ニューラル機械翻訳の品質向上に向けた原言語における言い換え

佐藤紗都 [*1]　上垣外英剛 [*2]　高村大也 [*2*3]　奥村学 [*2]
[*1] 東京工業大学工学院　[*2] 東京工業大学科学技術創成研究院　[*3] 産業技術総合研究所
{satosato, kamigaito, takamura, oku}@lr.pi.titech.ac.jp

1　はじめに

文化の異なる言語間の機械翻訳においては，翻訳先の言語には存在しない概念や意味を持つ単語を翻訳しなければならないという問題がある．例えば，日本語の「詠み人知らず」を英語の “author unknown” に正しく翻訳するためには，翻訳システムの訓練データ中に「詠み人知らず」が含まれている必要がある．しかし，和歌や俳句など特定のドメインにおいてのみ出現するこのフレーズが一般的なドメインを対象として作成された訓練データ中に含まれることは少ない．結果として，一般的なドメインを対象に訓練された翻訳システムが「詠み人知らず」のように特定のドメインにのみ出現する表現を正しく翻訳することは難しい．一方で，“author unknown” の直訳である「著者不明」は一般的なドメインを対象として作成された訓練データ中に含まれることが多いため，正しく翻訳することは容易である．そのため，「詠み人知らず」のように特定のドメインにのみ出現する表現を「著者不明」のように一般的なドメインに出現しやすい表現に言い換えることができれば，「詠み人知らず」を正しく翻訳することは容易になると考えられる．

前述の特定のドメインにのみ出現する表現は未知語として扱われることが多い．未知語の多い入力文に対応する手法は，機械翻訳システム自体を改良する手法と，機械翻訳システムの入出力を編集する手法の 2 つに大別できる．本研究では既存の機械翻訳システムを活用した後者の手法に着目する．具体的には図 1 のように，特定のドメインにのみ出現する，機械翻訳しにくい表現を含む「翻訳しにくい原文」を，一般的なドメインに出現しやすい表現で言い換えられた「翻訳しやすい原文」に前編集し，既存の機械翻訳システムに入力する．これによって，既存の機械翻訳システム内部を変更することなく，機械翻訳システムが学習しているドメインとは異なるドメインの翻訳精度の向上を図ることが可能になる．特に，本研究では歴史や文化に関する文書の日英ニューラル機械翻訳の精度向上を目的とし，特定のドメインにのみ出現する機械翻訳しにくい表現を一般的なドメインに出現しやすい表現へ，自動生成した辞書を用いて自動で言い換えるシステムを構築する．その結果，言い換えた文のみの自動評価で 0.32%の METEOR の向上が見られ，また情報性に関しての人手評価に有意差が見られた．

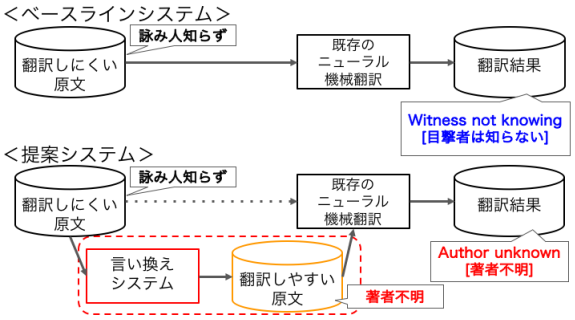

図 1: 提案システム概要

2　関連研究

機械翻訳システムの入出力を編集する手法には，入力文を編集する手法 (pre-edit) と，出力文を編集する手法 (post-edit) がある．宮田ら [1] によると，産業翻訳では pre-edit よりも post-edit が導入されることが多い．それは出力された不完全な機械翻訳文を修正する post-edit は最終的な翻訳品質を高める必須の作業であるためである．しかし，その一方で pre-edit は post-edit よりも上流工程で機械翻訳文を制御し，全体の生産性を高めることができるため，入力文を多言語に翻訳する際にメリットが大きいとされる．そのため，これまでにも様々な pre-edit を用いた手法が提案されているが，その多くは入力文中の語の区切りが明確な英語を対象とした手法が多い．

