

令和 6 年 5 月 30 日現在

機関番号：13901

研究種目：基盤研究(B)（一般）

研究期間：2021～2023

課題番号：21H03772

研究課題名（和文）日本法情報の国際的発信：ハイブリッド機械翻訳と法令LODによる法令改正への対応

研究課題名（英文）International Dissemination of Japanese Statutes Information: Responding to Amendments using Hybrid Machine Translation and Legal LOD

研究代表者

外山 勝彦 (TOYAMA, Katsuhiko)

名古屋大学・情報学研究科・教授

研究者番号：70217561

交付決定額（研究期間全体）：（直接経費） 10,800,000円

研究成果の概要（和文）：本研究は、社会のグローバル化に伴い必要な日本法に関する情報を迅速かつ国際的に発信するために、法令改正に伴う英訳法令の修正を支援する機械翻訳技術の開発を目的とする。その際、旧条文の改正部分に対する英訳だけを修正し、非改正部分に対する英訳は可能な限り修正しないで旧訳を用いるという訳文の変更極小性を要求する。そのために、新旧対照・日英対訳条文コーパスの構築、変更極小性の自動評価指標の設計、新訳文の生成時に旧訳文中の表現をコピーできるニューラル機械翻訳手法を開発した。実験の結果、提案手法は変更極小性について高い性能を示すことと、流暢性・妥当性について従来手法に匹敵することを明らかにした。

研究成果の学術的意義や社会的意義

社会のグローバル化に伴い、わが国の法情報の国際的発信が国内外から求められている。しかし、法令の新規制定や改正に対する英訳作業の遅れや訳語の一貫性の不徹底などの課題がある。本研究は、法令の一部改正に伴う英訳修正を支援する機械翻訳手法を開発し、英訳作業の効率化・高精度化に貢献した。また、訳文の妥当性と流暢性だけでなく、訳文の変更極小性も訳文の評価として新たに要求することにより、一過性の翻訳だけでなく、原文の修正に対して動的・継続的に対応する機械翻訳技術の開発という研究課題を新たに設定し、それに対応するためのコピー機能付きニューラル機械翻訳という新たな技術を開発した。

研究成果の概要（英文）：This research aims to develop a machine translation technology that supports the revision of English translations of Japanese statutes in response to their partial amendments, in order to quickly and internationally disseminate Japanese statutes information necessary for the globalization of society. In particular, we require focal translation: only the translation for the expressions that are amended in the statutory sentences should be revised without revising the others.

To establish this, (1) we compiled a corpus which consists of Japanese-English statutory sentences for before and after amendment, (2) we designed an automatic evaluation metrics for focality, and (3) we developed a machine translation technology that can copy expressions in the former translation when generating a new translation. Experimental results showed that the proposed method achieved high performance in focality, and was comparable to conventional methods using a large corpus in fluency and validity.

研究分野：自然言語処理

キーワード：法情報処理 法令翻訳 法制執務 リーガルテック 機械翻訳 自然言語処理

1. 研究開始当初の背景

社会のグローバル化に伴う法制度の透明化のために、わが国はその法情報を国際的に発信することを国内外から強く求められている。しかし、日本法令の英訳は従来、人手で個別に行われてきたため、信頼性や品質に難があり、内容の理解に支障を来す場合があった。そこで、日本政府（後に法務省所管）は「法令用語日英標準対訳辞書」（初版 2006）の整備や日本法令外国語訳データベースシステム（Japanese Law Translation Database System, JLT）の開発（2009）により、訳語・訳文の統一や英訳法令の一元管理・公開を図った。

しかし、JLT は所期の目的を十分に達成していない。実際、2020 年 1 月 7 日現在で効力のある 9,994 法令のうち、英訳が公開されているものは 697 法令（= 7.0%）に過ぎない。また、法令の翻訳決定から英訳公開までに平均 2 年程度を要し、法令の新規制定や改正に対して英訳作業は即応していない。さらに、頻出の定型表現である「この法律は、公布の日から施行する。」の英訳が 30 通りもあるなど、訳語・訳文の統一は徹底されていない。これは、対訳辞書や対訳法令を公開しても、英訳作業自体は依然として最初から人手に依存しているからであると考えられる。すなわち、日本法令の英訳作業を効率化・高精度化することが求められる。

