

# 表現と認識の構造対を用いた日本文生成システムの試作

1M-05

松尾 満

宮崎 正弘

新潟大学大学院自然科学研究科

## 1 はじめに

時枝誠記による「言語過程説」を発展的に継承した三浦つとむは、言語表現の背後に存在する話者の認識と表現の構造関係によって、言語を捉えることが重要であると述べている [1]。本稿では、三浦の言語過程説に基づき作成された表現と認識の構造対を用い、日本文を生成するプログラムを試作し、その有効性を示す。

## 2 認識構造と表現構造

言語の生成において、話者（書き手）は対象の客観的な姿を概念として捉える（客体認識）と共に、対象に対する主体的判断（主体認識）を重ね合わせる。それにより、話者は頭の中で対象のイメージ（認識）を作り上げる。このような客体認識と主体認識を融合し、構造化したものを認識構造と呼ぶ。話者は、このような認識構造から言語規範（文法や辞書など）に従って言語を生成し、それを表現する。この表現を構造化したものを表現構造と呼ぶ。表現構造は、入力文の統語的構造を表しているため、文生成においては、この表現構造上の単語を接続して出力する。

日本語文を統語解析し、このような表現構造と認識構造を出力する日本語文パーザ [2] が、一般化 LR 法（富田法）を Prolog 上に実現した SGLR パーザ [3] の拡張版である SGLR-Plus パーザ [4] 上に構築されている。

ここでは、このような日本語文パーザによって出力される表現構造と認識構造の対を構造対データとして蓄えておき、入力認識構造から日本語文を生成するために利用する。表現構造と認識構造の構造対データの例を図 1 に示す。

Japanese Sentence Generation Method using  
Pair of Speaker's Recognition and Expression  
Structure

Mitsuru Matsuo, Masahiro Miyazaki  
Niigata University

## 3 日本語文生成

あらかじめ、日本語文パーザで得られる 1 対 1 の表現構造と認識構造の対を構造対データとして蓄えておく。生成プログラムに認識構造が入力された時、この構造対データ中にある認識構造から最も類似した認識構造と、それと対になっている表現構造を抽出する。この時、入力側と構造対側の認識構造の差分も抽出する。

マッチングの第一段階では、入力認識構造と主体認識が同じ認識構造をマッチングの対象とする。

第二段階では、すべての候補に対してマッチングを試み、単語の差異数が小さい順に認識構造を並べる。

第三段階では、選ばれた最小差異数の認識構造の単語で、入力側と違うものは置換する。もし、差違数が 0 の構造対データ側の認識構造が見つければ、それを最も類似した認識構造として、その時点でマッチング処理を終了する。この時、主体認識と単語の対応テーブルなどを用いて、認識構造中にある話者の主体認識を表わす単語（助詞、助動詞など）を生成する。さらに、用言や助動詞などの語尾変化処理も行う。

これらの処理により、入力認識構造に対応する表現構造が得られる。この表現構造の単語を接続することにより、入力認識構造が表わす日本語文を生成する。図 2 に日本語文生成の流れを示す。

## 4 本方式の利点

構造対データには、日本語文パーザが出力する表現構造と認識構造を用いているため、解析用の文法や辞書を生成でも間接的に利用している。この点において、文法や辞書を解析、生成で併用することを実現している。

本方式では、入力認識構造に最も類似する認識構造を構造対データからマッチングにより見つけだすので、入力認識構造が表わす文を正しく生成できる。しかし、様々な種類の文を生成できるようにするためには、あらかじめ約 5,000 種類の表現構造と認識構造の対を構造対データとして蓄えておく必要がある。

『彼は落ちている缶を拾った。』の表現構造

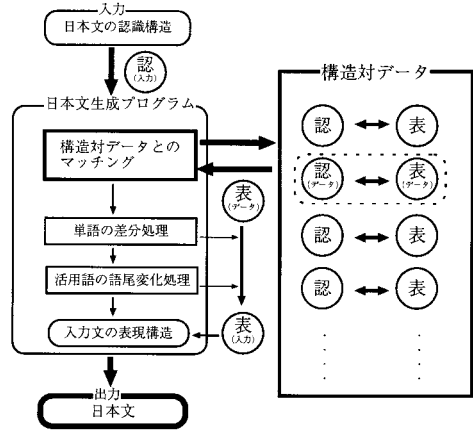
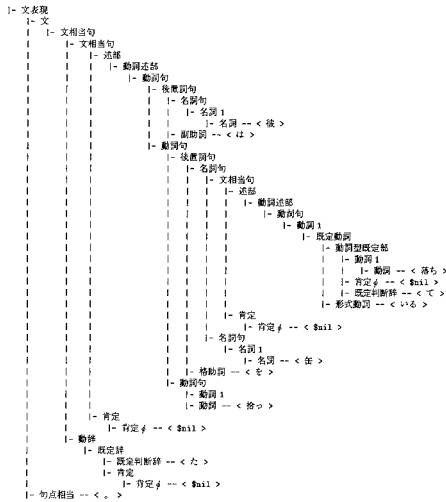


図2：日本語生成の流れ

『彼は落ちている缶を拾った。』の認識構造

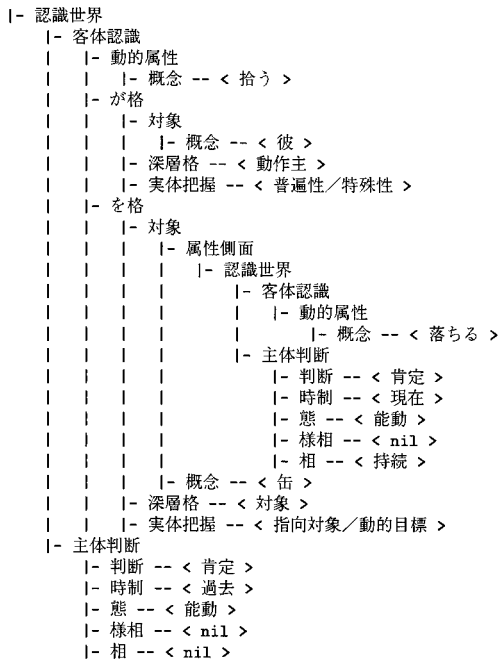


図1：表現構造と認識構造の例

5 おわりに

本稿では、計算機上で認識から表現を生成する過程の実現のために、認識構造と表現構造の対を用い、入力認識構造から日本語文を生成する生成プログラムを試作した。現段階においては、単文、複文が実現しているが、今後さらに、重文の生成についても検討する必要がある。

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] 藤石、宮崎：三浦文法による日本語パーザの試作、情報処理学会第51回全国大会、4H-9(1995)
- [3] 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)
- [4] 五百川、宮崎：痕跡処理のための GLR 法、自然言語処理、Vol.7,N0.3,pp.3-21(2000)