

基礎ソフトウェア・システム

(自然言語処理ソフトウェア)

ICOT

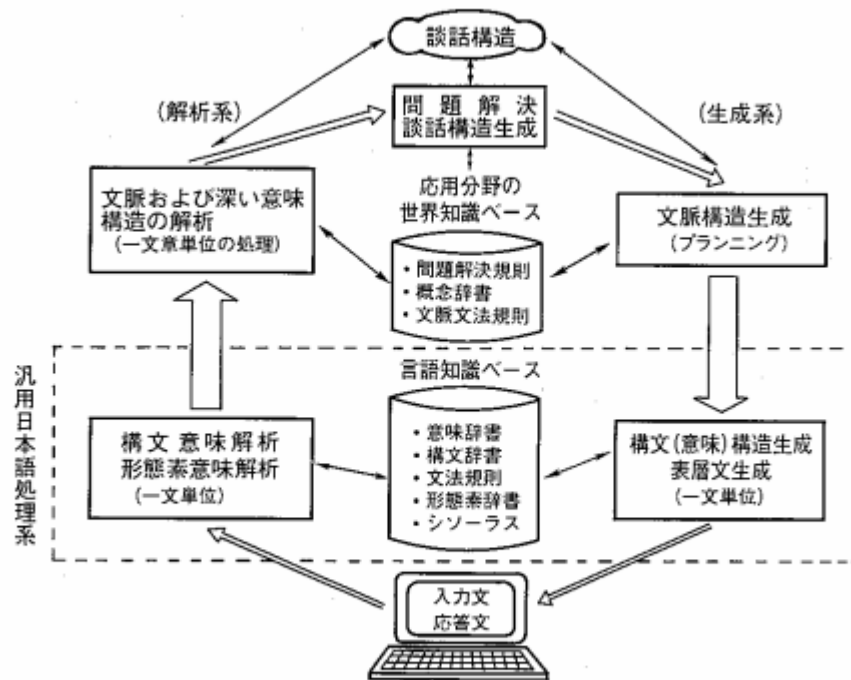
第2研究室

吉	岡	勉
橋	田	浩一
田	中	裕一
内	田	俊一

研究開発内容

1. 談話理解実験システム (DUALS)
2. 汎用日本語処理系 (LTB)
3. 要素技術実験パイロットシステム

談話理解実験システムと汎用日本語処理系



談話理解実験システムの研究開発

DUALS : Discourse Understanding

Aimed at Logic-based Systems

1. DUALS-I (59, 60年度)

- 規模：18文，200単語，基本言語：Prolog
- 形態素解析，構文解析などの統語レベル処理の論理プログラミング上での構築手法の確認

2. DUALS-II (60, 61年度)

- 規模：18文，200単語，基本言語：CIL
- 構文解析手法(SAX)の確立，意味記述言語CILの導入
- ESPによるPSI上での再構築

3. DUALS-Ⅲ (62, 63年度)

- 規模：200文，2,000単語，基本言語：CIL
- 生成を含む統語レベルの処理手法のまとめと汎用日本語処理系（LTB）としての共通ツール化
- LTBを実験環境とし，文脈処理や深い意味理解の研究への着手



- 状況意味論や制約論理プログラミングなどの基本的な枠組の有効性の確認
- 各モジュール間の共通的知识表現形式の必要性や，照応，省略，テンスなどの深い意味処理や文脈処理を扱う上での問題点などの明確化

