

ボトムアップ構文解析における左外置の処理

今野 聡、 田中 穂積 (東京工業大学 工学部)

1. まえがき

Prologによる自然言語処理の研究として、筆者らはETLとICOTで開発されたボトムアップ構文解析システムBUP[1]を用いて比較的大きな英語の文法を作成した[2]。これまで行なってきたDCG[3]による文法の記述では、Wh疑問文や関係節に関する規則を記述するのに、名詞句の欠けた動詞句や疑問文といったカテゴリをもうけた為、文法規則数が大幅に増加し、文法を作成する上での見通しが悪くなる問題点があった。ところで、英語などの言語では、関係節などにおける語の移動を、左外置(Left extraposition)をもちいて説明できる。この左外置は、あるカテゴリの構成要素の一つがそのカテゴリの左方に移動する際、移動前の場所に痕跡(trace)を残して移動すると考えるもので、これにより先程の問題点を解決できる。

本報告では、この左外置をBUP上で取り扱うための基本的なアルゴリズムとBUPに施した拡張について述べる。以後、このBUPの拡張版をBUP-XG[6]と呼ぶ。

2. 左外置と文法記述

従来行なってきたDCGによる文法記述例として、関係節に関する文法規則の一部を図2-1に示す。この例で分かるように、DCGで記述する場合には、主語の欠けた規則(1)や、目的語の欠けた規則(2)を一つ一つ書かななくてはならない。特に、目的語が欠けている場合には、vp2といったカテゴリを新たに設け、本来の他動詞句の規則(3)から目的語が欠けている規則(4)を記述する必要がある。

```
srel --> relpro, vp.           (1)
srel --> relpro, np, vp2.      (2)
vp --> v, {vt_obj}, np.       (3)
vp2 --> v, {vt_obj}.          (4)
```

図2-1 DCGによる文法記述例

それに対し、左外置を考慮に入れた場合について(5)の名詞句を例に考えてみる。この左外置によれば、動詞lovesの後ろには痕跡を持った名詞

句が存在し、その名詞句が動詞の目的語になっていると考えられる。したがって、she以降の埋め込み文は、文の構造から見て、完全な文であると言えるため、関係節の規則は(6)となる。これは、主語となる名詞句が欠ける場合も同じである。

```
a man that [S she loves t ]      (5)
t : trace
srel --> relpro, s.              (6)
```

3. 左外置処理の基本的アルゴリズム

BUP-XGでは一本のスタックを用意し、このスタックに、左外置した語句を支配していたカテゴリ名をプッシュ、ポップすることで、左外置の処理を行なっている。本章では、まずトップダウン法による場合について説明し、その後、BUP-XGでの場合について説明する。尚、これまで左外置を扱える文法として開発されたものに、PereiraによるXG(Extraposition Grammar)がある[4]。このXGは、DCGを拡張したもので、トップダウンパーザであるが、以下のトップダウン法の説明は、BUP-XGを説明する為のものであり、syntax sugarやスタックの構造及び使用法の点で、PereiraのXGとは異なっている。

3.1 トップダウン法による左外置の処理

まず、上の(5)の名詞句を左外置を考慮して解析する方法について説明する。左外置を考慮した場合、関係節の埋め込み文は、完全な文の構造を持っているので、関係節の為の規則は(6)のように記述できることは、2.で述べたが、構造が完全な文の形であることを示すには、動詞lovesの直後でnpの解析を成功させる必要がある。そのために、次のようなスタックを用いた処理を行なう。

①関係節の解析を始める前に、スタックにカテゴリnpをプッシュする。

②関係代名詞以下の埋め込み文を解析中にnpの解析が失敗した場合、スタックからnpをポップしてnpの解析を成功させる。このスタックのポップによる解析の成功を、カテゴリnpの下に痕跡を見つけたことと解釈する。

