

日本文書き替え処理における制御機能の構成について

4K-11

白井 諭<sup>†</sup> 池原 悟<sup>†</sup> 松尾三津恵<sup>‡</sup> 兵藤富子<sup>‡</sup>

<sup>†</sup>NTTコミュニケーション科学研究所

<sup>‡</sup>NTTアドバンステクノロジー

1 はじめに

日英機械翻訳の品質向上を狙いとして、筆者らは原文自動書き替え型翻訳方式を提案し[白井93a]、走行実験等を踏まえ、制御ルールを効率よく作成する方法[白井93b]と機能強化のための制御ルール体系の再構成[白井93c]について検討を進めている。本稿では、作成済みの制御ルールの分析[白井94b]から、既存の日本文書き替え処理が持つ基本機能と問題点を整理する。そして、制御ルールの4タイプ（14種類）への再構成に合わせて、処理系の構成と制御ルールの形式について提案する。

2 書き替え処理の基本機能

既存の制御ルール体系の再構成に伴う既存の940ルールの分類[白井94b]に引き続き、各ルールの処理に必要な書き替え基本機能を分析した結果、条件指定機能のほか、次の7項目に集約できることがわかった。

- ①連結：複数単語を1単語化する
- ②変更：各種単語属性を変更または付与する
- ③削除：単語・文節などを削除する
- ④補完：単語・文節などを補完する
- ⑤分割：1単語を複数単語に展開する
- ⑥入換：単語や文節の順序を変更する
- ⑦評価：条件を満たす表現の優劣を指定する

以上の基本機能が書き替えルールの種類別にどれだけ使われているかの分析結果を表1に示す。

また、その過程で条件指定機能や基本機能の一部には次のような問題があることがわかった。

- Ⓐ書き替え対象となる表現付近の文節や単語の有無を原則として記述する必要があるため、表現の揺らぎに弱い（条件指定）
- Ⓑ条件指定の内容に応じてAND/ORの解釈が固定されている（条件指定）
- Ⓒ表記を直接指定することにより単語連結の効果をj得ている（①）

表1 書き替えルールの基本機能構成

分類	基本機能 書き替えルール	連	変	削	補	分	入	評	ルール 数
		結	更	除	完	割	換	価	
解析後の 処理	形態素補正	25	32	2	4	13			60
	形態素多義絞り込み		1					15	16
	係り受け補正		19		1				19
	係り受け多義絞り込み		1					13	14
日本語内 の書き 替え	縮約展開		2	8	8				10
	冗長圧縮	21	17	33				1	55
	構文組み換え	10	20		8		12		41
	敬語の標準化			1					1
疑似的 日本語 への 書き 替え	助詞相当語	159	27	2	12	2			182
	副詞相当語	90	22	2	1				111
	連体詞相当語	9	7						15
	フレーズ	2							2
主 体的 な 書 き 替 え の 範 疇	接続様相時制	12	11	1					22
	様相時制	74	39	5	2	3	1		114
								その他	278
								合計	940

(注) 1ルールに複数機能を含む場合は、機能別に重複して集計したので、基本機能数の合計とルール数とは一致しない

Ⓐ変更または付与可能な単語属性が少なく、また指定方法も不統一となっている（②）

Ⓑ優先・非優先を間接的に実現している（⑦）

そこで、Ⓐ→正規表現に倣った曖昧指定の導入、Ⓑ→AND/OR指定の自由化と例外指定によるOR指定の簡素化、Ⓒ→文字列操作機能の導入、Ⓓ→対象とする単語属性の拡大と書式の統一、Ⓔ→評価点付与の導入、を考慮した。また、解析後の処理としての走行条件を変更した形態素解析や係り受け解析の再走行制御や、疑似的日本語への書き替えにおける訳語指定の強化のほか、係り受け解析処理[白井94a]と

A control structure of Japanese automatic rewriting functions

Satoshi SHIRAI<sup>†</sup>, Satoru IKEHARA<sup>†</sup>, Mitsue MATSUO<sup>†</sup> and Tomiko HYODO<sup>‡</sup>

<sup>†</sup>NTT Communication Science Laboratories (1-2356 Take, Yokosuka, 238-03) and <sup>‡</sup>NTT Advanced Technology Corporation

