Joint Parsing and Translation

Yang Liu and Qun Liu

Institute of Computing Technology
Chinese Academy of Sciences
Separate Parsing and Translation

(source → string → parse → tree → translate → string)

(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

1. 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

(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

(Liu et al., 2006; Huang et al., 2006)
Separate Parsing and Translation

- Transfer the tree into a target string

\[
\begin{align*}
&\text{VP} \\
&\quad \text{X}_1: \text{PP} \quad \text{X}_2: \text{VPB} \\
&\quad \rightarrow \quad \text{X}_2 \quad \text{X}_1
\end{align*}
\]

(Liu et al., 2006; Huang et al., 2006)
Separate Parsing and Translation

- Transfer the tree into a target string

\[ \text{Transfer the tree into a target string} \]

\[ \text{Bush juxing le huitan yu shalong} \]

(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

(Liu et al., 2006; Huang et al., 2006)
Separate Parsing and Translation

- Transfer the tree into a target string

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

(Liu et al., 2006; Huang et al., 2006)
Separate Parsing and Translation

- Transfer the tree into a target string

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

Bush held a talk yu shalong

\[
\begin{array}{c}
\text{PP} \\
X_1: P \quad X_2: \text{NPB} \\
\rightarrow \quad X_1 \quad X_2
\end{array}
\]

(Liu et al., 2006; Huang et al., 2006)
Separate Parsing and Translation

- Transfer the tree into a target string

Bush held a talk with (Liu et al., 2006; Huang et al., 2006)
Separate Parsing and Translation

- Transfer the tree into a target string

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

Bush held a talk with 

(Liu et al., 2006; Huang et al., 2006)
Separate Parsing and Translation

- Transfer the tree into a target string

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

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

This work
Joint Parsing and Translation

bushi yu shalong juxing le huitan

→ Bush
Joint Parsing and Translation

bushi yu shalong juxing le huitan

NPB

NR P

bushi [yu] shalong juxing le huitan

Bush with

P

yu

→ with
Joint Parsing and Translation

bushi yu shalong juxing le huitan

Bush with Sharon

→ Sharon
Joint Parsing and Translation

NPB

NR
bushi
Bush

P
yu
with

NR
shalong
Sharon

juxing
le
huitan

NN
talk

NPB

NN
talk

huitan
Joint Parsing and Translation

*busi yu shalong juxing le huitan*

**NPB**
- **NR**
  - *bushi*
  - *Bush*
- **P**
  - *yu*
  - *with*
- **NR**
  - *shalong*
  - *Sharon*
- **NN**
  - *juxing*
  - *le*
  - *huitan*
  - *talk*

**PP**
- **X1:P**
- **X2:NPB**

* → *X1 X2*
Joint Parsing and Translation

- NPB
  - NR
    - bushi
    - Bush
  - P
    - yu
    - with
  - NPB
    - shalong
    - Sharon
  - NPB
    - juxing
    - le
  - NPB
    - huitan
    - talk

