

## 意味解析向きの自然言語パーサ

内田裕士, 増山顕成, 小部正人  
(富士通研究所)

### 1. はじめに

計算機がより身近なものとなってくるにつれて、自然言語処理に対する関心が高まってきている。特に自然言語の解析は機械翻訳システムやマン・マシン・インタフェースを自然言語で実現するために必須のものなので、よく研究されており、種々のパーサが開発されてきている。

本格的な機械翻訳システムや自然言語理解システムのための自然言語の解析では、文法的な解析とともに意味的な解析を本格的に行う必要があるが従来のパーサにおいては、意味解析を統一的に行えるものではなく、ad hocなやり方に頼っていたとおもわれる。

本稿では、文法解析と一体化して標準的な意味処理が行えるパーサを紹介する。このパーサはESPERと呼ばれ、機械翻訳システムATLAS/Uの言語解析部で用いられている。

### 2. ATLAS/Uの言語解析部

ATLAS/Uの言語解析部は、自然言語文を解析し、その文の意味を意味ネットワークとして表現する。言語解析部は、図1に示すようにSEGMENT, ESPER, SNGENと呼ばれる3つのモジュールから構成されている。図1には各モジュールが使用する辞書類と各モジュールの処理結果がどのような形で表現されるかも同時に示されている。

SEGMENTは、入力文を構成する単語を抽出して解析用ノードリストを作成する。

ESPERは渡された解析用ノードリストを、文法規則に従って文法解析し、同時に意味ネットワークを作成するための意味処理を行って、その結果を解析木として出力する。

SNGENでは、この解析木を元にして入力文の意味を表わす意味ネットワークを生成する。

### 3. ESPERの特徴

自然言語パーサを設計するに当たって、次のようなことに留意した。

#### 1) 文法解析と意味解析を一体化する。

例えば、「私はパンは今日は食べてない。」のように単語間の関係が意味的に明確な文においては人間は文法に頼らず意味のみで理解することができるが、「熊は彼を食った。」というような文においては、単語間に潜在する意味的な関係のみでは理解できないので文法を用いて文章の意味を理解している。このように文法と意味は不可分なので自然言語を解析する以上、文法と意味を一体化した解析を行う必要がある。

#### 2) 文法規則や意味規則の記述しやすさをねらう。

いかに文法の記述能力が優れていても実際に自然言語を解析するための文法規則がだれにでも、記述できなければ意味がない。従って実際に記述できるような文法規則を実現することを目標にした。

#### 3) Context Sensitiveなものも記述可能にする。

ESPERの特徴は以下のようなものである。

#### 1) 文法規則によってpre-definedな意味関係処理が起動される。

ESPERの文法規則には種々の型があり、型によってあらかじめどのような意味処理が起動されるかが決められている。この意味処理の結果、文法規則の適用が決定されるようになっており、文法と意味の一体化された解析が実現されている。

#### 2) 文法的な処理メカニズムと意味的な処理メカニズムを分離した。

いわゆる文法属性に意味カテゴリを入れて意味を文法に反映した形での意味解析を行うと、文法規則の増加を招き、実質上細かな意味処理は困難である。

ESPERでは文法情報と意味情報は分離した形で扱われ、その処理は別々に記述され、別のメカニズムによって処理される。従って、詳細な意味処理を試みても、意味規則が増加するのみで文法規則は増加しない。また、文法規則や意味規則は各々意味的なことや文法的なことは考えずに記述できるため、形が単純であり、記述しやすく管理は容易である。

### 3) 文法情報を属性集合で表現している。

各々の単語の文法的特徴を記述するとき、品詞のようにあるひとつの文法カテゴリで表現するのではなく、いくつかの文法属性の集合で表現するようにした。この集合を項と呼ぶ。これはCFGの非終端記号に相当する。文法属性は、単語が持っている文法的な性質のある側面を表わしている。

文法規則は、項の列で定義される。文法規則は、後で詳しく述べられているように、規則に記述されている文法属性を含む単語に対して適用可能となる。このため、ひとつの文法規則が広い範囲の言語現象に対して適用でき、文法規則数を少なくすることができる。

