Probabilistic Transfer-based MT

Yvette Graham, Josef van Genabith
National Centre for Language Technology
School of Computing
Dublin City University
Transfer-based Machine Translation

- Transfer-rule Induction Algorithm & Packed Rule Data Structure
- Probabilistic Decoder for Transfer-based MT
Why LFG F-structure?

- encode dependencies between the content words of a sentence
- encode the linguistic information necessary to generate the TL surface form sentence
- provide an abstract linguistic structure for forming linguistically motivated generalizations about how to translate well from one language to another
- abstracting away from surface form to f-structure produces more general rules that can be applied to multiple surface form examples – reducing data sparseness problems
- reliable and accurate linguistic resources for parsing and generation are vital to the overall success of a transfer-based system
X₀ speige(t) die Vielfalt X₁ wider. ➞ X₀ reflect(s) the diversity of X₁.
Transfer Rule Induction

- Exponential number of possible transfer rules within a single f-structure pair

- Lots of those rules are complete rubbish!

The Plan: Constrain rule induction to eliminate unwanted rules
Constraining Rule Induction

- Contiguity Constraint (Riezler & Maxwell 06)

- Cross-structural Consistency Constraint:
  Use a 1-1 set of alignments between nodes & allow any SL-TL node pair with equivalent non-empty sets of aligned descendents be a transfer-rule root.

**Aligned Descendent** of a SL local f-structure n:
Any descendent of n that is aligned with a TL f-structure. If n itself is aligned with any TL local f-structure, then n is considered an aligned descendent of itself.

Note: by this definition the actual root nodes of every f-structure training pair is a rule root.
Example

Er hat die Mehrheit bei der Abstimmung bekommen.

It received a majority vote.

Rule Roots:
A, A
B, B
Er hat X bei der Abstimmung bekommen. It received a X vote.

Mehrheit ↔ majority

Abstimmung ↔ vote

die Mehrheit ↔ majority

die Abstimmung ↔ a majority vote.
Putting Variables into Transfer Rules

Variable Constraints:

1. The root of a transfer rule may never be a variable.

2. For any non-root node:
   - Iff it is aligned with a TL node it can be a variable
Example F-structure Transfer Rules

a. spiegeln \rightarrow reflect
   X0   X1   X0   X1

b. spiegeln \rightarrow reflect
   Sprache X0   language X0

c. spiegeln \rightarrow reflect
   X0   Vielfalt X0   diversity
   die X1   det   subj
   subj adj-gen

d. spiegeln \rightarrow reflect
   Sprache Vielfalt language diversity
   die X1   det   subj
   subj adj-gen
   det   the
   subj

e. spiegeln \rightarrow reflect
   X0   Vielfalt X0   diversity
   die X1   det   subj
   subj adj-gen
   det   the
   subj
   det   of
   subj
   X1

f. spiegeln \rightarrow reflect
   Sprache Vielfalt language diversity
   die X1   det   subj
   subj adj-gen
   det   the
   subj
   det   of
   subj
   X1
Example F-structure Transfer Rules

- g. **spiegeln** → **reflect**
  - spiegeln
  - die Vielfalt
  - die europäisch

- h. **spiegeln** → **reflect**
  - Spärche
  - Vielfalt
  - die Union
europäisch

- i. **Sprache** → **language**

- j. **Vielfalt** → **diversity**
  - die X0

- k. **Vielfalt** → **diversity**
  - die X0

- l. **Vielfalt** → **diversity**
  - die Union
europäisch

- m. **Union** → **of**
  - die europäisch

- n. **Union** → **European Union**
Encode all possible rules that can be induced from a SL TL f-structure pair given these constraints in a single packed representation.
Example Unpacking Rule

- Unpack a rule by assigning true or false to certain context variables.

\[
\begin{array}{|c|c|c|}
\hline
\text{Variable Instantiations:} \\
A_0 = 0 & A_3 = 1 & A_6 = 0 \\
A_1 = 0 & A_4 = 0 & A_7 = 1 \\
A_2 = 1 & A_5 = 0 \\
\hline
\end{array}
\]

- To unpack all possible rules assign all possible combinations of true and false to the context variables.
Experimental Evaluation of Packed Data Structure

- Extracted all rules from 217,666 f-structure pairs of length 5-15 of German-English Europarl
- Unpacked 10 sets of rules – each set consisting of rules induced from randomly selected sets of 1000 training pairs, on average 23955 rules per set

<table>
<thead>
<tr>
<th>Set</th>
<th>No. Rules</th>
<th>Disk Space</th>
<th>Write Time</th>
<th>Load Time</th>
<th>Retrieving/Unpacking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Enum</td>
<td>Packed</td>
<td>Enum</td>
<td>Packed</td>
</tr>
<tr>
<td>1</td>
<td>24121</td>
<td>96.37M</td>
<td>7.2M</td>
<td>144s</td>
<td>128s</td>
</tr>
<tr>
<td>2</td>
<td>24486</td>
<td>98.89M</td>
<td>7.16M</td>
<td>145s</td>
<td>127s</td>
</tr>
<tr>
<td>3</td>
<td>23650</td>
<td>93.58M</td>
<td>7.17M</td>
<td>142s</td>
<td>133s</td>
</tr>
<tr>
<td>4</td>
<td>23882</td>
<td>96.83M</td>
<td>7.22M</td>
<td>149s</td>
<td>118s</td>
</tr>
<tr>
<td>5</td>
<td>24146</td>
<td>98.03M</td>
<td>7.15M</td>
<td>148s</td>
<td>128s</td>
</tr>
<tr>
<td>6</td>
<td>23355</td>
<td>91.75M</td>
<td>7.1M</td>
<td>140s</td>
<td>128s</td>
</tr>
<tr>
<td>7</td>
<td>23620</td>
<td>94.55M</td>
<td>7.21M</td>
<td>141s</td>
<td>142s</td>
</tr>
<tr>
<td>8</td>
<td>23687</td>
<td>94.02M</td>
<td>7.11M</td>
<td>137s</td>
<td>124s</td>
</tr>
<tr>
<td>9</td>
<td>23534</td>
<td>94.95M</td>
<td>7.12M</td>
<td>142s</td>
<td>120s</td>
</tr>
<tr>
<td>10</td>
<td>25069</td>
<td>100.66M</td>
<td>7.26M</td>
<td>152s</td>
<td>231s</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Average</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Rules Estimate</td>
<td>5214189</td>
<td><strong>20.4G</strong></td>
<td><strong>1.52G</strong></td>
<td>8h43m</td>
<td>8h20m</td>
</tr>
</tbody>
</table>
Chart-based Decoder

Input: a pre-compiled chart for the source language f-structure containing

Output: TL f-structure or n-best TL f-structures

- Top-down beam search of transfer chart
- Log-linear model to combine feature scores
- MERT to adjust weights (still to be integrated into decoder)
Future Work

• Fully integrate MERT into decoder
• Implement a faster search algorithm for decoding

• Factor atomic features and values to investigate what level of rule specificity is best
• Automatically learn where atomic features get their values – SL v TL