

## 平成25年度北海道大学情報基盤センター共同研究成果報告書

1. 研究領域番号 A2 大規模問題解決の基盤技術
2. 研究課題名 機械学習と論理計算の融合による自然言語意味理解の研究
3. 研究期間 平成25年 4月 1日 ~ 平成26年 3月31日
4. 研究代表者

氏名	所属機関・部局名	職名	備考
麻生 英樹	産業技術総合研究所・知能システム研究部門	上級主任研究員	

## 5. 研究分担者

氏名	所属機関・部局名	職名	備考
赤間 清	北海道大学・情報基盤センター	特任教授	
小林 一郎	お茶の水女子大学・理学部	教授	
小池 英勝	札幌学院大学・社会情報学部	准教授	
Peerasak Intarapaiboon	Thammasat University	講師	

## 6. 共同研究の成果

下欄には、当該研究期間内に実施した共同研究の成果について、その具体的内容、意義、重要性等を、共同研究申請書に記載した「研究目的」と「研究計画・方法」に照らし、800字~1,000字で、できるだけ分かりやすく記載願います。文章の他に、研究成果を端的に表す図表を貼り付けても構いません。なお、研究成果の論文・学会発表等を行った実績（発表等の予定を含む。）があれば、あわせて記載して下さい。

WWWなどの普及に伴い、大規模なテキストデータやテキストと画像のペアデータなどが利用可能になっている。それに伴い、コーパスベースの統計的な自然言語処理の研究が盛んになり、形態素解析、係り受け解析、統計的機械翻訳など様々な問題で成功を収めている。しかしながら、それらの多くは統語解析に近いレベルであり、多様な文体の自然言語文の深い意味理解を実現するための技術は未だに確立されていない。自然言語の本質は無限のバリエーションを持つ現象を記述することができる組み合わせ的な生成性にあり、本課題研究では、そうした性質を捉えるために、統計的なアプローチと組合せ構造に対する論理計算によるアプローチを融合させ、新しい意味理解方式を確立することを目的とした研究を実施した。

こうした研究はまだあまり行われていないが、先行研究の一つとして、Percy Liangらは、依存構造ベースの意味表現（Dependency-based Combinatorial Semantics: DCS）を拡張して集合論的な意味内容を扱うための方法を提案し、質問応答タスクの質問文と応答文のペアから、表層言語文を意味表現に変換する手続きを学習する方法を提案している。この研究は、組み合わせ構造を持つ意味表現と統計的なアプローチを融合している点で興味深いものだが、そこで用いられている組み合わせ的な意味表現 DCS は、表層文に近い語順を持つ簡単なグラフ構造によって量化のスコップなどを扱うことを可能にしているものの、論理的な推論などには適していない。そこで、まず、Liangらの研究をベースとして、そこに含まれる論理的な構造変換処理の部分を、より強固な論理計算的基盤である等価変換理論に基づく論理計算体系の上で再構築することを目指した。

具体的には、Liang らが質問応答タスクの対象にしている米国地理情報データベースの情報（Prolog 様の言語で記述されているもの）を赤間が開発している等価変換理論に基づく計算言語 ET で取り扱えるデータ構造に変換した。また、タスクで用いられている英語の質問文データを日本語に翻訳し、その意味表現を、同じく ET で取り扱える意味表現構造によって記述した。さらに、これらのデータから、自然言語表現と意味表現の中の頻出対応関係を抽出し、自然言語表現の部分パターンに対応する意味表現候補を割り当てた辞書を、自動で構築する手法を提案した。これらによって、自然言語による質問文を、言語によらず、また事前知識を用いずに論理計算的に妥当な意味表現構造に自動変換し、それを用いてデータベース検索等の論理的推論を実行するための新たな枠組みを構築した。