# トップダウン法とボトムアップ法を融合した新しいパージング方式

A New Parsing Method Mixed by Top-Down and Bottom-Up Strategies

上脇　正、田中穂積（東京工業大学・工学部）

論理プログラミングの枠組みで、トップダウン法とボトムアップ法とを融合させた新しいパージング方式を考案したので、そのアルゴリズムと、文法規則をトップダウン規則とボトムアップ規則に分類する方法を述べる。

## 1．はじめに

トップダウンによる解析は、解析に必要になる予測情報ができるだけでなく、Prologの問題解決の戦略とも一致するため［Pereira 80］、Prolog上に簡単に高速なパーザを実現できる。しかし、トップダウン縦型探索による解析には、左再帰的規則があるとそこで無限ループに入るという欠点がある。一方、ボトムアップによる解析は左再帰の規則を扱えるが、予測を用いた解析が行えない。

従来の方式では、パーザが一つの書き換え規則をトップダウンの向きに解釈したり、ボトムアップの向きに解釈したりして解析を行うものであった。この方法によると、パーザが、どのタイミングでトップダウンかボトムアップの向きに規則を解釈すべきかが明確ではないため、必ずしも有効な方法ではなかった。

本論文では、トップダウン法とボトムアップ法の長所をできる限り取り入れるために、トップダウン法とボトムアップ法を融合した新しいパージングの方式を提案する。本方式では、文法規則をトップダウンに解析を行うために用いる規則とボトムアップに解析を行うために用いる規則とを明確に分類する。トップダウン規則はPereira の方法で解析を行ない、ボトムアップ規則はＢＵＰ節［松本83］を用いる。トップダウン法とボトムアップ法を融合したパージングは、トップダウン規則とボトムアップ規則が相互に呼出し通信し合いながら進行する。したがって、トップダウン規則とボトムアップ規則とが相互に呼合う機構が必要である。これについては、第２章で述べる。本論文で提案するトップダウン法とボトムアップ法の融合パージングではどのタイミングでトップダウンからボトムアップ、ボトムアップからトップダウンの構文解析に切り替えるべきかは、各文法規則が知っていることになる。このような文法規則の自動生成法の一つの方法を第３章で述べる。

４章でパージングの高速化について述べる。

図1-1がシステムの概略である。ユーザはＤＣＧを

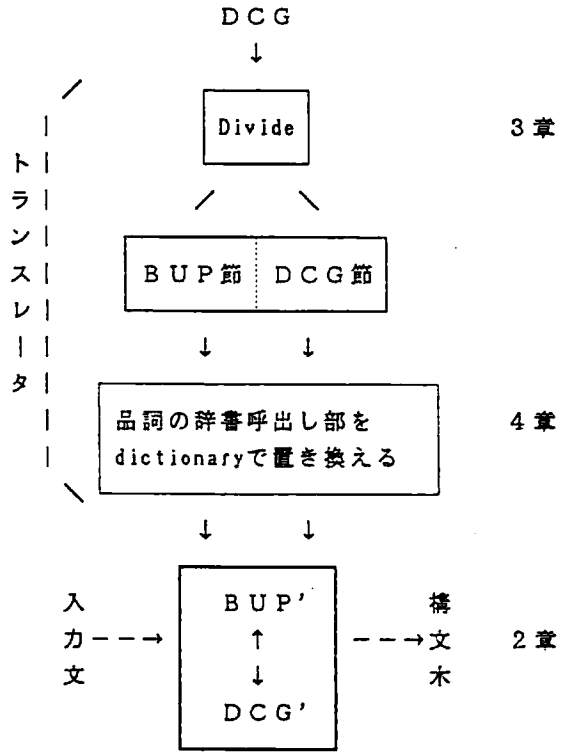


図1-1トップダウン法、ボトムアップ法融合パーザ

入力するとトランスレータによりＤＣＧはボトムアップ法で解析を行う規則とトップダウン法で解析を行う規則に分類され、それぞれＢＵＰ節とＤＣＧ節に変換される。その結果に更に高速化のための変換が行われる。解析時は両方の規則が互いに呼合いながら構文木を生成する。