ここで、(5)の名詞句を解析するための文法規則を(7) - (11)に示す。

s --> np, vp. (7)

vp --> v, np. (8)

np --> pron. (9)

np --> a, noun, srel../np. (10)

np --> virtual(np). (11)

npの規則(10)で". . /"はこの度加えたsyntax sugarで、srel../npは、スタックにnpをプッシュしてからsrelの解析を行なうことを示している。又、(11)中のvirtualは、引数として与えられたカテゴリがスタックトップにあれば、スタックからそのカテゴリをポップして成功する述語で、上記②の処理を行なう。

3.2 BUP-XGでの左外置の処理

純粋なボトムアップ法での左外置の処理に、トップダウン法と同様にスタックを使用する場合、次のような問題がある。

まず第一に、スタックにカテゴリをプッシュするタイミングが不明である。トップダウン法の場合には、解析木の中に痕跡があると思われるカテゴリ(srel等)の解析を始める時点で、その痕跡を持つカテゴリ(np等)をプッシュするが、ボトムアップ法の場合、そのタイミングが分からない。

第二にカテゴリをポップするタイミングが不明である。トップダウン法の場合には、スタックのカテゴリをポップできるのは、そのカテゴリの解析を行っている時だけである。それに対し、ボトムアップ法の場合には、ポップするタイミングが不明なため、スタックにカテゴリが存在すれば、入力文の全ての単語の間(及び文の前後)で、スタックをポップする可能性があり、その為、左外置を考えない場合に比べ、計算量が大幅に増加してしまう。

BUP-XGにおいては、ボトムアップ法における上記の問題点を、BUPの基本的な動作メカニズムを利用することで解決している。通常、BUPシステムでは、図3-1に示すように、ユーザがDCGで記述した規則(12)はトランスレータ[5]によってBUP節(13)に変換され、そのBUP節がボトムアップ解析プログラムとして動作するようになっている。実行に際しては、始めに入力文の最初の単語の辞書引きを行ない、その単語の品詞を

キーとしてBUP節を選択し解析を始める。すなわち、先頭の単語の品詞がbであるならば、(13)のようにbがheadの述語名となっているBUP節が選択され、述語linkでカテゴリaから現在のゴールGへの到達可能性を調べた後、続いてc、dがサブゴールとして予測され解析が進む。このように、BUPにおいては、始めに辞書引きによって見つかった単語の品詞からBUP節を選択する部分がボトムアップの動作、又、サブゴールを予測して解析を進める部分がトップダウンの動作となっている。尚、(13)の述語goalはBUPシステムに組み込まれているものを用いる。

a(ARG) --> b(ARG1), c(ARG2), d(ARG3). (12)

↓ 変換

b(G, (ARG1), I) --> {link(a, G), (13)

goal(c, (ARG2)),

goal(d, (ARG3)),

a(G, (ARG), I).

図3-1 DCGからBUP節への変換例

さて、スタックへのカテゴリのプッシュであるが、これにはBUPのトップダウンの動作を利用する。例えば、規則(12)のdをsrel../npとすると、カテゴリcの解析が終了(goal(c, (ARG2))が成功)し、srelの解析を始める時点でnpをスタックにプッシュする。詳しくは、4.2のトランスレータのところで述べる。

BUP-XGでは、スタックのポップが行なわれるタイミングに二つの場合がある。第一は、スタックトップにあるカテゴリがゴールとして呼ばれた時点で、この場合には、スタックをポップして単に成功させればよく、トップダウンの場合と基本的に同じである。第二は、スタックトップにあるカテゴリから、現在のゴールへの到達可能性がある場合である。通常のBUPでは、goal節で、単語の辞書引きを行ない、その単語の品詞から、現在のゴールへの到達可能性があるか述語linkで調べ、可能性がある場合にのみ、その品詞をキーとしてBUP節を選択し解析に移る。スタックのポップに関しても同様に、ゴールへの到達可能性がある場合にのみ、そのカテゴリをポップし、そのカテゴリをキーとして、解析を進める。これにより、純粋なボトムアップ法の場合に見られるような無駄なスタックのポップを避けることができる。

