

平成24年度 情報工学コース卒業研究報告要旨

外山 研究室	氏 名	村 神 匠
卒業研究題目	浅い構文解析に基づく法令ターミノロジーの構築	
<p>近年、グローバル化に伴い法令翻訳の需要が増大している。2009年4月には法令外国語訳データベースシステム（Japanese Law Translation; JLT）が法務省によって公開され、日本語法令の英訳が進行している。法令翻訳の際に利用される言語資源として、JLTが編纂している法令用語日英標準対訳辞書がある。しかし、この対訳辞書に登録されている語は、法学の専門用語の一部のみである。そのため登録語数は現在4,482語であり、法令翻訳に必要な語数には到底足りていない。対訳辞書に登録されていない用語の翻訳は訳者に依存するため、統一性のない翻訳の原因となっている。また、この対訳辞書には法令名は含まれていない。しかし、翻訳のためには法令名を固有名詞として登録することが望ましい。このように、JLTの対訳辞書には翻訳に必要な用語が十分に登録されていない。</p> <p>また、日本と世界各国の法令の国際的な比較研究のために、多言語法令ターミノロジーの構築が求められている。多言語法令ターミノロジーは、言語による用語や概念の違いを記述するものであるが、それには、まず日本語の法令ターミノロジーの構築が必要となる。</p> <p>ターミノロジーという言葉にはいくつかの定義が存在するが、本研究ではある特定の分野の専門用語の集合体と捉える。そこで、対象分野を法令に限定し、法令文に出現する用語すべてを専門用語と考え、法令ターミノロジーを構築する。これらの用語の訳語を決定することができれば、法令翻訳に使用可能な語数を増大させることが可能である。</p> <p>よって本研究では、日本語の法令ターミノロジーを構築する。また、法令文に出現する用語の特徴を明らかにし、その特徴がターミノロジーの構築に有用であるかを考察する。</p> <p>法令ターミノロジー構築に関しては、これまでに小林らによって構文パターンに基づく手法が提案されている。この手法は正規表現を用いた文字列検索によって、法令中で定義された用語を取得する。しかし、この手法では法令2,741本中から3,808個の定義語しか得られていない。すべての用語を網羅的に取得するには、形態素解析および構文解析が必要になる。しかし、法令文の構文情報は複雑であり、通常の構文解析で高い精度を得ることは容易ではない。そこで、本研究では新たな手法として、浅い構文解析、すなわちチャンキングに基づく手法を提案する。浅い構文解析を用いることにより、比較的良い精度で解析できる。これにより、従来手法より多くの用語が獲得できる。また、従来手法とは異なり、本手法はルールではなく、統計に基づいたものである。</p> <p>本手法ではまず、法令文に対して形態素解析を行った。その際、従来の形態素解析システムでは法令文特有の旧仮名遣いに対処できないため、前処理を導入した。次に、チャンキングによって形態素解析結果に対してタグを付けた。タグにはStart/End手法を発展させた9種類を使用した。チャンキングの学習には、憲法を除く六法とJLTの主要法令を合わせた計20本の法令に人手でタグを付けたデータを使用した。その後、チャンキング結果から名詞句と動詞句を判定して抽出し、各句の出現回数を求めた。なお、形態素解析にはChaSenを、チャンキングにはYamChaをそれぞれ利用した。</p> <p>以上の手法を法令221本（120,449文）に対して適用した。その結果、38,935種類の名詞句と8,847種類の動詞句を抽出した。そのうち名詞句の36.5%、動詞句の34.2%は1回しか出現していなかった。また、条項などの番号を示す句が名詞句に3,778種類含まれていることも明らかになった。</p>		