The SLE Example-Based Translation System

Pete Whitelock, Victor Poznanski
Sharp Laboratories of Europe Ltd.
Oxford
{pete, vp}@sharp.co.uk
Background

• 1970s – work by Sharp Corp. on E→J MT Duet, Power E/J, これ一本 (PC package)

• 1997-2003 – SLE’s Intelligent Dictionary おまかせ 訳振り (lightweight English analysis)

• 2003-2005 – SLE work on bilingual example retrieval and TM (~EBMT) functions for package

• 2005-2006 – SLE extend EBMT
  – Improve accuracy
    • include lightweight dependency analysis
  – Reduce resource utilisation for low power device
    • no language model
    • replace recursive use of TM with dictionary
Example

query → この 階 に コート が あります か
stored example → この 階 に 子供服 が あります か
Is this the floor for children’s clothes?
result → Is this the floor for coats?
Edit Model

alignment: \( s_i \leftrightarrow t_j \)

stored example

correspondence

\( edit_{src}: s_i \rightarrow q_k \)

replacement

\( edit_{tgt}: t_j \rightarrow trn(q_k) \)

translation

\( trn: q_k \rightarrow trn(q_k) \)
Alignment

- Sentence aligned bitext => (partial) alignment of (sets of) words
- Bilingual dictionary/thesaurus + dependency analysis
  - Constraint propagation
  - Use dependency proximity to resolve ambiguity
  - Annotate examples with semantic (thesaurus) codes of bilingual entries used
- Special routines
  - Number parser for English and Japanese
  - Readings of kanji proper names as found by Japanese morphological analyser
  - Katakana to English edit rules +dp spell check
Please change US dollars into Australian dollars.
Correspondence

- Example base is indexed by words, lemmas, semantic codes
- Retrieve candidate examples which share index terms with query
  - don’t prefer rare terms, so obtain templatic effect.
- Compute sequence of matched and unmatched stretches
  - determine score based on these
  - adjust score by semantic overlap in unmatched stretches
- Use example with best score as basis for translation
この階にコートがありますか
Is this the floor for children’s clothes?

この階に子供服がありますか
Is this the floor for children’s clothes?

この階にレストランがありますか
Is there a restaurant on this floor?
Translation

• Dependency-analyse query
• Find set of bilingual entries matching query
• Find maximum consistent (tiling) set
  – prefer entries covering more items
  – prefer entries with semantic codes consistent with those determined in correspondance
  – prefer entry used most frequently in corpus
• Pass the set to replacement phase
彼 は 5 時 に 戻り ます 。
彼 は 月曜日 に 戻り ます 。
He’ll be back on Monday.
He’ll be back at 5 o’clock
Replacement

- Determine positions in basis to be substituted
  - by examination of alignment relation
- Apply TL edit
  - deletion
    - delete words at positions in basis
    - delete determiners and prepositions when their head is deleted
  - substitution
    - substitute head position in basis by translations of words in query
    - delete words at all other substitution positions in basis
  - insert
    - insertion is adnominal – add head noun to both sides of source edit
      - assimilate insertion to substitution
    - insertion is adverbial – add translation to start or end of sentence
  - handle cases of boundary friction
    - a/an alternation, removal of multiple prepositions
    - apply number agreement
    - generate inflected comparatives and superlatives
    - copy features from basis to output
Will this flight arrive on time?
Will this 出発 (電車) 出発  on time?
Will this train  出発 on time?
Will this train  出発 on time?
## Results

<table>
<thead>
<tr>
<th></th>
<th>BBS</th>
<th>SLE Example Base (~13,000)</th>
<th>SLE+IWSLT Example Base (~53,000)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>EBMT Only</td>
<td>%age</td>
</tr>
<tr>
<td>test (asr 1best)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>test (correct)</td>
<td>.1797 [5.4599]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>