汎用日本語処理系の研究開発

LTB : Language Tool Box

文脈処理，意味処理の研究の共通ツールとなるソフトウェア群をライブラリ化

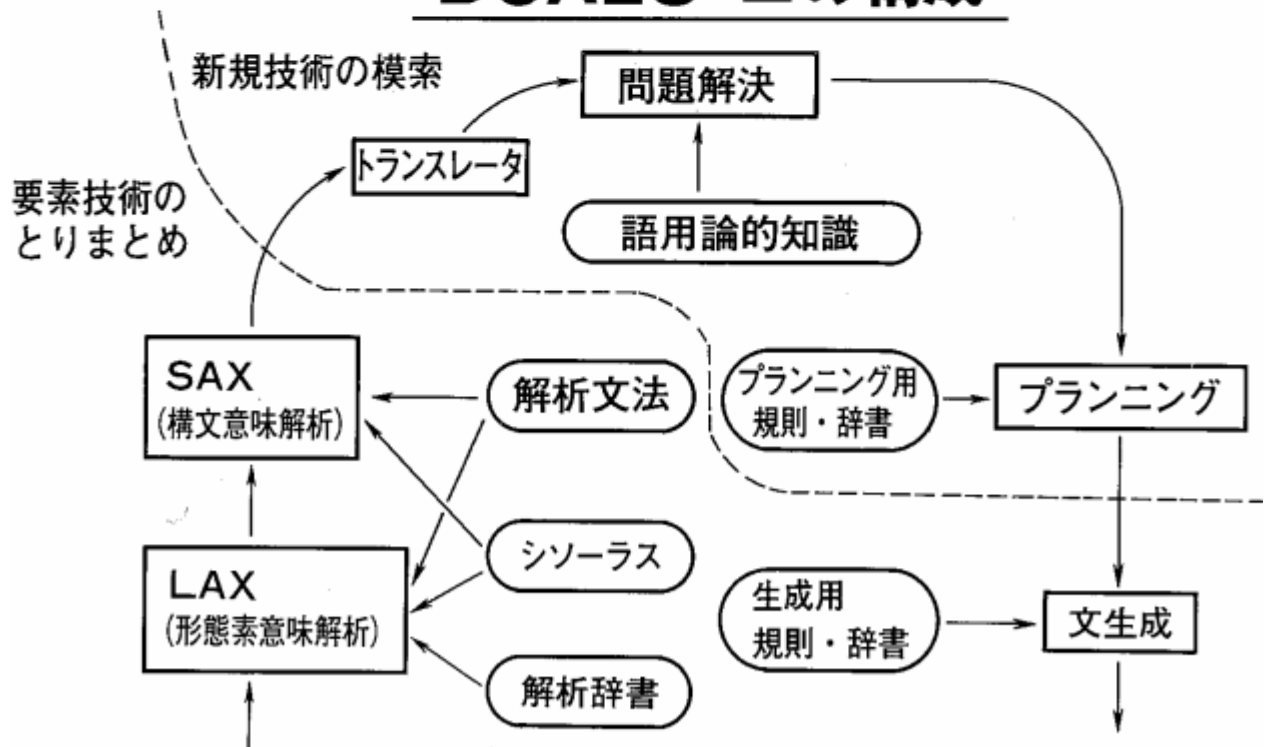
- 言語知識ベース，KWIC
- 意味記述言語(CIL)とそのプログラミング環境
- 形態素/構文意味解析系 (LAX/SAX)
- 生成系
- これらを統合する統合利用環境 (LTB-Shell)

要素技術実験パイロットシステム

談話モデル，意味処理，意味記述，解析 / 生成などの要素技術を，いろいろな用途を想定した実験システムを試作して研究開発する。

- 知的情報検索実験システム
- 意味談話理解のための知識表現実験システム
- 状況意味論に基づく対話モデル実験システム
- 文脈理解に基づく対話管理実験システム
- 世界知識に基づく要約支援システム
- 意味構造を用いた入力解析・修正支援システム

DUALS-Ⅲの構成



DUALS-IIIの成果

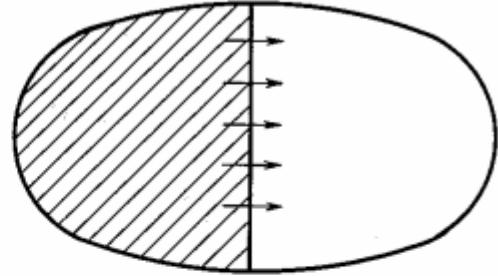
- 手続き的パラダイムの集大成
 - LAX
 - SAX
 - 文生成
- より高度な談話処理への助走
 - 問題解決
 - プランニング

