

# 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

4K-10

白井 諭<sup>†</sup> 池原 悟<sup>†</sup> 阿部さつき<sup>‡</sup> 松尾三津恵<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。一般利用者がこの制御ルールを作成可能になれば、日英機械翻訳の利用が加速されると思われるが、実際にはその作成は難しい。そこで本稿では、制御ルールのテンプレートを用意しておき、それに基づいて制御ルール作成を誘導する方法を提案する。具体的には、作成済みの制御ルール940件を14種類（[白井93c]の13種類+「係り受け多義絞り込み」）に分類するとともに、件数の多いものについて類型化を試み、作成支援の可能性について検討する。

## 2 書き替えルールの分類

最初の提案[白井93a]では、日本語内書き替え（縮約展開、冗長圧縮、構文組み換え）と疑似的日本語への書き替え（独立句の表現、様相時制表現、接続表現）の2タイプ（6種類）を定義した。そして、形態素解析及び係り受け解析によって得られる解析情報に対して書き替え処理を適用することとし、また、書き替え処理の適用を1箇所あたり1回以下に制限する条件下で940ルールを作成した。

その後の検討により、目的の異なるルールを複数個順に適用しても副作用の恐れがないこと、解析誤りの回復や多義の絞り込みにも応用できることがわかったので、書き替え処理の適用範囲を拡張するとともに、目的別に4タイプ（13種類）に分類することを提案した[白井93c]。

この提案に沿って、先の940ルールをルール本来の狙いという観点から再検討し、前記の13種類に「係り受け多義絞り込み」を加えた14種類に分類した結果を表1に示す。日本語内の書き替えは機械翻訳可能な表現に対応づけるため、ルールの作成効率が今ひとつあがっていないが、疑似的日本語への書

表1 既存の書き替えルールの分類

分類	書き替えルールの種類	備考	件数
解析後の処理	形態素補正	単語の分割誤り等を狙い撃ち救済	60
	形態素多義絞り込み	単語並びにより不要な解釈を除去	16
	係り受け補正	係り受けの誤解釈を狙い撃ち救済	19
	係り受け多義絞り込み	多項関係により不要な解釈を除去	14
日本語内書き替え	縮約展開	語尾の省略や短絡的な用法を回復	10
	冗長圧縮	訳出不要な表現を除去して簡潔化	55
	構文組み換え	情動的な表現を断定文などに変換	41
	敬語の標準化	敬語表現を通常の表現に変換する	1
疑似的日本語への書き替え	助詞相当語	英語の前置詞句に直接対応付ける	182
	副詞相当語	英語の副詞表現に直接対応付ける	111
	連体詞相当語	英語の形容詞句に直接対応付ける	15
	フレーズ	英語の定型表現へ直接対応付ける	2
主體的な書き替え	接続様相時制表現	文の接続の構造を固定的に捉える	22
	様相時制表現	文末の表現を英語対応に固定する	114
その他		複合的、分類困難、分類保留など	278
合計			940

き替えは英語表現を直接指定できるため、ルールの作成効率が比較的良いのを反映している。

## 3 書き替えルールの類型情報

本節では、表1のうち件数が多い助詞相当語と副詞相当語について類型化を検討した。様相時制表現は訳語指定時に助動詞などの語形変化情報を付与する必要があるので、今回の検討からは除外した。

助詞相当語および副詞相当語に対する類型情報を図1～図3に示す。類型化可能なルールの大部分は、助詞相当語では固定化する表現の直前および直後、副詞相当語では直後のみを認定条件として設定すればよく、半自動的に作成可能である。また、これら

### Acquisition of typical patterns in rules for automatically rewriting of Japanese sentences

Satoshi SHIRAI<sup>†</sup>, Satoru IKEHARA<sup>†</sup>, Satsuki ABE<sup>‡</sup> and Mitsue MATSUO<sup>‡</sup>

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

のルールが多くは翻訳家による訳例に基づいて作成したが、森田らが指摘している「複合辞性」の認定条件、①構成要素の緊密化の度合、②後の形式化の度合[森田89]を踏まえて、テンプレートとして記述すべき項目の検討を深めたい。

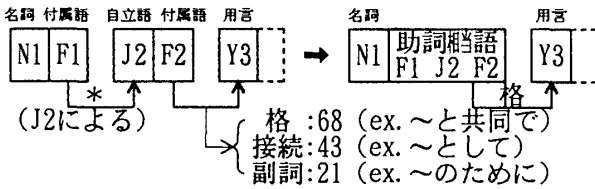


図1 助詞相当語の形式(連用型)

