

日英ニューラル機械翻訳における対訳句の追加

今仁優希*1 村上仁一*2

*1 鳥取大学大学院 持続性社会創生科学研究科

*2 鳥取大学大学院 工学研究科 情報エレクトロニクス専攻

{m18j4007z, murakami}@eecs.tottori-u.ac.jp

1 はじめに

近年、機械翻訳においてニューラル機械翻訳 (Neural Machine Translation; NMT)[1] が注目されている。NMT は Attention という仕組みに基づき入力文と出力文における語句の対応情報を学習する手法であり、従来の句に基づく統計翻訳 (Phrase-based SMT; PBSMT) などの手法と比較して流暢性の高い翻訳文を出力することが可能である。一方で、入力文中の語句に対する出力文中の語句の対応 (Attention) の学習に誤りがある場合には、翻訳精度が低下するという問題がある。従来の PBSMT において、語句の対応 (アライメント) を学習する目的で用いられる翻訳モデルに関しても同様の問題が指摘されており、アライメントの精度を向上させる手法が提案されている。Popovic ら [2] は、セルビア語英語間の PBSMT において、対訳文に対訳句を追加し、翻訳精度の向上を実現している。日野ら [3] および池淵ら [4] は同様の手法を日本語英語間の PBSMT に用い、翻訳モデルの学習データに対訳文から作成した対訳句や辞書データから抽出した対訳句を追加することで、翻訳精度の向上を実現している。

そこで、本研究では、日英 NMT において対訳句を利用する手法を提案する。日英 NMT の学習において、対訳文に対訳文より作成した対訳句を追加し、句レベルの対応情報を付与することで語句の対応情報をモデル化する Attention を強化し、日英 NMT の翻訳精度の向上を目指す。

また、本研究では、句レベルの対応情報を付与することのみにより翻訳精度が向上することを検証する。したがって、それ以外の情報 (語彙情報など) の拡大により翻訳精度に影響が生じる可能性を除くため、外部の情報に有する対訳句データは利用しない。

2 ニューラル機械翻訳における対訳句の追加

本研究では日英 NMT における Attention モデルの学習データとして、日英の対訳文に対訳文から作成した対訳句を追加したデータを利用する。日英ニューラル機械翻訳における対訳句追加の流れを図 1 に、手順を以下に示す。

- 手順 1 対訳文から対訳句を人手および自動で作成する。
- 手順 2 対訳文と作成した対訳句を組み合わせて対訳データを作成する。
- 手順 3 手順 2 で作成した対訳データを学習データとして Attention モデル [5] の学習を行う。
- 手順 4 手順 3 で学習したモデルを用いて日英ニューラル機械翻訳を行う。

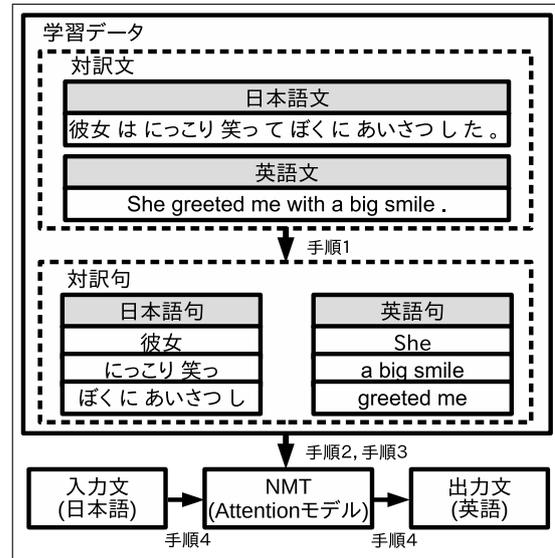


図 1 日英ニューラル機械翻訳における対訳句追加の流れ

3 実験環境

3.1 実験データ

本研究では、実験データとして日英重文複文コーパス [6] を用いる。実験で用いるデータの内訳を表 1 に、対訳文の例を表 2 に示す。

表 1 実験データの内訳

学習データ	日本語	英語
文数	100,000 文	
単語数	1,404,366 語	1,137,163 語
異なり単語数	38,530 語	37,035 語
平均文長	14.0 語	11.4 語
テストデータ	日本語	英語
文数	10,000 文	
単語数	140,671 語	113,586 語
異なり単語数	13,765 語	12,176 語
平均文長	14.1 語	11.4 語
開発データ	日本語	英語
文数	1,000 文	
単語数	14,237 語	11,591 語
異なり単語数	3,438 語	3,047 語
平均文長	14.3 語	11.6 語

表 2 対訳文の例

彼女はにっこり笑ってぼくにあいさつした。
She greeted me with a big smile .

