

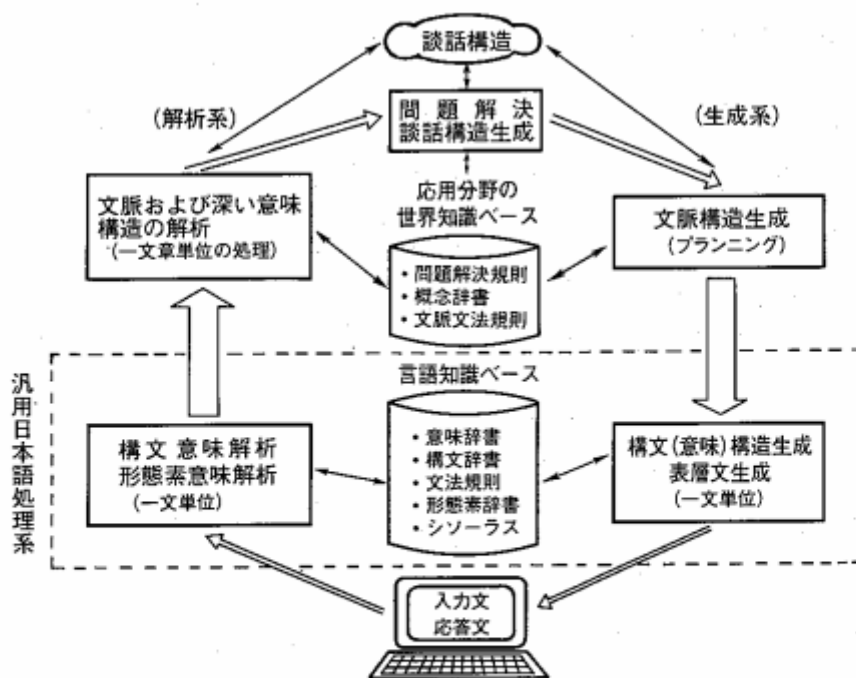
形態素意味解析システムLAX

ICOT

第2研究室

久保幸弘

談話理解実験システムと汎用日本語処理系



形態素意味解析とは

- 分かち書きされていない日本語の文章から文節(単語)を切り出す

(例)

“人間が地球で生きる。”