既存技術の改良すべき点

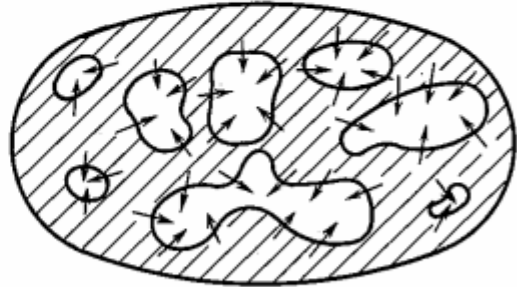
- 異種パラダイムの混在
 - 手続きパラダイム……………LTB
 - 制約パラダイム……………問題解決
- 処理モジュールの分割 ←— 手続きパラダイム
 - 情報の流れの限定
 - システムの複雑化
 - 共通であるべきデータの分割

情報の流れ

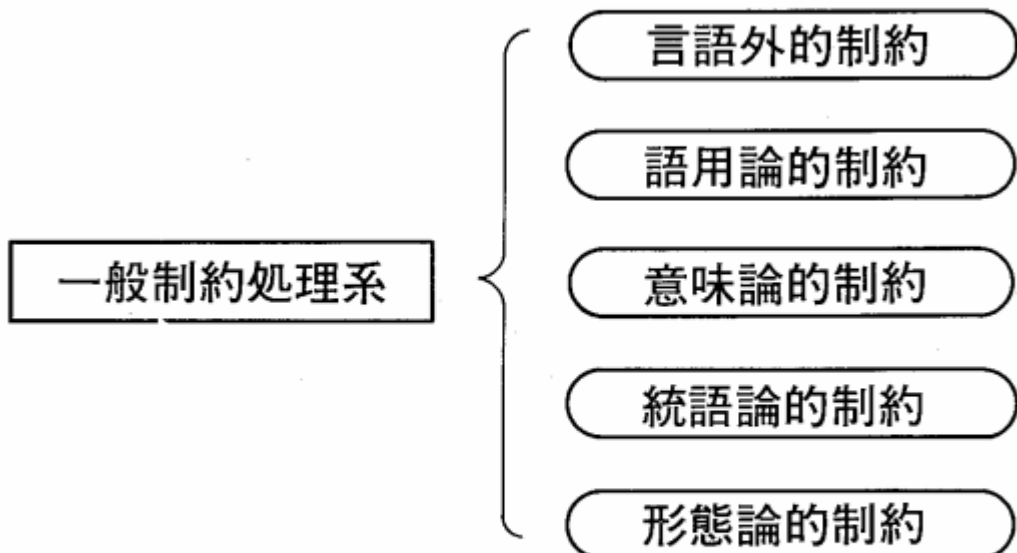
完全情報問題
↓
手続きパラダイム



部分情報問題
↓
制約パラダイム



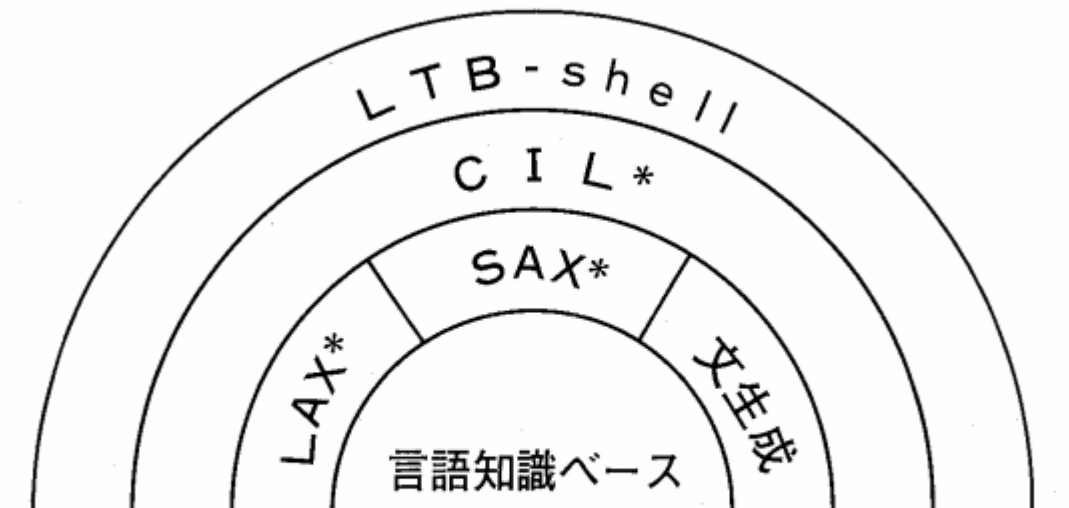
DUALS-IVの枠組



DUALS-IV

- 制約パラダイム
 - DUALS-IIIの問題解決モジュール
 - 文解析, 文生成, 推論等の統合