3.2 対訳句

本研究で利用する対訳句は 3.1 節において示した対訳文を用いて作成する。また、対訳句には対訳文から人手で作成した対訳句 (以下, 人手対訳句) および対訳文から自動的に作成した対訳句 (以下, 自動対訳句) を用いる。

3.2.1 人手対訳句

人手対訳句として, 鳥バンク [7] において人手により作成したパターンを対訳文と比較し, 変数部の品詞が *AJ*(形容詞), *V*(動詞), *N*(名詞), *ADV*(副詞), *AJV*(形容動詞), *AJP*(形容詞句), *VP*(動詞句), *AJVP*(形容動詞句) となるものを利用する。

本研究で用いる人手対訳句の内訳を表 3 に, 人手対訳句の例を表 4 に示す。

表 3 人手対訳句の内訳

	日本語	英語
句数	548,694 句	
単語数	1,058,949 語	1,031,265 語
異なり単語数	35,613 語	34,823 語
平均句長	1.93 語	1.88 語

表 4 人手対訳句の例

日本語句	英語句
彼女	She
にっこり笑っ	a big smile
ぼく	me
ぼくにあいさつし	greeted me
あいさつし	greeted

3.2.2 人手対訳句の精度

人手対訳句の精度を測るため, 全ての人手対訳句から無作為に 100 対を抽出し, それぞれの対応の適切さを人手により評価した。評価基準を以下に示す。評価結果を表 5 に, 評価例を表 6 に示す。

- 適切な対応をとる対訳句
- △ 部分的に適切な対応をとる対訳句
- × 不適切な対応をとる対訳句

表 5 人手対訳句の評価結果

○	△	×
86 句	10 句	4 句

表 6 人手対訳句の評価例

評価	対訳句例
○	条件 演説し あんな賢い子ども the conditions made a speech such a clever child
△	暇 歌詞を書け 勝負 will be free Fit the words victory
×	山田さん 自社 ブラウス he their it

3.2.3 自動対訳句

自動対訳句の作成手順 [8] を以下に示す。

- 手順 1 対訳文と GIZA++[9] を利用して単語辞書を作成する。
- 手順 2 対訳文と作成した単語辞書を用いて単語レベル文パターンを作成する。
- 手順 3 対訳文と単語レベル文パターンを用いて, 対訳句を作成する。

実験において得られた自動対訳句の内訳を表 7 に示す。

表 7 自動対訳句の内訳

	日本語	英語
句数	16,286 句	
単語数	69,856 語	56,925 語
異なり単語数	11,508 語	9,879 語
平均句長	4.29 語	3.50 語

表 3 および表 7 より, 作成された自動対訳句の数は人手対訳句の数に対して 3% 程度であり, 作成された自動対訳句は大幅に少なくなっていることが確認できる。

3.2.4 自動対訳句の精度

3.2.2 節と同様に, 全ての自動対訳句から無作為に 100 対を抽出し, それぞれの対応の適切さを人手により評価した。評価結果を表 8 に, 評価例を表 9 に示す。

表 8 自動対訳句の評価結果

○	△	×
36 句	44 句	20 句

表 9 自動対訳句の評価例

評価	対訳句例
○	容赦のない残忍さ 食事の用意をし 同じ道 remorseless brutality preparing the meals same path
△	の痛みが和らい はしごを掛け フラスコ pain from the a ladder flask and heat
×	入りて の卑しい で給水の Ogata him in birth on its

表 5 および表 8 より, 作成された自動対訳句の精度は人手対訳句と比較して低くなっていることが確認できる。

3.3 ニューラル機械翻訳の実験設定

ニューラル機械翻訳のツールキットには OpenNMT [10] を用い, モデルは Luong ら [5] により提案された Global Attention を用いる。Encoder, Decoder の LSTM は 2 層とし, ユニット数は 500, 単語の分散表現のベクトルサイズは 500 を設定する。ミニバッチサイズは 40 とし, モデルの訓練は最大 32 エポック行う。

3.4 評価実験

実験では NMT の学習データに対訳文のみを用いる手法をベースラインとし、対訳文に人手対訳句を追加した学習データを用いる手法 (以下, + 人手対訳句), 対訳文に自動対訳句を追加した学習データを用いる手法 (以下, + 自動対訳句) の 2 つの対訳句追加手法をそれぞれベースラインと比較する。

翻訳精度の評価方法として, 自動評価では実験に置いて得られた 10,000 文の出力文に対して, 4 つの自動評価指標 (BLEU, METEOR, RIBES, WER) に基づいて評価を行う。人手対比較評価では出力文 10,000 文より無作為に抽出した 100 文に対して, 正確性 (adequacy: 入力文の意味をどれだけ正確に翻訳文より読み取れるか) に基づいて人手対比較評価を行う。

