

2F-3

LAX-インタプリタと
意味構成部のデバッグ環境の実現

米田順美 久保幸弘† 白石智子 妙泉正隆
(財)日本情報処理開発協会 †(財)新世代コンピュータ技術開発機構

1 はじめに

ICOTでは、日本語処理システムの構築に必要な解析・生成などのモジュールの統合開発環境をLTB(Language Tool Box)[5]としてまとめている。形態素意味解析システムLAX[1]はこのLTBのツールの一つで、分かち書きされていない通常の日本語の文章から、文節の並びとその意味構造を解析するプログラムの開発環境である。

従来のLAXの開発環境[2]では、形態素解析部のデバッグのみをサポートしていたが、今回意味構成部のデバッグ環境を作成したので報告する。また、従来版では形態素意味解析を行うためにはエントリの修正毎に辞書をコンパイルする必要があり、大きな辞書の開発には不向きであったが、これも改善したので報告する。

2 LAXの構成

図1にLAXのシステム構成を示す。

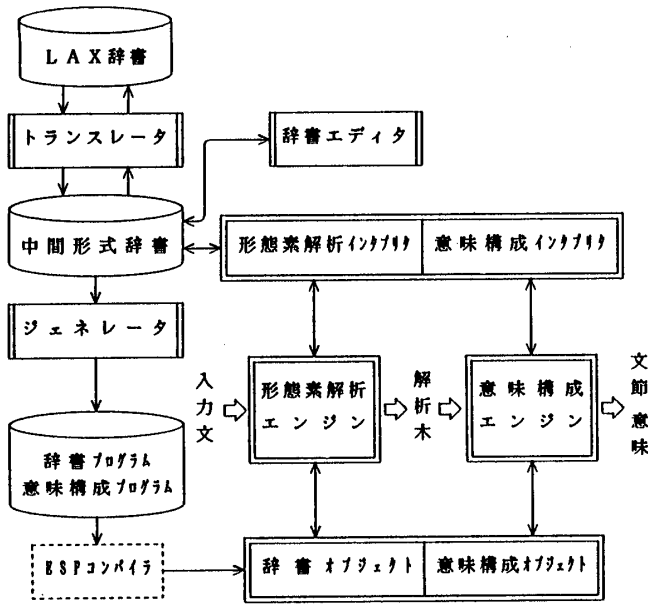


図1: LAXシステム構成

LAXでは、LAX辞書と呼ばれる辞書を作成し、トランスレータにより中間形式辞書ファイルを得る。以後LAXでの辞書の開発はこの中間形式辞書ファイルに対して行う。さらにジェネレータにより辞書プログラムと意味構成プログラムを得る。これらはESPにより記述

Implementation of the LAX-interpreter and the debugging environment for the semantic construction part.

Junmi YONEDA Yukihiro KUBO† Tomoko SHIRAISHI Masataka YOSHIZUMI JIPDEC †ICOT

されており、ESPコンパイラを用いてコンパイルすることにより辞書オブジェクトと意味構成オブジェクトを得る。従来の形態素意味解析プログラムでは、解析エンジンがこの辞書オブジェクトと意味構成オブジェクトを参照し、形態素解析と意味の構成を行う。今回開発した解析インタプリタでは、オブジェクトの代わりに中間形式辞書ファイルを参照する。LAX辞書へは逆トランスレータにより変換される。表1に従来のLAXの機能と今回実現した機能を示す。以下では、改良した点につ

表1: LAXの機能

モジュール名	形態素解析部	意味構成部
コンパイラ	○	◎
インタプリタ	◎	◎
エディタ	○	◎
デバッガ	○	◎

○: 従来の機能 ◎: 今回実現した機能

いて説明する。ただし、意味構成部のコンパイラ機能と構成の詳細については別途[3]発表するので省略する。

3 中間形式辞書を用いた形態素解析インタプリタ

中間形式辞書ファイルは図2に示すようにTRIE構造、辞書エントリファイル、左(右)方接続素性ファイル、接続素性ファイルおよび意味構成規則ファイルで構成され、それぞれ矢印が示すような参照関係にある。形

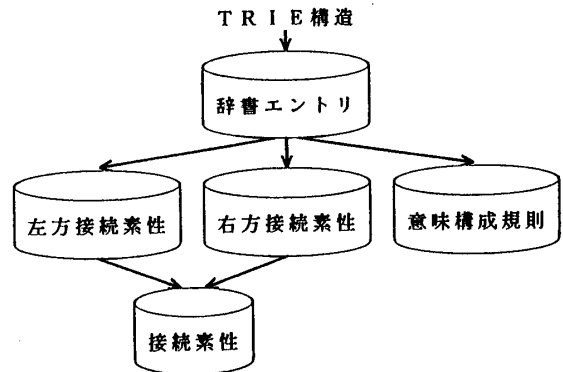


図2: 中間形式辞書ファイルの構成

態素解析インタプリタは辞書オブジェクトの代わりに中間形式辞書ファイルを用いて辞書引きを行う。辞書オブジェクトは作成時にTRIE構造化されている。一方形態素解析インタプリタでは、TRIE構造を中間形式辞書ファイルの構成要素の一つとして持ち、辞書エン

