

2B-4

機械翻訳システム P I V O T における 格パターン選択について

波多野 昇* 依田 透** 野村 直之** 村木 一至**

* 神戸日本電気ソフトウェア株式会社 ** 日本電気株式会社

1. はじめに

双方向機械翻訳システム P I V O T は、言語に依存しない中間表現を介して解析部/生成部が独立に処理を行う。概念依存構造で表現された中間表現からは、通常、複数の訳文が生成可能であり、これらの相違は「談話の流れ」(プラグマティクス)の差異であると説明できる。

生成処理においては、こうした「談話の流れ」を訳文に反映させることが訳文の品質に大きく影響してくる。そして、幾つかの可能性の中から原文の意味に忠実な訳文を得るためには、格パターン選択の戦略とそれを実現するための効率的なシステムがなければならない。(ここでいう格パターンとは、「が-を-に」のような深層格と対応した表層格助詞の「組合せ」のことである。)

本稿では、その一例として、P I V O T 日本語生成処理における格パターン選択の概要について述べる。

2. 格パターン

P I V O T の必須格処理を設計するにあたって、我々は日本語の用言を動詞68個、形容詞・形容動詞11個のパターンに分類・整理した。(例1)

各々のパターンには識別番号(JMAPS)を与え、これによって格パターンセットを一元管理している。また、辞書にはJMAPSを付与することで辞書をコンパクトにすることができた。

JMAPS	必須格の格パターン	例
VT11	AGTがOBJをTARに	返す、落とす
VT21	AGTがOBJをPRTと	つなぐ、区別する
VT67	OBJがTARに	つながる、育つ
AT67	OBJがTARに	優しい、厳しい
	AGT : AGenT	行為者格
	OBJ : OBJect	対象格
	TAR : TARget	目標格
	PRT : PaRTicipant	関与格

例1 用言の必須格パターンによる分類例

格パターンは、ルールや辞書とは独立に知識ベースとして管理している。個々の格パターンは、深層格と表層格助詞の対応を両者のペア(スロット)で表現し、格要素に対する意味的・統語的情報を記述している。また、スロットの並びには表層での標準語順の情報を反映させた。(図1)

JMAPS	表層格助詞	深層格	意味制限
VT11 YD	GA	AGT	-抽象事
	WO	OBJ	
	NI, HE, HETO	TAR	-時間
VT11 TD	NIYORU, NO	AGT	-抽象事
	NO	OBJ	-場所・時間
	HENO, TAISURU	TAR	-時間

図1 知識ベース

3. 必須格テーブル

知識ベースの格パターンは、実行時にテーブル(必須格テーブル)へジェネレートされる。

- 1) 辞書のJMAPSに対応した格パターンを生成
- 2) 態変換によって変形した格パターンを生成

1)は用言(概念ノードの訳語候補)の基本的な格パターンである。2)は受動、使役といった態変換によって変形された格パターンである。P I V O T ではこれら態変換も格パターンの変形として捉えている。1)2)で生成されたものが必須格テーブルであり、格パターン選択の対象となる候補である。(図2)

このように、我々は格パターン選択に先だて、可能性のある全ての格パターンを用意する戦略を取った。知識ベースと必須格テーブルはそのための方策である。また、必須格テーブルの各格パターンはその出どころである辞書とインデクス番号で対応付けられている。このため統語上の選択条件も格パターン選択のひとつとして扱うことが可能となった。

