Syntactic structure alignment based Statistical machine translation

Doctoral Thesis Proposal

Sun Jun
Chinese/English Bilingual Text

Chinese

Translation Model $P(f|e)$

Decoding algorithm $\arg\max_{e} P(e) \cdot P(f|e)$

Broken English

I am so hungry

Statistical Analysis

English Text

Statistical Analysis

English

Language Model $P(e)$
SMT in nutshell (cont.)
Syntax based----Why Syntax?

- Much more grammatical output
- Accurate control over re-ordering
- Appropriate insertion of function words
- Word translations need to depend on grammatically-related words
Syntax based

- **String-to-tree** (Galley et al., 2006; Marcu et al., 2006; Yamada and Knight, 2001):
  - Source sentence: a word sequence; Target sentence: a tree
  - Decoding as parsing (CYK)

- **Tree-to-string** (Quirk et al., 2005; Liu et al., 2006):
  - Source sentence: a tree; Target sentence: a word sequence
  - Tree2string rules
  - Decoding more easier (translate each internal node in a bottom-up manner)

- **Tree-to-tree** (Eisner, 2003; Graehl and Knight, 2004; Zhang et al., 2007):
  - Source sentence: a tree; Target sentence: a tree
  - Tree2tree rules
  - Decoding as Tree2string (bilingual parsing)
Motivations

- More & Variant Syntactic translational equivalences (TEs)
  - Continuous phrases vs. discontinuous phrases
  - Syntactic phrases vs. non-syntactic phrases
- Comparison of TEs are still not well studied
  - Local/Global Reordering rules
  - Continuous/discontinuous phrasal rules
  - Lexical/partial-lexical/non-lexical rules
- TEs acquired by word alignment
  - Only flat features adopted
Motivations (cont.)

- A more powerful syntax-based translation model
  - Global reordering ... (as former syntax-based models)
  - Improve grammaticality of target translation results (tree2tree)
  - Well model all kinds of TEs
    - Continuous phrases vs. non-continuous phrases
    - Syntactic phrases vs. non-syntactic phrases
- To acquire more effective TEs (syntactic structure alignment)
  - Overcome easy but non-accurate word alignment based methods
  - A structure alignment framework
  - Design more effective syntactic features
- Specified decoding algorithm & parameterization methods
Pre Work/TM

- Synchronous Tree Substitution Grammar
- Synchronous Tree sequence Substitution Grammar
- Synchronous discontinuous Tree sequence Substitution Grammar
Pre Work/TM/STSG

TE

[Diagram showing the syntax tree for the sentence: 把钢笔给(笔)(me).]

English translation:

Give the pen to me.

Chinese original:

把钢笔给(笔)(me).
Pre Work/TM/STSG (cont.)

TE

Give the pen to me. (give)

Verbs: VBP, VBP

NP PP

PP NN TO PRP PUNC.

VG VBP

Give ① give ①

给 ① give ①
Pre Work/TM/STSG (cont.)
Pre Work/TM/STSG (cont.)

E1: \[
\begin{align*}
VBA & \\
\text{把} & \text{我} \\
\text{NULL} & \text{(me)} \\
\end{align*}
\]

E2: \[
\begin{align*}
\text{钢笔} & \\
\text{pen} & \\
\end{align*}
\]

E3: \[
\begin{align*}
\text{给} & \\
\text{(give)} & \\
\end{align*}
\]
Pre Work/TM/STSG (cont.)

- **Pros**
  - Non-isomorphic tree alignment
  - Multi-level global structure distortion
  - Discontinuous phrases

- **Cons**
  - Non-syntactic phrases
Pre Work/TM/STSSG

Synchronous Tree sequence Substitution Grammar

ETS1:

\[
\begin{align*}
&E_
\text{TS1}: &
\begin{cases}
 &\text{NG} &\text{VG} \\
 &\text{pen} &\text{give}
\end{cases} &
\begin{cases}
 &\text{VBP} &\text{NP} \\
 &\text{Give} &\text{the} &\text{pen}
\end{cases}
\end{align*}
\]

ETS2:

\[
\begin{align*}
&E_
\text{TS2}: &
\begin{cases}
 &\text{VBA} &\text{VO} \\
 &\text{null} &\text{give} &\text{me}
\end{cases} &
\begin{cases}
 &\text{VBP} &\text{TO} &\text{PRP} \\
 &\text{to} &\text{me}
\end{cases}
\end{align*}
\]
Pre Work/TM/STSSG (cont.)

