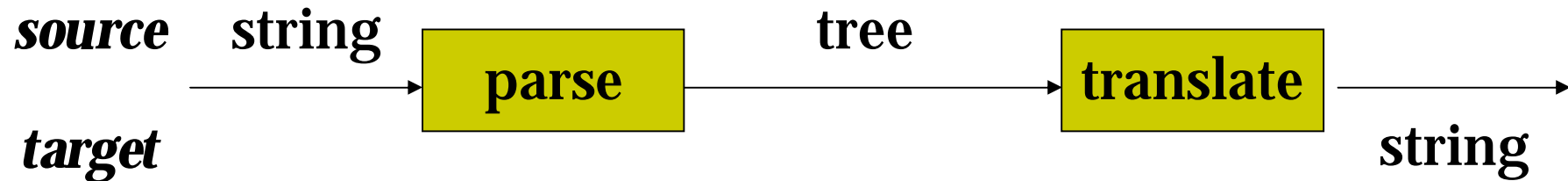# Joint Parsing and Translation

Yang Liu and Qun Liu

Institute of Computing Technology

Chinese Academy of Sciences

INSTITUTE OF COMPUTING TECHNOLOGY
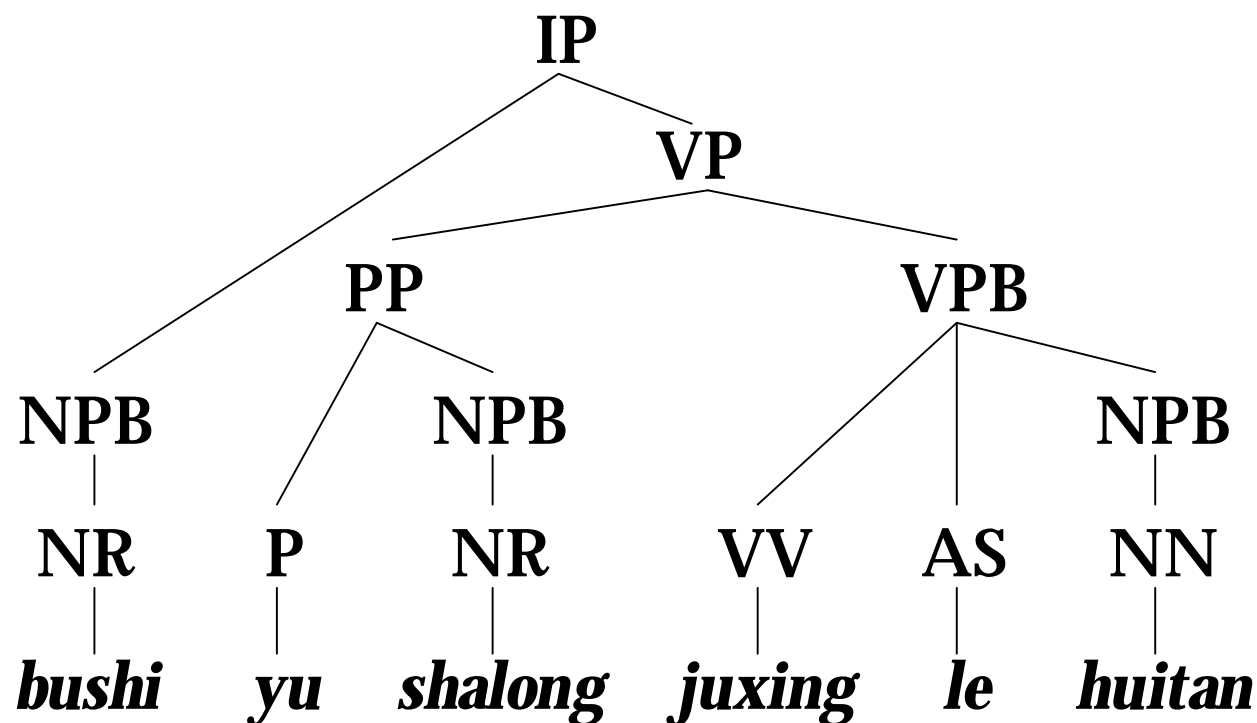
Separate Parsing and Translation

source string → [**parse**] → tree → [**translate**] → string

target

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

Separate Parsing and Translation

Parse the source sentence into a tree

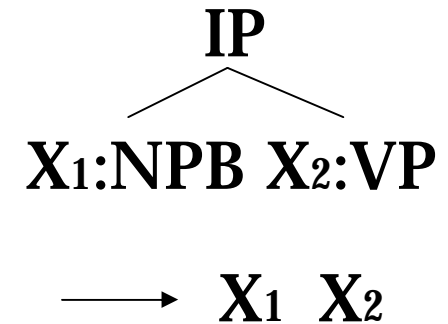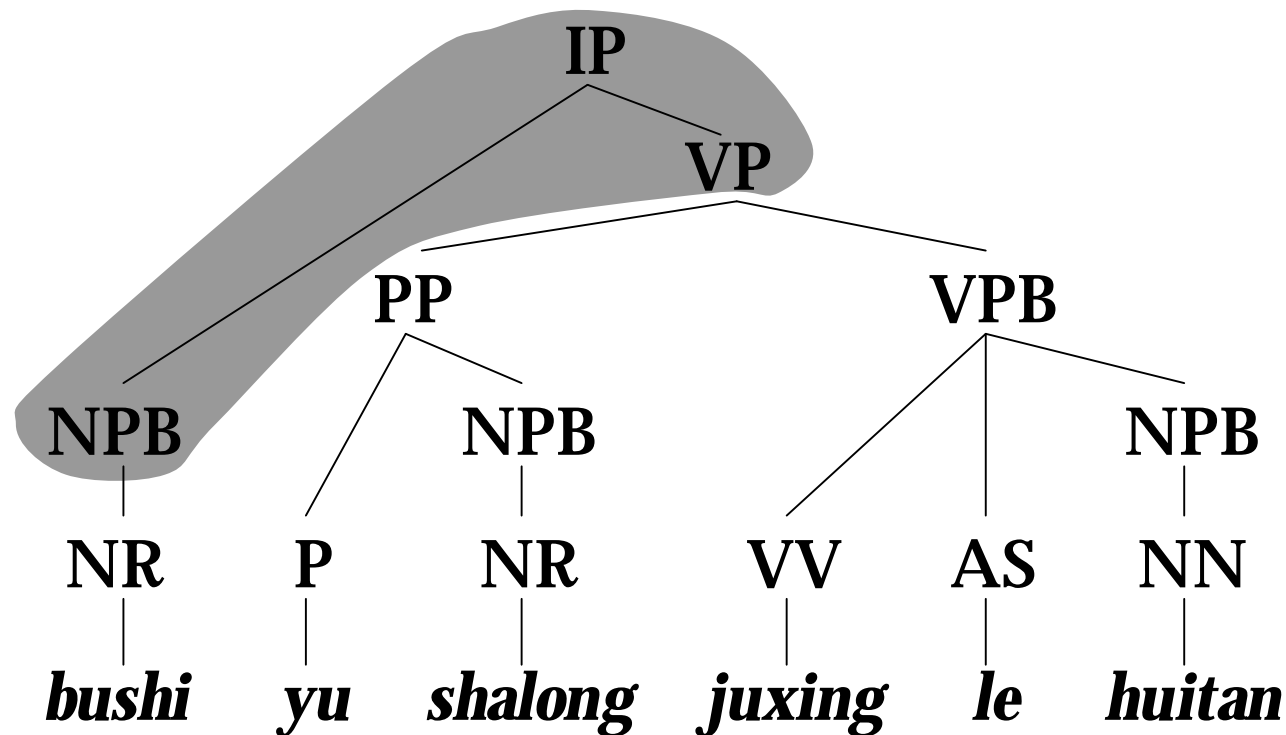*(Liu et al., 2006; Huang et al., 2006)*

Separate Parsing and Translation
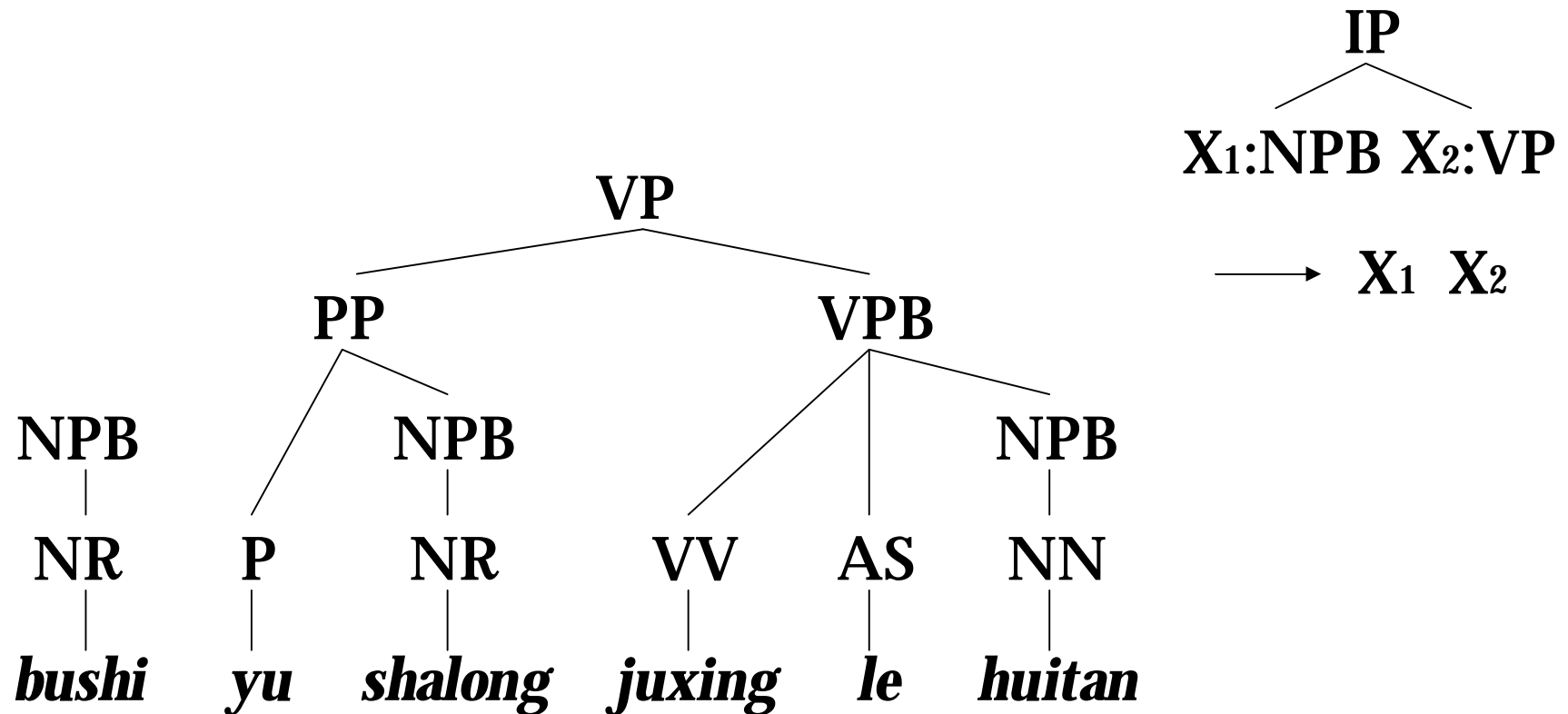
- Transfer the tree into a target string

IP
NPB ─ VP
PP ─ VPB
NPB
NR   P   NR   VV   AS   NN
bushi *yu* *shalong* *juxing* *le* *huitan*

