3E-4

自然な文章生成における規範
－機械翻訳への応用－

吉村裕美子, 平川秀樹, 天野真家
(株)東芝 総合研究所

1. はじめに

文法体系を異にする二言語間の翻訳において、第一言語の文章にある意味を最大限保持した訳文を得る鍵となるものの一つに、語順操作の問題がある。これに対する対処なくしては、意味は分かるが不自然な文、誤解を招く文、不可解な文等の生成を回避できない。本稿では、我々が開発した機械翻訳システムで用いている語順操作の方法について述べる。

2. 語順に作用する要因

第二言語の文法的枠組が語順操作の根本にある。そこで、以降、日本語に比べ、語順的拘束の多い英語を第二言語とした時をベースに考える。

まず、概念構造のレベルで、アーク名等により条件付け可能なものに次の3項目がある。

(a) 動詞・形容詞が強く結合してある概念を構成する副詞句は、動詞・形容詞に近い位置に生成する。
ex) He is independent of his parents ---.
(b) 時・場所を表す副詞句は、<場所・時>の順で通常文尾(時に文頭)に生成する。
(c) 一つの原語から分離独立した語は近い位置を好む。
ex)「駆使する」 use the tool freely ---.

これらの操作を前提にして、さらに次に挙げる要素がある。

(d) 個々の訳語が好む生起位置
(e) 各修飾句の持つスコープ
(f) 各句の重さ(生成後の語列の長さ)

(d)~(f)に対する処理の方法については、次の節で詳説する。

3. 生成過程での語順操作処理

3.1. 個々の訳語が好む生起位置

副詞・形容詞は、個々の英単語自体が、好む生起位置を持っている。同じ日本語に当てられる訳語にも、好む生起位置の異なるものが混在することがある。これに対しては、個々の訳語に当てる品詞を細分することで解決している。品詞の細分においては、意味的側面・生起位置からの側面・形態的側面のそれぞれの見方を取り混ぜているが、語順操作の効果が最大になるように品詞の割り当てを行う。

次に挙げるのは、副詞の細分の例である。番号数が増すにつれて優先度が下がる。

(a) 通常主語と動詞句の間に生成する。
(b) 通常文頭に生成する。
(c) 頻度を表す。
(d) 時を表す。
(e) 場所を表す。
(f) 動詞句から形成される。
(g) 前置詞句から形成されるか、副詞に後置就修飾句が付属している。
(h) その他。

この様な情報を各訳語の品詞に持たせることにより、第二言語のみの持つ特性を生成過程という閉じた中で操作することができ、かつユーザの訳語選択にも影響を受けない。

3.2. 各修飾句の持つスコープ

一つの文の中で、各々の修飾句は、その生起位置により修飾のスコープを異にしている。翻訳において、このスコーピングの状態を最大限生かすことが、望ましい訳文を得る大きな手助けとなる。日本語では、常に修飾句が被修飾句に先行する。一方、英語には前置修飾と後置修飾があり、各々被修飾句の外側に位置する。すなわち、両言語とも、被修飾句の核部分から遠ざかるほどスコープが広いと言える。そこで、入力文中の各語句の生起位置情報を利用した生成処理が有効であると考えられる。

Criteria of Natural Sentence Generation
－ Application to Machine Translation －
Yumiko YOSHIMURA, Hideki HIRAKAWA, and Shin-ya AMANO
TOSHIBA Corporation

概念構造のレベルでは、各サブ tree が将来どんな句としてどんな生起位置コントロールを受けるべきか、十分に予測されない。そこで、生成ルール中の任意の箇所で、入力文中の語順により、生成順を操作する記述ができる枠組みを考案した。

次に挙げるのは、一文中で、前置修飾するものについては入力文中の語順の先のものから後のものへ、後置修飾するものについては後のものから先のものへと、順に生成するためのルールの表記モデルである。

```
S = $PREADV,SUBJ,*,OBJ,$POSTADV;
$PREADV =  X,
           ADV{前置修飾条件},
           ADV{前置修飾条件},
           ADV{前置修飾条件};
$POSTADV = Y,ADV,ADV,ADV;
X = ;
Y = ;
```

X・Yは、形態上、SUBJ・OBJ等と同じく概念構造中のアーク名の形をとっているが、これらはプログラム中の予約項目である。"X"の記述により、以下の生成処理は、同じアーク名を持つサブtreeについては入力文中の語順の先のものから後のものへ、逆に、"Y"の記述により、語順の後のものから先のものへと生成を進める。

この枠組みにより、概念構造上では差を持たずに並列する要素に対する語順操作が容易に可能である。

3.3.各句の重さ

これまでに挙げた操作に加えて、さらに綿密性を求めるには、各サブ tree の重さ(生成後の語列の長さ)をも操作の条件に組み込む必要がある。英語では、end-weight の原理により、重い句・複雑な句は文尾の位置が好まれる。また、新情報は重い構造を持つことが多く、end-weight の原理が end-focus の原理と協同して句の外置を促す。(1)のような目的語と的格補語の語順の転換もこの原理から導かれる。

(1) He had called an idiot the man
on whose judgement he now had to rely.

一般に、関係節等の長い複雑な修飾句をサブ tree に持つ重い句を文尾以外に置くと、修飾のスコープの問題が絡み、その後に位置する句とのシンタクス上の関係を曖昧にしがちであり、慎重に対処する必要がある。

この問題に対しては、各ルール中の条件記述部に、任意のサブ tree のノード数を参照する記述をゆるすことにより、解決に大きく近付いた。次に、(1)を生成するためのルールの記述例を挙げる。

```
---, OBJ{pw<8 | _! REL{pw>6}.&
                _! NPP{pw>6}.&---},
     COMP,
     OBJ, ---
```

この例では、ノード数8 以上の目的語、あるいは、ノード数7 以上の関係節・前置詞句を従える目的語は目的格補語の後に生成される。厳密には、ノードの数ではなく、一つの訳語が複数の単語から構成されていれば、その単語数もカウントに加える。

この操作にはまだ課題が残されている。第一に、何単語以上であれば重い句であると言えるかが非常に曖昧である。次に、ルール中に数を指定する際には、実際には、pw より得られる値のほかに、生成過程で、独自に生成する語句、また、故意に未生成とするノードのあることを意識しなくてはならない。だが、その正確な測定はルール記述の段階では無理である。しかし、こういった問題も、条件部の記述をさらに詳細化していくことで、大半は克服されるものである。

4. おわりに

精度の高い訳文を得るための語順操作の方法として、異なった方向からの3つのアプローチについて述べた。これらは全て生成過程という閉じた中で操作を行うことを目的としている。

本方式は現に我々の機械翻訳システムに組み込まれ、良い成果を見せている。今後は、条件記述を詳細化することにより、さらに操作を精密にしていくつもりである。