Preordering and Word Order Freedom in Machine Translation
Reordering in PBMT
Reordering in PBMT
Reordering in PBMT

Preordering
+ Works well for English–Japanese, English–Korean, ...
Reordering in PBMT

Preordering
+ Works well for English–Japanese, English–Korean, ...
− Not that well for English–German, English–Czech, ...
Reordering in PBMT

Why the partial success?

- **Free word order** languages are difficult: many potential target word orders
Reordering in PBMT

Why the partial success?

- **Free word order** languages are difficult: many potential target word orders
- **Source side often not enough** to predict a unique target word order
Reordering in PBMT

Why the partial success?

- **Free word order** languages are difficult: many potential target word orders
- **Source side often not enough** to predict a unique target word order

Potential solution:

- Narrow down space of target word orders and let decoder decide
Reordering in PBMT

Why the partial success?

▶ **Free word order** languages are difficult: many potential target word orders

▶ **Source side often not enough** to predict a unique target word order

Potential solution:

▶ Narrow down space of target word orders and let decoder decide

▶ Prepare MT system for this kind of input
Outline

1. Word Order Freedom
   - What exactly do we mean by free word order?
   - Can we measure it?
Outline

1. Word Order Freedom
   - What exactly do we mean by free word order?
   - Can we measure it?

2. Translation with Permutation Lattices
   - Permutation Lattices
   - Lattice Silver Training
Outline

1. Word Order Freedom
   - What exactly do we mean by free word order?
   - Can we measure it?

2. Translation with Permutation Lattices
   - Permutation Lattices
   - Lattice Silver Training

3. Experiments with two language pairs
   - English–Japanese
     *Strict word order language*
   - English–German
     *Free word order language*
What Does Free Word Order Mean?

- Multiple word orders expressing the same meaning
What Does Free Word Order Mean?

- Multiple word orders expressing the same meaning
  → Fuzzy definition!
- Can we quantify and compare languages’ word order freedom?
Estimating Word Order Freedom: Our Bilingual Approach
Word Order Freedom

Graph showing bilingual head direction entropy for various languages, with Berber, Turkish, Hebrew, German, Russian, French, Spanish, Italian, Portuguese, Esperanto, Japanese, and Mandarin Chinese listed. The y-axis represents the languages, and the x-axis represents the entropy values ranging from 0 to 0.6.
Free Word Order

What to do about Free Word Order?

- Instead of choosing 1-best word order, work with $n$-best
Free Word Order

What to do about Free Word Order?

▶ Instead of choosing 1-best word order, work with n-best

Decoding with n-best permutations:

1) **Inference:** Compactly encode n-best WOs into lattice
2) **Training:** Estimate parameters for translating lattices
Permutation Lattices
Lattice Training

Q: How to extract the phrase table?

Gold Training

\[(\text{target-order} \ s', \ t)\]

+ compact
+ less noisy
– may not cover
  preordered input
Lattice Training

Q: How to extract the phrase table?

Gold Training
(target-order \( s', t \))

- + compact
- + less noisy
- - may not cover preordered input

Predicted Training
(predicted \( s', t \))

- ? less compact
- - more noisy
- + covers preordered input
Lattice Training

Q: How to extract the phrase table?

**Gold Training**
(target-order \(s', t\))

+ compact
+ less noisy
– may not cover preordered input

**Predicted Training**
(predicted \(s', t\))

? less compact
– more noisy
+ covers preordered input

**Lattice Silver Training**
(s’ from \(n\)-best most similar to gold, \(t\))

+ compact
+ less noisy
+ covers preordered input
Preordering Models

Reordering Grammar
Stanojević and Sima’an (2015)
- Hierarchical, probabilistic model
- Unsupervised (no syntax)
- Beyond ITG (based on PETs)

Neural Preordering
de Gispert et al. (2015); Jehl et al. (2014)
- Preorders dep. tree recursively
- NN to decide pairwise ordering of subtrees
- We extend search to $n$-best branch-and-bound
Translation Experiment: English → German

On WMT newstest 2015.

