# 局所・大域的制約による構造的曖昧性抑制機構を持つ日本語文パーザ

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

日本語の構文解析において、解析対象の文が長くなるにつれて構造的曖昧性は増加していく。この構造的曖昧性を適切な手段により抑制し、正しい解析木が得られるようにする必要がある。そのため、構文解析の前処理・後処理を組み合わせて最終的に出力される構文木の数を絞り込む。

まず、前処理の段階では、特定の語をキーとして事前に節候補を抽出しておくことで係り受けを局所化して解析時に発生する構造的曖昧性を抑制する。

さらに、構文解析の後処理として、構文木の作成後に用言の必須格や局所・大域的な呼応の関係などを利用して係り受けの適切性に対してコスト付けし、優先順位を与える。この優先順位に基づき最終的に構造的曖昧性を絞り込む。

本稿では局所・大域的制約による構造的曖昧性抑制機構を持つ日本語文パーザを試作し、その有効性を検証した結果について述べる。

## 2　構文解析システムの概要

入力文をまず日本語形態素解析システム Maja[1] で解析し、それを拡張型チャートパーザ Schart[2] に日本語文法を実装した Schart-J で構文解析する。構文解析の前後には構文解析前処理と構文解析後処理を行う。言語過程説 [3][4] に基づく日本語品詞体系 [5] および文法を採用した。

日本語構文解析システムの流れを図 1 に示す。各部の詳細を以下で述べる。

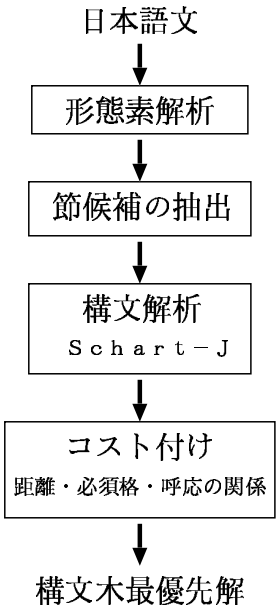


図 1: 日本語構文解析システムの流れ

## 3　構文解析前処理

### 3.1　節候補の抽出

ここでは、文を構成する単文および埋め込み文を含む複文相当の節に着目する。文をこのような節に分割することができれば、節の内部の係り受けは単純であるので、内部を先に構造化し、次に節同士の係り受けを決定すれば全体の構造が得られる。このように、抽出した節内で優先的に構造化することにより、構造的曖昧性を抑止する。節を抽出する際の節候補の判定基準について以下に示す。

### 3.2　節候補の判定基準

この判定基準は [6] を参考にしている。

以下に示す語の後ろ (右) を節境界とする。

- 節の接続語となるもの
  - 連用中止形の動詞、動詞型接尾語、形容詞、形容詞型接尾語
    例：山を 下り/村に着いた。
  - 連用中止形または仮定形の助動詞
    例：ちょっと横になったら/疲れが取れた。
  - 接続助詞、接続助詞相当の形式名詞
    例：ちょうど私が出る とき/彼が着いた。
- 遠くに係りやすいもの
  - 副助詞「は」
    例：彼 は/疲れていたが、彼女を手伝いに行った。
  - 文頭の副詞型名詞
    例：来月/沖縄へ行く予定だ。

また、以下の語の前 (左) を節境界とする。

- 節全体に係る
  - 節の切れ目となる助動詞
    例：台風が通過し /て ようやく風雨が収まった。

## 4 構文解析中の節候補による制約

構文解析処理の内部で部分木を構成する時点で、前述の節候補に基づいて、節の内部で優先的に係り受けを決定する。その上でそれぞれを組み上げる。

チャート法における不活性弧を監視して、節が完成する前の他の節からの係り受けを抑制する。

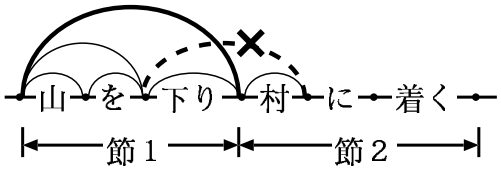


図 2: 節候補による制約

図 2 において、「山を」と「下り」から「山を下り」が完成する段階では、節の内部であるので許可されるが、「下り」と「村」は節が完成しないうちに他の節の弧同士が結合しようとしているため係り受け禁止となる。

