

単一化文法の並列処理の一提案

2S-6

菅井 猛 溝口文雄
(東京理科大学 理工学部)

1. はじめに

GPSGのような単一化文法[1]を扱うようなパーザを作るには、大量のデータを扱わなければならない。PAF/SFG(Parallel Analysis For Situation and Feature Grammar)[2]では、構文的な処理だけでなく、状況意味論[3]に基づく、意味的なデータ構造も扱っている。この素性構造の組み合わせ的な爆発が起こりかねない。この解決策として、並列処理という点からの考察を行なう。この時、並列処理を行なうと、単一化文法のユニフィケーション[4]がどのようなものがものであるかの検討を行なう。使用言語は、単一化文法と相性が高いことと並列処理を扱えるということから、並列論理型言語GHCを用いて、その言語の枠組み内でインプリメントしてある。

2. 単一化文法による並列処理

PAF/SFGでは、構文解析のデータフローにレイヤードストリームを採用しており、素性構造はパーザの中に埋め込まれた形となっている。図1にPAF/SFGの動詞の終端記号の文法の記述例を示す。図1の規則はトランスレータによって、図2のような、二分木の生成述語を含むをDCG形式に変換される。これは、内部的には、図3のように素性構造の二分木として埋め込まれた形になっている。

```
verb
+ type:subob      + sem:love
+ subj>gr:subj    + subj>sem:Subj
+ obj>gr:obj       + obj>pcase:wo
+ obj>sem:Obj      + obj2>gr:obj2
+ obj2>sem:0

+ subj>in>sit:Sit  + subj>in>loc:Loc
+ obj>in>sit:Sit   + obj>in>loc:Loc

+ in>sit:Sit        + in>loc:Loc
+ in>pol:Pol
- subcat:[sub,obj]
* anchor(love,subj,Subj)
* anchor(love,obj,Obj)
* in(Sit,Loc,[love,subj,obj],Pol)
```

-->[愛す].

図1：終端記号の記述例

図1の+オペレータはローカル素性、*オペレータはグローバル素性を示している。ローカル素性とは、カテゴリそれぞれが固有に有する素性である。一方、グローバル素性は、すべてのカテゴリが共有する素性で

あり、どのレベルからも素性の検索が可能になっている。この時、それぞれの句構造規則に対し、素性集合の処理を行なわねばならず、大きな計算量を要する。

句構造規則内の素性構造は、その句構造の中で、変数を含まない順序付きの二分木に変換される。この二分木への変換は、素性構造を探索する際に、探索空間を削減するものである。

subcat[sub,obj]は、下位範疇化の原則を表したものである。この規則は、M→D Hにおける句構造について、「Mのsubcatの値は、Hのsubcatの値から、Dと等しいものを取り除いたものに一致する」という制約で表せる。この他に、HFC(Head Feature Convention), CAP(Control Agreement Principle), FFP(Foot feature Principle)という単一化の原則を句構造規則中に埋め込んである。

統語的あいまいさは、素性構造の(二分木)の組み合わせによって表せる。この時、部分木が、ボトムアップに作られながら、部分的な値の割り当てが成立する。

```
word(愛す,In,Out):-true!
make_obt([type:subob, sem:love,
      subj>gr:subj,      subj>sem:var(c),
      obj>gr:obj,        obj>pcase:wo,
      obj>sem:var(d),
      obj2>gr:obj2,      obj2>sem:0,
      subj>in>sit:var(e), subj>in>loc:var(f),
      obj>in>sit:var(e), obj>in>loc:var(f),
      in>sit:var(e),      in>loc:var(f),
      in>pol:var(g)],H),
verb(In,H,[sub,obj],verb(愛す),
[situation(anchor(love,subj,var(c))),
 situation(anchor(love,obj,var(d))),
 situation(anchor(var(e),var(f),[love,subj,obj],
 var(g))],Out).
```

図2：トランスレータによって変換されたもの

3. 値の割り当ての実現

GHCでは、変数の具体化の方向性が決まっており、組み込み述語をそのまま使うことはできない。また、ある文脈として、1文1文をとられた場合、文脈間にまたがる共有変数の扱わなければならない。そこで、次のような解決策をとる。

図1の大文字で始まるSubj, Obj, Loc, Polなどは変数であり、共有変数によって素性構造の整合性を保っている。図2は、句構造規則内の共有変数が、同じグラントアトムとなっている。二分木への変換では、