## 2．トップダウン規則とボトムアップ規則の相互呼び出し

まず、'The boy walks' を例にトップダウン法とボトムアップ法による解析の順番を比べ、その後で呼出し機構の説明をする。

図2-1 がPereira の方法のトップダウンによる解析の順番である。縦型探索の典型的な順番の上から下、左から右の順番に解析が進む。

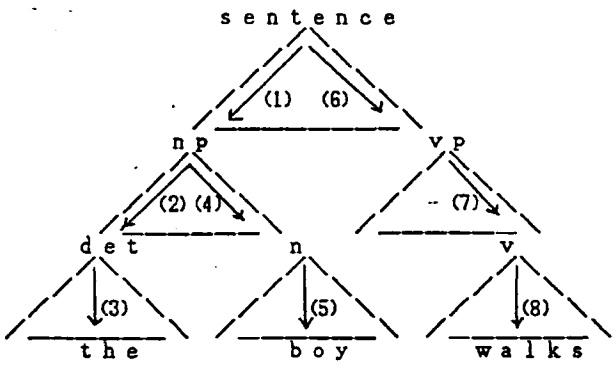


図2-1 トップダウンの解析順序

一方、図2-2 はＢＵＰのボトムアップによる解析の順番である。最初に辞書引きを行ない、それを手がかりに解析が進められる。このため左下から段々に部分木が形成され最後に全体の構文木となる。ただし、ＢＵＰでは、解析の効率を上げるためにトップダウン的予測を用いている。このため３、６では下向きに予測がなされている。つまり、６の場合、npの部分木が１－４により出来上がると、文法規則の中より右辺の先頭がnpである規則

```
sentence --> np,vp.
```

を捜し出し、その規則によりnpの右にvpが来ることを予測している。この予測の部分が６である。
そして、８により出来上がる部分木（この部分木を作るのにも６の予測が使用される。）が、６の予測と一致するとこの部分の解析が成功したことになる。

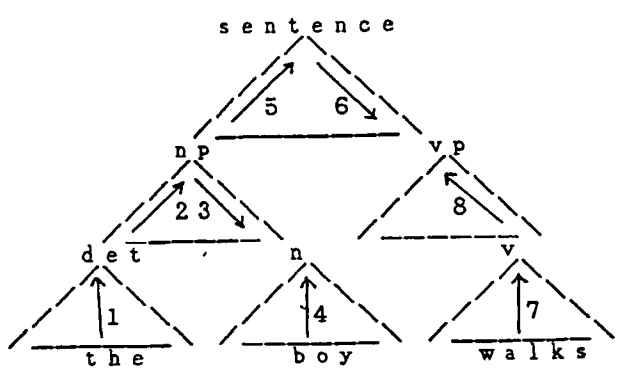


図2-2 ボトムアップの解析順序

３の部分も同様である。

このトップダウン的に予測されたカテゴリを利用するとボトムアップ規則とトップダウン規則で相互に呼び合うことが可能となる。

まず、ボトムアップ規則からのトップダウン規則の呼び出しを考えてみる。図2-2 で６の部分はトップダウン的予測を行なっている。即ち、この段階で既にその下に来るべきカテゴリ(VP)が分かっていることになる。よって、８による下からの部分木と６の予測との一致を見る代わりに、６で予測されたカテゴリを用いて、図2-1 の(7),(8) のようにトップダウンに解析を行なうことが可能である。このようしてに、ボトムアップ法からトップダウン法への切替がなされる。

実際のシステムでは、図2-1 の(6) の部分のＤＣＧの呼び出しは、

```
vp(Arg1,Arg2)
```

のようになっている。一方、ＢＵＰに於ける図2-2 の６の部分の予測付のボトムアップ規則の呼び出しは、goalという節を用いて次の様になっている。

```
goal(vp,(Arg1,Arg2))
```

