自然言語処理 81－7
（1991. 1. 18）

中間言語を得るための自然言語解析システム

内田 裕士　朱 美英

㈱富士通研究所

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はなく，ad hocなやり方に頼っていたと思われる．本稿では，日本電子化辞書研究所で開発中の電子化辞書の枠組みに基づいた辞書を用い，国際情報化協力センター(CICC)で進められている近隣諸国間機械翻訳プロジェクトで開発された中間言語（意味ネットワーク表現）を得るためのパーサを提案する．このパーサは，文法的な解析と一体化して本格的な意味処理を行うことができる．

A natural language parsing system to get interlingua from sentences

Hiroshi UCHIDA　Meiying ZHU

FUJITSU LABORATORIES LTD.
1015, KAMIKODANAKA NAKAHARA-KU, KAWASAKI 211, JAPAN

Semantic analysis is required for a natural language analysis system for a high-quality machine translation system and natural language understanding system basides grammatical analysis. However, in conventional parser semantic analysis could not be done systematically, it rather seemed to depend on the ad hoc way. In this paper, using a dictionary based on the framework of the electronic dictionary developed by EDR we propose parser to get the interlingua (semantic network representation) which was developed by the machine translation project among neibouring countries promoted by CICC.

This parser can do full-scale semantic processing unified with the grammatical analysis.

1.はじめに

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はなく，ad hoc なやり方に頼っていたと思われる．

本稿では，文法的な解析と一体化して本格的な意味処理が行えるパーサを提案する．このパーサは機械翻訳システムATLAS/U [1]の言語解析部で用いられたパーサESPER [2]を改良したものである．改良の主な外部的要求としては，日本の標準的辞書として開発が進められている日本電子化辞書研究所(EDR) の電子化辞書の枠組み[3][4]に基づいた辞書を用いるということ，および国際情報化協力センター(CICC)で進められている近隣諸国間機械翻訳プロジェクトで開発された中間言語[5][6] (意味ネットワーク表現) をパーサの出力として得るということである．

2.設計思想

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)形態素解析，構文解析，意味解析を一体化する．

文解析の最初のフェーズである形態素解析でさえ，構文的な情報や意味的な情報を利用しなければ解析の曖昧性を解消することはできない．また，構文解析における曖昧性も意味的な情報を用いなければ，その曖昧性を解消していくことはできない．このためには形態素，構文，意味解析を一体化して行い，そのどの解析においても，各種の情報を利用可能にする必要がある．

(2)言語依存性をなくする．

高度に言語依存になり易い形態素解析に到るまで言語依存性をなくし，すべて規則として記述可能にする．

(3)文法規則の記述のし易さを狙う．

いかに文法の記述能力が優れていても，実際に自然言語文を解析するための文法規則が誰にでも記述できなければ，意味がない．従って，実際に記述可能な文法規則を実現することを目標にする．

(4)文脈依存な言語現象にも対応できるようにする．

(5)国際情報化協力センター(CICC)で進められている近隣諸国間機械翻訳プロジェクトで開発された中間言語をパーサの出力として得ることを目標とする．

(6)日本電子化辞書研究所(EDR) の単語辞書および概念辞書を利用できるようにする．

このような設計思想に基づいて作成されたESPER Ⅱの特徴は以下のようなものである．

(1)構文操作と意味操作の対応した文法規則

ESPER Ⅱの文法規則には種々の型があり，型によって予めどうのような意味操作がなされるかが決まっている．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，文法処理と意味処理が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 Ⅱ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，記述し易く管理は用意である．

(3)文法属性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に，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できるようにした．この属性集合はCFG の非終端記号に相当する．文法規則はこの属性集合に対して定義される，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になる．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可能になり，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，新しくできる属性集合は，適用前の属性集合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．

3.ESPER Ⅱの機能

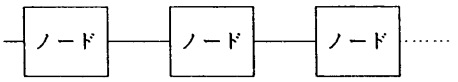
ESPER Ⅱは文字列から意味ネットワークで表現される中間言語を直接作りだす．

ESPER Ⅱの入力は，文頭記号，入力文字列，文末記号からなる解析用ノードリストである．ESPER Ⅱは解析用ノードリストに対して文法規則を適用し，解析木および中間言語を作りあげて行く．