4 実験結果

4.1 自動評価結果

各手法の出力文 10,000 文に関して自動評価を行った。評価結果を表 10 に示す。

表 10 自動評価結果

手法	BLEU	METEOR	RIBES	WER
ベースライン	0.1534	0.4147	0.7312	0.7107
+ 人手対訳句	0.1701	0.4430	0.7411	0.6873
+ 自動対訳句	0.1607	0.4223	0.7347	0.7032

表 10 より, 自動評価の結果では, + 人手対訳句, + 自動対訳句, ベースラインの順に翻訳精度が高くなっていることが確認できる。

4.2 人手対比較評価結果

各手法の出力文 10,000 文より無作為に抽出した 100 文に対して人手対比較評価を行った。評価結果を表 11 および表 12 に示す。また, 人手対比較評価に関する表記の説明を以下に示す。

- ベースライン○
比較する出力文に対してベースラインの出力文の方がより入力文の意味に近い
- + 人手対訳句○
比較する出力文に対して + 人手対訳句の出力文の方がより入力文の意味に近い
- + 自動対訳句○
比較する出力文に対して + 自動対訳句の出力文の方がより入力文の意味に近い
- 差なし
比較する両手法の出力文には正確性において差がない
- 一致
比較する両手法の出力文が完全に同一

表 11 ベースラインと + 人手対訳句の人手対比較評価結果

ベースライン○	+ 人手対訳句○	差なし	一致
14 文	44 文	38 文	4 文

表 12 ベースラインと + 自動対訳句の人手対比較評価結果

ベースライン○	+ 自動対訳句○	差なし	一致
13 文	30 文	53 文	4 文

表 11 および表 12 より, 人手対比較評価では, 2 つの対訳句追加手法はいずれもベースラインと比較して高い翻訳精度であることがわかる。また, 翻訳精度の向上は, + 人手対訳句の方が + 自動対訳句と比較して大きいことが確認できる。

4.3 人手対比較評価例

この節では, 4.2 節の人手対比較評価における具体的な評価例および評価の観点を示す。

4.3.1 ベースラインと + 人手対訳句の人手対比較評価例

ベースラインと + 人手対訳句との対比較評価において, + 人手対訳句○とされた例を表 13 に, ベースライン○とされた例を表 14 に示す。

表 13 + 人手対訳句○

入力文	彼は自尊心が強すぎてとても引き返せなかった。
参照文	He was far too proud to go back .
ベースライン	He was too proud for his pride .
+ 人手対訳句	He was too proud to turn back .

表 13 の例において, 入力文中の語句「引き返せ」に対して, ベースラインでは相当する翻訳が得られていない一方で, + 人手対訳句では比較的正しい訳語「turn back」が出力されている。

表 14 + ベースライン○

入力文	それを見て私はどきどきした。
参照文	I throbbed at the sight .
ベースライン	The sight made me feel nervous .
+ 人手対訳句	The sight thumped me .

表 14 の例において, 入力文中の表現「私はどきどきした」に対して, ベースラインでは比較的正しく翻訳できている一方で, + 人手対訳句では「thumped me」と誤った訳語が出力されている。

4.3.2 ベースラインと + 自動対訳句の人手対比較評価例

ベースラインと + 自動対訳句との対比較評価において, + 自動対訳句○とされた例を表 15 に, ベースライン○とされた例を表 16 に示す。

表 15 + 自動対訳句○

入力文	鉄は加熱するとやわらかくなる。
参照文	Iron softens with heat .
ベースライン	Iron hardens with heating .
+ 自動対訳句	Iron becomes tender when heated .

表 15 の例において, 入力文中の語句「やわらかくなる」に対して, ベースラインでは誤った翻訳となっている一方で, + 自動対訳句では比較的正しい訳語「becomes tender」が得られている。

表 16 + ベースライン○

入力文	食卓がいっぱいでパンを置く所がない。
参照文	There is no room on the table for the bread .
ベースライン	The table is full and there is no place for bread .
+ 自動対訳句	There is no room for bread .

表 16 の例において、入力文中の語句「食卓がいっぱい」に対して、ベースラインでは相当する訳語が得られている一方で、+ 自動対訳句では翻訳が抜け落ちている。

5 考察

5.1 対訳句利用の効果

結果として、2つの対訳句追加手法において翻訳精度の向上が確認された。利用した2種類の対訳句はいずれも対訳文より作成し、対訳文外の語彙などの情報を有しないものである。これより、文レベルの対応情報に加えて句レベルの対応情報を学習データに付与することがNMTにおいて語句の対応情報をモデル化するAttentionの強化につながり、翻訳精度が向上したと考えられる。

また、2つの対訳句追加手法のうち、句の対応が比較的適切かつ多量である人手対訳句を利用する手法がより高い翻訳精度を示した。したがって、翻訳精度の向上にはより高精度かつ多量の対訳句を利用する必要があると推察される。

