

表現構造と話者の対象認識構造を抽出する 日本語文パーザの試作

藪正樹

藤石知夏

宮崎正弘

新潟大学大学院自然科学研究科

1. はじめに

表現構造と話者の対象認識構造とは、時枝誠記による言語過程説を発展的に継承し関係意味論を唱えた三浦つとむの日本語文法[1]に基づく構造である。この三浦文法に基づく日本語構文解析用の文法をDCG形式で作成し、一般化LR文法(富田法)をProlog上に実現したSGLRパーザを用いて、表現構造から話者の対象認識構造を抽出する日本語文パーザを試作した。本稿では、試作した日本語文パーザについて報告し、その有効性を論じる。

2. 表現構造と対象認識構造

2.1 日本語表現構造

言語過程説によれば、言語表現は、話者の主観的な感情・要求・意志・判断などを直接表現した主体的表現と、話者が対象を概念化して捉えた客体的表現とに分けられる。膠着語である日本語は、主体的表現の語と客体的表現の語がそれぞれ別の単語として独立しており、表現ではそれらが結び付けられて用いられる。また、一人称・時制・否定などは、観念的な自己分裂によって生じた観念的な話者が観念的な世界を移動することで説明する。したがって日本語においては主体的表現の語を連結することによ

て次々に観念的な世界の多重化を表現する。

日本語文「花が咲かなかった」の表現構造を図1に示す。ここで、 ϕ は表現としては省略された肯定判断を表している。

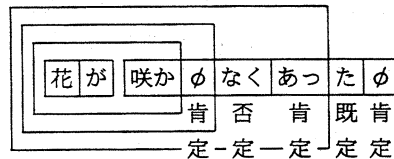


図1 「花が咲かなかった」の表現構造

2.2 対象認識構造

三浦文法では、言語の意味を対象-認識-表現の関係として捉える。人間が現実の世界を対象として言語表現を行う過程では、話者はまず着目した実体や属性を概念として捉え、実体と属性、あるいは実体と実体との間に関係付け、それに話者自身の判断を加えて表現する。またそれを一つの概念として捉え直し、順々にそれに主体的な判断を下すことにより、話者の観念的に多重化した世界を表現する。

日本語文「花が咲かなかった」の認識構造を図2に示す。まず話者は「花」という実体と「咲く」という動的属性を「が」という主体的な個別的判断によって関係づけて一つの概念と

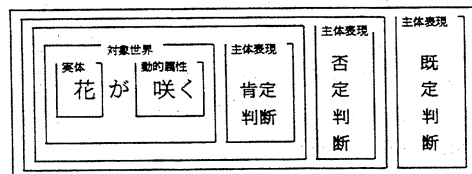


図2 「花が咲かなかった」の認識構造

An Implementation of Japanese Paser based on the Viewpoint of Speaker's Recognition
Masaki Yabu, Chinatsu Fujiishi, Masahiro Miyazaki
Niigata University

し、肯定判断を下す。次に、その観念的な世界を否定判断によって否定し、最後に、またその世界に既定判断を下すことによって観念的に三重の世界を構成する。

図1および図2に示したように、日本語においては、表現構造と認識構造は非常に似通った構造となる。

3. 日本語文パーザ

3. 1 品詞分類

三浦文法では、単語を文構成上の機能や単語が表す内容で分類するのではなく、対象の種類とその捉え方で分類する[2]。代表的な品詞に対する考え方を以下に述べる。

(1) 客体的表現

名詞は実体のあり方を捉えたものである。

動詞は動的属性を表し、形容詞は静的属性を表す。ともに実体の属性を概念化する。

(2) 主体的表現

助動詞は、観念的な自己分裂によって話者が見る対象世界に対する話者の感情・意志・判断などを直接表現したもので、肯定・否定・既定・未定の主体的表現を行う。

助詞は、実体に対する話者の捉え方を表現する。三浦文法では、「が」は個別的判断を表し、「は」は普遍判断または特殊判断を表すとされる[1]。格パターン辞書などへの参照を考慮し、試作した日本語文パーザが出力する認識構造には、判断に関する情報と、格に関する情報の両方を含ませた。例えば、「枝には」は【格—に格、判断—静的目標、特殊的】、「日本だけが」は【格—が格、判断—唯一的、個別的】と表現する。終助詞は話者の感情を伝達する。認識構造では、全体を対象世界として捉え直し、終助詞によって最終的な主体判断が下される。接続助詞は事象間の関係づけを行う。