解析用ノードリストのノードには2 つのタイプがある．文字列と非終端記号に相当する属性集合である．ESPER Ⅱは文字列タイプのノードに当たると形態素解析を行い，

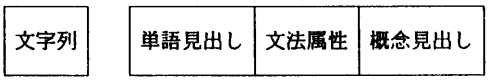
その一部あるいは全部を属性集合に変える．

解析ノードリスト



ノード

文字列型　　辞書項目型



部分木型

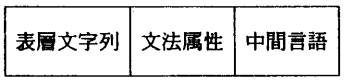


図1 解析ノードリストの構造

ESPER Ⅱの出力は解析木および中間言語である．中間言語は国際情報化協力センター(CICC)で進められている近隣諸国間機械翻訳プロジェクトで開発されたものであり，文の表す意味を，文の表す概念，その概念に対する話者の視点，概念を表現する話者の意図や概念に対する話者の判断，および文章の構造という観点から表現したものである．

中間言語は，基本的には概念間の二項関係および概念に付加された属性子の集合として表される．この中間言語の特徴としては，

(1)複合概念がひとつの概念としても，また2 つ以上の要素概念から構成された概念としても見ることのできるハイパーネットワークになっているということ，

(2)同一の実体に対してはひとつのノードが与えられ，種々のレベルから参照可能になっている

ことがあげられる．

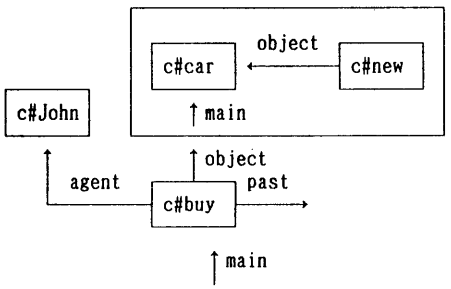


図2 中間言語のネットワーク表現

4.ESPER Ⅱの構造

ESPER Ⅱは状態スタック(status stack)，解析窓(analysis window),条件窓(conditional window)および制御部(control section) よりなり，単語辞書，連接規則 (行列),解析文法規則，概念辞書を用いて解析を行う．

状態スタックは解析中の状態をモニターするためのものである．状態スタックには：解析実行中，状態を示すための情報が蓄積される．解析窓は，文法規則適用の対象となる隣接した2 つのノードを見るためのものである．

状態スタック

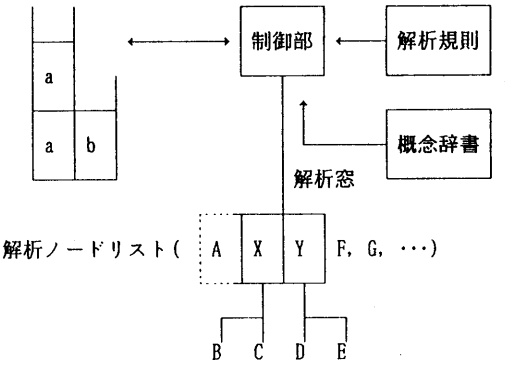


図3　ESPERⅡの構造

5.ESPER Ⅱの動作

ESPER Ⅱは解析用ノードリストをスキャンし，文法規則を適用し，1 つの解析木を作って行くと同時に，入力文の中間言語表現も作成する．解析木および中間言語は，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．再度ESPER を呼び出すと別の解析木および中間言語を出力することができる．

初期状態においては，解析窓は文頭記号を表すノードと，その右の入力文字列全体を表すノードの上に位置している．状態スタックは空である．

文法規則の適用は，状態スタックに示された状態と解析窓および条件窓を通して見える解析用ノードリストの上のノードを参照し，適用可能な文法規則を見付だし，それを適用して行く．

解析窓のノードが文字列のときは，その文字列に対して形態素解析が行われ，その後に文法規則の適用が試みられる．

適用できる文法規則がないときは，最も最近に文法規則の適用が行われる時点にバックトラックする．このとき，最も最近に適用した規則が解析規則ならば，その状態に戻り，次に適用できる規則を探すが，適用できる規

則がないなら更にバックトラックする．また，もし最も最近に適用した規則が形態素選択ならば，その状態に戻り次の形態素候補を選択する．他の形態素が存在しない場合は更にバックトラックする．

