

1 L-6

構文解析との部分融合による 日本語形態素解析の曖昧性解消法

柴山 威

宮崎正弘

新潟大学大学院自然科学研究科

1 はじめに

日本語形態素解析の曖昧性は単語候補を結ぶグラフ構造として表現される。従来、このグラフ構造を展開して得られる複数の候補からヒューリスティックスなどにより絞り込んだ一つの形態素解析結果のみを構文解析側に渡して解析していたが、このような絞り込みの方法では正しい解析結果に絞られるとは限らず、絞り込みの方法として十分とは言えない。

そこで、本稿では形態素解析結果の曖昧性を持ったグラフ構造をオブジェクト指向パーザ **POWER**[1] に入力し、構文解析と並行して形態素解析の曖昧性の解消を行なう。構文解析の段階で発見された格関係のある名詞と動詞や係り受けなどの「関係」に着目して、各々の関係について接続の確からしさの評価を行なう。そして、この評価を基に曖昧性のある候補に対して絞り込みを行なう方法を提案する。

2 POWER の解析機構

POWER はオブジェクト指向の考えを用いた、従来の句構造解析とは根本的に異なる日本語構文解析システムである。**POWER** は、それ自体自律的に働く単語オブジェクトの相互メッセージ通信によって解析を進めるパーザである。単語を単なる記号ではなく自律的な実体（オブジェクト）として扱っている。

Disambiguation of Japanese Morphological
Analysis by partial Integration of Morphological
Analysis and Syntactic Analysis
Takeshi Shibayama, Masahiro Miyazaki
Niigata University

3 構文解析との融合

日本語形態素解析の曖昧性は単語候補を結ぶグラフ構造として表現される。形態素解析結果のグラフ構造を直接パーザに入力し、構文解析と並行して曖昧性の解消を行なう。また **POWER** は形態素解析や意味解析を統合した解析を行なうことを目標としており、現段階でも一部に意味情報を解析に利用している。このため、必要ならば曖昧性の解消に意味情報も用いていく。形態素解析結果の曖昧性解消に **POWER** を用いることの利点として、以下の点が挙げられる。

- 係り受けなどの構文規則を用いることにより、より正しい曖昧性解消が行なえる。
- **POWER** のメッセージ通信機能を用い、グラフ構造を展開することなく処理を行なえる。

4 POWER の解析結果

図1のように形態素解析結果に曖昧性があり複数の候補が得られた場合、**POWER** の解析で得られた「関係」に着目し、これらの各々の「関係」によって確からしさの評価を行ない、評価値の比較によって曖昧性の除去を行なう。解析結果は図2に示すように、格関係や修飾関係などの関係の集合として表される。

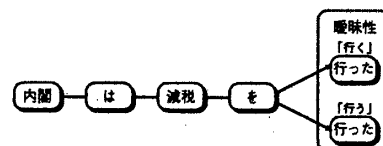


図1: 曖昧性のある形態素解析結果

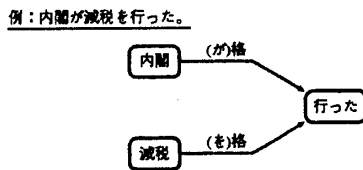


図 2: POWER の解析結果の表現

5 格関係による動詞の評価

格関係を用いた曖昧性の解消は図 3 に示す手順で行なわれる。名詞・動詞の格関係を発見すると、格パターン辞書 [2]・慣用表現データベースを参照し格チェックを行ない、各々の動詞へ格関係の数から確からしさの評価値を与える。

曖昧性の候補に対して与えられた評価値から絞り込みを行なう。ここでは、優先順位づけを行ない、明らかに誤りの候補については削除をする。

例えば図 2 の解析例では、動詞「行った」が「おこなった」か「いった」なのかの曖昧性が生じる。この場合は「減税」の(を)格の関係があるので、格チェックにより「おこなう」に高い評価が与えられる。

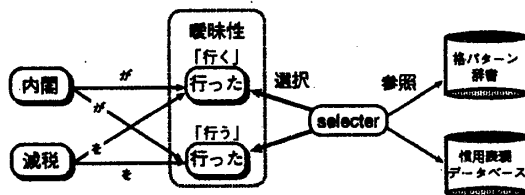


図 3: 格パターンチェックによる曖昧性の解消

5.1 慣用表現データベース

格パターン辞書による格チェックでは、制限を広くすれば絞り込みの手段としては有効性が無くなってしまい、厳しくすると「嫁に行く」など慣用的な表現などを認めないシステムとなってしまう。

これを補うため、慣用的な表現などはデータベース化しておき、格パターンチェックの評価手段とする。

5.2 評価方法

格チェックを通り、動詞に接続を認められたそれぞれの係り受けに重みづけを行ない、各々の動詞に接続する係り受けの重みの合計から、その動詞の確からしさの評価を得る。

6 現状と今後の課題

現状では、係り受けの評価を行なう格チェックの手法に関しては、検討中であるが、重みづけを行わず、単純に認められた格関係の数の比較だけでもある程度絞り込みの効果が期待できるものと思われる。また、格パターンチェックを行なう上で単語の意味の曖昧性によって、予期しない格パターンが通ってしまう事があるが、これは動詞と格関係にある複数の名詞間の関係から意味の絞り込みを行なう事によって防ぐことが可能であると思われる。

7 おわりに

本稿ではオブジェクト指向パーザ POWER 上で日本語形態素解析の曖昧性を解消する手段として、動詞の係り受けなどの「関係」を用いて動詞の候補の評価を行なう方法を提案した。今後は意味の曖昧性を絞り込む方法の検討を進め、実装とともに評価を行なっていく予定である。

謝辞

格パターン辞書として「一般パターン対辞書」を提供して下さった NTT コミュニケーション科学研究所の関係者各位に深謝いたします。

参考文献

- [1] 高橋、宮崎:オブジェクト指向パーザ POWER 情報処理学会第 51 回全国大会 No.4H-8(1995)
- [2] 池原、宮崎、横尾:日英機械翻訳のための意味解析用の知識とその分解能 情報処理学会論文集 Vol.34 No.8,PP.1692-1704(1993)