

# 邦字新聞における 近代文語体と現代口語体の自動翻訳の検討

稲見郁乃<sup>†1</sup> 竹本有紀<sup>†1</sup> 石川由羽<sup>†2</sup> 高田雅美<sup>†1</sup> 上田薫<sup>†3</sup> 城和貴<sup>†1</sup>

**概要:** 日本語で書かれた文章は時代や分野により文体が異なり、これらの文献から知見を得るには、専門知識を必要とする。そのため、異なる時代間の自動翻訳が求められている。邦字新聞は近代に刊行された多様な分野を有する新聞であり、分野に依存しない時代間翻訳が可能となる。そこで、本稿では、邦字新聞における近代文語体と現代口語体の自動翻訳を検討する。本翻訳において最適モデルを検討すべく、各種モデルの翻訳精度の違いを BLEU スコアにより比較検討する。また、各種モデルにおけるデータ数や文長の影響について考察する。

**キーワード:** 邦字新聞, 近代文語, ニューラル機械翻訳, Seq2Seq, ConvSeq2Seq, Transformer

## Automatic machine translation of modern classical style and present colloquial style in Hoji Shinbun

KANO INAMI<sup>†1</sup> YUKI TAKEMOTO<sup>†1</sup> YU ISHIKAWA<sup>†2</sup>  
MASAMI TAKATA<sup>†1</sup> KAORU UEDA<sup>†3</sup> KAZUKI JOE<sup>†1</sup>

**Abstract:** Sentences written in Japanese differ depending on the times and fields, and specialized knowledge is required to obtain knowledge from these documents. Therefore, automatic machine translation between different eras is desired. Hoji Shinbun are newspapers published in modern times with various fields, and can be translated between times without depending on fields. In this paper, we examine the automatic machine translation of modern classical style and present colloquial style in Hoji Shinbun. In order to examine the optimum model for this translation, we compare and examine the differences in translation accuracy of various models using BLEU scores. We also consider the effects of the number of data and sentence length on various models.

**Keywords:** Hoji Shinbun, Modern classical style, Neural Machine Translation, Seq2Seq, ConvSeq2Seq, Transformer

### 1. はじめに

日本語で書かれた文章は、時代や分野により文体が異なる。書き言葉の歴史は平安時代より始まり、明治時代半ばの言文一致運動にかけて大きく変化している。平安時代では表音仮名を基にした「雅文体」、鎌倉・室町から江戸・明治時代では和漢混淆の文語体である「候文体」が使用されている。その後、明治時代半ばの言文一致運動を経て徐々に文体が統一され、戦後に現代の文体へと変化している。

近代文語体とは、明治半ばから昭和初期に使用された文体である。この時代には、西洋の文化を導入する明治維新や文明開化がなされたため、大量の新単語が作成されている。しかし、多くの単語が定着しなかったことから、近代の文書には低頻度語が数多く存在する。また、文学作品や議会書など分野によっても文体が異なるため、近代の文書を容易に読むのは困難である。そのため、近代の文書から知見を得るためには、専門知識が不可欠となる。しかし、近代の文書は膨大に存在するため、専門知識を有する翻訳者の人手を数多く必要とする。したがって、近代文語体と

現代口語体の自動翻訳技術の開発が求められている。

近年、異言語間翻訳の分野では、流暢な訳文を生成できるという観点からニューラル機械翻訳が注目されている。ニューラル機械翻訳 (Neural Machine Translation, NMT) とは、人間の脳神経回路が情報伝達を行う仕組みを模倣した人工的なニューラルネットワークが情報を収集して自ら学習し、単語の意味として正しい可能性の高い訳語を定める手法である。従来のルールベース機械翻訳 (Rule Based Machine Translation, RMT) [1]や統計的機械翻訳 (Statistical Base Machine Translation, SMT) [2][3]では、単語や句の関係に着目して翻訳モデルを作成する。一方で、ニューラル機械翻訳は、文章全体を一つの塊として捉えて翻訳モデルを作成する。そのため、自然な語順を生成し、流暢な訳文の生成が可能となる。

