Preordering
Change source-side word order to make it more similar to target-side word order. Can improve word alignments and can be combined with any translation system.

Classifier Preordering
Predict the target word order by treating each permutation as a label in a multi-class classifier. Traverse the parse tree, reordering each family (head and children) and recursing.

Each classifier is trained on up to 15M training instances extracted from automatically aligned data. Limit to 20 most common permutations for ≥ 4 words.

1-Step Classifier

GradBoost

Features:
- Head wording
- Children and position
- Words/tags between children

Pairs of the above.

Separate classifiers for 2, 3, ..., 7 involved words.

2-Step Classifier
Decompose the search space: first determine the position of every child relative to the head (pivot) and then order the children before and after the head. Think QuickSort without recursion.

Reordering example from Genzel (2010), where the system learned to move RB over NN.

When multiple rules apply: Use All? Most Frequent? Most Specific? (worrisome 'Most Frequent' and 'Most Specific' are at opposite ends)

Related Work: Precedence Ranking
Assign a precedence score to each clause, then sort.

<table>
<thead>
<tr>
<th>Tag</th>
<th>(Label, Weight, Order)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VB*</td>
<td>(adv, 1, NORMAL)</td>
</tr>
<tr>
<td></td>
<td>(adverb, 1, NORMAL)</td>
</tr>
<tr>
<td></td>
<td>(prep, 0, NORMAL)</td>
</tr>
<tr>
<td></td>
<td>(adv, -1, NORMAL)</td>
</tr>
<tr>
<td></td>
<td>(prep, 0, REVERSE)</td>
</tr>
<tr>
<td></td>
<td>(adv, 1, REVERSE)</td>
</tr>
</tbody>
</table>

Precedence rules from Xu et. al (2009). Rules were extracted manually by a bilingual speaker.

Precedence score computed independently for each clause/word:
- No global consistency and hard to express priors
  (e.g., A-B-C-D and D-C-B-A are likely but C-A-D-B is not).

Results
Improvements relative to a phrase-based baseline for 1-step and 2-step classifier preordering and the system of Genzel 2010 (rule).

Analysis
Some reordering decisions depend heavily on lexical choice.

Conclusions
- New approach: Directly predict in the exponential permutation space.
- Decompose space via 2-step approach.
- Limit to 20 most likely outcomes.
- Provides an elegant framework to combining different "rules."
- Treat "rules" as features in a discriminative classifier.
- Can naturally express global consistency as priors.