

意味と親和性のある統語構造を出力する日本語文パーザ

武本 裕

宮崎 正弘

新潟大学大学院自然科学研究科

1 はじめに

三浦文法は、時枝誠記により提唱され三浦つとむにより発展的に継承された言語過程説に基づく日本語文法である。学校文法風の係り受け解析では意味との整合性に問題があるような場合にも、三浦文法を採用することで入れ子構造を用いて意味的にも正しい統語構造を表現できることが多い。本稿では、三浦文法に基づく文法記述を行なった、意味と親和性のある統語構造を出力する日本語文パーザについて述べ、その有効性を論じる。

2 三浦の言語モデルによる文の基本構造

時枝誠記 [1] によれば、言語は対象 認識 表現という過程的構造をもち、言語表現では対象のあり方がそのまま表現されるのではなく、話者の認識を通して表現されている、と説明している (言語過程説)。

また、三浦つとむ [2] は時枝の言語過程説およびそれに基づく日本語文法体系を継承し、言語の意味は表現自体が持っている対象と認識との間の客観的な関係であるとした (関係意味論)。

言語表現を行なうとき、話者は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を関係付ける。この対象認識世界に主体的な判断を加えることによって認識構造が形成される。またそれを一つの世界として捉え直し、次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ことができる。

時枝によれば、言語は話者の主観的な判断を直接に表現する主体的表現の語と、対象を概念化して表現する客体的表現の語に分かれる。文は主体的表現が客体的表現を包み込んだ形の入れ子型構造で表される (図 1)。

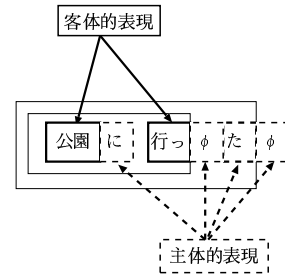


図 1: 主体的表現と客体的表現により構成される入れ子型構造

3 三浦文法に基づく日本語品詞体系

三浦文法に基づいた構文解析を行なうにあたって、日本語の品詞体系もまた言語過程説に基づく日本語品詞体系 [3] を採用する。その特徴としては、単語を詞 (客体的表現) と辞 (主体的表現) に分類している。客体的表現の語は、名詞、動詞、形容詞、連体詞、副詞、接辞。主体的表現の語は、助動詞、助詞、陳述副詞、接続詞、感動詞である。

また、学校文法との違いとしては、

- 接続助詞「て、で、たり」は既定判断の助動詞「た、だ」の連用形である
- 形容動詞は状態名詞 + 肯定判断の助動詞「だ」 / 格助詞「に」である

等があげられる。

4 拡張型 SGLR パーザ

構文解析システムには、一般化 LR 法を Prolog 上に実装した SGLR パーザ [4] を拡張した SGLR-plus [5] を用いた。拡張された機能としては痕跡処理、グラフ構造入力などがあげられる。本稿の議論は、特定のパーザに依存するものではないが、検証の際にはこれを利用した。

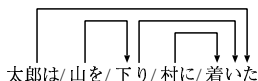
5 構文解析への三浦文法の適用

学校文法的な解析では、多義の爆発を防ぐために基本的に係り受け交差を禁止し、係り先も単一である。しかし、これでは結果が出力されても、その結果が意味的に適切でない統語構造の場合がある。以下で、そのような場合に三浦文法に基づいた入れ子構造を導入する有効性について述べる。また、実際にパーザを動作させるためには、文法をどのように記述し、統語構造をどのように表現すればよいかについて述べる。

5.1 一对多・多対一の係り受け関係

「太郎は山を下り、村に着いた」という文の場合、「太郎は」の節は直後に係るだけでなく、文全体に係る。さらに、助動詞「た」は直前の「着い」から係るだけでなく、文全体から係る(図2)。このように、複数の節からの係り受けが発生する例がある。その場合、入れ子構造を導入することで自然に表現できるようになる。

(a) 係り受け解析



(b) 三浦文法風

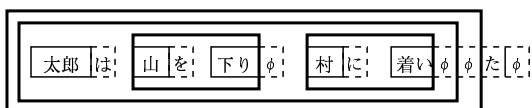


図2: 一对多・多対一の係り受け関係

三浦文法に基づいた入れ子構造の導入により、文節間の係り受けでは表現できない場合も解析できるようになるが、その分構造的多義の増加が予想され、その絞り込みが必要となる。

その際には、一つの方法として経験則を導入するというものがある。基本的には係り受けは直近を優先とするが、例外的に、助詞「は」は、広いスコープを持ち、助動詞「た」は、入れ子構造に対応して、それ以前の部分全体から係るというものである。