第１引数が予測されるカテゴリである。これらは解析手順は異なるが、全く同じ結果を得ることが可能であり、相互に入れ替えても支障がない。ＢＵＰ節中のgoal節をＤＣＧ節の呼び出しに置き換えることにより、ボトムアップ規則からトップダウン規則の呼び出しが可能である。

```
例：np(G,0,Info) --> {link(np,G)},
                     goal(vp,0),
                     sentence(G,0,Info).
                       ↓
    np(G,0,Info) --> {link(np,G)},
                     vp,
                     sentence(G,0,Info).
```

トップダウン規則からボトムアップ規則を呼ぶ場合は、この逆を行なえばよい。図2-1 の(6) のカテゴリ(vp)が予測されたとして、図2-2 の７、８のようなボトムアップの解析を行ない、(6) と８の解析の結果を比較する。実際のシステムでは、ＤＣＧの規則の中にgoal節を入れることにより、トップダウン規則からボトムアップ規則の呼び出しが可能になる。

```
例：sentence --> np,vp.
                   ↓
    sentence --> np,goal(vp,0).
```

## 3．ボトムアップ規則と
## トップダウン規則への分類

ボトムアップ規則とトップダウン規則の相互呼び出しを行なうためには、文法規則を予めボトムアップ規則とトップダウン規則に分けておく必要がある．人間がこの分類を行うのは大変なので、ある種の基準を設けて計算機に自動的に規則の分類をやらせることにする．この章ではその分類を自動的に行うアルゴリズムについて述べる．

まず、トップダウン規則とボトムアップ規則に分けるための基準を述べる．我々は以下に示す３つの基準を考えている．

基準１：スタート・シンボルはトップダウンに呼び出す．スタート・シンボルはトップダウン規則かボトムアップ規則のどちらかに確定しておかなければ、解析を開始することが出来ない．ここではスタート・シンボルはトップダウン規則で書くと約束する．

基準２：トップダウン規則は、左再帰の規則を含んではいけない．（ここの左再帰の規則とはa --> a,b,c. のような直接的な再帰だけでなく、a --> b.　b --> c.　c --> a.のような再帰も含む．）１章で述べたように、トップダウン＆総型探索で解析を行なう場合、左再帰の規則があると無限ループに入る．

基準３：ボトムアップ規則からトップダウン規則を呼ぶ場合、規則の右辺の最初の文法カテゴリを呼び出してはならない．図2-2 を見ると２を行なうためには、その下の部分木が完成していることが必要である．なぜなら、１の部分木が完成してなく、この部分のカテゴリが分からないと２、３でどの文法規則を使用するかが決定できないからである．下の部分の完成により上の部分の解析が始まるということはトップダウンの規則が使えないことを意味する．

以上の基準により、次のような分類アルゴリズムが考えられる．

1. 文法規則の左辺のカテゴリclと右辺の最初の要素crの関係をcl->crと表わす．

例：　sentence --> sdec,conj,sdec.
ならsentence -> sdec

2. sentence -> sdec -> subj　のような関係がある時、１回以上の'->'の関係により到達可能な関係を

sentence => sdec
sentence => subj
sdec　　=> subj

のように表す．

3. カテゴリの集合C_loopを次のように定義する．

C_loop = {c ; c => c}

この集合の要素は全てトップダウン法により解析を行おうとすると無限ループに陥るカテゴリである．

4. カテゴリの集合C_bottomを次のように定義する．この集合の要素についての規則はボトムアップ規則にする．

C_bottom = { c ; c1 => c, c1∈C_loop}

5. C_bottomの要素以外のカテゴリはトップダウン規則であるとする．

ただし、スタート・シンボルは予め無限ループにならないように文法規則を作成しておく必要がある．

以上に述べた文法規則の分類アルゴリズムを用いてトランスレータは、与えられたＤＣＧの文法規則を自動的にトップダウンとボトムアップに分類し変換する．（図1-1参照）

## 4．パージングの高速化手法

### 4.1. 品詞の直接辞書引き

２章で述べたようにＢＵＰではgoalという節を解析に用いている．この節は、

```
goal(G,A,X,Z) :- dictionary(C,A1,X,Y),link(C,G),
                 P=..(C,G,A1,A,Y,Z),call(P).
```