ニューラル機械翻訳のモデルでは、より高い翻訳精度を得るために数十万から数百万文の大規模な対訳コーパスを用いて学習し、翻訳が行われている。例えば、英独翻訳では、機械翻訳に関する国際会議 (Conference on Machine Translation, WMT) により作成された 450 万文の対訳コーパ

<sup>†1</sup> 奈良女子大学  
Nara Women's University  
<sup>†2</sup> 滋賀大学  
Shiga University

<sup>†3</sup> スタンフォード大学  
Stanford University

スが使用されている。しかし、近代文語体で書かれた文書のほとんどは、画像データとしてアーカイブされたものである。そのため、OCRによりテキストデータ化がなされているものの、不完全なものが多く、大量に対訳データを得るのは困難である。そこで、これまでに我々の研究室では、対訳データの取得が比較的容易な文学作品に着目し、近代文語体と現代口語体の自動翻訳を検討している[4]。しかし、1つの分野のみに着目していたため、議会書など文体の異なる他分野の文書に対する汎用性が低いという課題がある。そこで、本稿では、文体の異なる多様な分野を有する新聞に着目する。

邦字新聞デジタルコレクション[5]では、アメリカ大陸、アジアにおける海外在住の日本人や日系人が発行した海外日系新聞を集めた邦字新聞を、世界最大規模で提供している。提供されている邦字新聞は、明治から昭和初期にかけて刊行されたものであるため、近代文語体の特徴を有する。したがって、邦字新聞を対象とした近代文語体と現代口語体の自動翻訳を検討する。また、翻訳手法には、異言語間で用いられているニューラル機械翻訳の手法を用いて最適なモデルを検討する。その中でも、代表的な Sequence to Sequence モデル[6]、Convolutional Sequence to Sequence モデル[7]、Transformer モデル[8]の3つのモデルを検討し、最適な翻訳モデルの探索を行う。さらに、データ数と文長が翻訳精度に及ぼす影響について考察する。

以下、2章では、使用するモデルの詳細を紹介する。3章では、実験に用いた対訳コーパスの説明、取得方法、前処理について述べ、実験結果を述べる。

## 2. ニューラル機械翻訳

従来の統計的機械翻訳[2][3]では、フレーズテーブルや言語モデル等の独立した翻訳知識、翻訳時に考慮する特徴量の設計が必要であったが、ニューラル機械翻訳は、1つのニューラルネットワーク内で訓練と翻訳を行うことが可能である。そのため、異言語間翻訳において近年注目されている手法である。NMTは大きく分けて、エンコーダ、デコーダ、Attention 機構[9]の3つのパーツで構成されている。

エンコーダとは、入力文を符号化するものである。デコーダとは、エンコードされた情報を復号化し、所望の出力を得るものである。また、Attention 機構とは、長文対応を可能とするため、ある単語が出力文のどの単語と関連性が高いかを学習する機構である。特に、エンコーダとデコーダの2つのパーツから構成されるモデルをエンコーダデコーダモデルと呼び、文書要約、テキスト分類、機械翻訳など様々なタスクで応用されている。

以下、2.1節ではエンコーダデコーダモデルの代表的なモデルである再帰型ニューラルネットワーク (Recurrent Neural Network, RNN) [10]を用いた Sequence to Sequence モ

デルについて述べる。次に、2.2節で Attention 機構について、2.3節で Attention 機構と畳み込みニューラルネットワーク (Convolutional Neural Network, CNN) [11]を用いた Convolutional Sequence to Sequence モデルについて、2.4節で Attention 機構のみに着目した Transformer モデルについて説明する。

