Syntactic Reordering of Source Sentences for Statistical Machine Translation

Mohammad Sadegh Rasooli

Columbia University

rasooli@cs.columbia.edu

April 9, 2013
Overview

1. **First Paper: Collins, et al. (2005)**
   - The Role of Syntax in SMT
   - Syntactic Preprocessing Approaches
   - Clause Restructing
   - Experiments
   - Discussion

   - Approaches to Syntactic Reordering
   - Translation Between SVO and SOV Languages
   - Precedence Reordering Based on a Dependency Parser
   - Experiments
   - Discussion
In the original phrase-base SMT, syntax is not taken into account.

Phrase-based systems have limited potential to model word-order differences between languages.

- The word order differences between languages are considered as distortion.
- Each reordering rule adds distortion penalties to the overall score of the translation model.
I will pass on to you the corresponding comments, so that you can adopt them perhaps in the vote.

I will to you the corresponding comments pass on, so that you them perhaps in the vote adopt can.
Research on Syntax in MT

- Changing the word order of one of the languages or both, to make their word order more similar to each other.

- **Syntax-Based MT Approaches**
  - Make use of bitext grammars to parse both parts.
  - Change the syntax of target language alone.
  - Transform the translation problem into a parsing problem.

- **Reranking methods**
  - Select between N-best results of the phrase-based system, using syntactic information.

- **Preprocessing Approaches**
  - The source language sentences are modified before translation.
  - This approach is used in this paper.
I will pass on to you the corresponding comments, so that you can adopt them perhaps in the vote.

German

Ich werde Ihnen die entsprechenden Einwände übermitteln, damit Sie sie vielleicht im Stimmengenuss übernehmen können.

German (Preprocessed)

Ich werde Ihnen die entsprechenden Einwände mitteilen, damit Sie sie vielleicht im Stimmengenuss übernehmen können.
Clause Restructuring

Steps (both in training and decoding)
1. Parse the source sentence.
2. Apply reordering rules on the source sentence.
3. Use phrase-based models.
### Six Reordering Rules in German

<table>
<thead>
<tr>
<th>Transformation</th>
<th>Example</th>
</tr>
</thead>
</table>
| Verb Initial     | Before: Ich werde Ihnen die entsprechenden Anmerkungen *aushaendigen*, ...  
                  | After:  *Ich werde aushaendigen Ihnen* die entsprechenden Anmerkungen, ...  
                 | I shall be passing on to you some comments, ... |
| Verb 2nd         | Before: ... damit Sie übernehmen das eventuell bei der Abstimmung *koennen*.  
                  | After: ... damit *koennen Sie* übernehmen das eventuell bei der Abstimmung.  
                 | ... so that you could adopt this perhaps in the voting. |
| Move Subject     | Before: ... damit *koennen Sie* übernehmen das eventuell bei der Abstimmung.  
                  | After: ... damit *Sie koennen* übernehmen das eventuell bei der Abstimmung.  
                 | ... so that you could adopt this perhaps in the voting. |
| Particles        | Before: Wir fordern das Praesidium *auf*, ...  
                  | After: Wir *auf fordern* das Praesidium, ...  
                 | We ask the Bureau, ... |
| Infinitives      | Before: Ich werde der Sache *nachgehen* dann, ...  
                  | After: Ich werde *nachgehen* der Sache dann, ...  
                 | I will look into the matter then, ... |
| Negation         | Before: Wir konnten einreichen es *nicht* mehr rechtzeitig, ...  
                  | After: Wir konnten *nicht* einreichen es mehr rechtzeitig, ...  
                 | We could not hand it in in time, ... |

*Table 1: Examples for each of the reordering steps. In each case the item that is moved is underlined.*
Experiments

- Experimental setup
  - Data: Europarl Corpus.
  - 751,088 parallel sentence.
  - Evaluation on 2000 sentences.
  - Average sentence length: 28 words
  - Baseline: no reordering phrase-based system.

- Results (BLEU score)
  - Baseline: 25.2%
  - Reordering: 26.8%
Human Translation Judgments

- Two annotators judged 100 sentences (10 to 20 words in length; chosen at random).
- Three versions: Human, baseline, reordered.
- Judgments: Worse/better or equal.