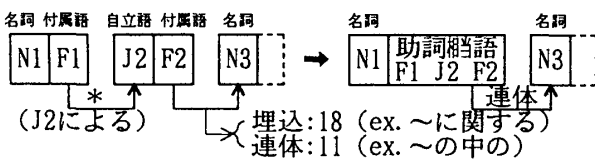


図2 助詞相当語の形式(連体型)

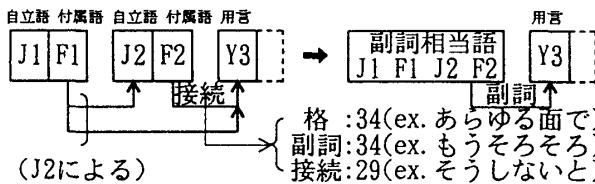


図3 副詞相当語の形式

#### 4 書き替えルールの作成支援

日本文書き替え処理は人手により行なわれている前編集を一部自動化するとともに、いくら書き改めても機械翻訳できない表現に対して英語の表現を直接指定する手段を併せて提供している。このため、日本文書き替え処理は機械翻訳の利用促進に効果を発揮すると思われるが、制御ルールを一般利用者が記述するのは困難である点が問題となる。

筆者らは、入力原文に対する解析情報と、目標とする訳文が得られる前編集された文に対する解析情報を組み合わせることにより、半自動的に書き替えルールを作成する方法を提案し[白井94b]、試作によりその有効性を確認しているが、これは日本語内書き替えに適した作成支援であると考えられる。

一方、表1のルールの件数から疑似的日本語の作成を効率化するのも作成支援の有力な選択肢と云える。前節の検討から助詞相当語と副詞相当語についてはいくつかのパターンがあることがわかったので、抽出されたテンプレートに従って設定すべき項目の入力を誘導していく方法が考えられる。そこで、日本文書き替え処理[白井94]との連携をとり、図4のような支援処理を提案する。すなわち、疑似的日本語への書替に対し、①利用者が固定的に扱う表現とその訳語を指定する、②システムは助詞相当語/副詞相当語等の種別を判定する、③テンプレートに基づき条件を設定したり必要な情報を利用者に問い合わせたりしてルールを完成させる、の手順となる。

#### 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を再検討し、助詞相当語と副詞相当語について類型情報を抽出しテンプレートを作成するとともにそれを利用する制御ルールの作成支援について提案した。今後は、一般利用者による制御ルールの作成に向けて、デフォルト情報の整備や作成支援処理の改良を進める予定である。

<謝辞>

本検討にご協力くださった中村三紀氏、兵藤富子氏を始めとするNTTアドバンステクノロジーの各位に感謝する。

<参考文献>

[森田89] 森田, 松木: 日本語表現文型, NAFL選書, アルク(1989.5)  
 [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, 横尾, 土倉: 日本文書き替え処理用制御ルールの作成支援, 1993年秋信学全大D-67(1993.9)  
 [白井93c] 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94] 白井, 池原, 松尾, 兵藤: 日本文書き替え処理における制御機能の構成について, 第49回情処全大4K-11(1994.10)

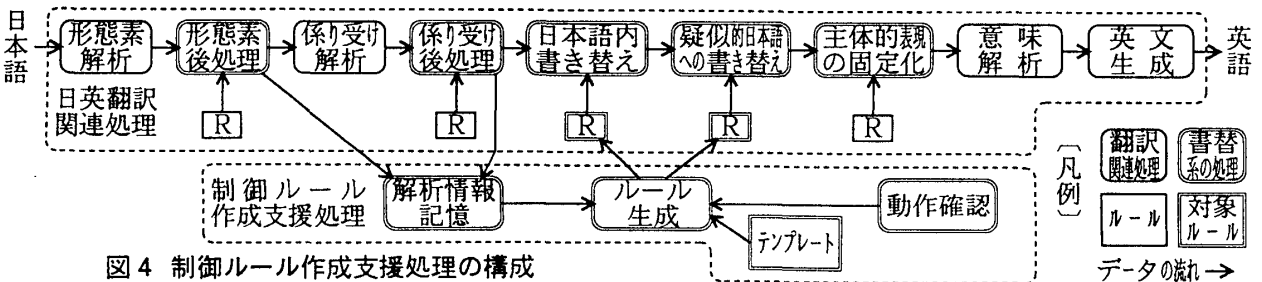


図4 制御ルール作成支援処理の構成

データの流れ →