新旧対照表の利用による法令の英訳修正

小酒井 款雄[1]　小川 泰弘[1,2]　大野 誠寛[1,2]　中村 誠[3]　外山 勝彦[1,2]
[1]名古屋大学 大学院情報科学研究科　[2]同 情報基盤センター　[3]同 大学院法学研究科
{kozakai, yasuhiro}@kl.i.is.nagoya-u.ac.jp

1　はじめに

社会・経済の国際化が進む今日において、国際取引の円滑化、対日投資の促進、法整備支援の推進などを図るために、わが国の法令を国際的に発信することが求められている。そのためには、法令を外国語に翻訳し、それを迅速に発信する必要がある。それを実現するため、2009 年 4 月に日本法令外国語訳データベースシステム (JLT) [1] が開設された [1]。

しかし、JLT にはいくつかの問題点がある。その一つが、法令の改正に英訳が追従できていないことである。JLT には 2016 年 9 月の時点で 397 本の法令とその英訳が収録されている。そのうち、現行法令の英訳は 116 本に過ぎず、残りの 281 本、すなわち、約 7 割の英訳が最新の改正に追従できていない。特に、商業登記法 (昭和 38 年法律第 125 号) は最新の改正が平成 26 年に行われているにもかかわらず、JLT には平成 17 年における改正を反映したものしか収録されていない。改正前の法令の英訳が JLT で公開されていると、利用者にそれ以降の改正が行われていないという誤解を与える原因になる。よって、JLT において法令の英訳を改正に追従させることは急務である。本稿では、この問題を解決するための機械翻訳手法を提案する。

改正された法令の翻訳を行うに当たり、本稿では改正前の法令 (以下、旧法令と呼ぶ。) の英訳の一部を修正することにより、改正後の法令 (以下、新法令と呼ぶ。) の英訳を作成する。これにより、通常の機械翻訳よりも精度の高い翻訳が可能となり、訳語の統一という効果も期待できる。

本稿では、この手法を実現するために Koehn らが提案した手法 [2](以下、Koehn の手法と呼ぶ。) を利用した。この手法では、翻訳メモリから抽出した入力文に類似する文の翻訳について、その一部を修正することにより入力文を翻訳する。

Koehn の手法では、入力文とその類似文の間の修正部分を編集距離の計算過程から求める。しかし、これは表層の文字列に基づくものであり、この単位が翻訳に適しているとは限らない。そこで、法令が改正される際に作成される新旧対照表 (図 1) に着目した。新旧対照表は人手により作成されたものであり、修正部分は意味的にまとまったものとなっている [3]。そこで、本稿では、修正部分を編集距離を用いて求める代わりに、新旧対照表を用いて求める手法を提案する。

[1]http://www.japaneselawtranslation.go.jp/

新法令	旧法令
（削る）	四　父に支給される公的年金給付の額の加算の対象となつているとき。
四　児童福祉法（昭和二十二年法律第百六十四号）第六条の三第一項に規定する里親に委託されているとき。	五　児童福祉法（昭和二十二年法律第百六十四号）第六条の三第一項に規定する里親に委託されているとき。
五　父に支給される公的年金給付の額の加算の対象となつているとき。	（新設）
六　父と生計を同じくしているとき。ただし、その者が前項第一号ハに規定する政令で定める程度の障害の状態にあるときを除く。	六　父と生計を同じくしているとき。ただし、その者が前項第三号に規定する政令で定める程度の障害の状態にあるときを除く。

図 1: 新旧対照表の例 (児童扶養手当法（昭和 36 年法律第 238 号) 第四条第二項の一部)

翻訳実験の結果、提案手法は Koehn の手法に比べて、BLEU[4] と RIBES[5] のスコアがともに向上した。

2　法令の改正と新旧対照表

法令は様々な時代背景の変化に対応するために改正されることがある。改正は、その規模により全部改正と一部改正に区別される。全部改正は法令の内容を全面的に改正する方式であり、一部改正は法令の一部を改正する方式である。一部改正の際には、法令改正の審議のための参考資料として、図 1 に示すような新旧対照表が作成される。

新旧対照表は、法令改正の詳細を視覚的に示したものであり、新法令中の条項と旧法令中の条項が対応付けられている。旧法令から新法令への修正部分は法令文に付された傍線により示される。図 1 の例について第六号 (「六」で始まる条項) に付された傍線部分に注目すると、旧法令中の「前項第三号」が新法令では「第一号ハ」に修正されていることがわかる。また、条を新たに挿入する場合や既存の条を削除する場合は、対応する条項がないので、それぞれ「(新設)」、「(削る)」のように記述される。