4. インプリメンテーション

4.1 Xlist

スタックは、Xlistと呼ばれる以下に示すような3つの要素を持つ関数xの構造体で実現されている。

$x(\text{category}, \text{argument}, \text{xlist})$

ここで、categoryはスタックトップのカテゴリ、argumentは、DCGで記述された文法規則でcategoryに付加されていた引数のリストである。又、xlistは、そのカテゴリをプッシュする前のスタックを表わすXlistである。空スタックは、[]で表わされる。例として、空スタックにnp(NP_A, NP_S, NP_T)をプッシュしたスタックを表わすXlistは、次のようになる。

$x(\text{np}, [\text{NP_A}, \text{NP_S}, \text{NP_T}], [])$

4.2 BUP-XGトランスレータ

1) Xlist用変数の付加

図4-1はBUP-XGトランスレータによる変換例である。本トランスレータは、変換後の全ての述語({)で囲まれたprologプログラムを除く)にXlist用の変数を付け加える。(15)においてXで始まる変数がXlistである。述語goal中の2つのXlistのうち、最初(1番目)のものは各述語が呼ばれる時点でのスタック、2番目はゴールとしてたてたカテゴリの解析が終了した時点でのスタックを表わしており、2番目のXlistは、後続する述語の1番目のXlistとなる。

$a(\text{ARG}) \rightarrow b(\text{ARG1}), c(\text{ARG2}), d(\text{ARG3}).$ (14)

↓ 変換

$b(\text{G}, (\text{ARG2}), \text{I}, \text{X0}, \text{X1}, \text{XR}) \rightarrow$ (15)

$\text{goal}(c, (\text{ARG1}), \text{X1}, \text{X2}),$

$\text{goal}(d, (\text{ARG3}), \text{X2}, \text{X3}),$

$a(\text{G}, (\text{ARG}), \text{X0}, \text{X3}, \text{XR}).$

図4-1 BUP-XGトランスレータの変換例(1)

2) スタックへのカテゴリのプッシュ

図4-2は、". . . /"を含む規則の変換例である。(17)において下線で示した部分が、スタックにカテゴリnpをプッシュしているところである。4.1で述べたことから分かるように、下線部のXlistは、nounの解析が終了した時点でのスタック(X2)に、カテゴリnpをプッシュした状態を表わしている。したがって、srelの解析を

始めるgoalには、npがプッシュされたスタックが渡されることになる。

$\text{np} \rightarrow a, \text{noun}, \text{srel}.. / \text{np}.$ (16)

↓ 変換

$a(\text{G}, \text{O}, \text{I}, \text{X0}, \text{X1}, \text{XR}) \rightarrow \{\text{link}(\text{np}, \text{G}),$ (17)

$\text{goal}(\text{noun}, \text{O}, \text{X1}, \text{X2}),$

$\text{goal}(\text{srel}, \text{O}, \underline{x(\text{np}, \text{O}, \text{X2})}, \text{X3}),$

$\text{np}(\text{G}, \text{O}, \text{I}, \text{X0}, \text{X3}, \text{XR}).$

図4-2 BUP-XGトランスレータの変換例(2)

尚、スタックにカテゴリをプッシュしてから解析を行なえるのは、変換後のBUP節において述語ゴールによって解析が始まるカテゴリに限られる。したがって、BUP-XGでは、DCGで記述した規則の右辺の第一カテゴリには、". . . /"のついたカテゴリは記述できない。

3) ユーザによるXlistの記述モード

このトランスレータにより、ユーザは、スタック(Xlist)を意識せずに文法を記述できるが、下のように、Xlistを直接記述したい場合がある。(18)は、関係節の中の埋め込み文が接続詞でつながれている場合であるが、スタックにプッシュされたカテゴリnpが"love"の目的語として一度ポップされるとスタックが空となるため、"dislike"の目的語としての痕跡が見つからず解析は失敗する。

She is a girl that I love but you dislike.

