 2010)

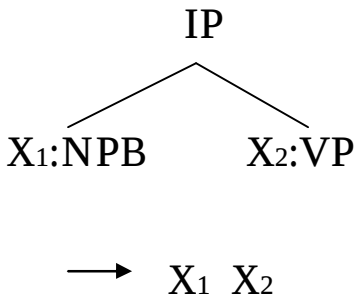
Tree-to-String Translation as Parsing



(Liu and Liu, 2010)

Tree-to-String Translation as Parsing

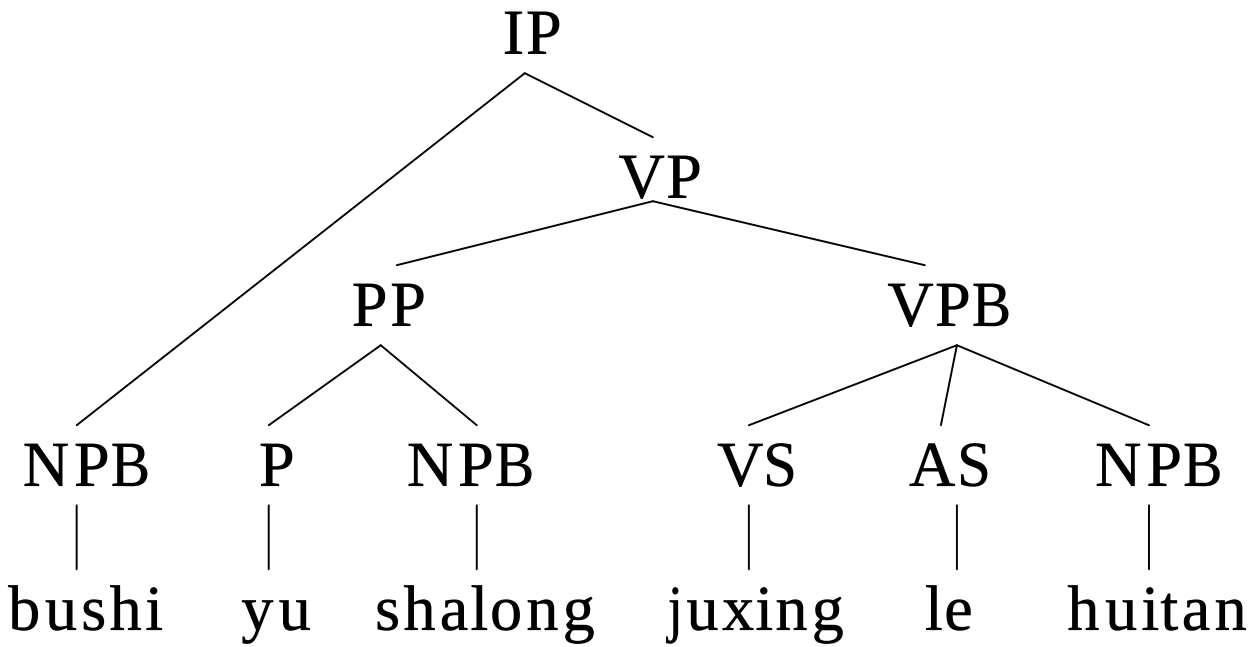


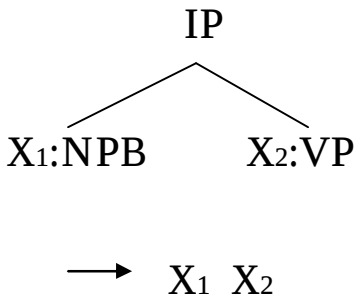


Bush held a talk with Sharon

(Liu and Liu, 2010)

Tree-to-String Translation as Parsing





Bush held a talk with Sharon

(Liu and Liu, 2010)

Translation Evaluation

algorithm	input	parsing model	rules	BLEU	time
matching	tree	none	1.2M	29.8	0.56
	forest	PCFG	1.9M	31.6	9.49
parsing	string	none	7.7M	32.0	51.41
		PCFG		32.4	55.52
		Lex		32.6	89.35
		PCFG+Lex		32.7	91.72

(Liu and Liu, 2010)

Parsing Evaluation

parsing model	F1	time
none	62.7	23.9
PCFG	65.4	24.7
Lex	79.8	48.8
PCFG + Lex	80.6	50.4

(Liu and Liu, 2010)