3　Koehn の手法

Koehn の手法 [2] は翻訳メモリ技術と統計的機械翻訳 (SMT) を統合した手法である。手法の概要を図 2 に

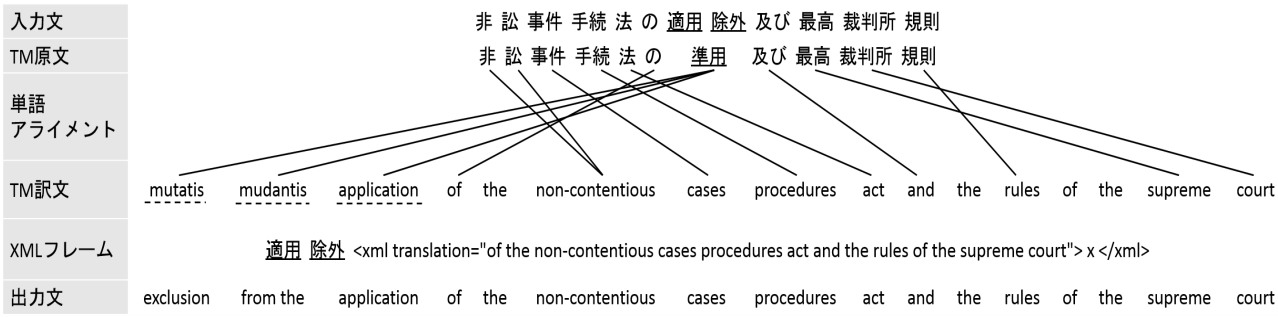

図 2: Koehn の手法の手順

示す。この手法では、まず、入力文に類似する原言語文と目標言語文のペア (以下、このペアにおける原言語文を TM 原文、目標言語文を TM 訳文と呼ぶ。) を翻訳メモリから抽出する。次に、TM 原文から入力文への修正部分を編集距離の計算過程から決定する (図 2 の下線部)。その後、TM 訳文における修正部分を単語アライメントにより決定する (図 2 の破線部)。最後に、修正部分を削除した TM 訳文と、入力文における修正部分を SMT により翻訳したものとを組み合わせることにより、入力文の翻訳を生成する。

Koehn の手法では Moses[6] を利用することを前提とする。Moses には、あらかじめ入力文の一部を XML タグで囲むことにより、翻訳を固定する機能がある。具体的には図 2 の 5 行目に示すような XML フレームを作成し、Moses への入力とする。これにより、入力文と TM 原文で異なる部分のみを翻訳するという手法を実現している。

4 提案手法

本稿における提案手法は、Koehn の手法に以下の二つの変更を加えたものである。

1. TM 原文・TM 訳文の代わりに、それぞれ旧原文・旧訳文を用いる。
2. 修正部分を編集距離の計算過程からではなく、新旧対照表から求める。

Koehn の手法では、入力文の類似文を翻訳メモリから抽出していたが、翻訳メモリに類似文が存在しない場合も考えられる。しかし、提案手法における入力は新原文であり、その類似文として旧原文が必ず存在する。そのため、提案手法では TM 原文・TM 訳文の代わりに、それぞれ旧原文・旧訳文を用いる。

また、Koehn の手法においては、旧原文から新原文への修正部分を編集距離の計算過程から決定していたが、提案手法では、修正部分を決定するために、2 節で示した新旧対照表を用いる。新旧対照表は人手により作成されたものであり、修正部分は意味的にまとまったものとなっている [3]。

例えば、図 1 では、第六号の修正部分は傍線部分のようになっている。これは、「前項第三号」を「第一号ハ」に置換するというものである。これを編集距離の計算過程から求めると、その修正は旧法令における「前項」を削除し、「三」を「一」に置換し、「ハ」を挿入するというものになる。

新旧対照表により決定した修正部分は、編集距離の計算過程から決定したものよりも翻訳の単位としては適切であると期待できる。なお、提案手法で利用する新旧対照表については、新法令と旧法令それぞれが別々に XML 文書化されており、傍線部分は、あらかじめそれを示す XML タグにより囲まれている。よって、提案手法では、この情報を用いて修正部分を決定する。

5 実験

提案手法の有効性を検証する翻訳実験を行った。この実験では、通常の SMT、旧訳文をそのまま新訳文とした場合、Koehn の手法の三つと提案手法を比較した。なお、Koehn の手法においては、TM 原文・TM 訳文として、それぞれ旧原文・旧訳文を利用した。

