Two Languages are Better than One
(for Syntactic Parsing)

David Burkett and Dan Klein
{dburkett, klein}@cs.berkeley.edu
Motivating Example

Whereas in 1992 this figure was only 1500 US dollars.
Motivating Example

Incorrect

```
S
   S
      SBAR
         FRAG
            PP
                NP
                   IN
                   IN
                   CD
Whichas in 1992
```
Motivating Example

Incorrect

```
SBAR
  FRAG
    PP
      IN
      IN
      CD
      Whereas in 1992
```

Correct

```
S

SBAR
  ADVP
    IN
    Whereas
  PP
    IN
    IN
    CD
    in 1992
```
Motivating Example

Incorrect

Correct

```
SBAR
  FRAG
    PP
      IN
      IN
      CD
      Whereas
      in
      1992
      而
      在
      一九九二年
      AD
      P
      NT
      NP
      PP
      IP

PP
ADVP
IN
IN
CD
Whereas
in
1992
而
在
一九九二年
AD
P
NT
NP
PP
IP
```
Joint Parsing
Joint Parsing

Target Sentence

Whereas in 1992

Source Sentence
Joint Parsing

Whereas in 1992

Target Sentence

Whereas in 1992

Source Sentence

Bitext
Joint Parsing

Target Sentence

Whereas in 1992

Source Sentence

Bitext

Target Parse Trees

Source Parse Trees
Joint Parsing

Target Parse Trees

Target Sentence

Source Sentence

Bitext

Target Parser

Word Aligner

Source Parser

Source Parse Trees
Joint Parsing

Target Parse Trees

Target Sentence

Word Aligner

Word Alignment

Source Sentence

Source Parse Trees
Joint Parsing

Target Parse Trees

Target Sentence

Target Parser

Word Aligner

Word Alignment

Source Sentence

Source Parser

Source Parse Trees

Bitext

Bilingual Reranker

Whereas in 1992

Source Parse Trees
Whereas in 1992

Joint Parsing

Source Sentence

Word Aligner

Source Parse Trees

Target Sentence

Target Parse Trees

Bilingual Reranker

JointlyParsedTrees

Source Parser

Target Parser

Word Alignment

Target Sentence

Word Alignment
Overview

• A log-linear model over aligned parse trees

• Training with latent tree alignments

• Improvements from joint parsing
Model
Model

Sentences

High levels of product and project

产品

项目

水平

高
High levels of product and project
Alignments

\[ \alpha \]

at most 1-1 matchings between tree nodes

High levels of product and project

产品、项目水平高
High levels of product and project

Feature function: $\phi(t, a, t')$
Model

Feature function: $\phi(t, a, t')$

Log-linear model:

$$P(t, a, t'|s, s') \propto \exp(w^\top \phi(t, a, t'))$$

High levels of product and project

产品、项目水平高
Baseline Features
Baseline Features

Source tree log likelihood: \( \log P(t|s) \)

Target tree log likelihood: \( \log P(t'|s') \)
Baseline Features

Source tree log likelihood: \( \log P(t|s) \)

Target tree log likelihood: \( \log P(t'|s') \)

“Default” is to use output of baseline parsers
High levels of product and project
Bilingual Features

A “bad” aligned node pair

High levels of product and project

产品
项目
水平
高
Features on Aligned Node Pairs

Node Label Indicator
Indicator[NP, NP] = 1.0

High levels of product and project

产品、项目水平高
Features on Aligned Node Pairs

Node Label Indicator
Indicator\[\text{NP, PP} = 1.0\]

High levels of product and project
Features on Aligned Node Pairs

Span Difference
SpanDiff = 0.0

High levels of product and project
Features on Aligned Tree Pairs

Span Difference

SpanDiff = 3.0

High levels of product and project

产品

项目

水平

高
Features on Aligned Node Pairs

Word Alignments

High levels of product and project
Features on Aligned Node Pairs

Word Alignments

InsideBoth 3.0

High levels of product and project

产品 项目 水平 高
Features on Aligned Node Pairs

Word Alignments

<table>
<thead>
<tr>
<th>InsideBoth</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>InSrcOutTrg</td>
<td>0.0</td>
</tr>
</tbody>
</table>

High levels of product and project

产品、项目水平高
Features on Aligned Node Pairs

Word Alignments

<table>
<thead>
<tr>
<th>InsideBoth</th>
<th>3.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>InSrcOutTrg</td>
<td>0.0</td>
</tr>
<tr>
<td>InTrgOutSrc</td>
<td>0.0</td>
</tr>
</tbody>
</table>

