中科院计算所
INSTITUTE OF COMPUTING TECHNOLOGY

Source Syntax-based Statistical Machine Translation Models and Approaches

Qun Liu
Institute of Computing Technology, CAS

Sino-Japanese Machine Translation Technology Cooperation Symposium
2011-09-26 Beijing, China

Outline

Background

Tree-to-String Model

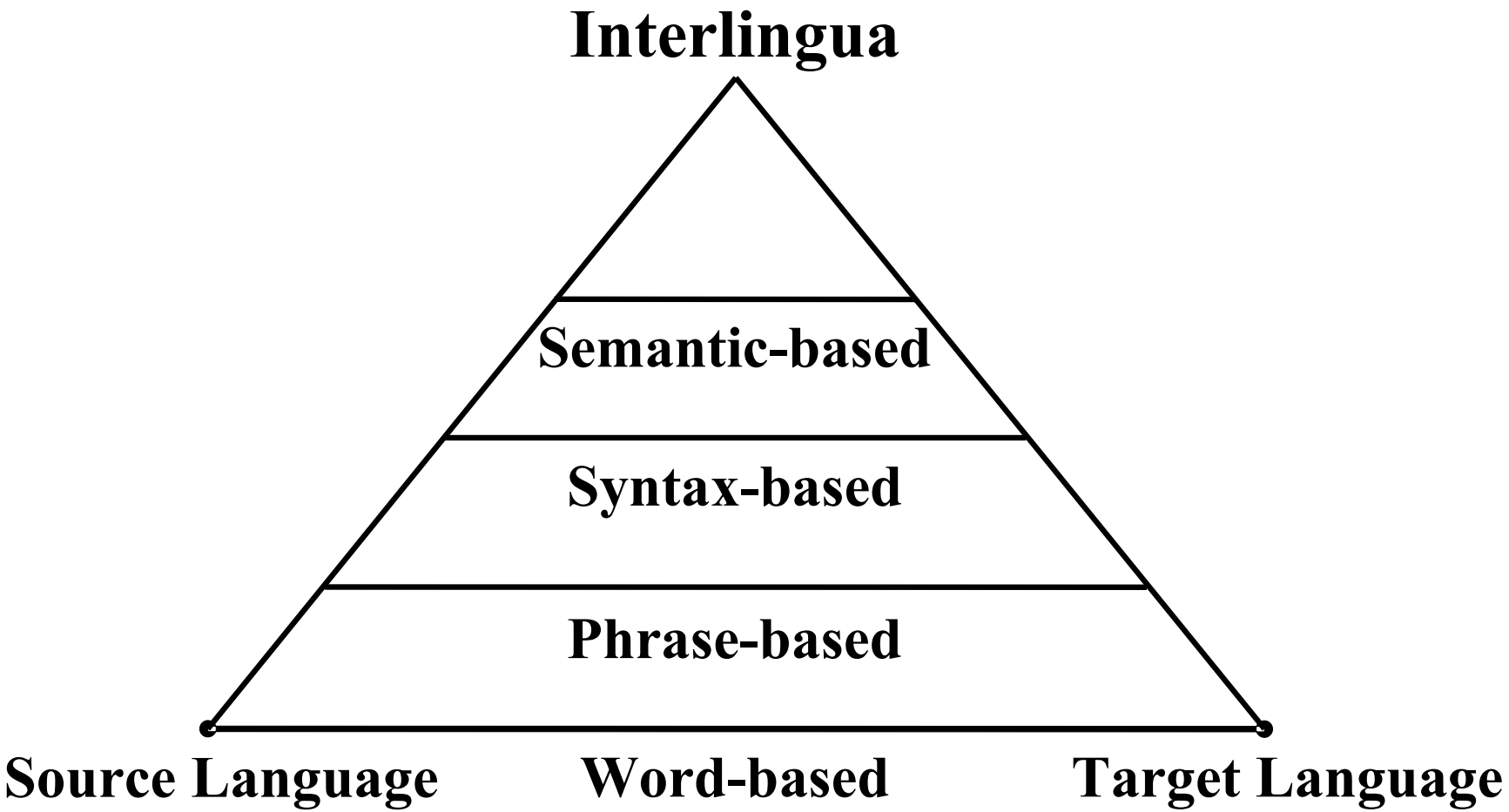
Conclusion

Statistical Translation Model

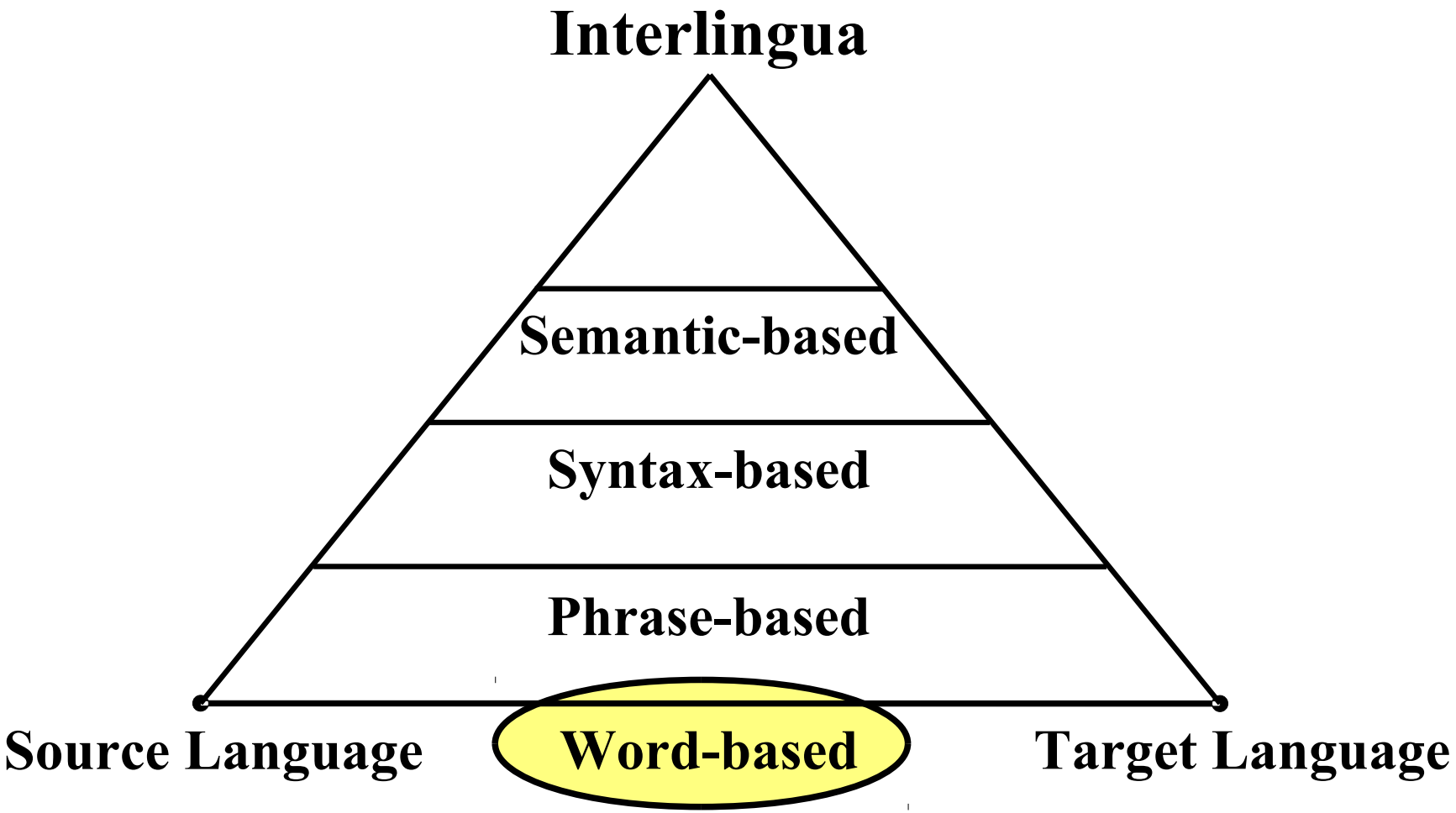
$$P(E|F)$$

$$\sum_{E} P(E|F) = 1$$

Translation Models



Translation Models



An Example

布什 与 沙龙 举行 了 会谈

bushi yu shalong juxing le huitan

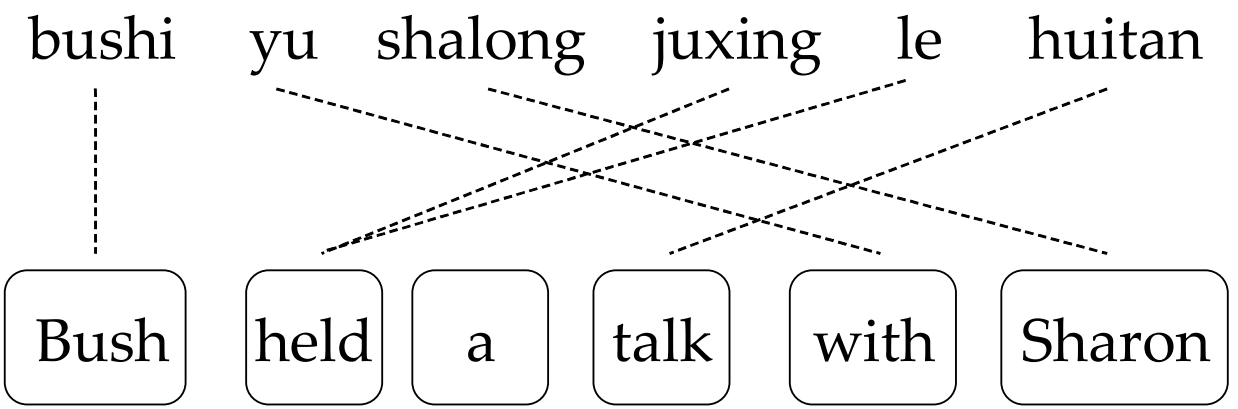
Bush held a talk with Sharon

Word-based Model

IBM Model 1-5

- Peter F. Brown, Stephen 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.

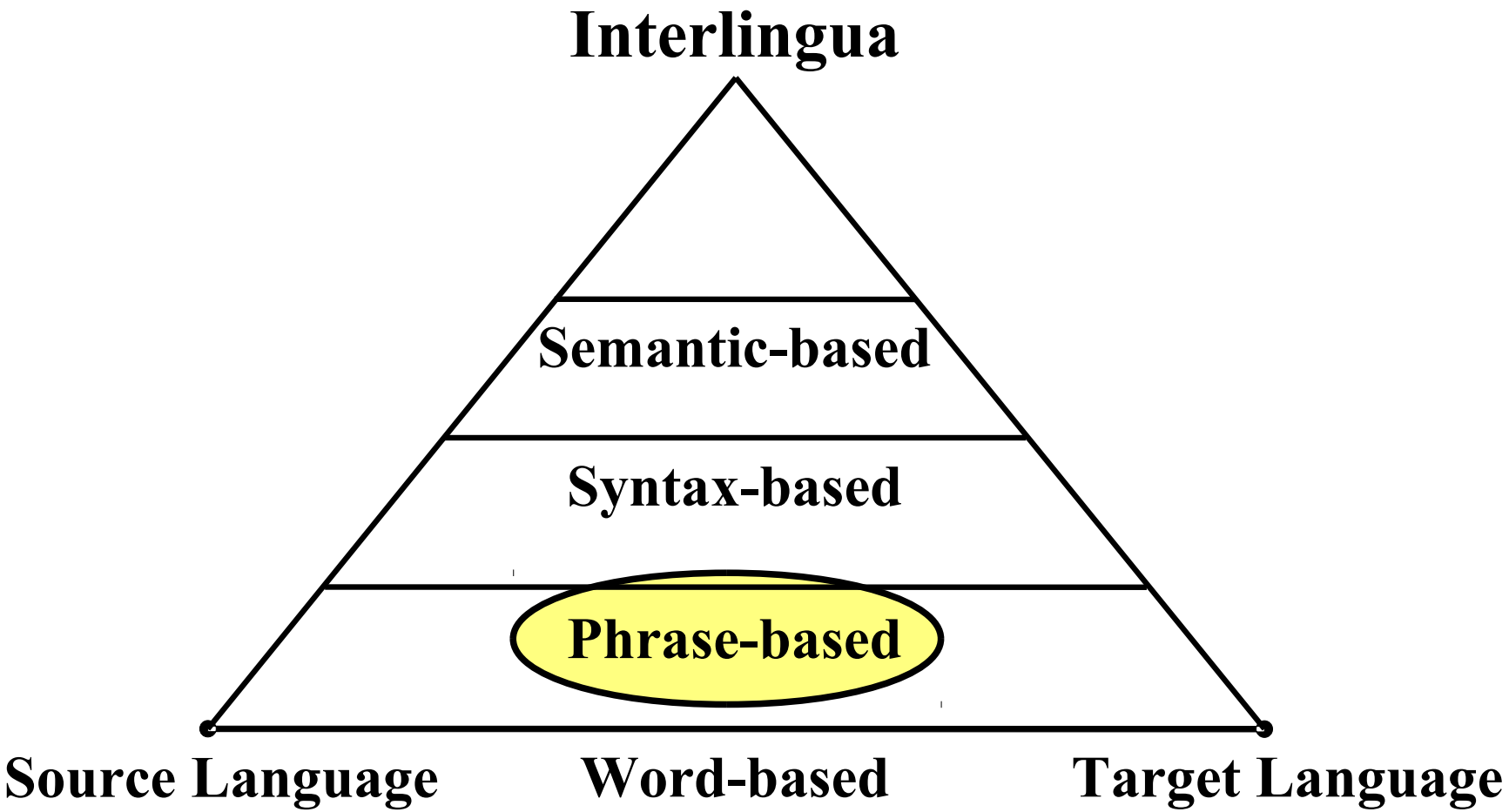
Word-based Model



Word-based Model

Source	Target	Probability
Bushi (布什)	Bush	0.7
	President	0.2
	US	0.1
yu (与)	and	0.6
	with	0.4
juxing (举行)	hold	0.7
	had	0.3
le (了)	hold	0.01

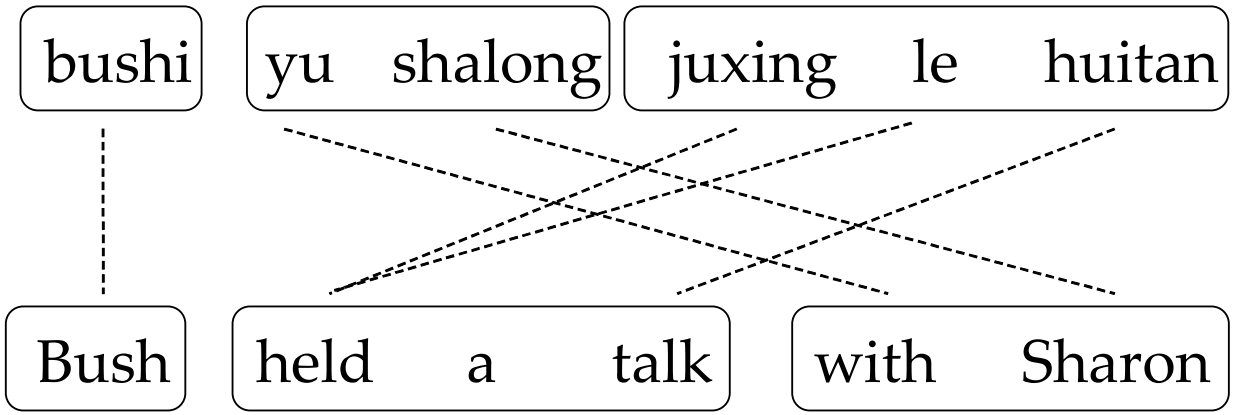
Translation Models



Phrase-based Model

- Franz J. Och and Hermann Ney. 2004. The Alignment Template Approach to Statistical Machine Translation. Computational Linguistics, 30(4):417-449.
- Philipp Koehn, Franz J. Och, and Daniel Marcu. 2003. Statistical Phrase-Based Translation. In Proceedings of the Human Language Technology and North American Association for Computational Linguistics Conference, pages 127-133, Edmonton, Canada, May.

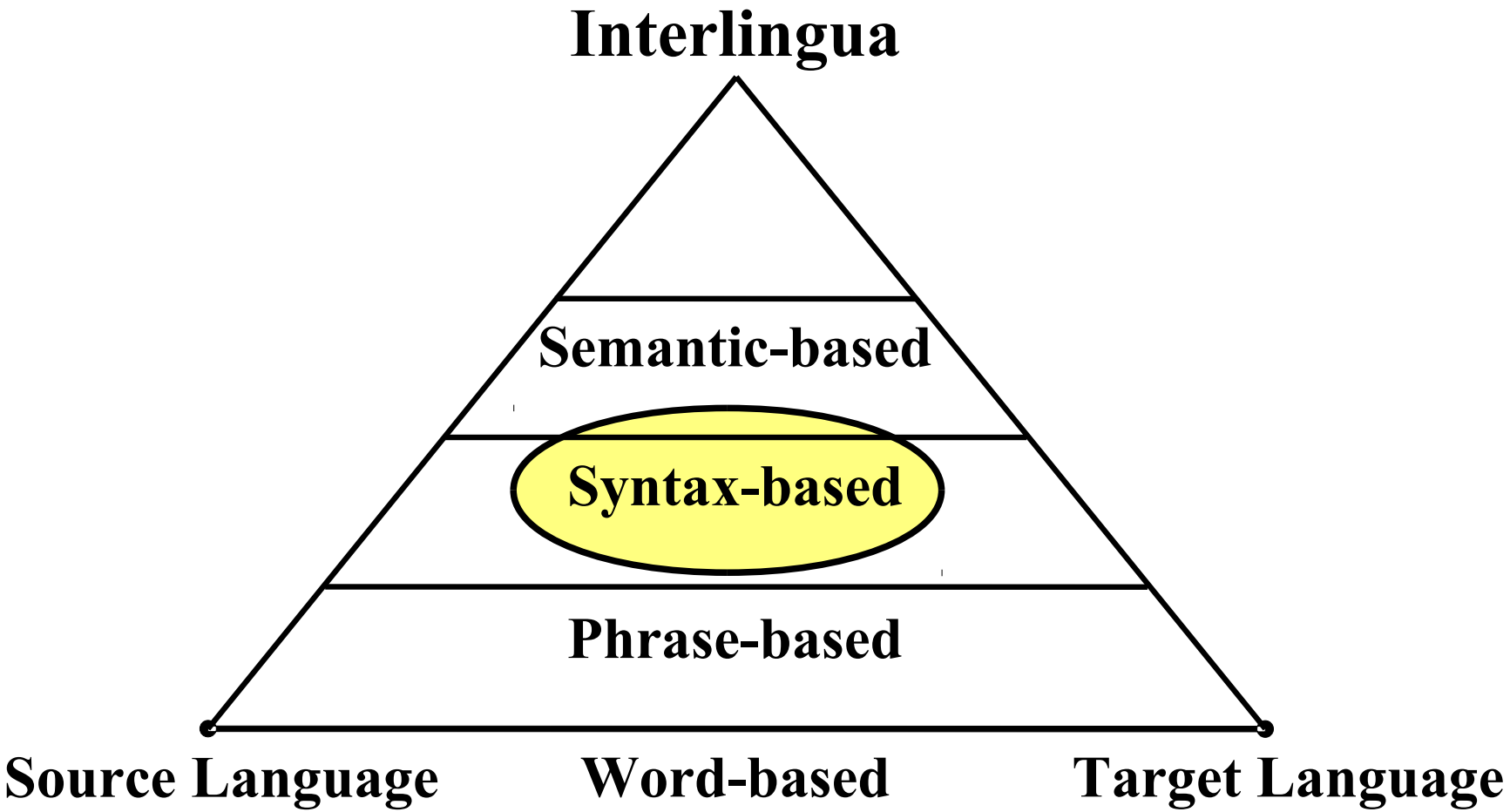
Phrase-based Model



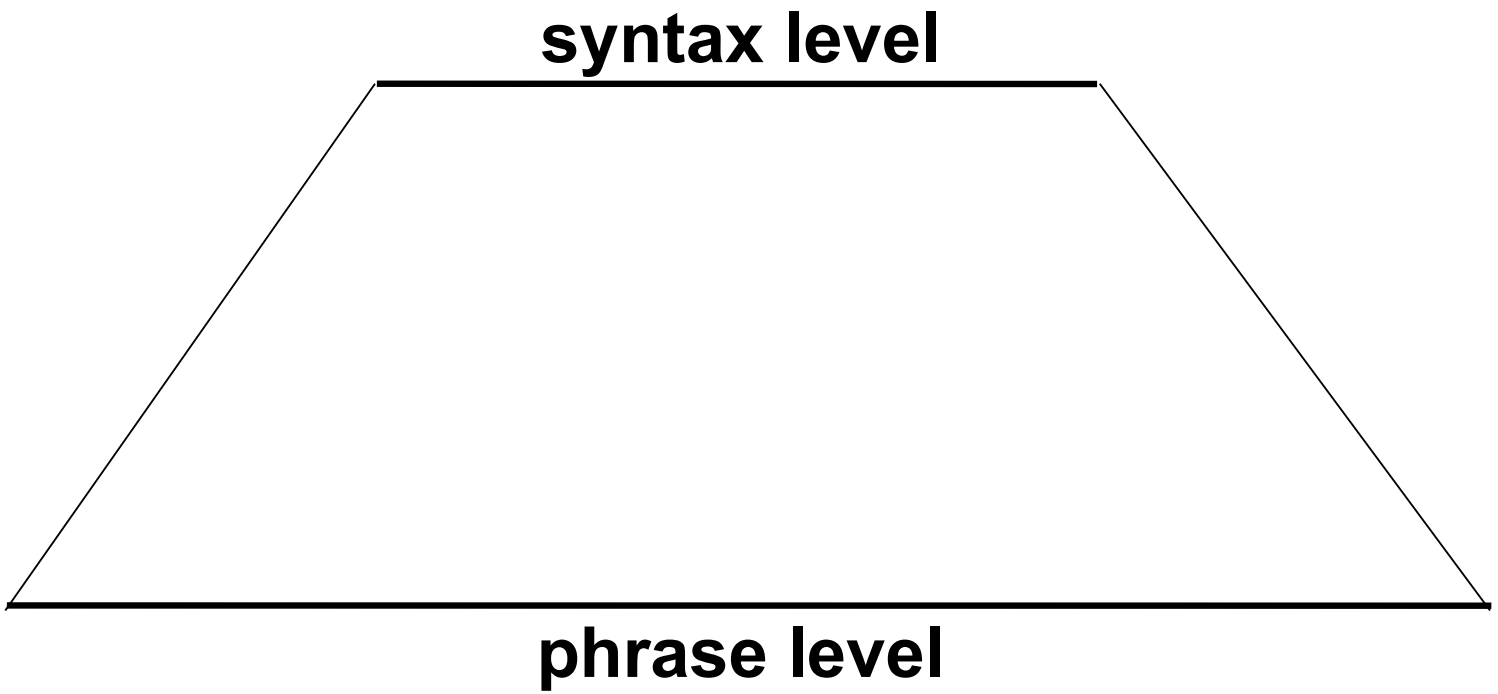
Phrase-based Model

Source	Target	Probability
Bushi （布什）	Bush	0.5
	president Bush	0.3
	the US president	0.2
Bushi yu （布什与）	Bush and	0.8
	the president and	0.2
yu Shalong （与沙龙）	and Shalong	0.6
	with Shalong	0.4
juxing le huiang （举行了会谈）	hold a meeting	0.7
	had a meeting	0.3

