A Unified Framework for phrase-based, Hierarchical and Syntax SMT

Hieu Hoang
Philipp Koehn
Adam Lopez

University of Edinburgh
Decoding methods

• Phrase Based
  – Alignment Template System (Och 2004)
  – Pharaoh (Koehn 2003)
  – Moses (Koehn et al 2007)

• Hierarchical
  – Hiero (Chiang 2007)
  – ITG (Wu 1997)

• Syntactic
  – ISI (Yamada and Knight 2001)
  – SAMT (Zollmann 2006)
Phrase-based Decoding Pipeline

Preprocessing
- tokenizer
- tagging
- lemmatization

Alignment

Phrase extraction

Tuning

Decoding

Postprocessing
- recasing
- detokenizer

Scoring
- BLEU score
Hierarchical Decoding Pipeline

Preprocessing
- tokenizer
- tagging
- lemmatization

Alignment

Phrase extraction

Tuning

Chart Decoding

Hierarchical phrase extraction

Postprocessing
- recasing
- detokenizer

Chart decoder

Scoring
- BLEU score
Syntactic Decoding Pipeline

Preprocessing
- tokenizer
- tagging
- lemmatization

Alignment

Parse
Phrase extraction

Tuning

Chart Decoding

Postprocessing
- recasing
- detokenizer

Scoring
- BLEU score

Linguistic information

Syntactic phrase extraction

Chart decoder
Decoding

Preprocessing
- tokenizer
- tagging
- lemmatization

Alignment

Phrase extraction

Tuning

Decoding

Postprocessing
- recasing
- detokenizer

Scoring
- BLEU score
Phrase-Based

• Translate contiguous phrases
  
  assumes || geht davon aus, dass
  with regard to || bezüglich
  translation system ||
  Übersetzungssystem

• Finite state machine decoding
  – Stack based
  – Beam search
Hierarchical

• Discontiguous phrases
  
  $X \rightarrow$ take $X_1$ into account $||$ berücksichtigt $X_1$
  $X \rightarrow$ must explain $X_1$ $||$ muss $X_1$ erklären
  $X \rightarrow$ either $X_1$ or $X_2$ $||$ entweder $X_1$ oder $X_2$

• CKY+ decoding algorithm
  
  – chart decoding
  
  – simultaneous parsing and generation
Syntax

• Discontiguous phrases
• Labeled non-terminals
  
  \[ VP \rightarrow \text{take } NP_1 \text{ into account} \mid \mid \text{berücksichtigt } NP_1 \]
  \[ VP \rightarrow \text{must explain } NP_1 \mid \mid \text{muss } NP_1 \text{ erklären} \]
  \[ S \rightarrow \text{either } S_1 \text{ or } S_2 \mid \mid \text{entweder } S_1 \text{ oder } S_2 \]

• CKY+ decoding
Similarities

• Trained using aligned corpus
• Phrase tables
• Linear scoring
• N-best list for weight tuning
• Dynamic programming
• Language model context
Phrase-based decoder

• Base functionality
  – Incremental scoring
  – LM context
  – Dynamic programming
  – Search graph

• Decoding
  – Stacks
    • 1 stack for number of words covered
    • Future cost for better intra-stack comparison.
  – Search strategy
    • Bottom up, least number of words first
Hierarchical decoder

• Base functionality
  – Incremental scoring
  – LM context
  – Dynamic programming
  – Search graph

• Decoding
  – Stacks
    • 1 stack for each source contiguous coverage
  – Search strategy
    • Bottom up, smallest span first
Syntax decoder

• Base functionality
  – Incremental scoring
  – LM context
  – Dynamic programming
  – Search graph

• Decoding
  – Stacks
    • 1 stack for each non-terminal, for each source contiguous coverage
  – Search strategy
    • Bottom up, smallest span first
Inherited from the Moses decoder

- Factored word representation
- Multiple language models
- Multiple phrase tables
- Multiple implementations of LM and phrase tables
Phrase-based Decoding Pipeline

Preprocessing
- tokenizer
- tagging
- lemmatization

Alignment

Phrase extraction

Tuning

Decoding

Postprocessing
- recasing
- detokenizer

Scoring
- BLEU score
Phrase-based rule extraction

• Heuristic algorithms
  – Phrase extraction

• Probability estimates
  – Phrase and lexical probabilities
  – Smoothing

• Filtering
Hierarchical rule extraction

- Heuristic algorithms
  - Phrase extraction
  - Replace subphrases with non-terminals
- Probability estimates
  - Phrase and lexical probabilities
  - Smoothing
- Filtering
Hierarchical rule extraction

Alignment
hat das Haus gekauft
bought the house

Extracted Phrase
Hierarchical rule extraction

Alignment

hat      das     Haus gekauft
bought  the   house

Extracted Phrase

X → Ich hat das Haus gekauft || bought the house
Hierarchical rule extraction

Alignment

hat  das  Haus gekauft
bought  the  house

Extracted Phrase

\(X \rightarrow \text{hat das Haus gekauft} \mid \mid \text{bought the house}\)

\(X \rightarrow \text{das Haus} \mid \mid \text{the house}\)
Hierarchical rule extraction

Alignment

hat das Haus gekauft
bought the house

Extracted Phrase

X \rightarrow \text{hat das Haus gekauft} \mid\mid \text{bought the house}

X \rightarrow \text{das Haus} \mid\mid \text{the house}

X \rightarrow \text{hat X gekauft} \mid\mid \text{bought X}
Syntactic rule extraction

- Heuristic algorithms
  - Phrase extraction
  - Replace subphrases with non-terminals
    - New rule
  - Labeled non-terminals
    - Constrain extraction
    - merge non-terminal symbols - SAMT
    - Binarization
- Probability estimates
  - Phrase and lexical probabilities
  - Smoothing
- Filtering
Syntactic rule extraction

Alignment

hat      das     Haus gekauft
bought  the   house

Extracted Phrase

X → hat das Haus gekauft || bought the house
X → das Haus || the house
X → hat NP gekauft || bought NP
Results

German-English

- WMT09 new commentary corpus
  - 82k sentences
  - 1.8m German, 1.7m English words

<table>
<thead>
<tr>
<th>Model</th>
<th>Rule count</th>
<th>BLEU %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phrase-based</td>
<td>6.2m</td>
<td>13.0</td>
</tr>
<tr>
<td>Hierarchical</td>
<td>59.1m</td>
<td>12.9</td>
</tr>
<tr>
<td>Target syntax</td>
<td>2.2m</td>
<td>12.5</td>
</tr>
<tr>
<td>SAMT syntax</td>
<td>35.1m</td>
<td>12.9</td>
</tr>
</tbody>
</table>
Summary

• Extend Moses toolkit
  – Synchronous CFG formalism
    • Hierarchical
    • Syntactic decoding
  – Decoding algorithm
  – Rule extraction
• Re-use mature SMT pipeline
• Comparison of different decoding models
  – Use the same training data
  – Use the same translation & language models
• Merge different models