Tailoring Word Alignments to Syntactic Machine Translation

John DeNero and Dan Klein

Presentation and paper:
http://nlp.cs.berkeley.edu/pages/WordAligner.html
Tailoring Word Alignments to Syntactic Machine Translation

**Setting:** Syntactic MT with tree transducers

**Problem:** Alignment errors that contradict constituent structure impede the rule extraction process

**Proposal:** Condition word alignment on syntactic structure
Translating with Tree Transducers

**Source:** Les emplois sont axés sur la carrière.
Translating with Tree Transducers

Source: Les emplois sont axés sur la carrière.

Gloss: The jobs are centered on the career.
Translating with Tree Transducers

Source: Les emplois sont axés sur la carrière.

Gloss: The jobs are centered on the career.

Transducer rule: \((NP \ (DT \ The) \ (NNS \ jobs)) \Rightarrow \ Les \ emplois\)
Translating with Tree Transducers

**Source:**

<table>
<thead>
<tr>
<th>NP</th>
<th>Les emplois</th>
</tr>
</thead>
</table>

**Gloss:**

<table>
<thead>
<tr>
<th>NP</th>
<th>sont axés sur la carrière</th>
</tr>
</thead>
</table>

**Transducer rule:**

\[
(NP (DT The) (NNS jobs)) \Rightarrow \text{Les emplois}
\]
Translating with Tree Transducers

Source: Les emplois sont sur la carrière.

Gloss: The jobs are centered on the career.

Transducer rule: \((\text{ADJP} (\text{NN} \text{ career}) (\text{VBN} \text{ oriented})) \Rightarrow \text{axés sure la carrière}\)
Translating with Tree Transducers

Source: Les emplois sont axés sur la carrière.

Gloss: The jobs are centered on the career.

Transducer rule: 

$$(S \ NP_1 (VP (VBP \ are) \ ADJP_2) (. .)) \Rightarrow NP_1 \ sont \ ADJP_2.$$
Extracting Transducer Rules

Extraction Procedure
(Galley et al., ‘04 & ‘06)

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Extracting Transducer Rules

Extraction Procedure
(Galley et al., ‘04 & ‘06)

1. Choose a constituent

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Extracting Transducer Rules

Extraction Procedure
(Galley et al., ‘04 & ‘06)

1. Choose a constituent
2. Choose a region around constituent alignments
Extracting Transducer Rules

**Extraction Procedure**
*(Galley et al., ‘04 & ‘06)*

1. Choose a constituent

2. Choose a region around constituent alignments

3. Verify that alignment is consistent with region

Les emplois sont axés sur la carrière.
Extracting Transducer Rules

Extraction Procedure
(Galley et al., ‘04 & ‘06)

1. Choose a constituent
2. Choose a region around constituent alignments
3. Verify that alignment is consistent with region
4. Extract phrase:
   \[(\text{NN career}) \Rightarrow \text{carrière}\]
Extracting Transducer Rules

Extraction Procedure (Galley et al., ‘04 & ‘06)

1. Choose a constituent

2. Choose a region around constituent alignments

3. Verify that alignment is consistent with region

4. Extract phrase:

\[(\text{NN} \text{ career}) \Rightarrow \text{carrière}\]
Extracting Transducer Rules

Extraction Procedure (Galley et al., ‘04 & ‘06)

1. Choose a constituent
2. Choose a region around constituent alignments
3. Verify that alignment is consistent with region
4. Extract phrase:

\[(VBN \text{ oriented}) \Rightarrow \text{axés sur}\]
Extracting Transducer Rules

Extraction Procedure
(Galley et al., ‘04 & ‘06)

1. Choose a constituent
2. Choose a region around constituent alignments
3. Verify that alignment is consistent with region
4. Extract phrase:

$(ADJP \; NN_1 \; VBN_2) \Rightarrow VBN_2 \; la \; NN_1$
Extracting Transducer Rules

Rule Combination
(Galley et al., ‘06)