バックトラックはこの他に，文法規則で陽に指定することもできる．

文解析は，解析用ノードリストが1つにのノードに成った時点，すなわち1つの解析木が作られた時点で終了する．

6.文法規則

文法規則の一般形は次の通りである．

```
"(" <S-COND> ")"
 ["<"  {"(" <PRE> ")" } ... ">" ]
"(" <NODE1> ")"
"(" <NODE2> ")"
 ["<"  {"(" <SUF> ")" } ... ">" ]
"="
 ["  { "]  {
 ["(" <NODE3> ")" ]
<TYPE>
 [ "R("  {<ARC> } ... ")" ]
 [ "SA(" <SEM-ACTION> ")" ]
 [ "(" <S-ACTION> ")" ]
 [";" ] } ...  [" }  "]
 [ "P(" <PRIORITY> ")"]
";"
```

<S-COND>では，文法規則適用するときの状態条件を指定する．

<PRE> では，<NODE1>で指定されたノードの左に隣接するノードを指定する．<SUF> では，<NODE2> で指定されたノードの右に隣接するノードを指定する．左右の隣接ノードは複数個指定することができる．

<NODE1>,<NODE2> では，構文，意味操作の対象となる隣接した2 つのノード（解析窓のノード）を指定する．

ノードは，論理和，論理積，否定演算子を含む文法属性の集合を各々示すことによって指定される．文法属性は単に文法属性だけでなく，その文法属性をもっている単語見出しや概念見出しも指定することができる．

<NODE3> では，規則の適用による構文操作によって新しく生成されるノードの文法属性を指定する．この文法属性は，<NODE1> および，<NODE2>で示されたノードの文法属性の一部あるいは全部を継承したり，新しく文法属性を追加したりすることによって指定される．

規則の適用によって，どのような構文，意味操作を行うのかは<TYPE>で指定する．タイプの種類およびそれぞれの機能については，図4 に示す．

規則適用前　(A B C D)
　　↑
　解析窓

型	生成部分木	中心概念	部分概念構造	規則適用後の解析窓の位置
＋合成 +	E B C b <c>	b,<c>	<c> b →	(A E D) B C
－合成 -	E B C <b> c	c,<b>	<b> → c	(A E D) B C
右修飾 →	E B C b,<r> c	c	<r> b → c	(A E D) B C
左修飾 ←	E B C b,<r> c	c	<r> c ← b	(A E D) B C
先読み F	E B b	b		(A E C D) B
後読み B	E C c	c		(A B E D) C
右ずらし R				(A B C D)
左ずらし L				(A B C D)
コピー C				(A B B C D
コピー生成CG				(A B B C D
交換 X				(A C B D)
バックトラック ?				
左特殊バックトラック 1?				(A B' C D)

右特殊バックトラック 2?				(A B C' D)
左削除 D1				(A C D)
右削除 D2				(A B D)
状態変更 N				(A B C D)

図4 解析規則のタイプ

<REL> では,<NODE1>と<NODE2> で指定されたノードの中心概念間の関係の候補を指定する．候補は複数個指定することができる．規則が適用されると概念辞書が参照され，複数の関係子の候補のうち，書かれた順に意味関係チェックが行われ，意味関係チェックが成功した関係子が選択され，どの関係子が選択されたかという情報が生成ノードの文法属性として追加される．どれも選択されなかった場合は，規則の適用が失敗する．指定した関係子が選択されないことによって規則の適用が失敗したときの処理は，"{" と "}" の中に記述できる．ここに記述された規則は，前から順に適用され，どれかひとつでも成功すれば規則の適用が成功する．

<SEM-ACTION>では，生成された部分意味ネットワークに対する付加的な意味操作を指定する．ここでできる意味操作は，①部分意味ネットワークの付加，②複合概念の作成，③中心概念の指定，④概念の関係付けである．意味操作の記述において，規則適用前の左ノード(NODE1) の中心概念は*1で，規則適用前の右ノード(NODE2) の中心概念は*2で，規則適用後の中心概念（生成ノードの中心概念）は* で表される．

意味ネットワークの記述形式は以下の通りである．

```
<SEM-ACTION>  ::= "(" <C-HEAD> ")" [<SUBNET>]
<C-HEAD>      ::= <CONCEPT>
<CONCEPT>     ::= ["(" ] { <HEADCONCEPT> | "*" |
                    *1 | *2 | <SEM-ACTION>} [")" ]
<SUBNET>      ::= " {" <ARC-CONCEPT>
                    ["," <ARC-CONCEPT> ... ] " } "
<ARC-CONCEPT> ::= <ARC> |
                    {<ARC> <CONCEPT> [<SUBNET>] }
<ARC>         ::= <ARC-DIR> <ARC-NAME>
<ARC-DIR>     ::= "<" | ">"
<ARC-NAME>    ::= <RELATION>
```

