

平成 20 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 松 秀 紀
卒業研究題目	対訳表現に対する統一性評価指標の性能向上	

従来の法令翻訳は、法令を所轄する府省や、法令出版社などによる個別の活動として行われてきた。そのため、同じ原文に対しても様々な訳し方が存在し、そうした訳し方の違いにより利用者に誤解を生じさせてしまう場合がある。そのため、このような翻訳においては対訳表現は適切なものに統一されていなければならない。そこで日本政府は、法令用語や頻出する言い回しに関して、標準対訳辞書を構築した。標準対訳辞書の利用により、単語レベルでの対訳表現の統一は一応解決した。しかし、同じ法令用語であっても文脈によって対訳表現を使い分けなければならない場合もある。全ての文脈に対して、どの対訳表現を用いるかを標準対訳辞書に記述するのは困難である。したがって、対訳表現の統一性を評価するのに、標準対訳辞書だけでは不十分である。これに対して、Ogawa et al.[1] は対訳表現に対する統一性評価指標 CIEL を提案している。

CIEL は BLEU[2] を拡張した手法であり、原文と類似した法令文の集合を対訳コーパスから獲得し、その訳文(以下、擬似参照訳と呼ぶ)と原文の翻訳を比較して、対訳表現の統一性を評価する。対訳コーパス中の日本語法令文集合を、編集距離を用いたクラスタリングにより類似文ごとのクラスタに分割し、原文との距離が最も小さいクラスタの訳文を擬似参照訳とする。Ogawa et al. は、労働基準法の政府翻訳、法令出版社翻訳、Google による機械翻訳の優劣を CIEL によって判別できる事を示した。

CIEL における問題点として、Ogawa et al. は擬似参照訳として獲得した法令文集合が原文と類似していない場合、その擬似参照訳を用いた評価結果は信頼できないことをあげている。Ogawa et al. は労働基準法のみを CIEL の評価しているが、それ以外の法令に対しても評価し、CIEL の性能を確認する必要がある。

本研究では、この 2 つの問題点の解決に関して以下のように取り組んだ。

前者の問題点に対しては、対訳コーパスを増大させることにより、多くの原文に対してより類似した法令文が獲得できると考えた。そこで、異なる大きさの対訳コーパスを複数用意し、それらを用いて労働基準法 242 文の評価実験を行った。その結果、対訳コーパス中の日本語異なり文が 4,465 文であった場合と 34,873 文である場合を比較すると、原文擬似参照訳間の距離の平均は 10%短縮する事を確認した。

後者の問題点に対しては、労働基準法以外の法令として、民法第 1 編～第 3 編 912 文、特許法 271 文をそれぞれ評価した。評価実験の結果、どちらの法令もその政府翻訳、法令出版社翻訳、Google による機械翻訳の優劣を CIEL によって判別できることを確認した。

参考文献

- [1]Yasuhiro Ogawa, Kazuhiro Imai, Katsuhiko Toyama: Evaluation Metrics for Consistent Translation of Japanese Legal Sentences, Proc. LREC 2008 Workshop on Semantic Processing of Legal Texts, pp.42-50, Marrakech, Morocco (2008).
- [2]Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu: “BLEU: a Method for Automatic Evaluation of Machine Translation”, In Proc. ACL’02, pp.311-318, 2002.