IP
X_1:NPB X_2:VP

$\longrightarrow$ X_1 X_2

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

Separate Parsing and Translation

Transfer the tree into a target string

IP
X₁:NPB X₂:VP

→ X₁ X₂

VP

PP   VPB

NPB   NPB   NPB

NR   P   NR   VV   AS   NN

bushi   *yu*   *shalong*   *juxing*   *le*   *huitan*

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

Separate Parsing and Translation

l **Transfer the tree into a target string**

NPB
|
NR
|
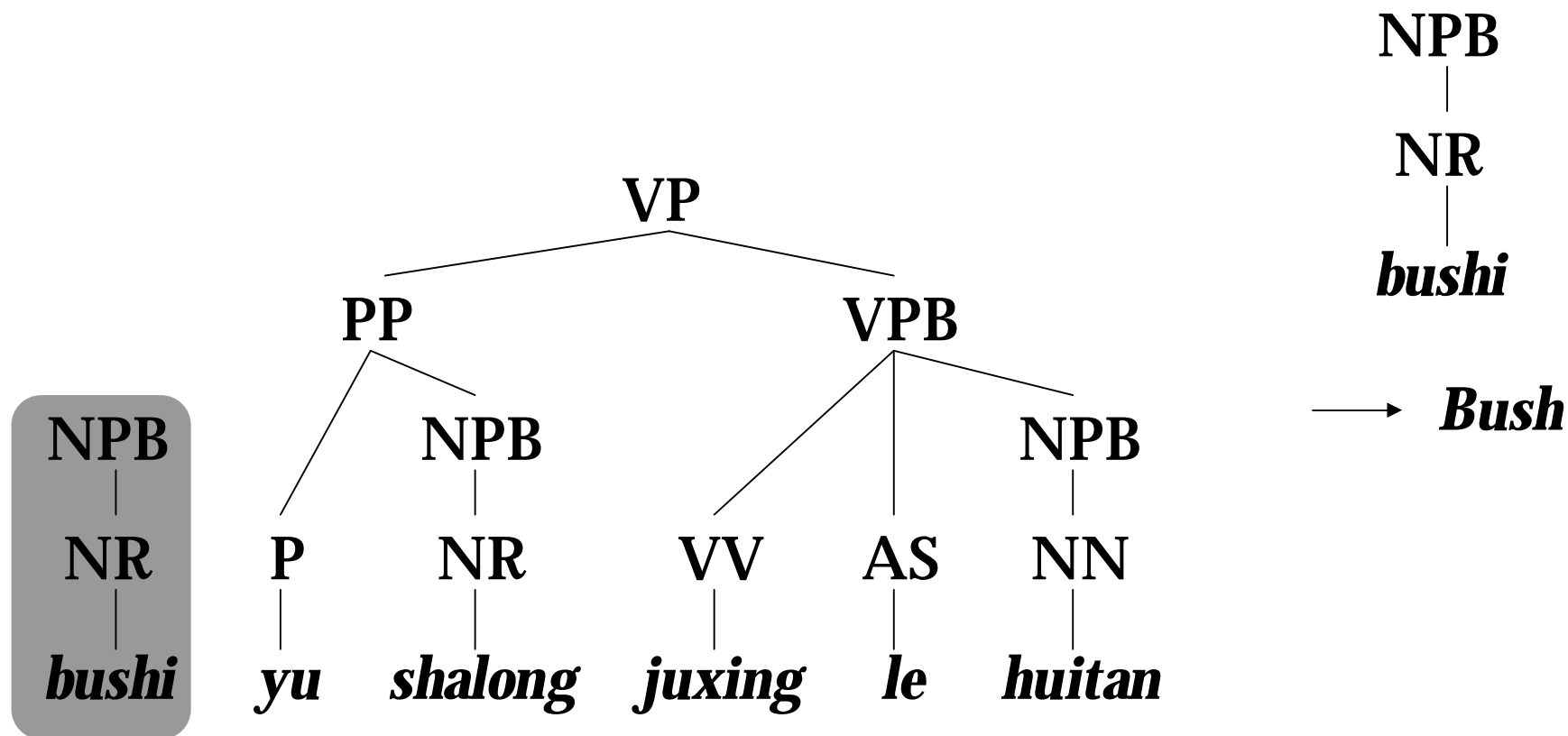*bushi*

→ *Bush*

VP

PP

NPB
|
NR
|
bushi

P
|
yu

NPB

NR
|
shalong

VPB

VV
|
juxing

AS
|
le

NPB

NN
|
huitan

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

Separate Parsing and Translation

- **Transfer the tree into a target string**

NPB

NR

bushi

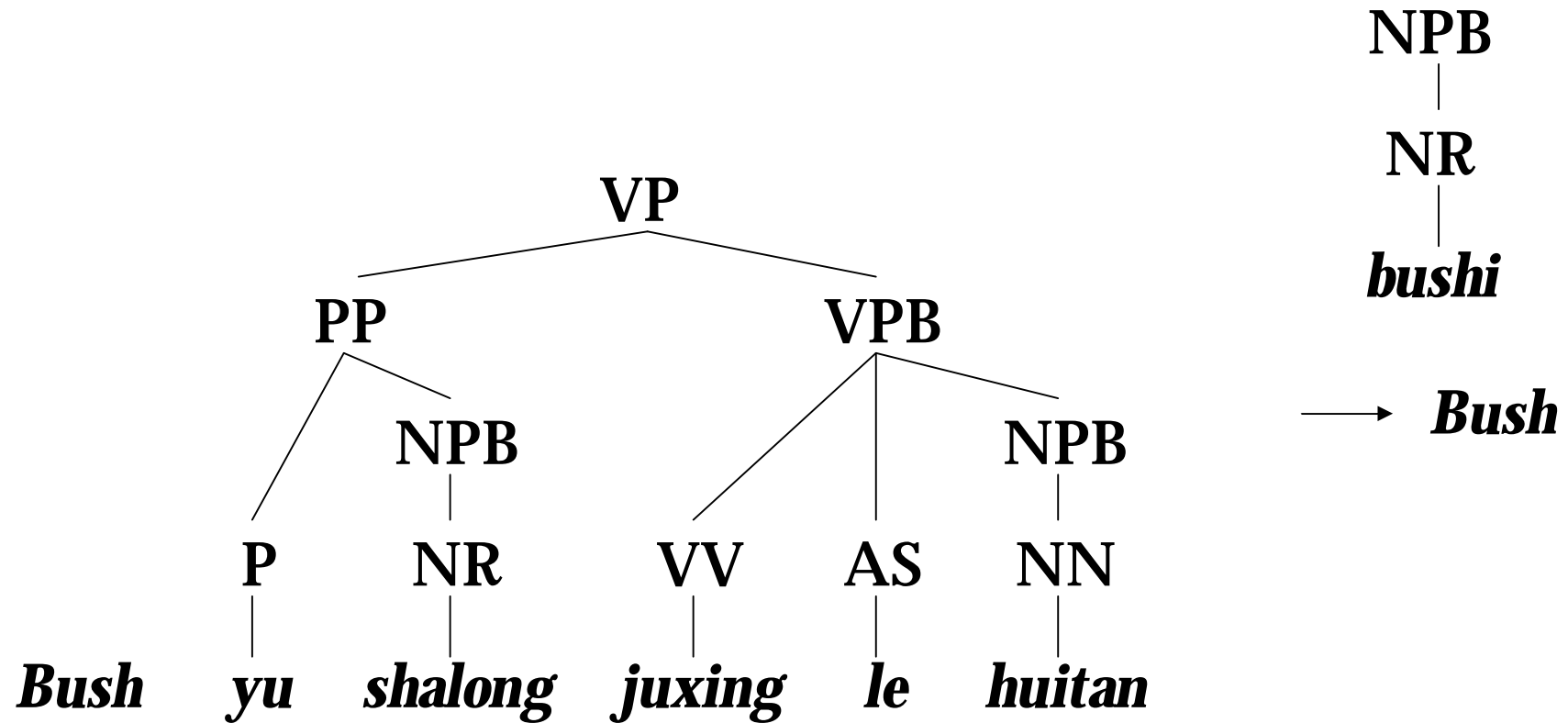VP

PP   VPB

NPB   NPB

P   NR   VV   AS   NN

Bush   yu   shalong   juxing   le   huitan

→ *Bush*

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

Separate Parsing and Translation

I **Transfer the tree into a target string**

$$VP \rightarrow X_1{:}PP \; X_2{:}VPB$$

$$\longrightarrow X_2 \; X_1$$

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

Separate Parsing and Translation

ꞮTransfer the tree into a target string

$$VP$$
$$X_1:PP \quad X_2:VPB$$
$$\longrightarrow \quad X_2 \quad X_1$$

VPB
NPB
PP
NPB

VV AS NN P NR

Bush juxing le huitan yu shalong

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

Separate Parsing and Translation

- Transfer the tree into a target string

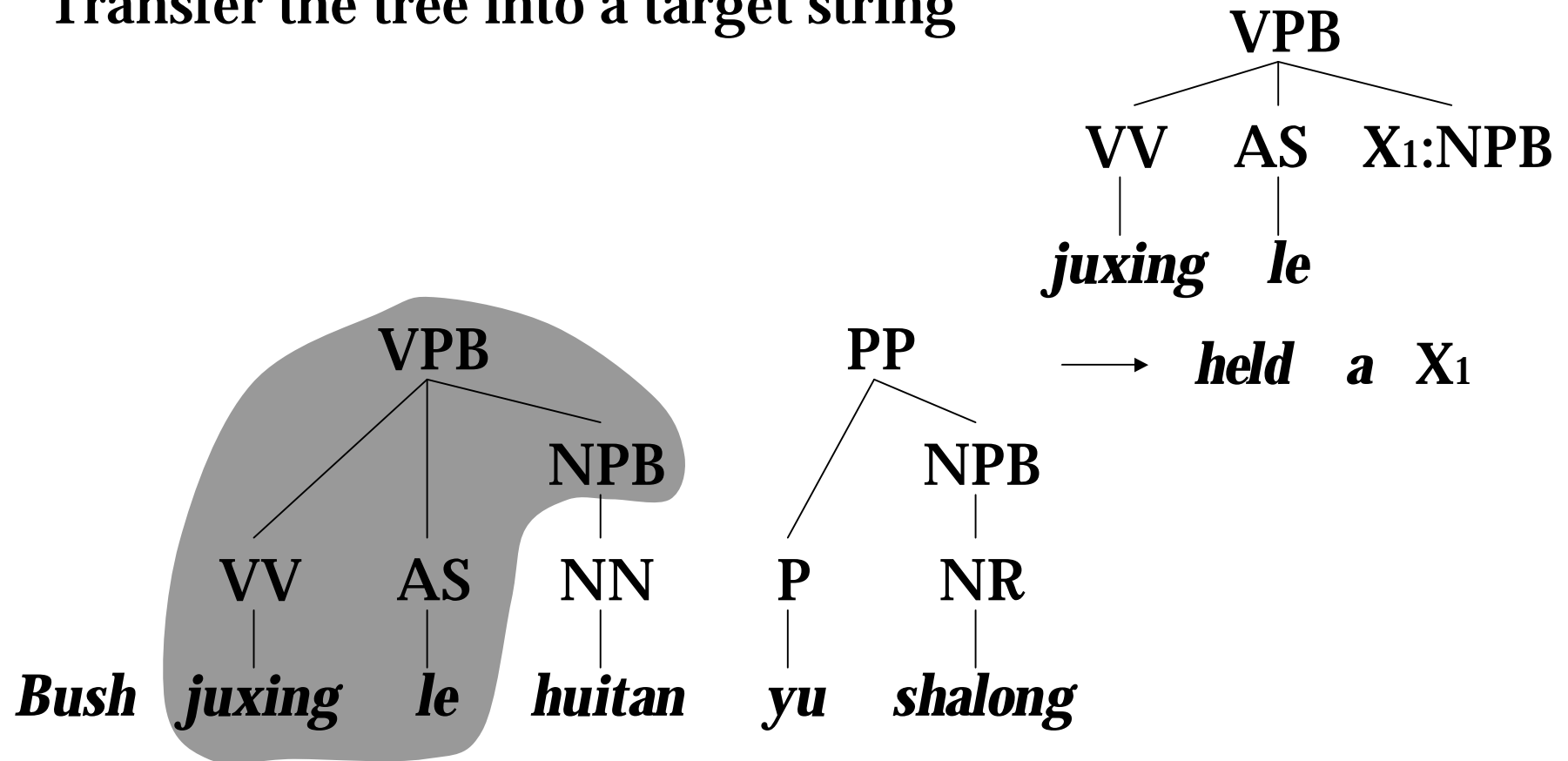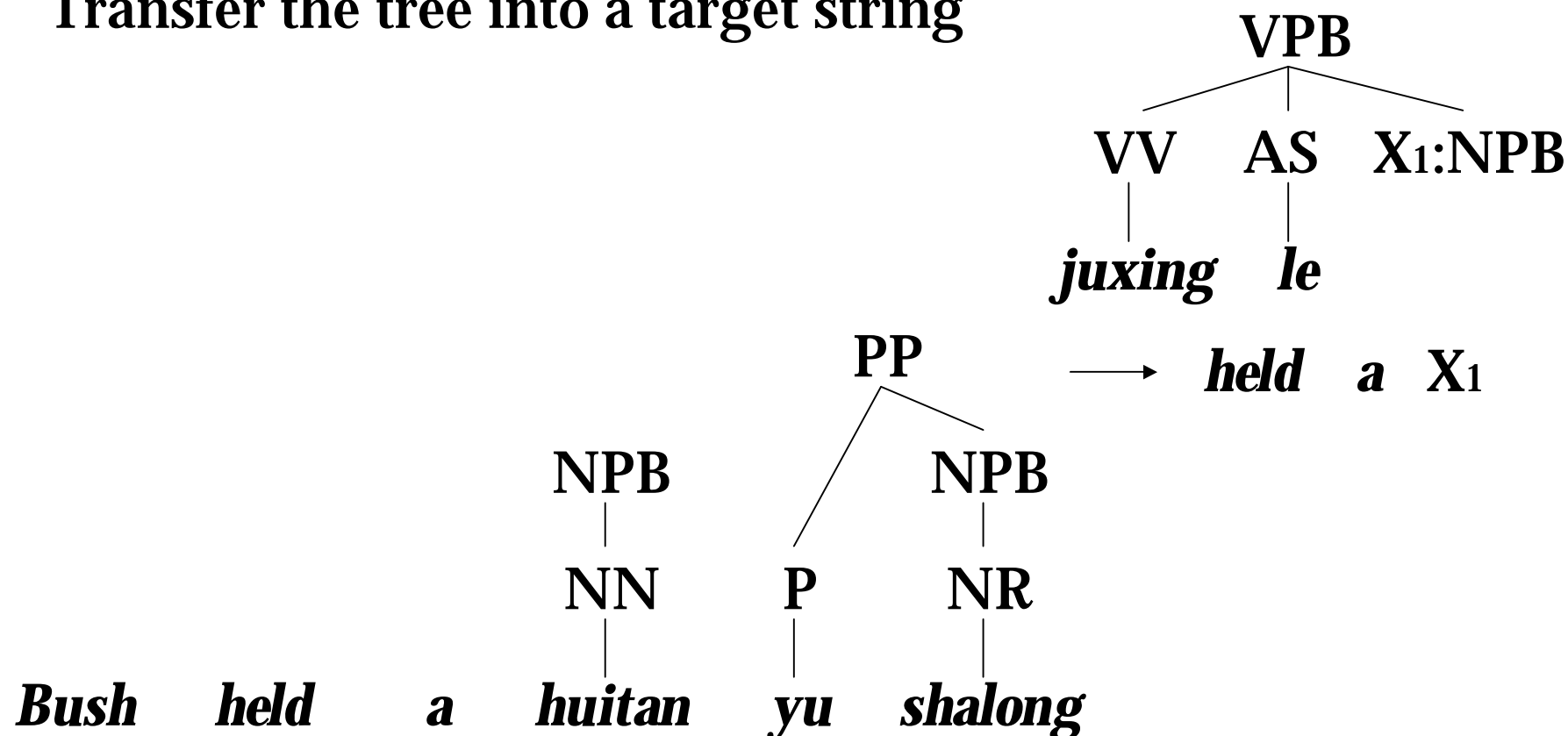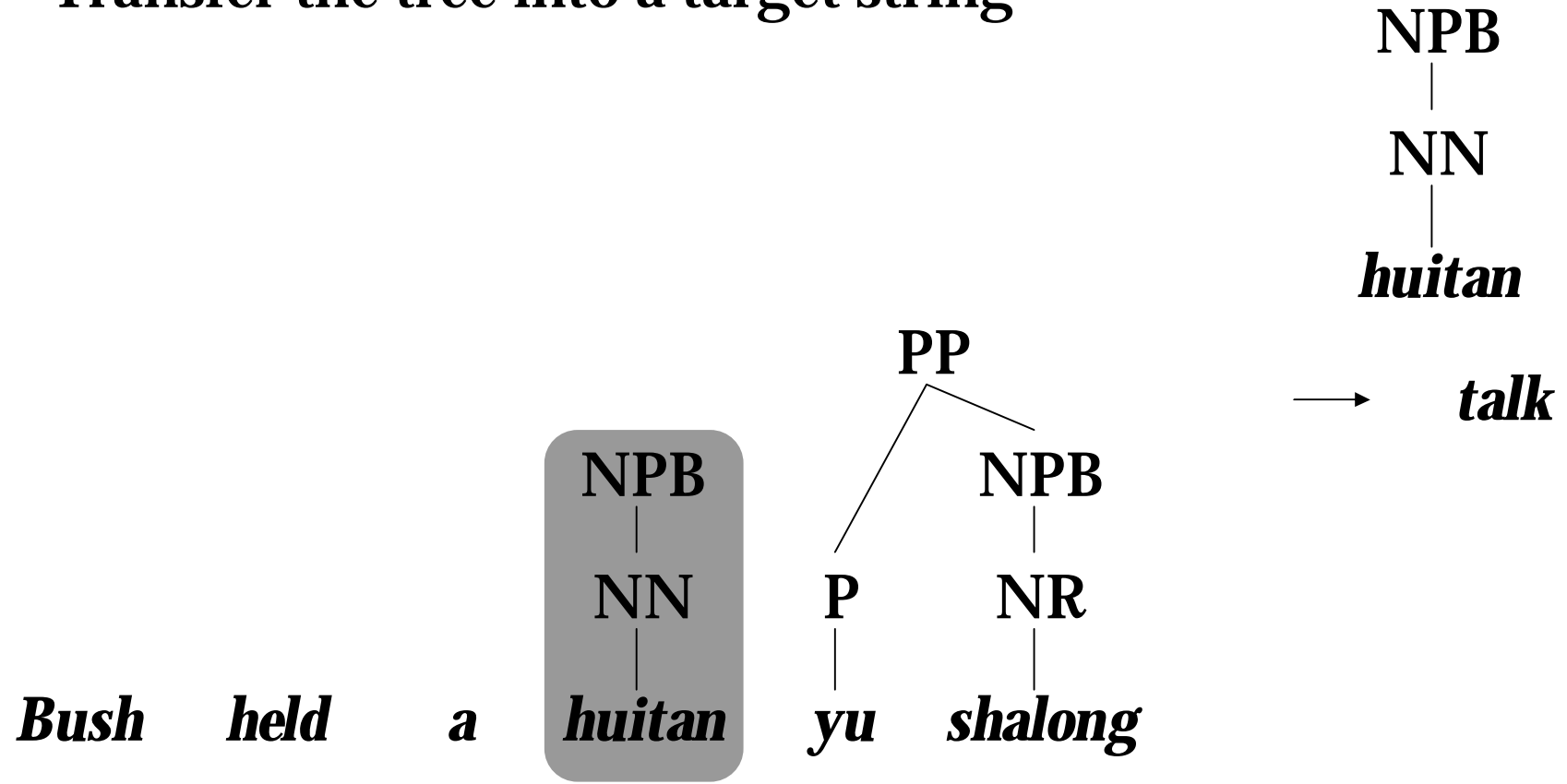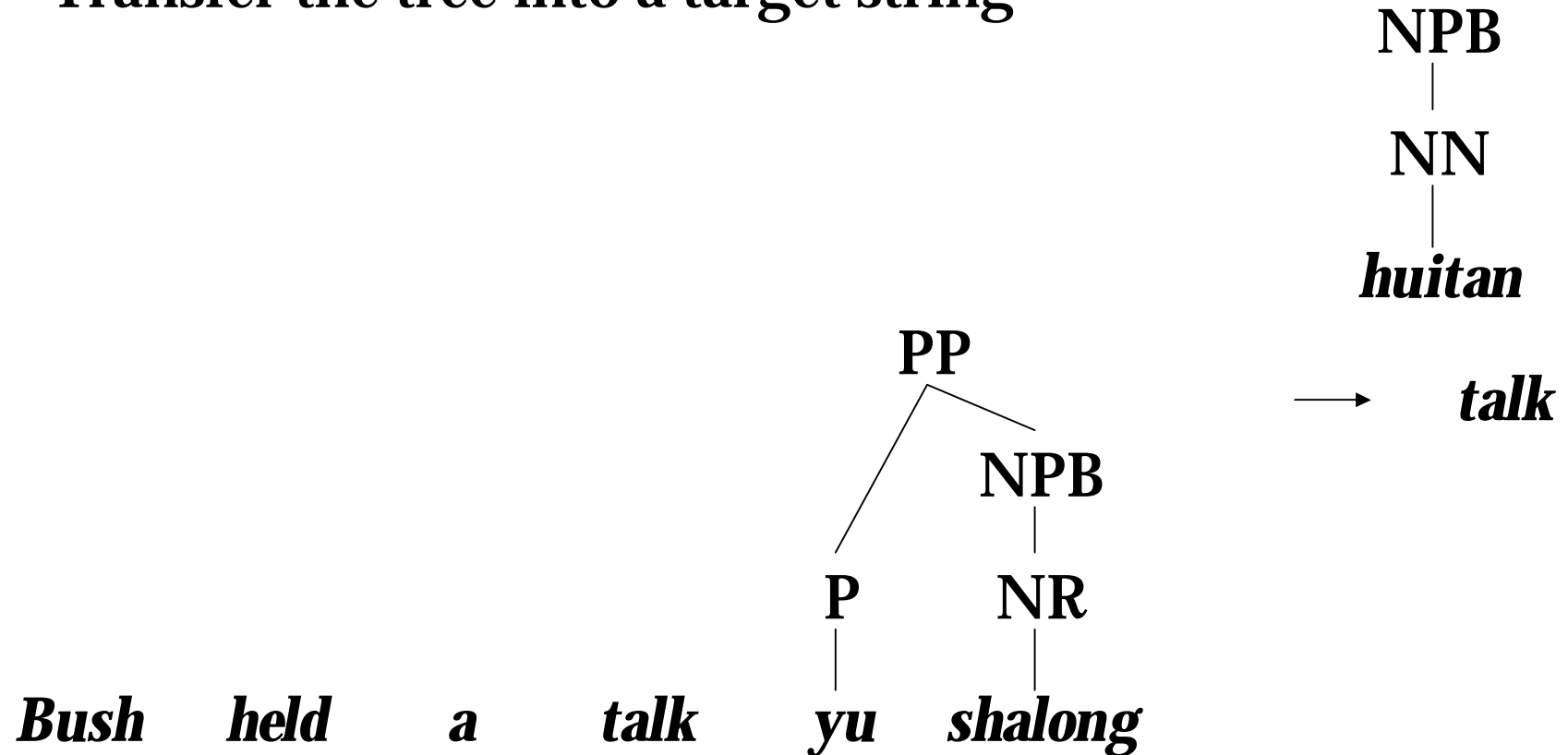(Liu et al., 2006; Huang et al., 2006)

