

# 単一化文法を用いた日本語文の構文解析における並列構造の処理

中村 健<sup>†</sup> 乙武 北斗<sup>‡</sup> 田辺 利文<sup>‡</sup> 吉村 賢治<sup>‡</sup>  
 福岡大学大学院 工学研究科<sup>†</sup> 福岡大学 工学部<sup>‡</sup>

## 1. はじめに

並列構造は1文中に同等の機能を持つ複数の単語や句を並べたものである。並列構造は文章の冗長性を無くし、簡潔に記述することに役立っている。しかし、並列構造は構文解析を困難にする要因の1つとなっている。

本稿では、日本語句構造文法(JPSG)<sup>[1]</sup>に基づく単一化文法を用いた構文解析<sup>[2]</sup>における並列構造の処理方法を提案する。なお、本稿の入力文は事前に形態素に分割されたものとする。

## 2. 並列構造

並列構造は次のような特性を持つ<sup>[3][4]</sup>。

- (1). 句を接続する働きを持つ文字列(並列キー)が存在する。
- (2). 並列キーの前と後に同じ形式の文要素(並列要素)が存在する。
- (3). 並列要素の単語数は、必ずしも同一ではない。
- (4). 文から並列キーと片方の並列要素を取り除いても意味が通る。

また、並列構造には次のような種類がある。

- (a). 名詞並列: 名詞や名詞句の並列
- (b). 述語並列
  - : 連用修飾語を共有する述語の並列
  - : 連用修飾語を共有しない述語の並列
- (c). 部分並列
  - : 複数の連用修飾語や名詞・名詞句からなる並列

さらに、並列構造には並列が入れ子になるもの(d)や3個以上の並列要素が接続するもの(e)が存在する。

- (d). 太郎が{(コーヒーに砂糖)、(水を氷)に入 re} {客に提供 s} ru
- (e). 太郎が(コーヒーに砂糖)、(水に氷)、(牛乳に蜂蜜)を入 re ru

## 3. 単一化文法を用いた構文解析

構文解析は CYK アルゴリズムで行う。『犬が吠える』という文を解析すると図1のような解析表が得られる。図1のセル中の数値はそのセルに存在する要素の数を表しており、各要素には素性構造とその他の構文情報が格納されている。

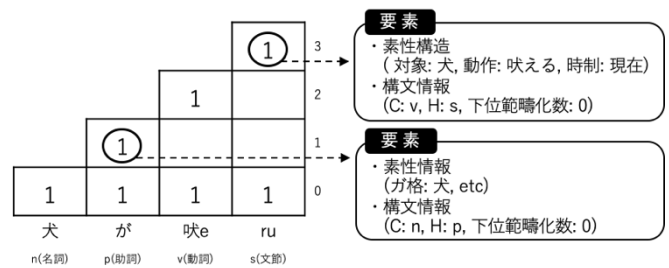


図1. 解析表

## 4. 並列構造の処理

構文解析を行なった結果、解析表の右端最上位セル(図1では(3,3))に要素が得られない場合、解析に失敗したとして、並列構造の処理を試みる。

この処理は文に対して、(1)並列要素の推定と(2)並列構造の分解を行うことで、図2のように文の並列構造を分解し、単文の生成を行う。



図2. 単文の生成

文に複数の並列構造が存在する場合、並列構造を分解する順番を考慮する必要があるが、並列キーが無くなるまで(1)、(2)の処理を再帰的に行うことで、同様に並列構造を分解し、単文を生成することが可能である。並列構造の処理の流れを図3に示す。

```

解析対象: target/ 単文群: result[]/出力: result_list[][]とする。
target = 入力文
for(全ての並列キーの解析順番)
| while(targetに並列キーが存在する)
| | [(1)並列要素の推定](任意の並列キーに対して)→ parallel[]
| | [(2)並列文の分解](parallel[]を用いて)
| | | (2)-1: result[]内の文に存在する並列構造の除去
| | | (2)-2: 並列構造を含むtarget文を分解
| | | a. 並列キーと後の並列要素を取り除いた文
| | | b. 前の並列要素と並列キーを取り除いた文
| | | bの文: result[]に追加, aの文: targetにセット
| result[]にtargetを追加, result_list[][]にresult[]を追加する。
result_list[][]を返り値として終了。
    
```

図3. 並列構造の処理

この処理が対象とする並列構造は、単一化文法の枠組みで処理ができない、名詞並列以外の並列構造である。

### (1) 並列要素の推定 (図3. (1))

並列要素の推定には、図4のような解析表とセルに登録されている構文情報(素性構造の形式で構成要素などが記述されている)を用いる。

**Processing of Coordinate Structure in Syntactic Analysis of Japanese based on Unification Grammar**  
 Ken Nakamura<sup>†</sup>, Hokuto Ototake<sup>‡</sup>, Toshifumi Tanabe<sup>‡</sup>,  
 Kenji Yoshimura<sup>‡</sup> Grad. Sch. of Electronics Eng. &  
 Computer Sci, Fukuoka Univ.<sup>†</sup>Dept. of Electronics Eng. &  
 Computer Sci, Fukuoka Univ.<sup>‡</sup>