5.2 翻訳結果と対訳句の分析

翻訳結果と対訳句の分析から、正しい対訳句を追加したことにより翻訳精度が向上したと見られる例が確認された。+ 人手対訳句の結果について、表 13 の例において、入力文中の単語「引き返せ」に対して、+ 人手対訳句では比較的正しい翻訳となっている。これに関して、追加した人手対訳句を調査すると、「引き返せ」に対して正しい訳語「turn back」を含む対訳句が存在し、正しい句レベルの対応を学習したために正確性が向上したと考えられる。また、+ 自動対訳句の結果についても同様のことが確認できる。表 15 の例において、入力文中の語句「やわらかくなる」に対して自動対訳句中に比較的正しい訳語「become tender」が含まれている。これにより、適切な語句を選択するように学習が強化され、より高い精度の翻訳が得られたと考えられる。

一方で、対訳句を追加することで翻訳精度が低下したと考えられる例も存在する。表 14 の例において、+ 人手対訳句では出力文中に「thumped me」という誤った翻訳が生じている。これに関して、入力文中の日本語単語「どきどき」を含む人手対訳句を調査すると、英語側に「thumped」を含むものが複数確認された。これは、「胸がどきどきする」という日本語に対して「my heart thumped」という翻訳を含む対訳文から作られた対訳句である。このような対訳句の対応は必ずしも誤りとは言えないものの、表 14 の例のような文中の用法としては不適切なものとなっている。このように、対訳句を追加する手法では、句レベルの対応を重視した結果、文全体としては誤った表現を導く可能性があることが課題

としてあげられる。

5.3 人手対訳句と自動対訳句の比較

人手対訳句と自動対訳句との比較として、自動対訳句と比較して人手対訳句の方が高精度かつ多量に得られており、翻訳精度についても人手対訳句を追加する手法は自動対訳句を追加する手法と比較してより高い結果を示した。しかしながら、人手対訳句の作成には大きなコストがかかる問題がある一方で、自動対訳句の作成におけるコストは小さい。また、自動対訳句はプログラムやパラメータの改良を行うことで、より適切な対訳句の作成や、より多くの対訳句の作成が見込めるという利点がある。

6 おわりに

NMTにおいて、Attentionの学習に誤りがある場合に翻訳精度が低下するという問題が指摘されていた。そこで、本研究では、日英NMTの翻訳精度を向上する目的で、NMTにおいて対訳文に対訳文より作成した対訳句を追加する手法を提案した。

出力文100文の結果として、人手対訳句を用いた手法の評価ではベースライン○が14文に対して+ 人手対訳句○が44文、自動対訳句を用いた手法ではベースライン○が13文に対して+ 自動対訳句○が30文となった。これより、ベースラインと比較して対訳句を追加する2つの手法では翻訳精度が向上することが確認された。本研究で用いた2種類の対訳句はいずれも対訳文から作成され、対訳文外の語彙などの情報を有しないものである。したがって、翻訳精度が向上した理由は、対訳句を追加することにより、NMTにおいて語句の対応をモデル化するAttentionが強化されたためであると考えられる。今後はより高精度かつ多量の対訳句を得る手法を検討し、更なる翻訳精度の向上を試みたい。

参考文献

- [1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *In Proceedings of ICLR*, 2015.
- [2] Maja Popović and Hermann Ney. Statistical machine translation with a small amount of bilingual training data. *5th LREC SALT MIL Workshop on Minority Languages*, pp. 25–29, 2006.
- [3] 日野聡子, 村上仁一, 徳久雅人, 村田真樹. 日英統計翻訳における対訳句コーパスの効果. 言語処理学会第19回年次大会, pp. 556–559, 2013.
- [4] 池淵堅斗. 対訳文から抽出した対訳句を用いた統計翻訳. 2015年度修士論文, 2015.
- [5] Minh-Thang Luong, Hieu Pham, and Christopher D. Manning. Effective approaches to attention-based neural machine translation. *In Proceedings of EMNLP*, pp. 1412–1421, 2015.
- [6] 村上仁一, 藤波進. 日本語と英語の対訳文対の収集と著作権の考察. 第一回コーパス日本語学ワークショップ, pp. 119–130, 2012.
- [7] 鳥バンク. <http://unicorn.ike.tottori-u.ac.jp/toribank/>.
- [8] 興梠玲架. パターンに基づく統計翻訳において変数部の確率の総和を使った対訳句の抽出. 2015年度鳥取大学卒業論文, 2015.
- [9] Franz Josef Och and Hermann Ney. A systematic comparison of various statistical alignment models. *Comput. Linguist.*, p. 1951, 2003.
- [10] Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander M. Rush. Opennmt: Open-source toolkit for neural machine translation, 2017.