Results on Tree-to-Tree

task	extraction	rules	features	BLEU
Chinese	string-to-string	440M	1K	23.7
	tree-to-tree	50M	5K	23.9
Arabic	string-to-string	790M	1K	48.9
	tree-to-tree	38M	5K	47.5

Results from (Chiang, 2010)

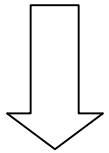
Outline

- *Part 1: Tree-based Translation*
  - *Overview and Motivation*
  - *Tree-to-String Model and Decoding*
  - *Tree-to-String Rule Extraction*
  - *Language Model-Integrated Decoding: Cube Pruning*
- *Part 2: Forest-based Translation*
  - *Packed Forest*
  - *Forest-based Decoding*
  - *Forest-based Rule Extraction*
- *Part 3: Extensions*
  - *Tree-to-Tree Translation*
  - *Tree Sequence-based Translation*
  - *Joint Parsing and Translation*
- Part 4: Conclusion

Conclusion

- Statistical machine translation
  - Word-based
  - Phrase-based
  - Syntax-based
    - String-to-String
    - String-to-Tree
    - Tree-to-String
    - Tree-to-Tree

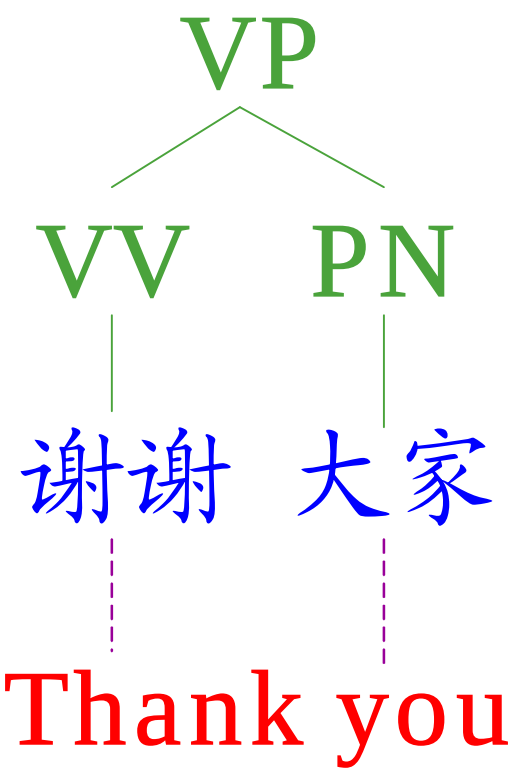
flat



hierarchical

Conclusion

- Tree-based translation
  - Pros: simplicity, faster decoding, expressive grammar, no need for binarization
  - Cons: commits to 1-best tree
- Forest-based translation
  - Compromise between tree-based and string-based, combining the advantages of both
    - Fast decoding, but does not commit to 1-best trees
    - Significant improvement of translation performance over tree-based


VP
VV
PN
谢谢
大家
Thank you

Bibliography

- Alfred V. Aho and Jeffrey D. Ullman. 1972. The Theory of Parsing, Translation, and Compiling, volume I: parsing. Prentice Hall, Englewood Cliffs, New Jersey.
- Peter F. Brown, Stephan A. Della Pietra, Vincent J. Della Pietra, and Robert L. Mercer. 1993. The mathematics of statistical machine translation: Parameter estimation. ***Computational Linguistics***, 19(2): 263-311.
- Sylvie Billot and Bernard Lang. 1989. The structure of shared forests in ambiguous parsing. In P***roceedings of ACL 1989***.
- David Chiang, 2005. A hierarchical phrase-based model for statistical machine translation. In ***Proceedings of ACL 2005***.
- David Chiang, 2007. Hierarchical phrase-based translation. ***Computational Linguistics***, 33(2): 201-228.
- David Chiang, 2010. Learning to translate with source and target syntax. In ***Proceedings of ACL 2010***.
- Trevor Cohn and Phil Blunsom. 2009. A bayesian model for syntax-directed tree to string grammar induction. In ***Proceedings of EMNLP 2009***.
- Michel Galley, Mark Hopkins, Kevin Knight, and Daniel Marcu. 2004. What's in a translation rule? In ***Proceedings of HLT-NAACL 2004***.
- Michel Galley, Jonathan Graehl, Kevin Knight, Daniel Marcu, Steve DeNeefe, Wei Wang, and Ignacio Thayer. 2006. Scalable inference and training of context-rich syntactic translation models. In ***Proceedings of COLING-ACL 2006***.