文法規則が適用された後、新しくできる項は、適用前の項の文法属性を明確な指定をすることなく引き継ぐことができる。これも、ひとつの文法規則が広い範囲に適用できる理由である。

### 4) 文法規則は2項間の規則にした。

2項関係は最も基本的であり、N項間の関係はすべて2項間に分解することができる。非終端記号に集合を導入していない一般のCFGにおいて、分解したときに非終端記号と文法規則が増加することについては3)により対処した。

また、解析の最終結果は意味ネットワークで表現されるので、そのための意味処理を文法規則と一体化して行うためには、2項間の規則が自然である。

自然言語、特に日本語の場合は文中の要素が省略されることが多い。N項間の規則では、省略される項を気にしながら文法規則を記述する必要があるので、文

法規則が複雑になり、記述性が損なわれる可能性が高い。

### 5) 非決定的な処理にした。

自然言語処理では、すべての場合を尽くし、そのそれぞれの場合に対して適切な処理を決定的に記述することは神でもなければ到底不可能である。

## 4. ESPERの入力と構造

ESPERは1次元的な単語配列(リスト構造)から解析木(木構造)を作り出す。

ESPERの入力は解析用ノードリストである。これは終端ノードの全順序集合であり、順序に従ったリスト構造を成している。この順序は、終端ノードに対応する単語の入力文における出現の順序である。ノードは文法情報と意味情報を持っている。文法情報は文法属性の集合として表わされ、意味情報は意味記号の列として表わされる。終端ノードは、このほかに、ノードに対応する単語の候補リストを持っており、このうちの1つの単語が最尤候補として選ばれている。初期状態では単語の候補リストの先頭の単語が最尤候補として選ばれている。終端ノードの文法情報と意味情報は、最尤候補の単語のものが採用される。

ESPERは状態スタック、解析窓及び文法規則並びに制御部よりなる。制御部は意味規則を用い、解析窓及び状態スタックで示された情報から解析木を構成する動作を行う(5. 参照)。

### 4. 1. 文法情報

文法情報は文法属性の集合として表わされる。文法情報の一般形は次の通りである。

( <文法属性> [ , <文法属性> ] . . . )

文法属性の一般形は次のとおりである。

[ <属性グループ名> . ] . . . <属性名>

属性グループ名、属性名はともに先頭が数字でない英数字列で表わされる。

### 4. 2. 意味情報

意味情報は意味記号の列として表わされる。意味情報

の一般形は次の通りである。

<意味記号> [ <意味記号> ] . . .

意味記号は大きく分けて2種類ある。

1) 概念を表わすもの

2) 概念間の関係を表わすもの

#### 4. 3. 状態スタック

ESPERは、解析中の種々の状態を記憶するために、状態スタックを持っている。

状態スタックは、状態名とカウンタの対の集合である。したがって、解析中の状態は状態名とカウンタによって表わされる。この状態は、文法規則の適用によって発生したり、消滅したりする。

#### 4. 4. 解析窓

解析窓は、解析用ノードリストの隣接した2つのノードを見るためのものである。文の解析中、ESPERは解析窓にある2つのノードのみを見ることができる。

#### 4. 5. 文法規則

文法規則は項の2項の列で定義するもので、ESPERの次の動作に対応する。即ち、1つあるいは2つのノードから新たなノードを作成し、それをルートノードとした解析木の部分木構造を作る。生成されたノードは解析用ノードリスト上のものとのノードと置き変わる。

文法規則の一般形は次の通りである。

```
<condition>
<gram1>+<gram2>
=<gram3><type>
<relation><action>
<priority>
```

### 5. ESPERの動作

ESPERは入力された解析用ノードリスト上をスキャンし、文法規則を適用し、1つの解析木を作っていく。解析木は複数ある可能性の中から最も適切であろうと判定したものを出力する。利用者が出力された解析木を不適当と判断した場合には再度ESPERを呼び出すこと

