文内コンテキストを利用した分割統治ニューラル機械翻訳

Divide-and-Conquer Neural Machine Translation Using Intra-Sentence Context

奈良先端科学技術大学院大学
石川　隆太

2024 年奈良先端科学技術大学院大学博士前期課程修了。在学中機械翻訳の研究に従事。

奈良先端科学技術大学院大学
加納　保昌

2022 年奈良先端科学技術大学院大学博士前期課程修了。2024 年同博士後期課程修了。博士（工学）。在学中機械翻訳の研究に従事。

奈良女子大学　教授／奈良先端科学技術大学院大学
須藤　克仁

2002 年京都大学大学院情報学研究科修士課程修了。同年 NTT 入社。2015 年京都大学博士（情報学）。2017 年より奈良先端科学技術大学院大学准教授、2024 年より奈良女子大学教授。

香港中文大学（深圳）　教授／奈良先端科学技術大学院大学　教授
中村　哲

1981 年京都工芸繊維大学工芸学部電子工学科卒業。シャープ、ATR、NICT を経て 2011 年より奈良先端科学技術大学院大学教授。2023 年より香港中文大学（深圳）教授を兼任。

① はじめに[1]

ニューラル機械翻訳 (Neural Machine Translation、NMT) は従来の統計的機械翻訳 (Statistical Machine Translation、SMT) を短期間で凌駕し、現在の機械翻訳技術の主流となった。SMT では、句単位の翻訳仮説を組み合わせながら複数の独立したモデルを統合して探索を行っていてしばしば流暢性に難があったのに対し、NMT は単一の大規模ニューラルネットワークを用いて、入出力の文内コンテキストを柔軟に参照しながら入力文を出力文に直接変換することで、自然で流暢な翻訳を実現している。一方で、NMT にも様々な課題があることが議論されており (Koehn and Knowles 2017)、その一つとして長文の入力に対する訳抜けや誤訳が発生が挙げられている。注意機構 (Bahdanau et al. 2014; Luong et al. 2015) はその軽減に一定の役割を果たしたが、非常に長い文や複雑な構造を持つ文に対しては十分でなく、NMT のためのモデルの主流である Transformer (Vaswani et al. 2017) でもこの問題は十分に解決できていないことが報告されている (Neishi and Yoshinaga 2019)。

長文の翻訳精度低下の課題に対して、統計的機械翻訳では、長文を統語構造上の節に分割して翻訳し、並べ替えて結合する分割統治的手法が提案されている (Sudoh et al. 2010)。また、Kano (2022) は NMT における長文のための分割統治的手法を提案している。Kano (2022) の手法による翻訳プロセスは大きく (1) 節単位の翻訳と (2) 節単位の翻訳結果の連結と書き換えの二つに分かれており、各プロセスはそれぞれ別の seq2seq モデルで実施される。しかしその性能向上は限定的であ

1　本稿は「自然言語処理」32 巻 1 号に掲載された著者らの同名の論文 https://doi.org/10.5715/jnlp.32.114 を CC BY 4.0 https://creativecommons.org/licenses/by/4.0/ に基づき再構成したものである。

り、その理由として (1) 節分割の方法、(2) 節分割後の節の翻訳精度の二つの課題が挙げられている。(1) は、(Sudoh et al. 2010) によるプレースホルダを使った階層的な分割ではなくフラットな構造のまま分割を行ったことで埋め込み従属節を含む節が分断されてしまうことに起因している。また (2) は、(1) による節の分断に加え、NMT の利点である文内コンテキストの参照が節単位の翻訳で活用できないことが理由に挙げられる。

本研究では、この二つの課題に注目し、英日翻訳における長文の翻訳精度の向上を試みる。具体的には、以下の二つを提案する。

（1）等位接続詞の前後に限定した節分割

（2）文内コンテキストを参照可能な節単位翻訳

これらにより、以下の効果が期待できる。

（1）埋め込み従属節による節の分断を避け、節としてほ完全な状態での翻訳

（2）各節を翻訳する際に、同一文内の他の節の情報を参照できることによる節翻訳の精度向上

提案手法は、多言語 BART モデル (Liu et al. 2020) を利用し ASPEC を対象にした英日機械翻訳実験において、41 単語以上の長い入力文に対してベースラインを上回る BLEU を達成した。また、提案手法が適用された文のみに限れば、より短い入力文に対してもベースラインを上回る翻訳精度が得られることが確認された。