Separate Parsing and Translation
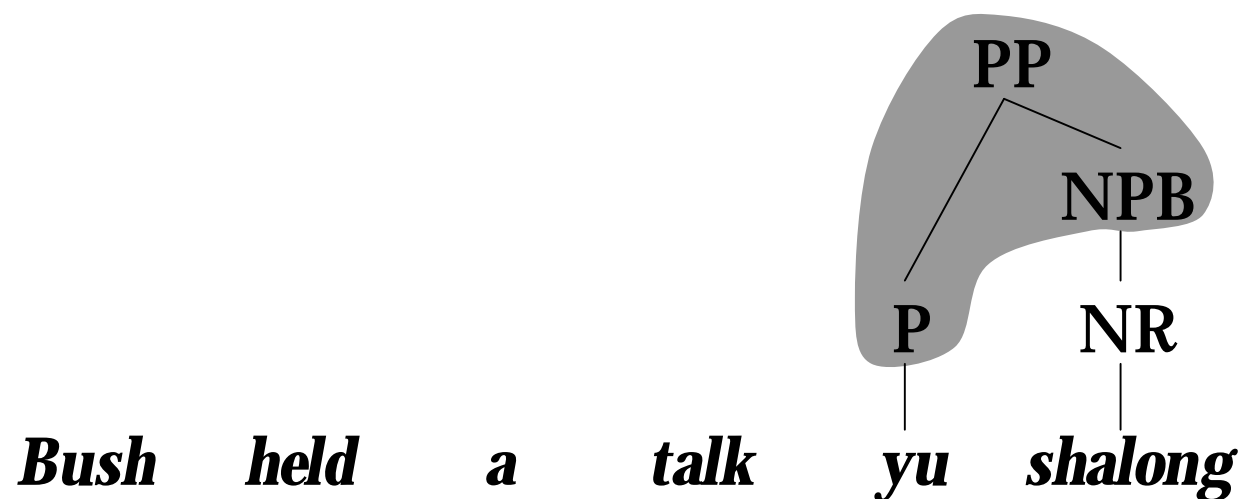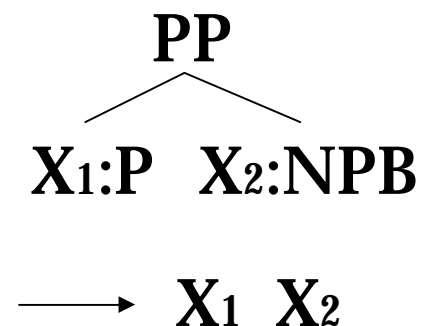
Transfer the tree into a target string

VPB

VV   AS X_1:NPB

juxing   *le*

PP $\longrightarrow$ *held   a X_1*

NPB   NPB

NN   P   NR

Bush   held   a   huitan   yu   shalong

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

Separate Parsing and Translation

- **Transfer the tree into a target string**

NPB
|
NN
|
huitan

PP

$\longrightarrow$ **talk**

NPB
|
NN
|
huitan

NPB

P   NR

Bush   **held**   **a**   *huitan*   *yu*   *shalong*

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

Separate Parsing and Translation

- Transfer the tree into a target string

NPB
|
NN
|
huitan

$\longrightarrow$ *talk*

PP
/ \
   NPB
    |
P   NR
| |

Bush   held   a   talk   yu   shalong

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

Separate Parsing and Translation

l **Transfer the tree into a target string**

$$PP$$

$$X_1{:}P \quad X_2{:}NPB$$

$$\longrightarrow X_1 \ X_2$$

PP

NPB

P　　NR

Bush　held　a　talk　yu　shalong

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

Separate Parsing and Translation

- **Transfer the tree into a target string**

PP

X$_1$:P   X$_2$:NPB

$\longrightarrow$ X$_1$ X$_2$

NPB

P   NR

Bush   held   a   talk   yu   shalong

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

Separate Parsing and Translation

- **Transfer the tree into a target string**

P
|
yu

$\longrightarrow$ *with*

NPB
|
P   NR
| |
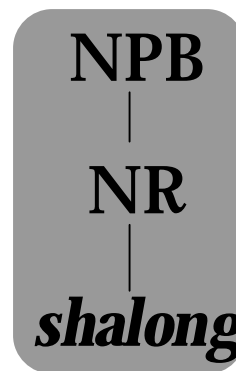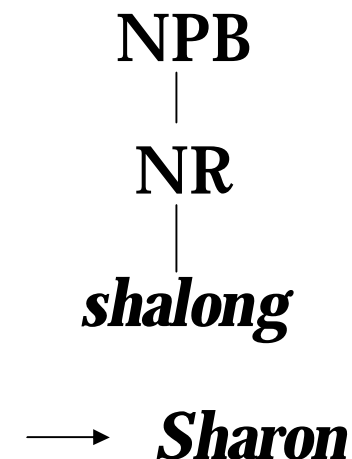**Bush   held   a   talk   yu   shalong**

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

Separate Parsing and Translation

- Transfer the tree into a target string

P
|
yu

⟶ *with*

NPB
|
NR
|

Bush   *held*   *a*   *talk*   *with*   *shalong*

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

Separate Parsing and Translation

l **Transfer the tree into a target string**

NPB
|
NR
|
shalong

→ **Sharon**

NPB
|
NR
|
shalong

Bush   held   a   talk   with   shalong

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

Separate Parsing and Translation

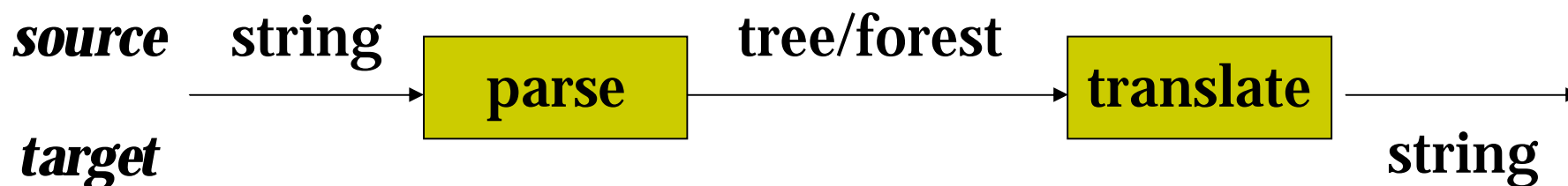- Transfer the tree into a target string

NPB
|
NR
|
shalong

→ **Sharon**

Bush held a talk with Sharon

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

Forest as Input

(Mi et al., 2008)

Joint Parsing and Translation

source string | parse | tree/forest

target translate

string
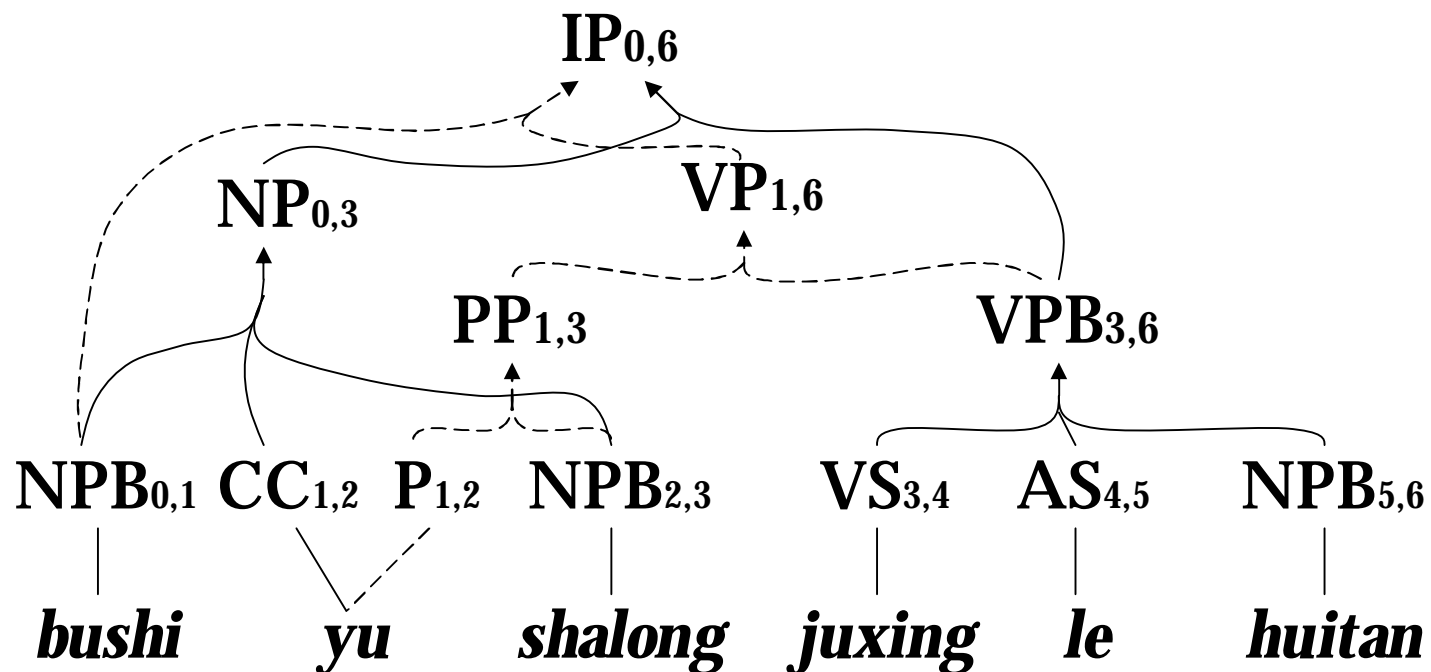
(Liu et al., 2006; Mi et al., 2008)