- **Pros**
  - Fully cover STSG
  - Non-syntactic phrases

- **Cons**
  - Discontinuous phrases only acquired from continuous TE
Pre Work/TM/SDTSSG

Synchronous discontinuous Tree sequence Substitution Grammar
Pre Work/TM/SDTSSG (cont.)

- **Pros**
  - Fully cover STSG & STSSG
  - Discontinuous phrases obtained from both continuous TEs and discontinuous TEs

- **Cons**
  - Solve a few language phenomena but not much
  - Noisy rules obtained from wrong word alignments
Pre Work/TM/SDTSSG (cont.)

- Why SDTSSG works?

  E.g. 两人 就 对簿公堂。

  - STSSG:
    - the two candidates would 对簿公堂

  - SDTSSG:
    - the two people can confront other countries at court leisurely manner

- Reasons: more effective TEs obtained

  - VV(对簿公堂) ||| VB(confront) NP(JJ(other),NNS(countries)) IN(at) NN(court) ... JJ(leisurely) NN(manner)
Why SDTSSG works?

E.g. 另一方面 英国 的 资料 则 预期 将 显示 工业 活动 增加。

STSSG:

- data while Britain is expected to show that the number of industrial activities.

SDTSSG:

- British data on the other hand, predicted that industrial activities will increase.

Reasons: more effective TEs obtained

- AD(将) ... VV(增加) ||| MD(will) VB(increase)
Pre Work/TM(cont)

- Experiment setting (basic setting’s same for all models)
  - Training data: FBIS (7.2M (Chinese)+9.2M(English) words)
  - Develop set: NIST-02
  - Test set: NIST-05

- Performance

<table>
<thead>
<tr>
<th>Model</th>
<th>BLEU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moses</td>
<td>23.86</td>
</tr>
<tr>
<td>STSG</td>
<td>24.71</td>
</tr>
<tr>
<td>STSSG</td>
<td>25.66</td>
</tr>
<tr>
<td>SDTSSG</td>
<td>26.17</td>
</tr>
</tbody>
</table>
Pre Work/Structure alignment

- Problem definition
  - Tree sequence alignment

- Sub-tree alignment
Pre Work/Structure alignment/Sub-tree alignment

- Previous Work
  - Heuristics
  - Modify the original aligned structures
  - Most work on dependency constraint
  - Ask for word alignment as a premise
Pre Work/Structure alignment/Sub-tree alignment (cont.)

- preservation of the given tree structures;
- word-level alignments not fixed a priori
- structure features
- statistical approach
Pre Work/Structure alignment/Sub-tree alignment (cont.)

- Two perspective of sub-tree structure
  - Inside-Outside View
  - Top-Down generative View
Pre Work/Structure alignment/Sub-tree alignment (cont.)

- Sub-Structure Extraction
  - Head nodes
  - Tokens
  - Trunk
    - POS tags
    - CFG rules
  - Path Generated Element (PGE)
    - Inside PGE: VP-vbz; VP-NP-art; VP-NP-nn;
    - Ourside PGE: S-INTJ-uh; S-com; S-NP-prp; S-fsp
Pre Work/Structure alignment/Sub-tree alignment (cont.)

