

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	東 内 悠 亮
卒業研究題目	英文官報からの対訳法令用語抽出	
<p>今日、社会のグローバル化が進み、国際取引の円滑化、対日投資の促進、後進国の法整備支援の推進、在日外国人の生活上の利便向上などを図るため、日本法令を翻訳する需要が高まっている。これに応えるため、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を開設し、法令用語の英日辞書である標準対訳辞書や、人手で英訳した日本法令を公開している。標準対訳辞書は、それまで個別に翻訳されてきた法令から法令用語の主要な用語や言い回しなどの日英対訳を自動抽出し、その対訳を法律専門家が追加・修正したもので構成されている。現在、この辞書に登録されている用語は4,786語であるが、この分量は十分ではなく、辞書に登録されていない用語を翻訳する場合、翻訳者によって異なった翻訳をする可能性があり、翻訳の統一性が損なわれる。これを解決するために、標準対訳辞書を拡充することが必要になる。</p> <p>そこで本研究では、英文官報に着目する。英文官報とは、戦後占領期に発行された官報を、GHQに提出するために英訳したものである。英文官報は、1946年4月4日から、日本がサンフランシスコ平和条約を締結して主権を回復する1952年4月28日までの約6年1ヶ月の間発行された。刊行当初は、形式、訳語の不統一といった問題が起こっていたため、政府は1949年8月の英文官報原稿翻訳に関する懇談会において、翻訳の細かな指示を出して、形式、訳語の統一を図った。このように、英文官報は質の高い翻訳がされていたことがわかる。しかし、標準対訳辞書を作成した時に、英文官報は考慮されていなかった。そこで、英文官報とそれに対応した邦文官報で構成される英文官報対訳コーパスから、新たな対訳法令用語を抽出することにより、標準対訳辞書を拡充することが期待できる。</p> <p>本研究では、標準対訳辞書の拡充を目指し、英文官報対訳コーパスから対訳法令用語を抽出する。具体的には、このコーパスに対してGIZA++を用いて、日本語単語と英単語の対応付けを行い、対応単語間の翻訳確率を求める。GIZA++は、2言語間で1単語対1単語の対応付けを行い、それぞれの共起回数をもとにEMアルゴリズムで翻訳確率を計算する。本研究では、日本語単語に対して、翻訳確率が最大となる英単語を対訳用語とし、それらに対訳ペア候補として抽出する。こうした対訳ペア候補のうち、出現回数と翻訳確率が閾値より大きなものを最終的な対訳ペアとした。なお、今回は抽出する対訳法令用語を名詞に限定した。</p> <p>法令1,624本155,051文からなる英文官報対訳コーパスに対して、翻訳確率の閾値を0.2、出現回数の閾値を30として本手法を適用したところ、1,555組の対訳法令用語を抽出できた。これらを評価したところ、正解率が63%(部分正解を含めると96%)であった。また、正しく抽出された対訳法令用語の内、標準対訳辞書に登録されていないものは624組あった。そのうち、日本語の見出し語は一緒であるが対訳は異なるものは198組あった。また、汎用的な辞書である英辞郎第8版と比較したところ、登録されていないものは94組であった。そのうち、日本語の見出し語は一緒であるが対訳は異なるものは6組あった。以上より、今回の実験で標準対訳辞書の拡充ができたといえる。</p>		