2. 研究の目的

本研究は、日本法令の英訳支援のための機械翻訳技術の開発を目的とする。特に、法令改正に伴う英訳法令の修正に対するものに焦点を当てる。

新規制定された法令の大半は、既存法令の一部を改正するものである。実際、法令のうち法律では、その数の約 70% が一部改正を目的としたものである。法令の改正に伴い、その英訳の修正も必要になる。しかし、改正に対応した英訳修正であっても、中には、旧訳文とは独立に新原文を英訳したため、改正部分だけが英訳修正の対象とならず、非改正部分の英訳も修正され、その結果、新旧訳文を参照しないで新旧訳文を比べたときに、改正部分と非改正部分の区別が不明確であるという事態が生じることになる。これは改正箇所について誤解を招くおそれがあり、好ましくない。法令の一部改正では、一つの条文内での文字列修正（置換・挿入・削除）が修正箇所の約 75% を占めるので、その英訳修正を支援すれば効果は大きい。

さらに、機械翻訳では一般に、訳文の妥当性（原文と訳文の意味は同じであること）と訳文の流暢性（訳文は目的言語文として自然であること）が要求される。本研究では、それらに加えて、新旧訳文において改正部分と非改正部分の区別を明確にするために、旧条文の改正部分に対する翻訳だけを修正し、非改正部分に対する翻訳は可能な限り修正しないで旧訳を用いること（訳文の変更極小性（focality））も要求する。

3. 研究の方法

本研究は、法令改正に伴う英訳法令の修正を支援する機械翻訳技術を開発し、計算機実験を通じて評価することにより、その有効性を明らかにする。特に、新訳文の生成時に、新原文に照らして、旧訳文中で用いられた表現を必要に応じてコピーできる機能を備えたニューラル機械翻訳手法を開発し、その有効性を示す。課題は次の三つである。

(1) 新旧対照・日英対訳条文コーパスの構築

旧原文、新原文、旧訳文、新訳文の四つ組からなる新旧対照・日英対訳条文コーパスを構築し、機械学習および評価実験のためのデータとする。

(2) 変更極小性の自動評価指標の設計

訳文の変更極小性を定量化するための指標を設計する。この指標は、旧訳文、新訳文、および参照訳に出現する単語や単語列の比較に基づく。

(3) コピー機能付きニューラル機械翻訳手法の開発・評価

新訳文の生成時に、新原文に照らして、旧訳文中で用いられた表現を必要に応じてコピーできる機能を備えたニューラル機械翻訳手法を開発する。(1)で構築したコーパス、(2)で設計した評価指標および機械翻訳システムに対する既存の評価指標（BLUE、RIBES）を用い、計算機実験により、提案手法の性能を評価する。

表1 法令文四つ組データの作成状況

年度	法令文四つ組の数
2020	1,351
2021	1,670
2022	1,477
2023	1,657
合計	6,155

4. 研究成果

本研究の主な成果は、次のとおりである。

(1) 新旧対照・日英対訳条文コーパスの構築

次の手順により、新旧対照・日英対訳条文コーパスを構築した。

- a) 法務省・日本法令外国語訳データベースシステム (JLT) (<http://www.japaneselawtranslation.go.jp/>) で公開されている法令日英対訳コーパスから法令データを取得した。
 - b) a) で取得した法令のうち、収録されている法令が現行改正バージョンでない場合、総務省・e-Gov 法令検索 (<https://elaws.e-gov.go.jp/>) から現行改正バージョンを取得した。
 - c) a), b) で取得した法令に対して、JLT 収録の最新改正バージョンの日本語原文 (旧原文)、その英訳文 (旧訳文)、および、e-Gov 法令検索収録の最新改正バージョンの日本語原文 (新原文) の3要素からなる法令文の組 (法令文三つ組データ) 11,809 組を作成した。
 - d) c) で作成した法令文三つ組データのうちから、新原文を手で翻訳して、新訳文を作成した。これを加えて、法令文四つ組データを構成した。なお、新原文の翻訳の際には、新原文だけでなく、旧原文と旧訳文を参照し、かつ、旧原文と新原文の差分を下線で示すとともに、旧訳文を編集することにより新原文を作成した。これにより、訳文の変更極小性を確保できるようにした。
- その結果、既存の分と合わせて、法令文四つ組データ 6,155 組を蓄積し、新旧対照・日英対訳条文コーパスを構築した。なお、この作業は、表1に示すように、年度ごとに進行した。