<table>
<thead>
<tr>
<th></th>
<th>Better</th>
<th>Equal</th>
<th>Worse</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotator 1</td>
<td>40%</td>
<td>40%</td>
<td>20%</td>
</tr>
<tr>
<td>Annotator 2</td>
<td>44%</td>
<td>37%</td>
<td>19%</td>
</tr>
</tbody>
</table>
Human

i think it is wrong in principle to have such measures in the european union.

Reordered

i believe that it is wrong in principle to take such measures in the european union.

Baseline

i believe that it is wrong in principle such measures in the european union to take.
Authors use sign test for statistical significance.

- \( f(x) \) is + if better than baseline, \( f(x) \) is - if worse; and \( f(x) \) is = if equal
- \( p_+ \): probability of \( (f(x) \) is +) and \( p_- \): probability of \( f(x) \) is minus

BLEU does not have per-sentence evaluation.

Authors create an artificial comparison:

- \( s \): baseline BLEU score
- \( s_i \): baseline BLEU score except the sentence \( i \) translated by the reordered model.

- \( f(x) \) is + is \( s_i > s \); \( f(x) \) is - is \( s_i < s \).

52.85% improved, 36.4% worse than baseline and 10.75% equal.

With 95% confidence, this method improves the baseline.
The method clearly improves the baseline.

The rules are language-specific (even cannot be used for English to German translation).

The authors did not try to learn reordering rules automatically.
Explicitly model phrase reordering distances; e.g. distance based distortion models.

Syntactic analysis of the target language into both modeling and decoding.

Reordering source sentences based on syntactic analysis

  This paper uses this approach
Subject-Verb-Object (SVO) and Subject-Object-Verb (SOV) are two common word order in the world languages.

English is SVO and Korean is SOV.

“John hit the ball.” vs. “John the ball hit.”

When the sentences get longer, the cost of moving structures during decoding (in phrase-based models) can be quite high.

“English is used as the first or second language in many countries around the world.”

→ “is used” should skip 13 words to go to the end of the sentence.
The children of each word have some relative ordering.

A **Precedence reordering rule** is a mapping from $T$ to a set of tuples \{(L, W, O)\}

- $T$: POS tag
- $L$: Dependency label
- $W$: Weight indicating the order (highest to lowest)
  - Children with the same weights are ordered according to the order defined in the rule.
  - **Why not explicitly pre-define unequal weights?**
- $O$: order type
  - NORMAL: preserve the original order
  - RESERVE: flip the order

If a node is not listed in the rules, $W = 0$ and $O = NORMAL$

Use **self** to refer to the head node itself.

Punctuations and conjugations disallow movements across them.
After apply precedence rule, this will be: *John the ball hit can.*

Table 1: Precedence Rules to Reorder English to SOV Language Order (These rules were extracted manually by a bilingual speaker after looking at some text book examples in English and Korean, and the dependency parse trees of the English examples.)
Novelties in This Work

1. This model is more efficient than its counterpart.
2. Outperforms the state-of-the-art (stronger baseline).
3. It is not restricted to one language pair.
4. It is possible to automatically learn precedence rules.
5. They use dependency parse trees rather than constituency trees.
Experiments

- English to 5 SOV languages.
- Baseline: Maximum entropy based lexicalized phrase reordering model.
  - Maximum allowed reordering: 10.
- Parser: Deterministic transition-based dependency parser.
  - Parses in linear time.
- Another baseline: Hierarchical phrase-based system.
  - Can capture long distance reordering by using a PCFG model.
  - Uses chart parsing during decoding: slower than deterministic dependency parser.
- 9.5K English sentences (from web) as evaluation data.
  - 3,500 sentences for dev (to perform MERT).
  - 1,000 sentences for test.
  - 5,000 sentences for blind test.
Experiments