複合概念の作成は，複合概念にしたい部分を "("と ")" で囲んでその範囲を示す．範囲の指定は，複合概念の中心概念を指定する方法と，複合概念にしたい部分の意味ネットワークを指定する方法がある．

規則適用後の中心概念は，規則のタイプによって決まっているが，中心概念を変更することもできる．中心概念を変更する場合は，

*= <中心概念としたい概念見出し>

と記述する．

既にできた意味ネットワークの概念と概念を関係付けることもできる．属性子を付加したいときは，どちらか一方の概念を省略する．概念間関係の指定は以下の形式に基づく．

R((<概念見出しの指定>)<ARC-DIR><ARC-NAME>
(<概念見出しの指定>))

意味操作は複数個指定することができる．意味操作の実行は，書かれた順に行われる．

<S-ACTION>では，規則適用後の状態を指定する．

<PRIORITY>では,2つ以上の規則が適用可能なときにどの規則を優先的に適用するかを指定する．

7.文法規則のタイプ

文法規則には図4 に示すように17のタイプがある．17のタイプのうち,4つは基本規則であり，残りの13つのタイプは基本規則の適用を制御するための規則である．

ESPER Ⅱの最も顕著な特徴は，全ての基本規則の構文操作が意味ネットワークに対する意味操作に対応していることである．この機能によって，構文操作の意味的な正しさが，生成された部分意味ネットワークが概念辞書に含まれるかどうかを調べることによって検証される．

図4 には全てのタイプの文法規則と，その文法規則が適用されたときに生成される部分木および2 つのノードから新しく生成される部分ネットワーク，規則適用後の解析窓の位置が示されている．

プラス合成(+) およびマイナス合成(-) は，自立語と非自立語，または非自立語と非自立語の合成に使う．

右修飾（→）では，左ノードが右ノードを修飾し，右ノードの中心概念が新しく生成されたノードの中心概念になる．左修飾（←）では逆に，右ノードが左ノードを修飾し，左ノードの中心概念が新しく生成されたノードの中心概念になる．

先読み(F) は，左ノードに対して情報を追加削除するときに使い，後読み(B) は，右ノードに対して情報を追加削除するときに使う．

右ずらし(R) は，解析窓を右にずらすときに使用し，左ずらし(L) は，解析窓を左にずらすときに使用する．

コピー(C) は，解析用のノードをコピーする．このとき，その解析用ノードが持っていたその時点の部分ネッ

トワークは共有される．コピー生成の場合は(CG)，解析用のノードとともにその時点の部分ネットワークがコピーされる．

交換(X) は，解析窓のノードの順序を入れ換える．

バックトラック(?) は，強制的にバックトラックをかける．

特殊バックトラック(1?/2?) は，指定した解析ノードを強制的に生成される前の状態に戻す．"1?"は左ノード，"2?"は右ノードの特殊バックトラックである．

単語削除(D1/D2) は，解析用ノードをノードリストから削除する．"D1"は左ノードを削除し，"D2"は右ノードを削除する．

状態変更(N) は，単に状態だけを変更する．

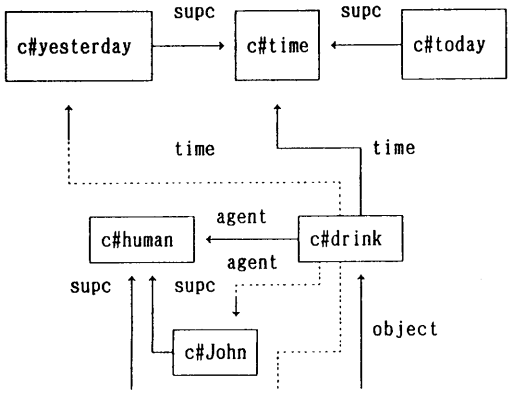
8.意味解析

意味解析は図5 に示すような概念辞書を用いて行われる．意味解析は主に単語の表す概念と，それらの概念間の関係を決定するために行われる．

ESPER Ⅱでは，文の文法的な解析と同時に，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操作を行うことができるようになっている．文法規則が適用されたときどのような意味操作がなされるかは図4 に示すように，文法規則の型によって決まっているし，それ以外の意味操作は文法規則の意味操作部に記述される．

単語の表す概念は概念見出しとして単語辞書エントリの中に定義されている．この概念見出しは，その単語が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．