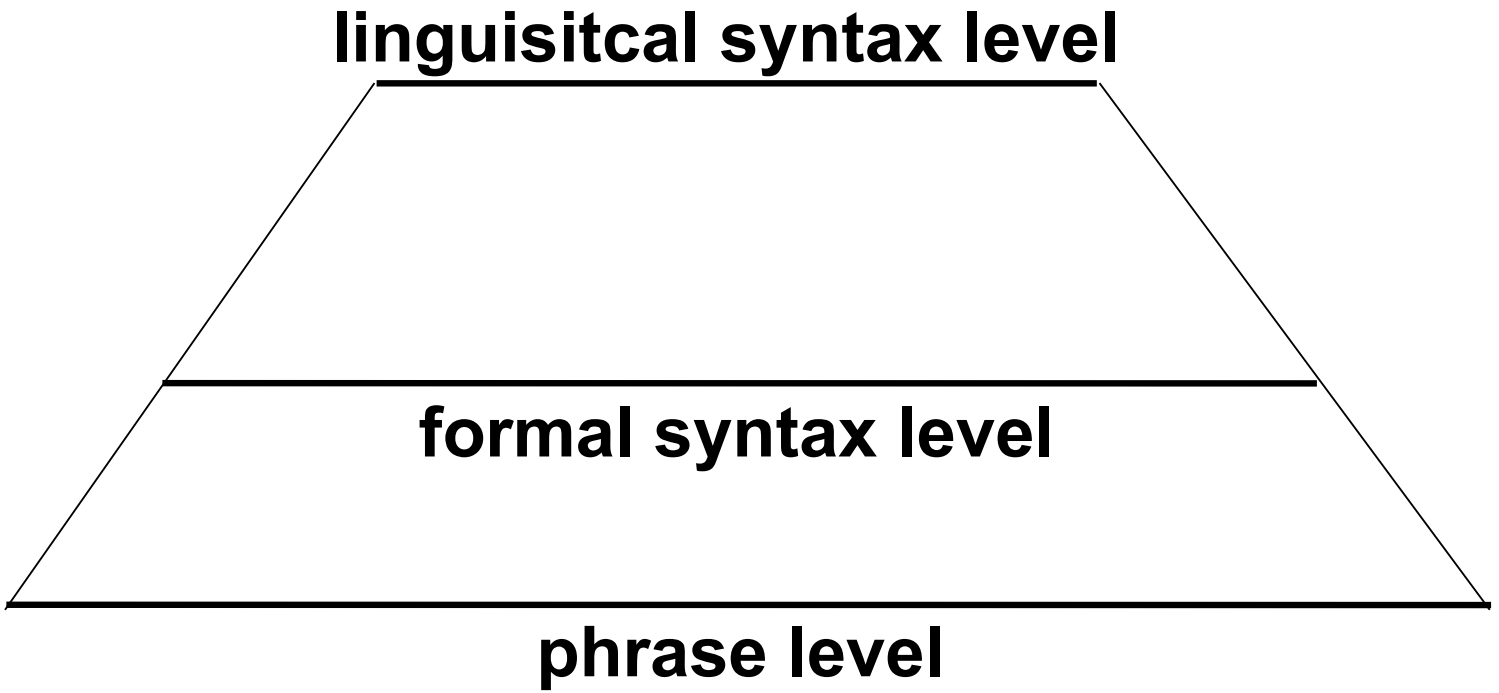
Translation Models



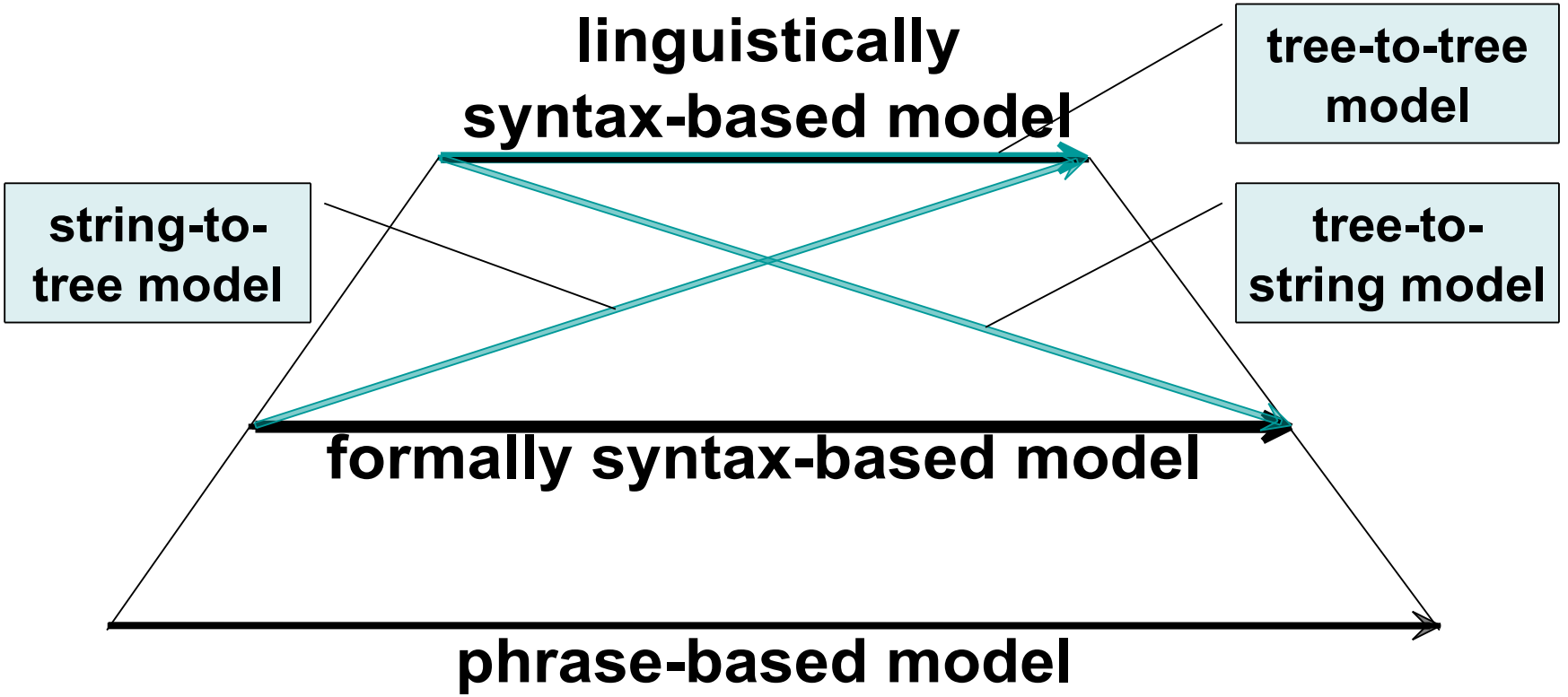
Syntax-based Translation Models #1



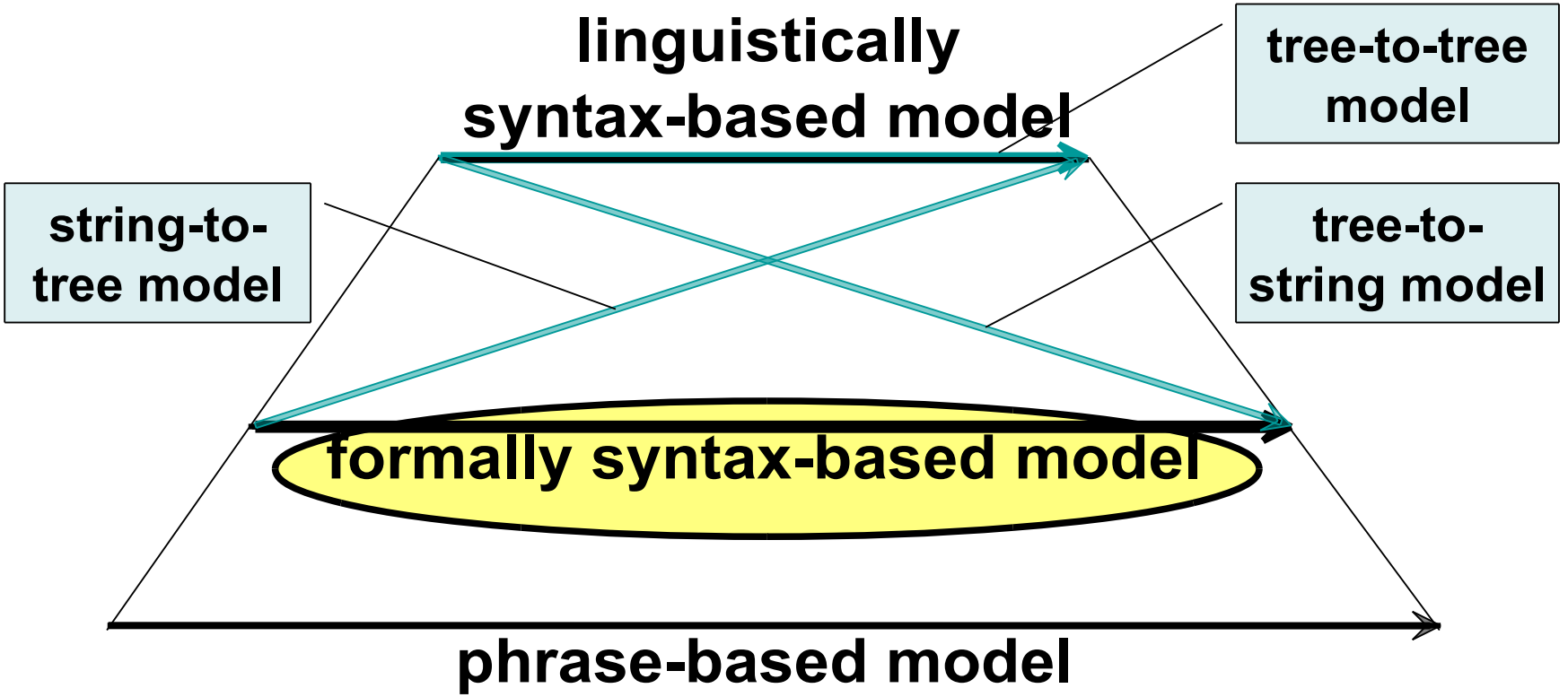
Syntax-based Translation Models #2



Syntax-based Translation Models #3



Syntax-based Translation Models #3



Formally Syntax-based Model

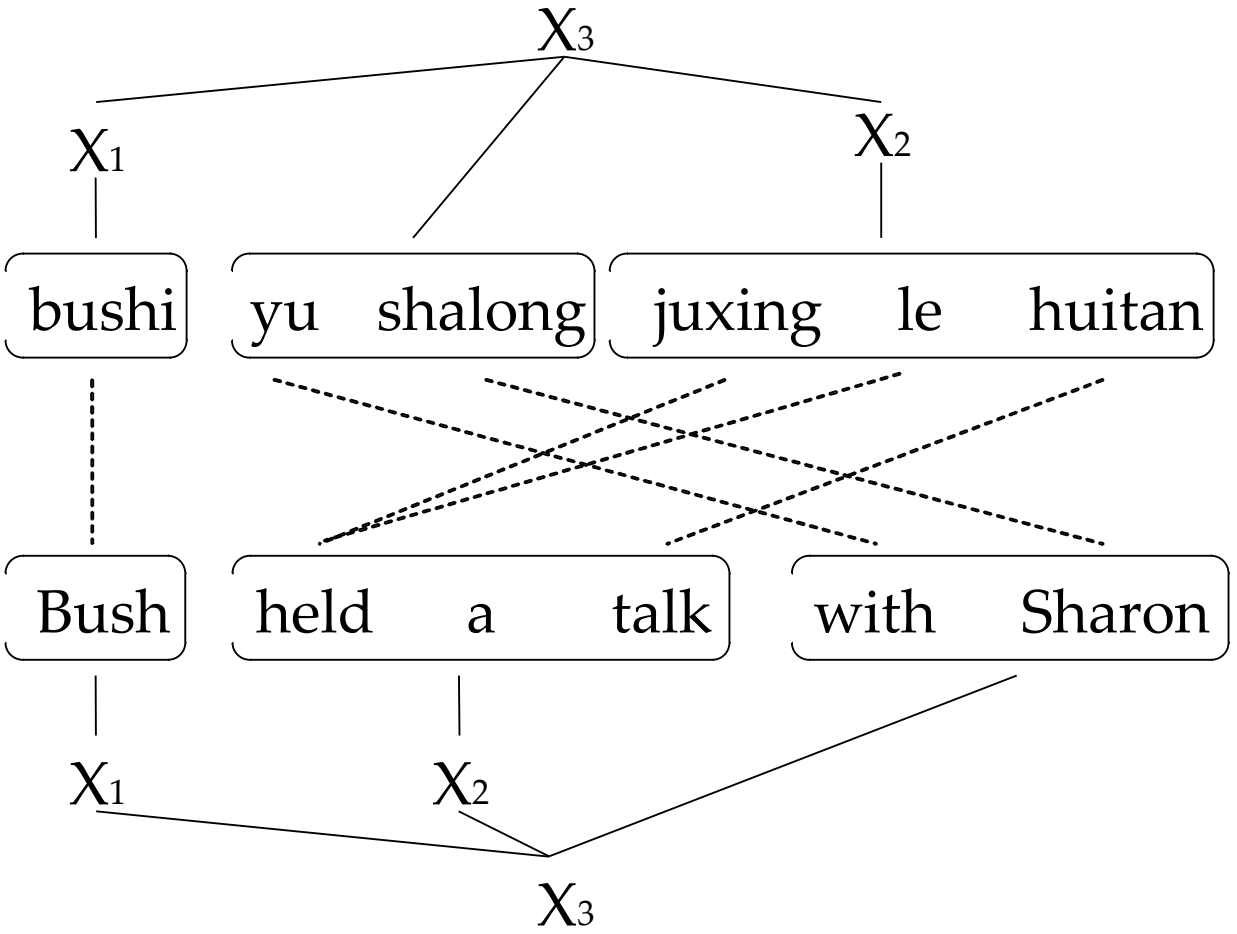
Hierarchical Phrase-based Model

- David Chiang, 2005. A hierarchical phrase-based model for statistical machine translation. In Proceedings of ACL 2005.

Maximum Entropy Bracketing Transduction Grammar Model

- Deyi Xiong, Qun Liu, and Shouxun Lin. Maximum Entropy Based Phrase Reordering Model for Statistical Machine Translation. COLING-ACL 2006, Sydney, Australia, July 17-21.

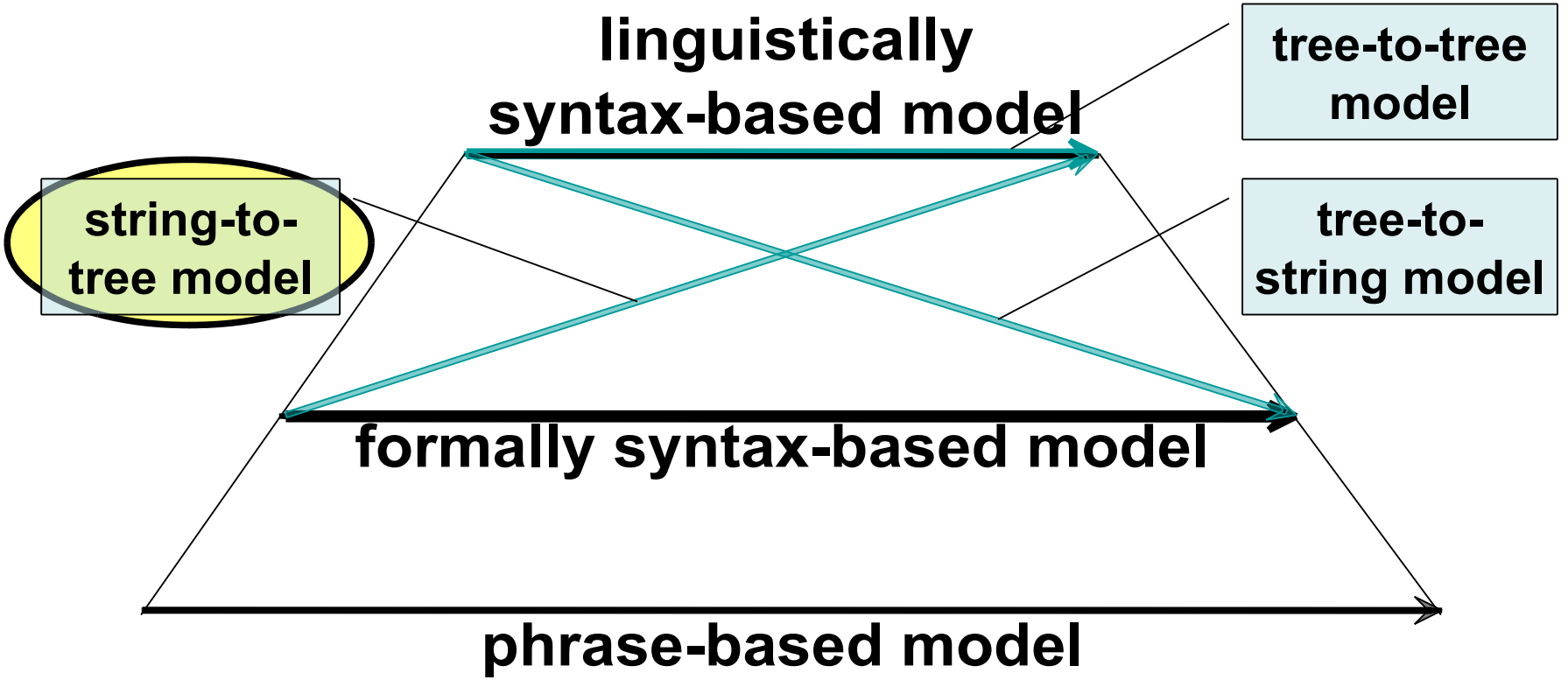
Hierarchical Phrased-based Model



Hierarchical Phrased-based Model

Source	Target	Probability
juxing le huiang （举行了会谈）	hold a meeting	0.6
	had a meeting	0.3
X huitang （X 会谈）	X a meeting	0.8
	X a talk	0.2
juxing le X （举行了 X）	hold a X	0.5
	had a X	0.5
Bushi yu Shalong （布什与沙龙）	Bush and Sharon	0.8
Bushi X （布什 X）	Bush X	0.7
X yu Y （X 与 Y）	X and Y	0.9

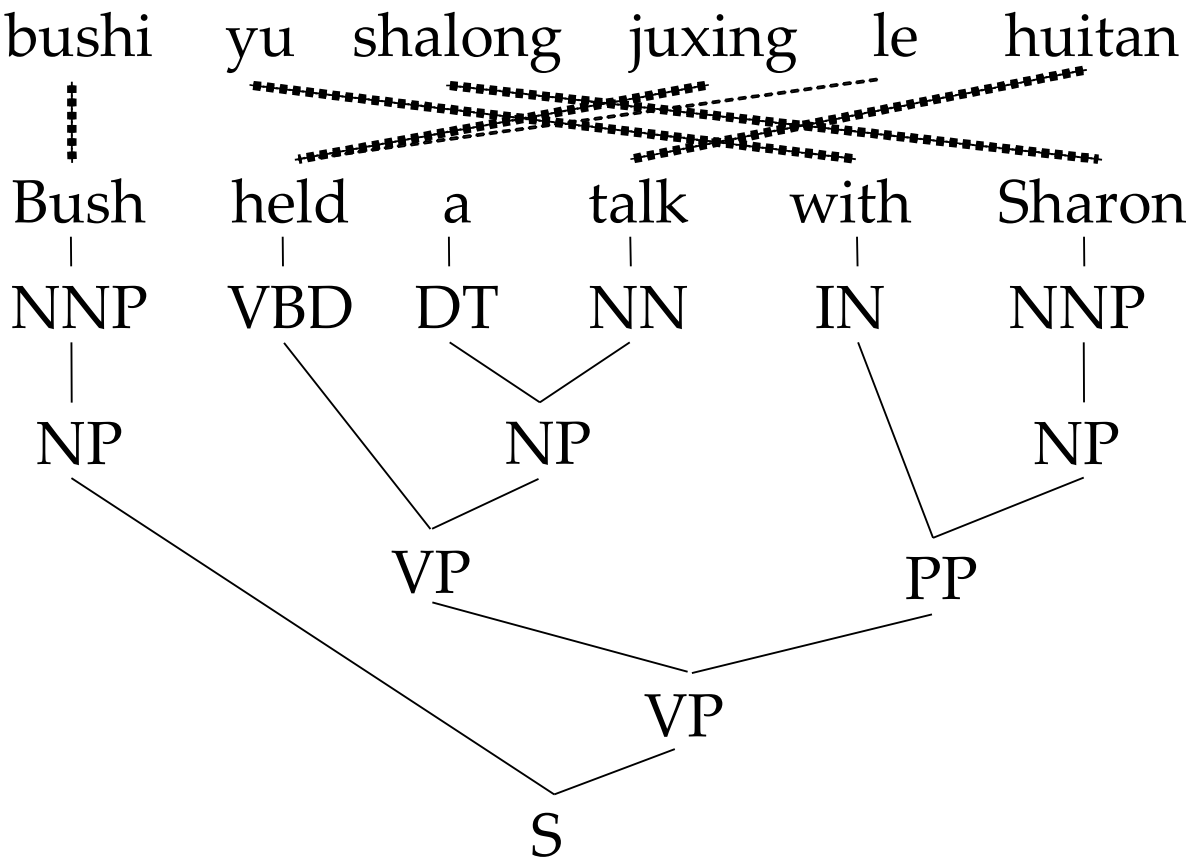
Syntax-based Translation Models #3



String-to-Tree Model

- Kenji Yamada and Kevin Knight. 2001. A syntax-based statistical machine translation model. In Proceedings of ACL 2001.
- Daniel Marcu, Wei Wang, Abdessamad Echihabi, and Kevin Knight. 2006. SPMT: Statistical machine translation with syntactified target language phrases. In Proceedings of EMNLP 2006.
- 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.

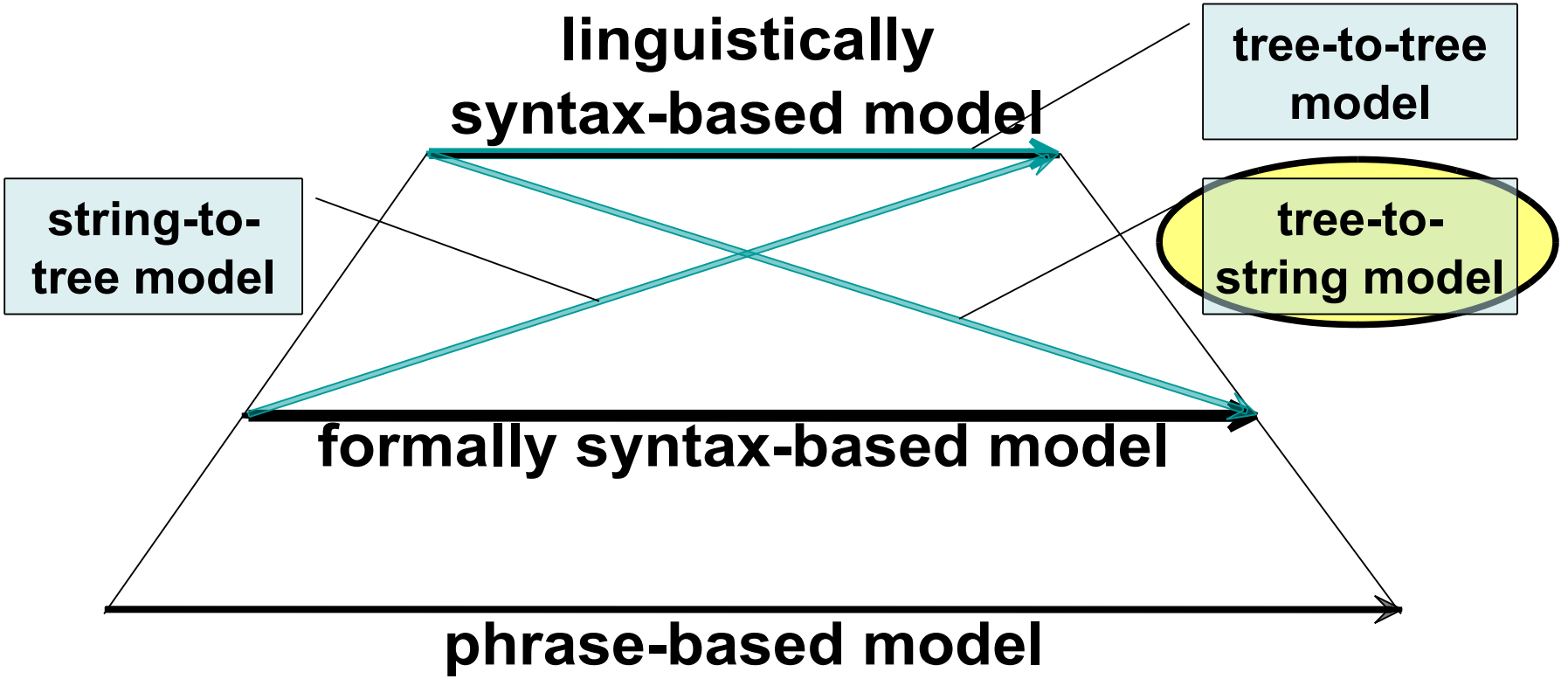
String-to-Tree Model



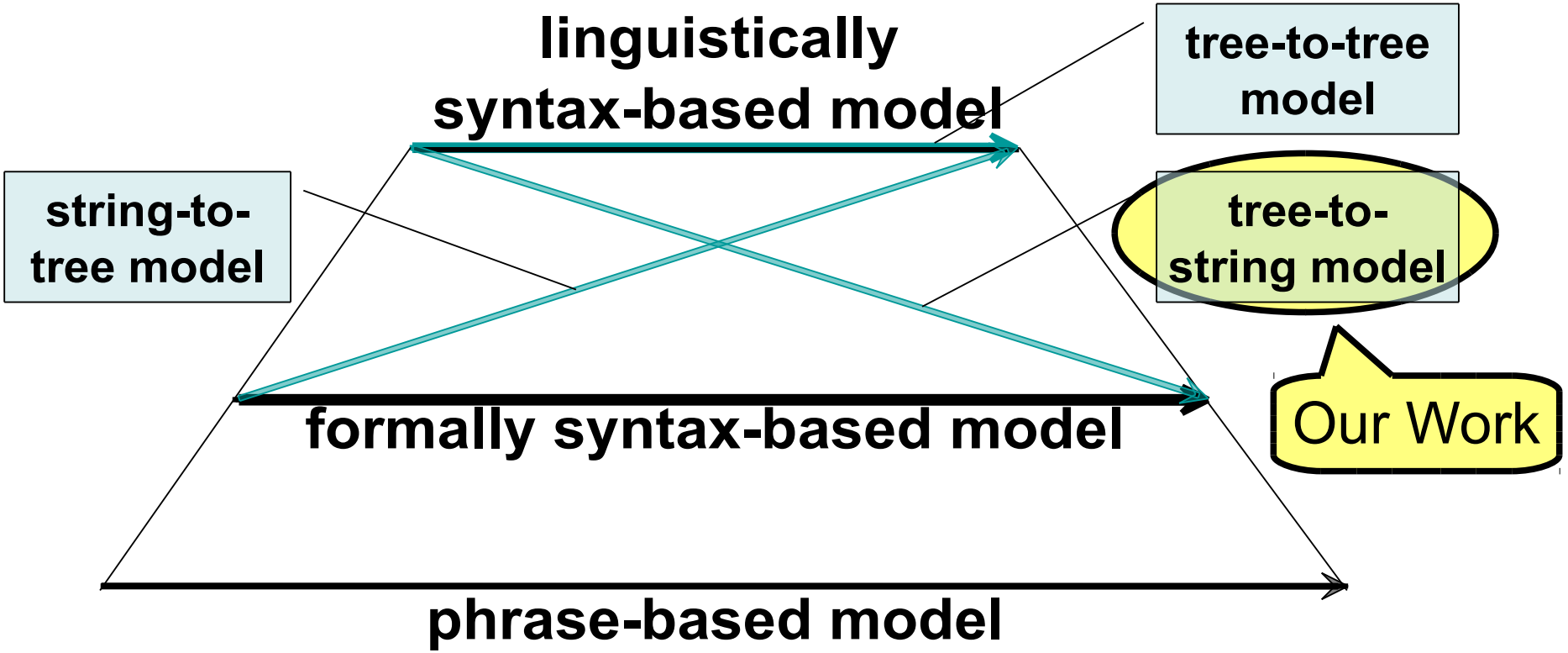
String-to-Tree Model

Source	Target	Probability
juxing le huiang （举行了会谈）	VP(VPD(hold) NP(DT(a) NN(meeting)))	0.6
	VP(VPD(had) NP(DP(a) NN(meeting)))	0.3
	VP(VPD(had) NP(DT(a) NN(talk)))	0.1
x_1 huitang （x_1 会谈）	VP(x_1:VPD NP(DT(a) NN(meeting)))	0.8
	VP(x_1:VPD NP(DT(a) NN(talk)))	0.2
juxing le x_1 （举行了 x_1）	VP(VPD(hold) NP(DT(a) x_1:NN))	0.5
	VP(VPD(had) NP(DT(a) x_1:NN))	0.5
x_1 yu x_2 （x_1 与 x_2）	NP(x_1:NNP CC(and) x_2:NNP))	0.9

Syntax-based Translation Models #3



Syntax-based Translation Models #3



Outline

Background

Tree-to-String Model

Conclusion

Our Work: Tree-to-String Model

Constituent-to-String Model

Our Work: Tree-to-String Model

Constituent-to-String Model

Tree-based Translation

Forest-based Translation

Joint Parsing and Translation

Our Work: Tree-to-String Model

Constituent-to-String Model

Tree-based Translation

Forest-based Translation

Joint Parsing and Translation

Dependency-to-String Model

Our Work: Tree-to-String Model

Constituent-to-String Model

Tree-based Translation

Forest-based Translation

Joint Parsing and Translation

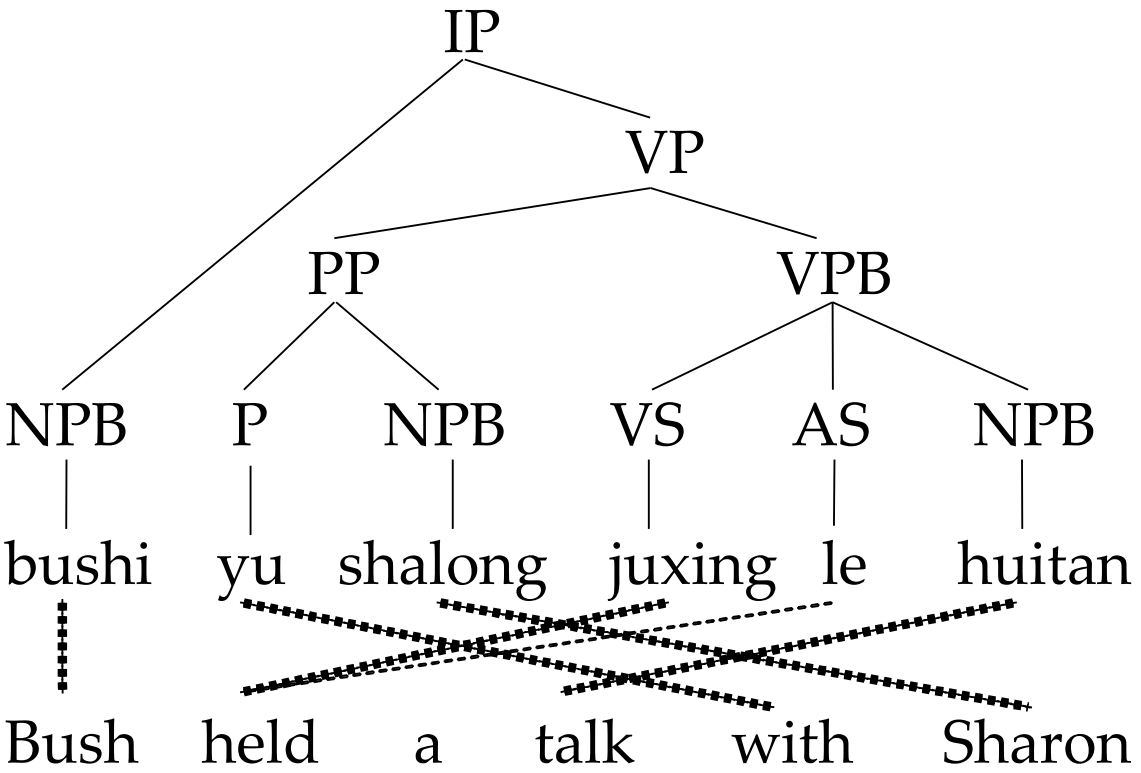
Dependency-to-String Model

Constituent-to-String Model

- Yang Liu, Qun Liu, and Shouxun Lin. 2006. Tree-to-String Alignment Template for Statistical Machine Translation. In Proceedings of COLING/ACL 2006, pages 609-616, Sydney, Australia, July.

