

## 4B-5

日英翻訳システムにおけるポスト  
エディット情報のフィードバック

土肥実久, 高松忍, 西田富士夫

(大阪府立大学 工学部)

## 1. まえがき

本稿では, 日英翻訳システムを対象とし, ポストエディタの修正情報のフィードバックによりシステムのもつ辞書やルールを半自動的に修正する手法について述べる. 修正されたターゲット文を英日翻訳システムにより逆方向に翻訳し, 得られた中間表現を日英翻訳システムの中間表現と比較して誤りパターンを同定する. 続いて, 誤りパターンに対する修正手順によりシステムの辞書やルールを修正する.

## 2. システムの概要

本手法では, 日英翻訳システム MTo が出力したターゲット文 To とポストエディタによる修正文 Tr から誤りパターンを求めるため, MTo と逆方向の英日翻訳システム MTr を用いる. ここで, 両システム MTo と MTr は, 格構造表現を中間表現とするトランスファ方式とし, 格体系と意味カテゴリ体系は同じものとする.

ポストエディタは, ターゲット文の各語に付けられた番号によって修正箇所を指示する. 例えば, コマンド:

```
replace n1 (or n1-n2) by  
"a new word sequence"
```

は, 番号 n<sub>1</sub> の語 (又は番号 n<sub>1</sub> から番号 n<sub>2</sub> までの語列) の "a new word sequence" による置換を指示する. 挿入, 削除, 移動についても同様なコマンドにより指示する.

修正されたターゲット文 Tr に対し, 次の手順で日英翻訳システム MTo の辞

書とルールを修正する.

- (1) Tr を英日翻訳システム MTr により解析し, Tr の中間表現 TTr を作る.
- (2) MTo による元のターゲット文 To の中間表現 TTo, ポストエディタによる修正情報, MTr による Tr の中間表現 TTr や解析に用いた書換え規則などから, 誤りパターンを同定する.
- (3) 誤りパターンに対して設けた修正手順により, 該当する辞書とルールの誤りの箇所を修正する.

以下に, 誤りパターンの同定とその修正について述べる.

## 3. 生成誤りの修正

構文生成の誤りは, ターゲット文 To と修正文 Tr の中間表現 TTo と TTr が構文上の表現の違いを除いて一致する場合に同定される. このとき, ポストエディタの修正情報と Tr の解析に用いた書換え規則などから生成辞書の誤りを修正する.

## 例 1

So : "プリズムには光線を屈折する能力がある."

```
TTo : (PRED:have, POSSESSOR:prism(...),  
      OBJ:ability(PRED:refract,  
                  OBJ:ray(...), INSTR:*) )
```

To : "A prism has the ability  
which refracts the ray."

Tr : "A prism has the ability  
to refract the ray."

Tr の解析によりえられる中間表現 TTr は TTo と同じである. また, Tr の修正箇所に用いた書換え規則は,

RR:ADJP2

<NP( $n_1$ (PRED:v, K1:\*, K2: $n_2$ , ...))>

::= <NP1( $n_1$ )>

<INF(PRED:v, K1: $n_1$ , K2: $n_2$ , ...)>

である。これより、次のように辞書を修正する。

HEADWORD:能力

EQUIV:ability, ...

MODifier<INF>, RR:ADJP2

#### 4. 構造変換誤りの修正

構造変換の誤りは、中間表現  $T_{10}$  と  $T_{1r}$  の格構造が一致しない場合に同定される。このとき、 $T_{10}$ を $T_{1r}$ に変換する一般的な構造変換規則を求め、規則に指定するように変換辞書を修正する。

##### 例 2

$T_{10}$ と $T_{1r}$ の比較により、例えば、原因の無生物を動作主体に変換する一般的な構造変換規則:

STR:CR1

(PRED: $v_1$ , OBJ: $n_1$ , CAUSE: $n_2$ )

→ (PRED: $v_2$ , AG: $n_2$ , OBJ: $n_1$ )

が求まったとする。ここで、 $v_1$ ,  $v_2$ は "obtain", "provide" のような因果関係にある動詞の組を示す。このとき、以下の下線のように辞書を修正する。

HEADWORD:得る

EQUIV:obtain, ...

CAUSE-V:provide, STR:CR1

#### 5. 語彙変換の誤りの修正

語彙変換の誤りは、中間表現  $T_{10}$  と  $T_{1r}$  の格構造が一致するが語彙が異なる場合に同定される。このとき、修正前の訳語と修正後の訳語の選定条件を修正する。

ソース言語の単語  $t_1$  の訳語  $t'_{11}$ ,  $t'_{12}$  の選定条件が次のように辞書に与えられているとする。

HEADWORD: $t_1$

EQUIV: $t'_{11}$  COND:( $K_1-C_1:t_1$ ,  $K_2-C_{21}:$ , ...)

EQUIV: $t'_{12}$  COND:( $K_1-C_1:t_1$ ,  $K_2-C_{22}:$ , ...)

ここで、 $K_i-C_i$  は格ラベルと意味カテゴリの組を表す。いま、ソース文の中

間表現:

$S_{10}$ : ( $K_1-C_1:t_1$ ,  $K_2-C_{23}:t_{23}$ , ...)

がえられ、 $C_{23}$ が $C_{21}$ のサブカテゴリであるとする。また、 $S_{10}$ から上の辞書により語彙変換して生成したターゲット文  $T_0$  に対し、ポストエディタが訳語  $t'_{11}$  を  $t'_{12}$  に修正したとする。

このとき、 $t_1$  の辞書における訳語選定条件を以下の下線部のように修正する。

HEADWORD: $t_1$

EQUIV: $t'_{11}$  COND:( $K_1-C_1:t_1$ ,  $K_2-\{C_{21}-C_{23}\}$ , ...)

EQUIV: $t'_{12}$  COND:( $K_1-C_1:t_1$ ,  $K_2-\{C_{22}+C_{23}\}$ , ...)

ここで、 $\{C_i \pm C_j\}$  は  $C_i$  と  $C_j$  の和 (差) を表わす。

#### 6. イディオム変換の誤り修正

始めに、ポストエディタの指示により修正されたターゲット文におけるイディオム表現の箇所を同定する。次に、そのイディオム表現に対応するソース文のイディオム表現: " $\alpha$  t  $\beta$ " とキーワード  $t$  を求める。そして、キーワード  $t$  を見出しとする辞書のイディオム欄に、" $\alpha$  t  $\beta$ ", 対応するターゲット言語の表現とそれらの格構造表現を登録する。

##### 例 3

HEADWORD:影響

IDIOM:に影響を与える

(PRED-POSSTRANS:与える,

OBJ:影響, GO(に):\_)

EQUIV:affect

(PRED-POSSTRANS:affect,

OBJ:\_)

#### 参考文献

- 1)V. Lawson(ed.): Practical Experience of Machine Translation, North-Holland Publishing Company(1982).
- 2)西田・高松:機械翻訳における中間表現の変換と文の合成手法, 情報処理学会誌, 26, 10, p. 1165-(1985).
- 3)F. Nishida et al: Feedback of Correcting Information in Postediting to a Machine Translation System, COLING-88(1988).