3. 1 単文解析

単文とは、主語・述語の関係が一回だけ成立

するもの[3]をいう。試作した日本語文パーザによって、日本語の単文「梅の花が咲く」を解析した結果を図3に示す。

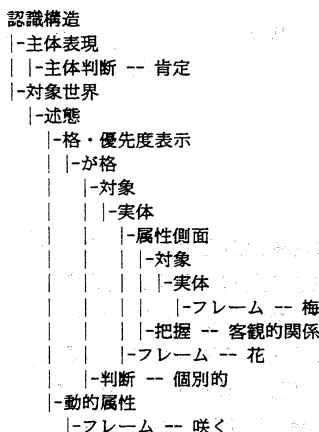


図3 「梅の花が咲く」の解析結果

3. 2 複文解析

複文とは、主語・述語が成り立っている文で、さらにその構成部分に主語・述語の関係が認められるもの[3]をいう。一文が名詞を修飾するものとして他の文に埋め込まれている場合、格要素の係り先に曖昧性が生じる。試作した日本語文パーザでは、

- ・ 「が」格は直後の属性に係りやすい
- ・ 二つ以上の「を」格は同時に一つの属性に係らない

という優先規則を用いて、曖昧性の制限を試みている。また、連体節には次節に述べる優先度の情報を含ませた。日本語の複文「子供が泣く声が聞こえる」を解析した結果を図4に示す。

3. 3 重文解析

重文とは、一文の中で、主語・述語の関係が二回以上成立し、しかも、その関係が対等に並列されている文[3]をいう。試作した日本語文パーザによって、日本語の重文「川が流れ、鳥が歌う」を解析した結果を図5に示す。

一文に複数の従属節がある場合、従属節の係り先に曖昧性が生じるが、従属節を表層表現によって六種類のタイプに分類して、それらを優

認識構造

```

|-主体表現
|-主体判断 -- 肯定
|-対象世界
|-述態
  |-格・優先度表示
  |-が格
  |-優先度 4 -- 連体節
  |-対象
    |-実体
      |-属性側面
      |-認識構造
        |-主体表現
        |-主体判断 -- 肯定
        |-対象世界
        |-述態
          |-格・優先度表示
          |-が格
          |-対象
            |-実体
              |-フレーム -- 子ども
              |-判断 -- 個別的
              |-動的属性
              |-フレーム -- 泣く
          |-フレーム -- 声
        |-判断 -- 個別的
      |-動的属性
      |-フレーム -- 聞こえる

```

図4 「子どもが泣く声が聞こえる」の解析結果

認識構造

```

|-主体表現
|-主体判断 -- 肯定
|-対象世界
|-述態
  |-格・優先度表示
  |-優先度 3 -- 連用節
  |-対象
    |-認識構造
      |-主体表現
      |-主体判断 -- 肯定
      |-対象世界
      |-述態
        |-格・優先度表示
        |-が格
        |-対象
          |-実体
            |-フレーム -- 川
            |-判断 -- 個別的
            |-動的属性
            |-フレーム -- 流れる
        |-判断 -- 中止
      |-係り先
      |-対象
        |-認識構造
          |-主体表現
          |-主体判断 -- 肯定
          |-対象世界
          |-述態
            |-格・優先度表示
            |-が格
            |-対象
              |-実体
                |-フレーム -- 鳥
                |-判断 -- 個別的
                |-動的属性
                |-フレーム -- 歌う

```

図5 「川が流れ、鳥が歌う」の解析結果

先順位付けし、

- ・ 優先度の低いものは高いものに係る
- ・ 優先度の高いものは低いものに係らない

という規則を用いることによって、従属節の係り先の曖昧性を高い精度で解決することができる[4,5]。三浦文法に基づく品詞分類にしたがった従属節のタイプとそれらの優先度を、表1に示す。

表1 従属節の分類

優先度	従属節のタイプ
1	「終止形」+「展開型」の接続助詞+読点 「～すると」などの引用節
2	「終止形」+「展開型」の接続助詞
3	「連体形」+形式名詞「の」+助動詞「だ」の連用形+読点 「連用形」+助動詞「た」の連用形+読点 「連用形」+読点 「～するよう」などの引用相当節 形式名詞に係る連体節
4	「連体形」+形式名詞「の」+助動詞「だ」の連用形 「連用形」+助動詞「た」の連用形 「連用形」 普通名詞に係る連体節
5	「連用形」+「同時型」の接続助詞+読点
6	「連用形」+「同時型」の接続助詞

試作した日本語文パーザが出力する認識構造には、節ごとに優先度の情報を含ませた。この優先度情報により、正しい係り先を決定する。係り先が異なる二種類の例（係り先が文末の節となる場合および係り先が直近の節となる場合）を、三浦文法風に解析した結果を図6および図7に示す。また、上記二種類の例に相当する日本語の重文「FDDを装備していたがHDDを組み込んでアクセス時間を短縮する」および「FDDを装備していたのでHDDを組み込んだがお金がかかる」を、試作した日本語文パーザによって解析した結果を図8および図9に示す。