(ADJP NN₁ VBN₂) => VBN₂ la NN₁
(VBN oriented) => axés sur
(NN career) => carrière
Extracting Transducer Rules

Rule Combination
(Galley et al., ‘06)

(ADJP NN₁ VBN₂) => VBN₂ la NN₁
(VBN oriented) => axés sur
(NN career) => carrière

Les emplois sont axés sur la carrière.

http://nlp.cs.berkeley.edu/pages/WordAligner.html
Extracting Transducer Rules

Rule Combination
(Galley et al., ‘06)

(ADJP NN₁ VBN₂) => VBN₂ la NN₁

(VBN oriented) => axés sur

(NN career) => carrière

Les emplois sont axés sur la NN₁

(ADJP NN₁ (VBN oriented)) => axés sur la NN₁

(ADJP (NN career) (VBN oriented)) => axés sur la carrière

...
Alignment Errors that Prevent Rule Extraction

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors that Prevent Rule Extraction

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors that Prevent Rule Extraction

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors that Prevent Rule Extraction

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors that Prevent Rule Extraction

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors that Prevent Rule Extraction

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors that Prevent Rule Extraction

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors that Prevent Rule Extraction

Net effect on extraction:
• 2 instead of 7 recursive rules can be extracted
Alignment Errors that Prevent Rule Extraction

Net effect on extraction:

• 2 instead of 7 recursive rules can be extracted

• Smallest recursive rule that can be extracted:
Alignment Errors that Prevent Rule Extraction

Net effect on extraction:

• 2 instead of 7 recursive rules can be extracted

• Smallest recursive rule that can be extracted:

\[
(S \ (NP \ (DT \ The) \ NNS_2) \\
\ (VP \ VBP_3) \\
\ (ADJP \ NN_4 \ VBN_5) \\
\) \\
\Rightarrow \ Les \ NNS_2 \ VBP_3 \ VBN_5 \ NN_4 \ .6
\]
Alignment Errors under the HMM Alignment Model

\[ p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-}) \]
Alignment Errors under the HMM Alignment Model

\[ p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-}) \]

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

\[ p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-1}) \]

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

\[ p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-}) \]

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

\[ p(f, a|e) = \prod_j p(f_j|e_{a_j}) \cdot p(a_j|a_{j-}) \]

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

\[ p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-1}) \]

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

\[ p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-}) \]

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

\[ p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-1}) \]

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

\[ p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-}) \]

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

\[ p(f, a|e) = \prod_j p(f_j|e_{a_j}) \cdot p(a_j|a_{j-1}) \]

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

\[
p(f,a|e) = \prod_j p(f_j|e_{a_j}) \cdot p(a_j|a_{j-1})
\]

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

\[ p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-}) \]

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

$$p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-})$$

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Alignment Errors under the HMM Alignment Model

Les emplois sont axés sur la carrière.

The jobs are career oriented.

http://nlp.cs.berkeley.edu/pages/WordAligner.html
The jobs are career oriented.

Les emplois sont axés sur la carrière.
The jobs are career oriented.

Les emplois sont axés sur la carrière.
Syntactic HMM Alignment Model

\[ p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-1}) \]
Syntactic HMM Alignment Model

\[
p(f, a | e) = \prod_j p(f_j | e_{a_j}) \cdot p(a_j | a_{j-1}, t)
\]

The jobs are career oriented.

Les emplois sont axés sur la carrière.
The jobs are career oriented.

Les emplois sont axés sur la carrière.
The jobs are career oriented.

Les emplois sont axés sur la carrière.
Syntactic HMM Alignment Model

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Syntactic HMM Alignment Model

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Syntactic HMM Alignment Model

The jobs are career-oriented.  

Les emplois sont axés sur la carrière.  

http://nlp.cs.berkeley.edu/pages/WordAligner.html
The jobs are career oriented.

Les emplois sont axés sur la carrière.
The jobs are career oriented.

