

4S-9

自然な日本語文を生成するための語順

野崎 康夫 湯村 武 西田 行輝
三洋電機(株) 情報通信システム研究所

1. はじめに

最近、機械翻訳などの自然言語処理システムにおいて、自然な文を生成することがますます重要になってきている。日本語生成処理においては、助詞の適切な選択、冗長表現の簡素化、語順の最適化などに負うところが大きい。本稿では自然な日本語を生成するための1つのアプローチとして語順について考察した。

従来、日本語の語順は比較的自由であり、深層構造が求まれば表層の日本語文を生成するのは容易であると考えられてきた。しかし、複数の修飾要素をむやみに並べていたのでは、出力文に自然さを欠くばかりでなく、原意の忠実な伝達ができなかったり、誤解釈をうける文が生成される可能性がある。

ここではまず、特に用言を修飾する要素間における、より自然な語順の決定方法について考察し、次に、これらの方法によって語順決定した後誤解釈の恐れのある用言修飾句の並びを検出するための条件と、それを補正する方法について述べ、両者を組み合わせることにより効果的な語順決定処理が行えることを示す。

2. 基本的な語順決定規則

文の意味内容をより正確に伝達し、読みやすく自然な文とするためには、最適な語順を決定する必要がある

ここでは、統語的、意味的な様々なレベルの情報から最適な語順を得る規則を提示する。

2-1. 表層情報に基づく語順決定規則

格助詞などの表層情報を用いて決定する。

[r1.1] ~は ⇒ ~で ⇒ ~に ⇒ ~を ⇒ 用言

*彼は電話で友人に礼を言った。

2-2. 深層情報に基づく語順決定規則

2-1は簡便な方法であるが、同じ格助詞でも用言に対する意味的役割が違えば、語順が異なる場合がある。

*彼は息子に立派な医者を紹介した。

*彼は息子を立派な医者に育てた。

そこで、用言に対する意味的役割(深層格)により、語

順を決定することが有効となる。[[1]]

[r1.2] 主題 ⇒ 時 ⇒ 場所(開始点 ⇒ 到達点)
⇒ 相手 ⇒ 対象 ⇒ 対象の内容・様相・結果
(補語) ⇒ 用言

*社長はあす講堂で優秀な社員に表彰状を贈ることにしている。

深層格は意味素性などをもとにした静的な属性であり、これにより決定された語順は、一般的には(一文単位で判断するかぎり)違和感のない自然なものとなる。

しかし、語順は状況により動的に変化するものであり、そのような現象を扱うには、より深い意味的あるいは文脈的な情報が必要である。次に、このような情報を利用した語順決定規則について述べる。

2-3. 情報の新旧、着目度を考慮した語順決定規則

日本語の特色として、用言修飾句は既知の古い情報から新しい情報へと並び、用言のそばにはもっとも新しい情報が配置されるという傾向がある。[[1][2]]これを利用した語順決定規則をいくつか示す。

新旧情報としては、主題(~は)と着目対象(~が)、定(既知)/不定(未知)、などが利用できる。

[r1.3] 主題 ⇒ 時 ⇒ 場所 ⇒ 着目対象

*きのう公園を(一匹の)犬が走っていた。

* (その)犬はおととい川岸を走っていた。

[r1.4] 定(既知) ⇒ 不定(未知)

*村人は(その)教授を(とある)山小屋へ案内した。

*村人は(その)山小屋へ(ひとりの)教授を招待した。

これらの文において、生成文の冗長さを軽減するという立場から、()内の語句は省略されることも多く、そのような場合、既知性・着目度といった概念を表現するために、語順を適切に配置することが有効となる。

2-4. 修飾句の長さによる語順決定規則

長い修飾句や従属節が文のなかほどに配置されると、読みにくく文意が通りにくい文章となる。そのため、これらは文頭へ配置するのが望ましい。

*拡張性があり適応性がある自然言語システムの発達は概念が効果的に表現される知識表現の枠組みに依存すると本論文は提案する。

ところが、ある程度長い要素であっても文頭に置かないほうがよい場合もある。例えば、

- a. 主題など広範囲にかかる要素の前には置かないほうが自然である
*彼は生まれつき話が上手であったのでこの仕事に向いていた。
- b. 長い要素が必須格である場合には[r1.2]の規則で定まる位置のほうが望ましい
*彼はきのう会議室で最近の企画には新鮮味がないということを主張した。
- 以上をまとめると次のような規則が追加される。