High levels of product and project
Features on Aligned Node Pairs

Word Alignments

High levels of product and project
Features on Aligned Node Pairs

Word Alignments

InsideBoth 0.0

High levels of product and project
Features on Aligned Node Pairs

Word Alignments

<table>
<thead>
<tr>
<th></th>
<th>InsideBoth</th>
<th>InSrcOutTrg</th>
</tr>
</thead>
<tbody>
<tr>
<td>InsideBoth</td>
<td>0.0</td>
<td></td>
</tr>
<tr>
<td>InSrcOutTrg</td>
<td>3.0</td>
<td></td>
</tr>
</tbody>
</table>

High levels of product and project

产品
项目
水平
高
Features on Aligned Node Pairs

Word Alignments

<table>
<thead>
<tr>
<th>InsideBoth</th>
<th>0.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>InSsrcOutTrg</td>
<td>3.0</td>
</tr>
<tr>
<td>InTrgOutSrc</td>
<td>1.0</td>
</tr>
</tbody>
</table>

High levels of product and project
Features on Aligned Node Pairs

Word Alignment Posteriors

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0.2</th>
<th>0</th>
<th>0.8</th>
<th>0</th>
<th>0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>0.7</td>
<td>0.4</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>0.3</td>
<td>0.2</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
<td></td>
</tr>
</tbody>
</table>

产品
项目
水平
高
Features on Aligned Node Pairs

High levels of product and project

Word Alignment Posteriors

<table>
<thead>
<tr>
<th>InsideBoth</th>
<th>3.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>InSrcOutTrg</td>
<td>0.4</td>
</tr>
<tr>
<td>InTrgOutSrc</td>
<td>0.4</td>
</tr>
</tbody>
</table>

产品

项目

水平

高
### Features on Aligned Node Pairs

#### Word Alignment Posteriors

<table>
<thead>
<tr>
<th></th>
<th>InsideBoth</th>
<th>InSrcOutTrg</th>
<th>InTrgOutSrc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.6</td>
<td>3.8</td>
<td>0.7</td>
</tr>
</tbody>
</table>

#### Sample Alignments

**Source Sentence:**

```
High levels of product and project
```

**Target Sentence:**

```
产品、项目水平高
```

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>0.2</th>
<th>0</th>
<th>0.8</th>
<th>0</th>
<th>0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
<td>0</td>
<td>0.8</td>
</tr>
<tr>
<td>0.7</td>
<td>0</td>
<td>0.4</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.8</td>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Word Alignment Posteriors:**

- **InsideBoth:** 0.6
- **InSrcOutTrg:** 3.8
- **InTrgOutSrc:** 0.7
Features on Aligned Node Pairs

Head Word Alignments

Hard 1.0
Soft 0.8

High levels of product and project

<table>
<thead>
<tr>
<th>0</th>
<th>0.2</th>
<th>0</th>
<th><strong>0.8</strong></th>
<th>0</th>
<th>0.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td><strong>0.9</strong></td>
<td>0.1</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
<td>0</td>
<td><strong>0.8</strong></td>
</tr>
<tr>
<td>0</td>
<td>0.7</td>
<td>0.4</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>0.8</strong></td>
<td>0.3</td>
<td>0.2</td>
<td>0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
</tbody>
</table>
Features on Aligned Node Pairs

Head Word Alignments

### High levels of product and project

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.2</td>
<td>0</td>
<td>0.8</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
</tr>
<tr>
<td>0.1</td>
<td>0</td>
<td>0.1</td>
<td>0.3</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0.7</td>
<td>0.4</td>
<td>0.2</td>
<td>0</td>
</tr>
<tr>
<td>0.8</td>
<td>0.3</td>
<td>0.2</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

### 产品、项目水平高

- **Hard**: 0.0
- **Soft**: 0.4
Overview

• A log-linear model over aligned parse trees

• Training with latent tree alignments

• Improvements from joint parsing
Training with Latent Alignments
Training with Latent Alignments

- Observed: \((s, s', t, t')\)
Training with Latent Alignments

- Observed: \((s, s', t, t')\)
- Optimizing weight vector:

\[
w^* = \arg\max_w P(t, t'|s, s')
\]
Training with Latent Alignments

- Observed: \((s, s', t, t')\)
- Optimizing weight vector:

\[
\begin{align*}
    w^* &= \arg \max_w P(t, t'|s, s') \\
    &= \arg \max_w \sum_a P(t, a, t'|s, s')
\end{align*}
\]
Training with Latent Alignments

- Observed: \((s, s', t, t')\)
- Optimizing weight vector:

\[
\begin{align*}
w^* &= \arg \max_w P(t, t'|s, s') \\
&= \arg \max_w \sum_a P(t, a, t'|s, s') \\
&= \arg \max_w \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))}
\end{align*}
\]
Training with Latent Alignments

• **Observed:** $(s, s', t, t')$

• **Optimizing weight vector:**

$$ w^* = \arg \max_w P(t, t'|s, s') $$

$$ = \arg \max_w \sum_a P(t, a, t'|s, s') $$

$$ = \arg \max_w \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))} $$

**Problem 1:** Infinite sum over tree pairs
Training with Latent Alignments

- **Observed:** \((s, s', t, t')\)