Bibliography

- Liang Huang and David Chiang. 2005. Better ***k***-best parsing. In ***Proceedings of IWPT 2005***.
- Liang Huang and David Chiang. 2007. Forest rescoring: Faster decoding with integrated language models. In ***Proceedings of ACL 2007***.
- Liang Huang, Kevin Knight, and Aravind Joshi. 2006. Statistical syntax-directed translation with extended domain of locality. In ***Proceedings of AMTA 2006***.
- Liang Huang. 2008. Forest reranking: Discriminative parsing with non-local features. In ***Proceedings. of ACL-HLT 2008***.
- E. T. Irons. 1961. A syntax-directed compiler for ALGOL 60. ***Comm. ACM***, 4(1): 51-55.
- Kevin Knight and Jonathan Graehl. 2005. An overview of probabilistic tree transducers for natural language processing. In ***Proceedings of CICLing 2005***.
- Dan Klein and Christopher D. Manning. 2001. Parsing and hypergraphs. In ***Proceedings of IWPT 2001***.
- Philipp Koehn, Franz Och, and Daniel Marcu. 2003. Statistical phrase-based translation In ***Proceedings of HLT-NAACL 2003***.
- P. M. Lewis and R. E. Stearns. 1968. Syntax-directed transduction. ***Journal of the ACM***, 15(3): 465-488.
- Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-string alignment template for statistical machine translation. In ***Proceedings of COLING-ACL 2006***.

Bibliography

- Yang Liu, Yun Huang, Qun Liu, and Shouxun Lin. 2007. Forest-to-string statistical translation rules. In ***Proceedings of ACL 2007***.
- Yang Liu, Yajuan Lu, and Qun Liu. 2009a. Improving tree-to-tree translation with packed forests. In ***Proceedings of ACL-IJCNLP 2009***.
- Yang Liu, Haitao Mi, Yang Feng, and Qun Liu. 2009b. Joint decoding with multiple translation models. In ***Proceeding of ACL-IJCNLP 2009***.
- Yang Liu and Qun Liu. 2010. Joint parsing and translation. Submitted to ***COLING 2010***.
- Daniel Marcu, Wei Wang, Abdessamad Echihabi, and Kevin Knight. 2006. SPMT: Statistical machine translation with syntactified target language phrases. In ***Proceedings of EMNLP 2006***.
- Jonathan May and Kevin Knight. 2007. Syntax re-alignment models for machine translation. In ***Proceedings of EMNLP 2007***.
- Haitao Mi and Liang Huang. 2008. Forest-based translation rule extraction. In ***Proceedings of EMNLP 2008***.
- Haitao Mi, Liang Huang, and Qun Liu. 2008. Forest-based translation. In ***Proceedings of ACL-HLT 2008***.
- Franz Och and Hermann Ney. 2004. The alignment template approach to statistical machine translation. ***Computational Linguistics***, 30(4): 417-449.

Bibliography

- Libin Shen, Jinxi Xu, and Ralph Weischedel. 2008. A new string-to-dependency machine translation algorithm with a target dependency language model. In Proceedings of ACL-HLT 2008.
- Ashish Venugopal, Andreas Zollmann, Noah Smith, and Stephan Vogel. 2008. Wider pipelines: ***n***-best alignments and parses in mt training. In ***Proceedings of AMTA 2008***.
- Wei Wang, Kevin Knight, and Daniel Marcu. 2007. Binarizing syntax trees to improve syntax-based machine translation accuracy. In ***Proceedings of EMNLP 2007***.
- Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical machine translation model. In ***Proceedings of ACL 2001***.
- Hui Zhang, Min Zhang, Haizhou Li, Aiti Aw, and Chew Lin Tan. 2009. Forest-based tree sequence to string translation model. In ***Proceedings of ACL-IJCNLP 2009***.
- Min Zhang, Hongfei Jiang, Aiti Aw, Jun Sun, Sheng Li, and Chew Lin Tan. 2007. A tree-to-tree alignment-based model for statistical machine translation. ***In Proceedings of MT Summit 2007***.
- Min Zhang, Hongfei Jiang, Aiti Aw, Haizhou Li, Chew Lin Tan, and Sheng Li. 2008. A tree sequence alignment-based tree-to-tree translation model. In ***Proceedings of ACL-HLT 2008***.