### 2.1 Sequence to Sequence モデル

Sequence to Sequence (Seq2Seq) モデル[6]は、2つのRNNを接続した構造のエンコーダデコーダモデルである。モデルの概略図を図1に示す。Seq2Seqモデルは、エンコーダに入力文が入力されると、文末記号EOS (End Of Sentence) タグが生起するまでデコーダで出力文を予測、生成する。図1に示す各ボックスはRNNセルを表す。本稿では、GRU (Gated Recurrent Unit) [12]を用いる。Sutskeverの論文[6]では、RNNセルにLSTM (Long short-term memory) [13]を採用することで、可変長な入力文を固定次元のベクトル表現にマッピングして学習することを実現している。LSTMは長期依存の特徴の学習を可能にしたセルであるが、計算コストが大きいという問題点がある。GRUセルは、LSTMと同様の性能を持ち、計算コストが小さく、より高速に学習を進めることを可能としたRNNの一種である。

### 2.2 Attention 機構

Attention 機構[9]は、デコード時に入力系列の情報を直接参照できるようにすることで、どこを注目すべきかを学習する仕組みである。Attention 機構の構造を図2に示す。2.1節で説明したSeq2Seqモデルでは、最後の隠れ状態のベクトルのみをデコーダに渡して参照している。そのため、入力系列の長さが大きい場合には、十分な情報が得られず、翻訳精度が低いという課題がある。この課題を克服したのが、Attention 機構である。Attention 機構は、エンコーダの各ステップの隠れ状態のベクトル( $h_1 \dots h_T$ )を記憶し、入力系列の  $t$  番目の単語が出力系列のどの単語と対応するかについての情報( $a_{t,1} \dots a_{t,T}$ )を算出する。この情報をもとに、各タイムステップごとにコンテキストベクトル $C_t$ を算出する。 $C_t$ は、次のデコーダの内部状態を推定する材料の1つとなり、現タイムステップ  $t$  の隠れ状態 $S_t$ を算出し、出力系列( $y_1 \dots y_T$ )を生成する。以上のAttention 機構を用いることで、入力系列の中のキーワードが距離に関わらず、情報を保持することができるため、長文への対応が可能となる。

### 2.3 Convolutional Sequence to Sequence モデル

Convolutional Sequence to Sequence (ConvSeq2Seq) モデル[7]は、単語列の処理をLSTMからCNNに置換することで、学習中にすべての要素の計算を完全に並列化することを可能としたモデルである。GPUハードウェアの有効活用により、非線形性の数が固定されて入力長に依存しないため、

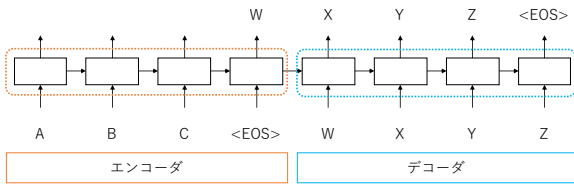


図 1 Seq2Seq モデルの概略図

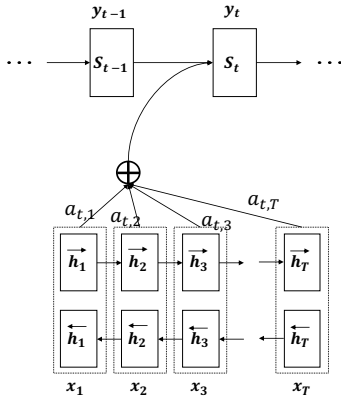


図 2 Attention 機構の構造図

最適化を容易にしている。モデルの概略図を図 3 に示す。ConvSeq2Seq モデルは、エンコーダとデコーダがそれぞれ GLU (Gated Linear Unit) [14]層と CNN 層により構成されている。GLU は、LSTM や GRU と類似の構造をもち、式 1 で表現されるゲート機構の役割をする。

$$f(x) \otimes \sigma(y) \quad (1)$$

ゲート機構は、 $\sigma(y)$ の出力により $x$ を出力するか判断できる仕組みのことである。

2.1 節で述べた LSTM や GRU では、入出力にシグモイドによるゲート機構を設けることで、RNN の勾配消滅問題に対処している。一方で、CNN を用いた場合、モデルの深さは時系列の長さではなく、畳み込みのスタック数に依存するため、RNN ほど勾配消滅は顕著ではない。そのため、GLU は勾配消滅問題の対策ではなく、Attention の目的で導入されている。