により別の解析木を出力することができる。

ESPERの初期状態においては、解析窓は文頭記号を表わすノードと、その右のノードを表わしており、状態スタックは空である。

ESPERは解析のために、状態スタックと解析窓を持っており、状態スタックに示された状態と解析窓を通して見える解析用ノードリスト上の2つのノードを参照し、適用可能な文法規則を見付け出し、それを適用して行く。

適用できる文法規則がないときは、解析窓は右へ1ノード分ずれる。文の解析は、解析用ノードリストが1つのノードに帰した時点（1つの解析木になった時点）で終了する（文が受理される）。解析窓が解析用ノードリストの右端に達した時点で、部分木を作成するような文法規則が適用できないときは、バックトラックが生ずる。

（バックトラックはこのほか、文法規則で指定されたときにも起こる）。

バックトラックは、最近に文法規則の適用が行われた時点に戻り、次の可能性を選ぶ。これは次のような手順で行われる。

- 1) 次の関係（文法規則の<relation>で指定されている）を選択する。関係の候補があればそれを採用し、解析を続ける。
- 2) 関係の候補がないときは、次に優先されるべき文法規則を適用し、解析を続ける。
- 3) 文法規則の候補がないときは、解析窓を右にずらし、解析を続ける。

文が受理されないときは、終端ノードの候補単語を変え、次の可能性をためす。候補単語は解析が及んだ最右端の終端ノードの候補から変えられる。これはすべての候補単語の組み合わせが試されるまで続けられる。候補単語の組み合わせのすべてを尽くしても受理されないときには、ESPERは解析を放棄する（入力文が不適格であったと認定する）。

#### 5. 1. 文法規則の適用

解析に際して、どの文法規則を適用するかは、状態スタックに示される状態と解析窓を通して見える解析用ノードリスト上の2つのノードの文法情報を参照して決定される。すなわち、状態スタックが示す状態と文法規則の<condition>に記述された適用条件とのマッチング及び、解析窓から見える2つのノードの文法情報と文法規則の項<gram1>及び<gram2>に書かれた左辺の2つの項の文法情報とのマッチングを行い、適用し得る文法規則を選び出す。

状態スタックに示される状態は、適用可能な文法規則を制限する。すなわち、文法規則の適用条件に状態スタックに示された状態を含む(適用条件中の状態名に“~”が付いているときは、その状態を含まない)文法規則のみが選択の対象となる。この状態は文法規則の<action>で作り出される。

文法規則の右辺に記述された項の文法属性は制限的に働く。すなわち、指定された文法属性を含む(文法属性の前に“~”がついているときは、その文法属性を含まない)ノードを選択する。従って、解析窓の2つのノードの文法属性以外の文法属性が指定されていない文法規則が選ばれる。

適用可能な文法規則が複数個ある場合は、文法規則の<priority>で示される優先度によって適用順序が決定される。文法規則の優先度が同じ場合は、より制限的な規則(指定された文法属性の数が多いもの)が優先される。

文法規則適用後に新しく生成されるノードの文法情報は文法規則の項<gram3>で示される。ここでは新たな文法属性を定義できるほか、子ノードの文法属性を全部あるいは一部引き継ぐことができる。

文法規則には図2に示されるような12個の型がある。文法規則の型によってESPERがどのような動作をするかが決まっている。それを図2に示す。

## 5. 2. 意味処理

ESPERでは、文の統語的な解析と一体化して、文

の意味を意味ネットワークとして表現するための意味処理を行うことができる。文法規則が適用されたときのような意味処理がなされるかは図2に示すように、文法規則の型によって決まっている。

単語の意味は、意味記号列として単語の辞書エントリの中に定義されている。この意味記号列は、その単語が最尤候補として選ばれたときに終端ノードの意味情報として反映される。文法規則には、意味処理のための記述が行えるようになっている。これは

<relation>で指定され、ノードの意味記号列に対する操作や、文法規則の適用によって新しく付加される意味記号列を記述することができる。意味処理はこの意味記号列を操作することによって行われる。ここに書かれた意味情報は、文法規則の型が+合成・-合成の場合は、文要素に付加される関係(修飾が起こったときに働く)を指示する。右修飾・左修飾の場合は、その修飾関係(意味)の候補を指示する。文法規則の型が先読み及び後読みの場合は、合成の場合と同じである。

