

# 汎用日本語処理系LTBの構成

杉村領一†, 赤坂宏二†, 奥西稔幸†, 久保幸弘†, 幡野浩司†, 田中裕一†, 渡塚孝志††

†(財)新世代コンピュータ技術開発機構 ††KDD研究所

## 1.はじめに

自然言語処理において、「談話処理」は重要な位置を占める。しかし、談話処理に工学的に取り組むには、質・量共に優れた種々のデータやツールが用意されている必要がある。

汎用日本語処理系(Language Tool Box 略称「LTB」)は、自然言語処理技術の研究開発に、「談話処理」の観点から取り組むための論理型プログラミングによる基本的で新しい種々の道具を、統一のとれたプログラム開発環境と共に提供する。

まず、LTBは談話処理に必要な日本語の①基本データを提供する。基本データは、深い意味処理を行う事を目指して設計されている。

意味処理は全て論理型言語CILにより記述される。LTBで提供される全ての処理系は、CILの提供する②基本機能とプログラム開発環境との上に構築されている。これにより、自然言語処理の種々の場面における意味処理を統一的に記述できる。

またLTBは、最近の言語理論の中心であるユニフィケーション文法に基づいた強力な解析/生成システムをLTBの③応用機能として用意している。

以上の①基本データ、②基本機能、③応用機能は、④統合操作環境「LTB-Shell」により、まとめられている。これによりLTBは要素技術の寄せ集め

ではなく、強力な研究開発環境となっている。

図1に、日本語文を解析し、意味構造を得た後、この意味構造から日本語文を逆に生成するように、①②③を用いた場合の関連構成を示す。

## 2.特徴

以下に、各ツールの特徴と研究開発の現状をまとめる。

### ①基本データ

【辞書】 日本語辞書として、解析/生成共通のマスター辞書第1版[1]と、シソーラスが用意されている。

【文法】 日本語形態素文法として、1) 森岡文法に基づいて、2) 正規言語で記述され、3) Unificationにより意味を組み上げる文法を約500ルール作成した[2]。現在、評価を行っている。

日本語構文文法としては、1) 文脈自由文法(Definit Clause Grammar)で記述され、2) Unificationにより意味を組み上げる文法を現在研究開発中である。

### ②基本機能

基本機能はCILにより提供される[3]。データ表現として、Partially Specified Term「PST」を記述できる。PSTは談話処理において中心的な役割を持つ部分意味情報を記述するのに適している。

また、談話における種々の制約を解決する機構と

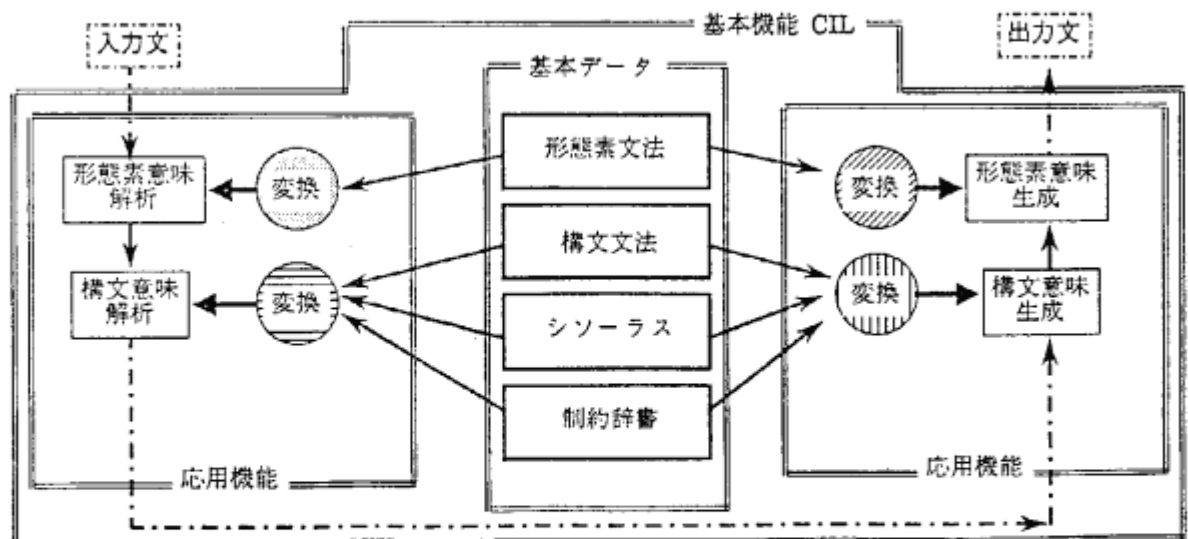


図1 汎用日本語処理系の関連構成

## Configuration of The Language Tool Box

Ryoichi Sugimura†, Koji Akasaka†, Yukihiro Kubo†, Kōzi Hatano†, Yuichi Tanaka†, Takashi Takizuka††

†Institute for New Generation Computer Technology (ICOT) ††KDD Laboratory

して、受動型の命題論理制約解決機構が組み込まれている。能動型の命題論理制約解決機構についても、現在検討が進んでいる。

型記述機構についても、検討を進めている。また、遅延実行型のプログラミングを行うための種々の言語プリミティブも用意されている。

### ③ 応用機能

【形態素解析「LAX」】 LAXは、正規言語による文法記述系と、これを解析実行系(プログラム)へ変換する変換系と、解析実行系の総称である[4]。解析実行系は、辞書引きにTRIE構造[5]を、解探索にはレイヤードストリーム[6]を用いて、高速な処理を決定的に行う。解探索は並列実行が可能である。また、未登録語処理を行える。

