

(続紙 1)

京都大学	博士（情報学）	氏名	中澤 敏明
論文題目	Fully Syntactic Example-based Machine Translation （構造的言語処理に基づく用例ベース機械翻訳）		
(論文内容の要旨)			
本論文は、言語の構造情報を利用した用例ベース機械翻訳について論じたものであり、対訳文アライメント手法として単語の依存関係に注目した統計的な手法と、言語資源を可能な限り利用した少量の対訳コーパスでも有効な手法の二つを提案している。 また翻訳手法として木構造化された用例の中から入力文構造木の各部分木に最適な用例を、用例の大きさや周辺の単語の類似度などを基準として選択し、それらを結合することで翻訳を行う手法を提案しており、全6章から構成されている。			
第1章は序論であり、機械翻訳研究の歴史や現在主流となっている3つの機械翻訳手法についてまとめ、それらの位置付けについて論じている。また現在の機械翻訳の質について、英語とフランス語のように対象とする二言語の言語的性質が似ている場合は実用レベルの翻訳が行えているが、日本語と英語や日本語と中国語など性質が大きく異なる言語対ではいまだに精度が低いという問題点を指摘し、これを解決するために言語の構造情報を利用したアプローチを提案しており、またその妥当性についても論じている。			
第2章では、統計的句アライメントモデルを提案している。言語的に近い言語対では文を単純な単語列として扱うシンプルなモデルであっても非常に高精度にアライメントを行えるが、そうでない言語対では高い精度を実現するのは困難である。 提案モデルは両言語文ともに構文解析を行い、木構造を利用したモデルである点、既存のモデルで問題となるモデルの方向性を排除し、双方向のモデルである点、単語だけでなく複数単語からなる句をアライメントの単位としている点の3点が特徴的である。実験によって言語的に遠い言語対に対して既存の単語列を対象とした方向性のあるモデルに比べてより精度の高いアライメントが行えることを示している。			
第3章では、複数の言語資源を利用したアライメント手法を提案している。 第2章で提案した統計的な手法が十分な精度を達成するためには比較的大きな対訳コーパスが必要であるが、現実的には小さな対訳コーパスしか利用できない状況は多々ある。そのような場合でも高精度なアライメントを行い、有用な翻訳知識を獲得するための手法を提案している。得られた対応候補の中から正しいものを選択するための指標として、アライメント全体の整合性をスコア化したものを用いている。言語資源として一般的に利用可能な対訳辞書を用いた場合に、統計的なモデルに比べて高い適合率を達成した。			
第4章では、第2章や第3章で獲得された翻訳知識を用いた木構造上での翻訳手法を提案している。基本となる戦略は、入力文の翻訳に利用可能な様々な用例のうち、より大きな用例を優先的に使うというものである。入力文の翻訳に利用可能な各用例のスコアの基準として、用例の大			

きさ、用例の周囲の単語の類似度、目的言語側の翻訳確率を採用しており、最も適切な用例の組み合わせを、スコアの総和が最も高くなるように、ビームサーチにより求めている。

用例どうしの結合の際には用例の周囲のノードをのりしろとして利用することにより、自然な語順の翻訳文が生成されるという特徴がある。また大きな用例を優先的に利用するため、翻訳文が安定しやすいという特徴もある。人間が翻訳結果を評価した結果、翻訳ルールを手で整備したルールベースの翻訳システムよりも最高評価の文を出力する割合が高いことが確認された。また、構築したシステムはEコマースサイトの商品説明文翻訳や多言語SNSで利用されている。

第5章では、種々の言語処理システムの研究開発を効率化するために、システムの解析過程、解析結果、そこで利用される言語処理リソース等を統合的に可視化する汎用ツールを提案している。言語処理システムにとって必要な可視化は多くの場合、表形式とハイパーテキストの参照関係によって表現できることに着目し、各言語処理システムのログを解釈してリンクを埋め込んだHTMLのTable形式に変換・表示する可視化ツールを設計し、構文・談話解析システム、機械翻訳システムを例にその有効性を議論している。

第6章は結論であり、本論文を総括している。

(論文審査の結果の要旨)

本論文は、構造的言語処理に基づく用例ベース機械翻訳について、翻訳用例の獲得の際に言語の構造情報を利用する方法と、依存構造木を利用した用例の選択・結合方式に関する研究をまとめたもので、得られた主要な成果は以下のとおりである。

1. 両言語の文を構文解析し、木構造をベースとした対訳文アライメント手法を提案し、言語構造が大きく異なる言語であってもその違いを柔軟に吸収することにより高精度なアライメントが行えることを実験により示した。またその精度は既存の単語列のアライメント手法と比較しても優れたものであった。さらに少量のトレーニングコーパスしか利用できない場合でも、既存の言語資源を利用し、アライメント全体の整合性を考慮することにより、高精度にアライメントが行える手法を提案した。

2. 入力文の翻訳の際により大きな翻訳用例を優先的に利用することを基本とし、用例の周囲の類似度や翻訳先の表現ごとの確率を考慮しながら適切な用例を選択するという戦略により、安定した翻訳を出力する用例ベースの翻訳手法を提案した。用例どうしをつなぎあわせる際には、用例の外側に位置するノードを置き換えることにより実現し、出力文の語順の安定性を実現した。翻訳結果を人手により評価した結果、商用の翻訳システムや統計的な翻訳システムよりも完全な翻訳を出力する割合が高いことが明らかとなった。評価の平均値でも、人手によりチューニングされている商用のシステムと同等であった。

3. 構築した用例ベース機械翻訳システムをEコマースサイトの商品説明文の翻訳や、京都大学情報学研究科で運営されている留学生の学生生活を支援するためのコミュニティーサイトなどで実際に利用し、社会的な貢献を果たした。

よって、本論文は博士（情報学）の学位論文として価値あるものと認める。

また、平成22年2月24日実施した論文内容とそれに関連した試問の結果合格と認めた。