② 関連研究

2.1 ニューラル機械翻訳における長文翻訳の課題

Long Short-Term Memory (LSTM) を利用した sequence-to-sequence の初期のモデル (Sutskever et al. 2014) は、RNN の回帰構造により文脈情報を一方向で蓄積するため、テキスト内の長期的な依存関係を捉えるのに限界があり、文が長くなるにつれて情報の保持が難しくなり翻訳精度が低下する (Bahdanau et al. 2014)。Cho et al. (2014) や Bahdanau et al. (2014) は RNN に注意機構を導入することで、局所的な入力の情報を柔軟に参照可能にし、注意機構が機械翻訳の性能向上に効果的であることを示した。さらに、Vaswani et al. (2017) は自己注意機構の導入によりこうした RNN の制約を排し、入力・出力それぞれの柔軟な活用を可能にした。しかしながら、長い入力文に対しては訳抜けが起こりやすく、実用上の大きな問題になっている。

また、NMT の実装においてはしばしば長文が学習コーパスから除外されることがある上、長文の対訳は対訳コーパス中では一般に量が少ないことから、学習で活用されづらい、また評価においても長文に対する性能が考慮されづらい面があったとも言える。本研究は、論文等の技術文書において頻出する長文の翻訳を主眼に、翻訳精度低下の問題の解消を図るものである。

2.2 分割統治的手法による長文の機械翻訳

統計的機械翻訳において、Sudoh et al. (2010) は長文を統語解析の結果に基づいて節単位に分割し、各節を個別に翻訳した上で埋め込み節のプレースホルダに基づき目的言語の語順に再構成する手法を提案している。この方法により、複雑な節単位の順序入れ替えを、プレースホルダの並べ替えに置き換え簡略化し、長文の翻訳精度を向上させた。

Pouget-Abadie et al. (2014) らは、入力文を前から順に分割する単純な分割統治型機械翻訳のための入力文の自動分割手法を提案した。この手法は原言語から目的言語への順方向の翻訳スコアとその逆方向の翻訳スコアを利用してセグメント対の対訳確信度を計算して適切なセグメント分割を求め、そこから入力文の自動分割モデルを学習するものである。英仏翻訳タスクでの実験において、特に長文に対する翻訳品質と未知語に対する頑健性の改善があることが確認されている。

この自動分割手法は句に基づく統計的機械翻訳における句と同様に統語構造上の要素の境界とは無関係に分割され、また前から順に分割・翻訳・連結を行うためセグメントの順序の入れ替えが不可能であるため、語順の違いの大きい言語対には適さない。

この手法に対応するために、Kano (2022) は Sudoh et al. (2010) の手法を参考にした、ニューラル機械翻訳（NMT）における分割統治的手法を提案した。Kano (2022) は、分割されたセグメントの翻訳を行う NMT モデルとは別に、セグメント単位の翻訳結果を文に再構成する NMT モデルを導入する手法を提案している。再構成のための NMT ではセグメント単位の翻訳結果を適宜並べ替えて自然に連結するための書き換えが行われる。これにより、Sudoh et al. (2010) が統計的

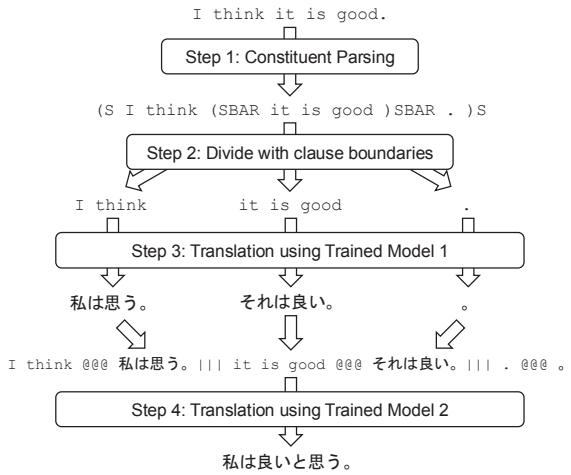

図 1　Kano (2022) の分割統治型ニューラル機械翻訳の手法の概略

機械翻訳で提案した分割統治的手法を NMT の手法として実現している。しかしながら通常の NMT からの制度改善は限定的で、その原因として節単位の翻訳の精度が低いことが挙げられており、本研究ではその解決を図る。

