Motivation
We observe that learning the unsupervised Convex-MST model on the English corpus and then directly applying it to parse other languages produces surprisingly good results.

Unsupervised Dependency Parsing
Dependency parsing pipeline (POS tagging step can be omitted):

I swam yesterday
Shanghai is beautiful
Empirical Method in NLP
PRP VB NN
NNP VB JJ
JJ NN IN NN
PRP VB NN
NNP VB JJ
JJ NN IN NN

Unsupervised dependency parsing:
- No access to golden trees.
- Usually taking POS tags as inputs.
- Typically incorporating different types of inductive biases.

Proposed Approach
We propose three regularization methods for bilingual grammar induction that encourage similarity between models of two languages in terms of model parameters, dependency edge scores, and parse trees respectively.

Empirical Results
Empirical results on multiple languages show that our methods outperform baselines.

Bilingual Knowledge Sharing
Three levels of knowledge:
- the parameter $w$,
- the edge score $s$,
- the parse $y$.

Vanilla Joint Training:

$J(w_1, w_2; X_1, X_2) = J(w_1; X_1) + J(w_2; X_2)$

Regularization of Weight Parameters (W-Reg)

$J(w_1, w_2; X_1, X_2) = J(w_1; X_1) + J(w_2; X_2) + \lambda \|w_1 - w_2\|^2$

Regularization of Edge Scores (E-Reg)

$J(w_1, w_2; X_1, X_2) = J(w_1; X_1) + J(w_2; X_2) + \lambda \|s(w_1(x, h, m) - s(w_2(x, h, m))\|^2$

Regularization of Parse Trees (T-Reg)

$J(w_1, w_2; X_1, X_2) = J(w_1; X_1) + J(w_2; X_2) + \lambda \|D(w_1; x, y) - D(w_2; x, y)\|_{T-Reg}$

Experiments
- Training data: Sentence length no more than 10.
- Testing data: Sentence length no more than 10 and all sentences.

Table 3: Transfer grammar induction from English to other languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>EN</th>
<th>DE</th>
<th>FR</th>
<th>ID</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>57.5</td>
<td>59.0</td>
<td>68.3</td>
<td>66.2</td>
<td>67.8</td>
</tr>
</tbody>
</table>

Table 4: Results of bilingual grammar induction on UD 1.4.

<table>
<thead>
<tr>
<th>Language</th>
<th>EN-ID</th>
<th>EN-ID</th>
<th>EN-ID</th>
<th>EN-ID</th>
<th>EN-ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy</td>
<td>57.5</td>
<td>59.0</td>
<td>68.3</td>
<td>66.2</td>
<td>67.8</td>
</tr>
</tbody>
</table>

Figure 1. Transfer grammar induction from English to the other language.

Figure 2. Results of bilingual grammar induction on test sentences no longer than 10.