source string | parse + translate | tree/forest

target

string
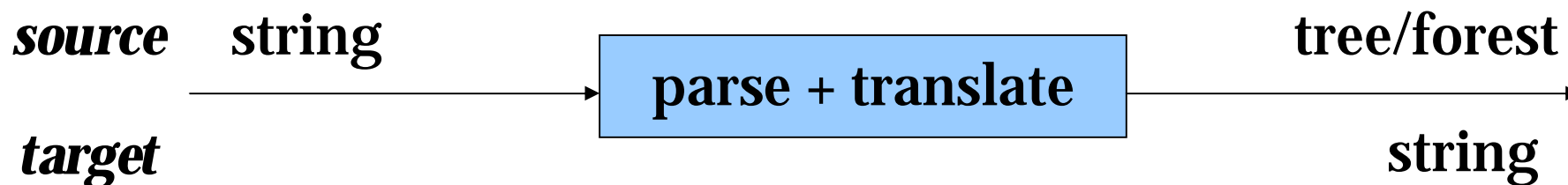
This work

Joint Parsing and Translation

NPB

NR

bushi

→ **Bush**

NPB

NR

bushi   *yu*   *shalong*   *juxing*   *le*   *huitan*

Bush

Joint Parsing and Translation
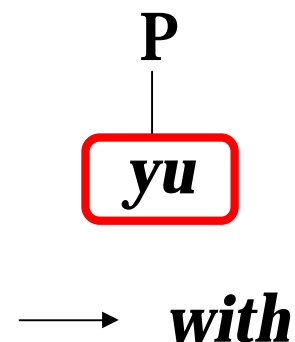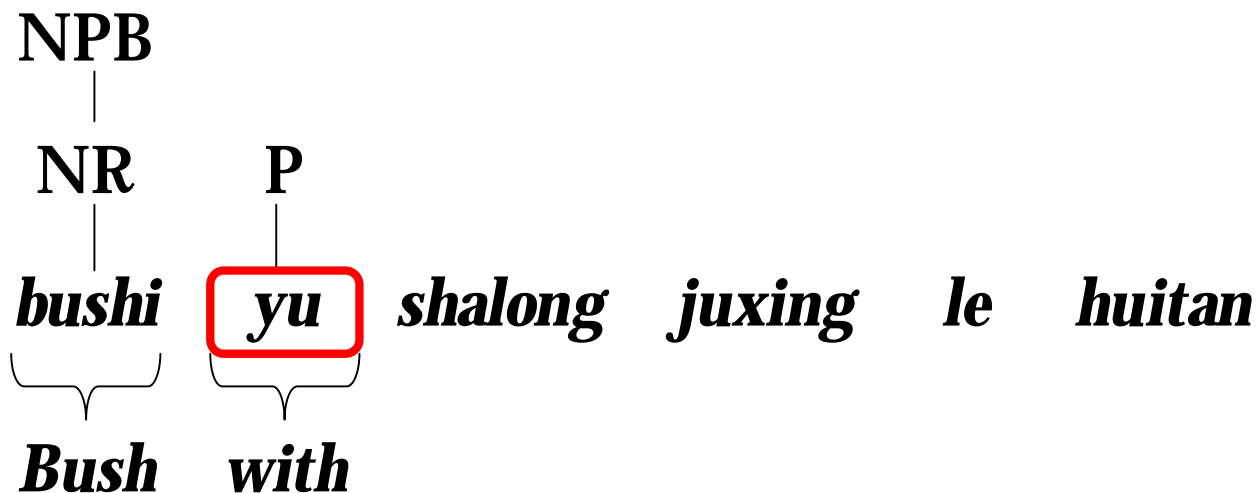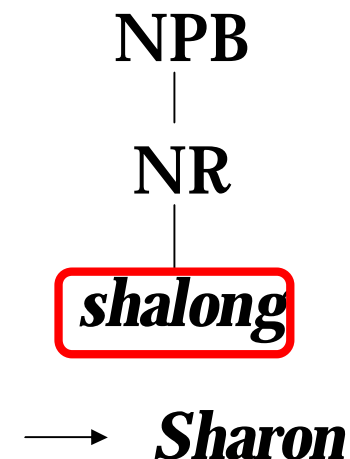
P

yu

$\longrightarrow$ *with*

NPB

NR   P

bushi   *yu*   *shalong*   *juxing*   *le*   *huitan*

Bush   *with*

Joint Parsing and Translation

NPB

NR

shalong

$\longrightarrow$ **Sharon**

NPB   NPB

NR   P   NR

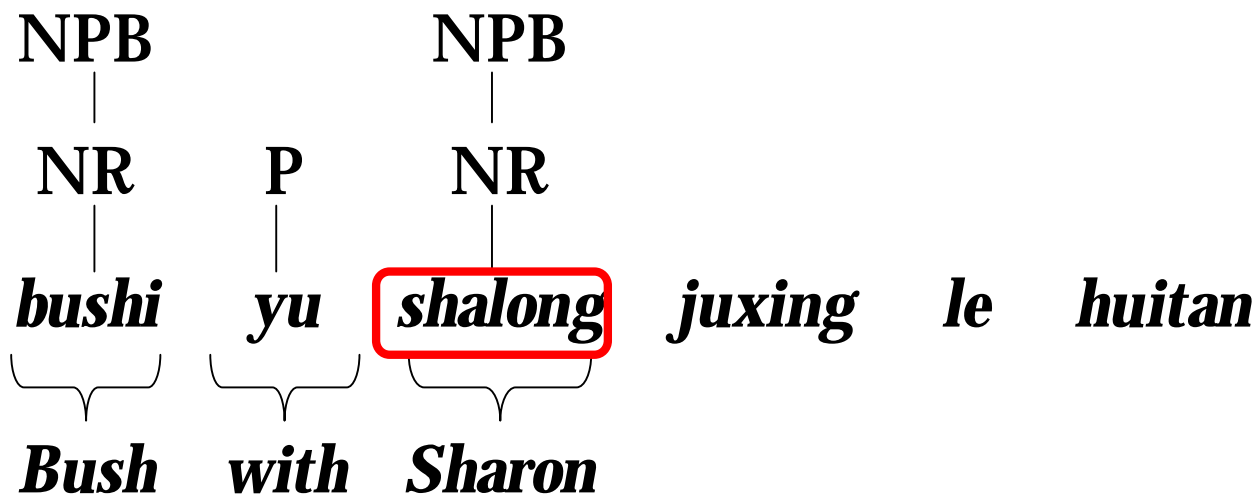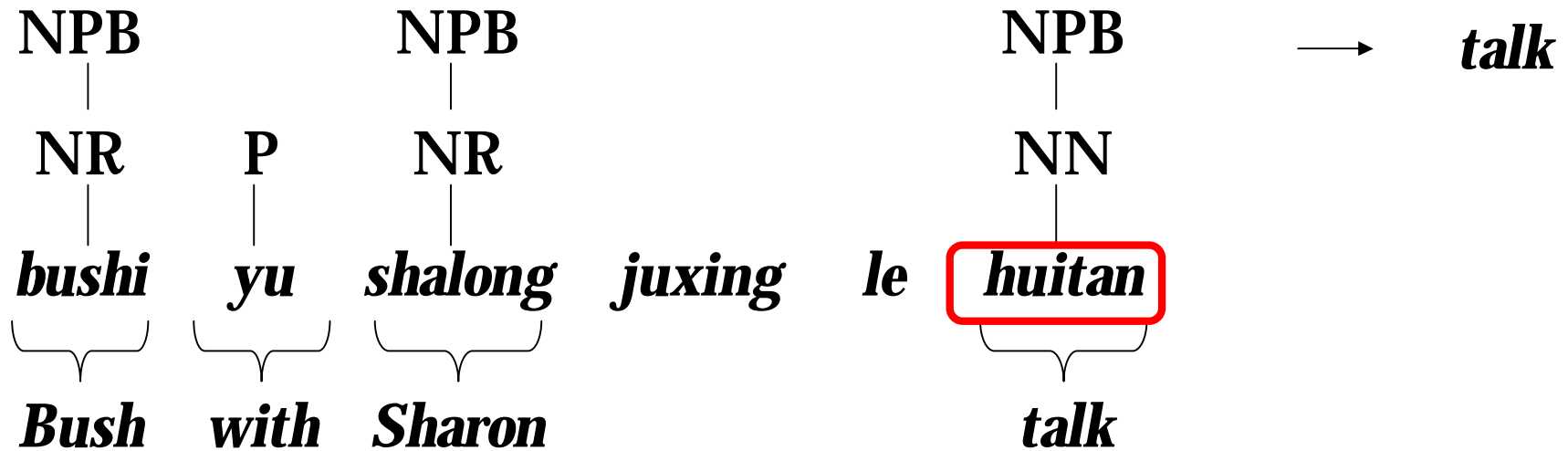***bushi   yu   shalong   juxing   le   huitan***