[r1.5] 相当長い修飾句、或いは主格を含む従属節
⇒ 主題 ⇒ 主格を含まない長めの従属節
⇒ その他の格要素(長めの必須格要素も含む)
⇒ 用言

3. 誤解釈回避のための規則

これまでに挙げた語順決定規則に従って生成された文はかなり自然で読みやすいものとなるが、修飾句の配置によっては生成文の解釈に曖昧さが生じ、誤解釈を起し得る場合がある。ここでは、その一例を示し、誤解釈を起し得ない文へ修正することを試みる。

- (1) 知らせに来(た)
├ (2) 彼(は) [主題]
├ (3) きのう [時]
└ (4) 退院し(た) (こと) (を) [対象]
 └ (5) 奥さん(が) [主体]
- 図1 誤解釈を生じる構造の例

図1に示した深層構造にこれまでの語順決定規則を適用すると、

*彼はきのう奥さんが退院したことを知らせに来た。という文が生成される。ところがこの場合、「きのう」の係り先として「退院した」と「知らせに来た」の2通り考えられる。

このような場合、

- a. 係り先の曖昧な要素の直後へ読点をうつ
⇒彼はきのう、奥さんが退院したことを知らせに来た。
- b. 本来の係り先の直前へ移動する
⇒彼は奥さんが退院したことをきのう知らせに来た。といった回避措置を講じることにより、生成文の意味的曖昧さを軽減できる。

[r1.6] 用言P1を修飾する格要素X2と従属節または埋め込み文S3が連続して配置され、S3中にX2と同

じ格要素がない場合、X2の直後に読点をうつ。またはX2をP1の直前に置き換える。

直前に置かれた格要素が埋め込み文中の用言に係り得る条件を検出するには、時制・アスペクトなどの一致を調べればよいが、厳密には主用言との相対的な時制を考えなければならない。次の文では、「先週」が「帰国するだろう」に係り得る。

*私は先週必ず彼が帰国するだろうと思っていた。

このように係り受け可能条件の厳密化が難しければ、この条件を緩和してもよい。この結果、曖昧さが生じ得ない文でも回避措置がとられることがある。

*彼はあした奥さんが退院したことを祝うつもりだ。
(「あした」が「退院した」に係る解釈は生じ得ない。)

⇒彼はあした、奥さんが退院したことを祝うつもりだ。しかし、これにより意味の通りが悪くなることはない。

したがって、この規則を用いれば意味の曖昧さの少ない日本語を生成することが比較的容易にできる。

4. 語順決定の指針

以上の考察をもとに、我々は英日機械翻訳の日本語生成処理において、次のような方針で語順の決定規則を構築した。

- (1) 意味素性などの静的な情報をもとに、深層格による語順決定を行なう。
- (2) 構文、意味、文脈から得られる「情報の新旧」を考慮して、既知情報から新情報へと並び換えを行なう。
- (3) 誤解釈の回避措置を講じる。

この結果、従来に比べてより自然でかつ曖昧さの少ない日本語文が得られるようになった。

5. おわりに

日本語の語順の決定をテーマとして、「自然な文の生成」と「正確な意味の伝達」について考察した。日本語の語順決定のための法則のなかには、情報の新旧、スコープなど文の伝達内容に本質的にかかわるものと、位置関係より生じる不都合(係り受け関係の曖昧さの発生など)を回避するための便宜的なものがあることがわかった。

英日機械翻訳において、より精度の高い訳文を生成しようとした際、この2つを明確に分離することにより、原文の意味を損なわない範囲で、より自然な文に仕上げることが可能である。

今後は、情報の新旧など、文脈・状況に関わる属性を精度よく抽出する方法についての研究をすすめ、より自然な文の生成を目指す。

【参考文献】

- [1]長尾 真 「言語工学」 昭光堂
[2]電子通信学会編 「日本語情報処理」