まず、エンコーダにより、入力系列が埋め込まれる。ここで、ConvSeq2Seq モデルでは、単語列の語順を絶対的な位置の情報としてモデルに与えるため、位置埋め込み (Position Embedding) がなされる。位置埋め込みは、単語列の埋め込み行列  $W = (w_1, \dots, w_T)$  に位置ごとに異なる変数  $P = (p_1, \dots, p_T)$  を加算し、 $E = (w_1 + p_1, \dots, w_T + p_T)$  を算出する。次に、CNN 層に移り、長期依存が捉えられ、GLU 層に移る。GLU 層では、CNN 層で捉えた情報が次の層へ送る情報かを判断し制御する。ConvSeq2Seq モデルでは、デコーダのすべての層で Attention を個別に計算する。従来の

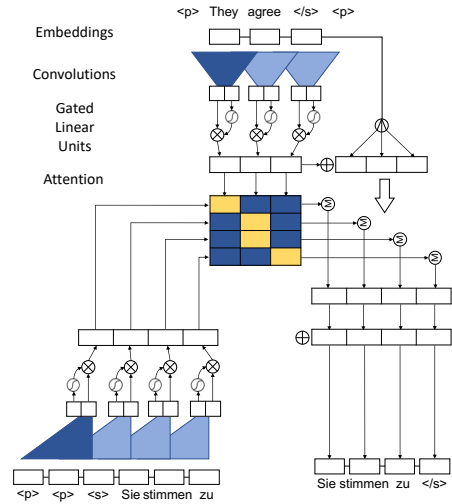


図 3 ConvSeq2Seq モデルの概略図

NMT では、デコーダの 1 系列の生成に 1 度のみ Attention を計算していたが、ConvSeq2Seq モデルでは、層ごとに Attention を再計算する複数ステップの Attention を使用する。これは、2.4 節で述べる Transformer モデルでも採用される。複数ステップの Attention により、ある層のある時刻  $t$  の Attention は、下層や過去の Attention の履歴を参照することが可能となる。

## 2.4 Transformer モデル

Transformer モデル[8]は、RNN や CNN を使用せず、Attention 機構のみを使用したモデルである。モデルの概略図を図 4 に示す。図 4 の左側がエンコーダ、右側がデコーダであり、Multi-Head Attention 層と Feed Forward 層により構成されている。まず、エンコーダで入力系列のトークンの埋め込み行列に対応する位置エンコーディングが要素ごとに加算され、エンコーダおよびデコーダに入力する。位置エンコーディングの行列  $PE$  の各成分は次式によって表される。トークンの位置を  $pos$ 、埋め込み行列の次元数を  $d_{model}$  とする。この場合、位置エンコーディングの  $i$  番目の次元の値  $PE_{(pos,i)}$  は式 2 より得られる。

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (2)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right)$$

位置エンコーディングは位置のみに依存した値であり、トークンの絶対位置を表す。次に、Attention の算出を行う。Multi-head Attention は、Attention を head と呼び、複数の Attention のことを表している。Transformer モデルでは、