 - 計算 = 制約の変換
- 領域知識としてのヒューリスティクス
- 理論的成果の具現化
 - 状況意味論
 - 日本語句構造文法

汎用日本語処理系LTBの構成



知識表現言語 CIL の特徴

- 部分項 (レコード)

例 $R = \{\text{関係} / \text{愛する}, \text{動作主} / \{\text{名前} / \text{健}, \text{性別} / \text{男}\},$
 $\text{対象} / \{\text{名前} / N, \text{性別} / \text{女}\}\},$

$X = R! \text{動作主}, Y = X! \text{名前}, \text{etc.}$

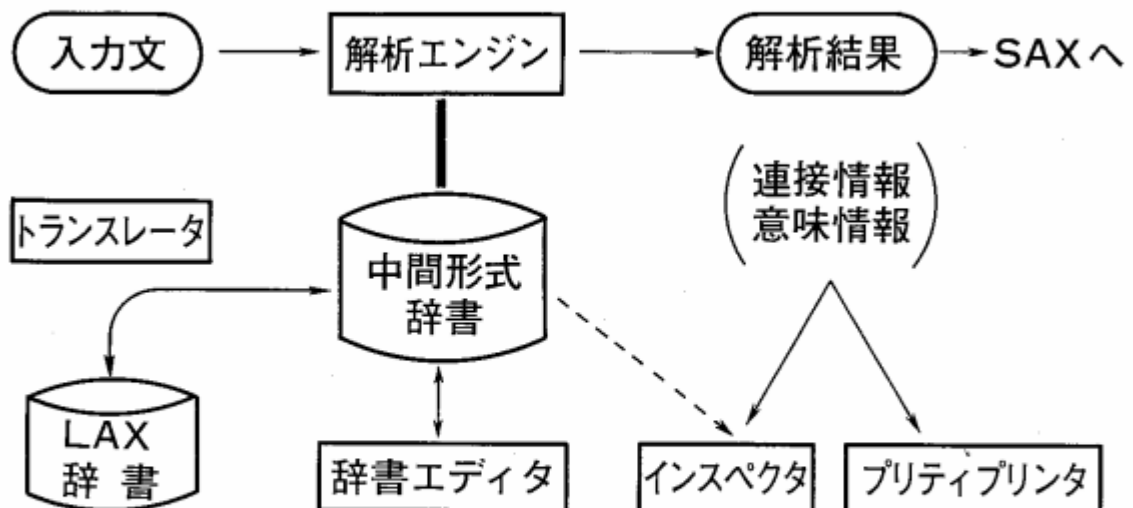
- 遅延実行制御 (フリーズ)