  Meritorious Asian NLP Paper Award

Constituent-to-String Model



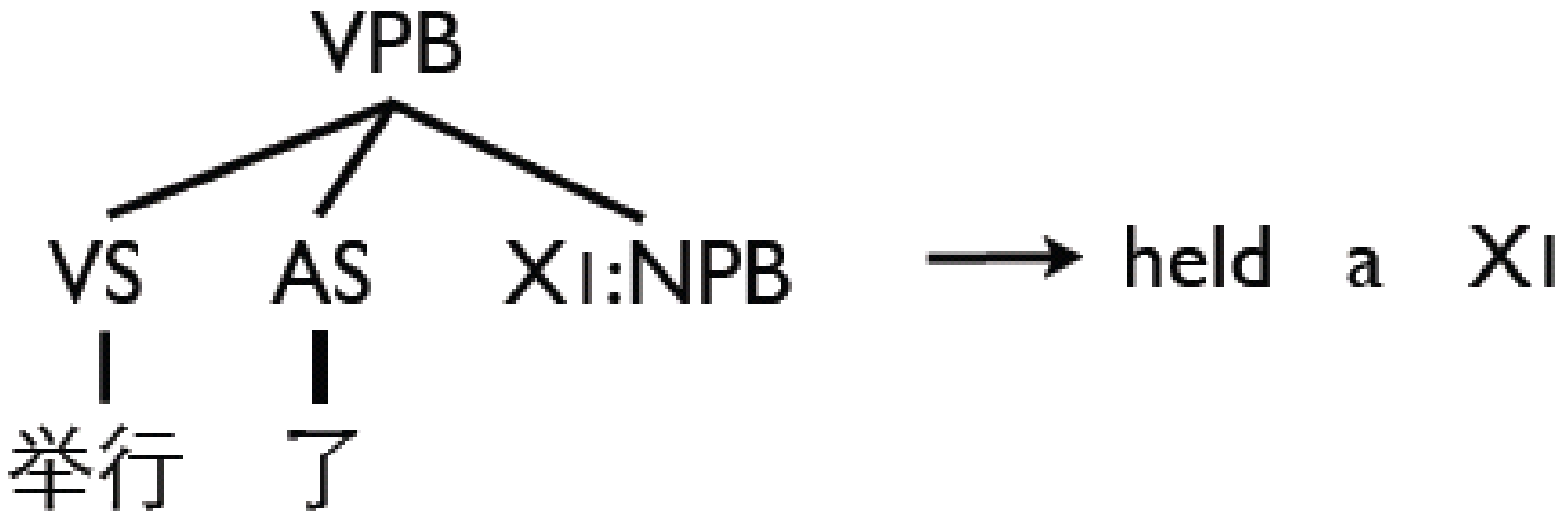
Constituent-to-String Model

Source	Target	Probability
VPB(VS(juxing) AS(le) NPB(huiang)) （举行了会谈）	hold a meeting	0.6
	have a meeting	0.3
	have a talk	0.1
VPB(VS(juxing) AS(le) x_1) （举行了 x_1 ）	hold a x_1	0.5
	have a x_1	0.5
VP(PP(P(yu) x_1:NPB) x_2:VPB) （与 x_1 x_2 ）	x_2 with x_1	0.9
IP(x_1:NPB VP(x_2:PP x_3:VPB))	x_1 x_3 x_2	0.7

Constituent-to-String Rule

- A constituent-to-string model is a statistical translation model built on constituent-to-string translation rules
- A constituent-to-string translation rules consist of:
  - A syntax subtree in source side, where the leaf nodes may be nonterminals or terminals (words)
  - A string of words and variables in target side
  - A one-to-one mapping between the nonterminal leaf nodes in source subtree and the variables in target string