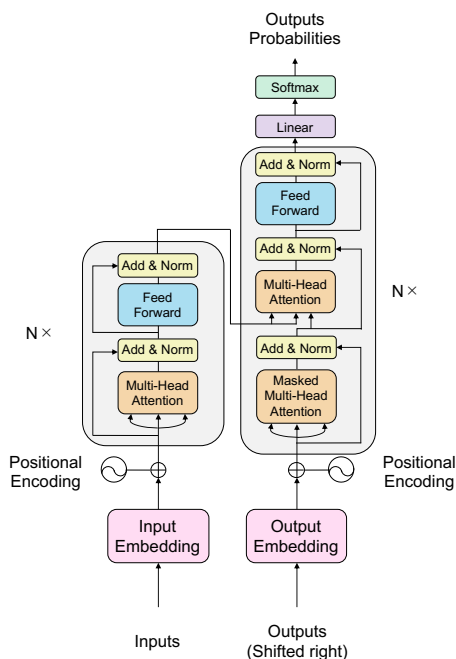


図 4 Transformer モデルの概略図

Attention を辞書オブジェクトとして扱う。目的言語を Query とみなし、原言語を Key と Value に分離して、従来の Attention を一般化している。概念的には、Key は Query の索引となり、Key をもとにメモリ内の対応する Value を取得する。ここで、Key と Value はそれぞれが 1 対 1 に対応する Key-Value のペア配列となるため、辞書オブジェクトとして扱うことが可能となる。したがって、Query と Key から関連度を表す Attention Weight を算出する。具体的には、まず Dot-Product Attention と呼ばれる Query と Key の内積をとることで Attention Weight を算出する。次に、算出した内積を Softmax 関数にかけ、Attention Weight の正規化を行う。正規化した Attention Weight は、各 Query が参照すべきメモリの情報を表す。最後に、Attention Weight にしたがって Value を取得し、出力を決定する。

### 3. 実験

本稿では、邦字新聞における近代文語体と現代口語体の自動翻訳に向けて、3 つのモデルを使用した実験を行い、モデル間の比較と翻訳を検討する。使用したモデルは、GRU を用いた Seq2Seq モデル、CNN を用いた ConvSeq2Seq モデル、Attention のみを用いた Transformer モデルの 3 つである。評価指標には、BLEU (BiLingual Evaluation Understudy) [15] を用いる。

3.1 節では、実験設定について述べる。3.1.1 項では、本稿で用いた実験データの詳細、3.1.2 項では実験環境について説明する。3.2 節では、各種モデルを比較検討するためにを行った実験の結果を示す。

### 3.1 実験設定

#### 3.1.1 実験データ

本稿では、多様な分野を有する近代の活字媒体として邦字新聞を用いる。具体的には、1908 年から 1941 年までの「日米新聞」、「日布時事」、「日本人」、「新世界」の 4 種類からランダムにサンプリングした 18 紙を対象とする。

上記の近代文語体で書かれた文書に対する現代口語体の対訳データは、邦字新聞デジタルコレクションには存在しない。そのため、本学文学部の学部生により翻訳された文を対訳データとして用いる。文学作品にのみ着目した手法[4]では、対訳データに意識が多く、正確な翻訳が困難であることが示されている。そのため、本稿では対訳データを直訳で作成することとしている。原言語文としては、邦字新聞の原文をテキスト化して用いる。邦字新聞の原文には多くの旧字体が含まれ、現在存在しない字体も存在する。その中でも、デジタルフォントとして存在する字体のみ旧字体を用いてテキスト化を行う。一方で、デジタルフォントとして存在しない字体は、現在使用している字体に変換する。また、仮名文字についても同様に平仮名に変換する。

ニューラル機械翻訳で使用するデータは、英語やドイツ語の文章のように単語間に余白が存在するため、トークン化の必要がないことが多い。一方で、日本語の文章は、文中に余白がなく、単語の境界の判別が困難である。そのため、実験用データに対して形態素解析によるトークン化が必要となる。このトークン化には、形態素解析エンジンである MeCab[16]を適用する。近代文語体のデータには、近代文語 UniDic、現代口語体のデータには現代書き言葉 UniDic の辞書を MeCab に用いる[17]。

実験に用いた対訳データ数は、全 8,467 文である。全 8,467 文のうち、訓練用 1,914 文、テスト用 239 文、検証用 240 文としたものをデータ 1 とする。全 8,467 文のうち、訓練用 6,770 文、テスト用 840 文、検証用 857 文としたものをデータ 2 とする。