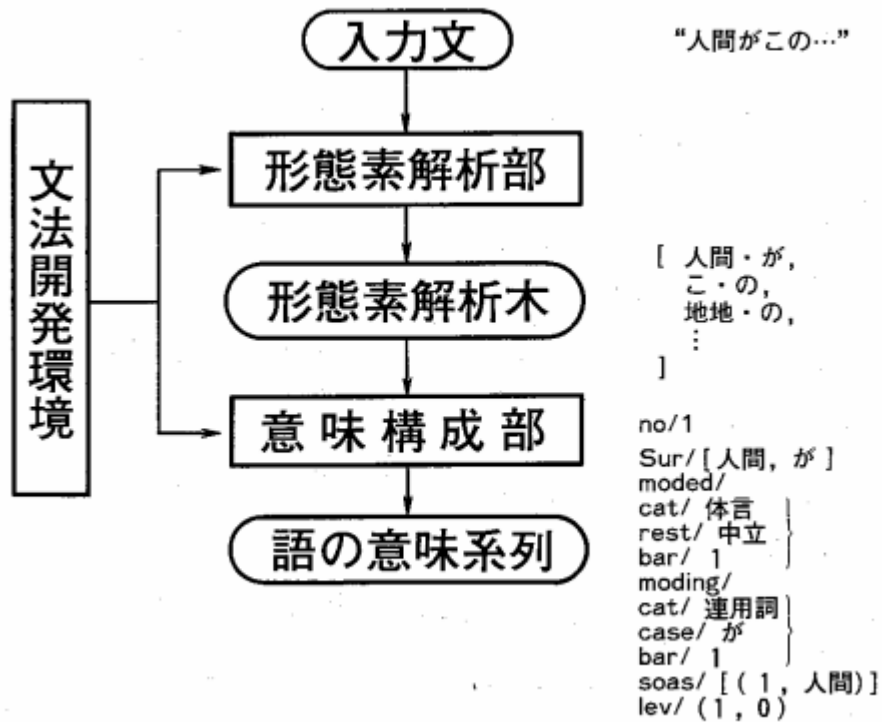
→人間・が/地球・で/生・き・る・。

- 各形態素の意味から文節の意味を構成する
文節の意味
 - 構文的情報
 - 意味的情報

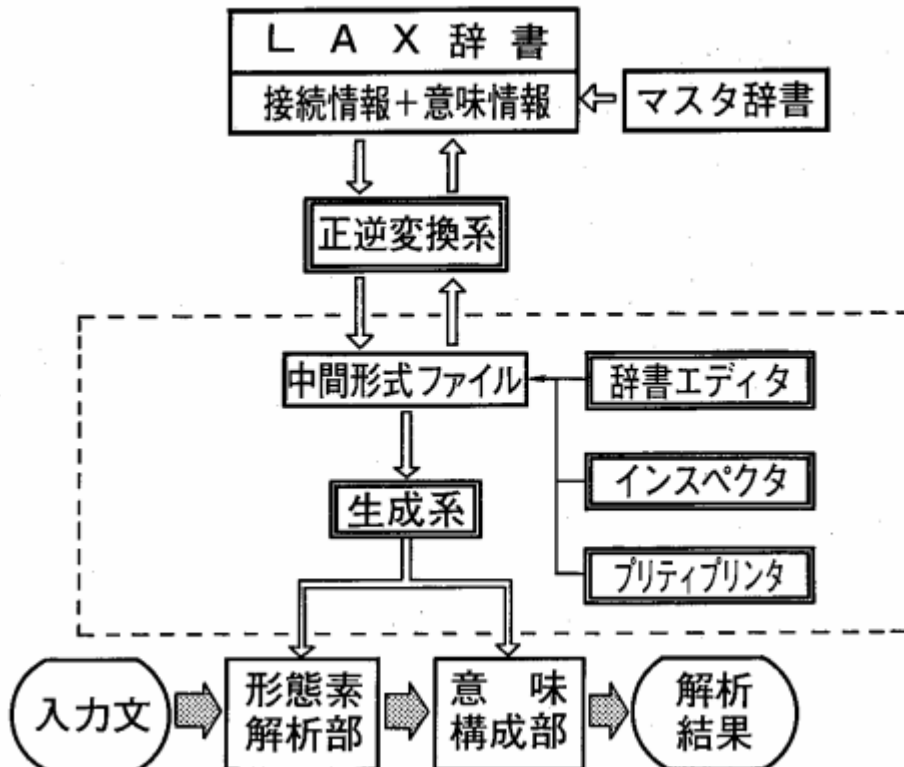
LAX(Lexical Analyzer for Syntax and Semantics)の特徴

- 逐次型推論マシンPSI-IIによる実行
- 汎用日本語処理LTB (Language Tool Box)のモジュール
 - 構文意味解析システムSAXとの併用
- 高速な形態素意味解析処理
 - 1文当たり100msec以下
- レイヤード・ストリーム方式による解探索
 - 決定的な解析処理で全解を算出
 - 並列論理型言語KL-1で実装 → 並列化可能
- TRIE構造による辞書検索
- ローカル・アンビギュイティ・パッキングによる解の縮退
- 未登録語の処理が可能
- 意味の構成が可能