Constituent-to-String Rule



VPB(VS(举行) AS(了) X1:NPB) → held a X1

Our Work: Tree-to-String Model

Constituent-to-String Model

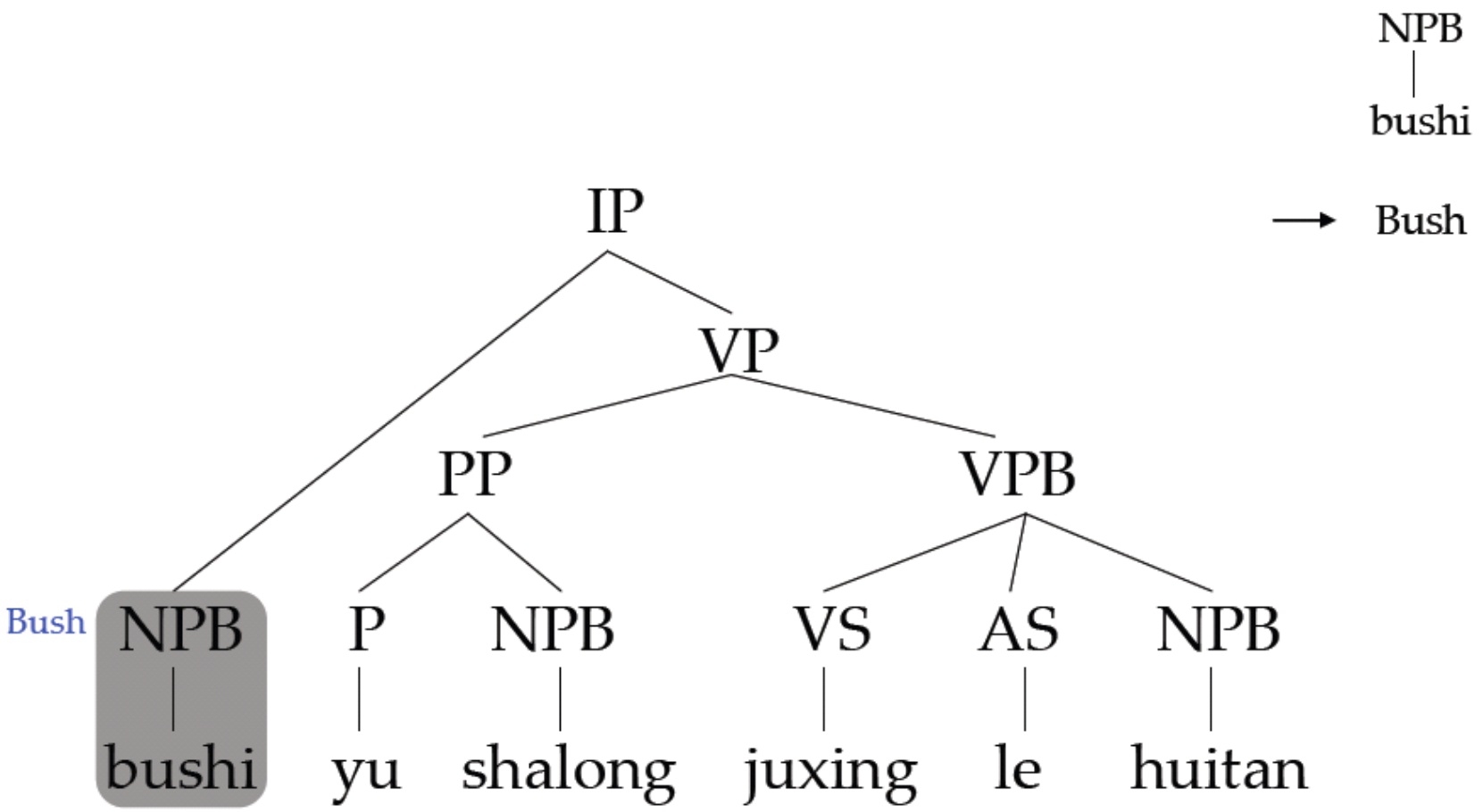
Tree-based Translation

Forest-based Translation

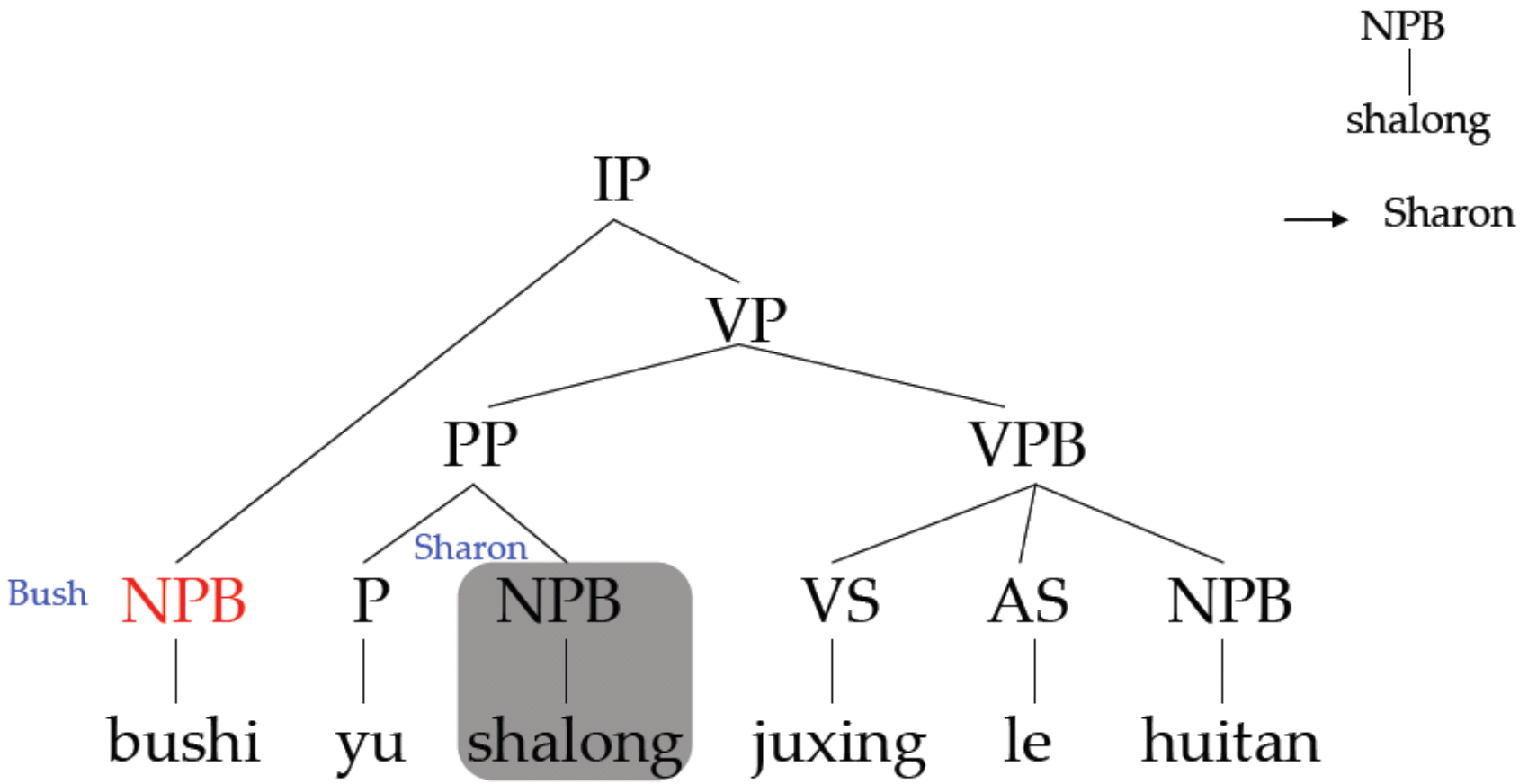
Joint Parsing and Translation

Dependency-to-String Model

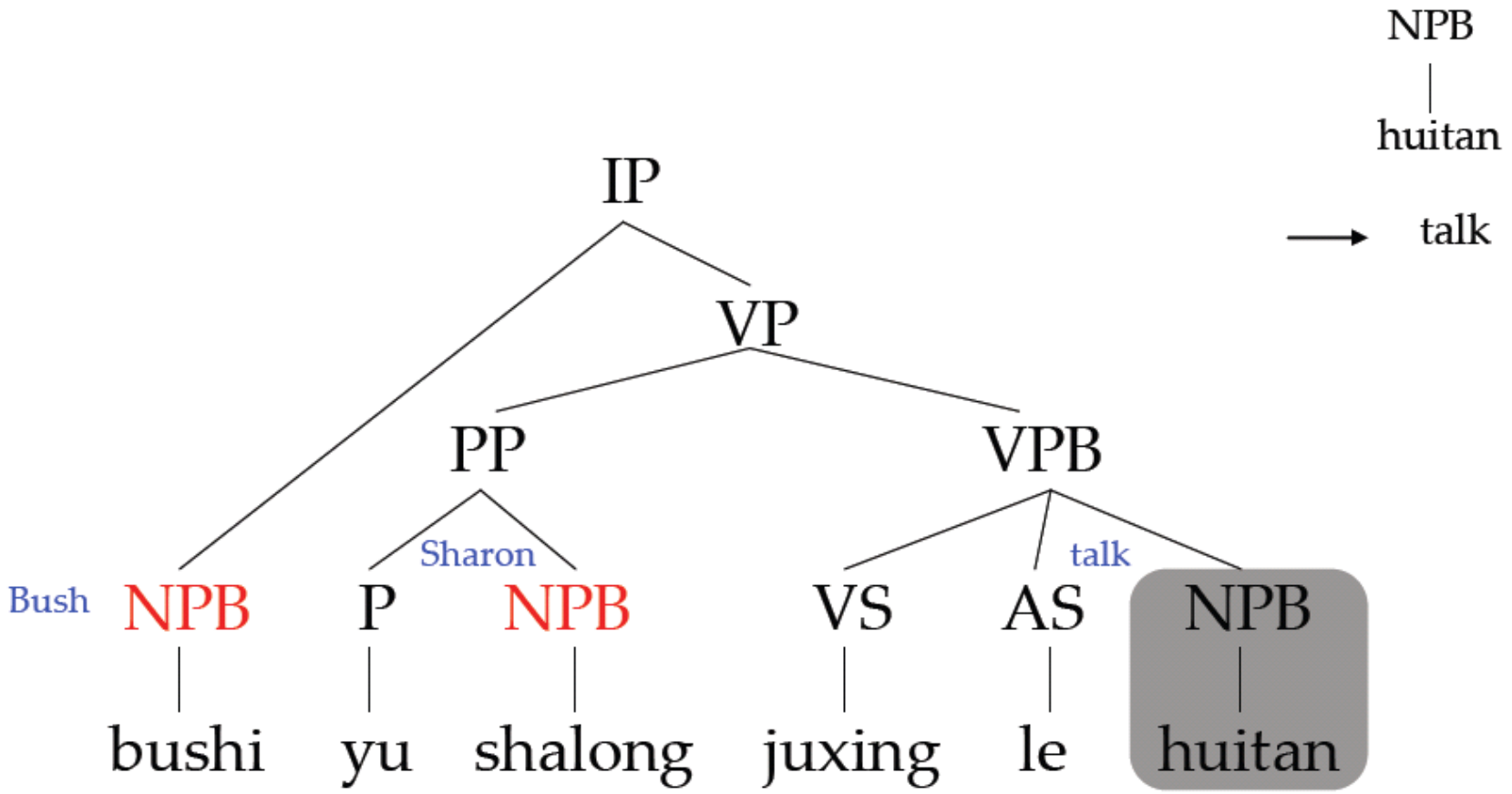
Tree-based Button-up Decoding



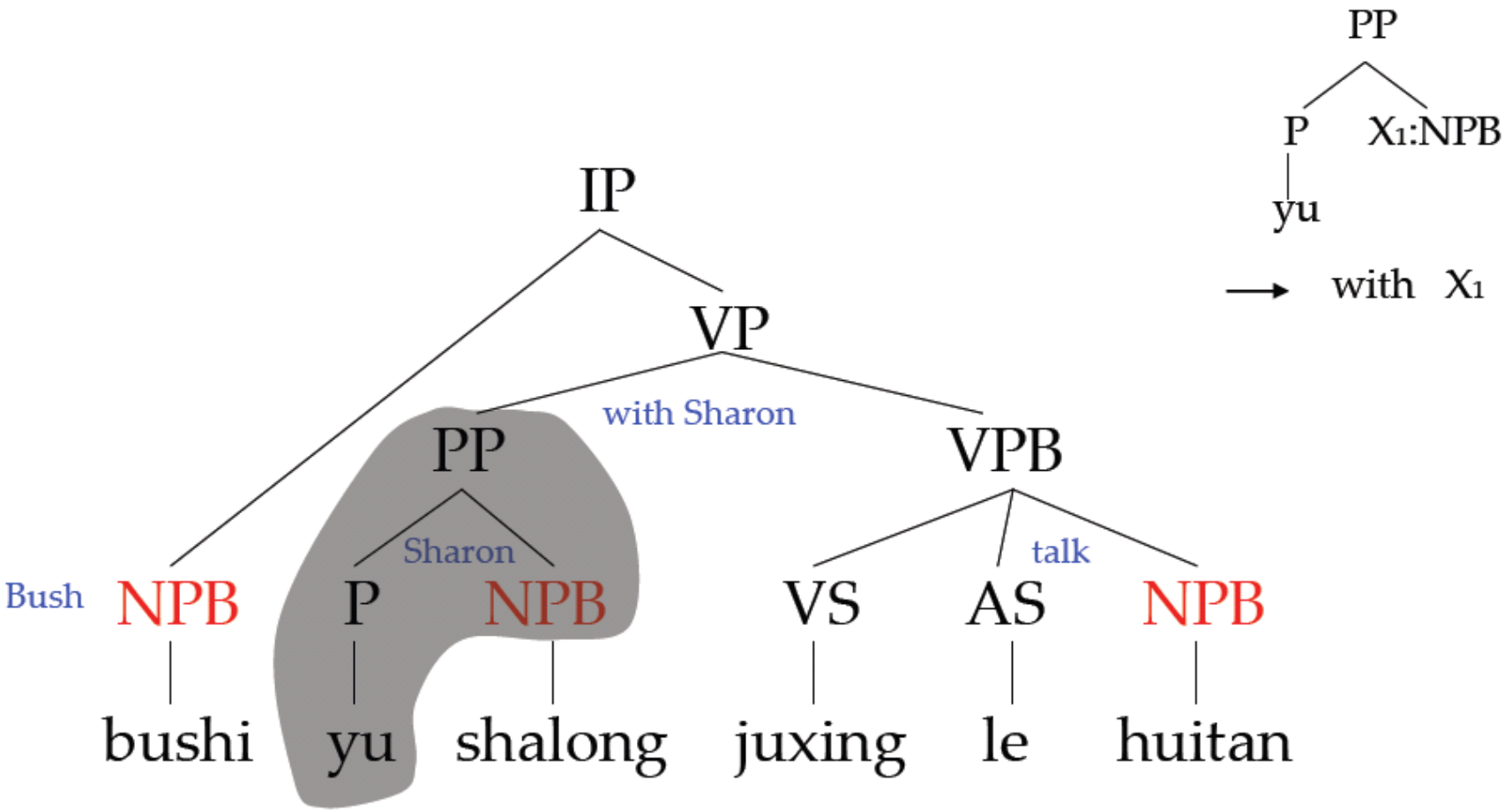
Tree-based Button-up Decoding



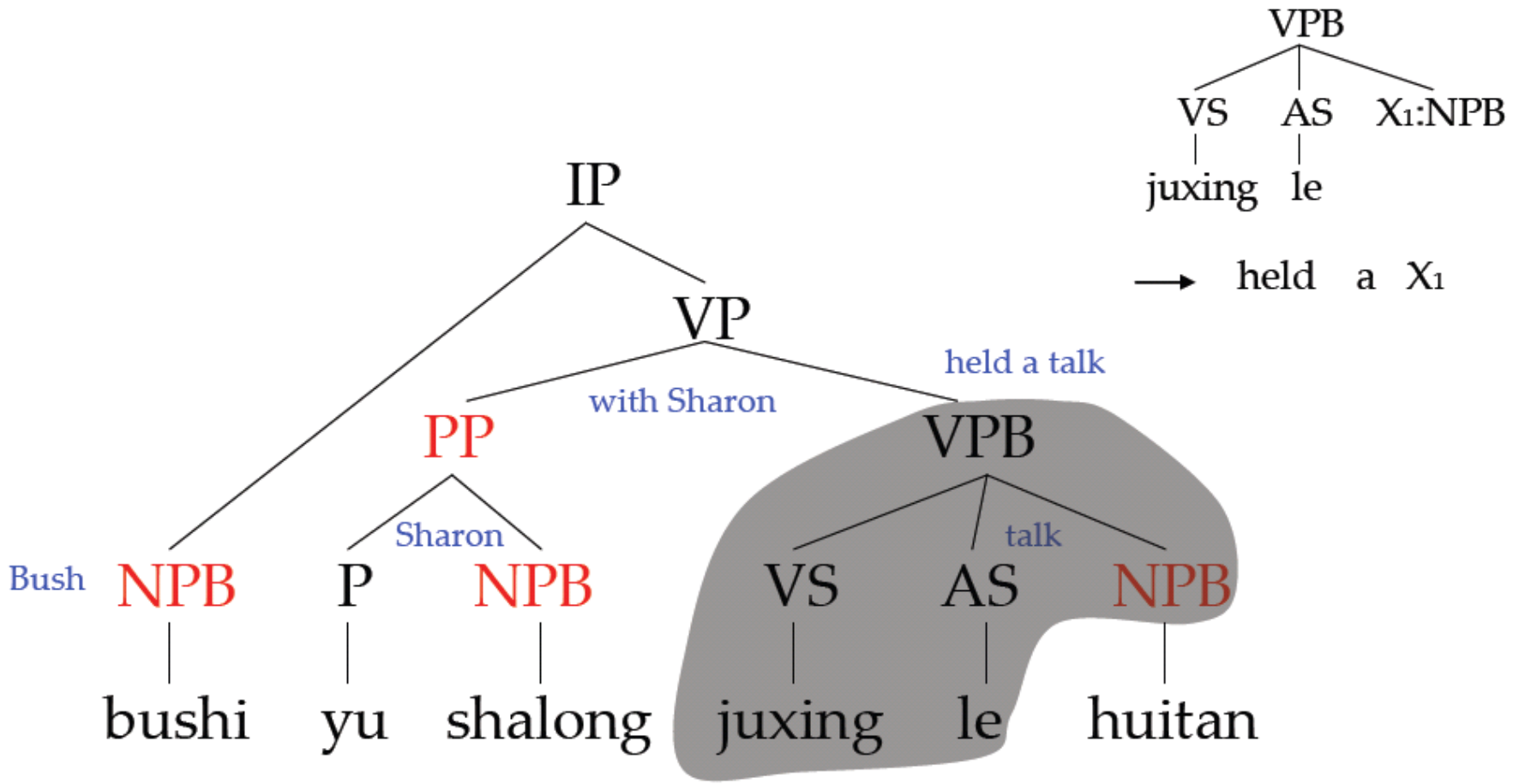
Tree-based Button-up Decoding



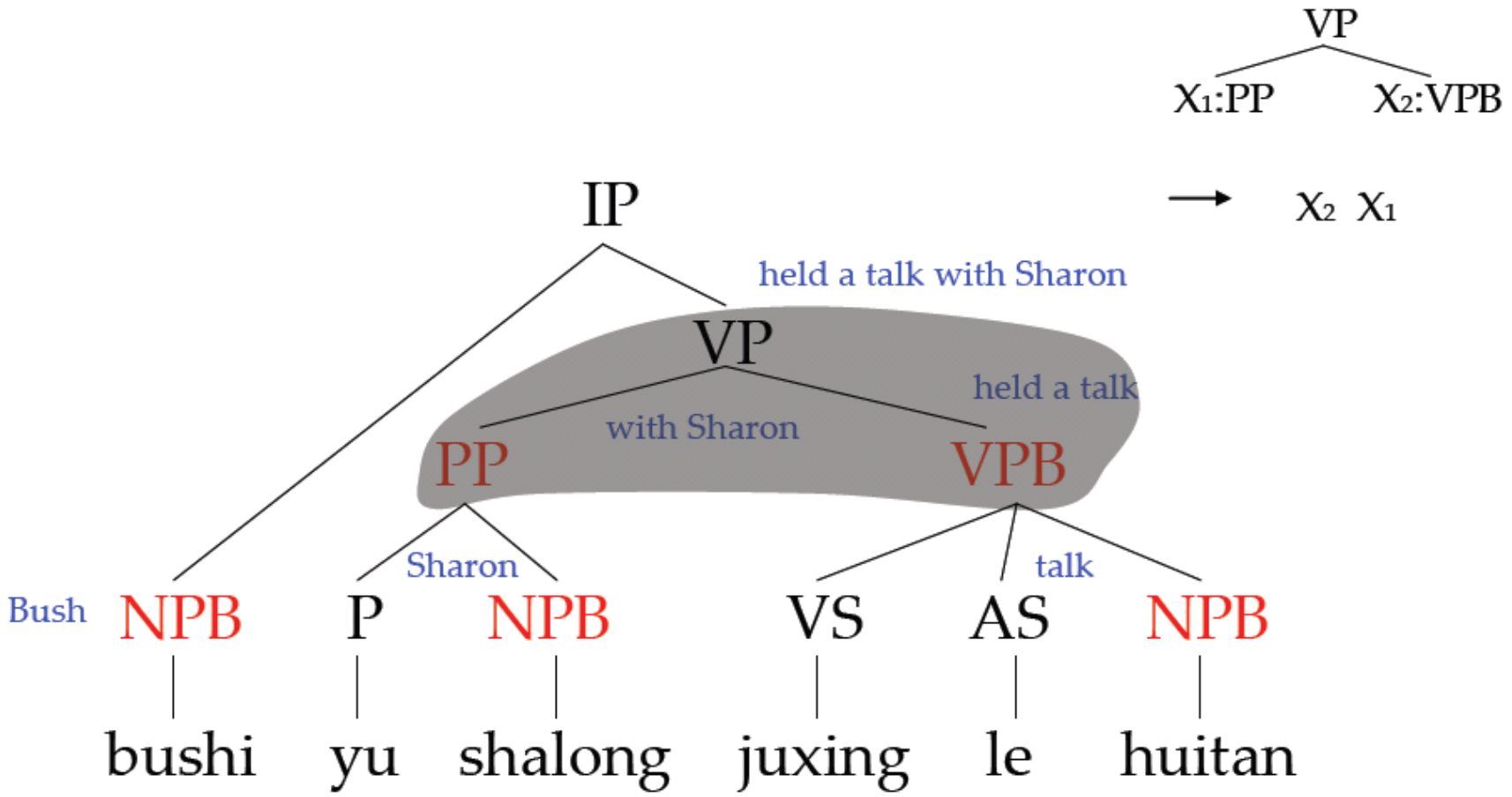
Tree-based Button-up Decoding



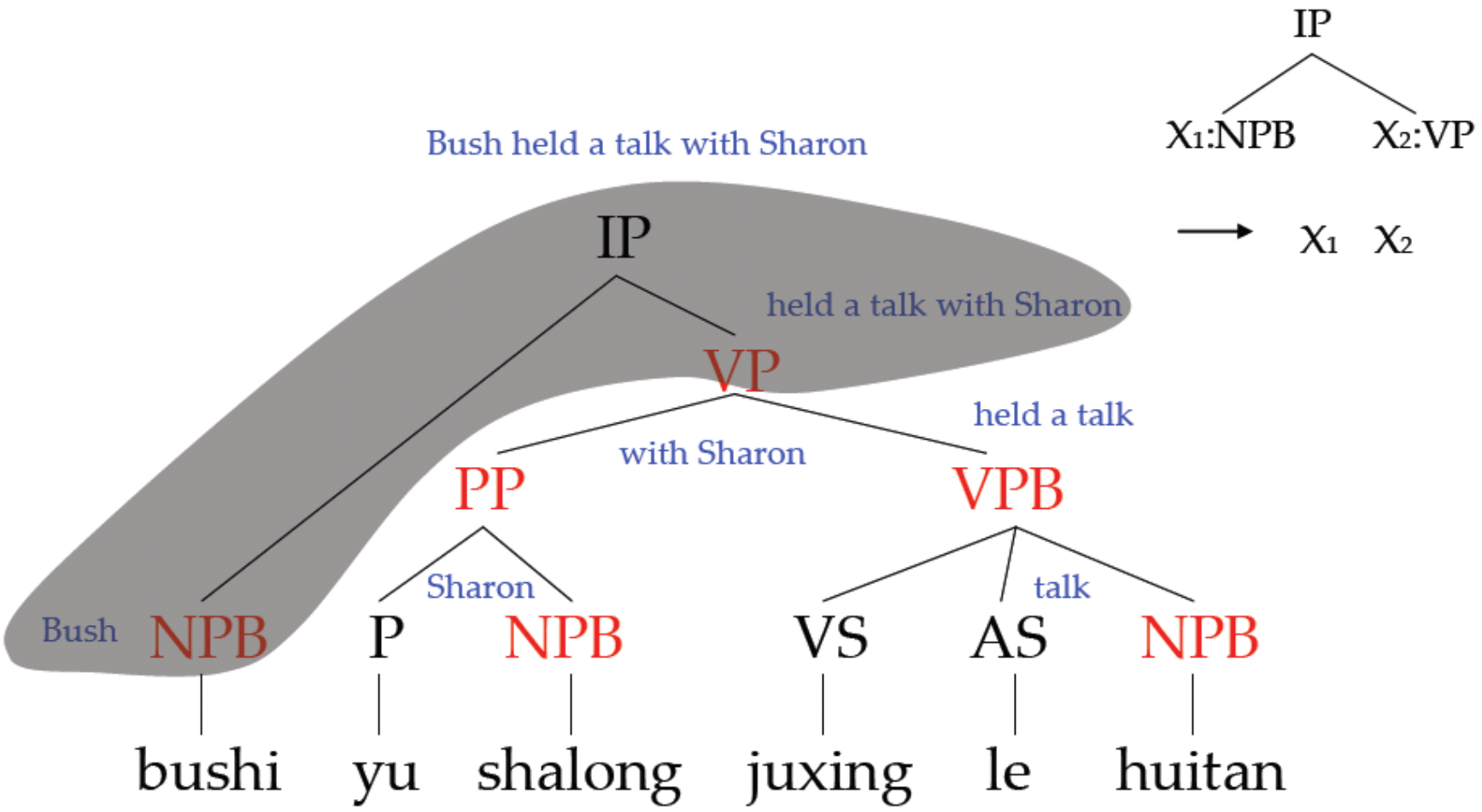
Tree-based Button-up Decoding



Tree-based Button-up Decoding

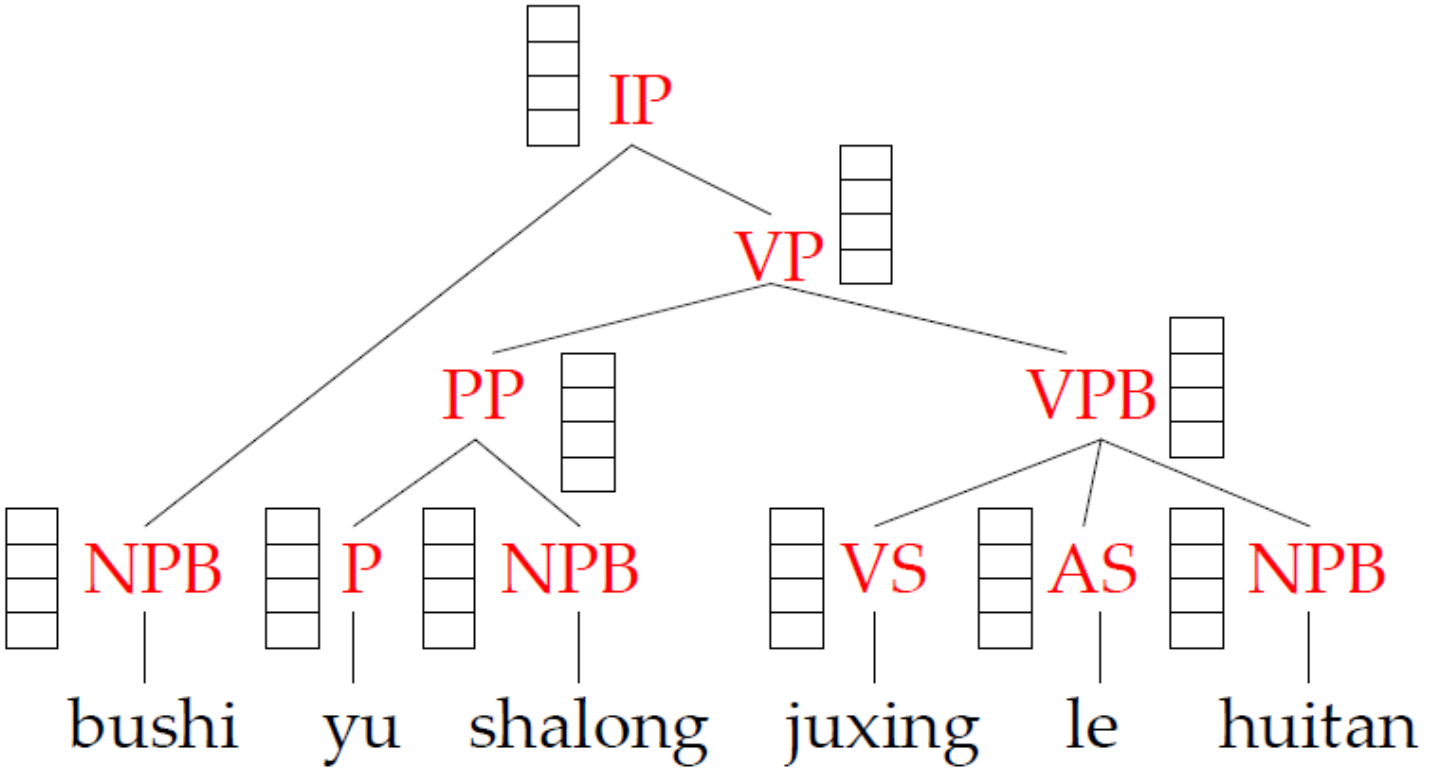


Tree-based Button-up Decoding



Tree-based Button-up Decoding

- Beam Search



Our Work: Tree-to-String Model

Constituent-to-String Model

Tree-based Translation

Forest-based Translation

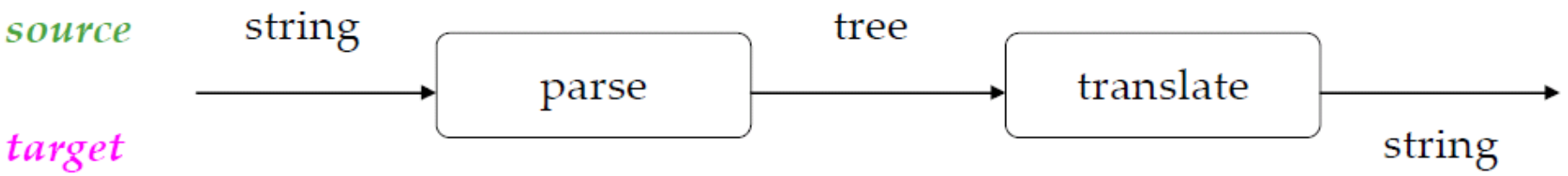
Joint Parsing and Translation

Dependency-to-String Model

Forest-based Translation

- Haitao Mi, Liang Huang and Qun Liu. Forest-Based Translation. In Proceedings of ACL 2008 Columbus, OH
- Haitao Mi and Liang Huang. Forest-based Translation Rule Extraction. In Proceedings of EMNLP 2008 ,Honolulu, Hawaii.
  Nominated for the best-paper award

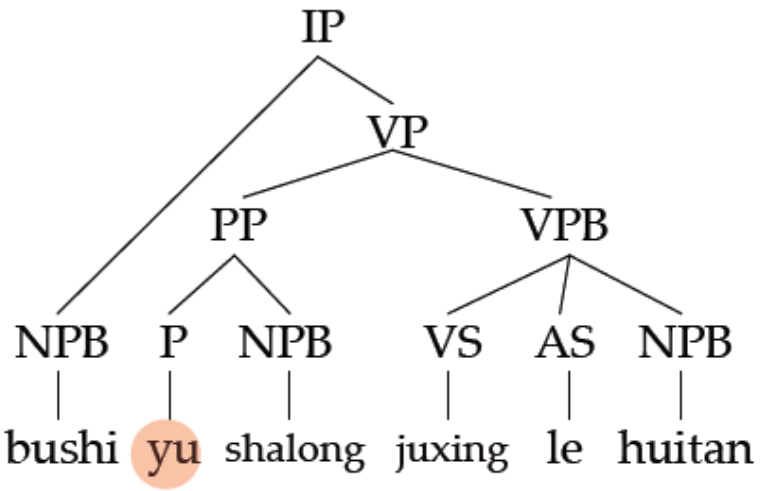
Parsing Mistake Propagation



parsing mistakes potentially introduce translation mistakes!

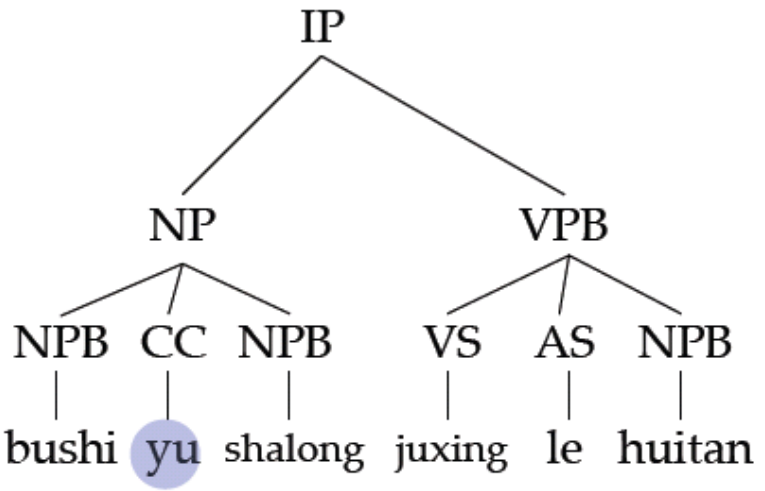
Syntatic Ambiguity

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



with

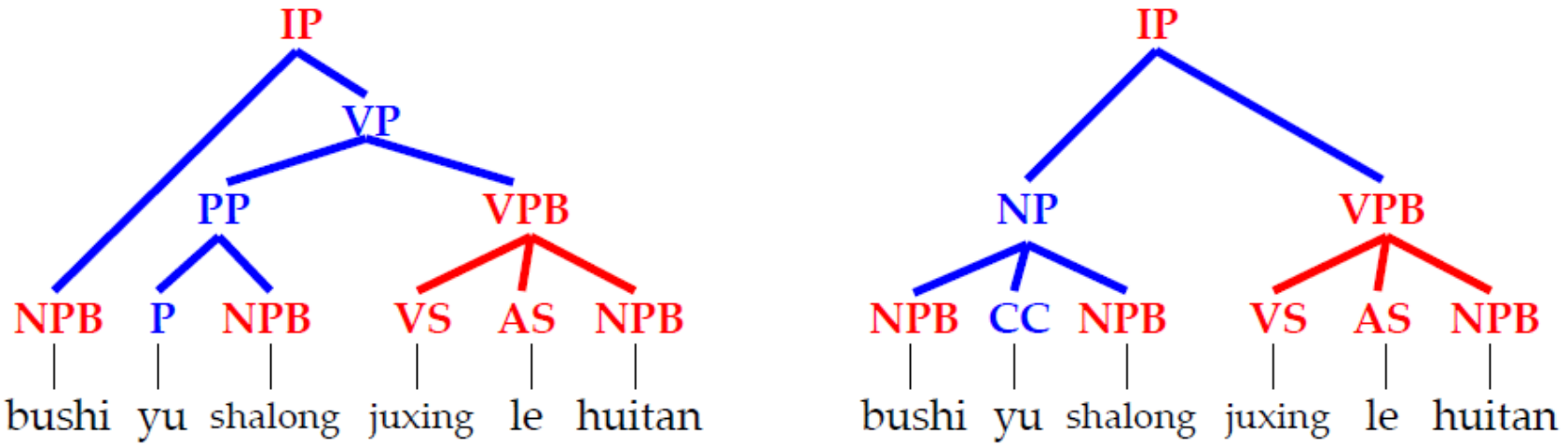
``Bush held a talk with Sharon''



and

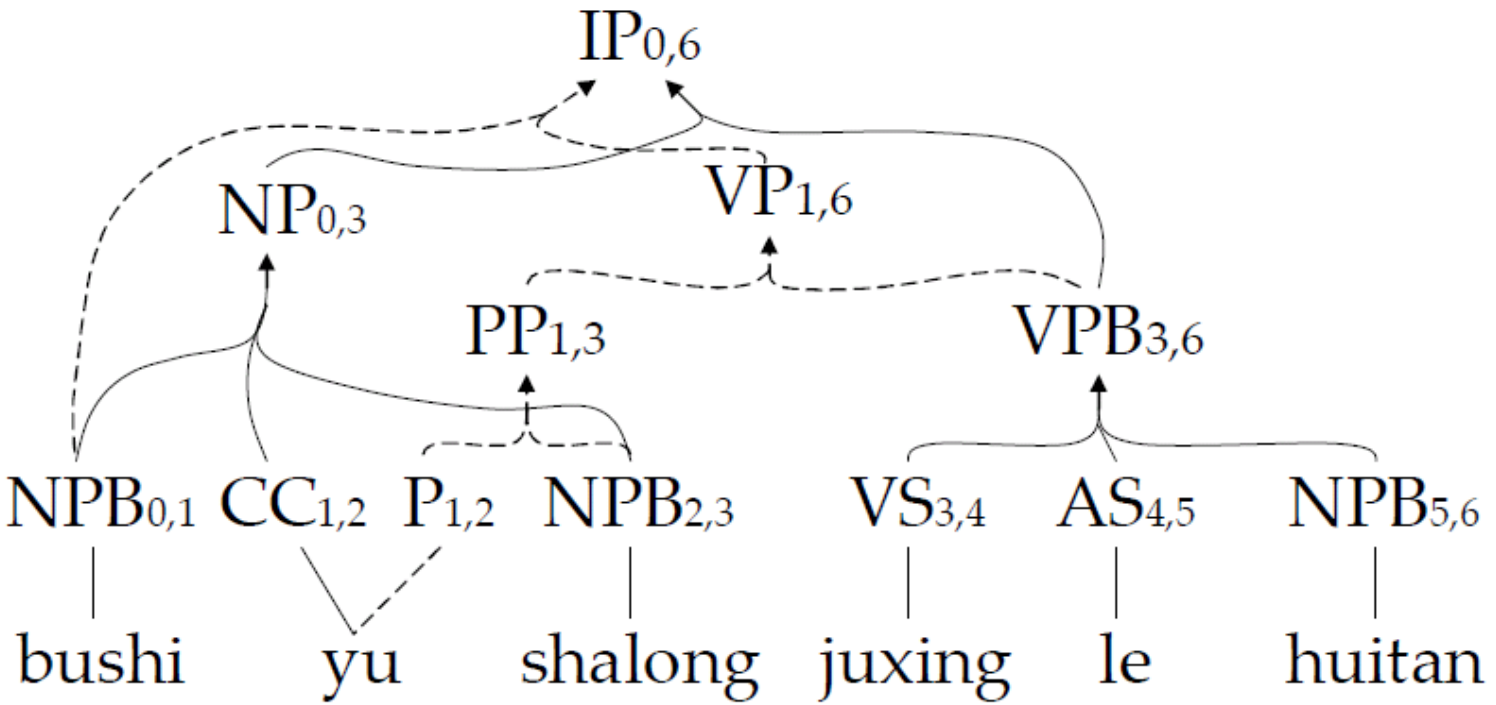
``Bush and Sharon held a talk''

1-best ➜ n-best trees?

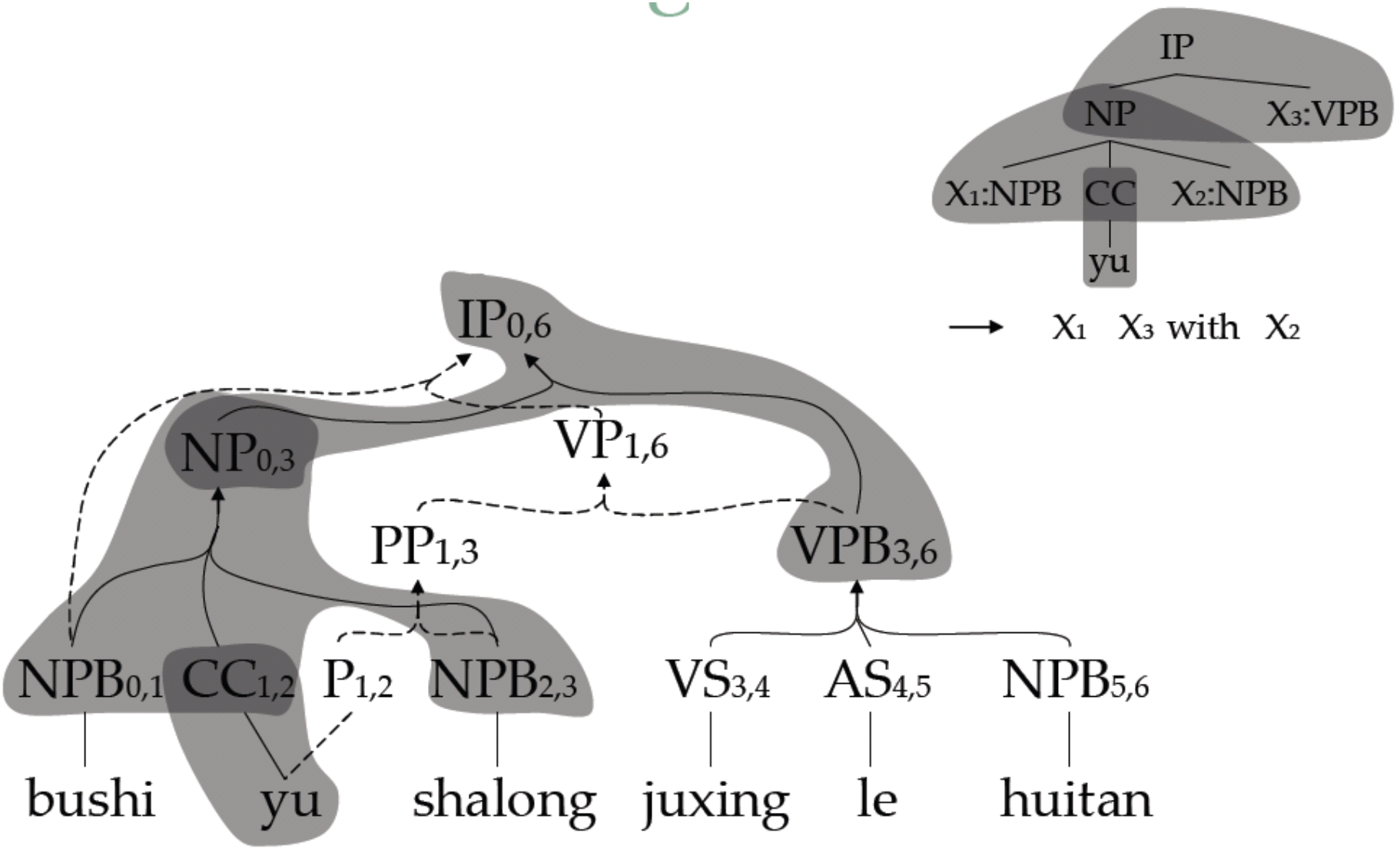


Very few variations among the *n*-best trees!

