

# 汎用日本語処理系 LTB

## ICOT

第 2 研究室

杉 村 領 一

## LTB(Language Tool Box)の特徴

---

- 基本データ
  - 日本語辞書(4,000語),
  - 日本語文法(Unification Based Morphology, Syntax, and Semantics)  
Morphology:500 rules, Syntax and Semantics: 100 rules
  - 日本語KWIC
- 基本機能(CIL)
  - データ表現(Partially Specified Term)とライブラリ
    - ◆複雑なデータ表現, 意味ネットワーク
  - 制約解決
    - ◆Constraint Solver for Propositional Logic: 受動型
  - ストリームプログラミング(Co-Routine)
- 応用機能
  - 形態素解析(LAX)
  - 構文解析(SAX)
  - 生成
- 統合操作環境(LTB-Shell)
  - 論理型基本言語CIL上の統合操作環境
  - デバッガ, トレーサ, インタプリタ
  - 簡易Shell

# 基本データ

- 日本語辞書

解析/生成共通のマスター辞書  
シソーラス

- 日本語文法

【形態素文法】

- 1) 正規文法(非決定性有限状態オートマトン)
  - 2) 森岡文法(構造主義言語学)
  - 3) Unification による意味構成
- ◆ 文法数: 500ルール

【構文文法】

- 1) 文脈自由文法(Definit Clause Grammar)
  - 2) 渡辺文法(依存構造文法)
  - 3) Unification による意味構成
- ◆ 文法数: 100ルール

- 日本語KWIC

## 基本機能(CIL)

- データ表現

Partially Specified Term

(例) (関係/会う, 主格/太郎, 目的格/花子, 場所/ICOT, 時間/6月1日, ...)

- 制約解決

Constraint Solver for Propositional Logic: 受割型

(例) >and(A, B, R), R=true.

A=true, B=true

>or(A, B, true), and(A1, C, A), and(B1, C, B)

A=Unbound, B=Unbound, C=Unbound,

A1=Unbound, B1=Unbound

遅延実行型プログラミング

(例) >and(A, B, true), (A=false; B=true), (B=false; B=true).

.....Test.....Generate.....

- ストリーム型プログラミング

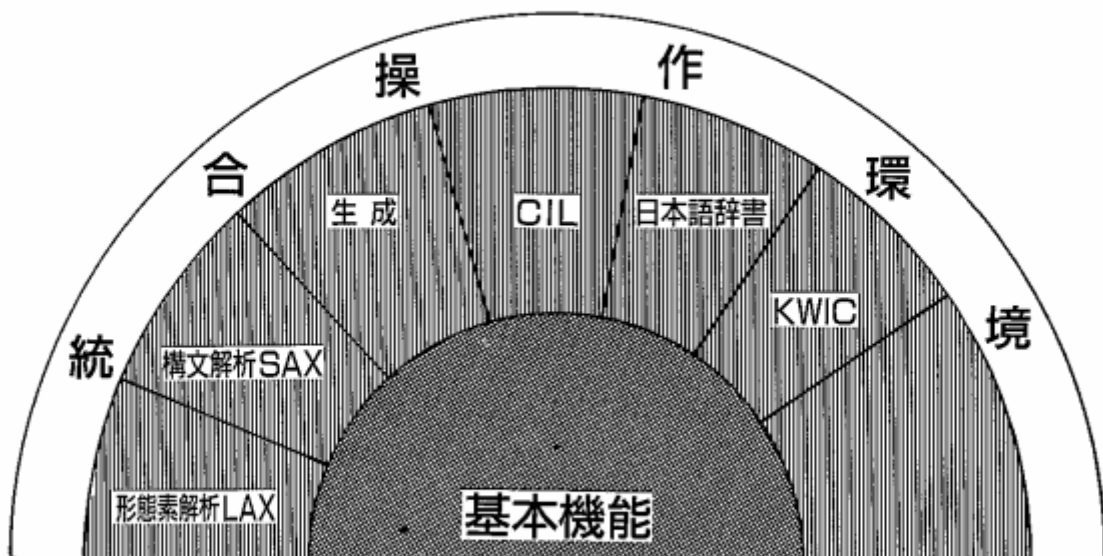
Co-Routine Programming

(例) >input(S1), lex(S1, S2), synsem(S2, S3), output(S3) .

# 応用機能

- 形態素解析LAX
  - 1) 正規言語による文法記述
  - 2) レイヤードストリームによる決定的な解探索：並列化が可能
  - 3) 高速な処理速度(30文/1秒)
  - 4) 解の縮退
  - 5) 未登録語処理
  - 6) 複数種の状態遷移表
- 構文解析SAX
  - 1) DCGによる文法記述
  - 2) レイヤードストリームによる決定的な解探索：並列化が可能
  - 3) 高速な処理速度
  - 4) Discontinuous Grammarの記述
  - 5) 解の縮退
  - 6) 静的●動的デバッグ
- 生成
  - 1) 入力：構文情報を Attribute/Value Pairの集合で表現したもの
  - 2) 表層情報(語順や語彙)の直接指定可能
  - 3) 文法、辞書の独立
- CIL(基本機能に加えて)
  - 1) 強力なマクロ機能
  - 2) 統合デバッグ環境

## 操作環境(LTB-Shell)



# 今後の課題

- 談話理解に必要な基本的な推論機能の研究開発と基本機能への追加

Type Inference  
Constraint Unification  
Boolean Gröbner Base

- 談話処理に必要な基本データの研究開発  
    Situation Theory Situation Semantics. (STASS)

- 並列化
- 操作環境の充実
- 検証のための、より充実したデータベース