- **Optimizing weight vector:**

\[
\begin{align*}
  w^* &= \arg \max_w P(t, t'|s, s') \\
  &= \arg \max_w \sum_a P(t, a, t'|s, s') \\
  &= \arg \max_w \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))}
\end{align*}
\]

Problem 2: \#P-hard sum over alignments
Approximating Sum over Tree Pairs
Approximating Sum over Tree Pairs

• Problem:

\[
P(t, t'|s, s') = \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{t,t'} \sum_a \exp(w^\top \phi(t, a, t'))}
\]
Approximating Sum over Tree Pairs

• Problem:

\[ P(t, t'|s, s') = \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t,t')} \sum_a \exp(w^\top \phi(t, a, t'))} \]

• Solution:
Approximating Sum over Tree Pairs

- Problem:

\[ P(t, t'|s, s') = \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t,t')} \sum_a \exp(w^\top \phi(t, a, t'))} \]

- Solution:
  - Train in reranking mode
Approximating Sum over Tree Pairs

• **Problem:**

\[ P(t, t'|s, s') = \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t,t')} \sum_a \exp(w^\top \phi(t, a, t'))} \]

• **Solution:**
  - Train in reranking mode
  - Candidate lists \((T, T')\) from baseline parsers
Approximating Sum over Tree Pairs

- **Problem:**
  \[
P(t, t' | s, s') = \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t')} \sum_a \exp(w^\top \phi(t, a, t'))}
  \]

- **Solution:**
  - Train in reranking mode
  - Candidate lists \((T, T')\) from baseline parsers

  \[
P(t, t' | s, s') \approx \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \sum_a \exp(w^\top \phi(t, a, t'))}
  \]
Tree Candidates
Tree Candidates

$k$ Source Candidates
Tree Candidates

$k$ Target Candidates

$k$ Source Candidates
Tree Candidates

$k$ Source Candidates

$k$ Target Candidates

$k^2$ Candidate Pairs
Tree Candidates

- New Problem:
  - Large number of candidate pairs
Training Set Pruning
Training Set Pruning

- Solution: Pruning the candidate lists
Training Set Pruning

- Solution: Pruning the candidate lists

- Start with ranked candidates from baseline parser
Training Set Pruning

• Solution: Pruning the candidate lists

• Start with ranked candidates from baseline parser

• Find gold tree
Training Set Pruning

• Solution: Pruning the candidate lists

• Start with ranked candidates from baseline parser

• Find gold tree

• Add $\epsilon$ more
Training Set Pruning

• Solution: Pruning the candidate lists

• Start with ranked candidates from baseline parser

• Find gold tree

• Add $\epsilon$ more

• Do same on target side
Training Set Pruning

- **Solution:** Pruning the candidate lists

- **Start with ranked candidates from baseline parser**

- **Find gold tree**

- **Add $\epsilon$ more**

- **Do same on target side**

- **Restrict sum to tree pairs in intersection**
Approximating Sum over Alignments
Approximating Sum over Alignments

- Problem:

\[ P(t, t'|s, s') \approx \frac{\sum_a \exp(\mathbf{w}^\top \phi(t, a, t'))}{\sum_{(t,t') \in (T, T')} \sum_a \exp(\mathbf{w}^\top \phi(t, a, t'))} \]
Approximating Sum over Alignments

- Problem:

\[ P(t, t' \mid s, s') \approx \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \sum_a \exp(w^\top \phi(t, a, t'))} \]

- Solution:

  - Replace summation with maximization
Approximating Sum over Alignments

• Problem:

\[ P(t, t' | s, s') \approx \frac{\sum_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t,t') \in (T,T')} \sum_a \exp(w^\top \phi(t, a, t'))} \]

• Solution:

• Replace summation with maximization

\[ P(t, t' | s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t,t') \in (T,T')} \max_a \exp(w^\top \phi(t, a, t'))} \]
Iterative Training
Iterative Training

- Problem: How to solve?

\[ P(t, t'|s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t,t') \in (T, T')} \max_a \exp(w^\top \phi(t, a, t'))} \]
Iterative Training

• **Problem:** How to solve?

\[
P(t, t' | s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \max_a \exp(w^\top \phi(t, a, t'))}
\]

• **Solution:** Iterative Procedure
Iterative Training

- **Problem: How to solve?**
  \[
P(t, t'|s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t,t') \in (T,T')} \max_a \exp(w^\top \phi(t, a, t'))}
\]

- **Solution: Iterative Procedure**

  1. Fix weights, optimize alignments

     \[
a^{(\tau)}_{(t,t')} = \arg \max_a w^{(\tau)^\top} \phi(t, a, t')
\]
Iterative Training