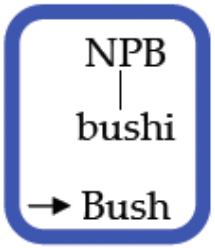
Packed Forest

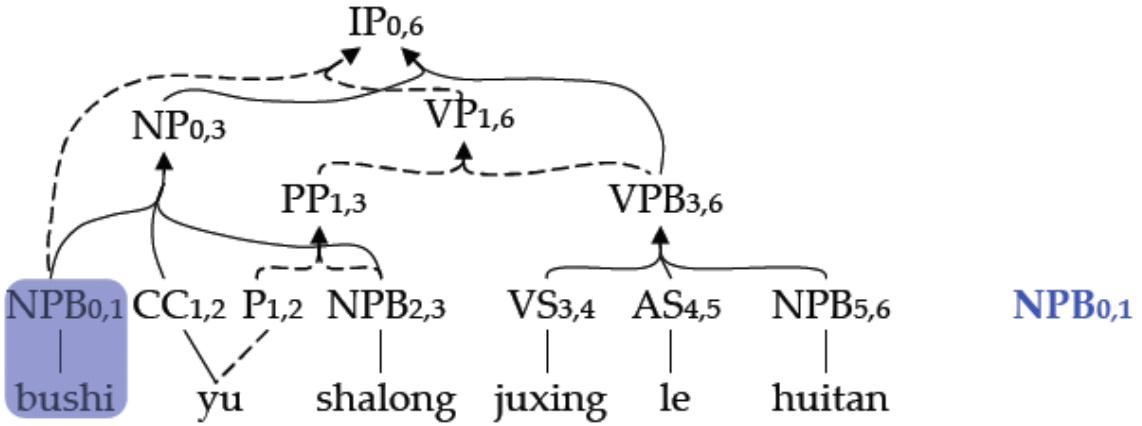


Patten Matching on Forest

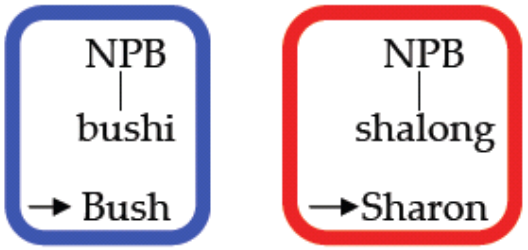


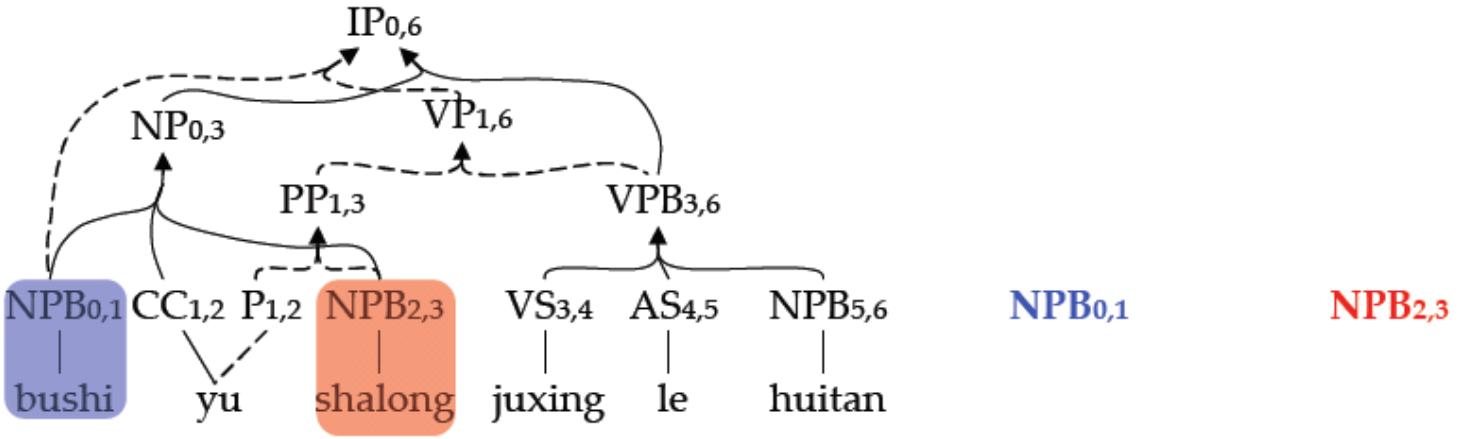
Translation Forest



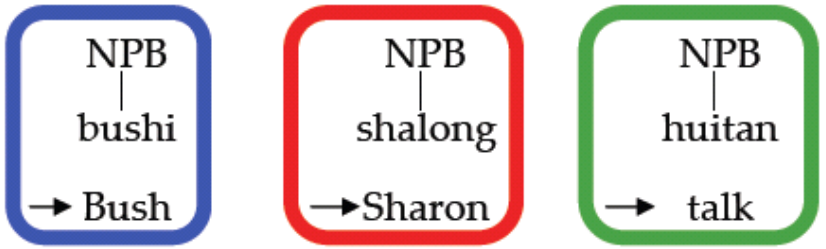


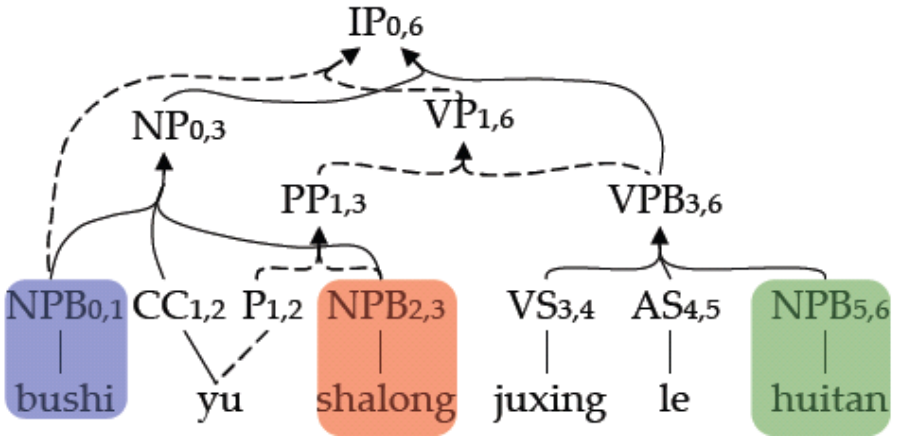
Translation Forest





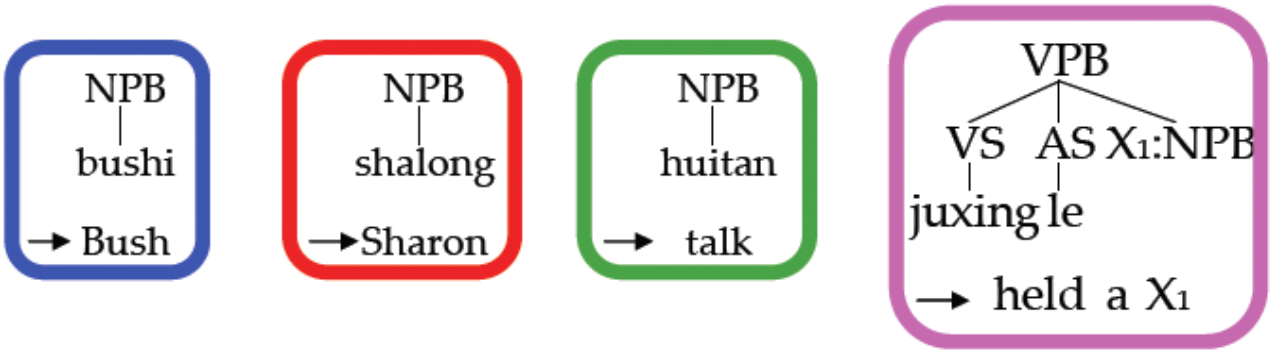
Translation Forest

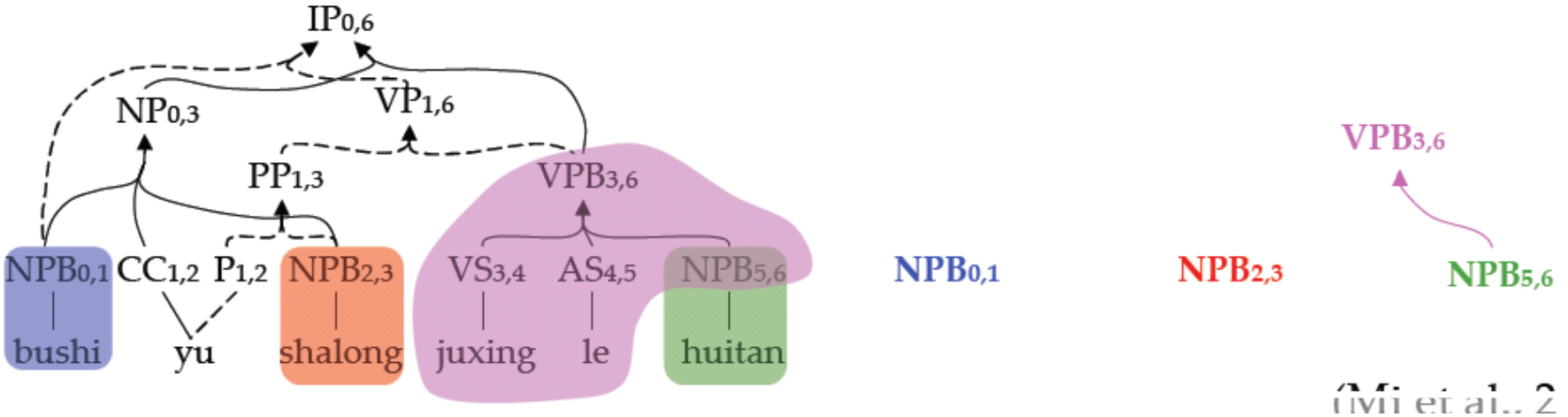




NPB0,1 NPB2,3 NPB5,6

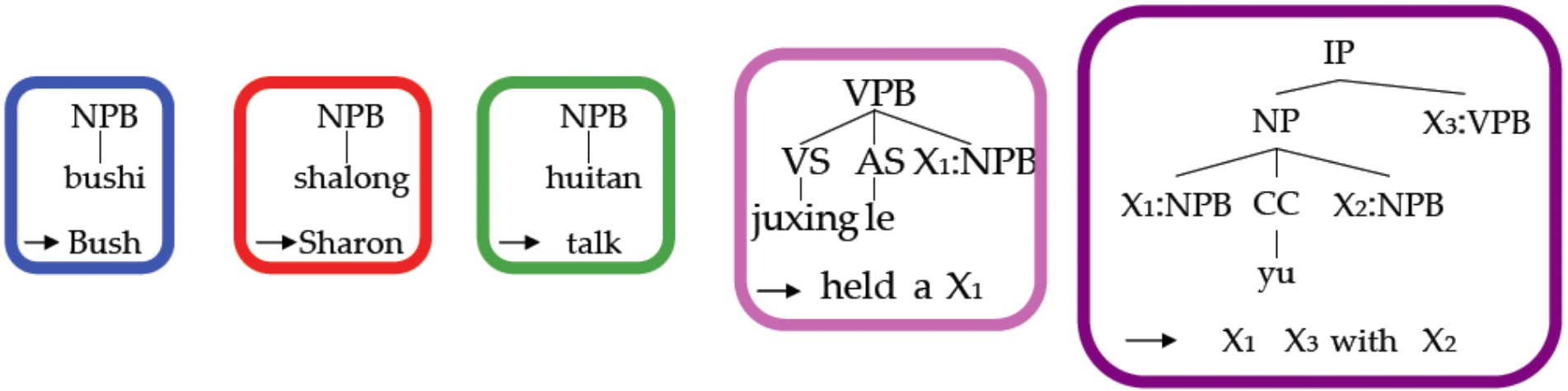
Translation Forest

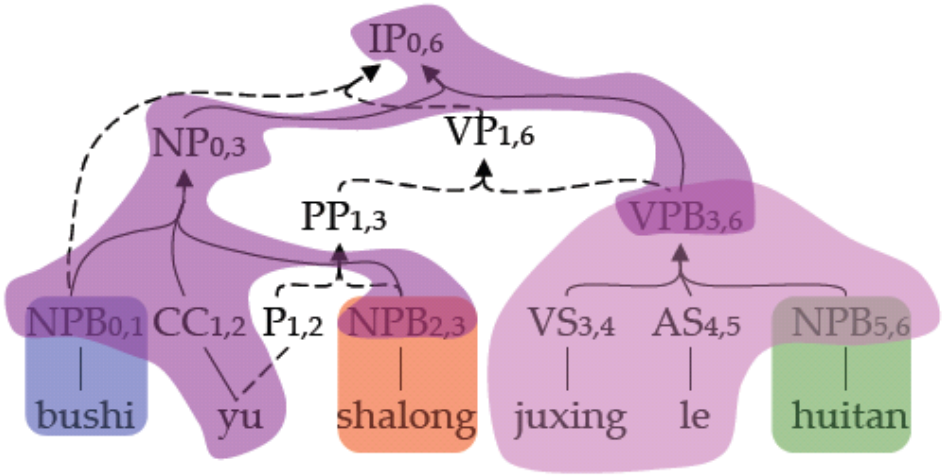


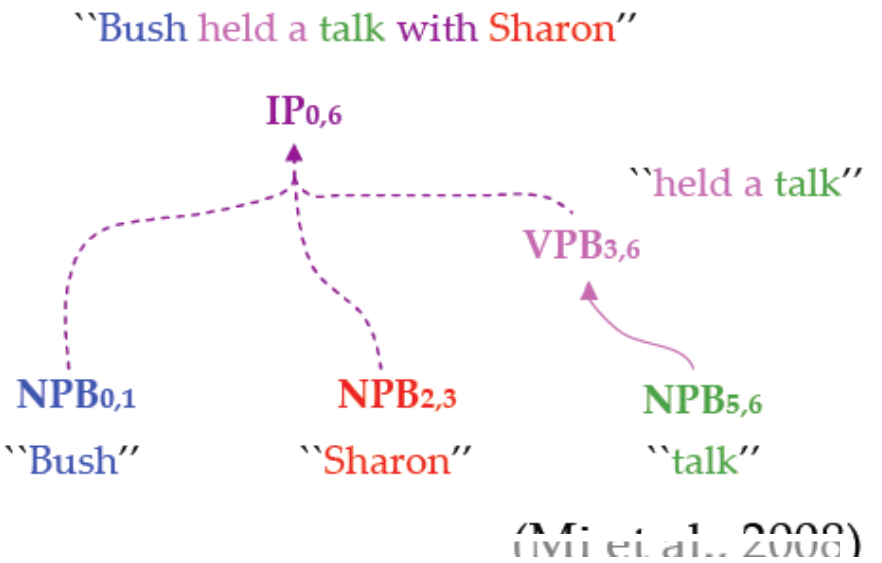


(Mi et al., 2

Translation Forest

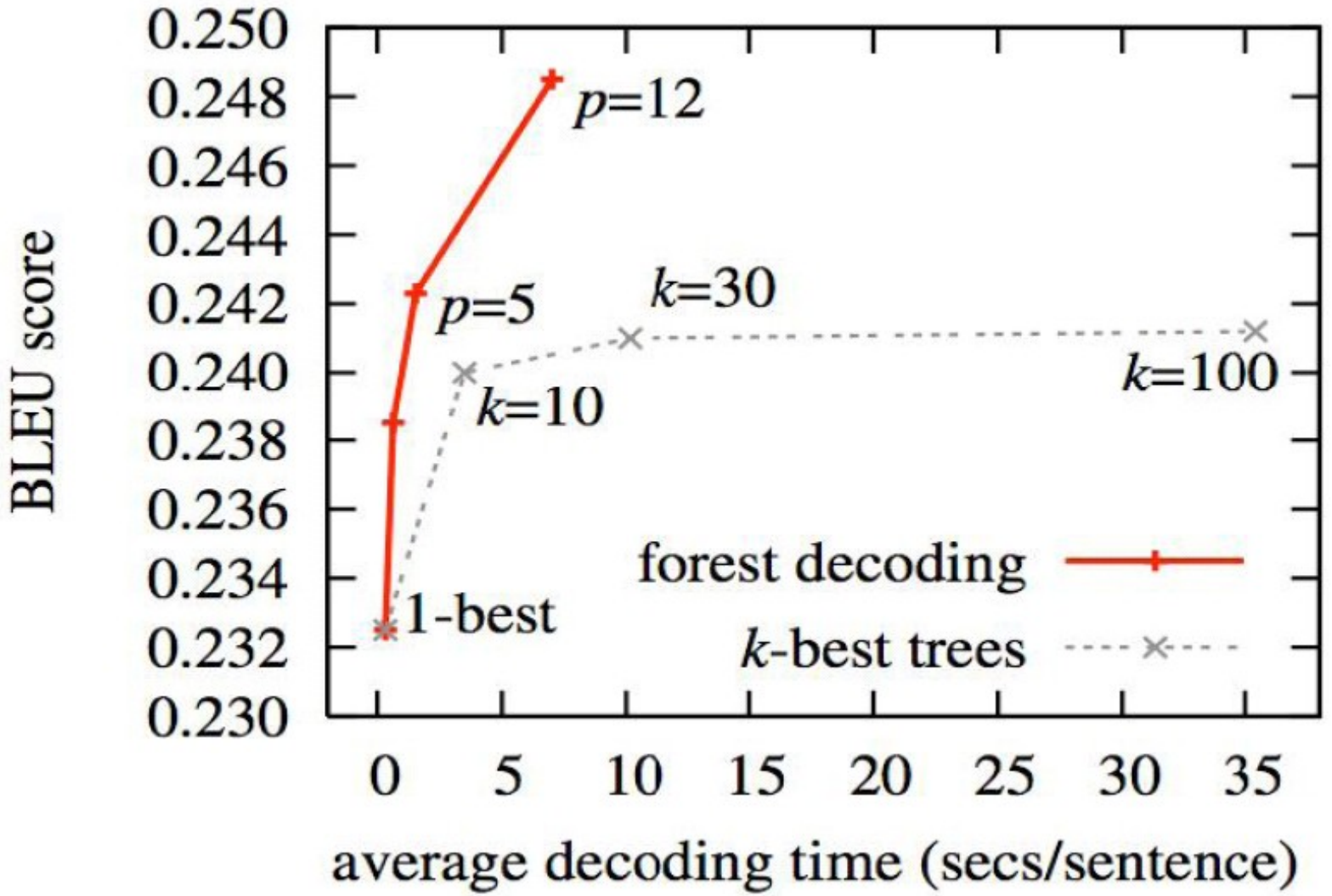






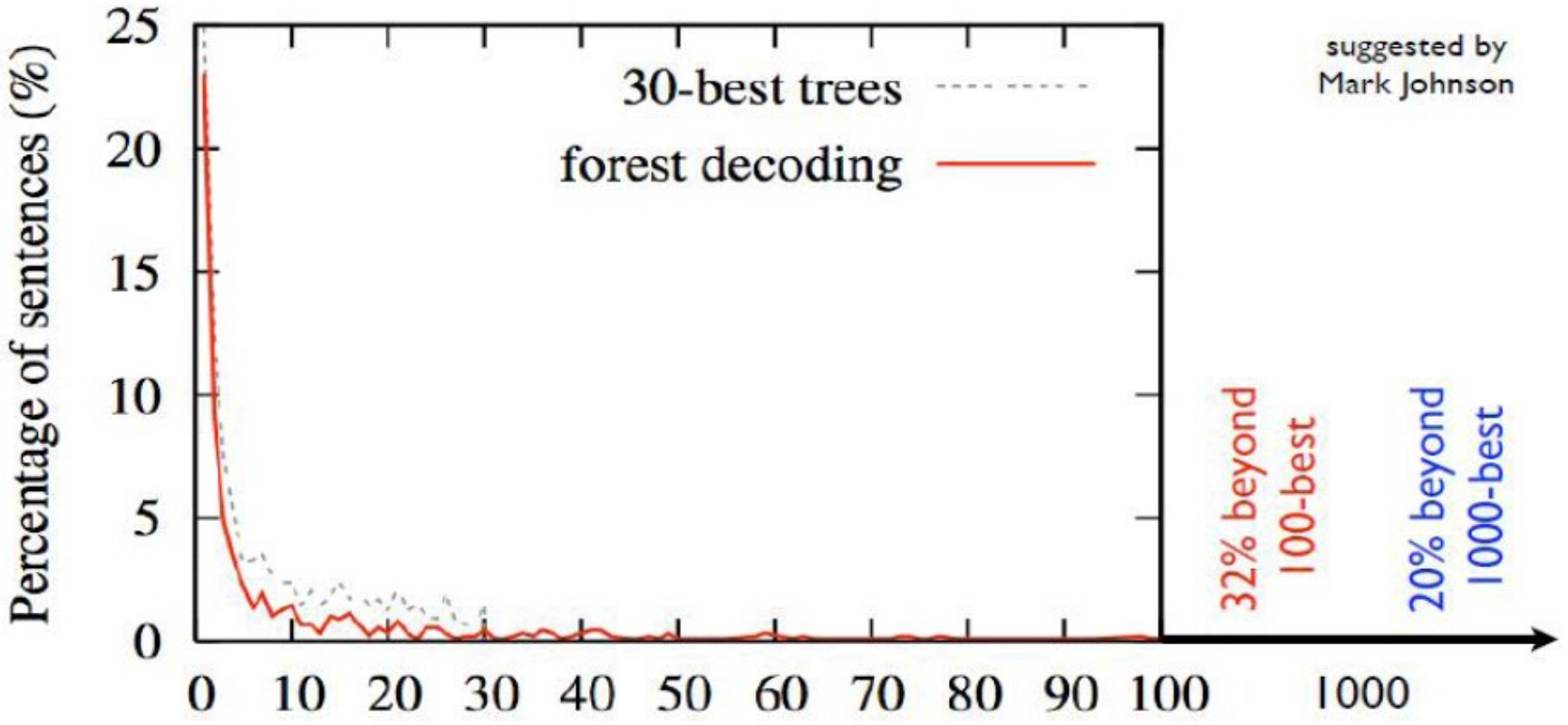
(Mi et al., 2008)

N-best Trees vs. Forest



Forest as Virtual ∞-best List

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



Our Work: Tree-to-String Model

Constituent-to-String Model

Tree-based Translation

Forest-based Translation

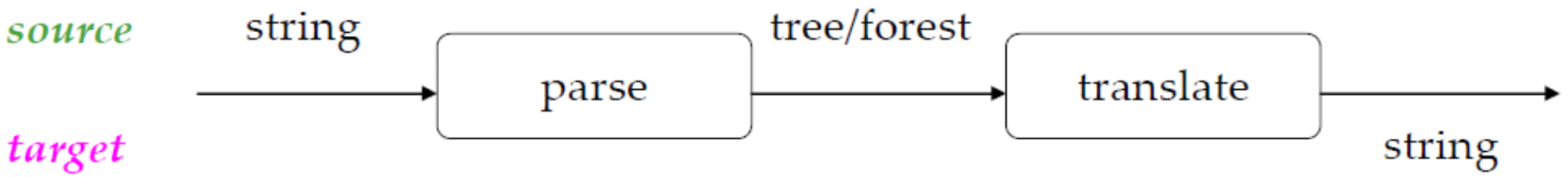
Joint Parsing and Translation

Dependency-to-String Model

Joint Parsing and Translatoin

- Yang Liu and Qun Liu. 2010. Joint Parsing and Translation. In Proceedings of COLING 2010, pages 707-715, Beijing, China, August.

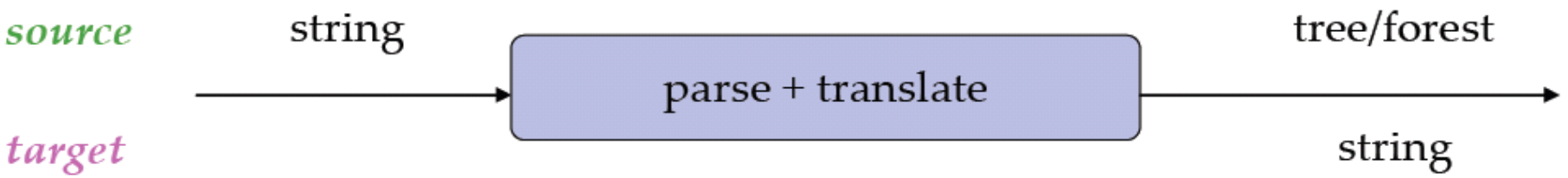
Seperate Parsing and Translation



☺ Separate grammar for parsing and translation

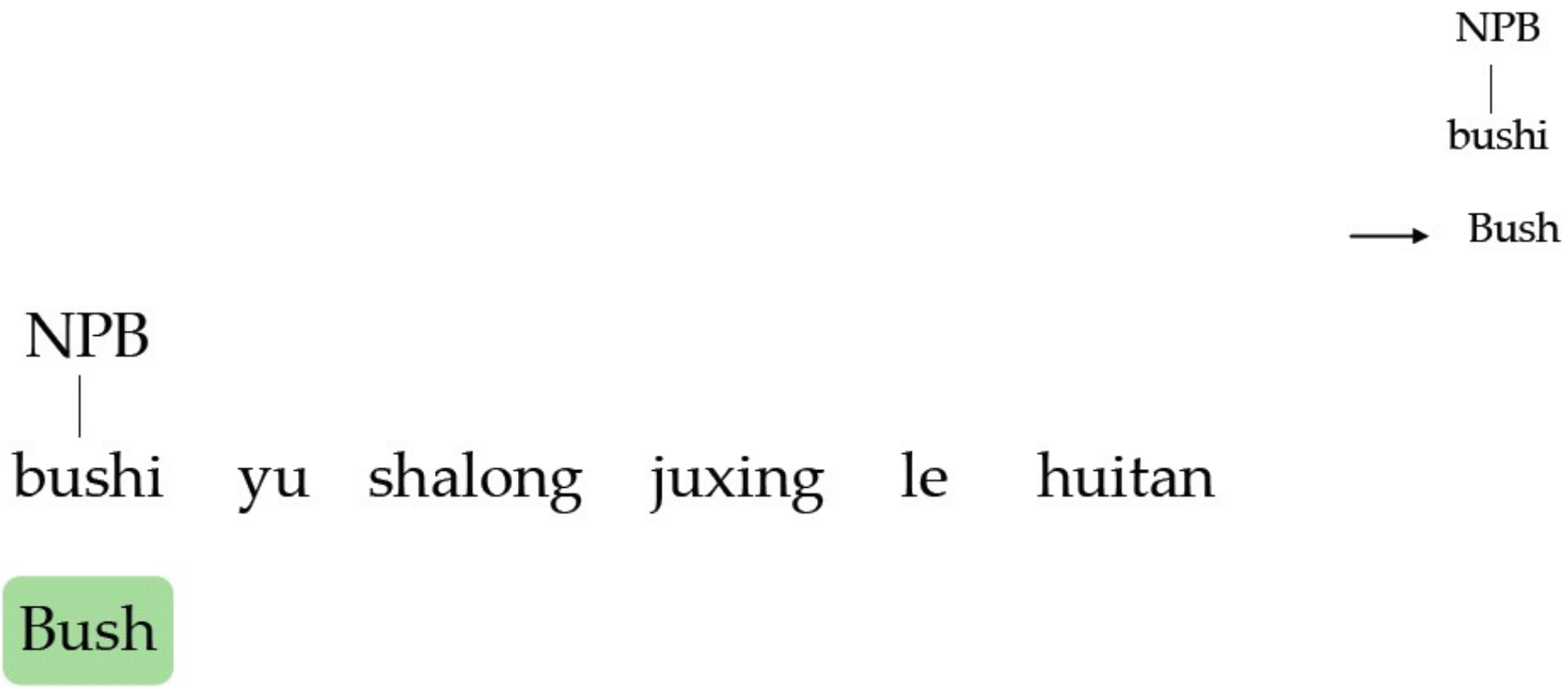
☺ decoding is fast!

Joint Parsing and Translation

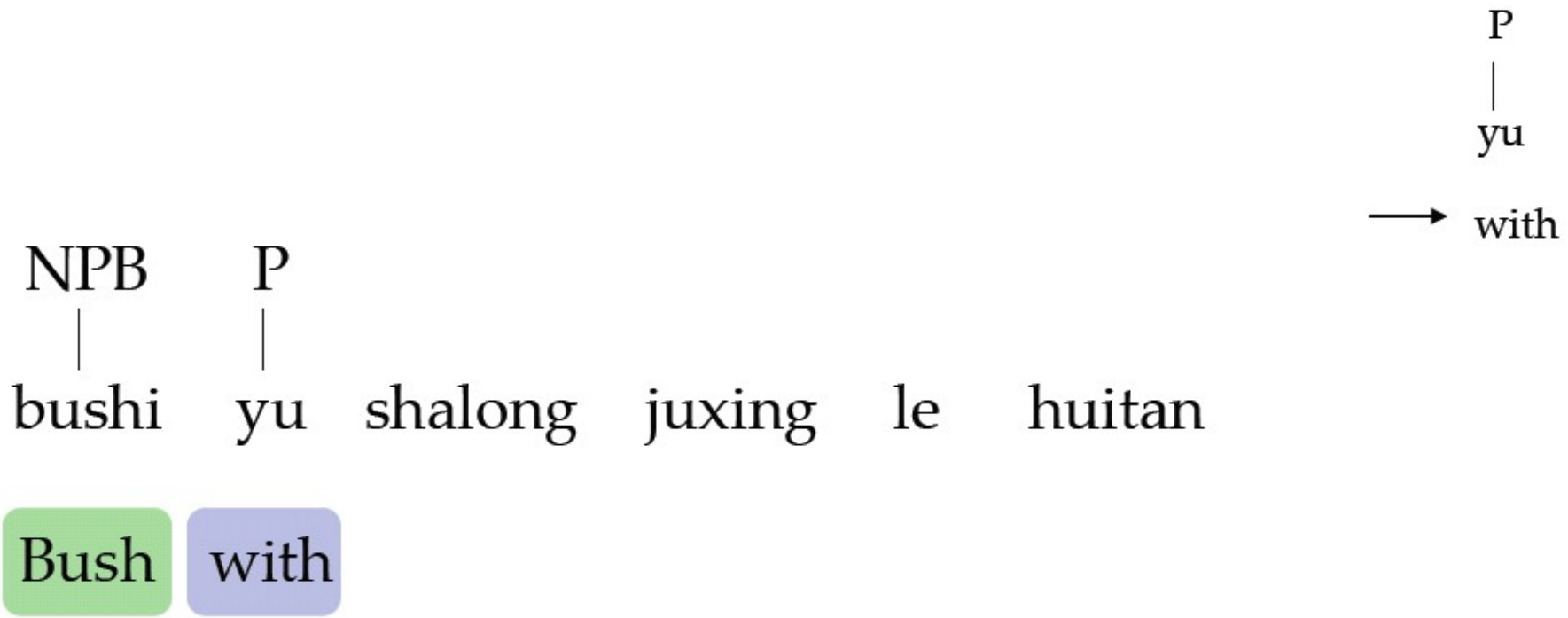


- Its search space is larger than tree/forest
- It is a translator as well as a parser
- Parsing interacts with translation

