

述語動詞の構文特徴に着目した中国語構文解析における曖昧性抑制法

王向莉

宮崎正弘

新潟大学大学院自然科学研究科

1 はじめに

中国語構文解析において、曖昧性を解消することが大きな問題となっている。中国語は孤立語であり、英語のような動詞や形容詞のような語尾変化、日本語のような用言の活用や格助詞などがいないため、従来の文脈自由文法に基づいて記述した文法規則は非常に緩やかであり、構文解析を行う際に曖昧性が爆発的に生じる [1]。

中国語構文解析における曖昧性には、曖昧性を係り受けによる曖昧性（係り受け曖昧性）と構文要素の境界区切りによる曖昧性（境界曖昧性）との二種があり、係り受け曖昧性は境界曖昧性の発生とともに発生すること、および境界曖昧性が非常に顕著であることを明らかにすると共に構文パターンを文法規則として用いて境界曖昧性の解消を図り、構文木の曖昧性を絞り込む手法を提案した [2]。また、動詞をその構文特徴により分類し、構文パターンに動詞や形容詞、介詞、特殊文の文型等の中国語文の構文制約を記述した構文パターンを文法規則として用いることによって、構文的曖昧性を大幅に抑制することを可能としている [2]。

しかし、二つ以上の述語動詞を持つ文（連続動詞句述語文）に対しては、その複数の述語動詞間の構文的関係を明確にしていないため、他の動詞述語文のように分類された述語動詞を中心にし、構文パターンを記述することができない。それに対応する構文パターンによる構文的曖昧さの抑制がうまく機能しないため、曖昧性が生じる原因となる。

本稿では、述語動詞の構文特徴に着目し、動詞とその直後の単語との接続ルールを作成した。構文解析を行う前の段階でこれらのルールを用いて、動詞という品詞をもつ単語の多品詞性を一部絞り込み、曖昧性を抑制する手法を提案し、その有効性を論じた。

2 連続動詞句述語文構文パターンによる曖昧さ

表 1: 記号と名称の対応表

記号	構文要素
s	文
Sn	名詞性主語
tp	時間句
jf	否定判断辞
d	副詞
zv	助動詞
zy	「"地"字フレーズ」などの状語
uv1	動態助詞"了"
Cj	結果補語
Cf	方向補語
uj1	構造助詞"的"
Os	文目的語
On	名詞性目的語

表 2: 動詞の小分類の記号とその意味

述語動詞の記述方式	意味
v	名詞性主語をとれる述語動詞
v_Cj	名詞性主語、結果補語をとれる述語動詞
v_Cf	名詞性主語、方向補語をとれる述語動詞
v_On	名詞性主語、名詞性目的語をとれる述語動詞
v_Cf_On	名詞性主語、方向補語、名詞性目的語をとれる述語動詞
v_Cj_On	名詞性主語、結果補語、名詞性目的語をとれる述語動詞
vb	補助動詞となれる動詞
vf	方向動詞となれる動詞

現在の構文解析システムの入力は一語ごとにスペースで分ち書きされた中国語文であり、品詞の絞り込みをしていない。中国語では多品詞性（一つの単語が複数の品詞を持つこと）が顕著であるが、構文パターンを用いて構文解析を行う場合、単語の多品詞を絞り込まなくても構文木を大幅に抑制できる。

「遅刻した学生はみんな手を挙げた」という例文を見ると、辞書引きの結果は以下のように、多品詞のままである。

```
consulted: [[v:遅刻], [uj1:的], [n:学生],
[d:都], [v:挙, v_On:挙, v_Cf_On:挙, v_Cf:
挙], [v:起, vf:起], [v:了, v_On:了, vb:
了, y:了, uv1:了], [n:手], [pd:.]]
```

この文は以下の構文パターン

```
"(s Sn tp* jf* d* zv* zy* v_Cf_On Cf uv1* On)"
```

だけにマッチングするため、「挙」という単語は述語動詞(v_Cf_On)、「起」は「挙」の方向補語(Cf)である方向動詞(vf)、「了」は動態助詞(uv1)と判断され、構文木数は図1に示すように一つだけとなる。

