

日英機械翻訳システムの語彙規則による自然な訳文の生成

3E-5

木村 真理子, 平川 秀樹, 天野 真家
(株)東芝 総合研究所

1.はじめに

日本語と英語を自然な出力を得ることを前提として、動詞文型について比較する時、二カ国語間の表層格構造の違いが指摘される。高品質な翻訳結果を得る為、機械翻訳システムには、この違いを吸収できる文法が必要である。我々が開発中の日英機械翻訳システムは、英日機械翻訳システム[1]と同様に、一般文法(純粋に統語的な解析を行う部門であり、拡張遷移ネットワーク文法(Augmented Transition Network文法)を用いている)と語彙規則から成る語彙遷移ネットワーク文法(Lexical Transition Network文法)を用いることにより、語彙に依存した意味処理を行う。本稿では、語彙規則の制御について、日本語と英語の表現の差異への対処を中心に述べる。

2.語彙規則の役割

語彙規則は、語の意味や性質を各語について記述するものである。その役割は、1)一般文法によって構文解析された統語構造内に残された曖昧性(係り受けや意味解釈に関する)を解消し、構文的構造を日本語概念依存構造に変換すること、2)その日本語概念依存構造を更に英語概念依存構造へ変換することである。

実際の処理プロセスでは、まず、入力文に対して構文解析が行われ、係り受け可能なノード間に可能仮想アークが付加される。次にそれらに対して、種々の preference に基づき、優先順位が決定される。ここでいう preference とは、動詞の格パターン、共起関係などの意味情報、語の位置情報などから得られるヒューリスティックな優先度である。そして、制約条件(深層格の多重占有禁止、係り受けの交差禁止)を満たし、かつ最高優先得点のアークが選択され、日本語概念依存構造が決定される。最後にその結果は、各語の訳語決定と格パターンの変換、および必須格の補充などの処理を経て英語概念依存構造に変換される。([2][3]参照)

3.語彙規則による同一概念の表層上の多様性の処理

日本語の格助詞の様な表層格と、深層格との対応

や、格を埋める要素の意味的な制限などは個々の動詞によって異なる。それらに関する情報は、格パターンとして語彙規則に記述される。

この格パターンは、基本的には、表層格と深層格との対応を表現しており、表層格助詞の多様性(学校へ行く=学校に行く)の解釈などが取扱われる。格フレーム中の深層格(agent, locativeなど)は、日英語彙変換ルールと対応付けられており、そこで日英対応関係が処理される。

以下、日本語概念依存構造から英語概念依存構造への変換処理の中から、二ヶ国語間の表層格構造の違いが顕著に現われる現象を列挙し、語彙規則の制御機能を説明する。

i) 原言語の表現の型式を目的言語の構造に反映させる

原言語の表現において多様性がある時、目的言語においても原言語の表層構造を保存した方が良い場合がある。

例)「充填する」

酸素がシリンダーに充填した。
-> Oxygen filled the cylinder.
シリンダーが酸素で充填した。
-> The cylinder is filled with oxygen.

動詞、「充填する」において、上の文で「が」格を埋める要素は、下の文で「で」格を埋める要素と同じ深層格に対応する。しかし、後者の日本語文においては、「が」格を用いることにより、「シリンダー」に焦点が当てられている。その様な表現を訳文中に活かす為、語彙規則では、表層格助詞の組み合わせを考慮して構文を決定する。

ii) 原言語において表層表現に本質的な多様性がない場合で、目的言語の表層構造に多様性がある時、より自然な候補を選択する

例)「広げる」

反射熱がコーティングを表面に広げる。
-> Reflected heat spreads coating on the surface.
-> Reflected heat spreads the surface with coating.

Generation of Natural English Output Realised by Lexical Rules for Japanese to English Machine Translation System

Mariko KIMURA, Hideki HIRAKAWA, Shin-ya AMANO
Toshiba Research and Development Center

この場合、どちらの構文も文法的には正しいが、上の方がより英語らしい表現である為、語彙規則はこの動詞において前者の表現を採用する。但し、それ程両構造の違いに差がない場合は、語彙規則が英語の構造を決定することはない。

例)「与える」

爆発は原子にエネルギーを与えた。
->The explosion gave the atom energy.
->The explosion gave energy to the atom.

上の例では、(強いていえば「原子」か「エネルギー」のどちらをより強調することになるか以外は)両構文は全く同義である。この場合、動詞の格関係に対応する構文の選択は、語彙的な意味ではなく、むしろ英文の理解容易性という概念に基づいて処理される。([4]参照)

iii) 主語が省略されている時に無生物主語を取ることが可能な動詞の場合、受動態を取らない

主語が明示的に存在しない時に道具格やその他の任意格を取る名詞を主語とした方が自然な文になる動詞がある。

例)「翻訳する」

この計算機で日本語を翻訳する。
-> This computer translates Japanese.
(Japanese is translated by this computer.)

例)「送る」

プロセスからプロセスIDを送る
->The process sends the process ID.
(The process ID is sent from the process.)

例)「述べる」(記述動詞)

本論文ではATNについて述べる。
-> This paper describes an ATN.
(An ATN is described in this paper.)

()内の訳文は、無生物主語化ルールが適用されなかった場合の出力である。この様な規則の適用には、種々の制約条件が付加されることが多い。例えば、agentが明示的に存在しないことは自然であるが、agentを要求しない構文的条件(命令文など)があれば、これらの規則の適用は制限される。

iv)ある種の属性を持つ動詞連体形よる名詞の修飾部が、英語において、関係節ではなく前置詞句に表現した方が自然な英語になる場合がある。

例)「持つ」

これは翻訳機能を持つ計算機である。
->This is a computer with translating function
(This is a computer which has translating function.)

例)「ある」

探知機が歪みのある所を発見する。
-> The detector finds the place with distortion.
(The detector finds the place where there is distortion.)

()内の訳文は、変換のルールが適用されなかった場合の出力である。この規則にもまた、適用に際して制限条件があるが、関係代名詞の多出を防いだり、文を簡単にする為、このような構文が選択されることがある。

4. おわりに

以上に述べたように、語彙規則によって様々な言語現象に応じた処理を行うことが可能になり、より英語らしい表現を得ることができた。

参考文献:

- [1]三池、平川、天野: 英日機械翻訳システムにおける語彙的意味の記述について (情報処理学会自然言語処理研究 62-1)
- [2]中里、堤、平川、天野: 日英機械翻訳システムにおける日本語解析について (情報処理学会 第32回 全国大会)
- [3]熊野、小山、平川、天野: 日英機械翻訳システムにおける構造変換について (情報処理学会 第32回 全国大会)
- [4]吉村、平川、天野: 自然な文章生成における規範 -機械翻訳への応用- (情報処理学会 第38回 全国大会)