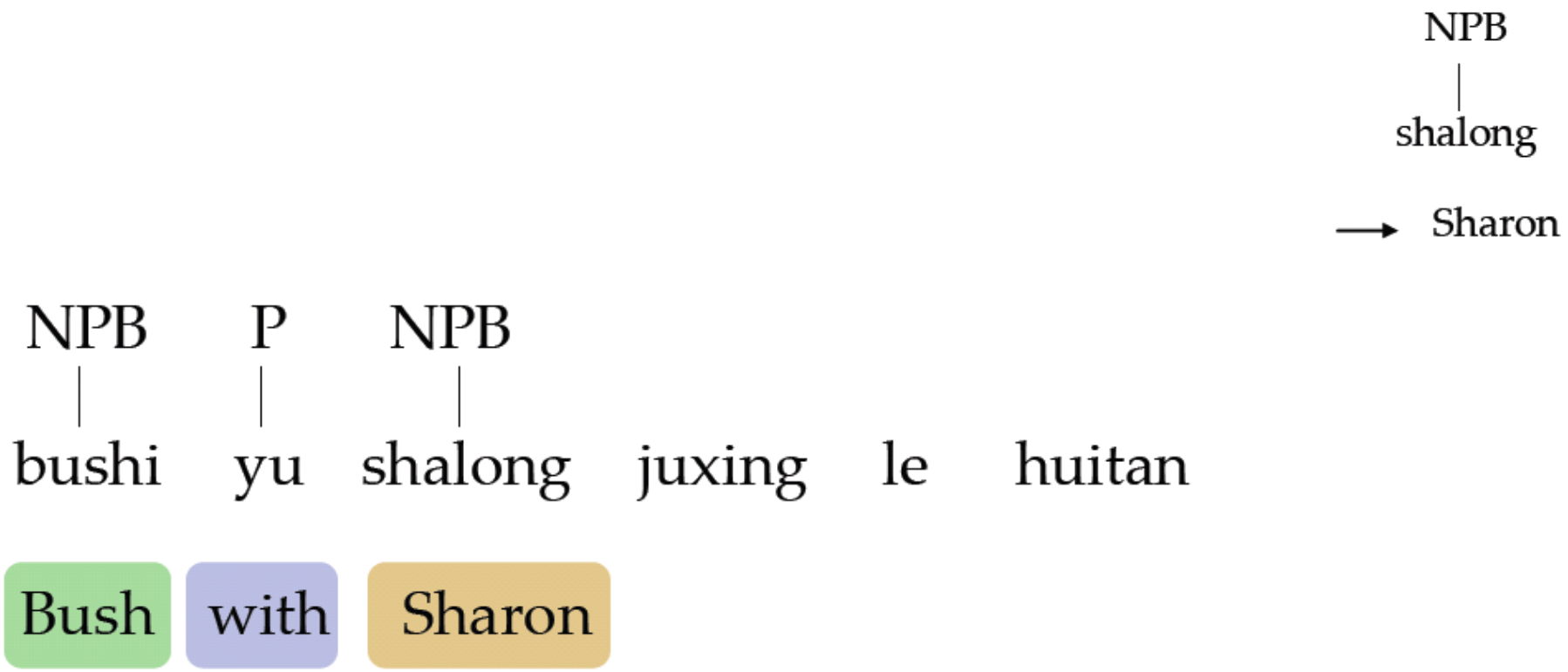
Joint Parsing and Translation



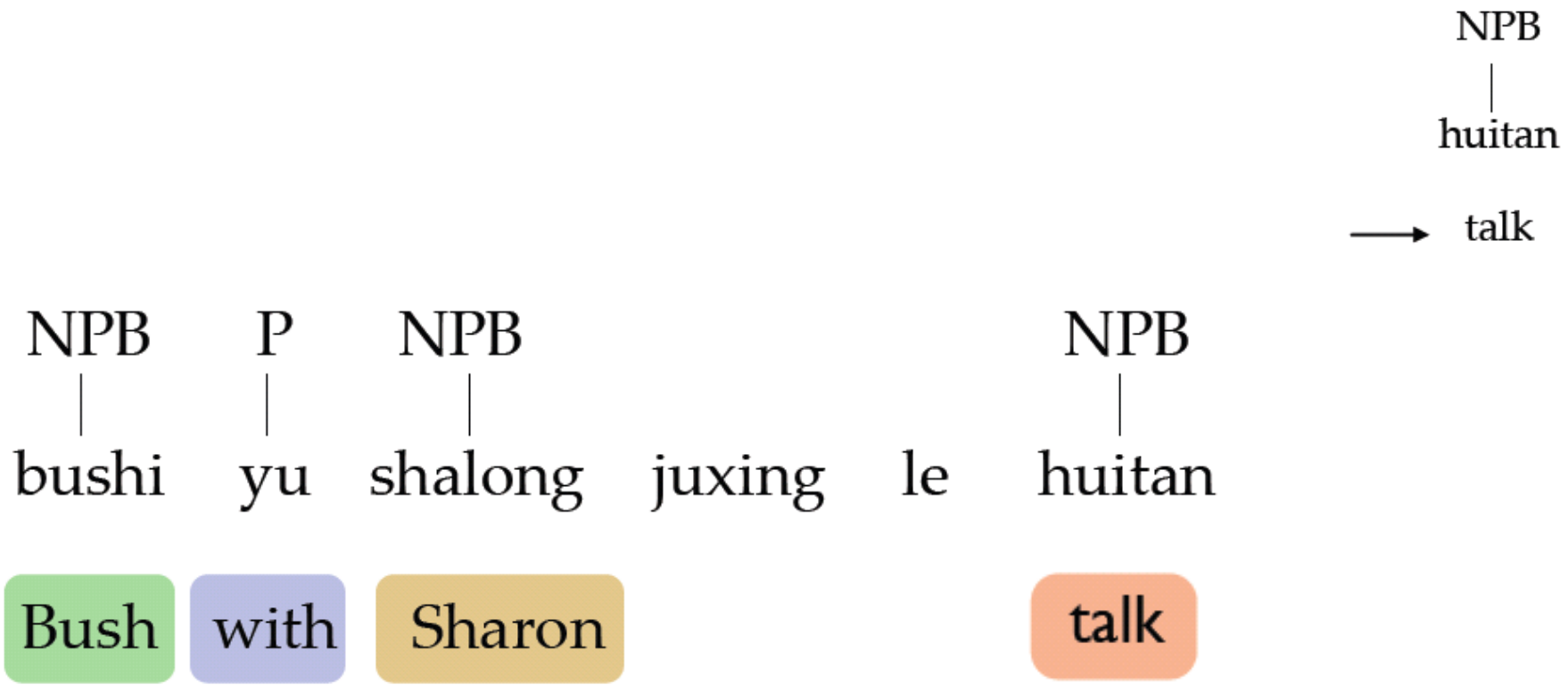
Joint Parsing and Translation



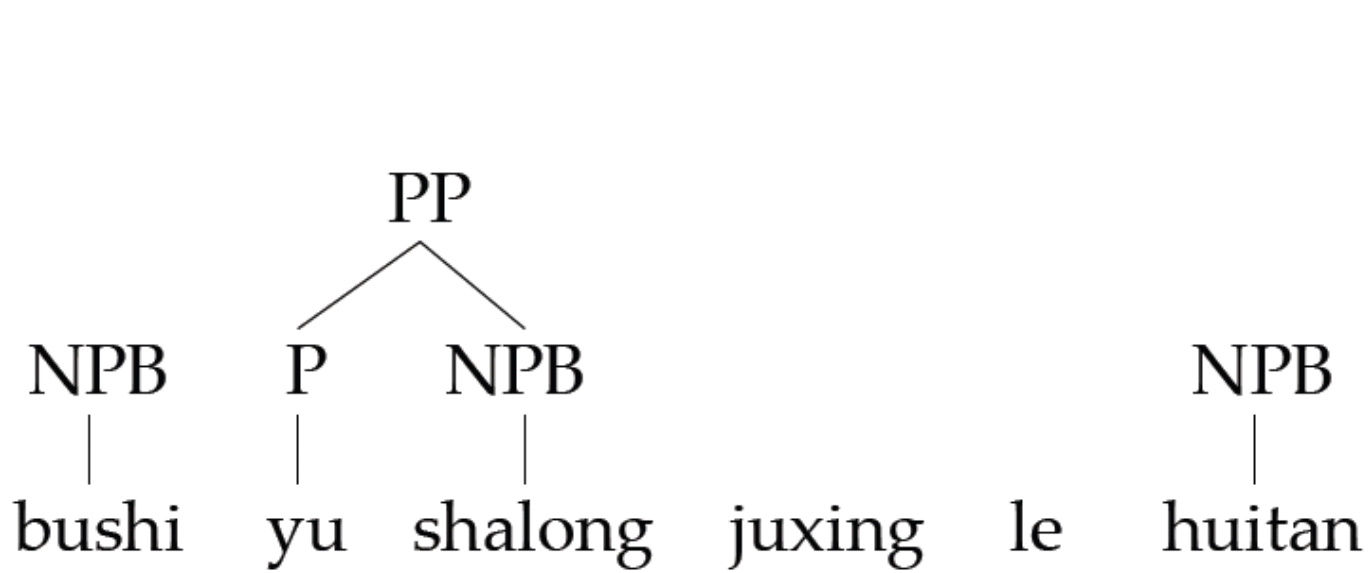
Joint Parsing and Translation

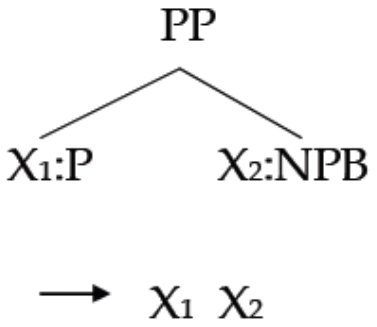


Joint Parsing and Translation

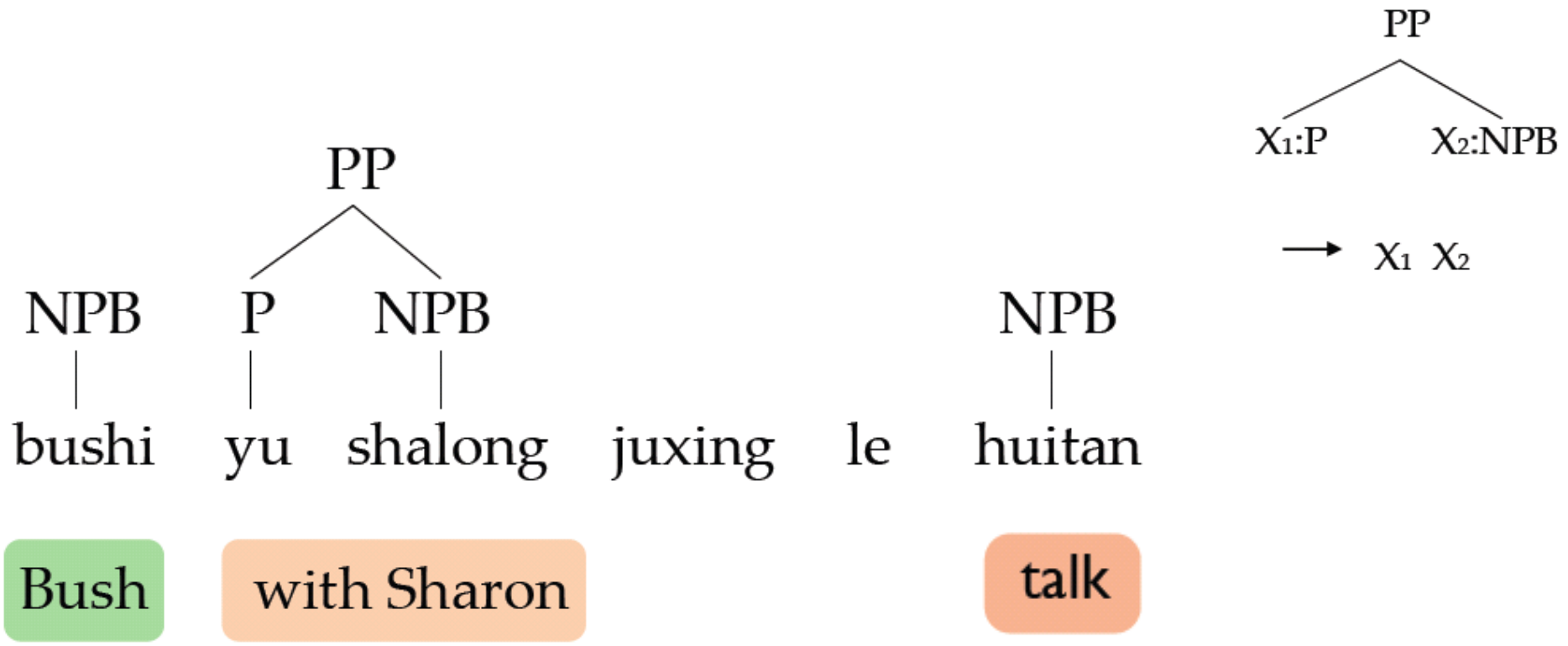


Joint Parsing and Translation

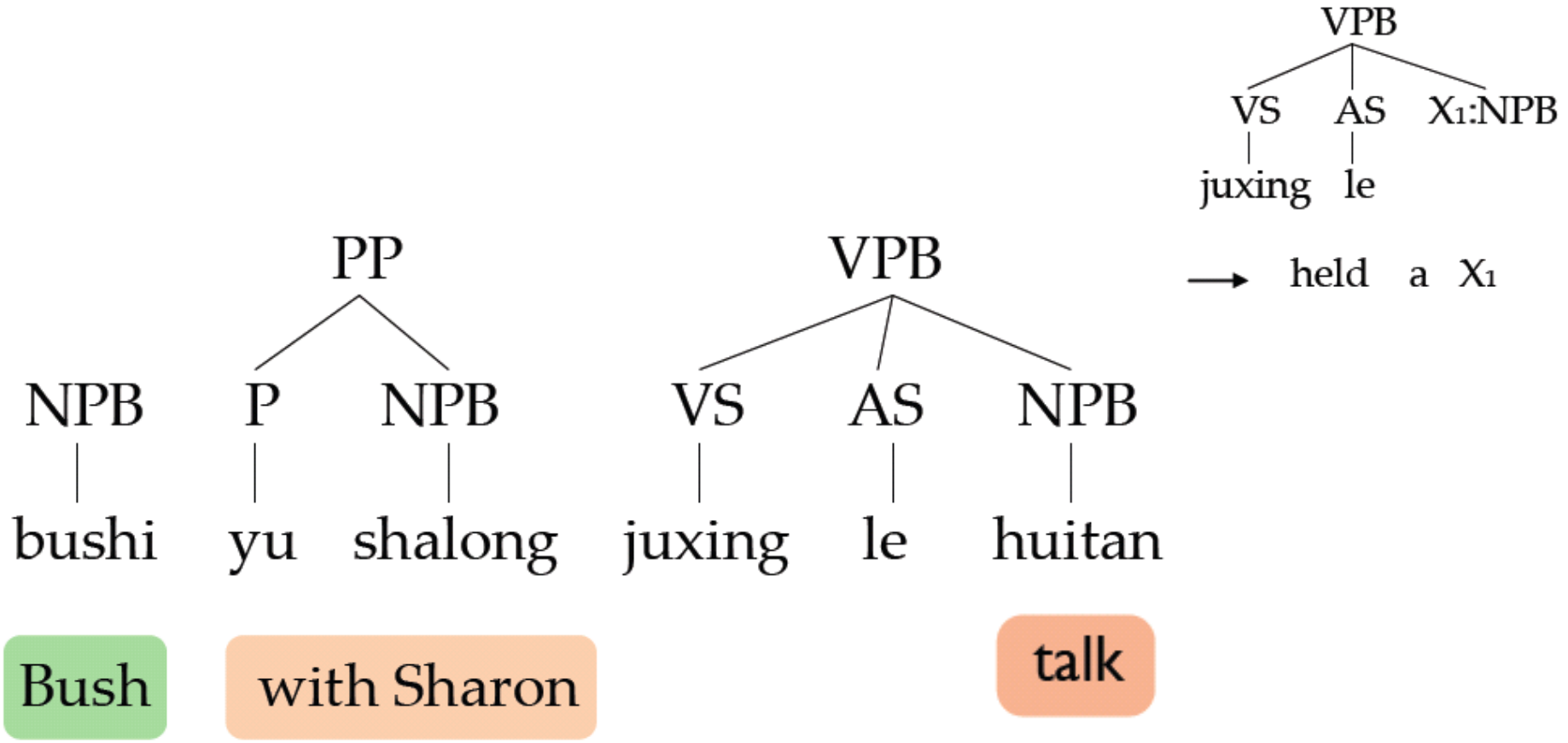




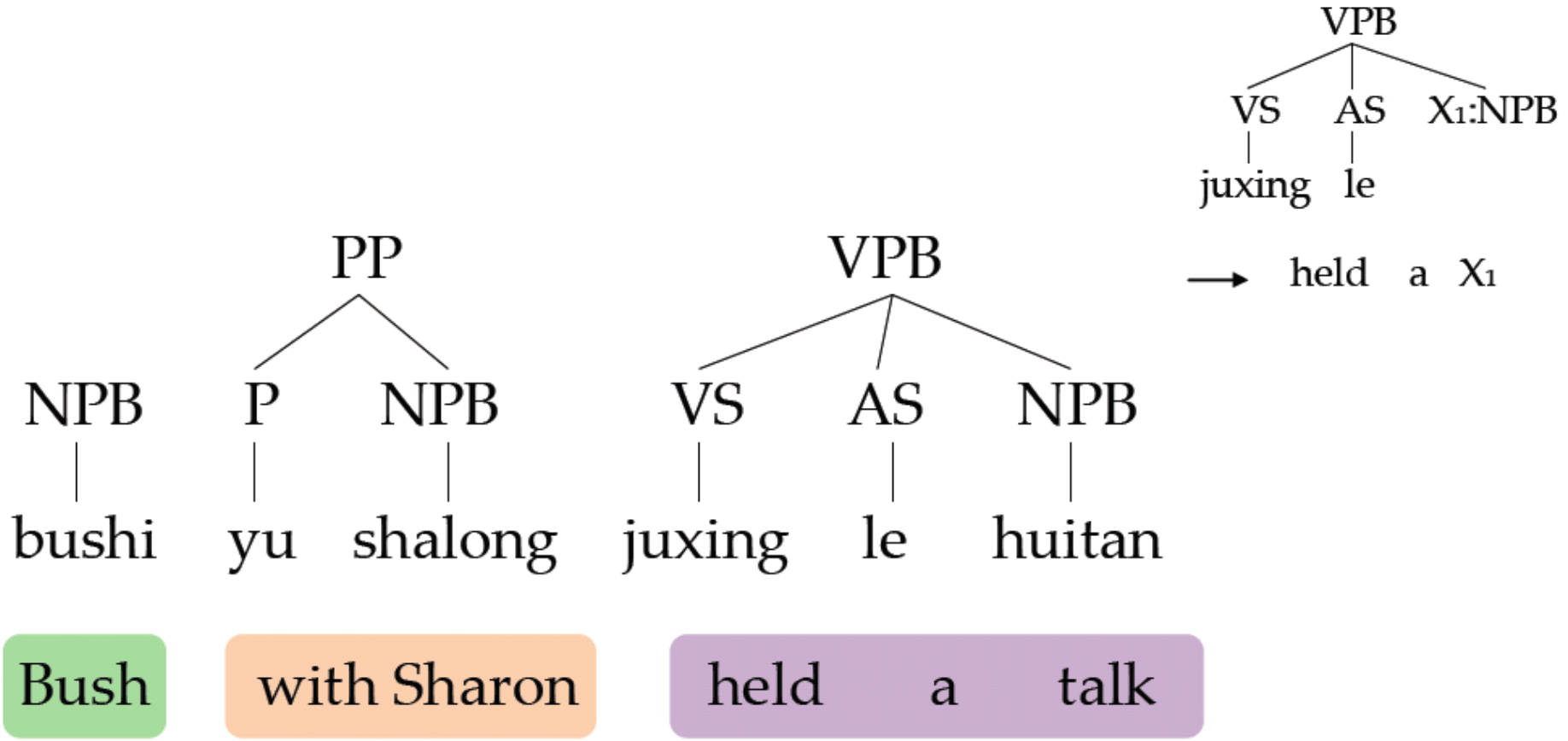
Joint Parsing and Translation



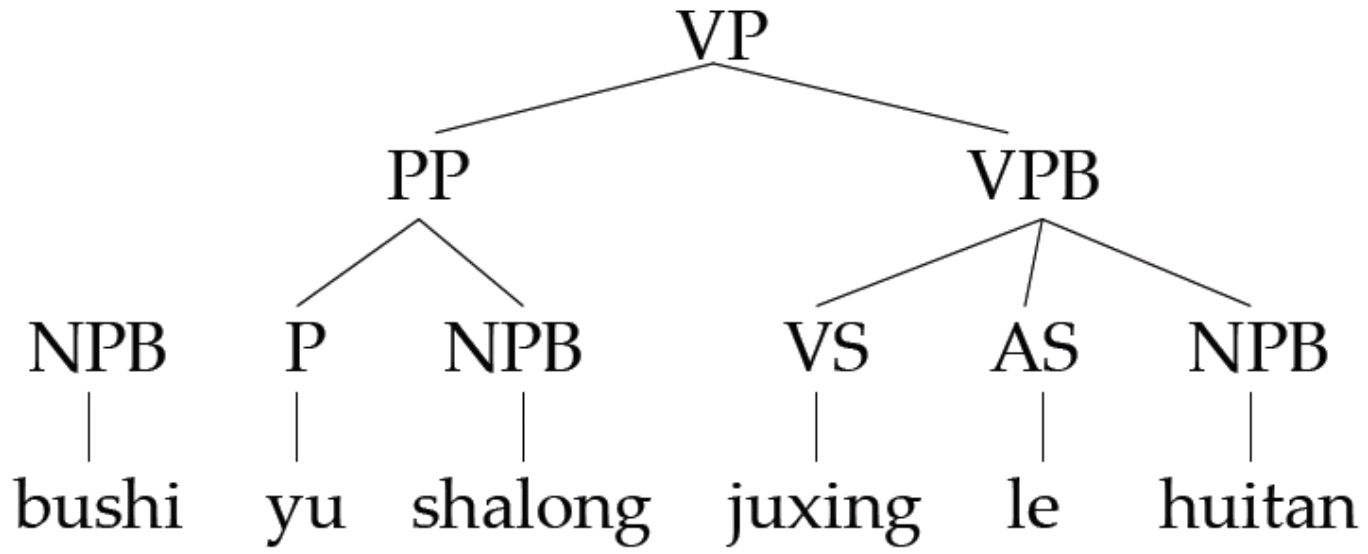
Joint Parsing and Translation

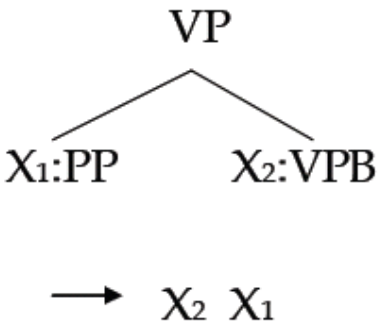


Joint Parsing and Translation



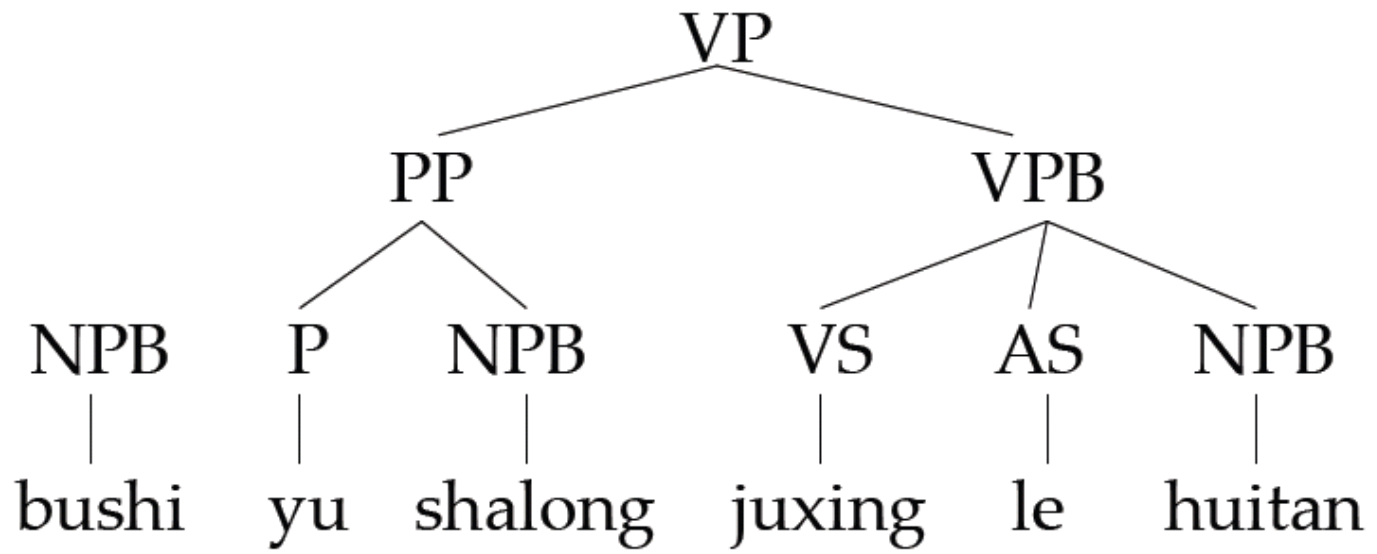
Joint Parsing and Translation

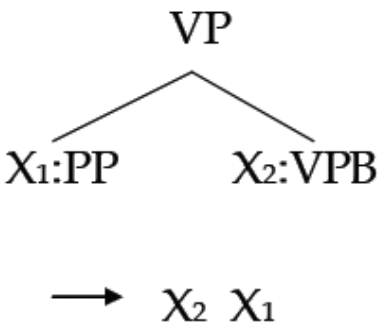




Bush with Sharon held a talk

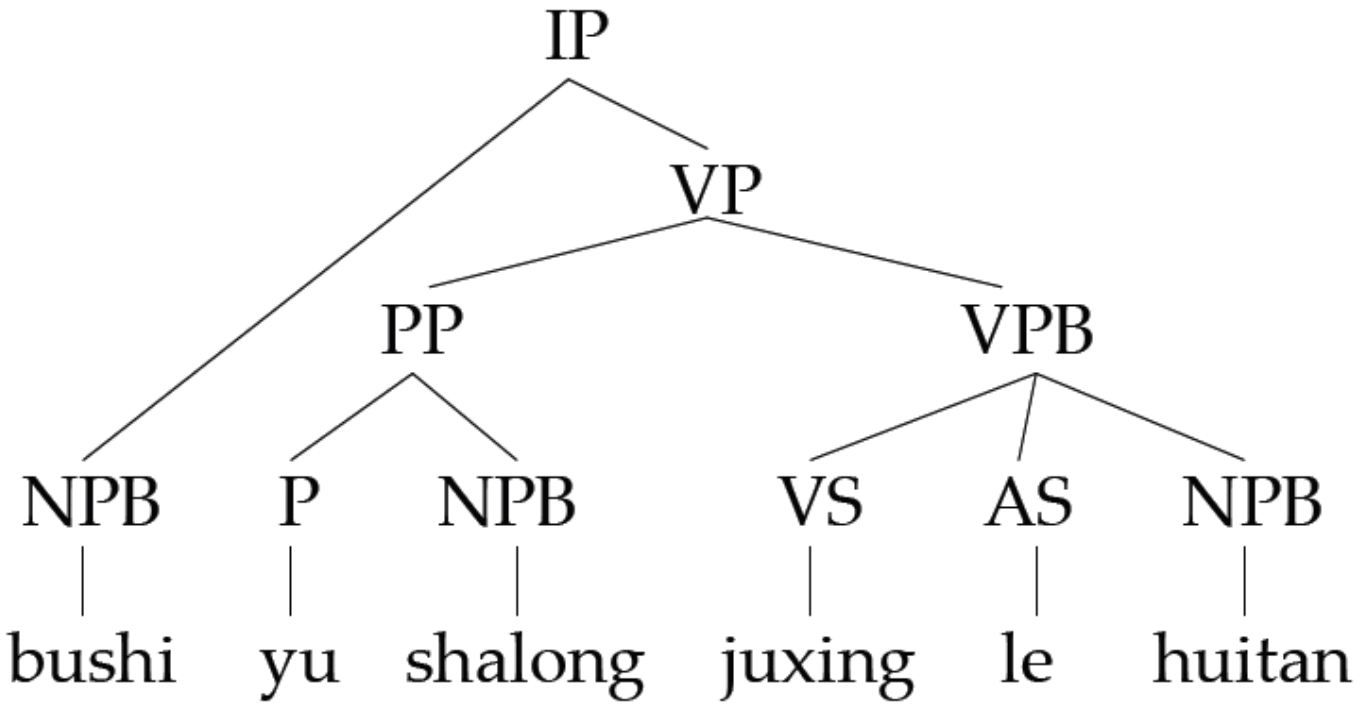
Joint Parsing and Translation

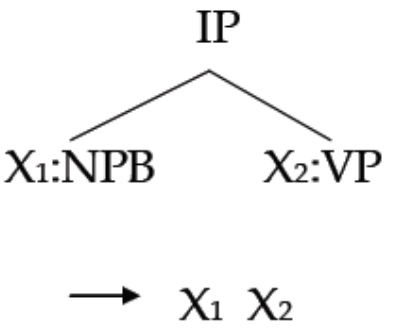




Bush held a talk with Sharon

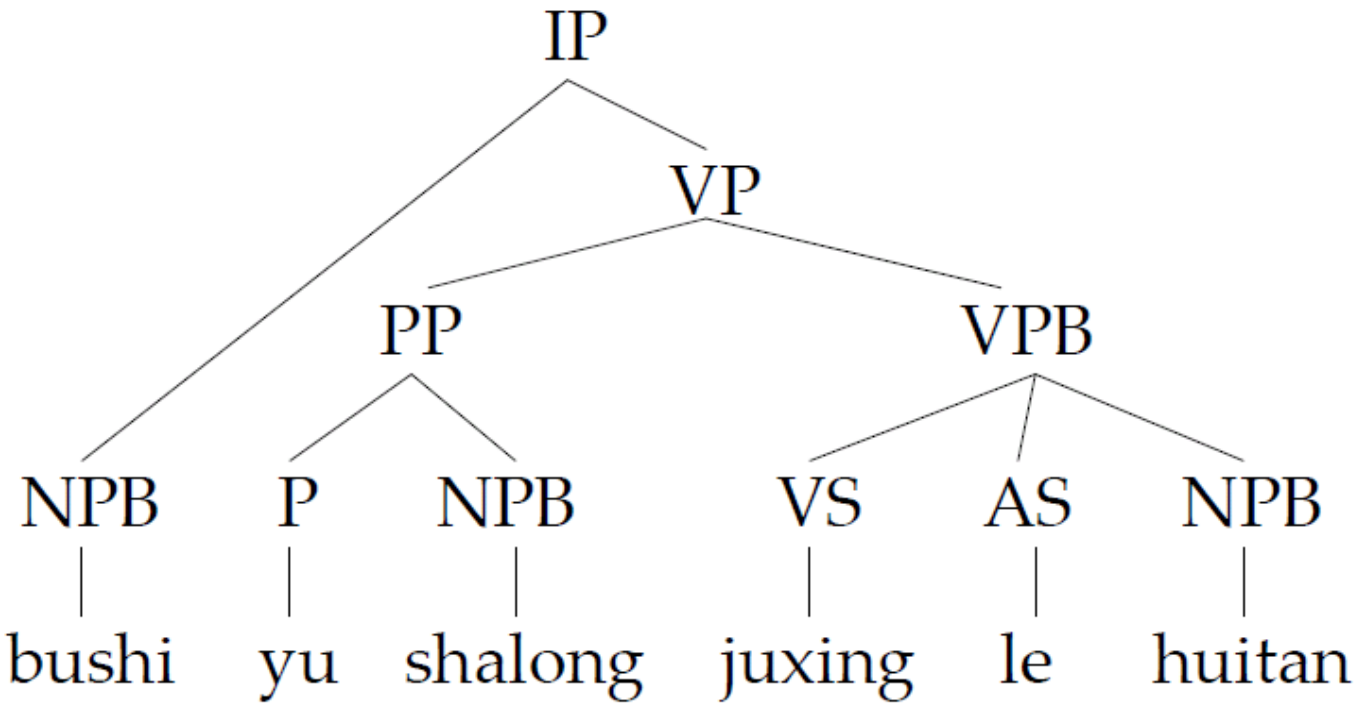
Joint Parsing and Translation

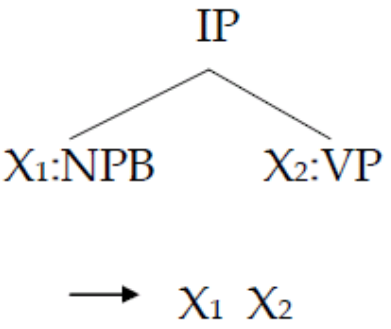




Bush held a talk with Sharon

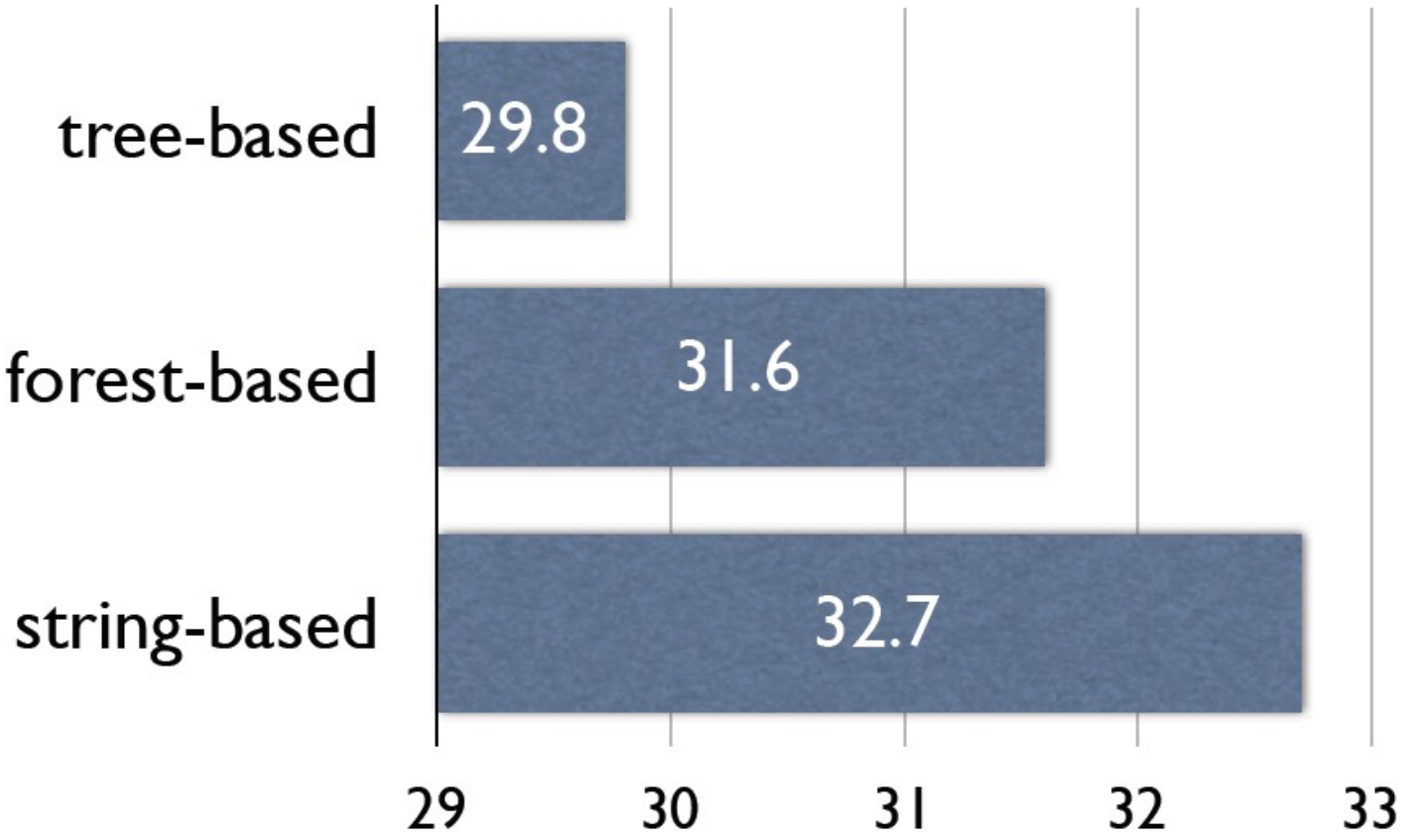
Joint Parsing and Translation





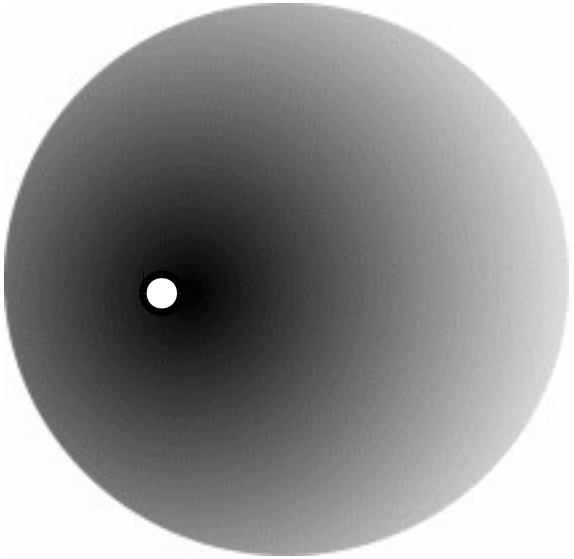
Bush held a talk with Sharon

Evaluation

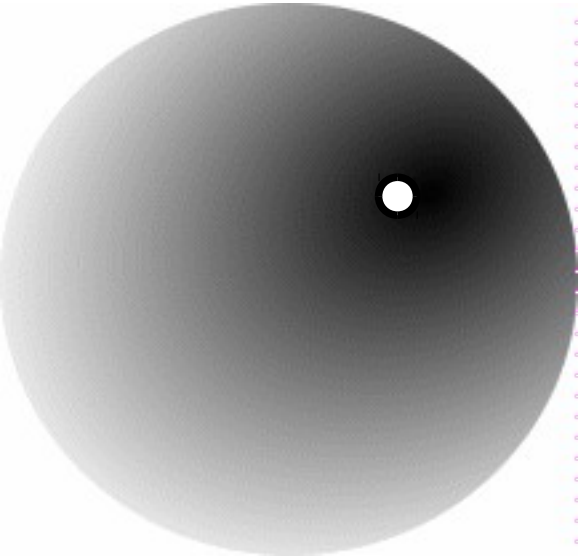


String-based Translation = Joint Parsing and Translation

Search Space Comparison



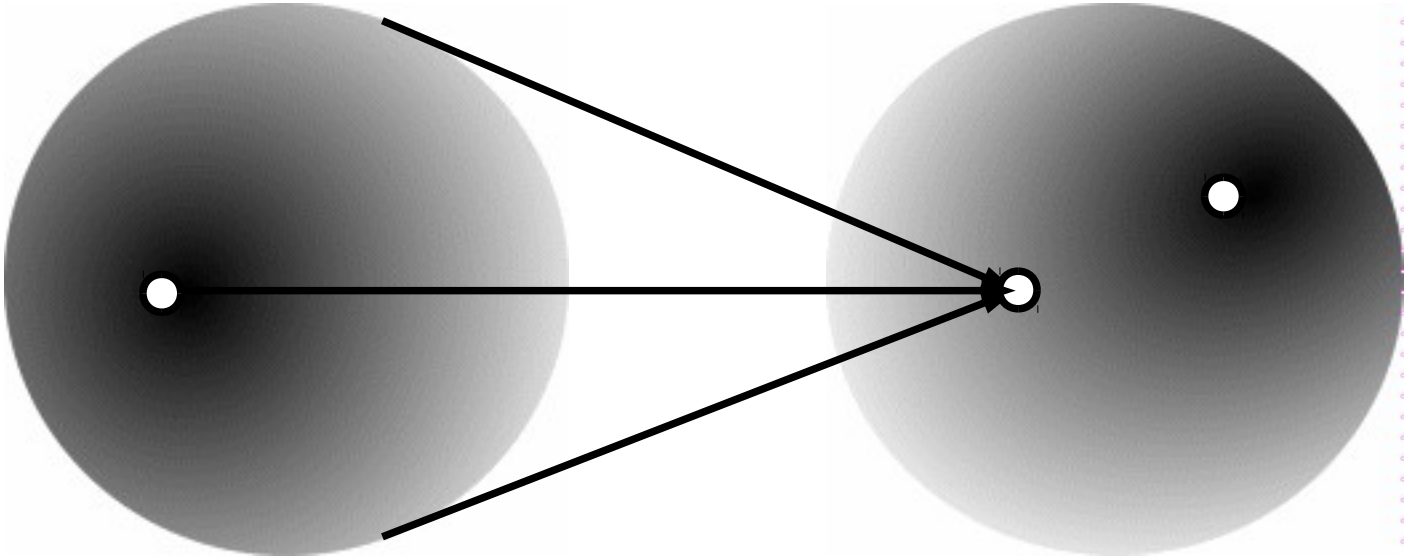
Probabilistic Distribution of Parsing Space



Probabilistic Distribution of Translation Space

Search Space Comparison

Tree-based Translation

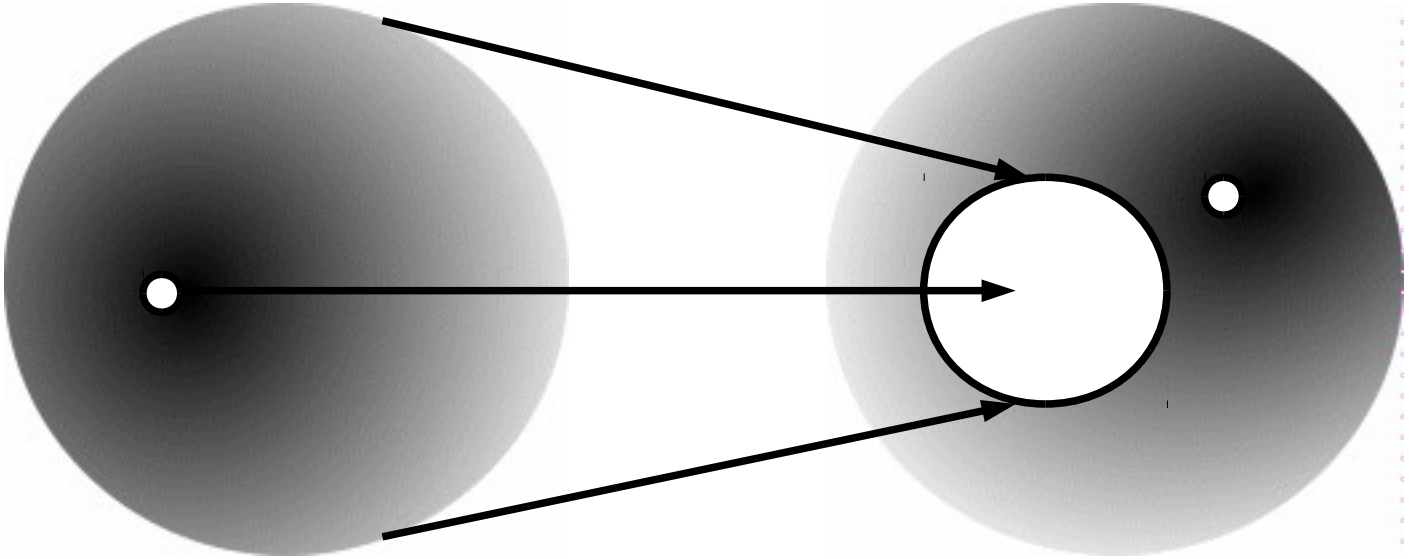


Probabilistic Distribution of Parsing Space

Probabilistic Distribution of Translation Space

Search Space Comparison

Forest-based Translation

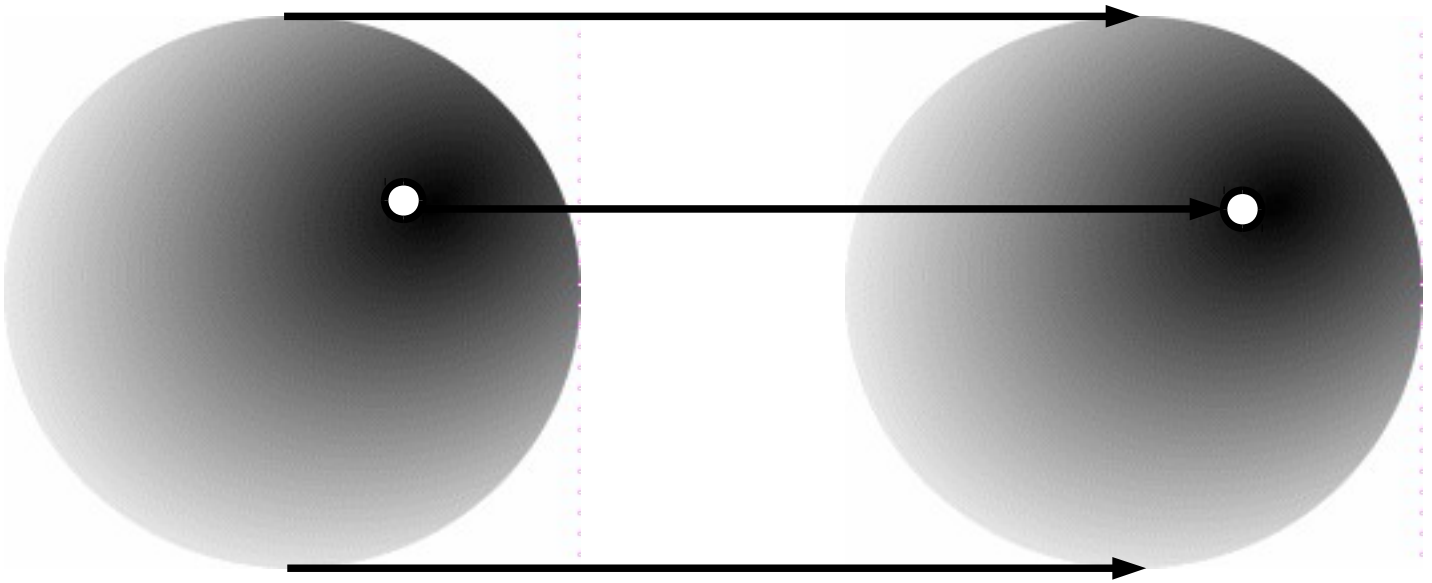


Probabilistic Distribution of Parsing Space

Probabilistic Distribution of Translation Space

Search Space Comparison

Joint Parsing and Translation



Probabilistic Distribution of Parsing Space