(18)

$\text{srel} \rightarrow \text{relpro}, \text{s}, \text{conj}, \text{s}.$

(19)

そこで、2つめのsの解析を始める時点でのスタックが、1つめのsの場合と同じであることを文法規則を記述する際に指定したい。そこで、ユーザが指定したXlistをそのまま使用してBUP節に変換するモードを用意した。下の(20)は、例文(18)が解析できるように、(19)をこのモードで書き直した規則である。

$\text{srel}(\text{ARG})(\text{X0}, \text{X}) \Rightarrow$ (20)

$\text{relpro}(\text{ARG1})(\text{X0}, \text{X1}),$

$\text{s}(\text{ARG2})(\text{X1}, \text{X}),$

$\text{conj}(\text{ARG3})(\text{O}, \text{O}),$

$\text{s}(\text{ARG4})(\text{X1}, \text{X}).$

5. 実験結果

5.1 文法の記述

左外置を扱えることで、関係節に関する規則を簡単に記述できることを示したが、これにより、文法規則数が減ると同時に、文法の見通しが良くなる。whatで始まるWh疑問文の規則は、(21)のように簡単に記述できる。ここで、whnpは、疑問代名詞のカテゴリである。又、be動詞で始まるYes-no疑問文を、宣言文の中のbe動詞が文頭に移動したものと考えると、これも同様に(22)のように記述できる。

```
swhq --> whnp, sq../np.          (21)
```

```
sq --> bep, s../bep.            (22)
```

5.2 構文解析例

図5-1に、本システムによる構文解析結果の一例を示す。構文解析木の中の"t"は、痕跡を表わしている。

```
He is the man that she loves.

1124 msec.
No. 1
|-sentence
  |-sdec
    |-sbas
      |-subj
        |-np
          |-pron -- he
        -aux
          |-bep
            |-be -- is
          -pred
            |-np
              -ddet
                |-det -- the
              -nomhd
                |-n -- man
              -ncomp
                |-srel
                  -relpro -- that
                -sbas
                  -subj
                    |-np
                      |-pron -- she
                  -vp
                    -v
                      |-v -- love
                      |-suffix -- s
                    -obj
                      |-np -- t

Total Time = 1511 msec.

number of wf_goal was : 10.
number of wf_dict was : 8.
number of fail_goal was : 72.
```

図5-1 構文解析例

6. おわりに

本研究では、英語の関係節や疑問文における語の移動を説明する左外置に注目し、これをボトムアップ構文解析システムBUP上で扱えるようにBUPを拡張した。文法の記述は拡張したDCG(新機能の為に加えたsyntax sugarを含む)で行なう。又、実際に英語の文法を記述して、その有効性を確認した。今後は、本システムを改良することで右外置変形(主語の関係節が文末に移動するもの)を扱えるかどうか検討していきたい。

謝辞

本研究を行なうにあたり、計算機を使用させていただきました。淵一博ICOT研究所長に深く感謝致します。又、日頃、御討論いただいた東京工業大学田中研究室の諸氏に感謝致します。

参考文献

- [1] Y. Matsumoto and et al. "BUP: A bottom-Up Parser Embedded in Prolog", New Generation Computing, 1,2,1983.
- [2] 田中穂積、他, "ボトムアップ構文解析システムBUP上での英語文法開発とBUPの評価", Proceedings of THE LOGIC PROGRAMMING CONFERENCE '84, 12-2, 1984.
- [3] Pereira, F and Warren, D. "Definite Clause Grammar for Language Analysis - A Survey of the Formalism and a Comparison with Augmented Transition Networks", Artificial Intelligence, 13, pp231-278, May 1980.
- [4] Pereira, F. "Extrapolation Grammar" AJCL Vol. 7, No. 4, 1981.
- [5] 松本裕治、他, "BUPトランスレータ", 電総研報, Vol47, 第8号, 1983.
- [6] 田中穂積、他, "ボトムアップ構文解析システムBUPでの左外置変形の処理", 情報処理学会自然言語研究会資料 44-1, 1984.