

## 審査の結果の要旨

氏名 荒牧英治

本論文は、「**Formalization and Realization of Example-based Machine Translation**」（用例ベース翻訳の定式化と実現）と題し、実用レベルの用例ベース機械翻訳を実現するために、用例データベースの頑健な構築手法、および、用例の大きさ、文脈類似度、確信度などを総合的に考慮する訳語選択手法を提案し、さらに、従来経験則に頼っていた用例ベース翻訳の確率論的な定式化を行い、実験的評価によりその有効性を論じたものであり、6章から構成されている。

第1章は「**Introduction**」（緒言）であり、機械翻訳の歴史について概観し、近年、大規模対訳コーパスに基づく用例ベース、統計ベースの手法が実現可能になってきたこと、これらの手法の妥当性および差異について論じ、本論文で取り扱う問題を明確化している。

第2章は「**Robust Dictionary-based Phrase Alignment**」（辞書に基づく頑健な句対応付け）と題し、対訳文間の句レベルの対応付けを頑健に行う手法について論じている。この句対応付けの結果が用例データベースとなり、用例ベース翻訳の直接の知識源となることから、これを高精度に行うことが用例ベース翻訳にとってのキーテクノロジーとなる。この処理を頑健に実現するために、構文解析によって得られる構文情報と、翻訳辞書によって与えられる情報を最大限に利用する。ここで問題となるのは、辞書によって得られる単語・句対応の曖昧性と、辞書にのっていない未知語の扱いである。この問題を解決するために、前者については句対応内部の整合性と句対応相互の整合性を用いる曖昧性解消手法を提案し、後者については構文情報を利用した句対応の推定手法を提案した。

第3章は「**Basic EBMT System**」（基本用例ベース翻訳システム）と題し、用例ベース翻訳システムの全体の構成を示し、特に用例、すなわち訳語選択を高精度に行う方法について論じている。句対応付けの結果は誤りを含むものであるため、大量の用例の中から適切なものを選ぶ必要がある。ここでは、従来から用いられている用例の大きさ、文脈の類似度に加え、用例の句対応の確信度を尺度に加えることで、用例選択の精度が向上することを示した。

第4章は「**EBMT based on Translation Probability**」（翻訳確率に基づく用例ベース翻訳）と題し、用例ベース翻訳の定式化を行っている。用例ベース翻訳では、基本的にできるだけ大きな用例を用いて翻訳を行うことで、大きな文脈に支えられる高精度の翻訳を実現しようとしている。しかし、従来の用例ベース翻訳は経験則に基づく指標によって用例を選択しており、現在、一方で盛んに研究が行われている統計ベース翻訳に比べてアドホックで不透明であるとの批判を受けてきた。そこで、用例ベース翻訳の中に確率の尺度を持ち込み、大きな用例を用いることでその翻訳が安定する、すなわち翻訳確率が大きくなる、という考え方に基づく用例選択の定式化を提案した。また、この定式化に基づくシステムを実際に構築し、その有効性を議論した。

第5章は「**Experiments**」（実験）と題し、第3章の基本翻訳システムと第4章の確率ベースの翻訳システムについて、2万文からなる旅行対話対訳コーパスを用いた大規模な評価実験について報告している。評価にあたっては、複数の自動評価手法と人手による主観評価の両面からの評価を行い、多角的な議論を展開している。また、第3章のシステムについては、機械翻訳の評価型国際ワークショップに

において上位の成績を収めたことを報告し、これらのシステムが実用的な翻訳精度に達していることを示した。

第6章「Conclusion」（結論）では、本論文の主たる成果をまとめるとともに、今後の課題と展開について述べている。

以上を要するに、本論文は、用例ベース機械翻訳において実用的なシステムを実現するために特に重要となる用例データベース構築法、および高精度な用例選択法においていくつかの新しい知見を示すとともに、用例ベース翻訳の確率論的な定式化を提案し、大規模対訳コーパスを用いてそれらの有効性の確認を行ったものであり、電子情報学上貢献するところが少なくない。

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。