5.2 2種類の「は」

「彼は金はないが、アイデアはたくさん持っていた」を例にあげる(図3)。「は」には、2種類の「は」があ

り、スコープが異なる。「彼は」の「は」は、いわゆる主題の「は」である。これは、広いスコープを持ち、「金はない～持ってい」に係る。「金は」および「アイデアは」の「は」は、対で用いられる対照の「は」である。それぞれ、対となる部分に限定して係る。つまり、「金は」は「無いが」、「アイデアは」は「たくさん持ってい」のみに係る。このように、「は」は、用法によりスコープが異なるため、それぞれのパターンに応じて対処する必要がある。

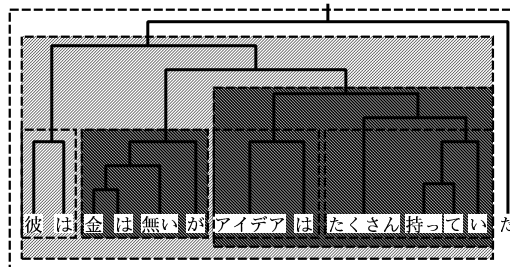


図3: 2種類の「は」

5.3 文中に局所的な入れ子構造をもつ文

「魚を釣りに行く」という文を例にあげる。ここで「釣り」は、内側の世界では格要素を取る動詞であると同時に、外側の世界では名詞の一部としての役割を持つというように、一文中で二つの品詞性を持つことになる(図4)。

このような場合でも、入れ子構造を想定すれば、自然な形でこれを導入することができる。

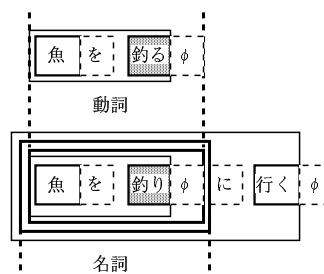


図4: 二つの品詞性のある語の取扱い

用言性(サ変名詞、連用名詞)に関しては二つの品詞性があるため、それに対応した文法が必要である。

文法を追加することにより構造的多義が増加するため、用言性(サ変名詞、連用名詞)とそれ以外の名詞を区別して記述す

ること、格パターンによって制約を行なうことが必要である。

図5は、「うなぎを食べに浜松に行く」の解析例である。2通りの解析結果を示した。図5は、三浦文法に基づく入れ子構造を導入したもので、「うなぎを」が「食べ」に係って局所的な構造を作っている。図6は、「うなぎを」が、「行く」に係るもので、意味的には適切でない。

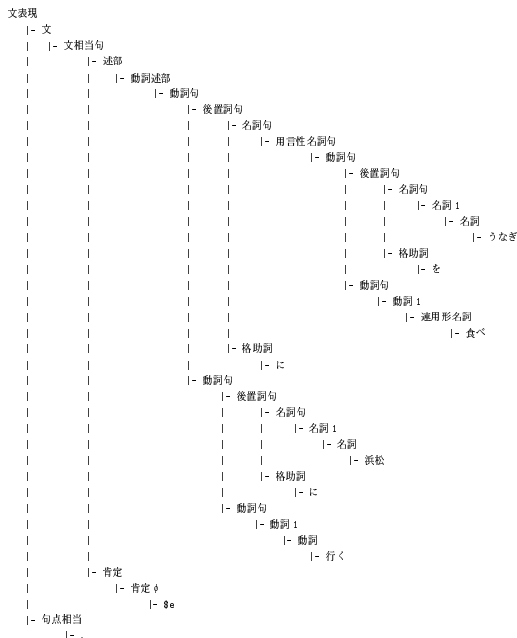


図 5: 「うなぎを食べに浜松に行く」の解析例 1

また、「尾張屋の在庫が潤沢な秘密はこれですよ」という文 [7] を例にあげる。この文中の「な」は、三浦文法では形容動詞連体形の活用語尾ではなく、断定の助動詞「だ」の連体形である。「潤沢な」が「秘密」を連体修飾しているのではなく、「尾張屋の在庫が潤沢」全体が「秘密」に係っている (図 7)。

