

大域的句型制約を組み込んだ 線形・非線形融合型英文生成方式

安藤 安明[†] 川辺 諭[‡] 宮崎 正弘[†]

[†]新潟大学大学院自然科学研究科 [‡]科学技術振興機構

1 はじめに

従来の機械翻訳では、原言語と目的言語間で単語や構文が単純に対応しない非線形の部分があるために、品質のよいこなれた訳文が得られないことが多い。

そこで本手法では、まず非線形に対応する部分を持つ日本語・英語表現に関して、大域的な日英文型パターンを準備し、日英翻訳において、日英文型パターンから英語表現の大域的な構造を得て、線形変換された節などの局所構造を埋め込むことで、こなれた英語表現を生成する。

本稿では、本手法によるこなれた日本語文表現からこなれた英文表現を生成する方法を提案し、その有効性を論じる。

2 要素合成の問題点

言語表現から表現と認識の2つの構造を抽出するために、富田法を prolog 上で実現したパーザ [2] を拡張した SGLR-plus [3] を使っている。

図1は、SGLR-plus を用いて実装された英文パーザ [4] 上で、英語表現 “He runs.” を解析した結果である。パーザの出力として、言語表現の持つ統語的な構造である“表現構造”と、話者の対象に対する客体的な認識と主体的な判断の情報で

Linear and Non-linear Unied English Sentence
Generation using Global Sentence Pattern
Yasuaki Ando[†] Satoru Kawabe[‡] Masahiro
Miyazaki[†]
[†]Niigata University [‡]Japan Science And Tech-
nology Agency

ある“認識構造”が得られる。

寺島らの英文生成法 [5] では、これらの構造対をあらかじめ蓄えておくことで、基本的な単文、複文、重文を要素合成的に生成する。

しかし、従来の翻訳システムで行われている要素合成による方式では、柔軟な翻訳を行うことはできず、不自然な翻訳結果となってしまう場合がある。例えば、「あまりに暑かったので、コートを脱いだ」のような文に対して、“so - that” 構文を使って表現することが出来ない。

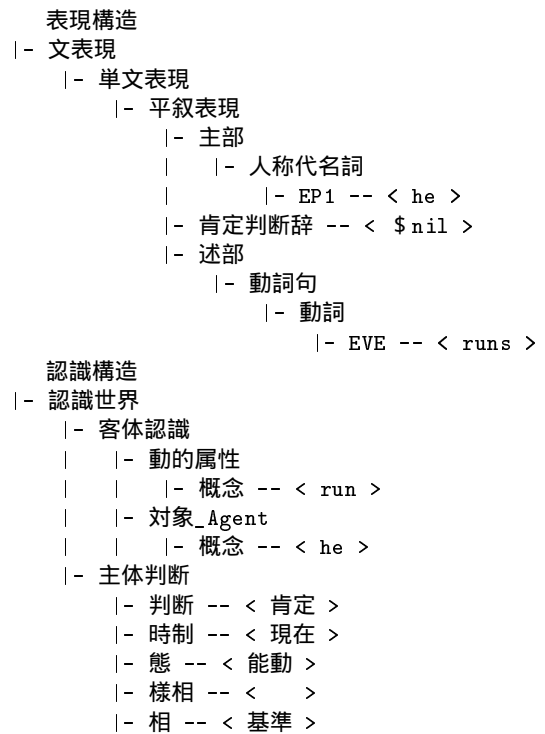


図1：表現構造と認識構造の例

3 日英翻訳処理のおおまかな流れ

複文、重文の日本語表現から単文の英語表現を生成する場合など、日英間が非線形に対応する英文表現を生成するために、要素が非線形に対応する複文、重文レベル(大域レベル)の日英表現の対を、日英文型パターン対として準備しておき、英文生成処理に利用する。処理の流れは以下の通り(図2)。

- 日本語入力文の形態素、構文解析を行う
- 日本語の統語構造を利用して、適合する大域レベルの日英文型パターンを検索する
- 検索に成功した場合は、パターンの英文構造を参照して出力文のおおまかな大域構造を決定する
- マッチングしなかった節などの差分について、さらにパターン検索を繰り返す
- 検索に成功した場合は、パターンの英文構造を参照して、すでに準備されている大域構造に、局所構造を埋め込む
- 最終的にマッチングしなかった部位は、要素合成処理を行う。