<table>
<thead>
<tr>
<th>System</th>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>English→Korean</td>
<td>303M</td>
<td>267M</td>
</tr>
<tr>
<td>English→Japanese</td>
<td>316M</td>
<td>350M</td>
</tr>
<tr>
<td>English→Hindi</td>
<td>16M</td>
<td>17M</td>
</tr>
<tr>
<td>English→Urdu</td>
<td>17M</td>
<td>19M</td>
</tr>
<tr>
<td>English→Turkish</td>
<td>83M</td>
<td>76M</td>
</tr>
</tbody>
</table>

Table 2: Training Corpus Statistics (#words) of Systems for 5 SOV Languages
### Results

<table>
<thead>
<tr>
<th>Language</th>
<th>System</th>
<th>dev</th>
<th>test</th>
<th>blind</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korean</td>
<td>BL</td>
<td>25.8</td>
<td>27.0</td>
<td>26.2</td>
</tr>
<tr>
<td></td>
<td>-LR</td>
<td>24.7</td>
<td>25.6</td>
<td>25.1</td>
</tr>
<tr>
<td></td>
<td>-LR+PR</td>
<td>27.3</td>
<td>28.3</td>
<td>27.5**</td>
</tr>
<tr>
<td></td>
<td>+PR</td>
<td>27.8</td>
<td>28.7</td>
<td>27.9**</td>
</tr>
<tr>
<td>Japanese</td>
<td>BL</td>
<td>29.5</td>
<td>29.3</td>
<td>29.3</td>
</tr>
<tr>
<td></td>
<td>-LR</td>
<td>29.2</td>
<td>29.0</td>
<td>29.0</td>
</tr>
<tr>
<td></td>
<td>-LR+PR</td>
<td>30.3</td>
<td>31.0</td>
<td>30.6**</td>
</tr>
<tr>
<td></td>
<td>+PR</td>
<td>30.7</td>
<td>31.2</td>
<td>31.1**</td>
</tr>
<tr>
<td>Hindi</td>
<td>BL</td>
<td>19.1</td>
<td>18.9</td>
<td>18.3</td>
</tr>
<tr>
<td></td>
<td>-LR</td>
<td>17.4</td>
<td>17.1</td>
<td>16.4</td>
</tr>
<tr>
<td></td>
<td>-LR+PR</td>
<td>19.6</td>
<td>18.8</td>
<td>18.7**</td>
</tr>
<tr>
<td></td>
<td>+PR</td>
<td>19.9</td>
<td>18.9</td>
<td>18.8**</td>
</tr>
<tr>
<td>Urdu</td>
<td>BL</td>
<td>9.7</td>
<td>9.5</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td>-LR</td>
<td>9.1</td>
<td>8.6</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>-LR+PR</td>
<td>10.0</td>
<td>9.6</td>
<td>9.6**</td>
</tr>
<tr>
<td></td>
<td>+PR</td>
<td>10.0</td>
<td>9.8</td>
<td>9.6**</td>
</tr>
<tr>
<td>Turkish</td>
<td>BL</td>
<td>10.0</td>
<td>10.5</td>
<td>9.8</td>
</tr>
<tr>
<td></td>
<td>-LR</td>
<td>9.1</td>
<td>10.0</td>
<td>9.0</td>
</tr>
<tr>
<td></td>
<td>-LR+PR</td>
<td>10.5</td>
<td>11.0</td>
<td>10.3**</td>
</tr>
<tr>
<td></td>
<td>+PR</td>
<td>10.5</td>
<td>10.9</td>
<td>10.4**</td>
</tr>
</tbody>
</table>

**Table 3:** BLEU Scores on Dev, Test and Blindtest for English to 5 SOV Languages with Various Reordering Options (BL means baseline, LR means maximum entropy based lexicalized phrase reordering model, PR means precedence rules based preprocessing reordering.)
Table 4: BLEU Scores on Dev, Test and Blindtest for English to 5 SOV Languages in Hierarchical Phrase-based Systems (PR is precedence rules based preprocessing re-ordering, same as in Table 3, while Hier is the hierarchical system.)
Reordering of languages with different word orders is essential.
The method seems to work fine for 5 languages.
Although authors claim that the rule can be extracted automatically,
they did not try.
The improvement of the basic over hierarchical phrase-based is not
significant.
Thanks!