例 $\text{print}(X?), N@\text{odd}, \text{etc.}$

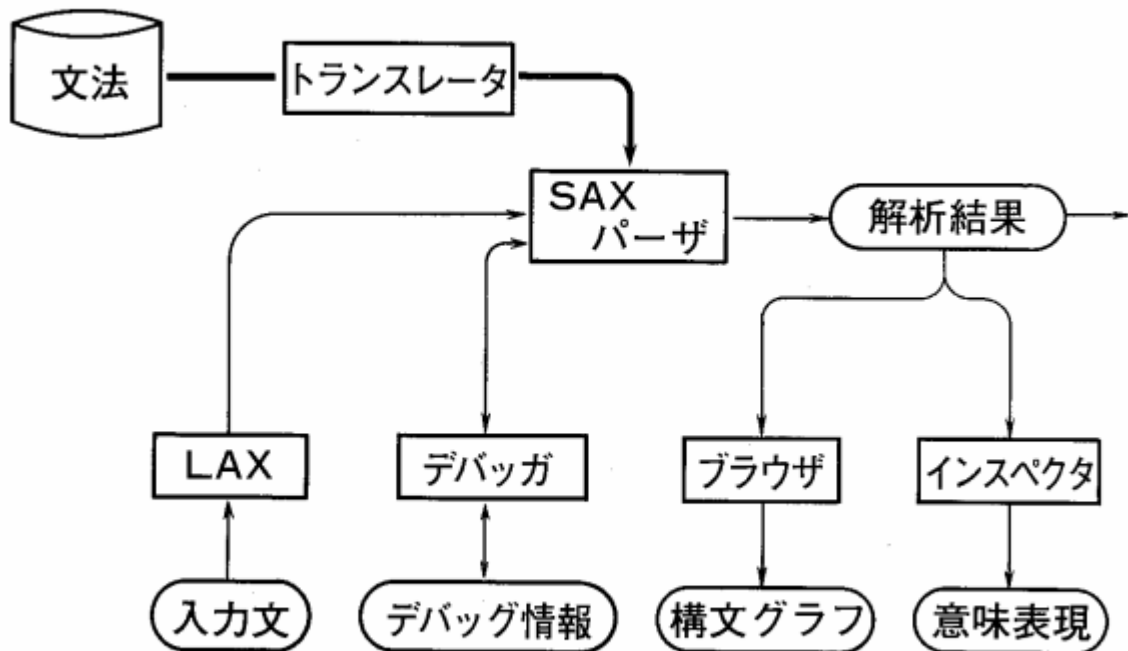
- プログラミング環境

デバッガ, インスペクタ, etc.

形態素意味解析システムLAXの構成



構文意味解析システムSAXの構成



第2 研究室後期計画の概要

1. 談話理解モデルの構築

談話に伴う言語現象の記述 / STASS とその計算モデル /
談話理解のための共通的枠組みの設定

2. DUALS の開発

STASS の実験モデル / 制約を利用した統一的処理の実現

3. LTB の改良・拡張・並列化

リリースツールとしての日本語処理系の改良・拡張 /
次世代 LTB 技術の研究 / 要素技術の並列アルゴリズム

談話理解モデルの構築

1. 談話理解のための共通的分フレームワークの設定
 - 共通的分知識表現の検討
 - 談話理解関連テーマの検討 (対象問題の設定とモデル化)
 - STASS を中心とした談話理解モデルの数学的分側面の検討
2. 個々の対象問題 (モデル) の記述
3. 小規模プログラムによる試験的分実証
4. DUALS を利用したモデルの検証

後期分の LTB の研究開発の進め方

1. リリース版 LTB の改良・拡張
 - 高速化・機能拡張
 - マンマシンインタフェースの向上
2. 要素技術の並列アルゴリズムの研究
 - LAX と SAX の並列化 (既存技術の応用)
 - 新しい並列アルゴリズムの研究
3. 次世代 LTB 技術の研究 (SIG)
 - 新アルゴリズムの開発
 - LTB 上での実験環境の構築

SIG の研究開発内容

1. 新アルゴリズムの開発

- 解析・生成アルゴリズム
- 統一的制約処理システム

2. LTB 上での実験環境の構築

- 文生成プランニング
- 質問応答システム
- 言語 KB 記述内容の検証