③ 分割統治型ニューラル機械翻訳

本セクションでは提案手法のベースとなる分割統治型 NMT について説明する。Kano (2022) の手法は図 1 に示す 4 つのステップで構成されている。

まず、Step1 で入力文を構文解析し、節 (ラベルが S または SBAR である統語要素) を同定する。続く Step2 では、各節の境界でセグメントに分割する。節が入れ子構造になっている場合もすべての節境界で分割を行うため、図中の例では I think / it is good / 。 のように実際の節よりも短い単位への分割が行われる。その後 Step3 で、各セグメントは文単位の対訳コーパスで学習された翻訳モデル (Trained Model 1) を用いて翻訳され、最後の Step4 で、セグメント単位の翻訳結果の統合と書き換えを行う翻訳モデル (Trained Model 2) を用いて、最終的な訳出を生成する。Step4 での入力は原言語側の各セグメントとその翻訳結果を特殊トークン" @@@" と" |||" を用いて結合したものになっている。これは、英日翻訳では Pouget-Abadie et al. (2014) のような単純なセグメント翻訳の連結ではなく語順の調整を行うこと、また文内文脈をこの時点で参照できるようにすることで訳語選択の調整を行うことを企図したものである。

④ 提案手法

前節で説明した Kano (2022) の手法は、セグメント単位の翻訳精度が不十分で、最終的に得られる文単位の翻訳精度の目立った向上が得られなかった。

その主な要因として以下の 3 つが考えられる。

（1）すべての節の境界で文を分割すると、文内の「I think」やピリオドのみでセグメントが構成されるような、過剰な分割が生じる。

（2）原言語の各セグメントを通常の文レベルの NMT モデルで翻訳する際、学習時と推論時の翻訳単位の不一致が生じ、過剰訳や訳抜けをもたらしやすい。

（3）原言語の各セグメントを独立して翻訳すると、文内のコンテキストが利用できないため、一貫性のない不適切な訳語選択が生じやすい。

そこで本研究では、これらの問題を解決する新しい分割統治型 NMT 手法を提案する。

4.1　等位接続詞を基準とした節単位の分割

本研究では、分割後のセグメントを統語上の節と合致させることで文として翻訳することができるよう、セグメントの過剰分割を抑制する。具体的には、統語上の節 (S) を連結する等位接続詞 (CC) に着目し、その前後でのみ分割を行う。

このセグメント分割のルールの詳細を以下の 2 つの文を例として示す。

(1) I like football and baseball.

(2) She wants to travel the world, but her job keeps her busy, and her savings are not enough for the trip.

これらを構文解析した結果が以下であるため。

(1)(S (S I like (NP (NN football) (CC and) (NN baseball)).))

(2) (S (S She wants to travel the world), (CC but) (S her job keeps her busy), (CC and) (S her savings are not enough for the trip).)

文 (1) には、2 つの名詞 (NN) を接続する and が含まれているが、これは上記のセグメント分割の条件を満たしていない。ここで分割を行うとすると、得られるセグ

メントは I like football と baseball. であり、前者は節と見なすことができるものの、後者は名詞と句点のみで構成されるセグメントであり、過剰分割であると考える。

文 (2) には、3 つの節を接続する but と and が含まれている。これらの接続詞の前後で分割を行うと、以下の通り 3 つの節と 2 つの接続詞の列が得られる

・(S She wants to travel the world,}
・(CC but)
・(S her job keeps her busy,)
・(CC and)
・(S her savings are not enough for the trip.)

分割された節は、図 1 の Step3 と同様に、NMT モデルを使用して翻訳することができる。そうすると等位接続詞は翻訳されずに残ることになるため、続く Step4 で節単位の翻訳結果を統合するための二段階目の翻訳を行う際に入力される。

4.2　節単位の対訳対応付け

4.1 の節単位の分割戦略に基づけば、文レベルの NMT を使用してセグメントを翻訳する際の翻訳単位の不整合の問題を軽減することが期待されるが、文の途中で分割を行う故に約物や文頭の大文字小文字の違いといった差異は残っている。その差異を解消するためには、本手法で実際に翻訳される翻訳単位での対訳データを用いて学習される、上述のように分割された節単位の入力に対応した NMT モデルを利用することが望ましい。そこで本研究では、節単位の入力に対応した NMT モデルの学習に利用する節単位の対訳対応付けのために、Prefix Alignment (Kano et al. 2022) で用いられた対訳接頭辞文字列組の抽出方法を参考に、4.1 の方法で分割した各原言語（英語）の節と目的言語（日本語）のセグメントとの間の対応付けを行う手法を提案する。この手法は以下の 5 つのステップから構成されており、処理の例を図 2 に示す。