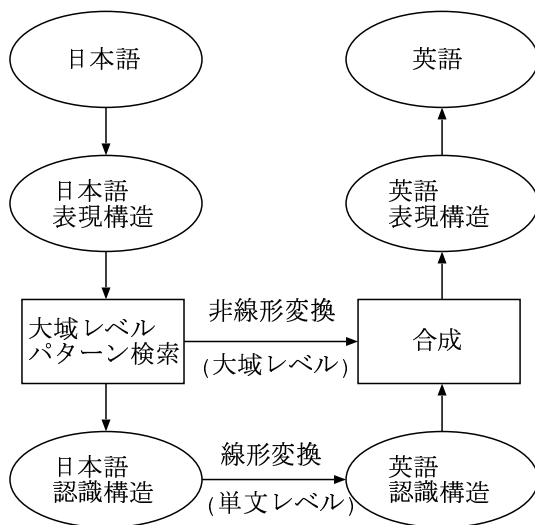


図2：英文生成処理の流れ

4 日英文型パターン対

表2は、日英対訳データ「彼が失敗したので私は心が痛む。」と”It ails me that he failed”から作成したパターンである。

表1：日英文型パターン対の例

日本語パターン	CL1 ので N2 は心が痛む
英語パターン	it ails N2.obj that CL1

N2.objの部分は「名詞を目的格に変形する」意を表す。

このパターンでは、構文要素(語、句、節)が汎化(変数化)される。変数は、対応付けのために番号が振られている。汎化のレベルは部分によって異なる。

以下は、日英文型パターン対の木構造である。

JP : (ss (s (cl "CL1") (aux "て")) (s (cl (vp (n "N2") (p "は")) (vp (np (n "心") (p "")) (v "痛む"))))))

```

|-ss
  |-s
    |-cl-<CL1>
    |-aux-<て>
  |-s
    |-cl
      |-vp
        |-np
          |-n-<N2>
          |-p-<は>
        |-vp
          |-np
            |-n-<心>
            |-p-<が>
          |-v-<痛む>

```

図3-1：日英大域パターン(日本語パターン)の例

EP : (s (cl (np (pron "it")) (vp (v "ail") (obj (np (n "N2")))) (that-cl (conj "that") (cl "CL1"))))

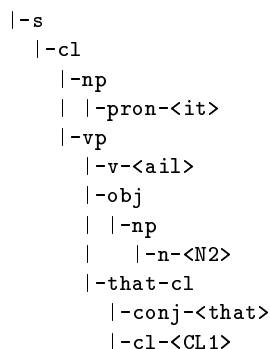


図 3-2 : 日英大域パターン (英語パターン) の例

5 線形パターン対

線形パターンとしては、単文パターンを用いる。これには、格パターン辞書を利用する。認識構造の客体認識をもとにしてデータベースから検索を行い、その情報を利用する。表 2 は格パターン辞書のレコードの例である。

表 2 : 線形パターン対の例

テキスト	(例)
表記	食べる
日本語パターン	N1 が N2 を 食べる
英語パターン	N1 eat N2
名詞意味属性	N1 4 人 535 動物 N2 838 食糧 534 生物

以下は、線形パターン対の木構造である。

JP : (s (cl (vp (np (n "N1") (p "が")) (vp (np (n "N2") (p "を")) (v "食べる"))))

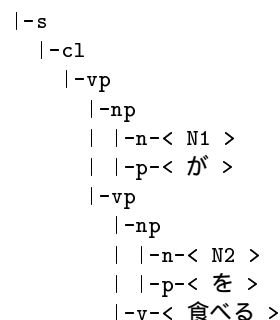


図 4-1 : 線形パターン (英語パターン) の例

EP : (s (cl (sbj (np "N1")) (vp (v "eat") (obj (np "N2"))))