語の区切りが不明確な日本語を対象とした pre-edit に関する近年の研究として，南條らと宮田らの研究があげられる．南條ら [2] の研究では複数の統計的機械翻訳 (SMT) システムを用いて英文を翻訳することで前編集用学習データ対を収集し，そのデータ対を用いて学習済みの言語モデルを SMT 向けの前編集システムとして提案している．ここで前編集用学習データ対とは翻訳コーパスの日本語文と，翻訳コーパスの英文の英日翻訳文の対のことである．それに対して本研究では，髙橋ら [3] が指摘するように SMT に比べ入力文の細かな差異による出力文の変化が大きいニューラル機械翻訳 (NMT) を対象とする点が異なる．また，

NMT は SMT に比べ，多少文法が破綻していてもある程度翻訳が可能であるため，南條らのように言語モデルを用いて文法なども含めて前編集するよりも，単純な単語の言い換えによって未知語を減らしたほうが効果的である可能性が高いと考えられる．そこで，本研究では単純な言い換えによる NMT の精度の変化を考察するため，自動構築した辞書による言い換え手法を用いる．

宮田らの研究 [4] では，地方自治体の手続き文書を SMT を用いて日英翻訳するための Web ベースの入力文書き換えサポートシステムを提案している．このシステムはユーザの入力文に対して細かな文脈依存の制限言語の規則違反などを指摘することによって，ユーザに入力文の書き換えを促し，SMT の翻訳精度向上を目指している．それに対し，本研究ではシステム側で入力文の自動編集を行う点，および対象とする翻訳手法が NMT である点の 2 点が異なる．

また，本研究と類似している研究として，梶原ら [5] のようなテキスト平易化が挙げられる．梶原らの研究と本研究が異なる点は，言い換えの目的が NMT の精度向上である点，および，文を言い換える判断を文がより平易になるかという基準ではなく，特定のドメインにのみ出現する機械翻訳しにくい表現であるかという基準で行う点の 2 点である．

3 提案手法

3.1 NMT 向け言い換えシステムの概要

提案手法は言い換え辞書の作成と言い換え辞書を用いた言い換え文生成の 2 つの過程で構成される．

言い換え辞書の作成は 3 段階で構成される．まず，商用 NMT モデルを用いて人手英訳文から翻訳しやすい日本語文を取得し，SMT モデルを用いて翻訳しにくい表現と翻訳しやすい表現のフレーズ対の抽出を行う．次に，フレーズ対の出現確率や，形態素や品詞などの情報，さらに共起指標を用いて，不要と思われるフレーズ対をフィルタリングし，言い換え辞書を構築する．

言い換え辞書を用いた言い換え文生成は 2 段階で構成される．まず，入力文を文節に区切り，その文節のうち言い換え辞書に登録されているものに対して言い換えを行い，入力文を前編集する．そして編集前後の文のうち，流暢な方を言い換え結果として出力する．

以上により生成した文を既存の NMT システムへの入力とする．次節から各システムについて説明する．

3.2 言い換え辞書の作成

3.2.1 フレーズ対の抽出

まず，翻訳しやすい原文の取得について説明する．本研究で対象とするデータは歴史や文化に関する文書の日英パラレルコーパスである京都フリー翻訳タスク (KFTT)[6] のデータである．このデータは京都関連の Wikipedia 記事が用いられているため，一般的なドメインを対象として作成された訓練データ中にはあまり出現しない表現が多く含まれている．また，翻訳の専門家が日英翻訳しているため，日本固有の表現を英語圏の表現に直して英訳されている．そのため，大規模データで学習された翻訳品質の高い NMT モデルにより，この人手英訳文を英日翻訳することで機械翻訳モデルにとって翻訳しやすい日本語の表現が取得可能であると考えられる．そこで，前述のような NMT モデルである Cloud Translation API (CTA)[1] を用いて人手英訳文から翻訳しやすい表現を含んだ日本語文の取得を試みる．しかし，CTA は固有名詞の誤訳が多いという問題がある．そのため，Stanford Named Entity Recognizer[2] を用いて，人手英訳文中の固有名詞を全て翻訳に影響しないタグに置き換えてから，CTA による翻訳を行う．例えば，Kotaro という人名は P1 に置き換えられる．