【構文解析「SAX」】 SAXは、DCGによる文法記述系と、これを解析実行系(プログラム)へ変換する変換系と、解析実行系の総称である[7]。文法記述系では、遅延実行項が記述できる。解析実行系(プログラム)は、前記レイヤードストリームにより、決定的な解探索を高速に行う。解探索は並列実行が可能である。

【生成】 生成は構文情報をAttribute-Value Pairの集合で表現したものを入力とする。現在文法、辞書の独立が図られている。また、表層情報(語順や語彙)の直接指定も可能である。

### ④ 統合操作環境「LTB-Shell」

統合操作環境から見た汎用日本語処理系の構成を図2に示す。LTB-Shellは、各基本データ、基本機能、応用機能を実現するプロセスの管理を行い、統一されたユーザインタフェースを提供する。

### 3. おわりに

以上、LTBの強力なツール群により、談話処理へ研究の焦点を絞る事が可能となった。

今後LTBとしては、談話理解に必要な基本的な推論機能(例えば、型推論、Conditioned

Unification、Boolean Gröbner Base)の研究開発と基本機能への追加、談話処理に必要な基本データの研究開発、及び、並列化、操作環境の充実、より充実したデータベース等を検討して行きたい。

これらの、研究開発の理論的なバックボーンとしては、状況理論と状況意味論(STASS)を中心に据えて取り組んで行きたい。

### 【謝辞】

LTBの研究開発にあたり、関連メーカーの方々には、多大なるご協力を頂きました。本紙面をお借りして感謝致します。

本研究開発の機会を下さいました、測所長、内田第2研究室室長、吉岡第2研究室室長代理に感謝致します。

### 【参考文献】

- [1] 瀧塚, 杉村, 田中, 赤坂, 佐野, 重永, LTBマスター辞書の構成, ソフトウェア科学会, 論理と自然言語研究会, 88年12月
- [2] 佐野洋, 赤坂宏二, 久保幸弘, 杉村領一: 語構成に基づく形態素解析, 情報処理学会第36回全国大会, 1988.
- [3] Mukai, K., A System of Logic Programming for Linguistic Analysis Based on Situation Semantics, Proceedings of the Workshop on Semantic Issues in Human and Computer Languages, Half Moon Bay, CSLI, 1987
- [4] 杉村領一, 赤坂宏二, 久保幸弘, 松本裕治, 佐野洋, 論理型形態素解析LAX, Proceedings of the Logic Programming Conference 88, pp213-222, 1988
- [5] 上脇正, 田中穂積: 辞書のTRIE構造化と熟語処理, Proceedings of the Logic Programming Conference 85', pp 329-340, 1985.
- [6] 奥村晃, 松本裕治: レイヤードストリームを用いた並列プログラミング, Proceedings of the Logic Programming Conference 87, pp223-232, 1987.
- [7] 松本裕治, 杉村領一: 論理型言語に基づく構文解析システムSAX, ソフトウェア科学会論文誌, 1986.

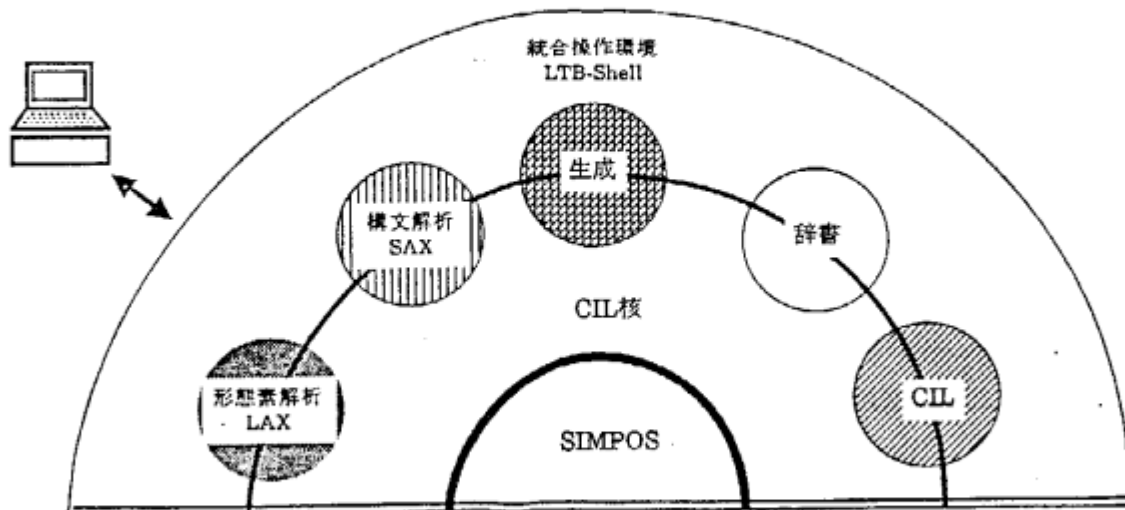


図2 汎用日本語処理系の構成