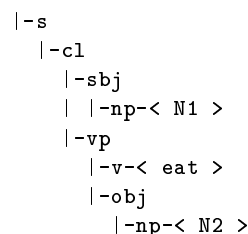


図 4-2 : 線形パターン (英語パターン) の例

6 パターン照合部

日本語入力文と準備されているパターンをマッチングする際に、マッチしなかった部分が差分として抽出される。

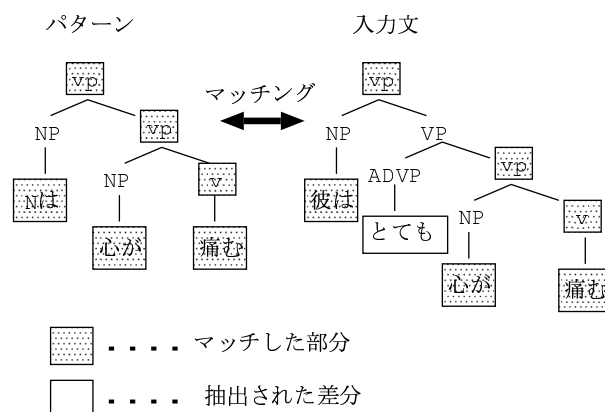


図 5 : パターン照合部

得られたパターンと抽出された差分を合成することにより、訳文が得られる。

7 変換の具体例

以下は日本語入力文「彼が失敗したので私は心が痛む」に関して、日英文型パターンを検索し、英語文を生成する例である。

入力文
 彼が 失敗し た ので 私は 心が 痛む
 適合した日英文型パターン
 JP : CL1 ので N2は 心が 痛む
 EP : it ail N2.obj that CL1
 大域構造決定
 -> it ail N2.obj that CL1
 単文生成
 J_CL1 = 彼が 失敗し た
 E_CL1 = he failed
 変数と単文の埋め込み
 -> it ail I.obj that he failed
 形態素調整
 -> it ails me that he failed.

図6は、大域構造に局所構造を埋め込む様子である。

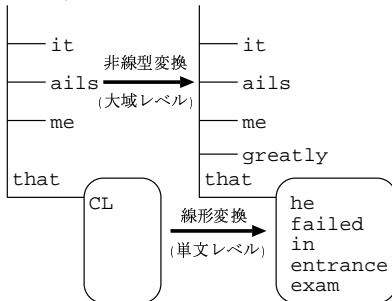


図6：局所構造の埋め込みの様子

8 システム概要

非線形変換において、試作中の日英翻訳システムの概要を図7に示す。

9 おわりに

本稿では、認識構造から生成される線形な英語表現と大域的なパターンを用いて生成される非線形な英語表現を合成して英文を生成する手法について述べた。本手法により、こなれた英文表現の生成が可能になる。

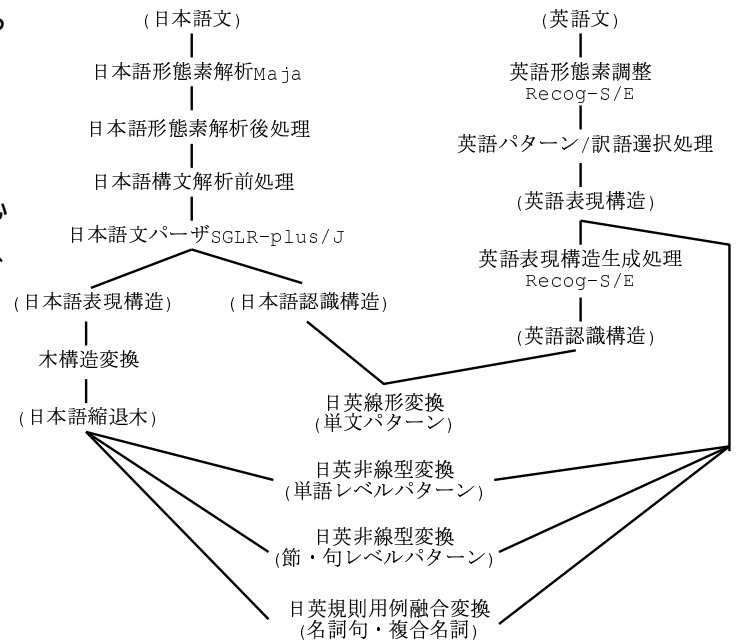


図7：システム概要

10 謝辞

この研究は、科学技術振興機構 (JST) の戦略的基礎研究事業 (CREST) の支援と、科学研究費補助金基盤研究 (B)(課題番号 13480091) を受けています。

参考文献

- [1] 三浦つとむ：日本語とはどういう言語か，講談社学術文庫 (1976)
- [2] 沼崎、田中：SGLR：逐次型一般化 LR パーザの Prolog による実現、情報処理学会論文誌、Vol.32, No.3, pp.396-403 (1991)
- [3] 五百川、宮崎：痕跡処理のための GLR 法の拡張、言語処理学会、Vol.7, No.3, pp3-21 (2000)
- [4] 加藤、宮崎：認識構造を抽出する英語文パーザの試作、情報処理学会第 64 回全国大会、3M-05 (2002)
- [5] 寺島、宮崎：表現と認識の構造対を用いた英文生成システムの試作、情報処理学会第 62 回全国大会、6M-5 (2001)