- **Problem:** How to solve?
  \[
P(t, t' | s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t, t') \in (T, T')} \max_a \exp(w^\top \phi(t, a, t'))}
\]

- **Solution:** Iterative Procedure
  1. Fix weights, optimize alignments
     \[
a^{(\tau)}_{(t, t')} = \arg \max_a w^{(\tau)} \top \phi(t, a, t')
\]
Iterative Training

• Problem: How to solve?

\[ P(t, t'|s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t,t') \in (T,T')} \max_a \exp(w^\top \phi(t, a, t'))} \]

• Solution: Iterative Procedure

1. Fix weights, optimize alignments

\[ a^{(\tau)}_{(t,t')} = \arg \max_a w^{(\tau)}^\top \phi(t, a, t') \]

2. Fix alignments, optimize weights

\[ w^{(\tau+1)} = \arg \max_w \frac{\exp(w^\top \phi(t, a^{(\tau)}_{(t,t')}, t'))}{\sum_{(t,t') \in (T,T')} \exp(w^\top \phi(t, a^{(\tau)}_{(t,t')}, t'))} \]
Iterative Training

• Problem: How to solve?

\[
P(t, t'|s, s') \approx \frac{\max_a \exp(w^\top \phi(t, a, t'))}{\sum_{(t,t') \in (T,T')} \max_a \exp(w^\top \phi(t, a, t'))}
\]

• Solution: Iterative Procedure

1. Fix weights, optimize alignments

\[
a^{(\tau)}_{(t,t')} = \arg \max_a w^{(\tau)} \top \phi(t, a, t')
\]

2. Fix alignments, optimize weights

\[
w^{(\tau+1)} = \arg \max_w \frac{\exp(w^\top \phi(t, a^{(\tau)}_{(t,t')}, t'))}{\sum_{(t,t') \in (T,T')} \exp(w^\top \phi(t, a^{(\tau)}_{(t,t')}, t'))}
\]
Overview

- A log-linear model over aligned parse trees
- Training with latent tree alignments
- Improvements from joint parsing
Parsing Experimental Setup

Chinese Treebank

English Translations

Penn Treebank
Parsing Experimental Setup

Chinese Treebank

English Translations

Penn Treebank

Baseline Parser Training
Parsing Experimental Setup

Chinese Treebank

English Translations

Penn Treebank

Baseline Parser Training
Bilingual Reranker Training
Parsing Experimental Setup

Chinese Treebank

English Translations

Penn Treebank

Baseline Parser Training

Bilingual Reranker Training

Development Set
Parsing Experimental Setup

Chinese Treebank

Baseline Parser Training

Bilingual Reranker Training

Development Set

Test Set

English Translations

Penn Treebank
Parsing Results

<table>
<thead>
<tr>
<th>Language</th>
<th>Baseline</th>
<th>Joint</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>83.50</td>
<td>85.25</td>
<td>+1.75</td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Parsing Results

F-measure

<table>
<thead>
<tr>
<th>Language</th>
<th>Baseline</th>
<th>Joint</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chinese</td>
<td>83.50</td>
<td>85.25</td>
</tr>
<tr>
<td>English</td>
<td>79.25</td>
<td>81.72</td>
</tr>
</tbody>
</table>

Baseline: +1.8
Joint: +2.5
Machine Translation Experimental Setup
• Syntax-based MT:

Translation rules derived from target-side trees
(Galley et al, 2004; 2006)
Machine Translation Experimental Setup

- Syntax-based MT:
  Translation rules derived from target-side trees (Galley et al, 2004; 2006)

- Comparison:
  Target trees from monolingual parser vs Target trees from joint parser
Machine Translation Experimental Setup

- Syntax-based MT:
  Translation rules derived from target-side trees (Galley et al, 2004; 2006)

- Comparison:
  Target trees from monolingual parser vs Target trees from joint parser

- Remaining data is identical
Machine Translation Results

Moses (Baseline)
Machine Translation Results

BLEU

Monolingual Parsing   Joint Parsing

Moses (Baseline)
Machine Translation Results

Monolingual Parsing: 18.7
Joint Parsing: Moses (Baseline)
Machine Translation Results

BLEU

Monolingual Parsing: 18.7
Joint Parsing: 21.1

Moses (Baseline) +2.4
Conclusions

- Bilingual constraints can be leveraged to improve parse quality
Conclusions

• Bilingual constraints can be leveraged to improve parse quality

• Parsing gets better
  • In-domain Chinese $F_1$ improves by 1.8
  • Out-of-domain English $F_1$ improves by 2.5
• Bilingual constraints can be leveraged to improve parse quality

• Parsing gets better
  • In-domain Chinese F₁ improves by 1.8
  • Out-of-domain English F₁ improves by 2.5

• Downstream machine translation gets better
  • BLEU improves by 2.4
Thank You