の処理系やルール形式の共通化についても検討した。

### 3 処理の機能構成

前節の検討を踏まえた日本文書き替え処理の機能構成を図1に示す。\*を付したものは注記の処理専用，それ以外は汎用の機能である。

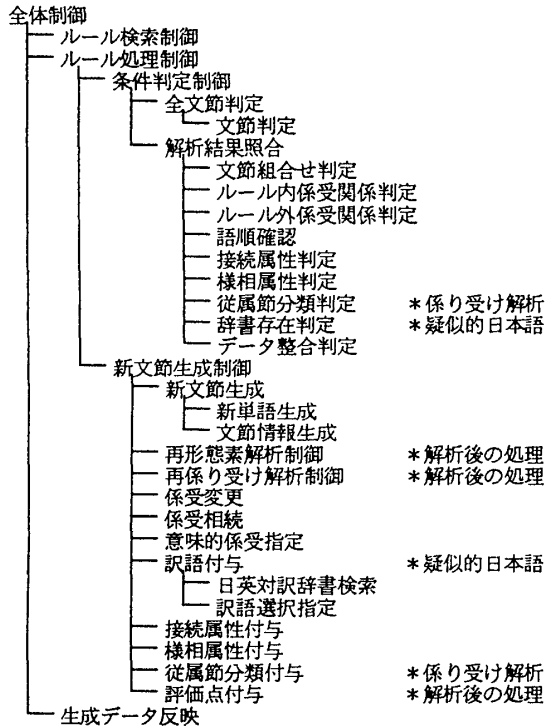


図1 日本文書き替え処理の機能構成

### 4 書き替えルールの形式

基本的には条件部と実行部があればよいが，適用可否の判定を容易にするためのルール検索部（検索キー，キーの所在文節，優先属性，専門分野などを指定する）と，評価点の導入に伴う解析多義の扱いを規定するモード指定部（通常書き替えでは対象となった解析情報と置換されるが，解析後の処理では評価点による多義の絞り込みを，また，係り受け解析では多義の展開を必要とする）を導入する。また，ルール保守のためのコメントなども必要に応じて記録する。図2にルールの記述例を示す。

```

( ("使う" B2 NIL) ;ルール検索部
  (条件部 ;条件部
    (文節 (@1)
      (B1 (JIRITSU (+1)) (FUZOKU (A2 (NHYO (o "を"))))
        (B2 (A3 (NHYO (o "使う")))) (A4 (NHYO (o "て"))))
        (?5 (NHYO (o ", " ", " "))))
      (@2))
      (係 (B1 B2 (o 格係)) (B2 B? (o 用用)))
      (受 (B* B1) (B* B2)))
      (処理 INS) ;モード指定部
      (生成 ;実行部
        (文節 (@1)
          (N1 (G1 +1)
            (G2 [ ] (NHYO "を使って") (HINS "7410")
              (YAKU (PREP ("USING")))))
            (@2))
            (係 (N1 (B2 KKR-BNST) 格係)
              (受 (N1 (B1 UKE) (B2 UKE))))
            (MEMO "R05280" "助詞相当語" "~//を//使って[/, ]"
              "試A60660" "天敵を使って害虫の駆除に成功した例は..."))
          )
        )
    )
  )

```

図2 書き替えルールの記述例

### 5 おわりに

本稿では，先に提案した書き替えルールの目的別構成に沿って既存のルールおよび処理系の問題点を再検討し，日本文書き替え処理の機能構成と制御ルールの書式について提案した。現在，本提案に基づいて試作した処理系（図3）の動作確認を進めており，今後は既存の制御ルールを本稿で提案した書式に移行・整理するとともに，制御ルール作成支援処理[白井94b]との連携により制御ルールの作成も開始する予定である。

#### <謝辞>

本検討にご協力くださった瀬下貴加子氏，阿部さつき氏を始めとするNTTアドバンステクノロジの各位に感謝する。

#### <参考文献>

[白井93a] S. Shirai, S. Ikehara and T. Kawaoka: Effects of automatic rewriting of source language within a Japanese to English MT System, TMI-93 Proceedings of the Conference, Kyoto(1993.7)  
 [白井93b] 白井, 池原, 河岡, 上田: 日本文書き替え処理における制御ルールの機能別構成, 第47回情処全大6P-4(1993.10)  
 [白井94a] 白井, 横尾, 木村, 小見: 従属節の依存関係を考慮した日本語係り受け解析の精度, 第49回情処全大1G-10(1994.10)  
 [白井94b] 白井, 池原, 阿部, 松尾: 日本文書き替え処理における制御ルールの類型情報の抽出, 第49回情処全大4K-10(1994.10)

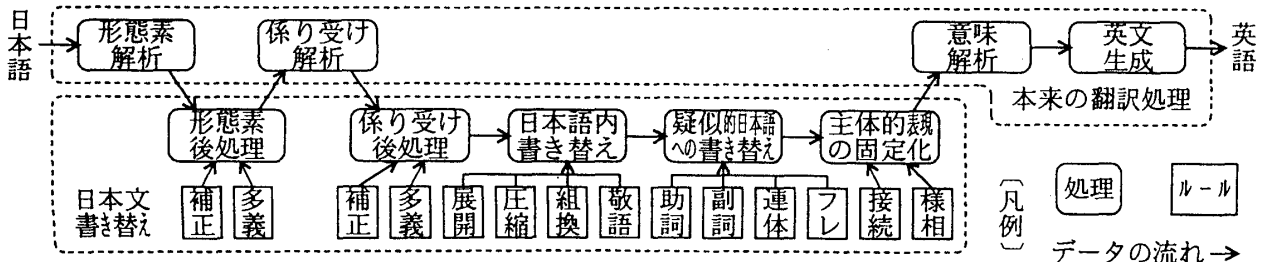


図3 日本文書き替え処理の位置づけ