Non-Projective Parsing for Statistical Machine Translation

Xavier Carreras and Michael Collins

MIT CSAIL
Phrase-based Translation

There is no higher level of discrimination between

Es gibt keine hierarchie der diskriminierung

Segmentation + phrase selection + distortion
There is no higher level of discrimination between

es gibt keine hierarchie der diskriminierung

Segmentation + phrase selection + distortion
Phrase-based Translation with TAG operations

There is no higher level discrimination between es gibt keine hierarchie der diskriminierung

Segmentation + s-phrase selection + adjunctions
Phrase-based Translation with TAG operations

segmentation + s-phrase selection + adjunctions
Phrase-based Translation with TAG operations

Contributions (I)

A TAG-based syntactic translation model. Properties:

- Retains the full set of lexical entries of a phrase-based system
- Straightforward integration of a syntactic language model
Reordering via Non-Projective Operations

segmentation + s-phrase selection + non-projective adjuctions
Reordering via Non-Projective Operations

we must also
diese kritik
we müssen auch
diese kritik

these criticisms
seriously
nehmen
take

take
die these kritiken
take
die These kritiken

take these criticisms

segmentation + s-phrase selection + non-projective adjunctions
Reordering via Non-Projective Operations

we must also take these criticisms seriously

wir müssen auch diese kritik ernst nehmen

we must also take these criticisms seriously

segmentation + s-phrase selection + non-projective adjunctions
Reordering via Non-Projective Operations

Contributions (II)

We model reordering with flexible non-projective adjunctions.

- How to control reorderings?
  - A discriminative model inspired by work in dependency parsing (e.g. [McDonald et al. 05])
  - Hard constraints

- How to decode efficiently?
  - A novel beam-search algorithm
Outline

Translation as TAG-based Parsing

Constraints on Reorderings

Efficient Decoding

Experiments
A TAG formalism  [Carreras, Collins and Koo 2008]

- Basic units are *spines*

- Spines are combined using *adjunction* operations
A TAG formalism [Carreras, Collins and Koo 2008]

- Basic units are **spines**
- Spines are combined using **adjunction** operations
A TAG formalism [Carreras, Collins and Koo 2008]

- Basic units are spines
- Spines are combined using adjunction operations
S-phrases: Syntactic Phrase-entries for Translation

An s-phrase consists of:
- Foreign words
- English words
- A syntactic structure
- An alignment
Extraction of S-phrases

Training example = source sentence + English sentence
+ English parse tree

We use phrasal entries from a standard phrase-based approach
Extraction of S-phrases

Training example = source sentence + English sentence + English parse tree

We use phrasal entries from a standard phrase-based approach
Extraction of S-phrases

Training example = source sentence + English sentence + English parse tree

We use phrasal entries from a standard phrase-based approach
Extraction of S-phrases

▸ Training example = source sentence + English sentence + English parse tree

▸ We use phrasal entries from a standard phrase-based approach
Extraction of S-phrases

- Training example = source sentence + English sentence + English parse tree
- We use phrasal entries from a standard phrase-based approach
A derivation:

wir müssen auch diese kritik ernst nehmen
A derivation:
- Step 1: segment the input sentence, and choose an s-phrase for each segment.
Derivations

A derivation:
- Step 1: segment the input sentence, and choose an s-phrase for each segment
- Step 2: connect s-phrases with adjunctions
Model

- Model score for a derivation $d$:

$$score(d) = score_{LM}(d) + score_{P}(d) + score_{SYN}(d) + score_{R}(d)$$

where

- $score_{LM}$ is a trigram language model
- $score_{P}$ is a sum of standard phrase-based scores
- $score_{SYN}$ is a syntactic language model [Charniak et al. 03] [Shen et al. 08] (probabilities are associated with adjunctions)
- $score_{R}$ is a sum of discriminative adjunction scores
Model score for a derivation $d$:

$$
score(d) = score_{LM}(d) + score_P(d) + score_{SYN}(d) + score_R(d)
$$

where

- $score_{LM}$ is a trigram language model
- $score_P$ is a sum of standard phrase-based scores
- $score_{SYN}$ is a syntactic language model [Charniak et al. 03] [Shen et al. 08] (probabilities are associated with adjuctions)
- $score_R$ is a sum of discriminative adjunction scores
Model

- Model score for a derivation $d$:

$$score(d) = score_{LM}(d) + score_{P}(d) + score_{SYN}(d) + score_{R}(d)$$

where

- $score_{LM}$ is a trigram language model
- $score_{P}$ is a sum of standard phrase-based scores
- $score_{SYN}$ is a syntactic language model [Charniak et al. 03] [Shen et al. 08] (probabilities are associated with adjunctions)
- $score_{R}$ is a sum of discriminative adjunction scores
Model score for a derivation $d$:

$$score(d) = score_{LM}(d) + score_P(d) + score_{SYN}(d) + score_R(d)$$

where

- $score_{LM}$ is a trigram language model
- $score_P$ is a sum of standard phrase-based scores
- $score_{SYN}$ is a syntactic language model \([\text{Charniak et al. 03}] \) \([\text{Shen et al. 08}] \) (probabilities are associated with adjunctions)
- $score_R$ is a sum of discriminative adjunction scores
Model

- Model score for a derivation \( d \):

\[
\text{score}(d) = \text{score}_{LM}(d) + \text{score}_P(d) + \text{score}_{SYN}(d) + \text{score}_R(d)
\]

where

- \( \text{score}_{LM} \) is a trigram language model
- \( \text{score}_P \) is a sum of standard phrase-based scores
- \( \text{score}_{SYN} \) is a syntactic language model [Charniak et al. 03] [Shen et al. 08] (probabilities are associated with adjunctions)
- \( \text{score}_R \) is a sum of discriminative adjunction scores
Two challenges

All permutations of s-phrases are possible.

