

日英翻訳システムALT-J/Eにおける
日英変換技術

6J-3

林 良彦 奥 雅博 石崎雅人
NTT電気通信研究所

1 はじめに

現在、開発中である日英翻訳システムALT-J/Eにおける日英変換技術について報告する。本システムの変換処理は、今回新たに提案する多段変換方式の考え方に基づいて設計されている。本稿では、多段変換方式の考え方、本システムにおける変換処理の位置付け・特徴、変換処理の概要及び、これに用いるパターン変換規則辞書の構成法について述べる。

2 多段変換方式の考え方

従来のトランスファ方式等による翻訳システムにおいては、対象とする文の集合を均一なレベルの解析/変換規則によって処理することが大原則となっている(文献[1])。しかしながら、現実の文の翻訳においては、均一なレベルへの解析、およびそこから汎用的な規則による変換処理では、適切な訳が得られない場合も多い。すなわち、均一な比較的深いレベルまでの解析が翻訳において意味を持たないか、あるいは、必要な表層上の情報が失われてしまう場合がある。

このような問題に対処するため、本システムでは、多段変換方式とよぶ新しい方式を用いている。多段変換方式は、「表現の構造は、意味の一部である。」という考えに基づき、原言語と目的言語の表現の密着度のレベルに応じて、個別の表現に対応する変換規則を用意し、それを変換処理ばかりでなく原言語の解析にも用いようとする方式である。現在開発中の日英翻訳システムALT-J/Eでは、このような考えに基づいて、(1)慣用表現変換規則、(2)意味的結合係パターン変換規則、および(3)汎用表現変換規則の3レベルの解析/変換規則を用意している。

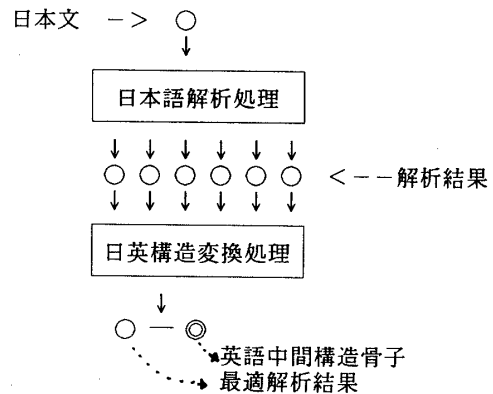
このように、多段変換方式においては、原言語解析の最終段階に目的言語への変換を意識した処理が組み入れられているため、解析と変換がある程度融合した方式であるといえる。

3 ALT-J/Eにおける日英変換処理

3.1 日英変換処理の入出力と基本的制御構造

日本語解析処理(文献[2])によって、単語認定の仕方ごとに構文構造が決定される。すなわち、日本語

析の結果としては、それぞれの組み合わせに応じて展開された多義が発生する。この展開された全ての解析結果が、日英変換処理への入力となる。日英変換処理では、そのなかの1つ1つの解析結果について並行して変換処理を試み、最後にそのなかから最適と考えられる1つの解析結果とそれに対応した生成すべき英語構造の骨子を入力する。(図1参照)



<図1:日英構造変換処理の位置付け>

3.2 日英変換処理の構成

日英変換処理を構成する主なサブ処理は次の2つである。

(1) 1つの構文解析結果に含まれる各単位文にたいする変換パターンを決定する処理

→ 単位文変換処理

(2) それぞれの組み合わせにたいする処理結果のなかから最適なものを決定する処理

→ 最適結果選択処理

(1)は、単位文の範囲のなかで、誤った解析結果を除去し、適切な変換が得られる解析結果を定める処理である。(2)は、入力文全体の範囲で、最適の変換を与える解析結果を選択する処理である。本システムの日英変換処理はこのような2段階の処理で解析多義を絞り、変換すべき英語構造を決定する。