#### 3.1.2 実験環境

本実験では、実行 OS に Ubuntu18.04LTS、GPU は NVIDIA GeForce RTX2080Ti を 1 枚用いる。

Seq2Seq モデルでは、エンコーダ、デコーダにそれぞれ 2 層の GRU を使用し、Seq2Seq モデルの各 GRU の隠れ状態は 512 次元とする。埋め込み層は 256 次元、dropout の確率は 0.25 とする。最適化手法には初期学習率を 0.0001 とした RMSprop を用いる。バッチサイズは 4 とする。

ConvSeq2Seq モデルでは、エンコーダ、デコーダにそれぞれ 2 層の CNN を使用し、ConvSeq2Seq モデルの各 CNN の隠れ状態は 512 次元とする。埋め込み層は 256 次元、dropout の確率は 0.25 とする。最適化手法には初期学習率を 0.0001 とした RMSprop を用いる。バッチサイズは 32 とする。

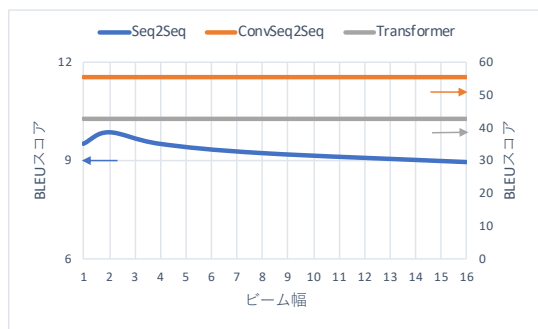


図 5 ビーム幅に対する各種モデルの BLEU スコア

Transformer モデルでは、エンコーダ、デコーダは共に 2 次元、並列数は 1、dropout の確率は 0.25 とする。最適化手法には初期学習率を 0.0001 とした RMSprop を用いる。バッチサイズは 32 とする。

### 3.2 各種モデルの検討

#### 3.2.1 ビームサーチの最適化

各種モデルに対し、ビーム幅が及ぼす影響についてデータ 1 を用いて検討する。ビーム幅を 1 から 16 とした場合の翻訳精度を BLEU スコアで比較する。その結果を図 5 に示す。図 5 より、ConvSeq2Seq モデル、Transformer モデルでは、ビーム幅による翻訳精度の違いは見られない。一方で、Seq2Seq モデルでは、ビーム幅の増加に伴い BLEU スコアが向上し、ビーム幅 2 のとき最大で 9.86 を示している。以上の結果より、本稿で検討した 3 つのモデルの中では、Seq2Seq モデルにおいてビームサーチによる影響が強いことがわかる。

#### 3.2.2 各種モデルの BLEU スコアの比較

3.2.1 項で検討した結果に基づき、各種モデルに対して最適な BLEU スコア及び処理速度について比較する。Seq2Seq モデルのビーム幅は 2 とし、ConvSeq2Seq モデル、Transformer モデルのビーム幅は 1 とする。その結果を表 1 に示す。表 1 では、各種モデルでバッチサイズが異なるため、バッチサイズを 32 とした場合の 1 エポックあたりの処理速度を比較する。表 1 より、Seq2Seq モデルは BLEU スコアが 9.86 と低いのにに対し、ConvSeq2Seq モデルでは BLEU スコアが 55.54 と最大であることがわかる。また、処理速度の観点からも、ConvSeq2Seq モデルが最速である。Seq2Seq モデルと比較し、ConvSeq2Seq モデルや Transformer モデルの BLEU スコアが高い理由としては、長文翻訳に対応した Attention 機構の有無が寄与していると考えられる。また、ConvSeq2Seq モデルと Transformer モデルの BLEU スコアに差異が見られた点について、文長とデータ数の影響の観点から考察する。

表 1 各種モデルの BLEU スコアと処理速度

モデル	BLEU	処理速度
		s/epoch/batch=32
Seq2Seq	9.86	3.79
ConvSeq2Seq	55.54	1.98
Transformer	42.92	2.23

