

「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

高度なネットワーク社会において言語バリアフリーの通信を実現することは、緊急かつ重要な問題となっている。しかし、従来の要素合成法を基本とする自然言語処理技術は、すでに技術的限界に近いレベルにあり、これを応用した機械翻訳の品質もほぼ限界と見られる。

本研究は、この限界を突破するため、新しい原理に基づく言語意味処理の基礎を確立しようとするものである。「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枠組みが存在し、それが言語表現に反映される」と言うセマンティック・タイポロジー（意味類型論）の観点から、言語表現の構造と意味の関係を意味類型化された言語知識データベースとして体系化し、言語の等価変換と生成の技術を実現する。

具体的には、「文型パターン辞書の研究開発」と「意味類型パターン辞書の研究開発」の2つの課題を実現する。第1の課題は、大量の日英対訳例文の線形要素をさまざまな方法で汎化し、日英対訳文型パターン化することであり、第2の課題は、得られた日英文型パターンを意味的に類型化し、マッピング関数によって意味類型間の対応関係を決定する方法を実現することである。

現在までの成果は以下の通りである。まず、約30種類のドキュメントから100万件の日英対訳文を抽出し、その中から述部2または3カ所を含む重文、複文、複重文を合わせて15万件抽出すると共に、文型パターン記述言語を設計した。それに基づいて、上記15万件の対訳例文を対象とする単語レベル、句レベル、節レベルの汎化作業を行い、「日英文型パターン辞書（文型数22万件）」を作成した。また、「文型照合プログラム」を試作して文型パターン辞書の被覆率特性を調査し、文法レベルでの文型パターン記述方法の問題点を分析した。その結果に基づき、「意味レベル文型パターン辞書」を試作した。

次に、意味類型化の研究では、まず、重文複文を構成する節の意味と節間の関係の意味を分類するための2種類の意味分類体系（いずれも深さ5段、分類数約370種類）を完成した。引き続き、全文型パターンにこれらの意味分類コードを付与することにより、「日英意味類型パターン辞書（暫定版）」を作成した。また、これらの意味分類コードを使用した

文型パターン意味検索プログラムを実現した。

現在、「意味レベル文型パターン辞書」の被覆率の向上を図るため、さまざまな改良検討とそれに基づくブラッシュアップ作業を実施しており、その成果が見え始めている。また、「意味類型パターン辞書」では、それを使用したさまざまな意味検索実験を実施中であり、今後その結果を分析し、残された期間内で可能な限りの改良を行う予定である。

2. 研究実施内容

今年度実施した主な研究項目は以下の通りである。

(1) 文型パターン辞書の被覆率向上対策

本プロジェクトの第1の課題である「文型パターン辞書の研究開発」では、文型パターン辞書の文法的意味的被覆率について様々な実験評価を行った結果、「意味的に不適切な文型パターンへの適合が多いこと」が分かった。実験結果の分析では、これは、「離散記号の付与基準に問題があること」、および、「変数の意味的制約条件の体系が適切でないこと」が大きな原因であると推測される。また、線形要素の汎化技術の検討から、時制様相など「半自動化が可能で意味的な被覆率の向上が期待できる項目が存在すること」が分かった。

そこで、文型パターン記述方式を見直し、大幅な改良を行った。これに伴い、文型パターン記述言語仕様を大幅に機能拡張し、文型パターンパーサへの機能追加と大幅な改良を行った。主な改良課題は以下の通りである。

○「文型適合率」向上の対策

- (1) 述部語尾表現（時制・相・様相）の汎化
- (2) 局所言い換え表現辞書による文型パターンの汎化の検討
- (3) 標準表記への統一と選択記号の適正化作業
- (4) 文型任意記号の適正化作業
- (5) 自己パターン適合性向上作業
- (6) 節変数化の適用範囲の見直し作業

このうち(1)は、半自動化を前提としたとき、残された汎化項目の中で、最も効果が期待できると予想される課題で、切り札と言えそうである。(2)(3)は、細かなレベルでの部分的表現の「揺らぎの解消」を狙ったものである。また、(4)~(6)は評価の過程で問題となった点に対する改良作業である。

○「適合文型意味正解率（P1）」向上の対策

- (1) 用言意味属性新体系の構築と付与作業
- (2) 名詞の意味属性新体系の構築と付与作業
- (3) 離散記号の付与基準の見直し作業

○英語文型パターンの改良準備

文型パターンの評価実験では、日本語側が適合した文型パターンでも英語側の文型パ

ターンの不適切な記述が原因で使用できないものがかなり存在することが分かった。この問題が解決すれば、「適合文型意味正解率（P 1）」は最大、20%程度向上する可能性があると予測される。従来は、英語文型パターンの記述方法についての詳細な検討を行っておらず、これは、中間評価後の大きな発見の一つである。そこで、

