「高度メディア社会の生活情報技術」
平成 13 年度採択研究代表者

池原　悟

（鳥取大学工学部　教授）

「セマンティック・タイポロジーによる言語の等価変換と生成技術」

1．研究実施の概要

本研究は、新しい原理に基づく言語処理の基礎を確立しようとするもので、「人間の対象把握作用には、思考形式とも言うべきある種のフレームワークが存在し、それが言語表現に反映される」と言うセマンティック・タイポロジ（意味類型論）の観点から、言語の等価変換と生成の技術を実現する。

本年度は、平成 14 年度からの言語知識ベースの構築作業開始に備えて、以下の2点について準備を進めると共に、本研究の基本方式について個別的な検討を進めた。

(1)　共同実験環境の整備

国内各地の研究機関に所属する研究者との共同研究を進めるために、共同実験環境の整備を行った。言語知識ベース開発用の計算機と言語変換等の実験用計算機などを設置すると共に、研究メンバー、および、アナリスト等の作業員と上記計算機システムとのネットワークの設定などを行った。

(2)　日英対訳コーパスのデータベース化

日英対訳標本（約 100 万文あまり）を収集し、そのうち、40 万文についてコード変換などを行った後、出典情報などを付与し、標本データベースを作成した。また、そのデータベースに対し文種別（単文、複文、複重文、疑問文、会話文、その他）に分類マークを付与したのち、重文、複文、複重文の3種類について、文字コードのチェックなどを行い、本研究で文型抽出用の標本セット（約9万文）を作成した。

(3)　個別的な検討

上記の作業と平行して、以下の検討を進めた。

①　方式概念の検討

日英両言語の意味類型パターン間で概念項を介した写像方式の具体的方法について検討した。

②　重文複文のパターン化例題検討

比較関係構文、因果関係構文の用例を対象とした意味類型パターン化の方法と意味的マッピング方式の検討を行った。

③　パターン記述方式と記述言語の関する検討

正規表現で記述した時の記述能力について検討した。

④　パターン検索アルゴリズムの検討

入力日本文に対して、離散的な表現パターン辞書から、該当するパターンをすべて同時に抽出するアルゴリズムの検討を行った。

来年度は、引き続き、重文・複文を対象とした日英対訳標本データの収集分類を進めると共に、今年度作成した標本データを対象に、第1次パターン化の作業を進める。また、第2次パターン化以降のパターン記述法を明らかにするための検討を開始する。

3. 研究実施体制

研究統括グループ

①　研究分担グループ長　　池原悟（鳥取大学工学部　教授）

②　研究項目　　基本方式の検討

等価変換システムグループ

①　研究分担グループ長　　宮崎　正弘（新潟大学工学部、教授）

②　研究項目　　日英等価変換方式

言語生成システムグループ

①　研究分担グループ長　　奥村　学　（東京工業大学、助教授）

②　研究項目　　日本語生成方式

意味類型知識ベースグループ

①　研究分担グループ長　　池田　尚志（岐阜大学工学部、教授）

②　研究項目　　知識ベース構成法