

## 審査の結果の要旨

氏 名 綱 川 隆 司

本論文は、「PIVOTAL APPROACH FOR LEXICAL TRANSLATION (中間言語を用いた辞書の翻訳)」と題し、5章より構成される。

言語横断的な自然言語処理において、対訳辞書は構築に手間がかかるが必須の言語資源である。本論文では機械翻訳のための対訳辞書の拡充を目的として、同義語の抽出および中間言語を用いた対訳の獲得について論じている。

第1章は、「Introduction (序章)」である。対象領域の対訳コーパス (parallel corpora) に基づき翻訳知識を獲得するデータ駆動型の機械翻訳の有効性を指摘するとともに、任意の2言語間で必ずしも十分な量の対訳コーパスが利用できるとは限らないことを問題点としてあげている。このような場合に、対訳関係が対応づけられていない2言語コーパス (comparable corpora) から分野固有の翻訳知識を獲得することが課題となる。これを受けて、特に日本語と中国語について、専門用語の翻訳モデルを獲得する手法を確立することを本論文の目標として設定している。続いて、統計的機械翻訳モデルを専門用語の翻訳に適用する際の課題を明らかにし、中間言語による機械翻訳アプローチ、2言語コーパスとしての電子化対訳辞書の利用、同義語と対訳語の選択など、本論文の中心となる考え方を解説して、本論文の構成を記している。

第2章は、「Background (背景)」と題し、機械翻訳の歴史と現状を概観するとともに、統計的機械翻訳の基礎を解説して、単語ではなく句を単位とする統計的機械翻訳手法について触れている。続いて、機械翻訳の評価手法について総説し、さらに、同義語抽出に関する過去の研究例についても記している。

第3章は、「A Chinese-Japanese Lexical Machine Translation through a Pivot Language (中間言語を用いた句に基づく中日機械翻訳)」と題し、英語を中間言語として中国語の専門用語を日本語に翻訳する手法について論じている。英日、中日それぞれの専門用語翻訳モデルから日中の専門用語翻訳モデルを生成し、さらに、中国語の漢字 (Hanzi) と日本語の漢字の対応づけに基づいて計算した翻訳確率を、特徴素として組み込む手法を提案している。評価では、数十万語規模の対訳辞書資源をコーパスとして用いて、BLUEスコアにより性能を比較し有効性を示すとともに、中間言語や日中漢字対応表の効果について詳細な分析を行っている。

第4章は、「Identifying Synonymous Translation Equivalentents in a Bilingual Lexicon (対訳辞書中の同義対訳表現の同定)」と題し、対訳辞書中の対訳用語ペアに基づく同義語と表記ゆれの抽出について論じている。具体的には、まず、対訳用語辞書から2言語間の対訳関係および単一言語内での同義語関係を抽出した後に、単語Nグラム類似度などの単語レベルの特徴素、文字バイグラムや正規化編集距離などの文字レベルの特徴素、さらに文脈類似度や共通対訳語の存在など多様な手掛かりを利用して、意味的な同一性を精度

高く判定する分類器を構成している。また、大規模な日英用語辞書を用いた評価により、提案手法が再現率およびF値において従来手法を大幅に改善することを示している。

第5章は、「Conclusion（結論）」であり、本論文の成果をまとめるとともに、課題や今後の展開の方向性について論じている。

以上を要するに、本論文では、機械翻訳システムのための対訳辞書の拡張手法を検討し、中間言語を介した専門用語対訳モデルの獲得および対訳用語辞書に基づく同義語獲得という2つの方向性を提示している。現実的なデータを用いた評価においても、提案手法は従来手法に対する優位性を示しており、日本語と中国語など対訳例文コーパスが必ずしも十分ではない言語間での機械翻訳システムの実現に貢献することが期待される。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。