

論文内容の要旨

博士論文題目 サブ言語に特有の文構造を用いたフォーマル文書の翻訳精度向上

氏名 富士秀

(論文内容の要旨)

日英間や日中間のような、語順の大きく異なる言語間における機械翻訳の精度は、利用可能なコーパスデータ量の増加に加え、構文解析にもとづく単語並べ替え技術の進展によって大きく向上した。しかし、法令文、特許文、技術マニュアルなどのフォーマルな文書は、内容が複雑でかつ長文で記述される場合が多いため、他の対象文書と比べても依然として質の高い翻訳が困難である。

これらフォーマルな文書は、分野や文種によって特有の表現形態を持っており、このためそれぞれが独自の「サブ言語」を構成しているとみなすことができる。これら文書は内容が複雑でかつ長文である反面、読者の理解を助けるために、サブ言語毎に特有の表現が発達し、執筆者は慣習的にこの表現を用いて文の構造をわかりやすく記述する場合が多い。このことから、サブ言語に特有の表現に着目した処理を機械翻訳に導入することで、文の構造を正確に把握して、翻訳精度を向上させることが考えられる。特に、語順の異なる言語間では、長文において、構文的な順番が異なるだけではなく文の構造も異なるため、正確な文構造の把握が重要となる。

本研究では、サブ言語を構成するようなフォーマルな文書の翻訳において、サブ言語に特有の表現を用いて入力文の構造を把握し、その構造を単位とした処理を導入することで高精度な翻訳を実現する。

第3章では、様々なサブ言語の中でも特に1文が特異的に長く、表現の規則性が高いという特徴を持つ特許請求項文のサブ言語を対象とした翻訳実験について述べる。特許請求項では、表現の規則性が極めて高いため、文の構造を認識するための規則を手で記述し、これをもとに統計的機械翻訳を動作させた。その結果、英日・日英・中日・日中という語順の大きく異なる言語間の翻訳において、極めて大きな翻訳精度の向上が認められた。

第4章では、表現の規則性はそれほど高くないが適用範囲の幅広い特有表現を持つ文が存在するサブ言語を対象とした実験を行った。ここでは、特有表現にある程度の揺れがあるため、人手による規則作成ではなく、大量のデータからの統計学習による方法を用いた。語順の異なる言語対を対象とし、原言語と目的言語間の構造部品の自動的な並べ替えを実現した。この構造部品の自動的な並べ替えを「大域的な並べ替え」と呼んでいる。特に、従来の構文解析による並べ替えの精度が低い日英翻訳において、大域的な並べ替えと従来の並べ替えの併用によって大きな翻訳精度の改善がみられた。

以上のようにして本研究では、特有表現の規則性の高さが異なるサブ言語を対象として、サブ言語に特有の文構造を導入した手法の有効性を確認することができた。

氏名	富士秀
----	-----

(論文審査結果の要旨)

平成29年1月24日に開催した公聴会の結果を参考に平成29年2月15日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

富士秀は、本博士論文において、特許文等の専門性が高いが定型性も高い文書の機械翻訳について主に2つの手法を提案し、その性能評価実験を行った。本論文の貢献は、以下のようにまとめることができる。

1. 専門性の高い文書は複雑な文が散見される反面、定型性が高くサブ言語を構成していると言われている。本論文では、サブ言語を構成する文書の典型例として、特許関連文書を対象にし、その性質とサブ言語の性質を考慮した機械翻訳法について詳細な検討を行った。
2. サブ言語の中でも特に長文が多く、文構造の定型性が高い特許請求項文を対象とした翻訳実験を行った。特許請求項の文構造を認識するための同期文脈自由文法の規則を人手で記述し、これをもとに統計的機械翻訳を実装した。その結果、英日・日英・中日・日中という語順の大きく異なる言語間の翻訳において、極めて大きな翻訳精度の向上を達成できることを示した。
3. 定型性が特許請求項ほど高くない一般的な特許文を対象に、より広いサブ言語を対象とした機械翻訳手法を提案した。人手により同期文脈自由文法の規則を記述するのではなく、大量の対訳データからの統計学習により、原言語と目的言語間の構造間の大局的な並べ替えを自動学習する方法を提案した。従来の構文解析による並べ替えの精度が低い日英翻訳において、大域的な並べ替えと従来の並べ替えの併用によって大きな翻訳精度の改善が得られることを示した。

専門性の高い特許関連文書の構造を規則および自動分割による事前並び替え手法を提案した本研究は、実用性と適応性に優れており、機械翻訳の分野において高い貢献があると評価する。

よって、本論文は、博士(工学)の学位論文として価値あるものと認める。