審 査 の 結 果 の 要 旨

氏　　名　ｱﾝﾄﾞﾗｰﾃﾞ ｼﾙﾌｧ ﾀﾞﾆｴﾙ ｹﾞｵﾙｸﾞ

本論文は、「Bayesian Statistical Methods for Extending Bilingual Lexicon Using Comparable Corpora(ベイズ統計によるComparable Corpora からの対訳ペアの獲得)」と題し、二言語間で対訳関係にある言語対を自動的に獲得する方法について論じている。

第1章では、専門的な用語に対する対訳辞書の必要性を述べ、話題は共通だが厳密な対訳関係にはない二言語コーパス（comparable corpora）から対訳辞書を自動的に拡張することを、研究の目標として設定している。このような二言語コーパスは、分野に依らず比較的簡単に入手できることから、対訳辞書の自動獲得手法が確立すれば、人手による辞書開発のコストが高いという従来の問題点を克服することができる。

第2章では、comparable corporaからの対訳関係の自動獲得についての従来研究を概観するとともに、単一言語コーパスからの同義語辞書の自動獲得などの関連分野に言及している。また、研究で用いるコーパスや評価法、比較手法などを説明し、論文中で扱う問題の定義を明確にしている。

第3章では、係り受け関係を含む文脈中での出現位置を手がかりとした、ピボット語の統計的抽出法の提案および評価を行っている。ここでピボット語とは、既存の対訳辞書中に存在する名詞などの内容語であり、文脈ごとして専門用語どうしの適切な対訳関係を発見するために用いられる。提案手法の第一のポイントは、語の共起の度合いを示す自己相互情報量（Point-wise Mutual Information、PWI）をベイズ法に基づく新しい手法によって推定し、重要な文脈語をより正確に発見する点である。第二のポイントは、それぞれの言語における単語どうしの係り受け関係の情報を言語間で対応づけることで、単語が現れる文脈をより的確に把握する点である。論文中では、様々な観点からの実験を通して提案手法の性能を分析し、既存手法に対する優位性を示している。

第4章では、係り受け関係に基づく異なる種類の文脈情報に対して、適切な重みづけを学習することで、対訳ペア獲得の精度を向上させる手法を提案している。この手法は、それぞれの言語に対する複数の文脈ベクトルを線形変換した後で、二言語間でコサイン類似度

を計算するものであり、従来手法におけるコサイン類似度の一般化になっている。提案手法では、ベイズ的な確率モデルによって線形変換の適切さを表現し、マルコフ連鎖モンテカルロ法を用いて変換行列の各パラメータを学習する。実験を通して、提案手法の適用により対訳獲得の精度が有意に向上し、最終的には、日本語-英語の言語対に対して既存の対訳獲得手法より最大10 ポイントの精度向上が得られることを示している。さらに、実験結果に対する考察から、日本語と英語の係り受け関係を対訳獲得に用いる際の、係り受け関係の重みづけについての知見を報告している。

第５章では、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、comparable corporaからの対訳関係の自動獲得について、ベイズ法に基づく確率推定を用いた主要な文脈語の抽出、係り受け関係に基づく一般化したコサイン尺度の導入と学習による重み調整という、２つのアプローチを示している。提案手法の枠組みは言語に依存せず、一般語の対訳辞書および各言語の構文解析器が存在すれば広く適用可能である。日本語-英語の組み合わせは、対訳獲得の対象言語対として難しい組み合わせの１つとされているが、提案手法はこの組合せに対して有効性を示していることから、その枠組みは他のさまざまな言語対に対しても有効であると考えられる。これらの研究成果は、comparable corpora を用いた対訳獲得に関する今後の研究において重要な基礎となるものである。

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。