次にフレーズ対の抽出について説明する．元々の日本語文から翻訳しやすい日本語文への日日翻訳 SMT モデルを学習する．それにより，元々の日本語文に含まれるフレーズと翻訳しやすい日本語文に含まれるフレーズが対となったフレーズテーブルを取得する．以下では元々の日本語文に含まれるフレーズを言い換え前のフレーズ，翻訳しやすい日本語文に含まれるフレーズを言い換え後のフレーズ，この 2 つの対をフレーズ対と呼ぶ．そして，フィッシャーの正確確率検定を利用したフィルタ [7] によって，そのフレーズテーブルから信頼性が低いものを削除する．

3.2.2 形態素情報によるフレーズ対のフィルタリング

はじめに前節で作成したフレーズテーブルのフレーズ対に対して，それぞれ形態素解析を行った．そして，その形態素解析結果を用いて，言い換えとして相応しくない下記の 4 種のフレーズ対を削除する．

- 固有名詞を含む
- 記号を含む
- 代名詞を含む
- フレーズ対の差異が，助詞のみである

3.2.3 翻訳確率および共起確率によるフィルタリング

前節で作成した言い換え辞書には「言い換え前後で意味が変わってしまうフレーズ対」,「言い換える前のフレーズが特定のドメイン固有のものではない，もしくは言い換え後のフレーズが特定のドメイン固有のものであるフレーズ対」といった本研究で有用でないフレーズ対が含まれている．そこで，これから定義する 3 つの指標を用いてフレーズ対をフィルタリングする．以下では言い換え前のドメイン k のフレーズを w_k，言い換え後のドメイン g のフレーズを w_g とする．また，

[1] https://cloud.google.com/translate/?hl=ja
[2] https://nlp.stanford.edu/software/CRF-NER.shtml

任意のフレーズが出現する事象を x，任意のフレーズが言い換え前のドメイン k のフレーズ w_k として出現する事象を y_k，任意のフレーズが言い換え後のドメイン g のフレーズ w_g として出現する事象を y_g とする．

はじめに「言い換え前後で意味が変わってしまうフレーズ対」をフィルタリングするための評価値 $PTP(w_k, w_g)$ について説明する．各フレーズ対について，その出現回数に基づく最尤推定からフレーズ翻訳確率 $\phi(w_k|w_g)$ と逆フレーズ翻訳確率 $\phi(w_g|w_k)$ が求まる．これらを用いて以下の式を定義する：

$$PTP(w_k, w_g) = -\log_e \phi(w_k|w_g) - \log_e \phi(w_g|w_k). \quad (1)$$

式 (1) が閾値より大きいフレーズ対を削除する．

次に「言い換える前のフレーズが特定のドメイン固有のものではない，もしくは言い換え後のフレーズが特定のドメイン固有のものであるフレーズ対」をフィルタリングする．自己相互情報量（PMI）と情報利得（IG）を指標の候補とした．

まず PMI を用いた指標 $Score_{\mathrm{PMI}}$ を定義する：

$$PMI(x, y) = \log_2 \frac{P(x, y)}{P(x)P(y)}, \quad (2)$$

$$Score_{\mathrm{src}}(w_k) = PMI(w_k, y_k) - PMI(w_k, y_g), \quad (3)$$

$$Score_{\mathrm{trg}}(w_g) = PMI(w_g, y_g) - PMI(w_g, y_k), \quad (4)$$

$$Score_{\mathrm{PMI}} = Score_{\mathrm{src}} + Score_{\mathrm{trg}}. \quad (5)$$

次に IG を用いた指標 $Score_{\mathrm{IG}}$ を定義する．ここで，$C = \{y_k, y_g\}$，$P(\overline{w}) = 1 - P(w)$ とする：

$$\begin{aligned} G(w) &= -\sum_{i \in C} P(i) \log_2 P(i) \\ &\quad + P(w) \sum_{i \in C} P(i|w) \log_2 P(i|w) \\ &\quad + P(\overline{w}) \sum_{i \in C} P(i|\overline{w}) \log_2 P(i|\overline{w}), \quad (6) \\ Score_{\mathrm{IG}} &= G(w_k) + G(w_g). \quad (7) \end{aligned}$$