また、この実験では図 1 の例にあるような条項全体を挿入、削除する場合は翻訳対象としなかった。条項全体を新たに挿入するという場合は、修正部分が文全体に及ぶため、通常の SMT と同じ問題となる。そのため、この実験では図 1 の第六号のように、文の一部を修正するという場合のみを翻訳対象とした。

5.1 実験設定

この翻訳実験では、JLT に収録されている法令文対訳コーパスから、テストデータと学習データを用意した。

JLT においては、法令の改正に英訳が追従した場合、改正前の英訳も残されている。ただし、JLT における英訳の修正は、改正一回ずつをその都度反映したものだけではなく、複数回の改正を一度に反映したものもある。後者については、対応する新旧対照表が作成できない場合がある。よって、この実験では、一回の改正にのみ対応して英訳を修正したものを抽出して

表 1: 翻訳実験の結果

	BLEU	RIBES
phrase-base	40.93	64.67
TM	54.11	**86.42**
Koehn の手法	60.36	82.56
提案手法	**60.73**	83.57

テストデータに用いた。今回の実験では 17 法令がこれに該当し、改正に伴う修正の対象となった 158 文をテストデータとした。また、この改正に対する新旧対照表は人手で作成して使用した。

学習データは、この 17 法令を除いた残りの法令ということになる。しかし、翻訳の対象は新法令であり、旧法令は事前に英訳されていることが前提となる。よって、この 17 法令の旧法令に関しては学習データに加えた。結果として、学習データには 407 法令 158,928 文を使用し、翻訳モデルと言語モデルを構築した。

SMT においては、単語分割には MeCab[7] を、単語アライメントには GIZA++[8] を、言語モデルの作成には SRILM[9] を、デコーダには Moses を利用した。また、翻訳結果は BLEU と RIBES により評価した。

5.2 結果

実験結果を表 1 に示す。ここで、phrase-base は通常の SMT を用いた場合のスコアを示し、TM は旧訳文をそのまま新訳文とした場合のスコアを示している。この表から、Koehn の手法と提案手法は phrase-base に比べて BLEU、RIBES ともにスコアが高く、また、TM に比べて BLEU のスコアが高くなっていることがわかる。RIBES については、TM が最もスコアが高かった。また、Koehn の手法と提案手法を比較すると、BLEU、RIBES ともに提案手法の方がスコアが高かった。

5.3 考察

通常の SMT との比較

Koehn の手法および提案手法を phrase-base と比較すると、Koehn の手法と提案手法の方が BLEU、RIBES ともにスコアが高かった。これについては、旧訳文を利用することにより正解の一部をそのまま翻訳に利用できたためであると考えられる。

TM との比較

Koehn の手法および提案手法を TM と比較すると、Koehn の手法と提案手法のほうが BLEU のスコアが高かった。これは、SMT により作成した翻訳を利用して、TM における正解と異なる部分を置換することにより、より正解に近い翻訳を作成できたためであると考えられる。

また、RIBES については TM のスコアのほうが高かった。これは RIBES の計算方法に起因する結果であると考えられる。RIBES ではシステムの出力と参照訳との間の語順の相関からスコアが計算される。ここで、相関の計算対象は、システムの出力と参照訳とで一致する単語のみである。TM の結果は旧訳文をそのままシステムの出力としたものであり、その語順は正しい。そのため、TM の結果と参照訳とで一致する単語の語順の相関を計算するとスコアが高くなる。これが、TM のスコアが高くなった要因だと考えられる。

Koehn の手法と提案手法の比較

Koehn の手法と提案手法を比較すると、BLEU、RIBES ともに提案手法のほうがスコアが高かった。これは 4 節で述べたように新旧対照表による修正部分は意味的にまとまったものとなっており、さらにその翻訳もまとまったものとなることがその要因だと考えられる。

提案手法による翻訳の方が良かった例を図 3 に示す。この図では、Koehn の手法における修正部分を波線、提案手法における修正部分を実線で示している。この例では、「高等海難審判庁」が「海難審判所長」に置換されている。これは新旧対照表から求めた修正部分であるが、修正部分を編集距離から求めると、「高等」を削除し、「庁」を「所長」に置換することになる。