Bush   with   Sharon

Joint Parsing and Translation

NPB

NN

huitan

→ **talk**

NPB NPB NPB

NR P NR NN

bushi **yu** **shalong** **juxing** **le** **huitan**

Bush **with** **Sharon** **talk**

Joint Parsing and Translation

PP

X_1:P X_2:NPB

$\longrightarrow$ X_1 X_2

PP

NPB   NPB   NPB

NR   P   NR   NN

bushi   *yu*   *shalong*   *juxing*   *le*   *huitan*

Bush   *with*   *Sharon*   *talk*

Joint Parsing and Translation

PP

X_1:P X_2:NPB

$\longrightarrow$ X_1 X_2

PP

NPB   NPB   NPB

NR   P   NR   NN

bushi   *yu*   *shalong*   *juxing*   *le*   *huitan*

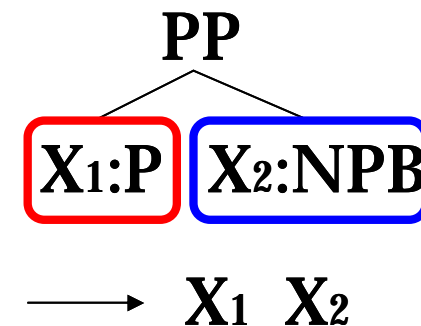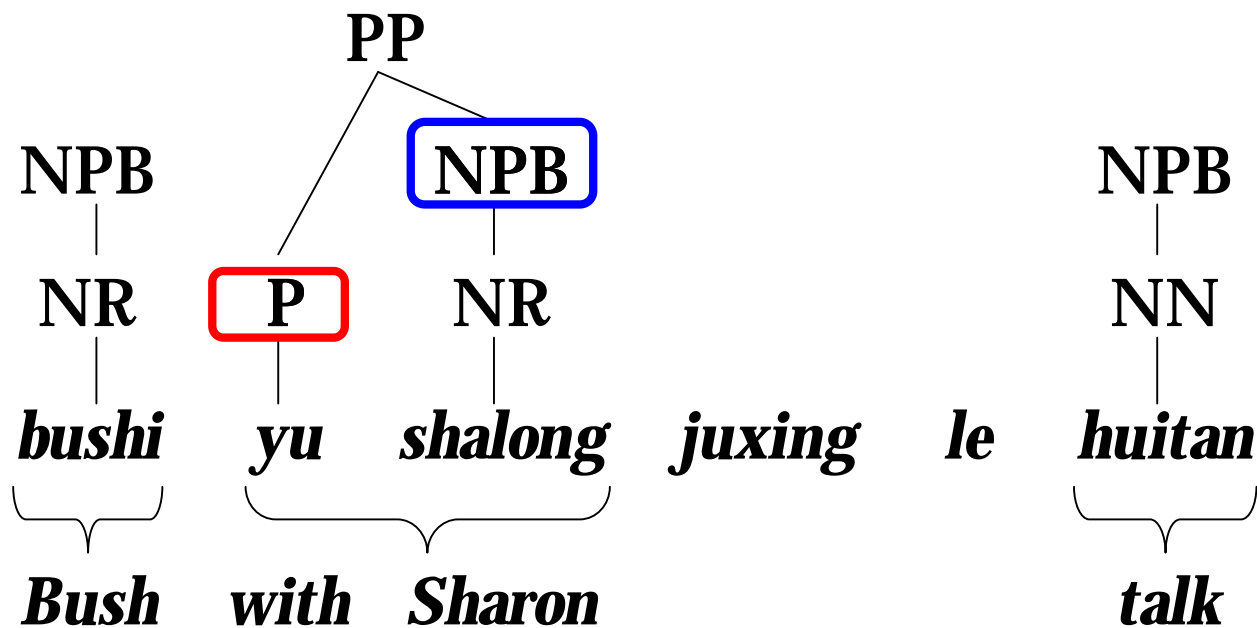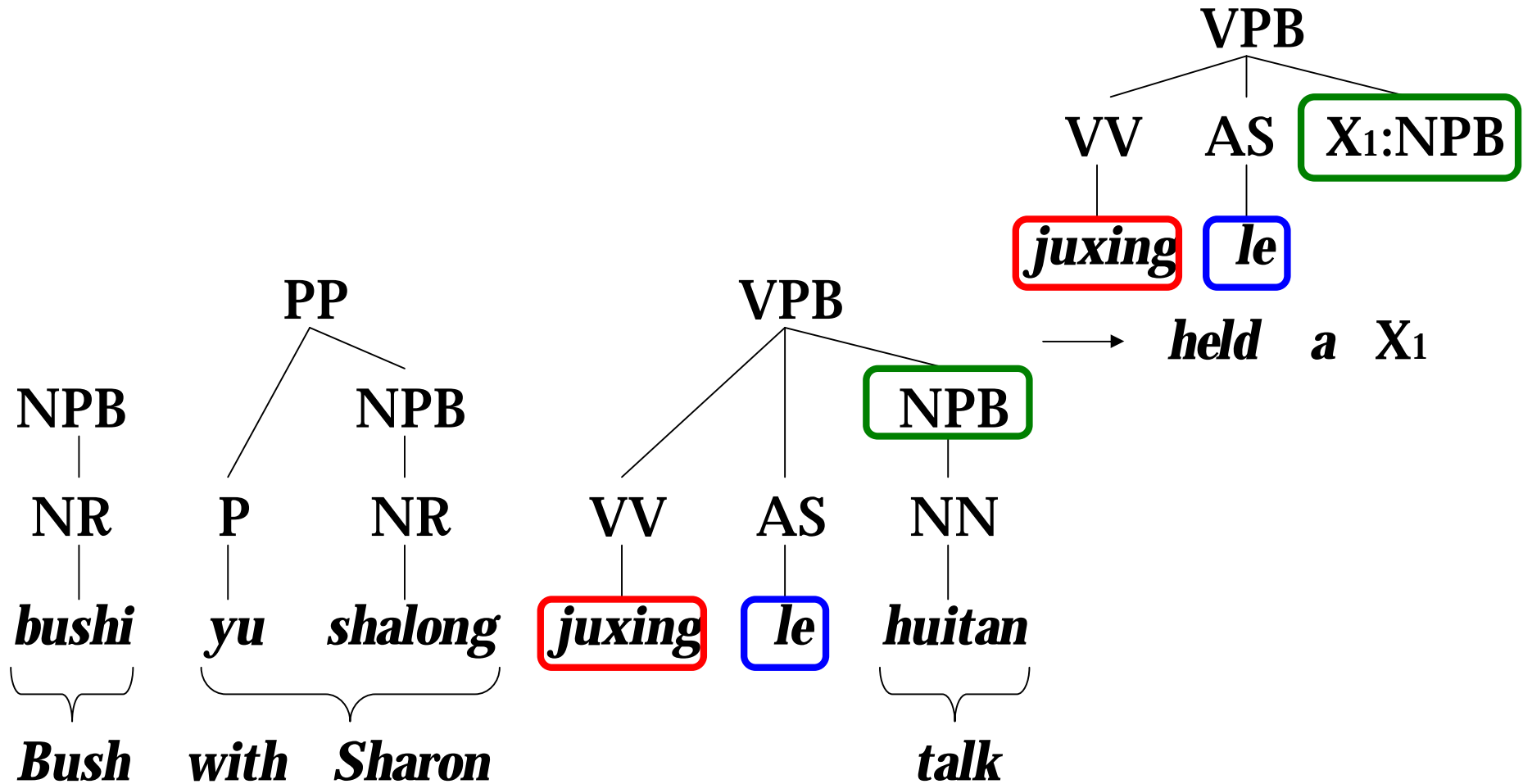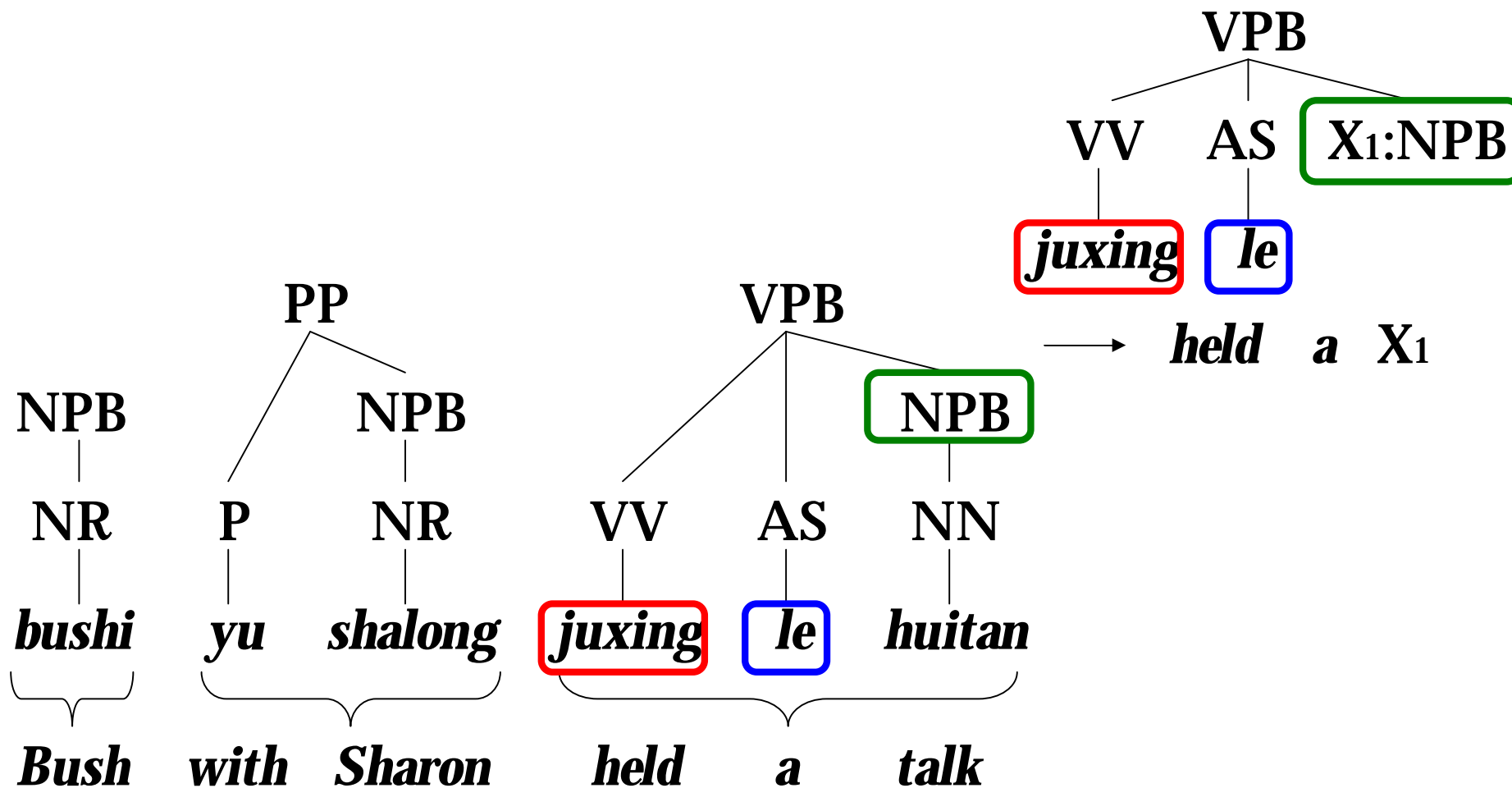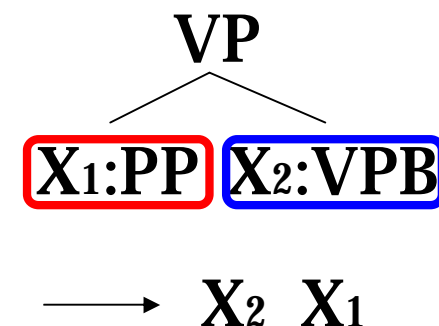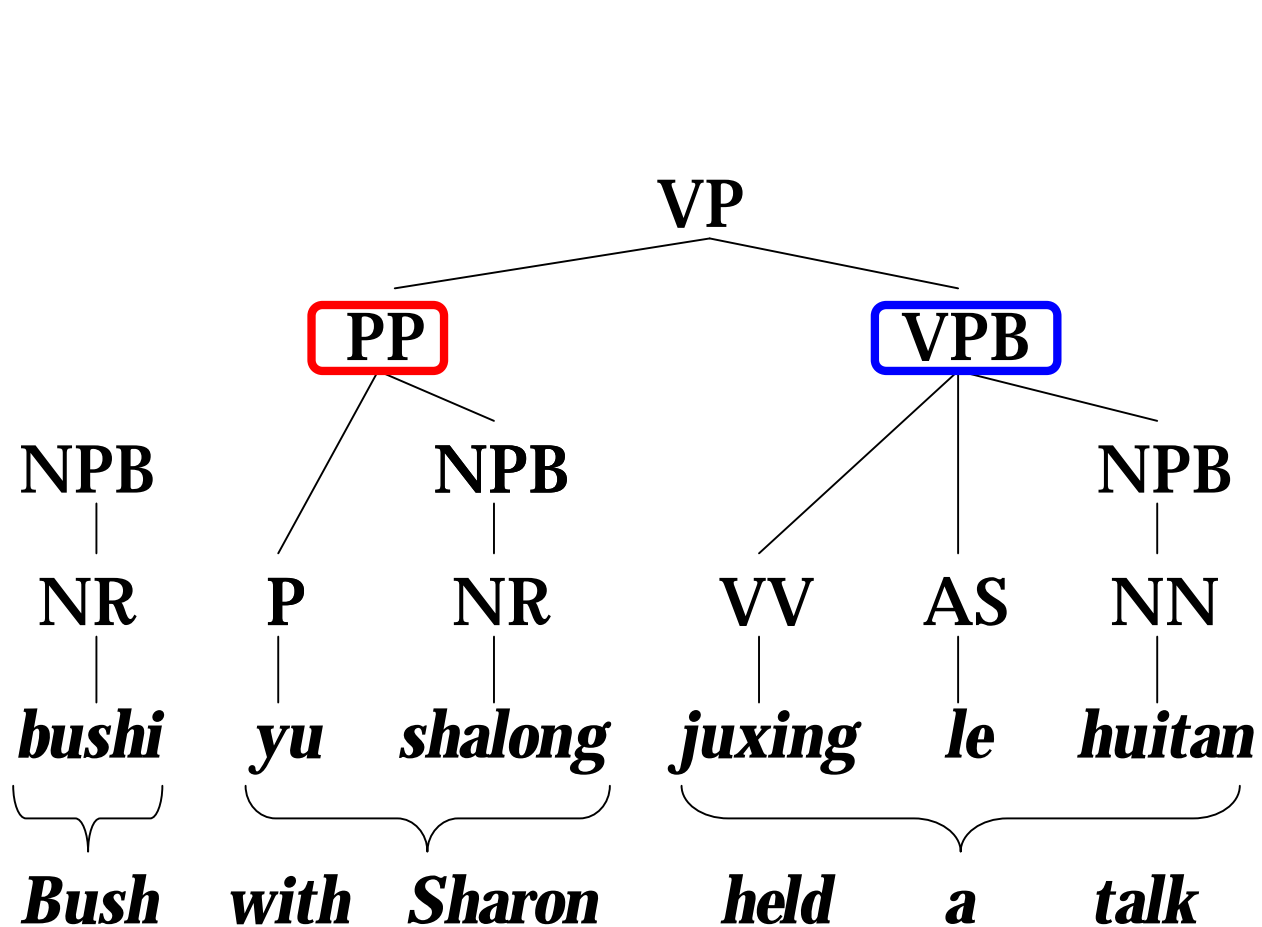*Bush*   *with*   *Sharon*   *talk*
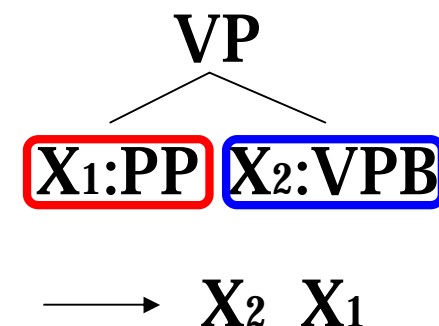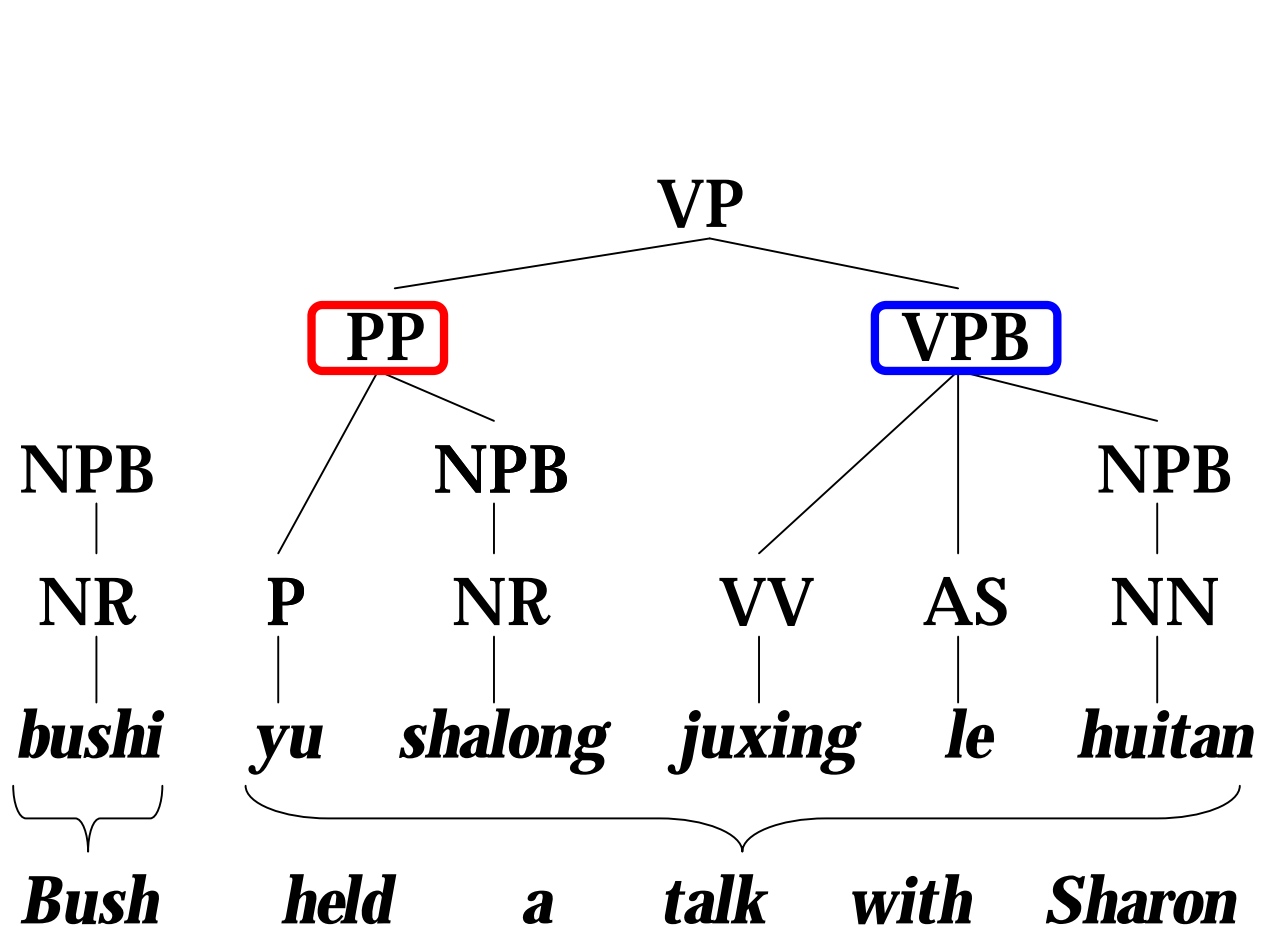
Joint Parsing and Translation

