

平成27年度 情報工学コース卒業研究報告要旨

外山 研究室	氏 名	平 田 雄 大
卒業研究題目	日英統計的機械翻訳における語順入れ替え範囲に関する研究	
<p>近年の国際社会のグローバル化に伴い、日本法の国際的発信の重要性が増している。これに対応するため、法務省は2009年に日本法令外国語訳データベースシステム(JLT)の運用を開始した。JLTでは人手で英訳された法令を提供している。しかし、人手による翻訳はコストがかかるため、公布された法令の英訳がJLTで公開されるまでには長い時間を要する。この問題を解決するための手法の一つとして、法令文を機械翻訳によって英訳することが考えられる。</p> <p>このような背景を基に、法令文を対象とした統計的機械翻訳(SMT)システムの研究が進められている。統計的機械翻訳では原言語の単語を翻訳し、語順を入れ替えて目的言語文を生成する。SMTシステムとして広く使われているMosesでは、語順入れ替え範囲を6と設定しているが、これは英仏翻訳などの語順が似ている言語間の翻訳のための設定である。それに対して、日英翻訳のような語順が大きく異なる言語間の翻訳では、入れ替え範囲が6では不十分であり、入れ替え範囲を無制限として翻訳することが考えられる。しかし、岡田らは語順入れ替え範囲を変化させる実験を行い、無制限ではなく、入れ替え範囲を24とした時が翻訳精度が最も良いことを示し、それにより翻訳精度を向上させた。しかし、この研究では全ての文に同じ入れ替え範囲を設定している。本研究では、文長に応じて語順入れ替え範囲を変化させることによって翻訳精度がさらに改善されるという仮説を立てた。</p> <p>この仮説を検証するために、語順入れ替え範囲を固定した翻訳結果と、文長に応じて変化させた翻訳結果とを比較する実験を行った。その結果、前者のRIBES値は66.93、後者のRIBES値が66.73となり、翻訳精度の優劣は認められなかった。</p> <p>この原因を調べるために、文長ごとに翻訳精度が最大となる語順入れ替え範囲を調べた。まず、文長が20語以下の場合、文長と同じ入れ替え範囲となった時、翻訳精度が最大となった。ところが、文長が20語を超えた場合は語順入れ替え範囲が20から30の間に翻訳精度が最大となる入れ替え範囲があった。この間では翻訳精度にほとんど差が見られなかった。さらに入れ替え範囲が30語を超えると翻訳精度は下がった。このような結果になった理由として、文長が20語を超える文では、ほとんどが複文であったことがいえる。複文における語順入れ替えは、文を構成するそれぞれの範囲内で行われるべきであるため、複文に対して広い入れ替え範囲を設定しても、翻訳精度の向上に寄与することはなく、むしろ複文では主節の主語が従属節の主語になるというような誤った語順入れ替えを引き起こす。</p> <p>以上の結果から、文長に応じた語順入れ替え範囲の変更は不要であることがわかった。そして、適切な入れ替え範囲は20から30程度で、翻訳速度を考慮するならば20程度がよいと考えられる。</p>		