Les emplois sont axés sur la carrière.
Syntactic HMM Alignment Model

The jobs are career oriented.

Les emplois sont axés sur la carrière.
The jobs are career oriented.

Les emplois sont axés sur la carrière.
The jobs are career oriented.

Les emplois sont axés sur la carrière.
Les emplois sont axés sur la carrière.

The jobs are career oriented.
Les emplois sont axés sur la carrière.

Syntactic HMM Alignment Model

The jobs are career oriented.
Evaluation:
Alignment Error Rate (AER)
Evaluation:
Alignment Error Rate (AER)

Test Conditions

• Chinese-English from MT-Eval 02 test set
• 100k training sentences from FBIS
• Initialized with agreement training for Model 1 (Liang et al., 06)
Evaluation: Alignment Error Rate (AER)

Test Conditions

- Chinese-English from MT-Eval 02 test set
- 100k training sentences from FBIS
- Initialized with agreement training for Model 1 (Liang et al., 06)

Results

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>AER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic HMM</td>
<td>81.6</td>
<td>78.8</td>
<td>19.8</td>
</tr>
<tr>
<td>Syntactic HMM</td>
<td>82.2</td>
<td>76.8</td>
<td>20.5</td>
</tr>
<tr>
<td>GIZA++</td>
<td>61.9</td>
<td>82.6</td>
<td>29.7</td>
</tr>
</tbody>
</table>
Evaluation:
Unproductive Constituent Rates

Classic HMM  Syntactic HMM

<table>
<thead>
<tr>
<th>NP</th>
<th>VP</th>
<th>PP</th>
<th>S</th>
<th>SBAR</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>30%</td>
<td>30%</td>
<td>30%</td>
<td>45%</td>
<td>30%</td>
<td>30%</td>
</tr>
</tbody>
</table>

http://nlp.cs.berkeley.edu/pages/WordAligner.html
Evaluation:
Unproductive Constituent Rates

The Syntactic HMM Reduces the Frequency of Unproductive *Interior* Nodes by 13%
Decoding Heuristic: Competitive Thresholding

Only the maximum posterior in each row or column and its neighbors can be included in the alignment.

### Alignment Example

<table>
<thead>
<tr>
<th>The</th>
<th>jobs</th>
<th>are</th>
<th>career</th>
<th>oriented</th>
</tr>
</thead>
</table>

| Les | emplois | sont | axés | sur | la | carrière | . |
Decoding Heuristic: Competitive Thresholding

Only the maximum posterior in each row or column and its neighbors can be included in the alignment.

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Decoding Heuristic: Competitive Thresholding

Only the maximum posterior in each row or column and its neighbors can be included in the alignment.

<table>
<thead>
<tr>
<th>The</th>
<th>Les</th>
<th>emplois</th>
<th>sont</th>
<th>axés</th>
<th>sur</th>
<th>la</th>
<th>carrière</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

http://nlp.cs.berkeley.edu/pages/WordAligner.html
Decoding Heuristic: Competitive Thresholding

Only the maximum posterior in each row or column and its neighbors can be included in the alignment.
Decoding Heuristic: Competitive Thresholding

Only the maximum posterior in each row or column and its neighbors can be included in the alignment.

The jobs are career oriented.

Les emplois sont axés sur la carrière.
Evaluation: Unproductive Constituent Rates

- Classic HMM
- Syntactic HMM
- Syntactic HMM with Competitive Thresholding

http://nlp.cs.berkeley.edu/pages/WordAligner.html
Evaluation: Unproductive Constituent Rates

Classical HMM
Syntactic HMM
Syntactic HMM with Competitive Thresholding

Unproductive Constituent Rates

Classic HMM: 48.5% overall reduction

NP  VP  PP  S   SBAR  All
Evaluation:
Quantity of Rules Extracted

Rules extracted per sentence
## Evaluation:
### Quantity of Rules Extracted

<table>
<thead>
<tr>
<th>Rules extracted per sentence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syntactic HMM + CT</td>
</tr>
<tr>
<td>Syntactic HMM</td>
</tr>
<tr>
<td>Classic HMM</td>
</tr>
</tbody>
</table>
Evaluation:
Quantity of Rules Extracted