構文パターンの中の記号とその意味との対応関係が表1と表2とに示す。「*」という記号は反復記号（非終端記号の一定回数の繰り返しを示す記号）である[3]。

しかし、連続動詞句述語文について、文の述部は複数の動詞句からなる。複数の述語動詞間の構文的関係を明確にしないため、動詞を中心にし、構文パターンを記述できない。それに対応する構文パターンを以下の記述のように、比較的緩やかである。

```
"(s Sn tp* jf* d* zv* zy* jf* vp vp)"
```

中国語では多品詞性の単語は動詞という品詞を持つものが多い。そのため、上記の構文パターンによって、曖昧性が生じる。「文章 修改 過 了（文章は修正されたのであった）」という例文を挙げる。図2では、no.001の構文木が正解で、no.002の構文木が誤りの構文木である。「過」という単語は動詞の品詞を持つため、「修改」と「過」をそれぞれ動詞とし、この文を連続動詞句述語文の構文パターンとマッチングし、誤りの構文木が生成された。

```
number of result > 001
* result no.001*
|-sen
  |-s
    | |-Sn
      | | |-np
        | |   |-cl
          | |   | |-v--< 遅刻 >
            | |   | |-uj1--< 的 >
              | |   | |-nf
                | |   | |-n--< 学生 >
                  | |   | |-d--< 都 >
                    | |   | |-v_Cf_On--< 挙 >
                      | |   | |-Cf
                        | |   | |-vf--< 起 >
                          | |   | |-uv1--< 了 >
                            | |   | |-On
                              | |   | |-np
                                | |   | |-nf
                                  | |   | |-n--< 手 >
                                    | |   | |-pd--< . >
```

図1: 「遅刻した学生はみんな手を挙げた」という文の解析結果

3 述語動詞の構文特徴と接続ルールによる多品詞の絞り込み

述語動詞とその直後の単語に着目すると、以下の特徴がある。

1. 述語動詞の直後にある「着」、「了」、「過」が動態助詞である。

中国語では「着」、「了」、「過」の三つの動態助詞がある。いずれも頻繁に使用され、しかも動詞の品詞を持つ。構文解析の結果から見ると、動態助詞が動詞とし、連続動詞句述語文の構文パターンにマッチングし、解析されるケースが多い。本構文解析システムの単語辞書では「着」、「了」、「過」それぞれ以下の品詞を持つ。

```
了 uv1/v/v_On/vb/y
着 uv2/v/v_On/vb
過 uv3/v/v_On/vb
```

その直前が動詞であるならば、品詞を動態助詞に絞り込む。

```
[[v_On] [v_Cf] [v_Cf_On] ... ] + 「了」
==>了 [uv1]
[[v_On] [v_Cf] [v_Cf_On] ... ] + 「着」
==>着 [uv2]
[[v_On] [v_Cf] [v_Cf_On] ... ] + 「過」
==>過 [uv3]
```

2. 結果補語と接続できる動詞の直後に補助動詞であれば、その単語の品詞を補助動詞とする。

述語動詞と結果補語となる構造は動補構造の一種である。結果補語になれる単語は形容詞のほか、数少ない動詞もある。結果補語になれる動詞は補助動詞である。補助動詞は普通の動詞でもある。「他都看見了(彼はすべてを目に入った)。｣という文では「見」は補助動詞(vb)であり、結果補語(Cj)である。「他見過我(彼は私に会ったことがある)」という文の中で、述語動詞(v_On: 名詞目的語と接続できる動詞)となる。

前述の特徴によって、結果補語と接続できる動詞の直後に結果補語となれる動詞であれば、この単語の品詞を補助動詞(vb)に絞り込む。

```
[[v_Cj] [v_On] ... ] + [[vb] [v_On] ... ]
==> [[v_Cj] [v_On] ... ] + [vb]
[[v_Cj_On] [v_On] ... ] + [[vb] [v_On] ... ]
==> [[v_Cj_On] [v_On] ... ] + [vb]
[[v_Cj_Os] [v_On] ... ] + [[vb] [v_On] ... ]
==> [[v_Cj_Os] [v_On] ... ] + [vb]
など
```