- Big oracle WO gain: +4.92 over baseline (21.76 BLEU)
- 1-best with DL 6 performs poorly
- Lattices with Silver Training and DL 0 on par with baseline +0.12

Software: Moses (with extended lattice support). Optimizer: Batch MIRA (15 iterations, BLEU averaged over 3 runs).
Translation Experiment: English $\rightarrow$ Japanese

On NTCIR Patent translation task.

- Big oracle WO gain: $+4.57$ over baseline (29.65 BLEU)
- 1-best with DL 6 performs well: $+2.49$
- Lattices with DL 0 enables additional improvement: $+2.85$

Software: Moses (with extended lattice support). Optimizer: Batch MIRA (15 iterations, BLEU averaged over 3 runs).
Summary of our Contributions

**Bilingual Head Direction Entropy:**
- Information-theoretic measure of word order freedom

**Permutation Lattices** produced from preordering models:
- Valuable tool to address uncertainty in word order
- Strict word order (En→Ja)
  - Additional improvements over 1-best preordering with DL 0
- Free word order (En→De)
  - Lattice Silver Training important for model estimation
  - Enable automatic preordering with DL 0
Thank You!

Questions?


# Word Order Freedom: Micro Level

<table>
<thead>
<tr>
<th></th>
<th>verb</th>
<th>noun</th>
<th></th>
<th>verb</th>
<th>noun</th>
</tr>
</thead>
<tbody>
<tr>
<td>noun</td>
<td>79.7%</td>
<td>79.5%</td>
<td>det</td>
<td>78.6%</td>
<td></td>
</tr>
<tr>
<td>- Sbj</td>
<td>63.3%</td>
<td></td>
<td>- Sbj</td>
<td>66.3%</td>
<td></td>
</tr>
<tr>
<td>- Obj</td>
<td>82.6%</td>
<td></td>
<td>- Obj</td>
<td>22.7%</td>
<td></td>
</tr>
<tr>
<td>- Adv</td>
<td>84.3%</td>
<td></td>
<td>- Adv</td>
<td>35.4%</td>
<td></td>
</tr>
<tr>
<td>adv</td>
<td>54.5%</td>
<td></td>
<td>adv</td>
<td>41.3%</td>
<td></td>
</tr>
</tbody>
</table>

(a) English–Japanese

(b) English–German
Number of Permutations vs. Lattice Size

<table>
<thead>
<tr>
<th>Permutations</th>
<th>Kendall $\tau$</th>
<th>Lattice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monotone</td>
<td>83.78</td>
<td>States</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>123</td>
<td>123</td>
</tr>
</tbody>
</table>

Table: Permutations and lattice size (En–De).
## Experiments: Neural Reordering on English-German

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Word order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DL</td>
<td>BLEU</td>
</tr>
<tr>
<td>Baseline</td>
<td>6</td>
<td>21.76</td>
</tr>
<tr>
<td>Oracle order</td>
<td>6</td>
<td>26.68</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>26.41</td>
</tr>
<tr>
<td>First-best</td>
<td>6</td>
<td>21.21$^A$</td>
</tr>
<tr>
<td>Lattice (silver)</td>
<td>0</td>
<td>21.88$^B$</td>
</tr>
</tbody>
</table>

Software: Moses (with extended lattice support).
Optimizer: Batch MIRA (15 iterations, BLEU averaged over 3 runs).
Training data: 4.5m sentence pairs (WMT16). LM (5-gram): 189m sentences (WMT16).

# Experiments: Reordering Grammar on English-Japanese

<table>
<thead>
<tr>
<th></th>
<th>Translation</th>
<th>Word order</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DL</td>
<td>BLEU</td>
</tr>
<tr>
<td>Baseline</td>
<td>6</td>
<td>29.65</td>
</tr>
<tr>
<td>Oracle order</td>
<td>6</td>
<td>34.22</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>30.55</td>
</tr>
<tr>
<td>First-best</td>
<td>6</td>
<td>32.14$^A$</td>
</tr>
<tr>
<td>Lattice</td>
<td>0</td>
<td>32.50$^{AB}$</td>
</tr>
</tbody>
</table>