表 2 データ数に対する各種モデルの BLEU スコア

モデル	BLEU		変量
	データ1	データ2	
Seq2Seq	9.86	7.28	-2.58
ConvSeq2Seq	55.54	60.25	4.71
Transformer	42.92	55.82	12.9

#### 3.2.3 データ数の違いによる BLEU スコアの比較

ニューラル機械翻訳では、データ数の増加に伴い、翻訳精度が向上することが知られている。本項では、翻訳精度向上を目的とし、各種モデルに対するデータ数の影響を検討する。本実験は、データ 1 とデータ 2 を用いて行う。その結果を表 2 に示す。表 2 より、データ 2 の場合においても、ConvSeq2Seq モデルの BLEU スコアが最大で 60.25 である。また、BLEU スコアの増加量の観点から見ると、他の 2 種と比較し、Transformer モデルが最大で 12.9 ポイント増加していることがわかる。これは ConvSeq2Seq モデルの約 2.7 倍の増加量である。以上の結果より、ConvSeq2Seq モデルと比較し Transformer モデルの翻訳精度は、データ数に大きく依存することが示唆される。

#### 3.2.4 文長の違いによる BLEU スコアの比較

本項では、3.2.3 節で検討したデータ 2 の各種モデルを用いて、各種モデルに文長が及ぼす影響について検討する。test データを原言語文の文長に基づいて分割し、算出した BLEU スコアを図 6 に示す。図 6 より、Seq2Seq モデル、Transformer モデルでは文長が長くなるにつれて翻訳精度が減少することがわかる。一方で、ConvSeq2Seq モデルでは、文長が 11 から 70 の範囲において一定の翻訳精度が得られている。以上の結果より、一定の文長を有する小規模データに対しては、ConvSeq2Seq モデルが適していることが示唆される。

#### 3.2.5 各種モデルが予測した翻訳文の比較

本項では、入力された原言語文に基づき、各種モデルが予測した翻訳文と参照訳文を比較する。本実験では、3.2.3 節で検討したデータ 2 の各種モデルを用いる。各種モデルが予測した実際の翻訳例を表 3 に示す。表 3 より、BLEU スコアが低かった Seq2Seq モデルは、文頭や文末部分は正確な出力がなされているが、文中部分の出力の精度は不十分である。一方で、ConvSeq2Seq モデル及び Transformer モ

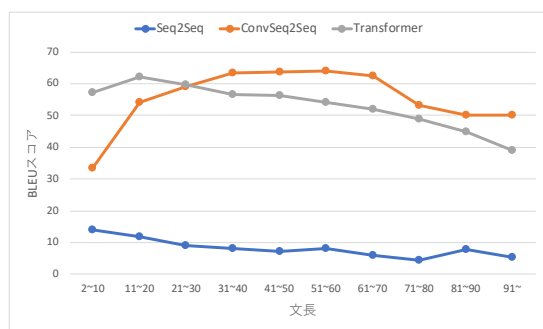


図 6 文長に対する各種モデルの BLEU スコア

表 3 実際の翻訳例

入力	目下英政府は之が歓迎の方法を講じつゝあり。
正解	現在イギリス政府はこの歓迎の方法を講じつつある。
Seq2Seq	現在政府ははがががののををがある。
ConvSeq2Seq	現在、英政府はこれが歓迎の方法を講じている。
Transformer	目下英政府はこれが歓迎の方法を講じつつある。

モデルの翻訳結果は、文頭や文末部分だけでなく、文中部分も翻訳がなされており、意味を理解できるレベルであるといえる。以上の結果より、各種モデルの評価指標として用いた BLEU スコアと実翻訳結果には妥当性があることが示唆される。