のようなものであり、辞書引きを行い、それを使って上へ木を伸ばしている．ＢＵＰ節に変換された規則の高速化をはかるために次のような性質を満たすカテゴリ(Hinsi)を予測に持つgoal節の呼出しを直接dictionary節の呼出しで置き換えることができる．

Hinsi = {h; ∀t　h -> t, h ∈VN,
　　t ∈VT}

VN：非終端記号　　VT：終端記号

これらのカテゴリは一般に辞書項目に記載された品詞なっている．

従来のＢＵＰでは、予測したカテゴリがたとえ集合Hinsi に含まれていても、goal節を呼んでgoal節内部でdictionary節を呼んでいた．しかし、このようなgoalは、直接辞書引きを行うdictionaryの呼出しに置き換えることができる．

集合Hinsi の作成方法は、文法規則（辞書項目は除く）の左辺に一度も出て来ない文法カテゴリをHinsiの要素とすればよい．これも図1-1のトランスレータが自動的に検出し、goalの呼出しをdictionaryに置き換えることができる．

```
det(G,0,Info) --> [link(det,G)],
        goal(n,0),
        np(G,0,Info).
              ↓
det(G,0,Info) --> [link(det,G)],
        dictionary(n,0),
        np(G,0,Info).
```

図4-1 品詞の直接辞書引き

4.2トップダウンの成功部分木の登録

ＢＵＰは、解析時の再計算を避けるために一度成功や失敗したgoalは、成功ゴール、失敗ゴールとして登録している［松本83B］。同様なことがトップダウンに於いても行なえる。ＤＣＧの呼出しの部分をnp(Arg)の形からtop(np,Arg)の形にし、top節の定義を［松本83B］のgoal節の定義とほぼ相似な図4-2のようにしておけばよい。

```
top(G,A,X,Y) :- (wf_top(G,X,_,_);
                 fail_top(G,X),!,fail),!,
                wf_top(G,X,A,Y).
top(G,A,X,Z) :- P=..[G,A,X,Z],call(P),
                assertz(wf_top(G,X,A,Z)).
top(G,A,X,Z) :- (wf_top(G,X,_,_);
                 assertz(fail_top(G,X))),!,fail.
```

図4-2 top節

## 5．結論

文法規則全体を、ボトムアップに解析を行なう規則とトップダウンに解析を行なう規則に自動的に分類し、それらの相互呼び出しを可能にすることにより、ボトムアップ法とトップダウン法を融合した新しいパーザの方式を提案した。また、文法カテゴリ中でいわゆる品詞（必ず終端記号となるカテゴリ）とそれ以外のカテゴリを区別するアルゴリズムを述べ、それらについて直接辞書引き行ない高速化をはかることができることを示した。

現在、我々の研究室で使用しているシステム（文法規則数約４００）に適用したところ、トップダウン規則とボトムアップ規則がほぼ半々になり、SUN-II Work-Station上のC-Prologインタープリタによる解析速度がボトムアップ法のみによるものと比べて３割程度向上した。

左外置の処理はボトムアップでも行なえるようになっているが、トップダウンによる解析法が有効であると考えている。今後は左外置の処理への応用を考えてみたい。［Pereira81］［今野84］

参考文献


［Pereira 80］：
Pereira, F. C. N. and Warren, D. H. D.,
"Definite Clause Grammar for Language Analysis -A Survey of the Formalism and a Comparison with Augmented Transition Networks",
Artif. Intell., 13,pp 231-278,May,1980

［Pereira 81］：
Pereira.F,
"Extraposition Grammar",
Am. J. Comput. Linguist., Vol.7, No.4,1981.

［松本83A］：
松本裕治、田中穂積、平川秀樹、三吉秀夫、
安川秀樹、向井国昭、横井俊夫、
"Prologに埋め込まれたボトムアップパーサ：ＢＵＰ"、
Proc. of the Logic Programming Conference '83,
Mar.1983.

［松本83B］：
松本裕治、清野正樹、田中穂積、
"ＢＵＰの高速化"、
情報処理学会自然言語処理研究会資料39-7,1983

［今野84］：
今野聡、他、
"ボトムアップ構文解析における左外置の処理"、
日本ソフトウェア科学会第１回大会論文集、
pp223-226, 1984