(2) 変更極小性の評価指標の設計

新旧訳文間の変更極小性を評価するための指標 ISDIT (Inclusive Score for Differential Translation) を次のように定義した。

$$ISDIT(S_{\text{新訳文}}; S_{\text{旧訳文}}, S_{\text{参照訳}}) = \text{Pre}(S_{\text{新訳文}}; S_{\text{参照訳}}) \cdot \text{Rec}(S_{\text{新訳文}}; S_{\text{旧訳文}}, S_{\text{参照訳}})$$

ここで、 $\text{Pre}(S_{\text{新訳文}}; S_{\text{参照訳}})$ は参照訳に対する新訳文の単語 n -gram 適合率であり、 $\text{Rec}(S_{\text{新訳文}}; S_{\text{旧訳文}}, S_{\text{参照訳}})$ は参照訳と旧訳文に共通して出現する単語 n -gram に対する新訳文の単語 n -gram 再現率である。ISDIT はそれらの積である。それらの項目の詳細は、Yamakoshi ら [2021] に示す。すなわち、参照訳の表現が新訳文において多く出現するほど、また、旧訳文に存在する表現が新訳文に多く出現するほど、ISDIT の値は高くなる。

(3) コピー機能付きニューラル機械翻訳手法の開発・評価

a) 手法の概要

訳文の流暢性、妥当性に加えて、訳文の変更極小性の鼎立を目指した機械翻訳手法として、Copiable Translation Transformer (CTT) を開発した。CTT の構造を図1に示す。CTT は、旧訳文と新原文を入力とし、新訳文を生成する Transformer [Vaswani, et al., 2017] をベースとした手法であり、旧訳文中の単語をコピーできる機能を備えている。具体的には、単語出力の際に、語彙中の各単語の「生成」の尤度と、旧訳文中の各単語の「コピー」の尤度を求め、もっとも高い尤度の単語を出力する。その際、「生成」と「コピー」の別の情報、また、「コピー」の場合は、旧訳文中の何番目の単語をコピーしたのかの情報を併せて出力する。

Transformer をベースとしてコピー機能を有する手法には、Junczys-Dowmunt らの手法 [2018] やそれを改良した Huang らの手法 [2019] などがある。これらの手法は、語彙中の各単語について、生成の尤度とコピーの尤度を合算したうえで最終的な尤度を求めている。そのため、出力文中の各単語からは、その単語が生成されたのかコピーされたのかを区別できない。一方、CTT は、出力文中の各単語について、「生成」と「コピー」の別、後者の場合はコピー元の単語の場所も出力する。この情報を用い、旧訳文中においてコピーされなかった単語と新原文中において生成された単語をマーキングすることにより、新訳文を旧訳文との差分 (下線で表示) とともに生成できる。

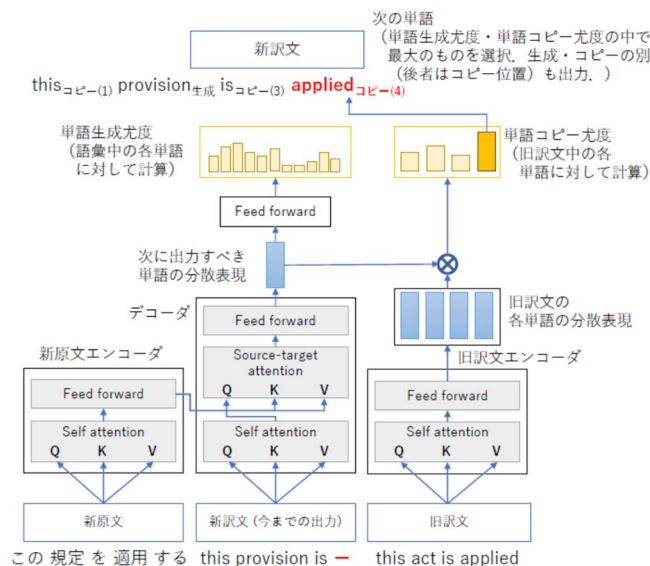


図1 Copiable Translation Transformer (CTT) の構造

b) 手法の評価

機械翻訳手法として広く用いられている Transformer [Vaswani, et al., 2017], 変更極小性に強い手法である小酒井らの手法 [2017] (Kozakai), Transformer ベースでコピー機能を有する手法である Huang らの手法 [2019] (Huang) を提案手法 (CTT) と比較する。比較する手法の一覧を表 2 に示す。ここで、小酒井らの手法は、この手法は、原文と関連する文の訳文中の単語をコピーする Koehn らの手法 [2010] を改正後法令文の翻訳に特化させたものである。すなわち、旧訳文中の単語をコピーする手法であり、旧原文と新原文の間の不変部分特定し、その不変部分と対応する旧原文中の翻訳を固定し、その他の部分のみを統計的機械翻訳 (SMT) によって新たに翻訳するものであり、変更極小性に強い。