トリファイル、左(右)方接続素性ファイル、接続素性ファイルへのポインタを参照しながら辞書引きを行う。表2にコンパイルされた辞書オブジェクトと、中間形式辞書ファイルを用いた解析速度の比較を示す。これは、約1000語の形態素を有する辞書に対しいくつかの例文を解析させたものである。インタプリタ・モードでは

表2: 解析時間の比較

入力文		解析時間 (ms)		
文字数	解析木数	インタプリタ i	コンパイラ c	i/c
16	8	517	62	8.3
45	3	366	42	8.7
64	16	730	97	7.5
83	32	700	126	5.5
88	96	1188	291	4.1

コンパイラ・モードの10倍程度の解析時間で結果を得る。1文当たり1秒またはそれ未満で解析できることから辞書のデバッグに対しては十分実用に適する。

インタプリタの実現により、辞書の変更にもなるオブジェクト生成およびコンパイル作業が不要となり、辞書開発速度を向上することができた。

4 意味構成規則のデバッグ環境の作成

従来の意味構成部の機能は意味構成オブジェクトの生成のみであったが、今回以下の機能を追加した。

1. エディット機能

従来の辞書エディタに、意味構成規則の修正機能を追加した。これにより辞書エディタの機能は、意味構成規則を含む形態素の検索、追加、置換および辞書環境値の変更となった。意味構成規則は中間形式辞書ファイルの構成要素である意味構成規則ファイルに格納され、後述の意味構成インタプリタにより参照される。

図3に辞書エディタのウインドウを示す。

LAX 辞書エディタ				
"自然を守る"	追加	置換	検索	初期値変更 終了
<体言語基>	削除	復活	UNDO	TOP RECURSIVE
私たち	表	層:	自然	
自分たち	左方接続:		end([文節])	
人間	右方接続:		end([文節]),	
地球	意味規則:		派生([体言語基]),	
			句読点([点])	
			[[no(No),	
			morph(Tok,Cat),	
			out(SEM)	
"自然"を追加しました。				

図3: LAX 辞書エディタ

2. 意味構成インタプリタ

意味構成エンジンは形態素解析エンジンの出力結

果である形態素解析木をもとに各形態素が持つ意味規則を順に適用し文節の意味を構成する。意味構成規則は変数宣言と意味処理記述から成る。意味構造とそれに対する操作は意味データ型としてあらかじめ定義しておく。意味処理記述部に記述される操作は全て意味データ型に定義されているものである。インタプリタは、各形態素の意味規則を意味構成規則ファイルから取り出し、変数宣言に記述されている変数を形態素解析木から得られる値で束縛する。次に、意味処理記述の処理をインタプリティブに実行する。

3. トレース機能

意味構成部のデバッグツールとして意味データの操作メソッドレベルでインタラクティブに実行を行うトレース機能を提供した。

図4に意味構成部のモジュール構成を示す。

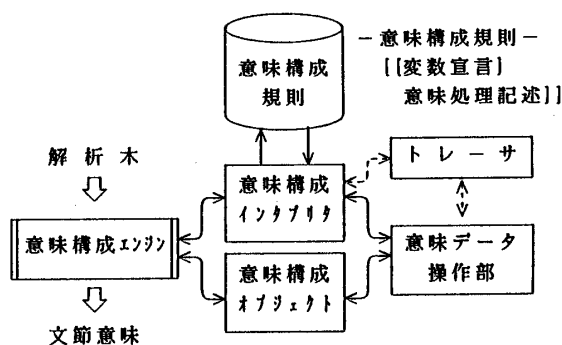


図4: LAX 意味構成部

5 おわりに

本稿では、LAXの解析インタプリタ、意味構成部のエディット機能、同トレース機能について報告した。これらのデバッグ機能を追加することにより、辞書記述の変更後にそれをコンパイルすることなく解析試験が行えるようになった。これにより辞書開発速度を向上することができた。

今後の課題として意味構成部のインタプリタの高速化がある。

参考文献

- [1] 久保: 形態素意味解析システム LAX, 第7回第五世代コンピュータに関するシンポジウム予稿集, 1989年6月
- [2] 久保, 妙泉, 佐野, 赤坂, 杉村: LTB- 形態素解析システム LAX の開発環境, 情報処理学会第37回全国大会, 1988
- [3] 久保: 形態素意味解析システム LAX による語の意味構成, 情報処理学会第39回全国大会, 1989
- [4] 杉村, 赤坂, 久保, 松本, 佐野: 論理型形態素解析 LAX, Proc. of the Logic Programming Conf. '88, pp213-222, 1988
- [5] Sugimura, et al.: A Software Environment for Research into Discourse Understanding Systems, Proc. of the International Conf. on FGCS, pp285-295, 1988