Probabilistic Distribution of Translation Space

Our Work: Tree-to-String Model

Constituent-to-String Model

Tree-based Translation

Forest-based Translation

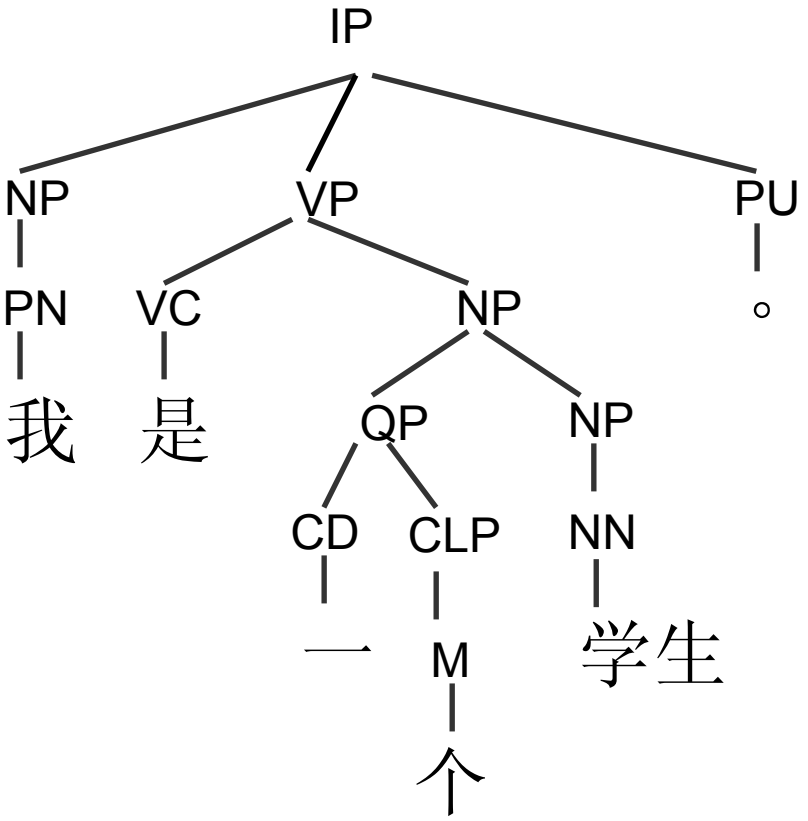
Joint Parsing and Translation

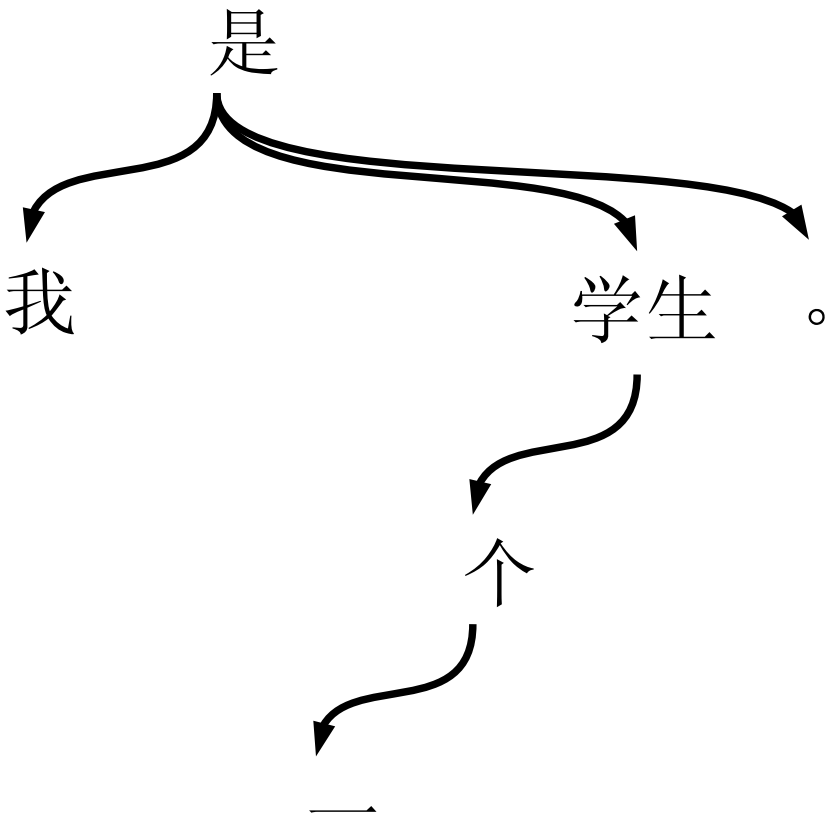
Dependency-to-String Model

Denpendency-to-String Model

- Jun Xie, Haitao Mi and Qun Liu, A novel dependency-to-string model for statistical machine translation, in Proceedings of EMNLP2011, July 27–31, 2011, Edinburgh, Scotland, UK.

Constituent vs Dependency





Constituent vs Dependency

	Constituent	Dependency
Node	Category	Word
Head Word	No	Yes
Number of Nodes	2*N	N
Parsing	Slow	Fast
Nonterminals	Yes	No

Dependency-based Models

- Dependency is regarded as a promising model because of its simplicity and its direct description of the relations between words
- Dependency is also regarded as a bridge between the syntax structure and the semantic structure
- Dependency has been successfully used to resolve many different NLP problems
- However, the attempt to build a dependency-based statistical translatoin model is not successful

Dependency-based Translation Models

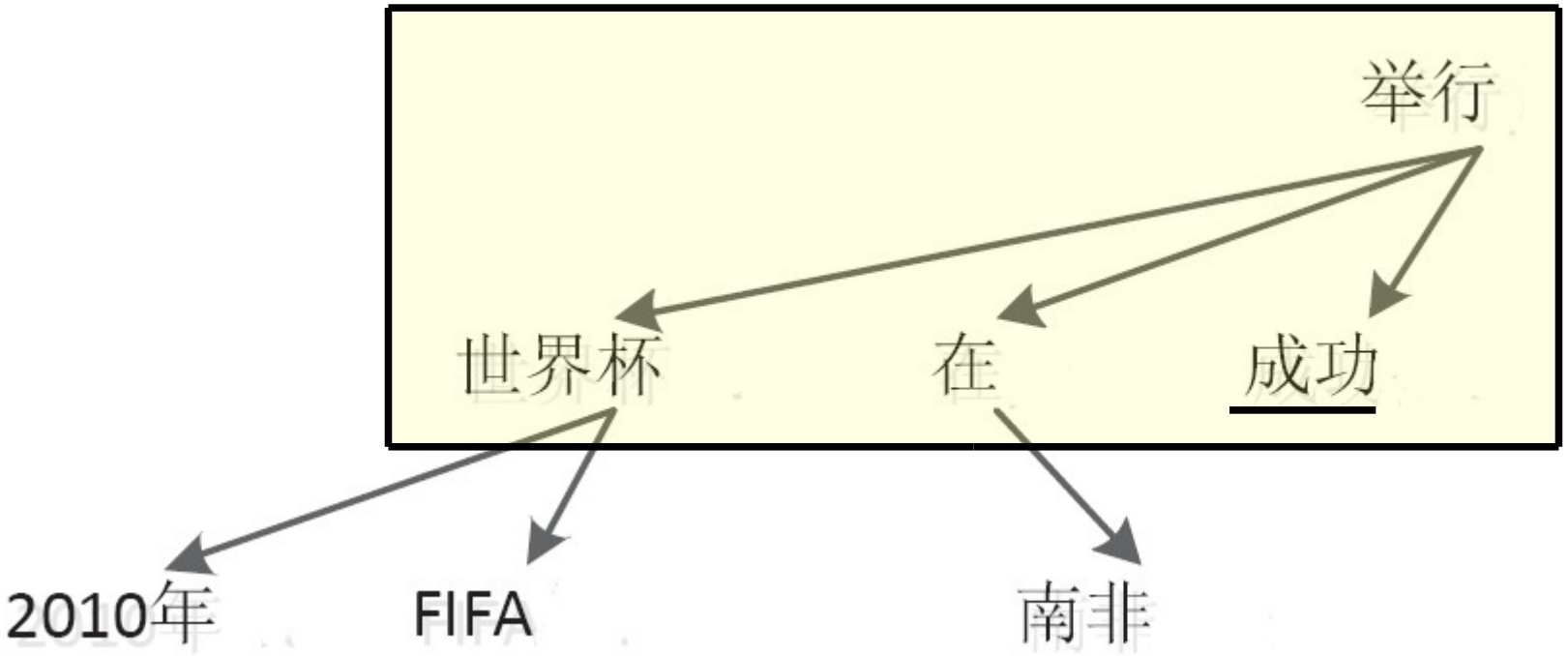
Chris Quirk, Arul Menezes, and Colin Cherry. 2005.Dependency treelet translation: Syntactically informed phrasal smt. In Proceedings of ACL 2005, pages 271–279.

This model is too complicated and cannot be repeated by others

Deyi Xiong, Qun Liu, and Shouxun Lin. 2007. A dependency treelet string correspondence model for statistical machine translation. In Proceedings of the Second Workshop on Statistical Machine Translation, pages 40–47, Prague, Czech Republic, June.

This model is over-flexible and of low performance

Dependency-to-Tree Rule

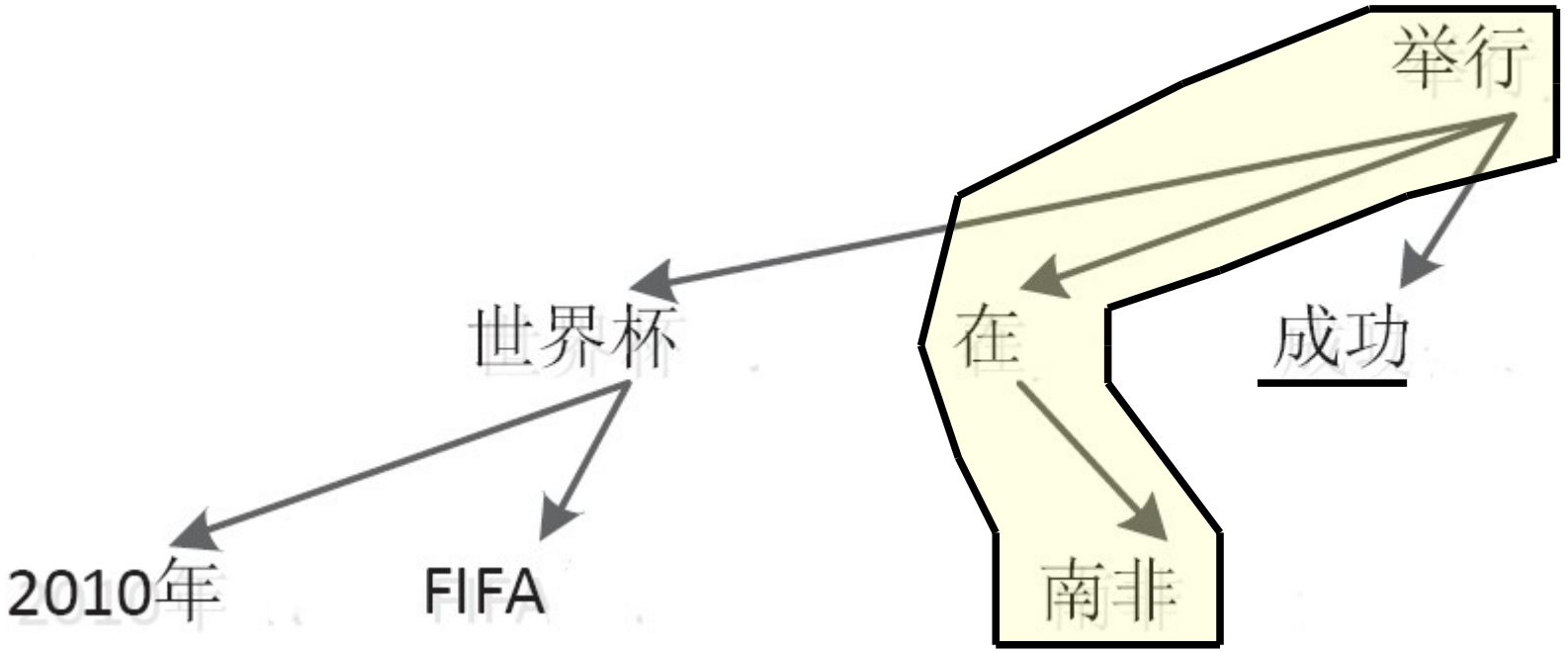


(* 世界杯 *) (在 *) (成功) 举行 →

(*FIFA*) was (successfully) held (*)

This rule is a really good rule, but it is too specific.
This kind of rules may seldom be matched.