Step1: 多言語 BART のトークナイザーを使用して、日本語文をトークン化する。そして、英語の節との対応付け候補として、日本語の最後のトークンから最初のトークンに向かって一つずつトークンを連結した接尾辞文字列の集合を作成する。

Step2: 4.1 の方法に基づいて英文を節と等位接続詞に分割し、各節を文単位の英日 NMT モデルを利用して翻訳したものを、擬似的な節単位の日本語訳とする。

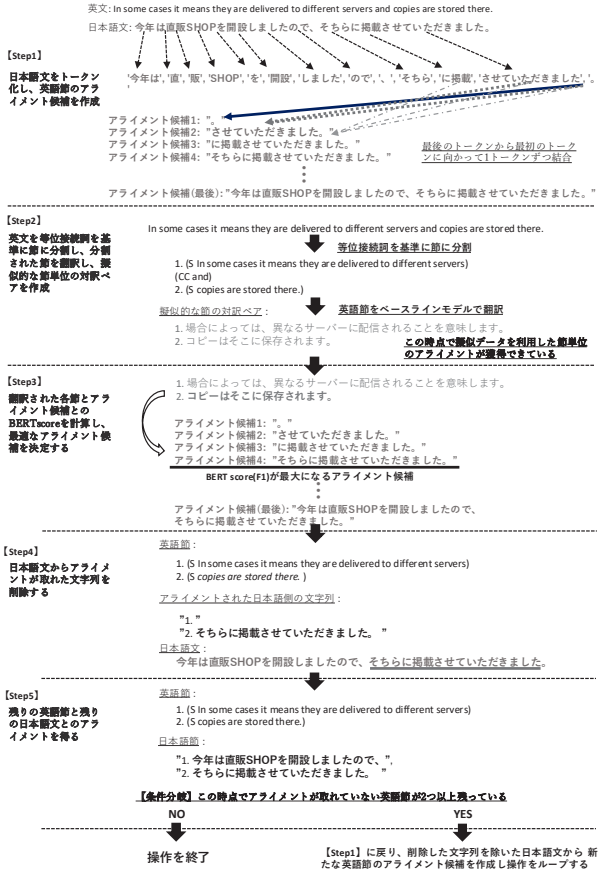

図 2　提案手法による英日の対訳文ペアから、分割された英語節と対応する日本語節のアライメントを得

Step3: この擬似的な日本語訳を、文末のものから順に Step1 で作成した対応付け候補と付き合わせる。その基準には BERTscore の F1 スコアを用い、最も高い F1 スコアを対応付け候補を英語側の節と対応する日本語訳とする。

Step4: Step3 で得られた対応付け候補の文字列を元の日本語文から削除する。

Step5: 英語側で対応付けされていない節が残り 1 つになっていれば、それを残りの日本語文字列と対応付けて処理を終了する。2 つ以上の英語の節が残っている場合は、残った日本語文字列を新しい日本語文として、Step1 に戻る。

図 2 の例では、最初に二番目の英語節 “copies are stored there.” に対する対応付けを求め、残りの英語節 “In some cases...” と残りの日本語文字列が対応付けられるものとする。Step4 においてすべての英語の節の対応付けが得られる前に日本語の文字列が尽きた場合は、対応付けが失敗したものとみなし、当該対訳文全部を節対応付けから除外するものとする。この手法を使

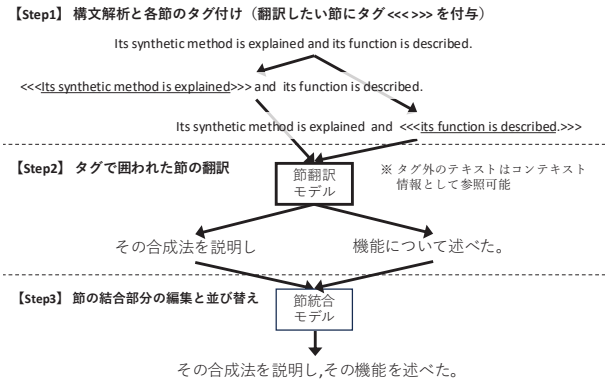

図３　提案手法による翻訳処理の概要

<<<En-clause1>>> CC1 En-clause2 CC2 En-clause3	Ja-clause1
~~En-clause1 CC1~~ <<<En-clause2>>> CC2 En-clause3	Ja-clause2
En-clause1 CC1 En-clause2 CC2 <<<En-clause3>>>	Ja-clause3