図 3 について提案手法の翻訳と Koehn の翻訳を見比べると、提案手法の翻訳には「長」に対応する “the president of” が含まれているのに対し、Koehn の手法の翻訳にはそれが存在しないことがわかる。この「長」に対応する翻訳がないという失敗は、Koehn の手法における問題に起因するものである。Koehn の手法において「長」に対応する “the president of” を加えるならば、それは “the japan” の前に挿入されるのが正しい。しかし、Koehn の手法において作成する XML フレームでは、XML タグの属性により指定された翻訳は一つのフレーズとして扱われる。そのため、そのフレーズ内の単語は並び替えることができず、単語を挿入することもできない。これが Koehn の手法の翻訳に「長」に対応する翻訳が出現しなかった原因である。その点、提案手法の翻訳では翻訳の単位が「海難審判所長」まで拡張されており、このような問題は発生しない。

一方、提案手法による翻訳の方が明らかに悪かったという例は見つからなかった。しかし、Koehn の手法、提案手法のどちらにおいても不十分であった例として、「ものの外」を「もののほか」に修正する、すなわち、日本語の表記を修正するという改正が存在した。この場合、英訳を修正する必要はないが、提案手法においては原文の修正に合わせて英訳も修正してしまう。そのために余分な修正を行い、かえって結果が悪くなってしまった。

入力	新原文	理事官又は受審人は、国土交通省令の定めるところにより、海難審判所長に管轄の移転を請求することができる。
	旧原文	理事官又は受審人は、国土交通省令の定めるところにより、高等海難審判庁に管轄の移転を請求することができる。
	旧訳文	an investigator or an examinee may request the japan marine accident inquiry agency (the second tribunal , tokyo) to make the disposition of the change of jurisdiction of a case , as provided for in the ordinance of the ministry of land , infrastructure , transport and tourism .
参照訳	新訳文	an investigator or an examinee may request the commissioner of japan marine accident tribunal to make the disposition of the change of jurisdiction of a case , as provided for in the ordinance of the ministry of land , infrastructure , transport and tourism .
Koehn の手法	新訳文	an investigator or an examinee may request the japan marine accident inquiry office (the second tribunal , tokyo) to make the disposition of the change of jurisdiction of a case , as provided for in the ordinance of the ministry of land , infrastructure , transport and tourism .
提案手法	新訳文	an investigator or an examinee may request **the president of** the japan marine accident inquiry office (the second tribunal , tokyo) to make the disposition of the change of jurisdiction of a case , as provided for in the ordinance of the ministry of land , infrastructure , transport and tourism .

図 3: Koehn の手法に比べて提案手法による翻訳の方が良かった例

6 おわりに

本稿では、JLT の英訳が法令の改正に追従できていないという問題点を解決するための方法として、Koehn の手法において旧法令の対訳と新旧対照表を用いる手法を提案した。それにより、新法令の翻訳について通常の統計的機械翻訳や Koehn の手法よりも精度の高い翻訳を実現した。

今後の課題としては、テストデータの拡充と 5.3 節で示した日本語の表記の修正のみの改正に対応することなどが挙げられる。

参考文献

[1] 外山，齋藤，関根，小川，角田，木村，松浦: 日本法令外国語訳データベースシステムの設計と開発, 情報ネットワーク・ローレビュー, Vol.11, pp. 33-53, 2012.

[2] Koehn, P., Senellart, J.: *Convergence of Translation Memory and Statistical Machine Translation*, In Proc. of AMTA Workshop on MT Research and the Translation Industry, pp. 21-31, 2010.

[3] 前田: ワークブック法制執務. 株式会社ぎょうせい, 2005.

[4] Papineni, K., Roukos, S., Ward, T., Zhu, W.: *BLEU: a Method for Automatic Evaluation of Machine Translation*. In Proc. of ACL 2002, pp. 138-145, 2002.

[5] 平尾, 磯崎, 須藤, Duh, K., 塚田, 永田: 語順の相関に基づく機械翻訳の自動評価法. 自然言語処理, 21(3), pp. 421-444, 2014.

[6] Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constanin, A., and Herbst, E.: *Moses: Open Source Toolkit for Statistical Machine Translation*. In Proc. of ACL 2007, pp. 177–180, 2007.

[7] Kudo, T., Yamamoto, K., and Matsumoto, Y.: *Applying Conditional Random Fields to Japanese Morphological Analysis*. In Proc. of EMNLP 2004, pp. 230-237, 2004.

[8] Och, F.J., Ney, H.: *A Systematic Comparison of Various Statistical Alignment Models*. Computational Linguistics, 29(1), pp.19-51, 2005.

[9] Stolcke, A.: *SRILM - An Extensible Language Modeling Toolkit*. In Proc. of ICSLP 2002, pp. 901-904, 2002.