Joint Parsing and Translation

Joint Parsing and Translation

Joint Parsing and Translation

VP

X₁:PP X₂:VPB

→ X₂ X₁

VP
PP VPB
NPB NPB NPB
NR P NR VV AS NN
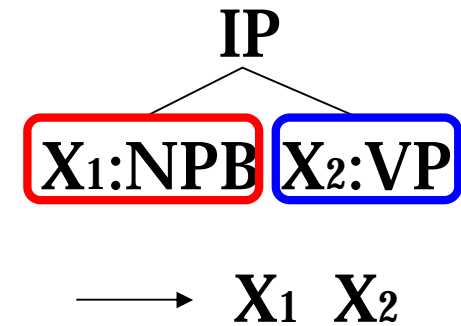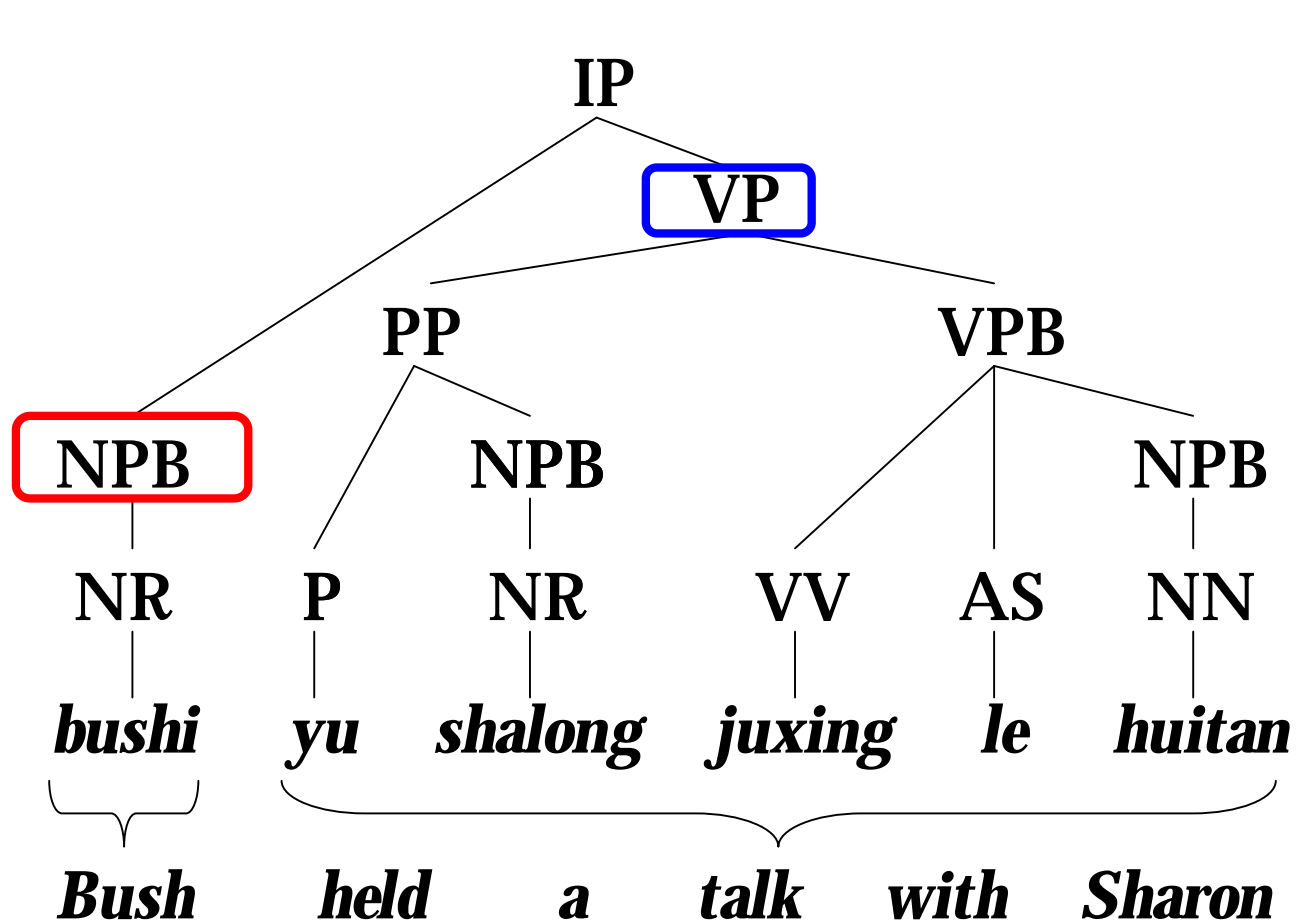bushi yu shalong juxing le huitan

Bush held a talk with Sharon

Joint Parsing and Translation

Joint Parsing and Translation

IP

VP

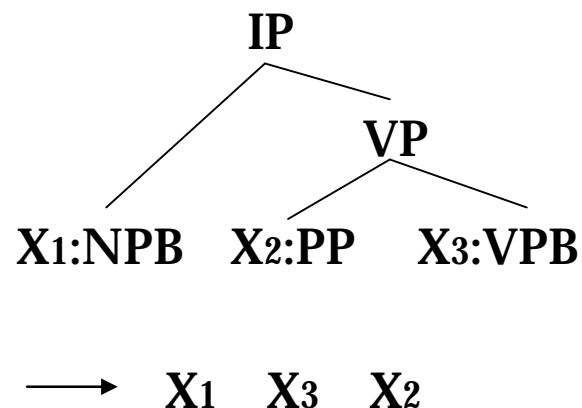PP   VPB

NPB   NPB   NPB

NR   P   NR   VV   AS   NN

bushi   *yu*   *shalong*   *juxing*   *le*   *huitan*

Bush   *held*   *a*   *talk*   *with*   *Sharon*

IP

X₁:NPB X₂:VP

⟶ X₁ X₂
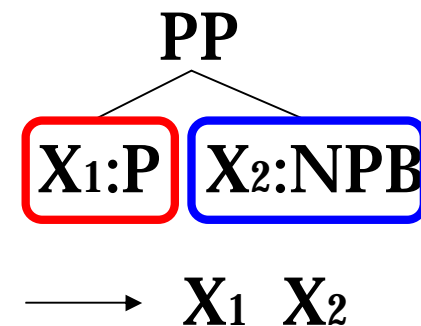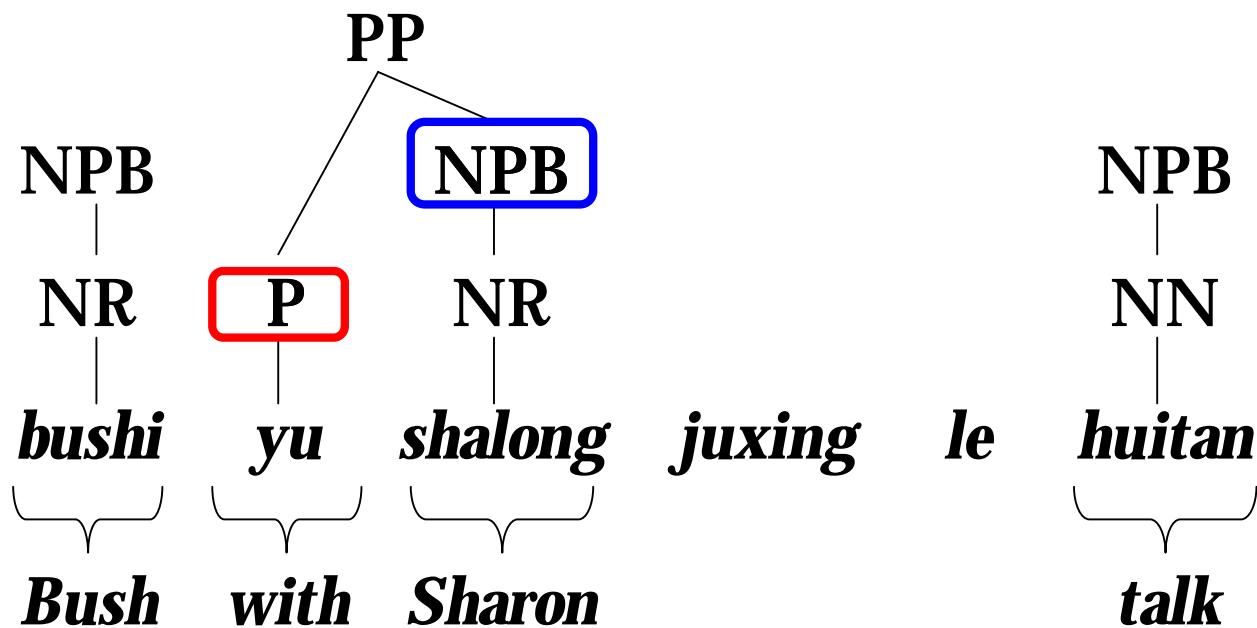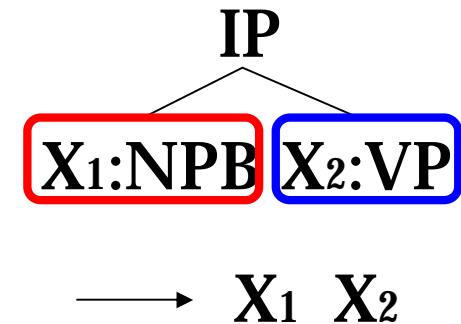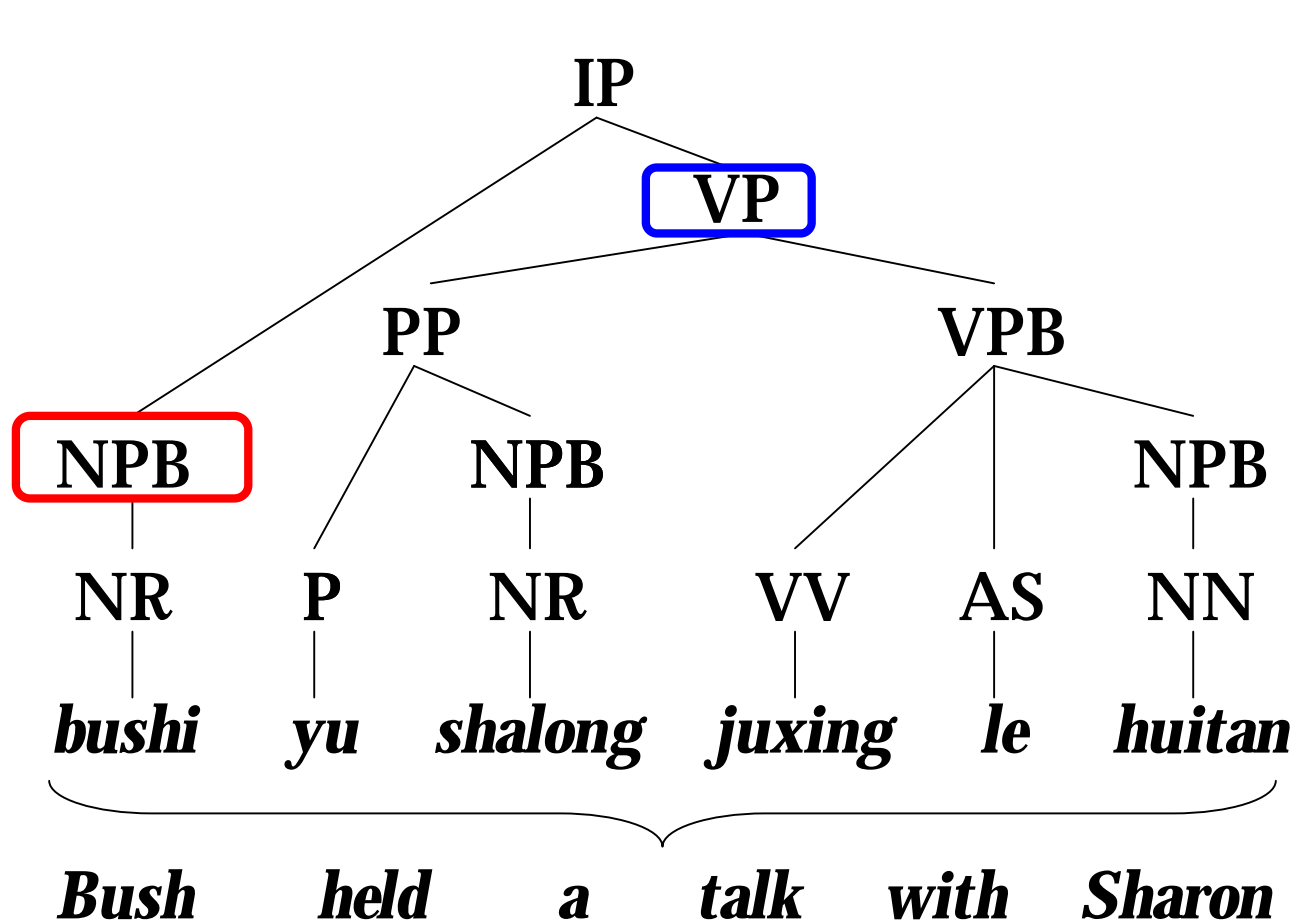
Synchronous Binarization

IP
— VP
— X1:NPB   X2:PP   X3:VPB

$\longrightarrow$ X1   X3   X2

$$IP \rightarrow \left\langle T_{[\![1]\!]}, T_{[\![1]\!]} \right\rangle$$

$$T \rightarrow \left\langle NPB_{[\![1]\!]}\ PP\text{-}VPB_{[\![2]\!]},\ NPB_{[\![1]\!]}\ PP\text{-}VPB_{[2]} \right\rangle$$

$$PP\text{-}VPB \rightarrow \left\langle PP_{[\![1]\!]}\ VPB_{[\![2]\!]},\ VPB_{[\![2]\!]}\ PP_{[\![1]\!]} \right\rangle$$

(Zhang et al., 2006; Huang et al., 2009)

