

審査の結果の要旨

氏名 辻 慶太

本論文は、多言語シソーラスの自動構築手法を新語の処理に焦点を当てた新しい理論的枠組みに基づいて開発し、既存手法との比較評価を踏まえて提案したものである。本論文は全部で5章から構成されている。

第1章では、情報検索、特に言語横断検索や問合せ拡張検索における基本ツールとして、また言語学習のための辞書編纂等の基礎資料としての研究成果の意義を述べ、次いでシソーラス構築手法の観点から、言語対非依存の出現頻度に基づく抽出手法と言語対依存の翻字パターンに基づく抽出手法を組み合わせた統合的な訳語対抽出手法の独自性と有効性を主張している。さらに研究で用いた日英対訳コーパスと日英対訳用語リストの情報源としての特徴および抽出対象語に関する技術的な方針を説明している。第2章では、主として自然言語処理とのかかわりで試みられてきた同義語抽出手法および訳語対抽出手法の先行研究を詳細に検討している。従来の抽出手法が頻度に基づく統計的アプローチに偏っており、そのためコーパス中の出現頻度が本来的に低い語である新語を適切に扱えないという問題点があることを確認するとともに、その背後に潜む新語の重要性に対する認識の欠如を指摘している。その上で、新語の適切な処理には日英の言語対固有の性質を利用した手法の導入が有効であることを強調している。

第3章では対訳用語リスト EDICT を既存多言語シソーラスのひとつとみなし、抽出実験で用いたコーパス中の低頻度訳語対の多くが EDICT に収録されていない事実を明らかにすることで、新語の訳語対抽出をめぐる問題が低出現頻度語の問題と密接に関係していることを検証している。さらに低頻度訳語対の多くは従来の手法では抽出が難しいことを、「内部完全共起」と「外部完全共起」の概念を用いて実証し、また日英の訳語対には借用語（カタカナ外来語）系の訳語対が多いことに着目し、借用語系の訳語対を翻字パターンに基づいて自動抽出することの有効性を示している。以上のような基礎調査を踏まえて、①翻字による借用語系訳語対を抽出・除外し、②残った候補訳語対から出現頻度に基づいて訳語対を抽出し、③抽出した全訳語対からグラフ理論を用いてグラフを作成し、同じ意味の語をクラスター化する、という独自のシソーラス自動構築手法を提案している。この手法に基づいて、第4章では学会発表論文データベース（5分野）における抄録とタイトルの対訳コーパスを用いてシソーラス自動構築の実験を試みている。その結果、翻字と頻度に基づく二つの訳語対抽出手法を組み合わせることにより、訳語対抽出の精度と再現率が従来の手法に比べて顕著に高くなることを実証した。最後の第5章では研究の総括を行い、今後の課題をあげている。

以上のように、本論文は多言語シソーラスの自動構築の理論的側面および技術的側面の双方において新たな知見を提示し、図書館情報学および言語研究の領域に重要な貢献をもたらすものであると評価された。よって、本論文は博士（教育学）の学位を授与するにふさわしいものと判断された。