解析中、ESPERは意味記号列の操作や付加を行うが、その際に意味関係規則を参照し、一般的な言語の知識や常識と矛盾しないかどうかを検証する。

## 6. おわりに

現在までにESPERを用いて解析を試みた自然言語文は、日本語の計算機マニュアル、商業通信文、及び英語の中学校教科書などである。

これらの文を解析するための文法規則の数は日本語では約150、英語では約100である。日本語の文法規則数は今後それほど増加すると思えず、英語の方も日本語に比べて5割増し程度で済むものとおもわれる。最終的な文法規則数は日本語では約200、英語では約300程度が予想される。従って文法規則を少なく抑えるという目的は達成されたものと考えられる。

文法規則の記述しやすさに関しても、全くの素人が日本語の文法規則を1日で記述できるようになったことから習熟性及び記述性は十分とおもわれる。

意味関係規則は更に簡単で誰でも書けている。このことから解析のための規則を文法規則と意味関係規則を分けたことは、記述性という面から見て非常に効果があったとおもわれる。

解析速度は約20 msec/語であり、意味解析を十分に行っているわりには早いと思われるが、今後改良したい点である。

謝辞

本研究の論文を作成するに当たって、いろいろな助言をいただいた東京理科大学井上謙蔵教授及び本研究の機会を与えていただき、研究の進め方についていろいろ助言していただいた林達也ソフトウェア研究部長に感謝いたします。