Metrics

\[
\text{precision} = \frac{|A \cap S|}{|A|} \quad \text{recall} = \frac{|A \cap S|}{|S|}
\]

\[
\text{AER}^* = 1 - \frac{|A \cap S| + |A \cap P|}{|A| + |S|} \quad \text{AER} = 1 - \frac{2|A \cap S|}{|A| + |S|}
\]
<table>
<thead>
<tr>
<th>Exp ID</th>
<th>sub-model combination</th>
<th>Prior</th>
<th>sub-structure model</th>
<th>precision</th>
<th>recall</th>
<th>AER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>word</td>
<td>root</td>
<td>trunk</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Prior_word_root_POS</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>POS</td>
<td>0.5824</td>
</tr>
<tr>
<td>2</td>
<td>Prior_word_root_CFG</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>CFG</td>
<td>0.6004</td>
</tr>
<tr>
<td>3</td>
<td>word</td>
<td>--</td>
<td>Y</td>
<td>--</td>
<td>--</td>
<td>0.4488</td>
</tr>
<tr>
<td>4</td>
<td>root</td>
<td>--</td>
<td>--</td>
<td>Y</td>
<td>--</td>
<td>0.4862</td>
</tr>
<tr>
<td>5</td>
<td>PGE</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>PGE</td>
<td>0.4683</td>
</tr>
<tr>
<td>6</td>
<td>word_root</td>
<td>--</td>
<td>Y</td>
<td>Y</td>
<td>--</td>
<td>0.5838</td>
</tr>
<tr>
<td>7</td>
<td>word_PGE</td>
<td>--</td>
<td>Y</td>
<td>--</td>
<td>PGE</td>
<td>0.5930</td>
</tr>
<tr>
<td>8</td>
<td>root_PGE</td>
<td>--</td>
<td>--</td>
<td>Y</td>
<td>PGE</td>
<td>0.4905</td>
</tr>
<tr>
<td>9</td>
<td>word_root_PGE</td>
<td>--</td>
<td>Y</td>
<td>Y</td>
<td>PGE</td>
<td>0.5796</td>
</tr>
<tr>
<td>10</td>
<td>Prior_word</td>
<td>Y</td>
<td>Y</td>
<td>--</td>
<td>--</td>
<td>0.4751</td>
</tr>
<tr>
<td>11</td>
<td>Prior_root</td>
<td>Y</td>
<td>--</td>
<td>Y</td>
<td>--</td>
<td>0.4915</td>
</tr>
<tr>
<td>12</td>
<td>Prior_PGE</td>
<td>Y</td>
<td>--</td>
<td>--</td>
<td>PGE</td>
<td>0.5378</td>
</tr>
<tr>
<td>13</td>
<td>Prior_word_root</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>--</td>
<td>0.6045</td>
</tr>
<tr>
<td>14</td>
<td>Prior_word_PGE</td>
<td>Y</td>
<td>Y</td>
<td>--</td>
<td>PGE</td>
<td>0.6138</td>
</tr>
<tr>
<td>15</td>
<td>Prior_root_PGE</td>
<td>Y</td>
<td>--</td>
<td>Y</td>
<td>PGE</td>
<td>0.5361</td>
</tr>
<tr>
<td>16</td>
<td>Prior_word_root_PGE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>PGE</td>
<td>0.6066</td>
</tr>
<tr>
<td>17</td>
<td>skip1_s1_span1</td>
<td></td>
<td>Baseline1</td>
<td>--</td>
<td></td>
<td>0.5354</td>
</tr>
<tr>
<td>18</td>
<td>skip2_s1_span1</td>
<td></td>
<td>Baseline2</td>
<td>--</td>
<td></td>
<td>0.5720</td>
</tr>
</tbody>
</table>
Some more to propose

- For syntax-based models
  - Specific decoding & training algorithm
  - Better parameterization methodology
- For structure alignment
  - More effective syntactic features
  - Discriminative training
Potential Contributions

- Formally define the syntactic TE from different granularities
- Explore the applicability and effectiveness of different TE to syntax-based models
- A series of structure alignment based model with varied modeling ability based on different grammars
- Comparative studies among these synchronous grammars to access their ability in describing parallel data
- An unsupervised framework for syntactic structure alignment
- Benefit syntactic structure alignment to MT
The End