- ①英語表現の骨格が分かり易い記述法にすること
- ②文型パターンの意味的な被覆率を向上させること

を目標に、英語文型パターンの記述方式を体系的に見直すこととした。これらの目標を達成するには、英語パターンの定義方法について大幅な改良作業が見込まれる。そこで、それらの作業が半自動的に実施できるようにするため、（１）英語構文情報の付与と、（２）英語文型パターンの適正化作業を実施した。

（２）文型パターン辞書の意味類型化

意味類型化の目的は、本プロジェクトの第２の目標としている「日本文（重文複文）に意味的に対応する複数の英語表現を発見する仕組みを実現すること」である。最終的には、マッピング関数によって意味類型間の対応関係を動的に決定する方法の実現を狙っている。検討の結果、いくつかの文型パターンの意味分類体系を作成したが、それらの意味コードを文型パターン付与するには、人手によらざるを得ないため、工数的に困難であることが明らかとなった。そこで、半自動的に付与可能な意味分類コード体系として、①節間の意味分類体系と、②構文意味コード体系を作成してすべての文型パターンに付与した。この結果、意味類型化方式検討のための実験環境が実現されたので、今後は、意味分類体系の改良を図っていく予定である。

（３）文型パターン検索プログラムの試作

任意の日本語入力文に対して、文型パターン辞書からそれに適合する文型パターンを検索するためのプログラムは、文型パターン辞書を改良していく上でも必須のツールである。このようなツールとして、以下の２種類の作業を実施した。

- (1) 文型パターンパーサの機能拡充
- (2) 文型パターンの意味検索プログラムの新規試作

すでに、入力文と文型パターンを照合するための「文型パターンパーサ」については、一通りの試作を終了し、それを使用してきたが、時制様相を初めとする文型パターン記述方式の改良に伴って、文型パターン記述言語仕様を大幅に拡張した。(1)は、これに合わせて、文型パターンパーサの大幅な機能拡張を行ったものである。

（４）翻訳簡易実験システムの試作

簡易実験システムの試作を開始した。このシステムは、変数、関数などの部分に対する英語を生成するもので、文型パターン辞書を使用することによって生成される英文の骨格

を確認するために使用する。

そのために今回実施した課題は以下の5項目である。

- (1) 文型パターン照合条件の最適化の検討
- (2) 最適文型パターン選択方式の検討
- (3) 変数にバインドされた表現の翻訳方法の検討
- (4) 大域文型制約を用いた英文生成部の検討
- (5) 構造化された日英文型パターンを用いた日英構造変換部の検討

文型パターンは入力文のすべての要素に対する解釈を与えるものでないため、任意化された要素などに対して、現実の機械翻訳では、別途翻訳する仕組みを実現することが必要である。また、冠詞などの形態素生成も必要であるが、それらの機能については、「等価変換システムグループ」での試作を予定している。

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、教授）
- ②研究項目：基本方式の検討

「等価変換システム研究」グループ

- ①研究分担グループ長：宮崎 正弘（新潟大学、教授）
- ②研究項目：言語等価変換システムの試作実験

「言語生成システム研究」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：言語生成システムの試作実験

「意味類型知識ベース研究」グループ

- ①研究分担グループ長：池田 尚志（岐阜大学、教授）
- ②研究項目：意味類型知識ベースの研究開発

4. 主な研究成果の発表

(1) 論文（原著論文）発表

- Nguyen My Chau, 池田尚志：日本語―ベトナム語機械翻訳における連体修飾構造の翻訳, 自然言語処理 Vol.12, No.3, pp.145-182 (2005.3)
- ト朝暉, 浅井良信, 宇野修一, 池田尚志：日中機械翻訳における連体修飾節構造の翻訳について, FIT2005(第4回情報科学技術フォーラム), pp.73-76 (2005.9)
- Katsuyuki Shibata: Dependence on Context in case of English-Japanese Machine

- Translation I-1, Fukuoka University Science Reports, 2005. Vol.35, No.1, pp.75-79
- Katsuyuki Shibata: Dependence on Context in case of English-Japanese Machine Translation I-2, Fukuoka University Science Reports, 2005. Vol.35, No.2, pp.85-88
 - 池原悟：意味類型論による言語の等価的変換方法の研究、TELECOM FRONTIER、SCAT-Technical Journal, No.48、pp.4-12, 2005 Summer
 - 新田義彦：Web 学習における教材提示方法の検討（How to Manipulate Teaching Materials in Web-Learning Environment），日本大学経済学部、経済集誌、Vol. 75, No. 1, pp.1-20, Apr-2005