式 (5) もしくは式 (7) が閾値より小さいフレーズ対を削除する．

3.3 言い換え辞書を用いた言い換え文生成

最初に，日本語係り受け解析器を用いて入力文を文節で区切る．次に，その文節が言い換え辞書に登録されているフレーズと一致した場合は言い換えを行う．この操作を文頭から文末まで行う．最後に，Web からクローリングしたデータで学習した RNN 言語モデルにより言い換え前の文と言い換え後の文について文の生成確率を計算し，さらに許容度 SLOR[8] を計算する．言い換え後の文の許容度の方が高い場合に，言い換えを実行する．

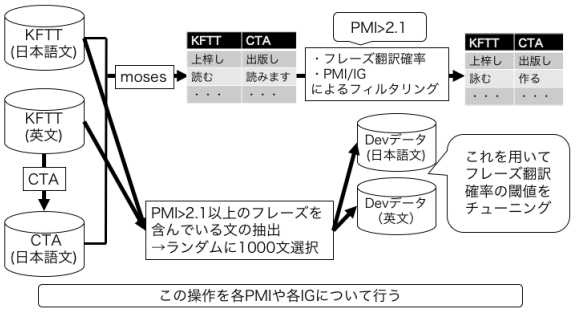

図 2: 実験手法

4 評価実験

4.1 実験設定

「言い換え前後で意味が変わってしまうフレーズ対」の指標と「言い換える前のフレーズが特定のドメイン固有のものではない，もしくは言い換え後のフレーズが特定のドメイン固有のものであるフレーズ対」の指標の 2 つの閾値によって言い換え辞書を作成し，既存の NMT モデルでの評価を行い，閾値をチューニングしていく．「言い換え前後で意味が変わってしまうフレーズ対」の指標として前述の $PTP(w_k, w_g)$，「言い換える前のフレーズが特定のドメイン固有のものではない，もしくは言い換え後のフレーズが特定のドメイン固有のものであるフレーズ対」の指標として前述の $Score_{\mathrm{PMI}}$ もしくは $Score_{\mathrm{IG}}$ を用いる．

本研究では SMT モデルとして moses [9]，形態素解析器として MeCab[3]，MeCab の追加辞書として Neologd[4]，日本語係り受け解析器として CaboCha[5]，RNN 言語モデルとして RNNLM Toolkit[6] を用いた．本研究で対象とするデータは前述したように京都フリー翻訳タスク (KFTT) のデータとした．KFTT で配布されているテストデータでは本システムで言い換えの対象とするフレーズが 30 文程度しか含まれていない．よって，KFTT の訓練データからランダムに 1 万文抽出し，残りを訓練データとした．また，抽出した 1 万文から訓練データに同一文が存在するものを削除し，残った 9,216 文をテストデータとした．

テストデータと同様に，KFTT で配布されている開発データには言い換え辞書に含まれているフレーズが僅かであった．そこで図 2 のように，$Score_{\mathrm{PMI}}$ のある値 (もしくは $Score_{\mathrm{IG}}$) に関してチューニングを行う場合，その値以上の評価値をもつフレーズ対を含む文を訓練データからランダムに 1,000 文抽出し，開発データとして用いた．そして，開発データをそのまま翻訳した結果の評価値と本システムを用いて言い換

[3] http://taku910.github.io/mecab/
[4] https://github.com/neologd/mecab-ipadic-neologd
[5] https://taku910.github.io/cabocha/
[6] http://www.fit.vutbr.cz/ imikolov/rnnlm/

表 1: テストデータ全文（9,216 文）に対するベースラインと提案手法の自動評価

	BLEU[%]	METEOR[%]
ベースライン	11.45	22.10
提案手法	11.45	**22.11**

表 2: テストデータ中の言い換え文（228 文）に対するベースラインと提案手法の自動評価

	BLEU[%]	METEOR[%]
ベースライン	**10.41**	21.69
提案手法	10.30	**22.01**

えてから翻訳した結果の評価値を比較し，その差分がもっとも大きいものを閾値として採用した．最終的なシステムの評価実験として，その結果を用いてテストデータを翻訳し，自動評価と人手評価を行った．