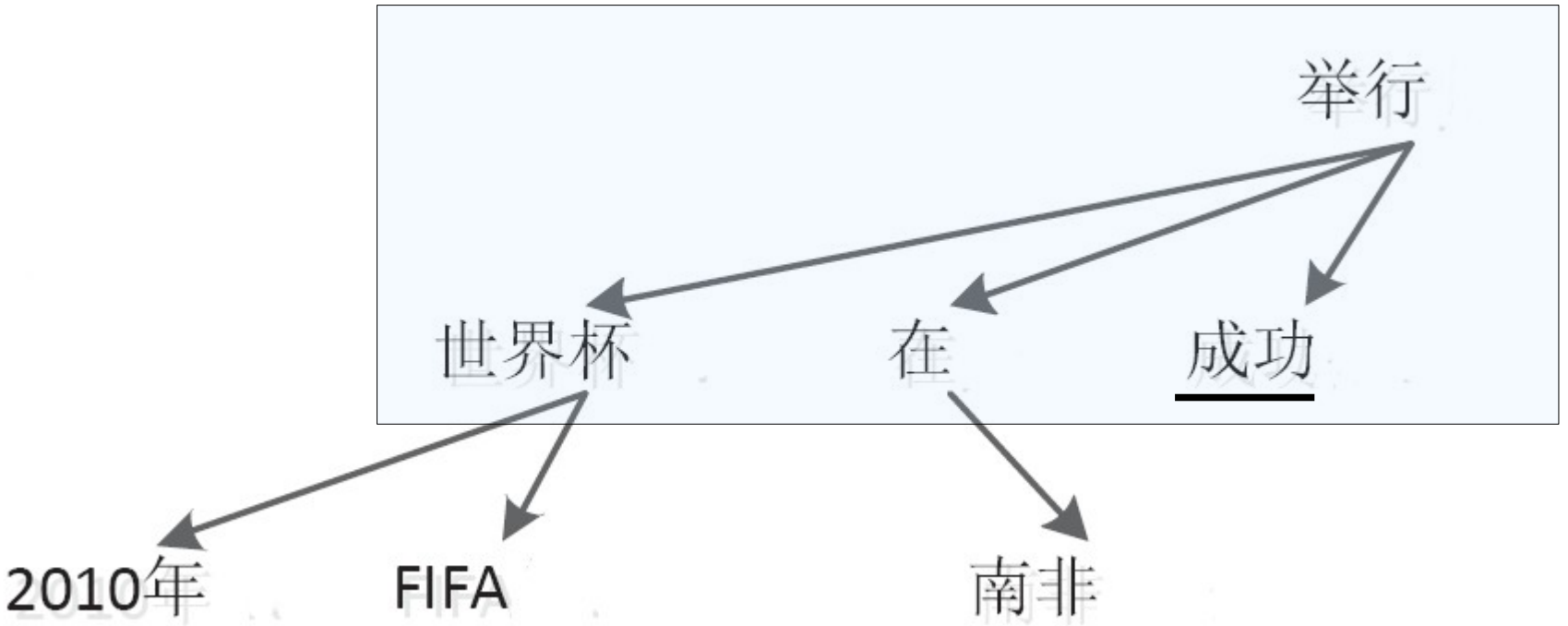
Dependency-Treelet-to-String Rule



* 在 * 南非 * 举行 * → was hold in South Africa

This rule is too flexible.
The target order of translation is difficult to be modeled.

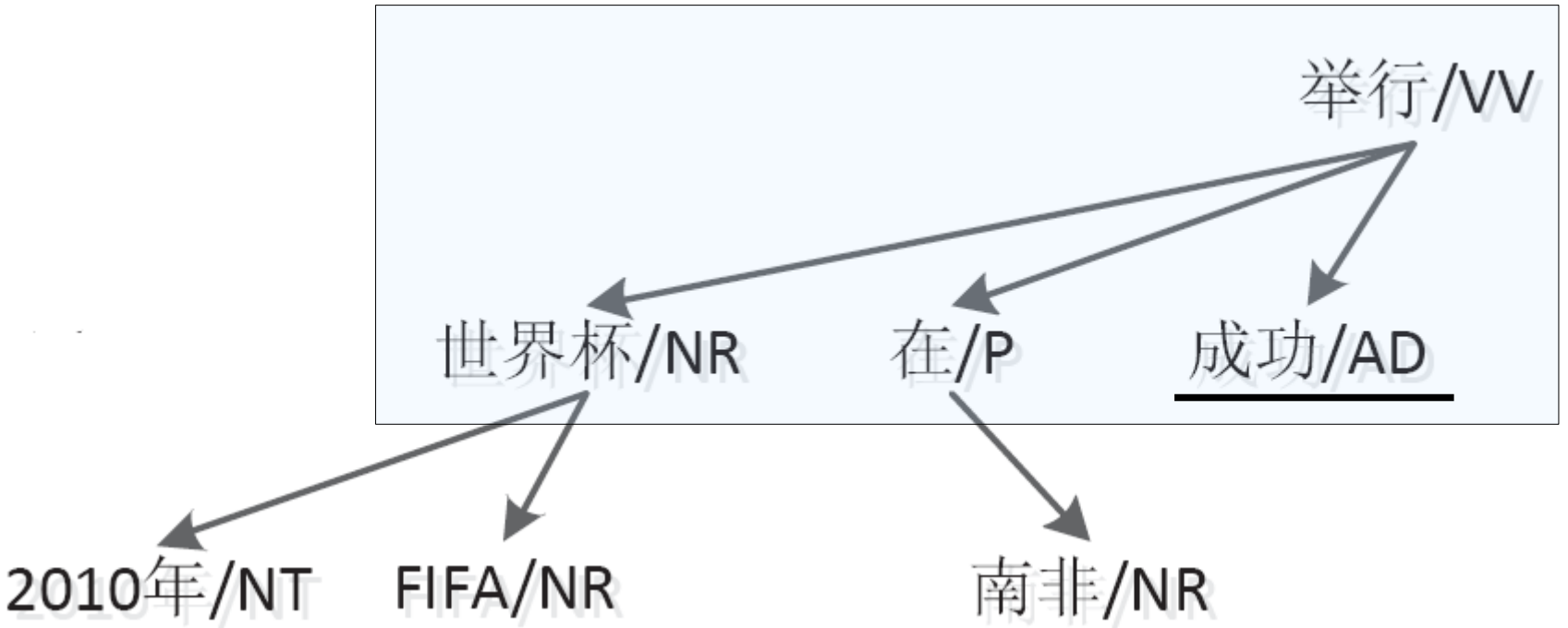
A novel Dependency-to-String Model



Our New Approach：

- Exact a rule on a whole subtree, rather than a treelet
-

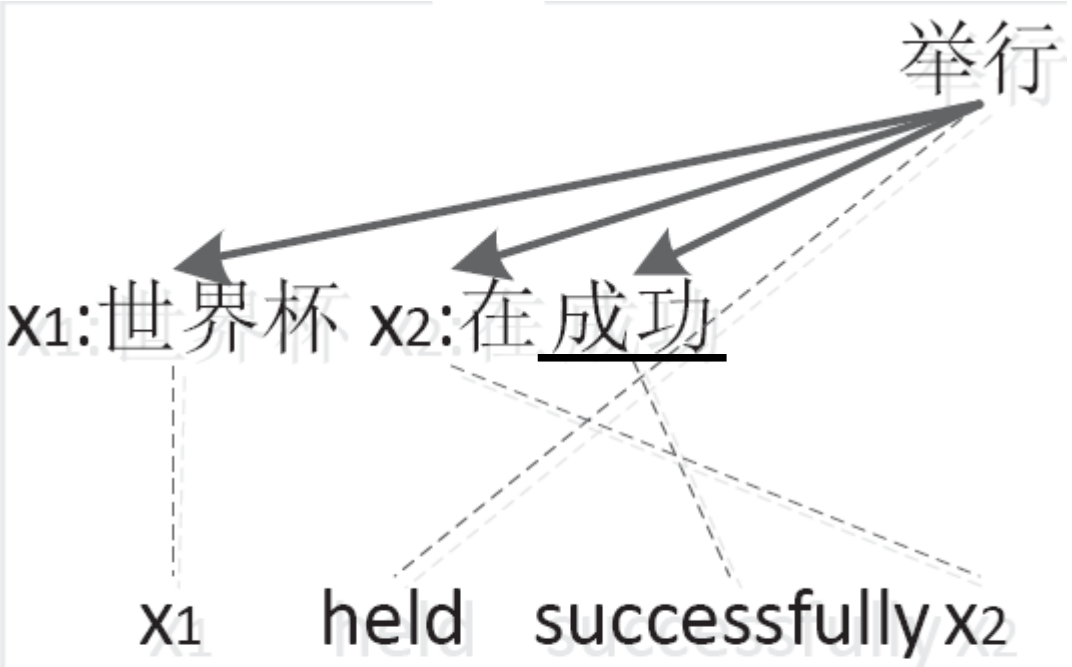
A novel Dependency-to-String Model



Our New Approach：

- Exact a rule on a whole subtree, rather than a treelet
- Generalize the rule using POS tags

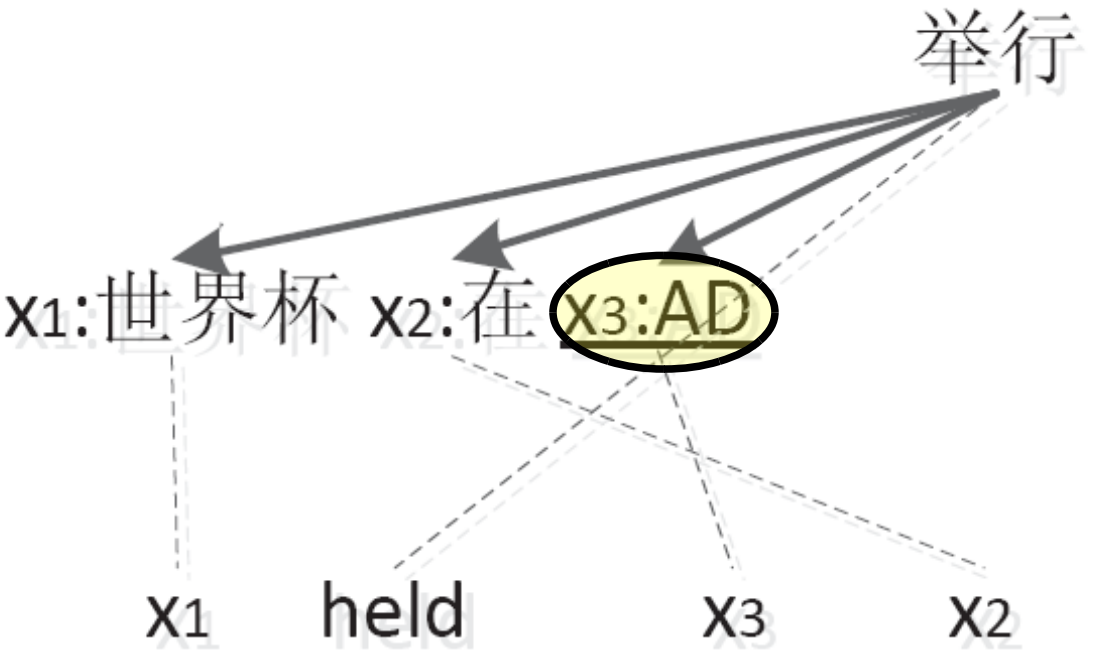
Original Rule



(x1:世界杯)(x2:在)(成功) 举行

→ x1 held successfully x2

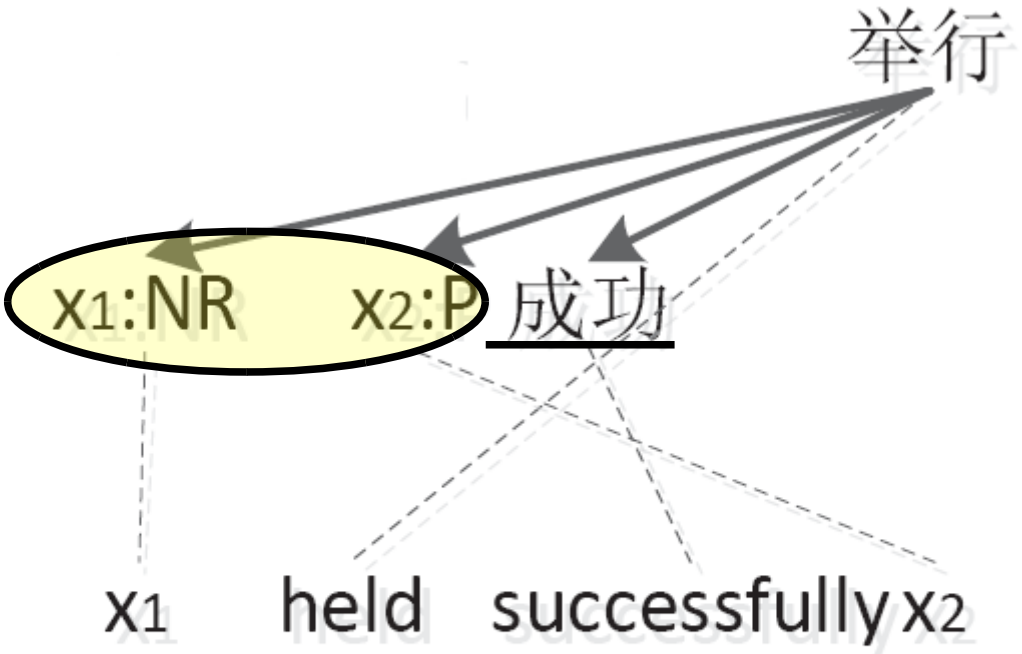
Generalized Rule #1



(x1:世界杯)(x2:在)(x3:AD) 举行

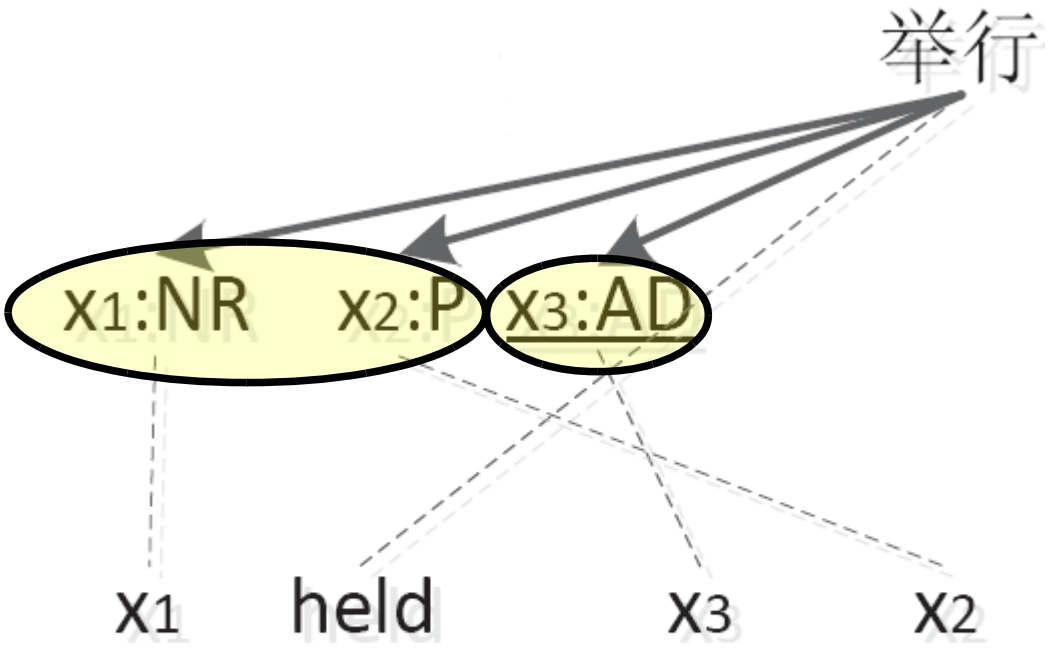
→ x1 held x3 x2

Generalized Rule #2



(x1:NR)(x2:P)(成功) 举行
→ x1 held successfully x2

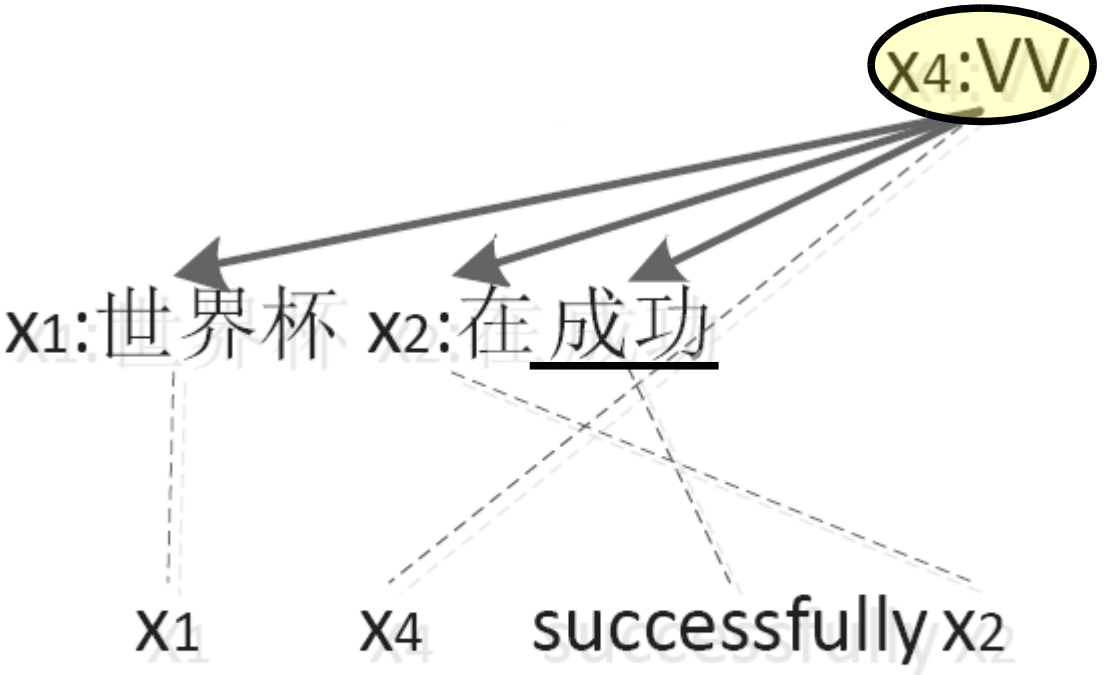
Generalized Rule #3



(x1:NR)(x2:P)(x3:AD) 举行

→ x1 held x3 x2

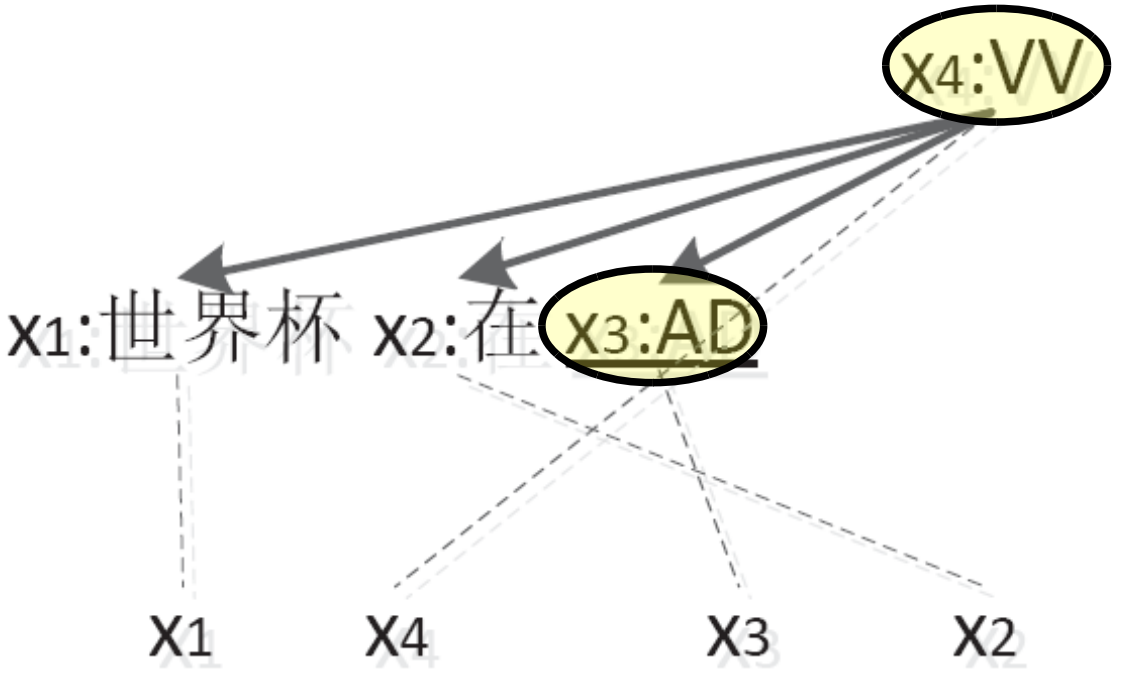
Generalized Rule #4



(x1:世界杯)(x2:在)(成功) x4:VV

→ x1 x4 successfully x2

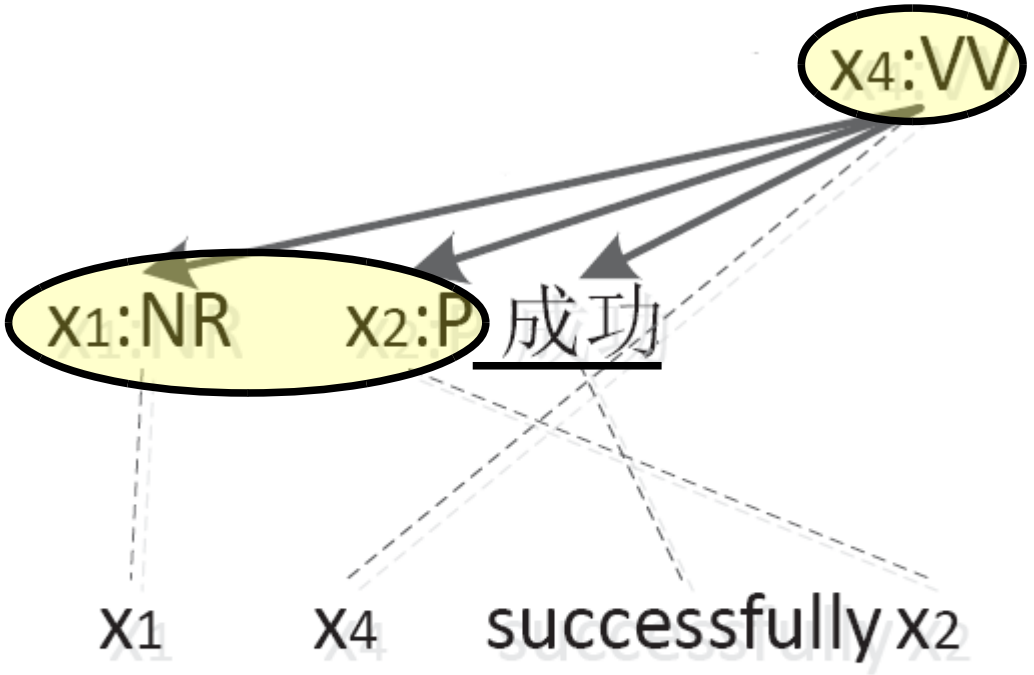
Generalized Rule #5



(x1:世界杯)(x2:在)(x3:AD) x4:VV

→ x1 x4 x3 x2

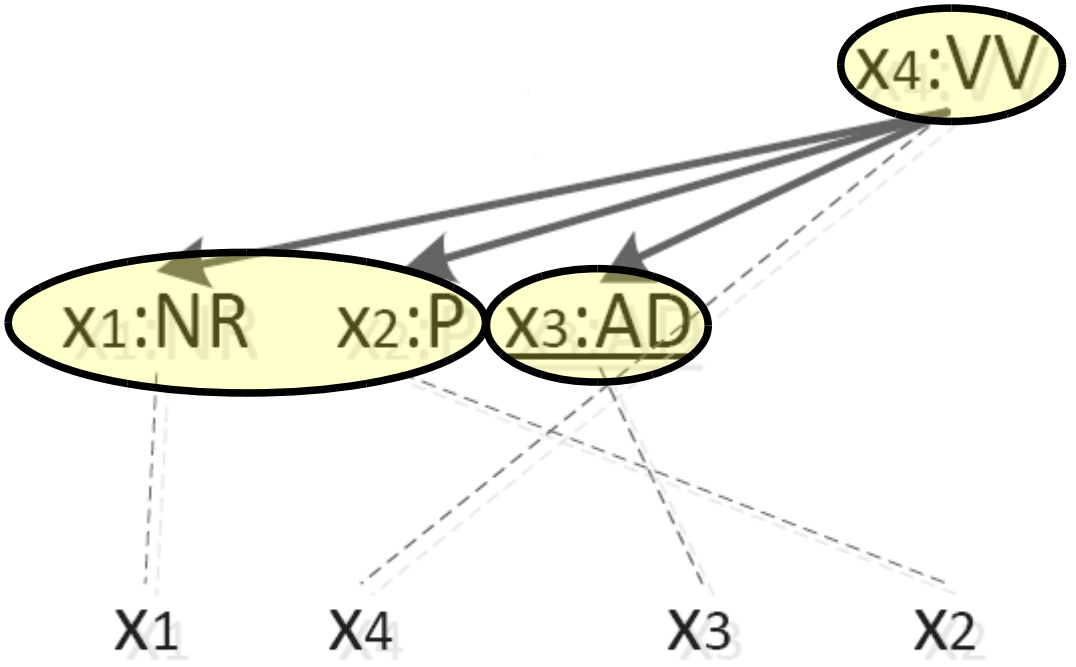
Generalized Rule #6



(x1:NR)(x2:P)(成功) x4:VV

→ x1 x4 successfully x2

Generalized Rule #7



(x1:NR)(x2:P)(x3:AD) x4:VV

→ x1 x4 x3 x2

Experiment Results

System	Rule #	MT04(%)	MT05(%)
cons2str	30M	34.55	31.94
hiero-re	148M	35.29	33.22
dep2str	56M	**35.82**$^{*+}$	**33.62**$^{+}$

Outline

Background

Tree-to-String Model

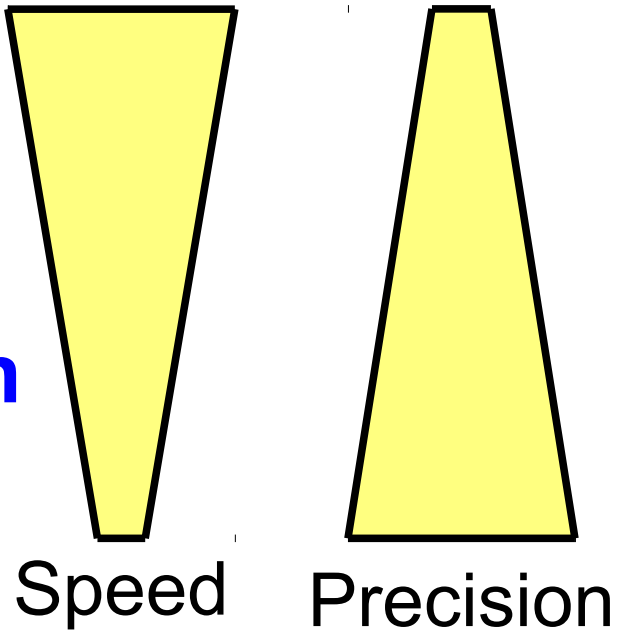
Conclusion

Conclusion

- We proposed two kinds of tree-to-string translation model based on source side syntax structure:
  - Constituent-to-String Model
  - Dependency-to-String Model

Conclusion

- For constituent-to-string model, we proposed three translation approaches:
  - **Tree-based Translation**
  - **Forest-based Translation**
  - **Joint Parsing and Translation (String-based translation)**



Future Work

- Semantic-base Translation Model
- Structural Language Model

Thanks

http://nlp.ict.ac.cn