Parsing and Translation Features

source sentence | source tree | feature function

$$score(f,p,e) = \sum_{k=1}^{K} l_k h_k(f,p,e)$$

(Och, 2003)

target sentence | feature weight
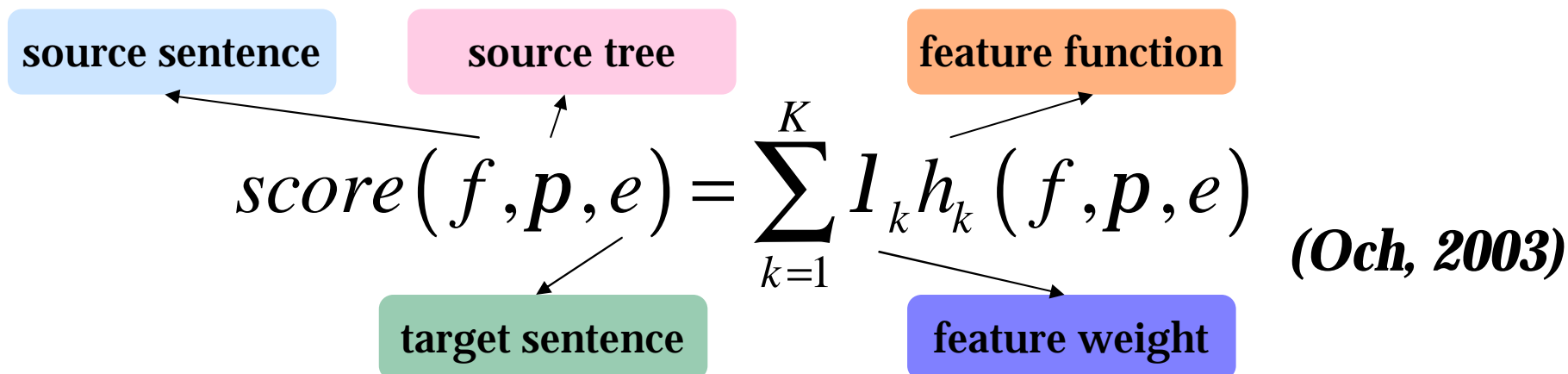
parsing features | **translation features**

PCFG | tree-to-string model

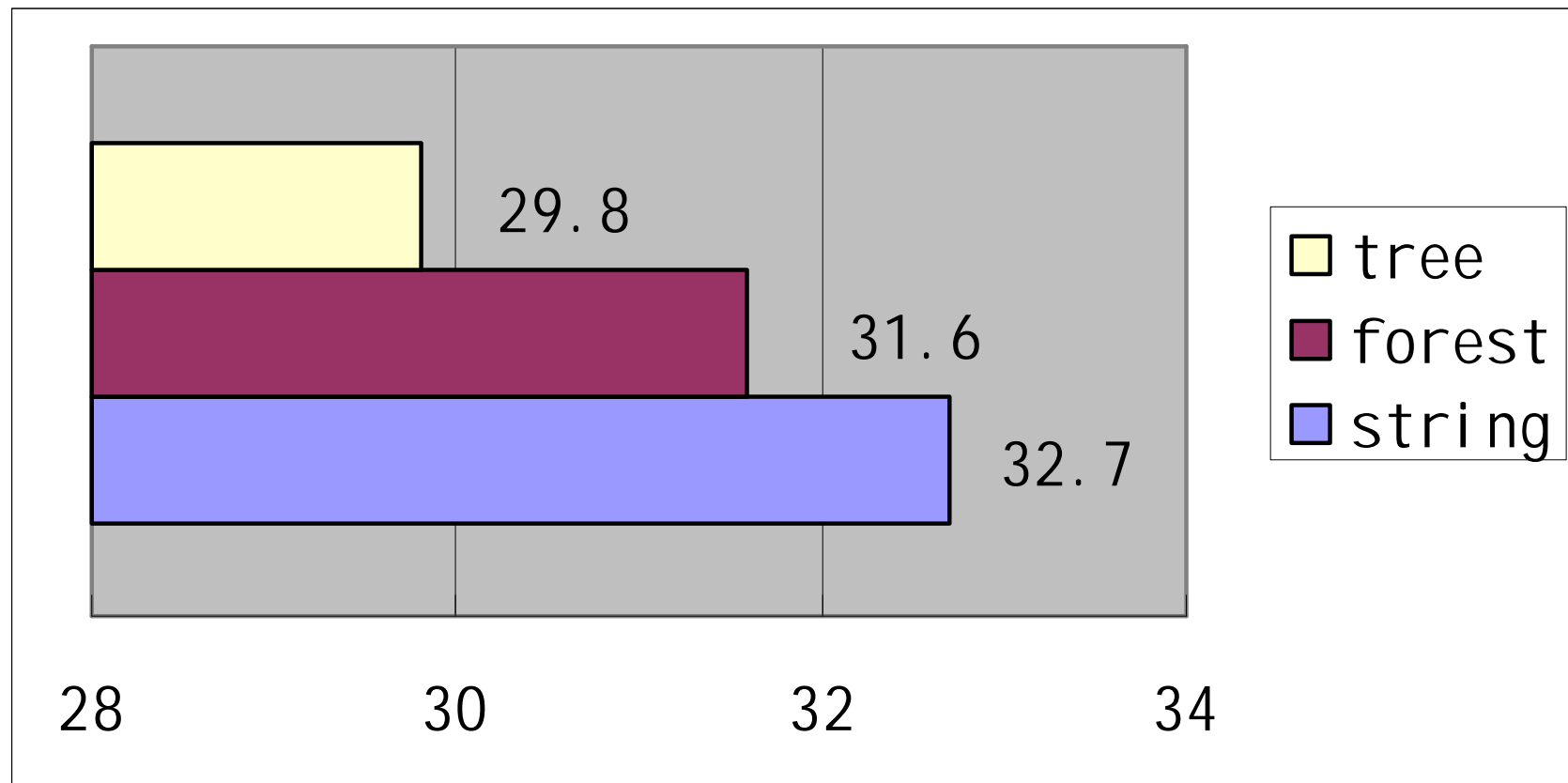Collins' model 1 | n-gram language model
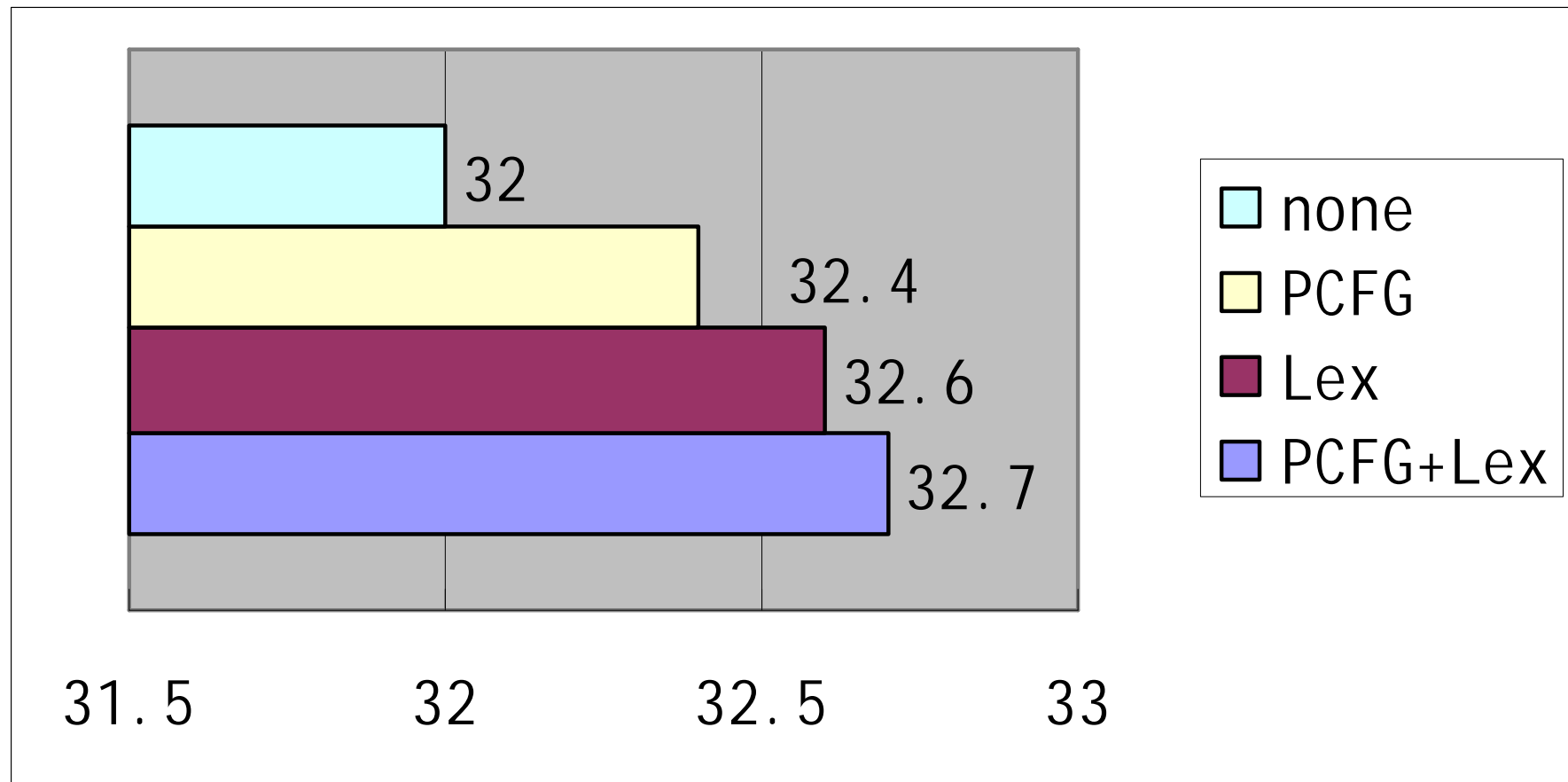
| word penalty

Translation Evaluation

- **Training sets**
  - Tree-to-string translation model: 251K sentences
  - N-gram language model: GIGAWORD Xinhua
  - Parsing models: Penn Chinese Treebank V5

- **Development set**
  - NIST 2002 Chinese-English

- **Test set**
  - NIST 2005 Chinese-English

- **Evaluation metric**
  - Case-insensitive BLEU4
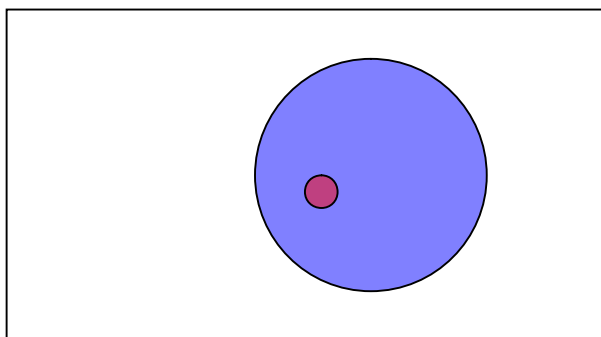
max-BLEU training

Tree-based, forest-based and string-based
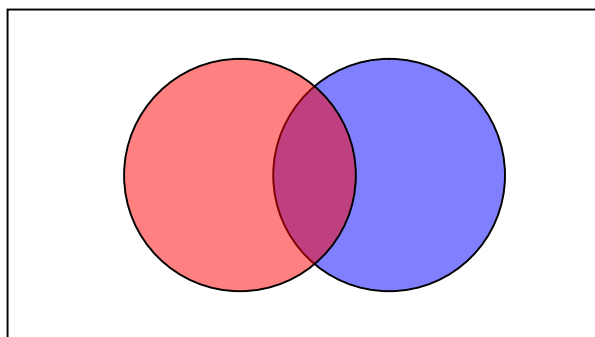
Effect of Parsing Models on Translation Quality
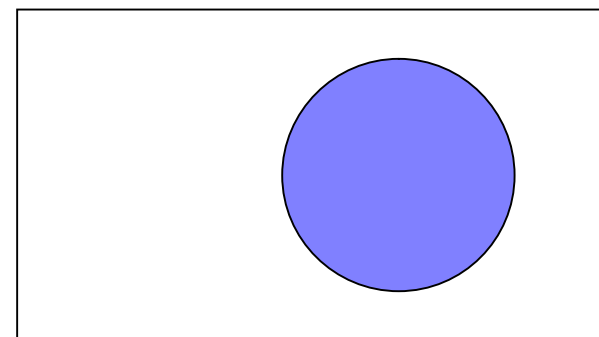
Why Different?

tree-based   forest-based   string-based
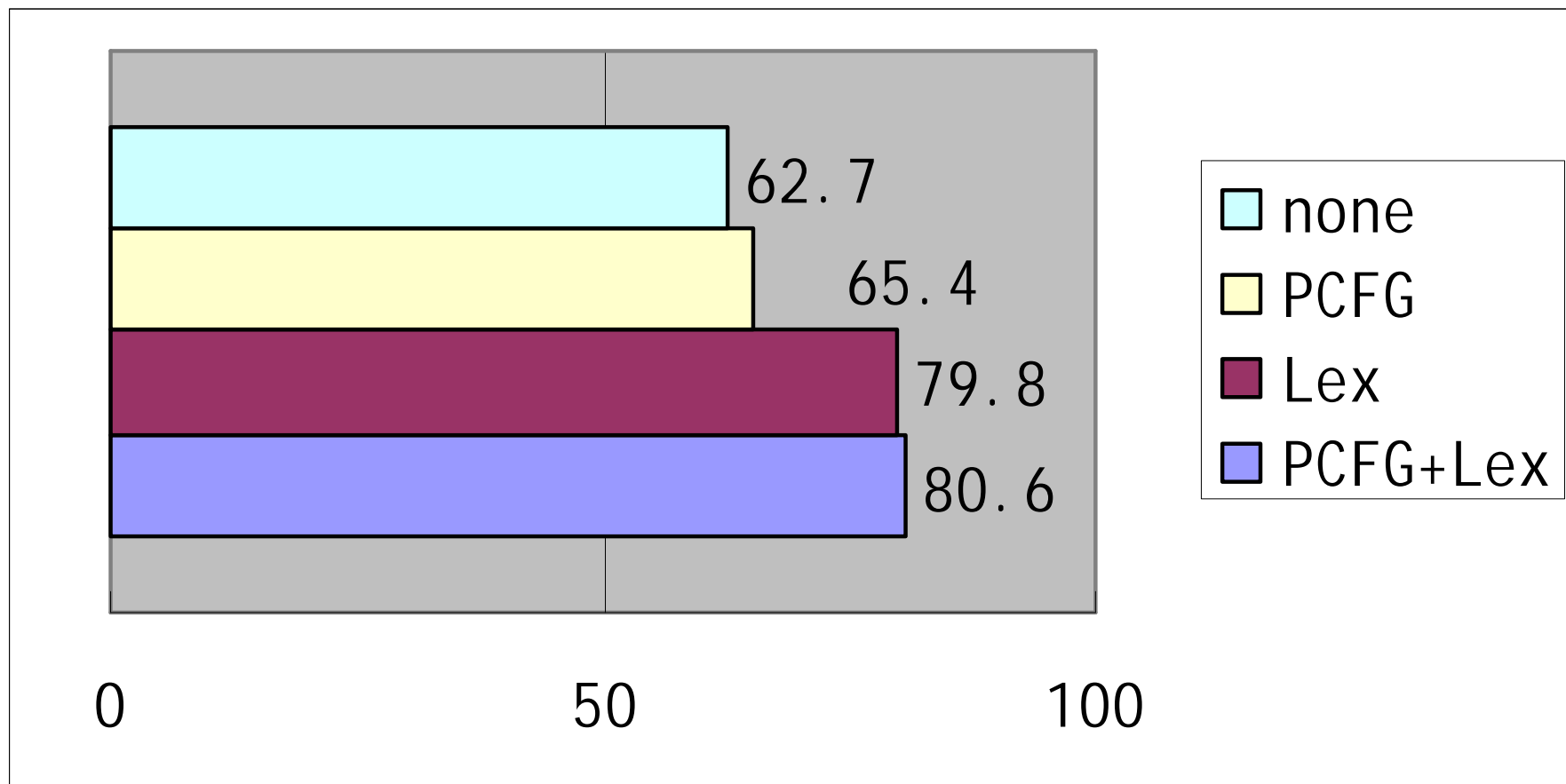
Parsing Evaluation