## 5 節同士の係り受け

前処理で抽出した節が 3 つ以上になる場合には、節同士の係り受けに関して構造的曖昧性が生じる。そこで、接続優先度という指標を導入して節同士の係り受けを決定する。これは [6] を参考にした。接続優先度を表 1 に示す。節間で接続優先度を比較し、自分より優先度が低い節であればその上を飛び越えるが、自分より優先度が高ければ飛び越えないものとして係り受けを決める。

表 1: 接続優先度

| 接続優先度 | 節の分類 |
|---|---|
| 7 | 主節 |
| | 「展開」の接続助詞＋読点 |
| 6 | 「展開」の接続助詞 |
| | 副助詞「は」＋読点 |
| 5 | 「条件」の接続助詞＋読点 |
| | 連用中止形＋読点 |
| | 用言の仮定形＋読点 |
| | 体言止め＋読点 |
| | 副助詞「は」 |
| | 格後置詞句＋読点 |
| | 副詞句＋読点 |
| 4 | 「条件」の接続助詞 |
| | 連用中止形 |
| | 用言の仮定形 |
| | 体言止め |
| 3 | 「同時」の接続助詞＋読点 |
| 2 | 「同時」の接続助詞 |
| | 格後置詞句 副詞句 |
| | 形式名詞＋の |
| | 形式名詞に係る「名詞＋の」 |
| | 名詞＋読点 |
| 1 | 名詞＋「の」（優先度２のもの以外） |
| | 連体詞 |

## 6 構文解析後処理における構文木の絞り込み

格要素と用言 (あるいは名詞述語) の間の係り受けに曖昧性がある場合にそれぞれの係り受けに対してコストを与え、最終的な絞り込みに利用する。前処理の導入により構造的曖昧さが抑止されるため、絞り込み対

象となる構文木の数は少ない。

### 6.1 係り受けの距離による制約

「格要素は最も近い用言に係りやすい」という経験則を制約として用いる。つまり、格要素が用言に係るとき、距離が「近い」係り受けを「遠い」係り受けよりも優先する。

ただし、副助詞「は」などが作る後置詞句に関しては遠くに係りやすいという性質を持つため例外的に扱う。

### 6.2 必須格に基づく制約

用言に関して必須格として一般にどのような格をとりうるかが知られている。これを一つの用言を中心にどのような組合せが存在するかをまとめたのが格パターンである。格パターンでは格要素に対し意味カテゴリによって制約している。ただし、意味カテゴリによる制約は緩すぎたり、厳しすぎたりして不完全であるためここでは用いない。

ここでは、格要素が係り先の用言の必須格を満たすかどうかを調べる際に、格助詞の字面のみ (「が」,「を」,「に」等) を使用する。必須格を「満たす」係り受けを「満たさない」係り受けよりも優先する。

### 6.3 呼応の関係に基づく制約

呼応の関係とは文中である語が先行した場合に、ある決まった語が要求されるという関係のことである。ただし、呼応の強さには確実性が非常に高いもの (強い呼応) と比較的高いもの (弱い呼応) とがある。

まず、単文中の局所的制約として、格要素の主名詞が特定の字面の語である強い呼応 (油を売る) や弱い呼応 (辞書を引く) がある。そして、大域的制約として、「決して〜ない」のような陳述副詞による強い呼応や「〜とは〜だ」のような弱い呼応がある。

ここでは特に格要素と用言との関係で、格要素が係り先の用言と呼応の関係にあるかどうか調べ、呼応が「成り立つ」係り受けを「成り立たない」係り受けよりも優先する。

## 7 最優先解の選択

上に示した、距離による制約・必須格による制約・呼応の関係による制約に基づいてそれぞれの解析木にコストを与える。複数の制約が適用される場合には、そのコストの総和を求める。このコストが最小となる順に優先順位を与える。