解析中，ESPER Ⅱは意味操作の結果として作成される部分概念関係表現を得るが，その際に，その部分概念関係表現が概念辞書中の知識と矛盾しないかどうかが確認される．この結果により，矛盾のある概念関係表現を作成するような概念や概念間の関係の選択が棄てられることにより，正しい概念家概念間の関係が選択される．



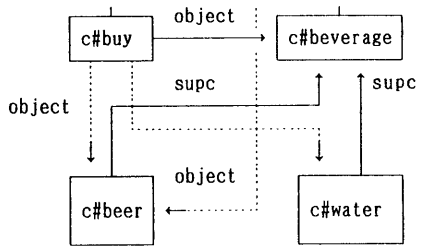


図5 概念辞書

9.並列句の解析

自然言語文の解析において，最も難しい問題の一つが並列句の解析である．ESPER Ⅱにおいては解析の結果として得られた関係子and で結ばれた概念が，それ以外の関係を他の概念ともつ場合に，and で結ばれた概念の全てがその概念と同じ関係をもつことができるかを概念辞書を使用して調べ，同じ関係をもてる部分についてのみ並列句であると認定している．

10. 省略句の処理

複文や重文において各々の文に共通する主語や目的語，同じ主語や目的語を共有する述語などは省略されることが多い．ESPER Ⅱにおいては，このように文中に省略句がある場合，省略される可能性のある主語や目的語あるいは述語等をコピー機能を用いて予めコピーしておくことによって，省略句を補うことができるようになっている．また，同一文中に全くないような概念が省略されているような場合は，解析規則の意味操作部で省略概念を補うことができるようになっている．

11. 慣用句の解析

「決して～ない」のように慣用句のなかに他の句を挿入することができるような慣用句の処理を，慣用句の検出を文字列の照合だけで行うのでは精度の点で問題がある．ESPER Ⅱにおいては，このような慣用句の処理を形態素，構文，意味処理を一体化して行うようにして精度をあげている．

EDR の単語辞書においては，このような慣用句は挿入可能な句を含めて，慣用句のとり得る構文構造が定義されている．ESPER Ⅱでは，この情報をもとにして，構文的制約を課して，慣用句に関して再帰的にESPER Ⅱを呼び出すことによって，解析を行えるようにしている．

12. 数字の処理

一般に数字は未知語として扱わなければならない．中間言語表現においては，数表現はその値で示される．こ

のためには文章中の数表現がどういう値を示しているのかを知る必要がある．パーサの言語独立性を保つためには，これを文法規則を用いて行わなければならない．このために文法規則のプラス合成およびマイナス合成規則には数字処理のために以下のような拡張機能が設けられている．

"+/-:PLUS": 左右のノードの概念見出しを加算する．
"+:TIMES": 左右のノードの概念見出しを掛算する．
"+:SUB": 左右のノードの概念見出しを引算する．
"+:DIV": 左右のノードの概念見出しを割算する．
"+:CONCAT": 左右のノードの概念見出しを合成する．

8.おわりに

日本電子化辞書研究所(EDR)の電子化辞書の枠組みに基づいた辞書を用いて，国際情報化協力センター(CICC)で進められている近隣諸国間機械翻訳プロジェクトで開発された中間言語を得るために必要な機能について概観した．

目標とした言語独立性についても本稿では触れなかったが，還移的連接関係と二方向連接文法を使用した形態素解析法により解決できたと考えている．

本パーサは解析のために必要な機能という面からのアプローチによって作成されたものであり，解析速度という面からの考慮はなされていない．今後，インプリメンテーションの改良を通じて，機能を犠牲にすることなくスピードアップを図っていく予定である．

参考文献

1)内田裕士，小部正人，西野文人，増山顕成，松井くにお：日英機械翻訳システムATLAS/U,自然言語処理研究会資料,29-3(1982.1.22)
2)内田裕士：意味解析向き自然言語パーサ，自然言語処理研究会資料,34-3(1982)
3)EDR 電子化辞書，TR-17,日本電子化辞書研究所(1990)
4)Uchida,H.:Electronic Dictionary, Proceedings of International AI Symposium 90, pp89-94(1990.11)
5)内田裕士，朱美英：多言語翻訳のための中間言語の構成法，自然言語処理研究会資料，72-9(1989.5.19)
6)Uchida,H.,Zhu,M.:Interlingua, Manuscripts of International Symposium on Multilingual Machine Translation'90(1990.11.5)