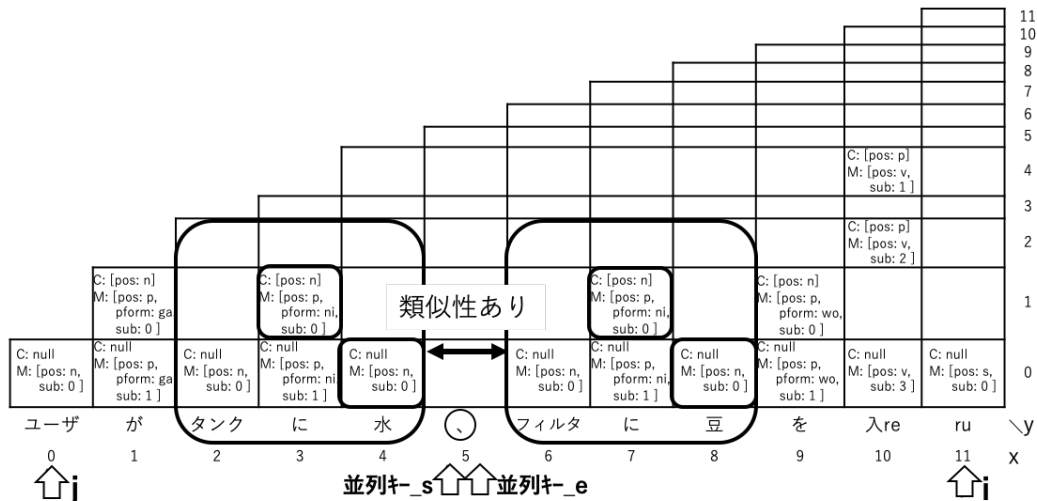


図4. 解析表

並列キーを中心とした、前後の任意の範囲の文要素において、それぞれが持つ部分木の根の類似性を調べ、類似性のある組を検出する。部分木の根は、その文要素で構成可能な部分木の最上位階層である。(例:図4の『タンクに水』の部分木の根は、『タンクに』が形成する座標(3,1)と『水』が形成する座標(4,0)である)。類似性は任意の構文情報同士が単一化可能であるかを調べ、単一化が可能であれば類似性があると判定し、不可能であれば類似性がないと判定する。並列要素の推定の流れを図5に示す。

```

並列要素推定結果: parallel[]
for (j=0; j<並列キー_s; j++)
  | for (i=解析表の大きさ-1; 並列キー_e<i; i--)
  | | if(類似判定(j から[並列キー_s-1]の部分木の根,
  | | | [並列キー_e+1]から i の部分木の根) == True)
  | | | then <座標(j, (並列キー_s-1))と 座標((並列キー_e+1),i) >
  | | | | | をparallel[]に追加
return parallel[]
    
```

図5. 並列要素の推定

図4の文に対し、並列要素の推定を行うと、『タンクに水』と『フィルタに豆』が並列要素であると推定する\*1。

(2) 並列構造の分解 (図3.(2))

並列構造の分解は、(1)並列要素の推定の結果に基づいて、並列キーとそれぞれ並列要素を取り除いた、2つの文(a)と(b)の生成を行う。(図3.(2)-2)

- (a) ユーザがタンクに水を入 re ru
- (b) ユーザがフィルタに豆を入 re ru

そして、文(b)は単文リスト(図3.の result)に追加し、文(a)も文中に並列構造が無ければ result に追加し、処理を終える。

文(a)に並列構造が含まれている場合は、文(a)に対して再度並列構造の処理を行う。

さらに、並列構造の分解の過程で次の処理を追加することで、resultに残っている並列構造を取り除く。(図3.(2)-1)

- (i) result内の文に[並列キーと後の並列要素]が全て存在する場合、その文から[並列キーと後の並列要素]を取り除く。
- (ii) result内の文に[前の並列要素と並列キー]が全て存在する場合、その文から[前の並列要素と並列キー]を取り除く。

5. おわりに

本研究では、単一化文法を用いた構文解析を開発し、本稿の並列構造の処理の導入を行なった。そして、単一化文法の枠組みで対応できなかった「述語並列」、「部分並列」、「並列の入れ子」、「並列の接続」の並列構造を含む文に対し、検証を行なった。その結果、期待した並列要素の推定結果を得ることができ、その推定を基に並列構造を含む文を分解し、単文の生成ができることを確認した。

6. 謝辞

本研究は文科省科研費(JP17K00116)の助成を受けたものである。

参考文献

- [1] 三吉, 郡司, 白井, 橋田, 原田: 日本語句構造文法-JPSG, コンピュータグラフィクス, Vol. 3, No. 4 (1986)
- [2] 吉村: 日本語単一化文法による形態素解析と構文解析の融合, 福岡大学工学集報, pp. 15-21 (2017)
- [3] 首藤, 吉村, 津田: 日本語技術文における並列構造, 情報処理学会論文誌, Vol. 27, No. 2, pp. 183-190 (1986)
- [4] 黒橋, 長尾: 長い日本語文における並列構造の推定, 情報処理学会論文誌, Vol. 33, No. 8, pp. 1022-1031 (1992)

\*1: 『水』と『フィルタ』も類似性ありと判定されるが名詞並列であるため無視するように設定している。