表 2 中の学習データのうち、「四つ組コーパス」は、(1) で構築した新旧対照・日英対訳条文コーパスに蓄積した法令文四つ組であり、全 6,155 組のうち、最大 5,954 組を学習データとして用いた。なお、Huang および CTT の学習には、新原文、新訳文、旧訳文を用い、Transformer の学習には、新原文と新訳文を用いた。また、表 2 中の学習データのうち、「対訳コーパス」は、法務省・日本法令外国語訳データベースシステム (JLT) から取得した対訳文 391,758 組から構築し、すべて学習データとして用いた。さらに、学習データの規模による性能差を調査するため、Transformer は Trm_S と Trm_L の 2 モデルを、また、CTT も CTT_S , CTT_L の 2 モデルをそれぞれ学習させた。学習時のハイパーパラメータなど実験の詳細は、山腰ら [2024] に示す。一方、テストデータは、上述の 6,155 組のうち、上述の学習データとは異なる 201 組を用いた。

各手法の評価には、機械翻訳 (流暢性, 妥当性) の評価指標 BLEU, RIBES と、本研究で設計した変更極小性の評価指標 ISDIT (2) 参照) を用いた。

実験結果を表 2 に示す。BLEU, RIBES, ISDIT において最大スコアを記録した手法は、それぞれ、Kozakai, Trm_L , CTT_L であった。Kozakai が BLEU, ISDIT で高いスコアを獲得した理由は、改正の対象外と判断した旧訳文の表現を固定し、当該表現を出力文中で確実に反映できたためと思われる。一方、RIBES は語順の一致度により評価するため、流暢な文を出力できるニューラル機械翻訳の手法 Trm_L において高い値を獲得したと考えられる。一方、CTT は、規模の小さいコーパスを用いて学習したにもかかわらず、大規模対訳コーパスを用いた Kozakai や Trm_L に匹敵する結果となった。

以上の結果から、提案手法 CTT は、特に訳文の変更極小性について高い性能を発揮することが示され、法令改正に伴う英訳法令の修正を支援する機械翻訳技術として有効であることを昭告した。流暢性, 妥当性について性能 (BLUE の値) を改善するためには、 CTT_S と CTT_L の学習データの規模差とその性能差から、コーパスの大幅な増強が必要であると考えられる。

表 2 実験結果

略称	手法	学習データ種類	学習データサイズ	BLEU	RIBES	ISDIT
Trm_S	Transformer	四つ組コーパス	4,297 組	24.63	57.19	12.40
Trm_L	Transformer	対訳コーパス	391,758 組	80.31	93.44	71.36
Kozakai	小酒井ら	対訳コーパス	391,758 組	82.79	92.04	77.53
Huang	Huang ら	四つ組コーパス	4,297 組	70.33	92.27	72.78
CTT_S	CTT (提案手法)	四つ組コーパス	4,297 組	74.46	93.19	77.64
CTT_L	CTT (提案手法)	四つ組コーパス	5,954 組	75.82	93.42	78.35

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 3件 / うち国際共著 0件 / うちオープンアクセス 5件）