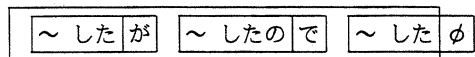


図6 係り先が文末の節となる場合の三浦文法風の解析結果

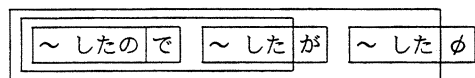


図7 係り先が直近の節となる場合の三浦文法風の解析結果

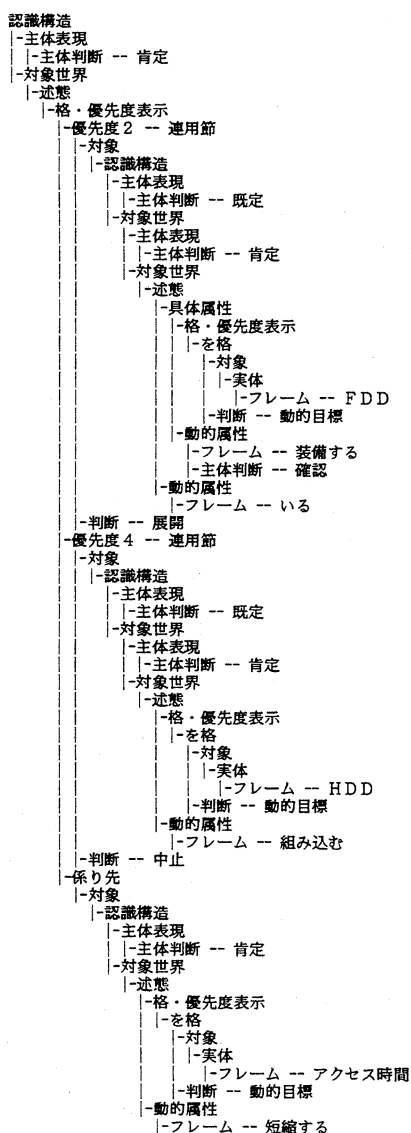


図8 「FDDを装備していたがHDDを組み込んでアクセス時間を短縮する」の解析結果

4 おわりに

三浦文法に基づく日本語文パーザを、一般化LR文法(富田法)をProlog上に実現したSGLRパーザを用いて試作した。本稿の最後に示した重文の解析結果については、今後さらなる検討が必要である。

参考文献

- [1] 三浦つとむ：日本語はどういう言語か，講談社（1976）
- [2] 宮崎，白井，池原：言語過程説に基づく日本語品詞の体系化

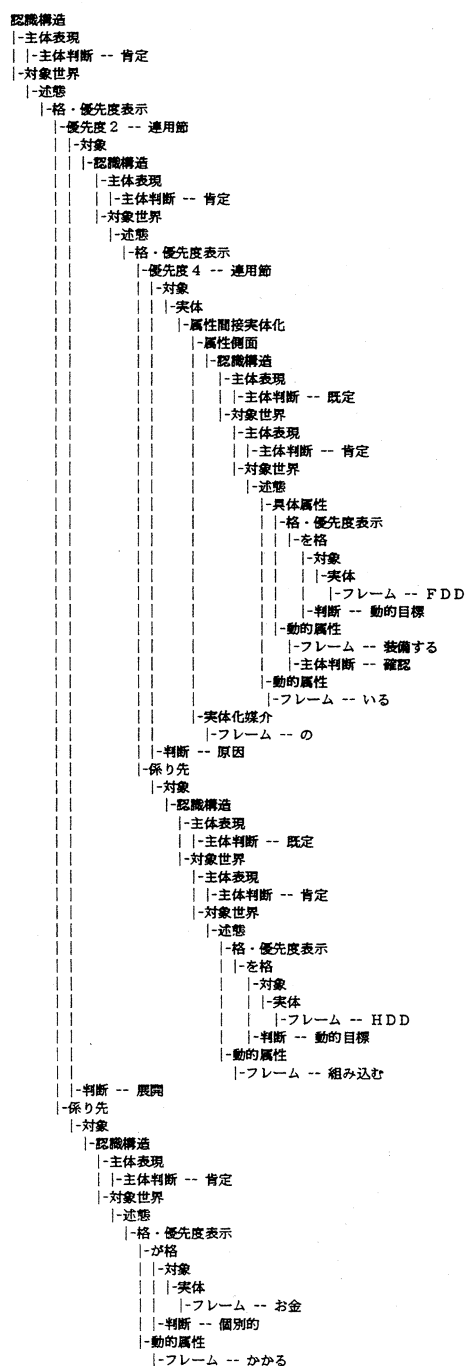


図9 「FDDを装備していたのでHDDを組み込んだがお金がかかる」の解析結果

- とその効用，自然言語処理，Vol.2，No.3，pp3-25（1995）
- [3] 梅村，金田一，阪倉，日野原監修：日本語大辞典，講談社（1989）
- [4] 白井，池原，横尾，木村：階層的認識構造に着目した日本語従属節間の係り受け解析の方法とその精度，信学技報，NLC95-1，pp1-8（1995）
- [5] 長尾真編：自然言語処理，岩波書店，pp187-188（1996）