

4C-5

山崎 重一郎、 弘田 直人、 赤坂 宏二
富士通 日本情報科学研究所 ICOT

1. はじめに

ICOTでは、談話処理システム研究のベースとなるソフトウェアツール群を汎用日本語処理系(LTB)として研究開発を進めている。SAXは、LTBの構文解析ツールであり、DCGによる文法をもとに上昇型の解析を行う並列構文解析システムAXの逐次版システムである。本稿では、LTBにおける、SAXと形態素解析システムLAXや意味処理言語CILとの連携方法と、SAXの実行及びデバッグの環境について述べる。

2. LAXとの連携

LAXは並列形態素解析システムの逐次実行版システムである。LTBは利用者とのインターフェースとしてLTBシェルを備えており、LTBの各ツールはLTBシェルのパイプを通じてデータを交換できる。LAXによる形態素解析の結果は、LTBシェルのパイプを通じてSAXに送られる。このLAXからの入力、形態素解析の結果得られた統語範疇の並びであり、SAXによる構文解析はこれらの統語範疇の並びより上の構造の解析を行う。LAXは、形態素解析に曖昧性がある場合、その全ての解を束状に縮退した統語範疇の並びを結果として与えるが、SAXはこの形式の入力をもとに解析を実行することができる。

LAXの解析結果の例

人間がこの地球の上で生きつづけていくためには、...

- [体言(語彙/人間)],
格関係(語彙/が),
連体(語彙/この),
...
(統叙成分(語彙/[生き, 続け, て, い, く]), ...),
[統叙成分(語彙/[生き]), ...],
統叙成分(語彙/[続け, て, い, く])...),
接続(語彙/ため),
...].

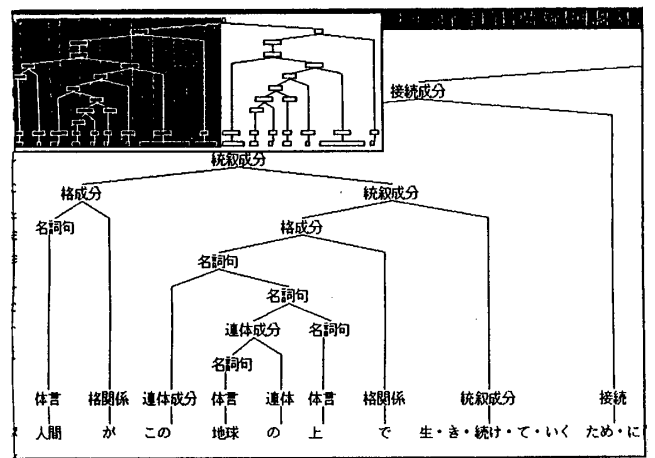
3. CILとの連携

CILは、部分項というデータ構造に対する単一化と遅延評価による制約記述によってPROLOGを拡張した論理型言語である。CILの部分項は、ラベル/データ対を

節とし横方向に順序を持たない無限木構造に対応するデータ構造であり、部分項による記述方法や部分項の単一化は自然言語の意味処理に適している。SAXは、DCGの補強項としてCILのゴール列を記述することや、DCG規則の中にCILのプログラム節を記述することを許しており、これらの補強項で意味処理を行うことによってSAXによる構文解析と同時に平行して意味処理を行うことが可能になる。また、SAXの解析アルゴリズムは決定的であるので、意味処理に非決定的な部分がある場合多重環境が必要になるが、遅延実行の利用により環境の複写を最小限にすることができる。これらのCILプログラムの評価は、SAXの解析実行部が、補強項の評価が必要となった時点で、CILインタプリタを呼び出すことによって実現している。

4. 解析実行環境

SAXの解析実行環境は、LAXからの入力をもとに全ての可能な構文木を求め、得られた木構造の数を表示し、それらの木の縮小表示をメニューとして表示する。利用者はこの縮小表示をマウスで選択することによって詳細な木構造をウィンドーに表示させることができる。また解析結果の木構造が巨大なためにウィンドーに一度に表示しきれない場合は、縮小表示によって表示範囲を指定することも可能である。また、解析の結果得られた意味構造は、部分項のプリティプリントとして表示させることができる。



The parsing system SAX in LTB

Shigeichiro YAMASAKI¹, Naoto HIROTA², Kouji AKASAKA³

(1)Fujitsu Ltd. (2)Information Science Labs. (3)ICOT