Two challenges:

1. Constraining reorderings
2. Search
Outline

Translation as TAG-based Parsing

Constraints on Reorderings

Efficient Decoding

Experiments
**score** \(_R\): A Discriminative Dependency Model

\(score_R(d)\) is a **discriminative dependency model** (related to work in dependency parsing (e.g. [McDonald et al. 05])).
\textit{score}_R: \text{A Discriminative Dependency Model}

\textit{score}_R(d) \text{ is a discriminative dependency model (related to work in dependency parsing (e.g. [McDonald et al. 05])}
\(\pi\)-constituent constraint

**Define** \(\pi\)-constituent: a head spine with all its descendants

**Constraint** any \(\pi\)-constituent must be aligned to a contiguous substring in the source sentence

**Satisfied:**

\[
\begin{align*}
S & \rightarrow NP \rightarrow VP \\
& \rightarrow ADVP \\
& \rightarrow \text{we must also}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow ADVP \\
& \rightarrow \text{these criticisms}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow ADVP \\
& \rightarrow \text{take}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow VP \\
& \rightarrow \text{wir müssen auch}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow ADVP \\
& \rightarrow \text{diese kritik}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow ADVP \\
& \rightarrow \text{ernst}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow ADVP \\
& \rightarrow \text{nehmen}
\end{align*}
\]

**Violated:**

\[
\begin{align*}
S & \rightarrow NP \rightarrow VP \\
& \rightarrow DT \rightarrow \text{there is}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow PP \\
& \rightarrow DT \rightarrow \text{no}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow PP \\
& \rightarrow \text{hierarchy of}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow PP \\
& \rightarrow \text{discrimination}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow PP \\
& \rightarrow DT \rightarrow \text{es gibt}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow PP \\
& \rightarrow DT \rightarrow \text{keine}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow PP \\
& \rightarrow DT \rightarrow \text{hierarchie der}
\end{align*}
\]

\[
\begin{align*}
S & \rightarrow NP \rightarrow PP \\
& \rightarrow DT \rightarrow \text{diskriminierung}
\end{align*}
\]
Outline

Translation as TAG-based Parsing

Constraints on Reorderings

Efficient Decoding

Experiments
Decoding as Parsing

Projective parsing: each constituent has an associated **span**

A generalization: each constituent has a **bit-string** recording which foreign words have been translated

Beam search strategy: ensures that the top $N$ analyses for each foreign word are explored at each stage
Outline

Translation as TAG-based Parsing

Constraints on Reorderings

Efficient Decoding

Experiments
Experiments

German to English using Europarl data (750K training sentences)

Development:

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax-based</td>
<td>25.2</td>
</tr>
<tr>
<td>Syntax (no disc. model)</td>
<td>23.7 (-1.5)</td>
</tr>
<tr>
<td>Syntax (no $\pi$-c constraint)</td>
<td>24.4 (-0.8)</td>
</tr>
</tbody>
</table>
Experiments

German to English using Europarl data (750K training sentences)

Development:

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax-based</td>
<td>25.2</td>
</tr>
<tr>
<td>Syntax (no disc. model)</td>
<td>23.7 (-1.5)</td>
</tr>
<tr>
<td>Syntax (no π-c constraint)</td>
<td>24.4 (-0.8)</td>
</tr>
</tbody>
</table>

Test:

<table>
<thead>
<tr>
<th>System</th>
<th>BLEU score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrase-based system (Pharaoh)</td>
<td>24.58</td>
</tr>
<tr>
<td>Syntax-based system</td>
<td>25.04 (+0.46)</td>
</tr>
</tbody>
</table>

significant \( (p = 0.021) \) under paired bootstrap resampling [Koehn 04] close to significant \( (p = 0.058) \) under the sign test [Collins et al. 05]
Ref: Now, however, we are seeing that president Putin is pursuing a policy of openness towards the west.

Now, however, we see that mr president Putin is pursuing a policy of openness towards the west.

We are, however, now that president Putin a policy of openness to the west out of blackmail.
Human Evaluations

Ref: Now, however, we are seeing that president Putin is pursuing a policy of openness towards the west.

Syn: Now, however, we see that mr president Putin is pursuing a policy of openness towards the west.

PB: We are, however, now that president Putin a policy of openness to the west out of blackmail.

<table>
<thead>
<tr>
<th></th>
<th>Syntax</th>
<th>PB</th>
<th>=</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntax</td>
<td>51</td>
<td>3</td>
<td>7</td>
<td>61</td>
</tr>
<tr>
<td>PB</td>
<td>1</td>
<td>25</td>
<td>11</td>
<td>37</td>
</tr>
<tr>
<td>=</td>
<td>21</td>
<td>14</td>
<td>67</td>
<td>102</td>
</tr>
<tr>
<td>Total</td>
<td>73</td>
<td>42</td>
<td>85</td>
<td>200</td>
</tr>
</tbody>
</table>

both results are significant with \( p < 0.05 \) under the sign test
Summary

A TAG-based syntactic translation model

Non-projective adjunctions for reordering:

- Arbitrary reorderings
- Discriminative dependency model
Summary

A TAG-based syntactic translation model

Non-projective adjunctions for reordering:

- Arbitrary reorderings
- Discriminative dependency model

Future work: Condition on syntactic structure of the source string