1. 著者名 Takahiro Yamakoshi, Yasuhiro Ogawa, Katsuhiko Toyama	4. 巻 13859
2. 論文標題 Differential-Aware Transformer for Partially Amended Sentence Translation	5. 発行年 2023年
3. 雑誌名 Lecture Notes in Computer Science	6. 最初と最後の頁 5-22
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-031-29168-5_1	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 佐野 智也、外山 勝彦、駒水 孝裕、増田 知子	4. 巻 7
2. 論文標題 歴史情報としての法令データベースの構築	5. 発行年 2023年
3. 雑誌名 デジタルアーカイブ学会誌	6. 最初と最後の頁 s142-s145
掲載論文のDOI (デジタルオブジェクト識別子) 10.24506/jsda.7.s2_s142	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 佐野 智也、外山 勝彦、増田 知子	4. 巻 6
2. 論文標題 近代日本の法律・勅令を踏まえた法令標準XMLスキーマの提案	5. 発行年 2022年
3. 雑誌名 デジタルアーカイブ学会誌	6. 最初と最後の頁 s226-s229
掲載論文のDOI (デジタルオブジェクト識別子) 10.24506/jsda.6.s3_s226	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Yamakoshi Takahiro, Komamizu Takahiro, Ogawa Yasuhiro, Toyama Katsuhiko	4. 巻 12758
2. 論文標題 Differential Translation for Japanese Partially Amended Statutory Sentences	5. 発行年 2021年
3. 雑誌名 New Frontiers in Artificial Intelligence: JSAI-isAI 2020 Conference and Workshops, Revised Selected Papers, Lecture Notes in Computer Science	6. 最初と最後の頁 162-178
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/978-3-030-79942-7_11	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Takahiro Yamakoshi, Takahiro Komamizu, Yasuhiro Ogawa, Katsuhiko Toyama	4. 巻 -
2. 論文標題 Evaluation Scheme of Focal Translation for Japanese Partially Amended Statutes	5. 発行年 2021年
3. 雑誌名 Proc. 8th Workshop on Asian Translation	6. 最初と最後の頁 124-132
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計11件 (うち招待講演 0件 / うち国際学会 1件)

1. 発表者名 山腰貴大, 小川泰弘, 外山勝彦
2. 発表標題 変更極小性を考慮した改正後法令文翻訳手法
3. 学会等名 言語処理学会第30回年次大会
4. 発表年 2024年

1. 発表者名 前田郁勝, 小川泰弘, 外山勝彦
2. 発表標題 民事第一審判決書のXMLデータ化
3. 学会等名 言語処理学会第30回年次大会
4. 発表年 2024年

1. 発表者名 加藤優吾, 小川泰弘, 外山勝彦
2. 発表標題 用語制約が多い翻訳に対する手法の提案
3. 学会等名 言語処理学会第30回年次大会
4. 発表年 2024年

1. 発表者名 平松悠太, 小川泰弘, 外山勝彦
2. 発表標題 決算短信における業績予測文のブートストラップを用いた抽出と極性付与
3. 学会等名 第20回テキストアナリティクス・シンポジウム
4. 発表年 2023年

1. 発表者名 澤柳翔太, 小川泰弘, 外山勝彦
2. 発表標題 個人向けテキスト難易度測定システムの評価
3. 学会等名 第20回テキストアナリティクス・シンポジウム
4. 発表年 2023年

1. 発表者名 佐野智也, 外山勝彦, 駒水孝裕, 増田知子
2. 発表標題 法令変遷の連続的把握のための法令データベースの構築
3. 学会等名 情報ネットワーク法学会第23回研究大会
4. 発表年 2023年

1. 発表者名 平松悠太, 小川泰弘, 外山勝彦
2. 発表標題 決算短信における業績予測文の抽出と極性付与
3. 学会等名 令和5年度電気・電子・情報関係学会東海支部連合大会
4. 発表年 2023年

1. 発表者名 澤柳翔太, 小川泰弘, 外山勝彦
2. 発表標題 個人向けテキスト難易度測定
3. 学会等名 令和5年度電気・電子・情報関係学会東海支部連合大会
4. 発表年 2023年

1. 発表者名 Takahiro Yamakoshi, Yasuhiro Ogawa, Katsuhiko Toyama
2. 発表標題 Differential-aware Transformer for Partially Amended Sentence Translation
3. 学会等名 16th Int. Workshop on Juris-informatics (国際学会)
4. 発表年 2022年

1. 発表者名 駒水孝裕, 外山勝彦, 河口信夫, 佐野智也
2. 発表標題 法令間の関係を利用したモビリティ関連法令検索に関する一考察
3. 学会等名 人工知能学会セマンティックウェブとオントロジー研究会
4. 発表年 2022年

1. 発表者名 駒水孝裕, 安田理恵, 外山勝彦
2. 発表標題 移動イノベーションに向けた法情報活用基盤
3. 学会等名 令和3年度電気・電子・情報関係学会東海支部連合大会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	山腰 貴大 (YAMAKOSHI Takahiro)		
研究協力者	小川 泰弘 (OGAWA Yasuhiro) (70332707)	名古屋市立大学・データサイエンス学部・准教授 (23903)	
研究協力者	駒水 孝裕 (KOMAMIZU Takahiro) (30756367)	名古屋大学・数理・データ科学教育研究センター・准教授 (13901)	
研究協力者	松浦 好治 (MATSUURA Yoshiharu) (40104830)	名古屋大学・大学院法学研究科・名誉教授 (13901)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------