A Pattern-Based Machine Translation System
— Yakushite Net MT Engine

Miki Sasaki and Toshiki Murata

Oki Electric Industry Co., Ltd.
2-5-7 Hommachi, Chuo-ku, Osaka 541-0053, JAPAN
{sasaki234, murata656}@oki.com
Machine Translation by OKI inc.

- Rule-based MT → Pattern-based MT
  - Rule-based MT (PENSEE 1980s ~ 1990s)
  - Pattern-based MT (implemented with Java 1997 ~)
  - Collaborative translation environment
    (Yakusite Net 2001 ~)

- Pattern-based MT method
  - All the knowledge needed for translation are treated as **translation patterns**
  - Grammars and word dictionaries can be registered in the same way to our system because they are both treated as **translation patterns**
Yakushite Net

- Pattern-based MT
- Collaborative translation environment
  - Users collaborate to improve the translation accuracy

To improve the translation accuracy;
  - Our system has various communities
  - Each community has a dictionary
  - Users register dictionary data to dictionaries of relevant communities
Structure of Communities

tree structure

Root

General dictionary

science
technology
computer
hobby

... ... ...

science dictionary
technology dictionary
computer dictionary

electronics

hardware
technology dictionary

software
electronics dictionary

programming

java
computer dictionary

perl
dictionary

perl dictionary

electronics dictionary

programming dictionary
Structure of Communities

tree structure

- Root
- General dictionary
- science
  - science dictionary
- technology
  - technology dictionary
- computer
  - computer dictionary
  - hardware
    - hardware dictionary
- electronics
  - electronics dictionary
- programming
  - programming dictionary
- java
  - java dictionary
- perl
  - perl dictionary
- hobby
  - hobby dictionary

...
Technologies in Yakushite Net

- Automatic dictionary acquisition
- Determination of dictionaries, texts and communities
- Multilingual processing
The sentence is parsed using the translation patterns in the dictionaries.
Translation Patterns

- Rules of Context-free Grammar (CFG) are paired
  - CFG is a formal grammar in which every production rule is of the form “V -> w”
  - Examples of CFG rules
    Japanese : S -> Sintr
    English : S -> Sintr ?
  - Examples of translation patterns
    [ja:S [1:SIntr:*] ]
    [en:S [1:SIntr:*] ?:pos=punc];
    - The mandatory numerical index allows elements between source and target patterns to be related

- Source language patterns are used for analysis.
  (In Japanese-English translation, “ja” is source language and “en” is target language)
Parsing and generating method

Source

Target
Parsing and generating method

Source

\[ S \]

Target

\[ S \]
Parsing and generating method

Source

S

VP か

VP 行く

Target

S
Parsing and generating method

Source

Target
Parsing and generating method

Source

Target
Parsing and generating method

Source

S

Target

S
Parsing and generating method

Source

Target
Parsing and generating method

Source

Target

S

S
Parsing and generating method

Source

Target

S

S
Parsing and generating method

Source

Target
Parsing and generating method

- Word sequences are reduced to a root of a parse tree ("S") by applying patterns.
- When word sequences reach "S", the source parse tree is completed:
  - each node using the corresponding target language pattern is converted.
- Generation of the target parse tree is carried out immediately after the parse tree is completed.
Priority Control of Translation

- A parsing tree
  - prioritized by the combination of criteria (ex. number of selected patterns)
- A translation pattern
  - prioritized with an priority control mark
- Failure Recovery Dictionary
  - becomes active only when the normal parsing process failed
The Results for IWSLT2005

- Description of the planned training methods
- Results
  - Performance for training data
  - Result for test data
- Examples of registered translation pattern and translation results
Description of the Planned Training Methods

- Not cover much of expressions seen in BTEC
- We manually made translation patterns that are highly generalized
  1. we manually extracted **frequently used expressions** in the IWSLT05 training corpus
  2. we **patternized** those expressions and gave them appropriate translations
  3. we made **corrections** to the existing patterns
  4. we **registered** the new patterns to our system
(1) Before registering new patterns
(2) After registering them
(3) After we extracted the parallel texts with one Japanese sentence from IWSLT05 training corpus and IWSLT04 test corpus, and registered them

<table>
<thead>
<tr>
<th></th>
<th>BLEU</th>
<th>NIST</th>
<th>WER</th>
<th>PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.1918</td>
<td>6.2283</td>
<td>0.6470</td>
<td>0.5640</td>
</tr>
<tr>
<td>(2)</td>
<td>0.2179</td>
<td>6.7882</td>
<td>0.5989</td>
<td>0.5183</td>
</tr>
<tr>
<td>(3)</td>
<td>0.7616</td>
<td>12.5216</td>
<td>0.2216</td>
<td>0.1894</td>
</tr>
</tbody>
</table>
(1) Before we registered the new patterns  
(2) After we registered the new patterns  
(3) After we extracted the parallel texts with one  
Japanese sentence from IWSLT05 training corpus  
and IWSLT04 test corpus, and registered them

<table>
<thead>
<tr>
<th></th>
<th>BLEU</th>
<th>NIST</th>
<th>WER</th>
<th>PER</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>0.1918</td>
<td>6.3279</td>
<td>0.6749</td>
<td>0.5624</td>
</tr>
<tr>
<td>(2)</td>
<td>0.2222</td>
<td>6.8913</td>
<td>0.6314</td>
<td>0.5258</td>
</tr>
<tr>
<td>(3)</td>
<td>0.2639</td>
<td>7.3585</td>
<td>0.6066</td>
<td>0.5065</td>
</tr>
</tbody>
</table>
Examples of Registered Translation Pattern and Translation Results(1/2)

IWSLT05_JE_training:
Japanese: ボール (booru) を (wo) よく (yoku) 見 (mi) て (te)。
Translation result (1): You see a ball well and.
English: Watch your ball carefully.
Japanese: つかまえ (tsukamae) て (te)。
Translation result (1): It catches it and.
English: Catch him.

Extracted expression:
-te form of verbs (conjugated form that leads declinable words) + particle "te(て)" or "de(で)" make imperatives.
Registered translation pattern:
[en:SImp [1:VP:*:conjug=bare] ];

IWSLT05_JE_TESTSET:
Japanese: 警察 (keisatsu) を (wo) 呼ん (yon) で (de)。
Translation result (1): It calls police and.
Translation result (3): Call police.
Japanese: 芝生 (shibahu) に (ni) 入ら (haira) ない (nai) で (de)。
translation result (1): It does not enter a lawn and.
translation result (3): Do not enter the lawn.
Conclusion

- We presented our pattern-based MT method
  - Enables easier registration of phrasal expressions and grammatical knowledge
- We described how we dealt with the task
  - We dealt with the task mainly manually

- Future study
  - Adoption of an automatic dictionary acquisition technology
Example of Translation (1/3)

Japanese: 「彼はどこに行くか」
English: “Where does he go?”

[ja:VP 行く:*:pos=ds]
[en:VP go:*:pos=v];

[ja:VP:jSentenceType=interrogative か :pos=ej]
[en:VP [1:VP:*]];
Example of Translation (2/3)

Japanese: 「彼はどこに行くか」
English: “Where does he go?”

[ja:NP:personNum=3sg 彼:*:pos=ms]
[en:NP he:*:pos=prn];

[en:NP [2:NP:*:case=subj] ];

[ja:FsIntr どこに :*:pos=fs]
[en:AdvIntr where:*:pos=adv];
Example of Translation (3/3)

Japanese: 「彼はどこに行くか」

English: “Where does he go?”


[ja:S [1:SIntr:*] ]
[en:S [1:SIntr:*] ?:pos=punc];