参考文献

- 1) 内田他：日英機械翻訳システムATLAS/U, 自然言語処理29-3, 1982

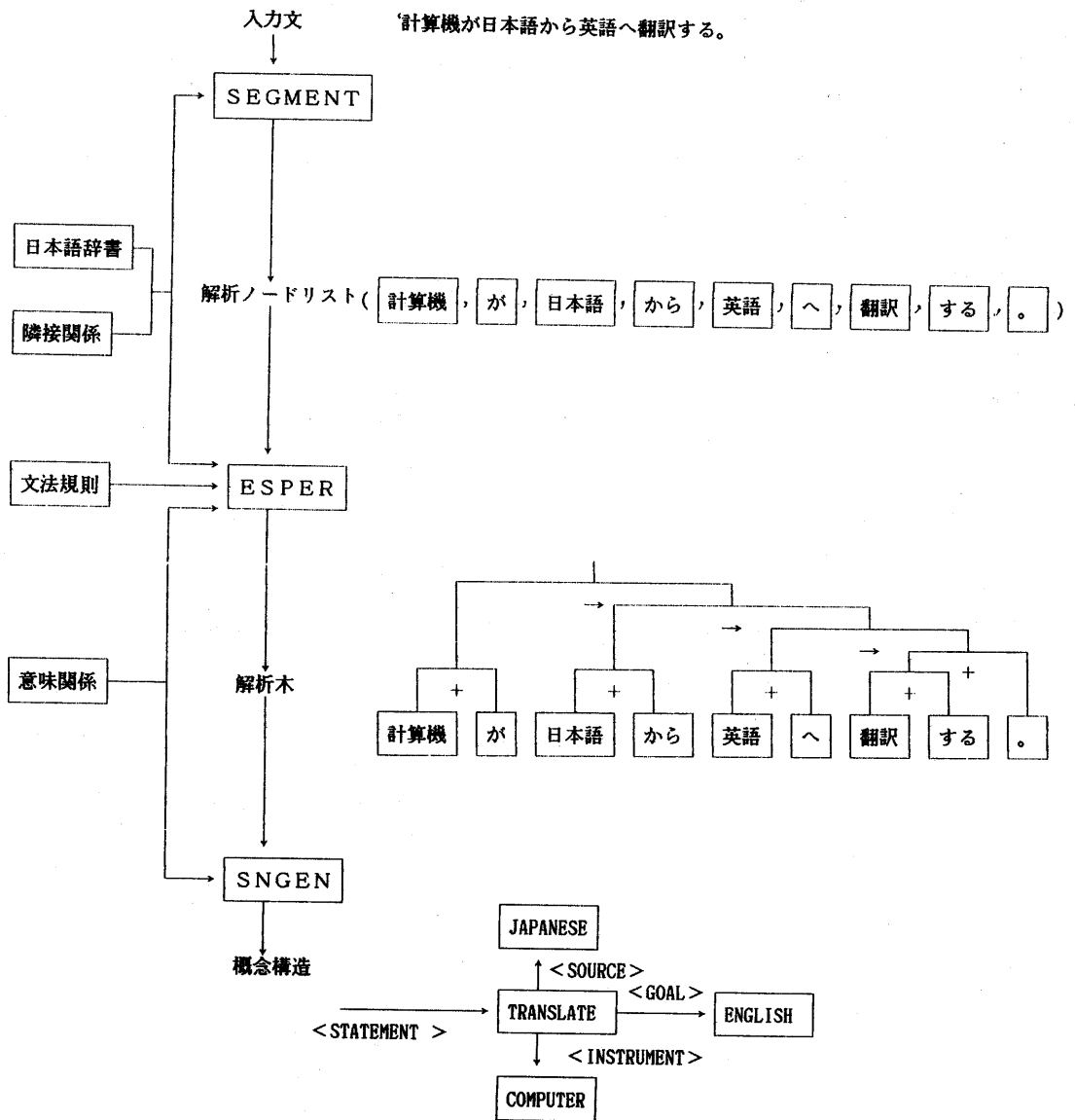


図 1 ATLAS/Uの言語解析部

処理前: ( A , B , C , D )



文法規則のタイプ	生成される木	参照される意味関係	生成されたノードの意味記号	処理後の窓の位置
+合成 (+)		(b, *, <c>)	b, <c>	( <span style="border: 1px solid black; padding: 2px;">A</span> , <span style="border: 1px solid black; padding: 2px;">E</span> , D ) 
-合成 (-)		(c, *, <b>)	c, <b>	( <span style="border: 1px solid black; padding: 2px;">A</span> , <span style="border: 1px solid black; padding: 2px;">E</span> , D ) 
右修飾 (->)		(b, c, <r>)	c	( <span style="border: 1px solid black; padding: 2px;">A</span> , <span style="border: 1px solid black; padding: 2px;">E</span> , D ) 
左修飾 (<-)		(c, b, <r>)	b	( <span style="border: 1px solid black; padding: 2px;">A</span> , <span style="border: 1px solid black; padding: 2px;">E</span> , D ) 
先読み (F)		-----	b	( <span style="border: 1px solid black; padding: 2px;">A</span> , <span style="border: 1px solid black; padding: 2px;">E</span> , C , D ) 
後読み (B)		-----	c	( A , <span style="border: 1px solid black; padding: 2px;">B</span> , <span style="border: 1px solid black; padding: 2px;">E</span> , D ) 
右ずらし (R)	-----	-----	-----	( A , B , <span style="border: 1px solid black; padding: 2px;">C</span> , <span style="border: 1px solid black; padding: 2px;">D</span> )
左ずらし (L)	-----	-----	-----	( <span style="border: 1px solid black; padding: 2px;">A</span> , <span style="border: 1px solid black; padding: 2px;">B</span> , C , D )
交換 (X)	C B	-----	-----	( <span style="border: 1px solid black; padding: 2px;">A</span> , <span style="border: 1px solid black; padding: 2px;">C</span> , B , D )
複写 (C)	B B C	-----	-----	( <span style="border: 1px solid black; padding: 2px;">A</span> , <span style="border: 1px solid black; padding: 2px;">B</span> , B , C )
バックトラック (?)	-----	-----	-----	-----
単語変更 (!)	B C'	-----	-----	( A , <span style="border: 1px solid black; padding: 2px;">B</span> , <span style="border: 1px solid black; padding: 2px;">C'</span> , D )

図 2 文法規則の型