3. 方向補語と接続できる動詞の直後に方向動詞があれば、その単語の品詞を方向動詞とする。

述語動詞と方向補語となる構造も動補構造の一種である。結果補語になれる単語は「出来」、「出去」などわずかである。方向動詞と呼ぶ。方向動詞は述語動詞にもなれる。「他從教室里走出来了(彼は教室から歩き出た)」という文では「出来」は方向動詞(vf)であり、方向補語(Cf)である。しかし「他從教室里出来了(彼は教室から出てきた)」という文では「出来」は述語動詞である。前述の特徴によって、方向補語と接続できる

動詞の直後に方向補語となれる動詞であれば、この単語の品詞を方向動詞(vf)に絞り込む。

```
[[v_Cf] [v_On] ... ] + [[vf] [v] ... ]
==> [[v_Cj] [v_On] ... ] + [vf]
[[v_Cf_On] [v_On] ... ] + [[vf] [v] ... ]
==> [[v_Cf_On] [v_On] ... ] + [vf]
など
```

```
consulted: [[n:文章], [v:修改, v_On:修改],
[uv3:過, v:過, v_On:過, vb:過], [v:了, v_On:了, vb:了, y:了, uv1:了], [pd:.]]
```

```
number of result > 002
```

```
* result no.001*
```

```
| -sen
| -s
| | -s
| | | -Sn
| | | | -np
| | | | -nf
| | | | -n--< 文章 >
| | | -v--< 修改 >
| | | -uv3--< 過 >
| | -y--< 了 >
| -pd--< . >
```

```
* result no.002*
```

```
| -sen
| -s
| | -s
| | | -Sn
| | | | -np
| | | | -nf
| | | | -n--< 文章 >
| | | -vp
| | | | -v--< 修改 >
| | | -vp
| | | | -v--< 過 >
| | -y--< 了 >
| -pd--< . >
```

図 2: 前処理抜きの場合に「文章 修改 過了」という文の解析結果

4 曖昧性抑制に対する有効性

述語動詞の構文特徴によって、動詞とその直後の単語の接続ルールを作成し、プログラムに実装し、構文解析の前処理を行った。前文に挙げた例文「文章 修改 過 了（文章は修正されたのであった）」を前処理プログラムに通して、「過」という単語の品詞を図2に示すような [uv3:過, v:過, v_On:過, vb:過] の四つから、図3に示すような [uv3:過] に絞り込み、誤りの構文木を解消した。

```
consulted: [[n:文章], [v:修改, v_On:修改],
[uv3:過], [y:了], [pd:.]]
number of result > 001
* result no.001*
|-sen
  |-s
    | |-s
      | | |-Sn
        | | | |-np
          | | |   |-nf
            | | |     |-n--< 文章 >
              | | |       |-v--< 修改 >
                | | |         |-uv3--< 過 >
                  | | |           |-y--< 了 >
                    | | |             |-pd--< . >
```

図 3: 前処理を行った場合に「文章 修改 過 了」という文の解析結果

ここでは、既に論じた述語動詞の構文特徴が現れる6文に対して、前処理を適用する前後の構文木数の比較を表3に示す。

<例文>

1. 小王 發現 了 他 的 秘密
2. 車 慢 慢 地 開 着
3. 他 都 看 見 了
4. 他 打 敗 了
5. 他 從 教 室 里 走 出 來
6. 大 家 一 起 唱 起 來

表 3: 曖昧性抑制効果

文番号	適用前	適用後	正解を含むか
1	11	4	ある
2	3	1	ある
3	2	1	ある
4	3	2	ある
5	3	1	ある
6	2	1	ある

5 おわりに

本稿では、述語動詞の構文特徴に着目し、動詞とその直後の単語との接続ルールを作成した。これを前処理プログラムに実装し、その有効性を検証した。今後、動詞の品詞を持つ介詞、助動詞、方位詞などの他の品詞の絞り込みをどのように実現するか今後の課題となる。

参考文献

- [1] 楊頌明, 堂下修司, 西田豊明: 中国語解析システムにおけるヒューリスティックな知識の利用, 情報処理論文誌, Vol.25 No.6, pp.1044-1054 (1984)
- [2] 王向莉, 宮崎正弘: 構文パターンを用いた文解析, 言語処理学会第11回年次発表論文集 (2005)
- [3] 川辺諭, 宮崎正弘: 構造を含む生成規則を扱える拡張型チャートパーザ, 言語処理学会第11回年次発表論文集 (2005)