3.3 ALT-J/Eの日英変換処理の特徴

以上より、次のような特徴があげられる。

(1) 英語への変換を意識して、日本語解析の多義を解消する。

(2) 日本語解析の多義を解消すると同時に、生成する英語構造を決定する。

本システムでは、4. に示すように1つの用言の1つの用法ごとに対応するローカルな解析・変換規則を多数持つ。これにより個別の言語現象に対応したキメの細かい訳出が可能となる。また、1つの規則の追加/修正が原則として、他の用言に対する規則に影響しないため、規則辞書のメンテナンスが容易となる。

4 単位文変換処理とパターン変換規則辞書

4.1 単位文変換の基本

本システムで扱う日本語単位文は大略次のような平板な構造を持つ。

単位文： := 修飾要素 + ・ ・ ・ + 修飾要素 + 述部

修飾要素： := 副詞修飾要素 | 格修飾要素

格修飾要素： := 名詞句 + 格助詞表現

単位文変換処理では、4. 2 に示すパターン変換規則を用いて、英語中間構造の骨子を決定する。修飾要素の翻訳結果は、この骨子に部分構造として取り込まれ、英語中間構造が生成される。(文献 [3])

4.2 パターン変換規則

パターン変換規則には、以下に示すような3つのクラスがある。このうち、(1)、(2)は、1つの規則は、「ある1つの用言の1つの用法」について整理したものである。

(1) 慣用表現パターン

1つ以上の格要素と、述部 (= 用言 + 様相属性等)の組み合わせをキーとして、単文の構造を整理したものである。

(例) Aが油を売る

→ A idle away one's time

(2) 意味的結合価パターン

用言をキーとして、単文の構造を整理したものである。

(例) AがBを暗記する

→ A learn B by heart

(3) 汎用表現パターン

これは、一種のデフォルトである。用言の分類ごとに、典型的な変換パターンを用意している。

以上の3つのクラスのうち、(1)は、いわゆる要素合成原理の壁を打ち破ろうとするもので、本システムにおける特徴の1つである。

4.3 慣用表現パターン変換規則

このクラスに属する変換規則はさらに次の2つに分類される。

(1) 慣用句

格要素、述部双方の組み合わせから英語表現が導かれ

るものと定義する。よって、一般に慣用句とよばれるものの以外のもを含む。

(例) Aは背が高い

→ A be tall

(2) 機能動詞結合 (文献 [4])

格要素、述部双方の組み合わせから英語表現が決まるが、格要素中の主名詞が用言性の名詞であり、その用言形の結合価パターンへのマッピングが可能なもの。

(例) AがBの非難を浴びる

→ AがBから非難される

→ BがAを非難する (+受け身)

4.4 辞書体系とパターンの適用

以上のような多レベルの変換規則についての辞書体系を図2に示す。慣用句については、キーとなる格要素にたいする制約条件を、インデックステーブルに記述する。機能動詞結合については、結合価パターンへのマッピング条件をインデックステーブルに記述する。

また、パターン変換規則の適用順序は慣用表現、結合価、汎用表現の順であり、自由格要素については、汎用表現パターンを混在して用いている。

慣用句インデックステーブル

用言	制約条件	
	

機能動詞結合インデックステーブル

用言	マップ条件	
	

結合価パターンインデックステーブル

用言	.

慣用句パターン対
...
...

結合価パターン対
...
...

<図2:パターン変換規則辞書の体系>

5 おわりに

変換処理の機能確認、変換規則の充実を翻訳実験をとおして行う。また、単位文の範囲を超えた部分にもパターン変換の考え方を拡張していく予定である。

<参考文献>

[1] 内田: 日経エレクトロニクス (1984.12.17)
 [2] 宮崎他: 情報処理学会第33回全国大会講演論文集 6J-2 (1986)
 [3] 岡本他: 情報処理学会第33回全国大会講演論文集 6J-4 (1986)
 [4] 村木: 「日本語の機能動詞表現をめぐって」 国立国語研究所報告 (1980)