## 8 検証

以上の内容に関して、拡張型チャートパーザに日本語文法を実装した Schart-J を用いて検証を行った。

### 8.1 節候補抽出の効果

- 平均文字数 19.4 字
- 重文・複文が各 50 文、合計 100 文

を例文として節候補抽出の効果を確認した。節候補抽出前処理を行うものと行わないものとで平均結果数を比較した。また、前処理を行った場合に正解が含まれるがどうか調べた。実験の結果を表 2 に示す。

表 2: 節候補抽出による平均結果数の変化と正解含有率

| | 前処理なし | 前処理あり | 正解含有率 (%) |
|---|---|---|---|
| 重文 | 13.0 | 4.8 | 95.8 |
| 複文 | 21.1 | 5.9 | 94.6 |
| 全体 | 17.1 | 5.4 | 95.3 |

正解含有率 (%) = (前処理後に正解を含む文数)/(前処理なしで正解を含む文数) * 100

節候補抽出前処理の結果、平均結果数が減少していることが確認できた。

### 8.2 制約を用いた最優先解の選択

具体的に例文と構文木を示してその効果を確認する。

#### 係り受けの距離

例として「うなぎを食べに行く。」について示す。

以下の係り受けの可能性がある (図 3)。

- うなぎを → 食べ (距離 1)
- うなぎを → 行く (距離 3)

距離が近いものを優先としているため、
「うなぎを → 食べ」が優先となる。
最終的には点線で示した木構造が最優先解となる。

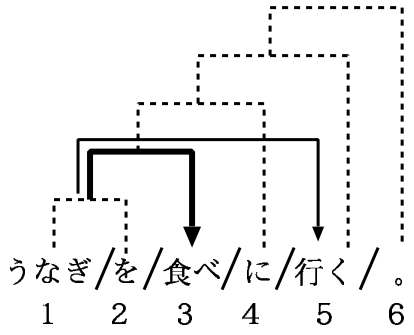


図 3: 係り受けの距離による制約

**必須格**

例として「懸賞金を受け取る権利がある。」について示す。

文中の用言 (受け取る, ある) の必須格を示す。

- 受け取る → が、を、から
- ある → が、に、から、まで

以下の係り受けの可能性がある (図 4)。

- 懸賞金を → 受け取る (必須格)
- 懸賞金を → ある

必須格を満たすものを優先としているため、
「懸賞金を → 受け取る」が優先となる。
最終的には点線で示した木構造が最優先解となる。

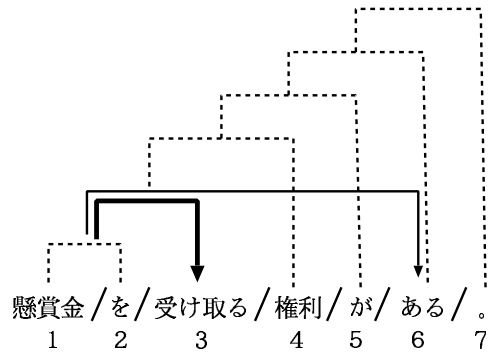


図 4: 必須格による制約

このように係り受けの距離・必須格等の制約を用いることで構造的曖昧性がある場合に最優先解を選択できる。

## 9 おわりに

本稿では局所・大域的制約による構造的曖昧性抑制機構を持つ日本語文パーザを試作し、その有効性を検証した。

今後は、曖昧性の抑制の効果および正解率に関して大量の文で定量的に評価する予定である。

## 参考文献

[1] 高橋、大川、尾嶋: 日本語形態素解析システム Maja、
http://www.nlp.ie.niigata-u.ac.jp/nlp/maja/

[2] 川辺、宮崎: 構造を含む生成規則を扱える拡張型チャートパーザ – Schart パーザの実装 –、言語処理学会第 11 回年次発表論文集、pp.911〜914(2005).

[3] 時枝誠記:日本文法 口語篇、岩波全書 (1950).

[4] 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).

[5] 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).

[6] 須田、宮崎: 大域的制約を利用した構造的曖昧さの抑止機構を持つ日本語文パーザ、言語処理学会第 10 回年次発表論文集、pp.357〜360(2004).