Rules extracted per sentence

- Syntactic HMM + CT
- Syntactic HMM
- Classic HMM

65
Evaluation: Quantity of Rules Extracted

Rules extracted per sentence

- Syntactic HMM + CT: 73
- Syntactic HMM: 73
- Classic HMM: 65
Evaluation: Quantity of Rules Extracted

Rules extracted per sentence

- Syntactic HMM + CT: 90
- Syntactic HMM: 73
- Classic HMM: 65
Evaluation:
Comparing Gold and Induced Rules

Evaluation Metric Idea:
Compare rules from gold alignments and induced alignments on both precision and recall.

Analog to the consistent phrase error rate (CPER) metric of Ayan & Dorr (06)
Evaluation Metric Idea:
Compare rules from gold alignments and induced alignments on both precision and recall.

Analog to the consistent phrase error rate (CPER) metric of Ayan & Dorr (06)

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic HMM</td>
<td>40.4</td>
<td>33.9</td>
<td>36.8</td>
</tr>
<tr>
<td>Syntactic HMM</td>
<td>41.3</td>
<td>36.7</td>
<td>38.9</td>
</tr>
<tr>
<td>Syntactic HMM + CT</td>
<td>39.6</td>
<td>41.1</td>
<td>40.3</td>
</tr>
</tbody>
</table>
Evaluation: Comparing Gold and Induced Rules

Evaluation Metric Idea:
Compare rules from gold alignments and induced alignments on both precision and recall.

Analog to the consistent phrase error rate (CPER) metric of Ayan & Dorr (06)

<table>
<thead>
<tr>
<th>Model</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic HMM</td>
<td>40.4</td>
<td>33.9</td>
<td>36.8</td>
</tr>
<tr>
<td>Syntactic HMM</td>
<td>41.3</td>
<td>36.7</td>
<td>38.9</td>
</tr>
<tr>
<td>Syntactic HMM + CT</td>
<td>39.6</td>
<td>41.1</td>
<td>40.3</td>
</tr>
</tbody>
</table>

F1 Increase: 9.5% in Chinese; 18.7% in French
Summary

• Tree transducer extraction systems should be wary of constituent-violating alignment errors
Summary

• Tree transducer extraction systems should be wary of constituent-violating alignment errors

• Conditioning the HMM alignment model on a parse tree corrects some such errors
Summary

• Tree transducer extraction systems should be wary of constituent-violating alignment errors

• Conditioning the HMM alignment model on a parse tree corrects some such errors

• Decoding heuristics correct even more
Summary

• Tree transducer extraction systems should be wary of constituent-violating alignment errors

• Conditioning the HMM alignment model on a parse tree corrects some such errors

• Decoding heuristics correct even more

• The resulting rules are more faithful to the rule set that should be extracted
Summary

• Tree transducer extraction systems should be wary of constituent-violating alignment errors

• Conditioning the HMM alignment model on a parse tree corrects some such errors

• Decoding heuristics correct even more

• The resulting rules are more faithful to the rule set that should be extracted

• Future work: end-to-end translation (BLEU)
Coming 07/07/07:
BerkeleyAligner Software Package

• Agreement training of IBM models, which reduces AER 32% relative to GIZA++ (Liang et al., 06)
• Syntactic distortion model (this paper)
• Posterior decoding heuristics (this paper)
• Evaluation code: searches for posterior thresholds, compares decoding methods, & tracks AER during training
• Easy integration with the Berkeley Parser
• Pure Java 1.5 will run on any platform

Check it out:
http://nlp.cs.berkeley.edu/pages/WordAligner.html
Thank You
denero@berkeley.edu