形態素意味解析の流れ



文法開発環境



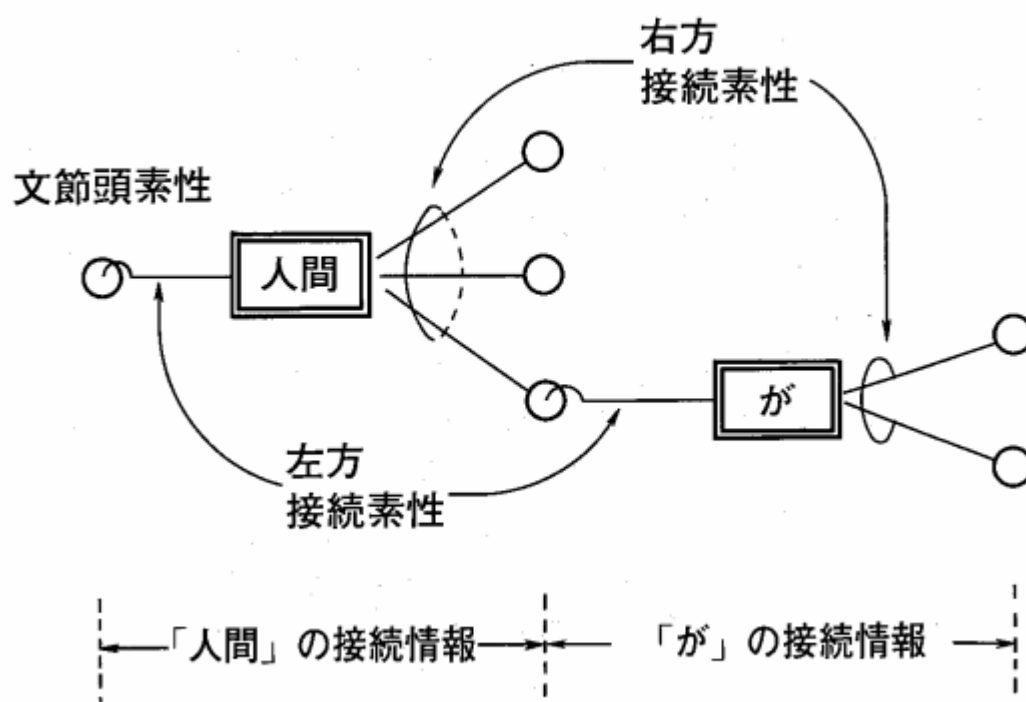
文法開発ツール

- **インスペクタ**
 - 形態素解析結果の簡単な表示
 - 形態素の候補とその辞書記述の表示
 - 接続試験
- **専用辞書エディタ**
 - 形態素文法と意味構成規則の編集
 - エントリの検索，更新，削除，新規登録
 - インスペクタと連携したデバッグ
- **プリティプリンタ**
 - 形態素解析結果のプリティ・プリント

形態素文法記述の特徴

- **森岡文法(構造主義言語学)**
 - 形態素：有意義の最小単位
 - 語(文節)：語基，接辞，助辞から構成
- **正規言語に基づく日本語向け形態素文法記述**
 - 各形態素に左方接続素性，右方接続素性を定義
 - 非決定性有限状態オートマトンの状態遷移関数の定義
 - 複数の状態遷移表
- **ユーザ定義のデータタイプによる意味構成**
 - 形態素の接続状況に応じたきめ細かい処理

形態素の接続可能性の判定



文法記述例 — 形態素解析部 —

人間：： end([文節])
 && [体言派生([接辞, 準体辞, 副接辞]),
 体言屈折([格助辞]),
 体言準用([準用辞]),
 情態派生([ダ系体言]),
 屈折([副助辞, 連体助辞, 係助辞,
 終助辞]),
 句読点([点, マル]),
 end([文節])]

文法記述例 — 意味構成部 —

```
$$ [ [ no(No), morph(Mor, Cat), out(SEM) ],  
      make(SEM),  
      role (SEM, [no], No),  
      role (SEM, moded, cat], 体言),  
      role (SEM, [moded, rest], 中立),  
      get_id(ID),  
      tail(SEM, [soas], (ID, Mor)),  
      cond( { [ 句読点([マル]) ],  
              [ role(SEM, [moded, bar], 4) ] },  
            { [ end([文節]), 句読点([点]) ],  
              [ role(SEM, [moded, bar], 1),  
                role(SEM, [moding, bar], 1) ] } ) ] ]
```

形態素解析の評価

○ 使用マシンおよびOS

- PSI-II
- SIMPOS 4. 2 版

○ 対象テキスト

- 『自然を守る』 82文
- 『キュリー夫人』 126文

○ 形態素辞書

- 付属語 500語
- 自立語 1,230語
- 『自然を守る』 450語
- 『キュリー夫人』 780語

課 題

- **意味構成部のデバッグ環境**
 - － LTB標準意味操作モジュールの設計
 - － 同インタフェースと意味構成部デバッガの作成
- **マスタ辞書へのフィードバックと文生成部との辞書の共通化**
- **大語彙の形態素辞書作成と評価**
 - － 素性の見直しと分類
 - － 60,000語の辞書作成
- **各種アプリケーションへの対応**
 - － DUALS
 - － KWIC
- **並列化アルゴリズムの検討**
 - － マルチPSIへのインプリメント