これも三浦文法の入れ子構造でうまく処理できる例である。

5.4 陳述副詞による呼応

「決して読まない」のように陳述副詞による呼応を含む文では入れ子破りが発生する (図 8)。

この場合、単純に木構造を用いて表現することができないが、それに対する一つの手法として痕跡の考え方を導入する。

図 9 の例では、意味的には、「決して」が直後の「読

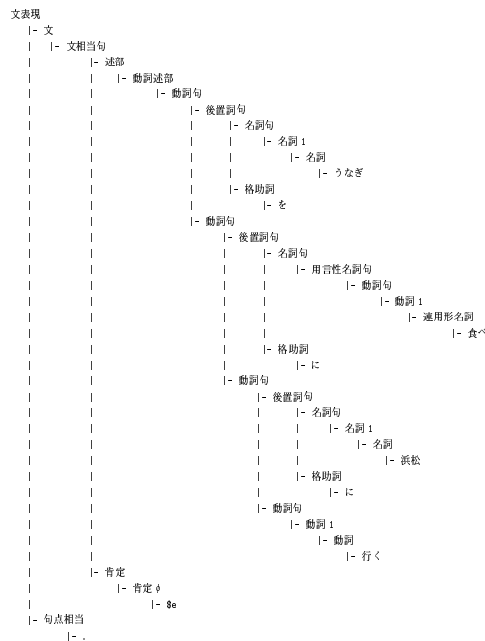


図 6: 「うなぎを食べに浜松に行く」の解析例 2

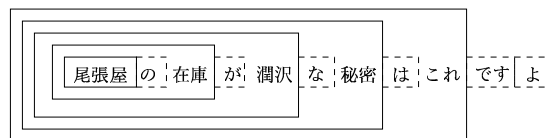
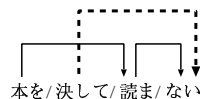


図 7: 断定の助動詞「だ」の導入

(a) 係り受け解析



(b) 三浦文法風

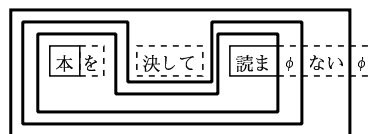


図 8: 陳述副詞の呼応

ま」ではなく、呼応の関係により「ない」に係っていると判断できることから、自然な係り受けの位置として、「読ま」と「ない」の間に痕跡として移動させる。このようにして、交差しない係り受け関係を構成することができる。この係り受け関係は、図 8(b) に対応している。

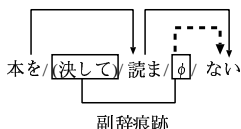
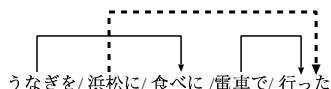


図 9: 痕跡を利用した入れ子破りへの対処

5.5 格要素の移動

「うなぎを浜松に食べに電車で行った」という例の場合も入れ子破りが生じる。これは格要素が通常的位置から移動したものであると見なすことができる(図 10)。この場合にも、入れ子構造を用いて解析することができる。

(a) 係り受け解析



(b) 三浦文法風

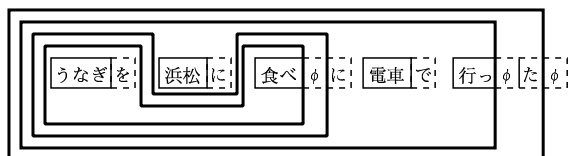


図 10: 格要素の移動

格要素が移動している文に関しては、痕跡を導入して交差なくとも係り受けが表現できる形に変形することになる。この場合、変形すべきかどうかは係り先と係り元の関係が適切かどうかということを意味的に判断することが必要となる。その際には、格パターンを利用することが考えられる。これにより、格要素と用言との関係が意味的に適切かどうかを判断することになる(図 11)。

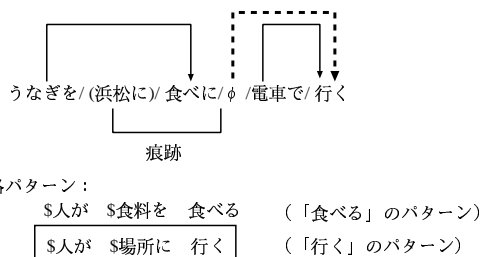


図 11: 格要素の移動への対処

6 おわりに

本稿では、意味と親和性のある統語構造を出力する日本語文パーザについて述べ、その有効性について論じた。本方式では、学校文法風の係り受け解析では困難な場合にも、多くの場合三浦文法に基づいた入れ子構造で表現できることを示した。文法の記述方法に関しては、その方針を一部示したが、今後もより詳細な検討が必要である。

参考文献

- [1] 時枝誠記:日本文法 口語篇、岩波全書 (1950).
- [2] 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).
- [3] 宮崎、白井、池原:言語過程説に基づく日本語品詞の体系化とその効用、自然言語処理 Vol.2 No.3、pp.3 ~ 25(1995).
- [4] 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).
- [5] 五百川、宮崎:痕跡処理のための逐次型一般化 LR パーザ SGLR の拡張、言語処理学会第 4 回年次発表論文集、pp.314 ~ 317(1998).
- [6] 藪、藤石、宮崎: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、言語処理学会第 3 回年次発表論文集、pp.205 ~ 208(1997).
- [7] 水谷静夫:意味・構文の関係を考へる九十例、計量国語学、Vol.19 No.1(1993).