PP

X₁:P

X₂:NPB

→ X₁  X₂
Joint Parsing and Translation

bushi yu shalong juxing le huitan

Bush with Sharon held a X1 with talk

NPB
  NR  bushi
  P    yu
  NN   shalong

NPB
  NR  juxing
  P    le
  NN   huitan

VPB
  VV  held
  AS
  NPB

X1:NPB
Joint Parsing and Translation

bushi yu shalong juxing le huitan

VPB
VV AS X₁:NPB
juxing le

VPB
VV AS NNP

held a X₁

NPB
P

bushi
yu

Bush with Sharon

held a talk
Joint Parsing and Translation

bushi yu shalong juxing le huitan

Bush with Sharon held a talk

NPB
NR
bushi
Bush

NPB
P
yu
with

NPB
NR
shalong
Sharon

VPB
VV
juxing
held

X1:PP
X2:VPB

VP

→

X2 X1
Joint Parsing and Translation

bushi yu shalong juxing le huitan

Bush held a talk with Sharon
Joint Parsing and Translation

Bushi yu shalong juxing le huitan

Bush held a talk with Sharon

IP

VP

PP

NPB

NR

bushi

P

yu

NPB

NR

shalong

VV

juxing

AS

le

NN

huitan

→ X₁ X₂
Joint Parsing and Translation

Bush held a talk with Sharon
Synchronous Binarization

\[
\begin{array}{c}
\text{IP} \\
\downarrow \quad \text{VP} \\
\text{X}_1: \text{NPB} & \text{X}_2: \text{PP} & \text{X}_3: \text{VPB}
\end{array}
\]

\[ \rightarrow \quad \text{X}_1 \quad \text{X}_3 \quad \text{X}_2 \]

\[ \text{IP} \rightarrow \langle T_{[1]}, T_{[1]} \rangle \]

\[ T \rightarrow \langle \text{NPB}_{[1]}, \text{PP-VPB}_{[2]}, \text{NPB}_{[1]}, \text{PP-VPB}_{[2]} \rangle \]

\[ \text{PP-VPB} \rightarrow \langle \text{PP}_{[1]}, \text{VPB}_{[2]}, \text{VPB}_{[2]}, \text{PP}_{[1]} \rangle \]

(Zhang et al., 2006; Huang et al., 2009)
Parsing and Translation Features

\[ \text{score}(f, \pi, e) = \sum_{k=1}^{K} \lambda_k h_k(f, \pi, e) \]  

(Och, 2003)

### Parsing Features
- PCFG
- Collins’ model 1

### Translation Features
- Tree-to-string model
- \(n\)-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

![Bar chart showing the effect of different parsing models on translation quality.

- **none**
- **PCFG**
- **Lex**
- **PCFG+Lex**

The chart compares the translation quality scores with different parsing models. **PCFG+Lex** shows the highest score of 32.7, followed by **Lex** with 32.6, then **PCFG** with 32.4, and finally **none** with the lowest score of 32. The x-axis represents the scores ranging from 31.5 to 33.
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**
  - \( \text{max-F}_1 \) training
  - \( F_1 (\leq 40 \text{ words}) \)
Parsing Performance

![Bar chart showing parsing performance with different methods: none, PCFG, Lex, PCFG+Lex.](chart.png)
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: $F_1=80.6$

- 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

0.56 9.49 91.72

tree-based
forest-based
string-based

0 20 40 60 80 100
``Separate” Parsing and Translation

source string → parse + translate → tree/forest string

target

tree/forest string

source string → parse forest → translate string

target

parse with TSG!
Effect of Bilingual Corpus Size on Parsing Performance

![Graph showing the effect of bilingual corpus size on parsing performance. The x-axis represents the bilingual corpus size, and the y-axis represents the F-measure. As the corpus size increases, the F-measure also increases.]
Effect of Language Model Size on Parsing Performance
Better Trees = BetterTranslations?

<table>
<thead>
<tr>
<th>criterion in MERT</th>
<th>BLEU</th>
<th>$F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLEU</td>
<td>32.7</td>
<td>70.1</td>
</tr>
<tr>
<td>$F_1$</td>
<td>27.3</td>
<td>80.6</td>
</tr>
</tbody>
</table>
``Rule'' Tree in ``Parser'' Forest?

<table>
<thead>
<tr>
<th>Forest size</th>
<th>Exact match (%)</th>
<th>Precision (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.55</td>
<td>41.5</td>
</tr>
<tr>
<td>390</td>
<td>0.74</td>
<td>47.7</td>
</tr>
<tr>
<td>5.8M</td>
<td>0.92</td>
<td>54.1</td>
</tr>
<tr>
<td>66M</td>
<td>1.48</td>
<td>62.0</td>
</tr>
<tr>
<td>105M</td>
<td>2.22</td>
<td>65.9</td>
</tr>
</tbody>
</table>

* sub-sentence division was used to reduce decoding complexity