4.2 評価手法

評価としては自動評価と人手評価を行った．

自動評価は機械翻訳の自動評価として一般的な BLEU[10] および METEOR[11] で行った．

人手評価は Amazon Mechanical Turk[7]を用いて，米国在住者 5 人に人手翻訳文・ベースラインでの機械翻訳結果・本システムを用いた機械翻訳結果を提示し，以下の 2 点でより良い機械翻訳結果を選択させた．

- 情報性：人手翻訳文の意味と自動翻訳文の意味が近いか（文の意味の欠落がないかに注目）
- 文の平易性：理解しやすい簡単な文であるか

5 実験結果

5.1 自動評価

テストデータにおける BLEU と METEOR での評価結果を表 1 と表 2 に示す．全文での評価では効果が見られなかったが，言い換えた文のみでの評価では 0.32%の METEOR の向上が見られた．BLEU はどちらの評価でも向上が見られなかったが，METEOR の評価には同義語辞書が用いられているため，METEOR が向上していることは，本手法では意味的な内容が訳文でも保持されている可能性があることを示唆している．また，全文での評価では効果が明確でなかった理由は，辞書を用いた言い換えが実際に行われた文がテストデータ全体の 2%程度であったためであると考えられる．

5.2 人手評価

表 3 に示した人手評価結果からわかるように，提案システムの出力が高く評価されることの方が多かった．情報性に関しては Permutation Test によって統計的に有意であることが示された ($p < 0.05$)．また，図 3 に示す例のように名詞句の言い換えによって翻訳結果が改善しているものが多くみられた．

[7]https://www.mturk.com/

表 3: 人手評価

	提案システムが高評価の割合 [%]	文ごとの評価の標準偏差 [%]
情報性	53.30	32.50
文の平易性	51.89	31.37

<正解英文>
His secular surname was Okumura.

<言い換え前の日本語文> <言い換え前の日本語文の英訳>
俗姓は奥村氏。 The suicide name is Mr. Okumura.
<言い換え後の日本語文> <言い換え後の日本語文の英訳>
世俗的な姓は奥村氏。 The worldly surname is Mr. Okumura.

図 3: 実際に高評価を受けた例

6 おわりに

本研究では歴史や文化に関する文書の日英ニューラル機械翻訳の精度向上を目的とし，自動生成した辞書を用いて，特定のドメインにのみ出現する機械翻訳しにくい表現を一般的なドメインに出現しやすい表現に言い換えるシステムを構築した．その結果，言い換えた文のみの自動評価で 0.32%の METEOR の向上が見られ，また情報性に関しての人手評価に有意差が見られた．

参考文献

[1] 宮田玲. ほか. 機械翻訳向けプリエディットの有効性と多様性の調査. 通訳翻訳研究への招待, No. 18, pp. 53–72, 2017.

[2] 南條浩輝. ほか. 機械翻訳の品質向上のための対訳コーパスからの統計的前編集システムの自動構築. 情報処理学会論文誌, Vol. 53, No. 6, pp. 1644–1653, jun 2012.

[3] 髙橋寛治. ほか. 機械翻訳システムの出力安定性の評価. 人工知能学会論文誌, Vol. 32, No. 5, pp. D–H33:1–4, 2017.

[4] R. Miyata, et al. Mutual: A controlled authoring support system enabling contextual machine translation. In *Proceedings of COLING*, 2016.

[5] 梶原智之. ほか. 平易なコーパスを用いないテキスト平易化. 自然言語処理, Vol. 25, No. 2, pp. 223–249, 2018.

[6] Graham Neubig. The Kyoto free translation task. http://www.phontron.com/kftt, 2011.

[7] H. Johnson, et al. Improving translation quality by discarding most of the phrasetable. In *Proceedings of EMNLP-CoNLL*, 2007.

[8] J. Lau, et al. Unsupervised prediction of acceptability judgements. In *Proceedings of ACL and IJCNLP*, pp. 1618–1628, 2015.

[9] P. Koehn, et al. Moses: Open source toolkit for statistical machine translation. In *Proceedings of ACL*, ACL '07, pp. 177–180, 2007.

[10] K. Papineni, et al. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of ACL*, pp. 311–318, Stroudsburg, PA, USA, 2002.

[11] Satanjeev B., et al. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of ACL*, 2005.