

オブジェクト指向パーザ POWER における

5 B - 8

日本語形態素解析の曖昧性解消機構

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、本稿では曖昧性をグラフ構造として持つ形態素解析結果をオブジェクト指向パーザ POWER[1] に入力し、POWER の持つメッセージ通信機能を利用して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）としてとらえたところが既存のシステムと大きく異なる点である。

2.1 オブジェクトとクラス

POWER ではオブジェクトと呼ばれる基本単位を用いて解析を進める。オブジェクトは基本的には

Disambiguation of Japanese Morphological Analyse with the Object-oriented Parser, POWER.

Takeshi Shibayama, Masahiro Miyazaki
Niigata University

単語に相当する。各オブジェクトはそれ固有の手続きと変数を持ち、メッセージにそれぞれのやり方で反応する。オブジェクトの記述は一連のメソッド（メッセージに対応する手続き）の記述からなる。また、同じような文法的な働きをするオブジェクトを名詞や動詞の様なクラスとして分類し、それぞれクラス毎にメソッドを記述する。

2.2 メッセージ通信

POWER は基本的に隣接のオブジェクトとのみメッセージ通信できる。隣接メッセージ通信は、POWER で一般的に使われる通信方法であり、メッセージの相手に隣接を指定すると、そのオブジェクトと隣接関係にあるオブジェクトに対しメッセージが送られる。また、離れたオブジェクトに対して通信を行なうためには、途中の単語に中継してもらう必要がある。

2.3 解析結果

解析結果は非交差などの制約のない二項関係の集合で表現される。

3 POWER と形態素解析の融合

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。従来、構文解析を行なう前に、ヒューリスティックスなどにより曖昧性を取り除いた形態素解析結果を構文解析側に渡していたが、この方法では正しい解析結果が除去される可能性があり有効な方法とは言い難い。そのため、POWER では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 は意味解析を組み込みうる柔軟性を持ち、将来的には意味解析による絞り込みも可能であるので、構文解析レベル

では絞り込めない曖昧性については絞り込みを行わず、さらに意味解析まで曖昧性を持ち越す。

形態素解析結果の曖昧性解消に **POWER** を用いることに利点として、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。
- 複雑なグラフ状のオブジェクト間のデータの授受に、**POWER** のメッセージ通信機能を用いることができる。
- 構文解析と並行して曖昧性の解消を行なうので、グラフの展開による処理量の爆発的増加を抑えられる。

4 曖昧性解消法

POWER による曖昧性の解消の方法は、構文規則に不適格な候補の除去と path の優先順序の判断による選択的な曖昧性の除去がある。

4.1 構文規則による曖昧性解消

POWER ではオブジェクト間のメッセージの送受によって解析が行なわれるが、このメッセージ通信が失敗した場合、図 1 に示すように不適格な候補と判断し除去する。例えば、連用格助詞「が」の後には格接続を受ける動詞や助動詞が来なければならない。もし、存在しなかった場合はこの候補を含む Path は形態素解析のグラフから除去する。このように、まず明らかに間違っている候補の除去を行なう。

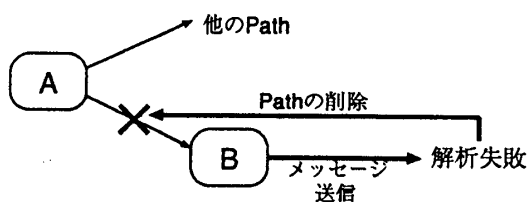


図 1: 文法に矛盾する Path の削除

4.2 優先順序の判断機構

図 2 に示すように複数の Path からメッセージを受けとったオブジェクトは、それぞれの Path の情報を持ったオブジェクトに分裂し、個々のオブジェクトで処理を行なう。セレクタはこの分裂したオブジェクトに関して、どの Path が優先されるかを判断し不要と判断した Path を削除する。このように並列に存在する複数の Path から候補を絞る。

優先順序の判断基準としては現在のところ以下に挙げたもの考えている。

格接続

格接続を受けるオブジェクト（動詞など）に注目し、そのオブジェクトに自然な接続が成立する。

係り受けの成立

修飾関係などの係り受けがある。

動詞の成立

格接続を受ける動詞が成立する。

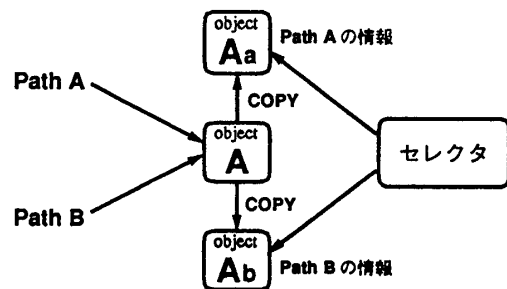


図 2: セレクタによる優先順序の判断

5 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方法を提案した。文法面も含め、現在のところ曖昧性の解消機構はまだ実装作業中であるので、今後は解消機構の実装とともに具体的な曖昧性の評価規則の作成をする予定である。

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)