- **Training set**
  - Tree-to-string translation model: 251K sentences
  - N-gram language model: GIGAWORD Xinhua
  - Parsing models: Penn Chinese Treebank V5 (Articles 1-270 and 400-1151)
- **Development set**
  - Articles 301-325
- **Test set**
  - Articles 271-300
- **Evaluation metric**
  - F_1 (<= 40 words)

max-F1 training

Parsing Performance

Related Work

- **Translation as parsing**
  - String-to-tree *(Galley et al., 2006; Shen et al., 2008)*
  - String-to-string *(Chiang et al., 2007; Plunsom and Osborne, 2008)*
  - Tree-to-tree *(Chiang, 2010)*
- **Syntax-based language model**
  - Constituency *(Charniak et al., 2003)*
  - Dependency *(Shen et al., 2008)*
- **Joint parsing** *(Smith and Smith, 2004; Burkett and Klein, 2008)*
- **Bilingually constrained monolingual parsing** *(Huang et al., 2009)*
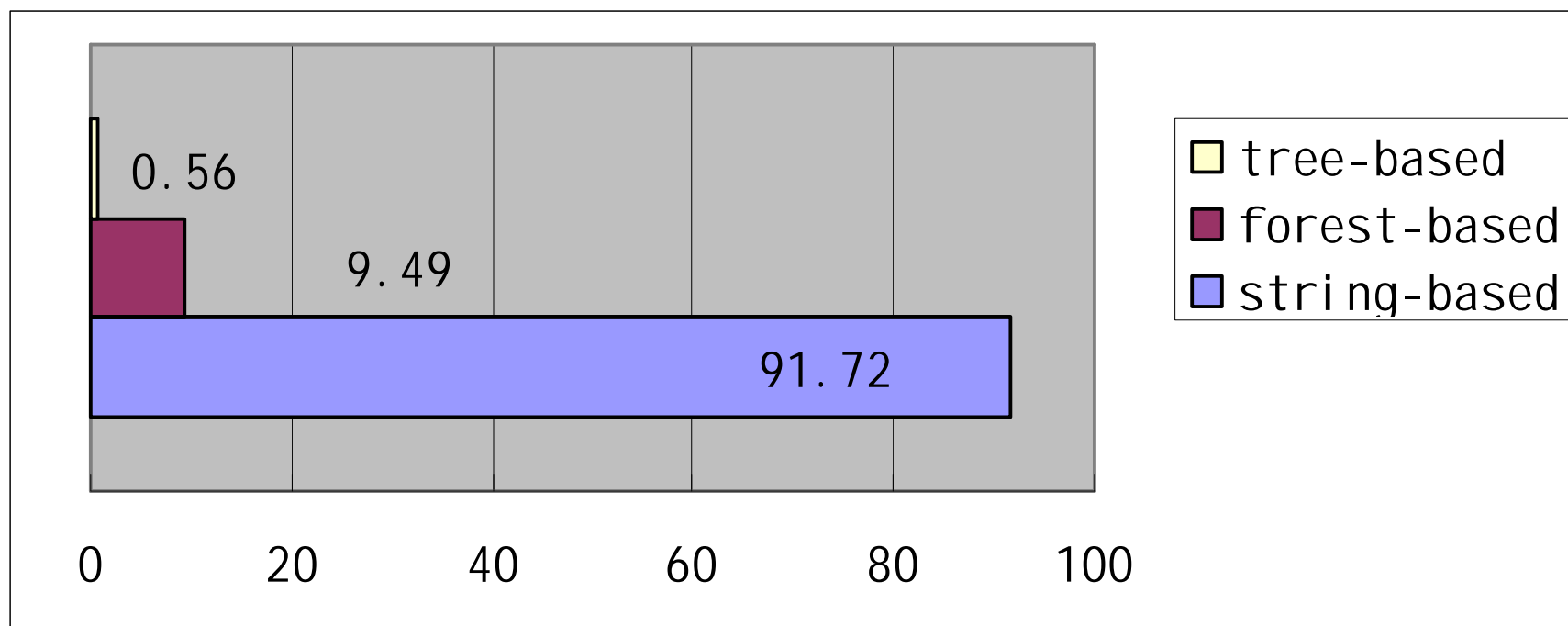- **Joint Tokenization and Translation** *(Xiao et al., 2010)*

Conclusion and Future Work

- **Joint parsing and translation by casting tree-to-string translation as a parsing problem:**
  - Translation: **+1.1 BLEU** over forest-based
  - Parsing: $\mathbf{F_1=80.6}$
- **Future directions**
  - Efficient decoding algorithm
  - In-depth investigation of forest matching
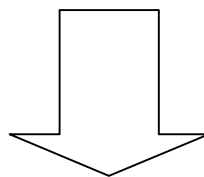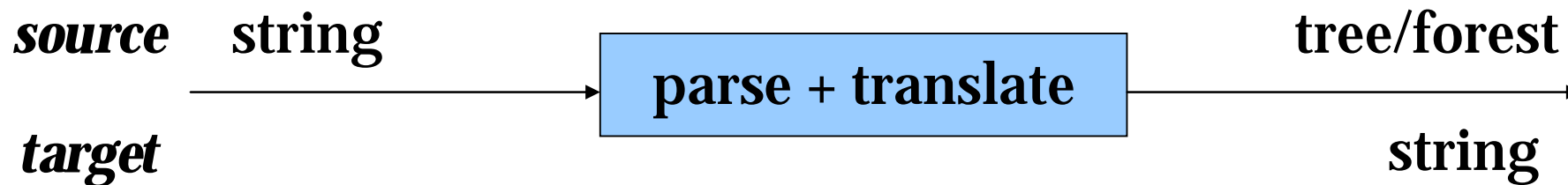 - Try on better parsing models
  - Scale up to larger data

Thank You!

Backup Slides for Questions

Average Decoding Time

``Separate'' Parsing and Translation

source   string   parse + translate   →   tree/forest

target   string

source   string   **parse**   forest   **translate**   →
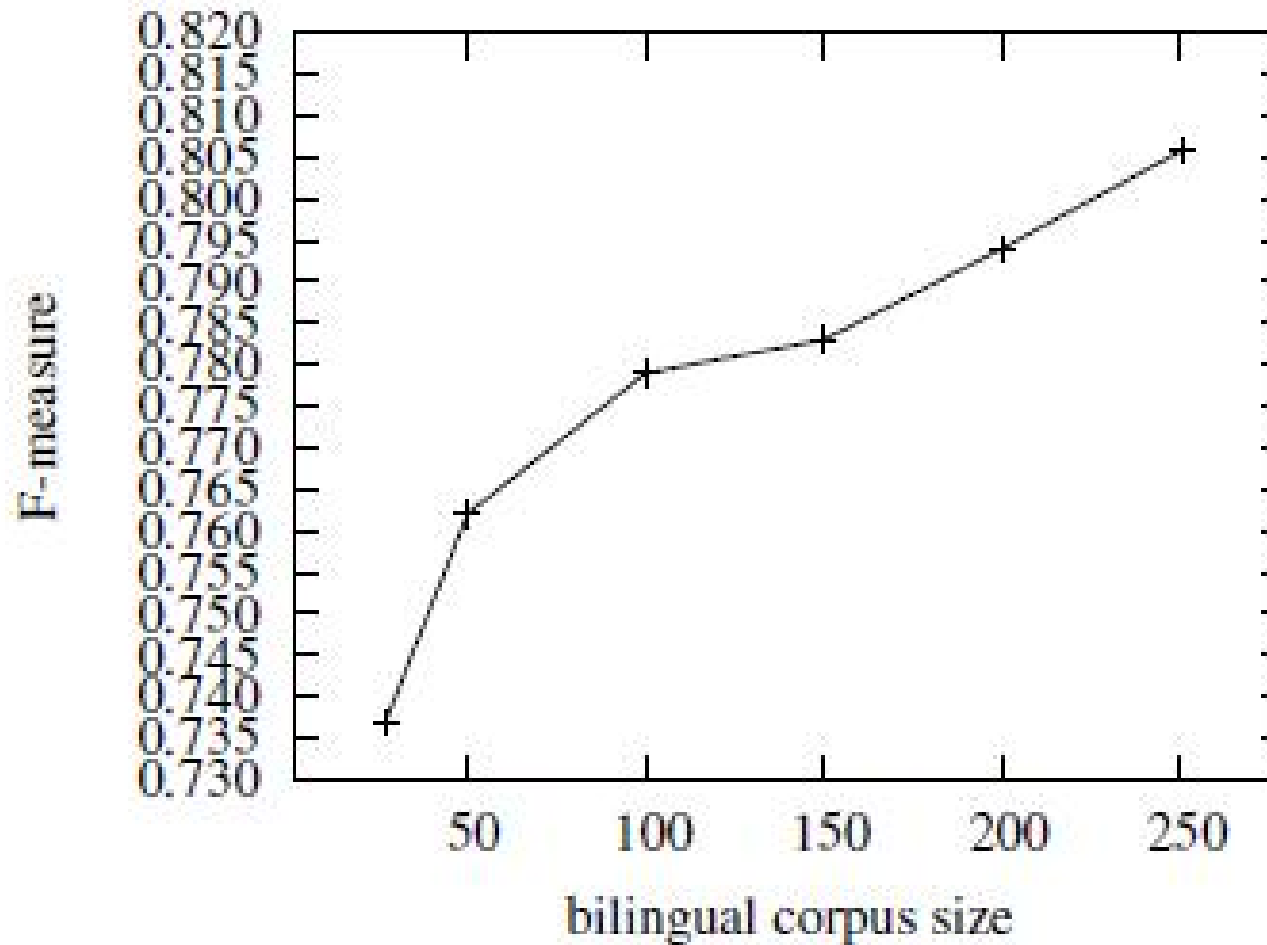target   string
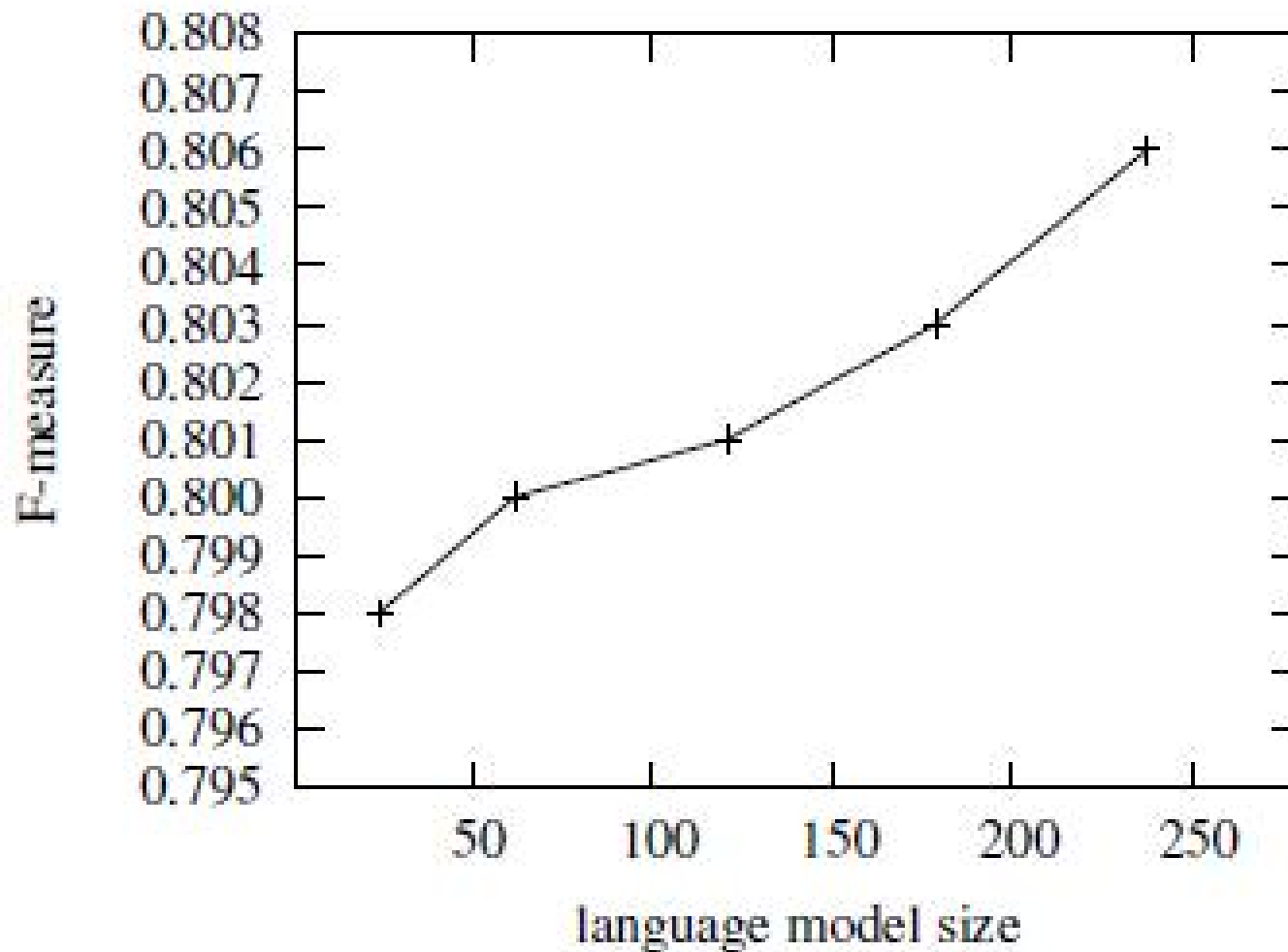
parse with TSG!

Effect of Bilingual Corpus Size on Parsing Performance

Effect of Language Model Size on Parsing Performance

Better Trees = Better Translations?

criterion in MERT	BLEU	F_1
BLEU	32.7	70.1
F_1	27.3	80.6

``Rule'' Tree in ``Parser'' Forest?

Forest size	Exact match (%)	Precision (%)
1	0.55	41.5
390	0.74	47.7
5.8M	0.92	54.1
66M	1.48	62.0
105M	2.22	65.9

* sub-sentence division was used to reduce decoding complexity