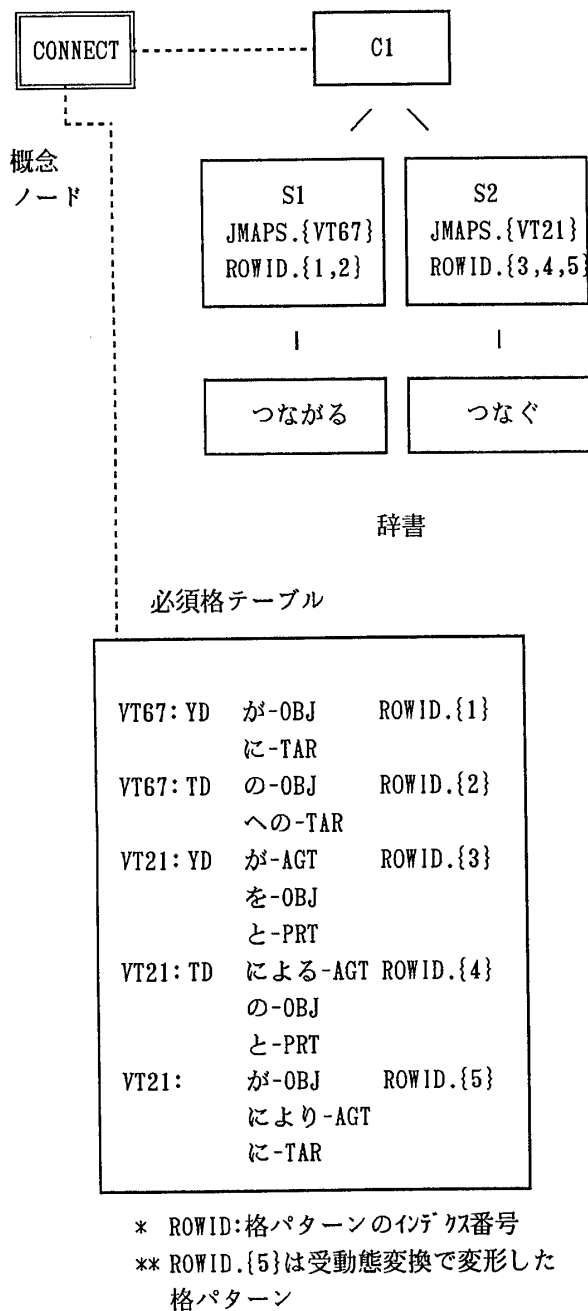


図2 概念ノード、必須格テーブル、辞書の関連図

4. 格パターン選択

格パターン選択は、各々の格パターンが持っている優先度の値を増減させることによって行う。条件Aに対して増減値aを設定しておき、条件を満足したものにその増減値を加算する。こうして、全ての条件を全ての格パターンに対して適用し、最終的にトータルして最も値の大きなものを最適格パターンとして選択する仕組みである。

この方法により、格パターン選択の規則は非常に柔軟なものにすることができた。つまり規則によって候補を絞り込んでゆく方法とは異なり、最後まで全ての候補を保持しておけるのである。また、設定する増減値の大きさに条件の優先順位を決定でき、規則のチューンナップも容易に行える。

PIVOTでは、現在、原文の主語、態などをそのまま訳文に再現させる方針を取っている。これは、談話情報を正確に抽出しえない現状にあって、その代替情報として現文の表層語順の情報が有効であるからである。

例2に選択条件と増減値の例を示す。

- 原文で主語だった要素の格が
日本語の「が」格となるもの : +160
- 原文が受動態のとき、
動作主格を含むもの : +10
- 原文が能動態のとき、
受動態変換で変形したもの : -10
- 肯定のとき、
否定の意味を持つ用言 : -100
- 動作性のアスペクトのとき、
状態性の用言 : -100

例2 選択条件とその増減値の例

5. おわりに

本稿では、PIVOT日本語生成における格パターン選択の仕組みと戦略について述べた。今後は、談話情報や格要素の意味制限を格パターン選択処理の中に取り込み、より柔軟できめ細かい選択が可能となるよう機能強化を続ける予定である。

本システムの開発にあたっては、神戸日本電気ソフトウェアの壁谷祐美子氏の協力があつた。ここに、謝意を表する。

参考文献

- 1)村木;"知識ベースと、言語に独立の中間表現とを用いた日英機械翻訳システム",日経エレクトロニクス1984. 12.17,pp195-220
- 2)壁谷、波多野、依田;"PIVOTE-J:日本語生成の概要",昭和63年電子情報通信学会秋期全国大会