#### 4. おわりに

本稿では、Seq2Seq モデル、ConvSeq2Seq モデル、Transformer モデルを用いて、邦字新聞における近代文語体と現代口語体の自動翻訳を検討している。さらに、データ数や文長が各種モデルの翻訳精度に与える影響について考察している。各種モデル比較検討の結果、ConvSeq2Seq モデルにおいて、BLEU スコア、処理速度の両面で良好な結果が得られている。その上、データ数の増加に伴い BLEU スコアは向上し、最大で 60.25 を示している。また、Transformer モデルではデータ数の増加に伴う BLEU スコアの増加量が、ConvSeq2Seq モデルの約 2.7 倍を示し、データ数に大きく依存するモデルであることが示唆される。最後に、文長に関して比較検討を行った結果、ConvSeq2Seq モデルは広範囲において良好な BLEU スコアが得られることがわかる。したがって、一定の文長を有する小規模データを用いた本翻訳に関しては、ConvSeq2Seq モデルが適していると考えられるが、更なるデータ数の増加により Transformer モデルの有効性も期待できる。今後は、更なる翻訳精度の向上にむけてデータ数を増加し、各種モデルに対する検討を行う。

**謝辞** 本稿の対訳データの作成に携わって頂いた本学文学部の学部生の皆様に、謹んで感謝の意を表します。本稿は、MEXT 科研費 JP20H04483 の助成を受けたものです。

#### 参考文献

- [1] Sholom M. Weiss, Nitin Indurkha : Rule based Machine Learning Methods for Functional Prediction. In Journal of Artificial Intelligence Research 3, 1995, pp.383-403.
- [2] Peter F. Brown, Stephen A. Della Pietra, Vincent J. Della Pietra, Robert L. Mercer : The Mathematics of Statistical Machine Translation: Parameter Estimation, In Computational Linguistics, Vol. 19, Number 2, June 1993, pp.263-311.
- [3] Philipp Koehn, Franz J. Och, Daniel Marcu : Statistical phrase-based translation. Statistical phrase-based translation. In Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics, Vol.1, pp.127-133, 2003.
- [4] 林英里香, 竹本有紀, 石川由羽, 高田雅美, 城和貴 : 近代文語体と現代口語体の自動翻訳への試み, 情報処理学会研究報告. 数理モデル化と問題解決 (MPS), Vol.2018-MPS-121, No.18, pp.1-6
- [5] 邦字新聞デジタルコレクション (online) <https://hojishinbun.hoover.org/?l=ja>, (参照 2020.11.17)
- [6] Sutskever, Ilya, Vinyals, and Le, Quoc V. : Sequence to Sequence Learning with Neural Networks, In Advances in Neural Information Processing Systems (NIPS 2014), 2014, pp.3104-3112.
- [7] Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, Yann N. Dauphin. : Convolutional Sequence to Sequence Learning, In International Conference on Machine Learning (ICML), 2017.
- [8] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin : Attention is all you need, In Advances in Neural Information Processing Systems, arXiv, 2017.
- [9] Dzmitry Bahdanau, KyungHyun Cho, Yoshua Bengio : NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE, arXiv, 2015.
- [10] Learning out of time series with an extended recurrent neural network, In 1996 IEEE Workshop on Neural Network Workshop for Signal Processing VI, 1996, pp.170-179.
- [11] Tejas D. Kulkarni, Will Whitney, Pushmeet Kohli, Joshua B. Tenenbaum : Deep Convolutional Inverse Graphics Network, In CoRR, 2015.
- [12] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio : Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp.1724-1734.
- [13] Felix A. Gers, Jürgen Schmidhuber, Fred Cummins : Learning to Forget: Continual Prediction with LSTM, In Neural computation, 12.10 (2000), pp.2451-2471
- [14] Yann N. Dauphin, Angela Fan, Michael Auli, David Grangier : Language Modeling with Gated Convolutional Networks, arXiv, 2016
- [15] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu : BLEU: a Method for Automatic Evaluation of Machine Translation, In ACL, 2002.
- [16] 工藤拓 : MeCab Yet Another Part-of-speech and Morphological Analyzer, (online) <https://sourceforge.net/projects/mecab/>, (参照 2020.11.17)
- [17] Unidic : (online) <https://unidic.ninjal.ac.jp/>, (参照 2020.11.17)