図４　英語側が３つの節で構成されていた場合の節単位対訳データの形式。翻訳対象となる節が先頭および末尾以外の場合は、その節に先行する文字列を削除する

用することで、Kano (2022) のような機械翻訳による擬似的な節単位対訳ではなく、元の対訳文に基づいた質の高い節単位対訳が得られることが期待できる。なお、この節対応付けは Prefix Alignment と同様に文頭から行うことも可能だが、予備実験において節対応付けの性能が低かったため文末から行うこととした。

4.3　文内コンテキストを利用した節単位翻訳

ここまでに示した方法により得られた節単位の英日対訳に基づき、節単位の翻訳の際に同じ文の残りの箇所を文内コンテキストとして参照する仕組みを導入する。図３にこの節単位の翻訳の概要を示す。図に示すように、節単位の翻訳はタグ <<< >>> で囲われた箇所を訳出するように学習用の対訳データを構築し、節翻訳モデルはそれを用いて学習する。これにより、節翻訳モデルはタグで囲われた箇所のみをその他の箇所を文内コンテキストとして参照しながら翻訳することを学習することが期待される。

表１に、図２で得られる対応付けに基づく節単位対訳データの例を示す。上で述べた英語の各節の対訳関係を学習させるため、4.2 で得られた節の対応付けに基づき、英語側の節をタグ <<< >>> で囲い、その節に対応する日本語の部分文字列を対訳として与える。ここで、一般に文は複数の節に分割されるため、翻訳対象となる（タグが付与される）節は文中の任意の位置に存在し得

表１　提案手法における節単位対訳データの例

原言語（英語）	目的言語（日本語）
<<<In some cases it means they are delivered to different servers>>> and copies are stored there.	今年は直販 SHOP を開設しましたので,
In some cases it means they are delivered to different servers and <<<copies are stored there.>>>	そちらに掲載させていただきました

るが、次節で述べるデータ拡張との整合のため、タグが付与される節は英語側の先頭もしくは末尾に配置するものとする。そのため、文が三つ以上の節に分割され、先頭および末尾以外にある節にタグ付けを行う場合には、その節に先行する文字列を削除する（図４）。これは学習時、推論時の双方で適用される。

4.4　モデルの学習

提案手法では Kano (2022) と同様に節翻訳モデルと節翻訳統合モデルの二つのモデルが必要となるため、それぞれ以下のように学習を行う。

4.4.1　節翻訳モデルの学習

節翻訳モデルは、4.3 で述べた文内コンテキストの参照に対応するための学習を行う必要がある。しかしながら、等位接続詞で節が連結されている事例は対訳コーパス全体の数割に過ぎず、学習事例の不足が懸念される。そのため、文脈情報が保持されている文単位対訳データを用い、図５に示すような形式で前後の文の情報をコンテキストとして参照可能な対訳データとして利用するデータ拡張を行う。これは文脈依存型機械翻訳において 2-to-1 と呼ばれるアプローチ (Tiedemann and Scherrer 2017) と類似するが、提案法では先行文、後続文のどちらを翻訳するかをタグを用いて表現したものである。なお、同様に３文以上を組み合わせて中間の節を翻訳させる事例を得ることもできるが、３つ以上の節が等位接続詞で連結される文が少数に限られること、またデータ拡張の組み合わせ数が膨大になることから、本研究では２文の組み合わせのみを用いることとした。節翻訳モデルの学習はまずこのデータ拡張によっ

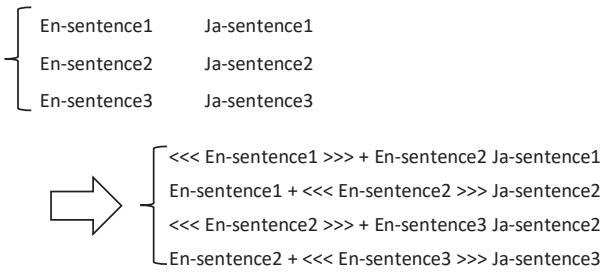

図５　文脈情報が保持されている文単位対訳データを用いた、文内コンテキスト依存節翻訳モデル学習のための データ拡張の模式図

て生成された先行文または後続文をコンテキストとして参照可能な文単位対訳データを用いて行い、続いて前述の文内コンテキストを参照可能な節単位対訳データを用いて追加学習する。

4.4.2　節翻訳統合モデルの学習

節翻訳統合モデルは、節翻訳の結果と、節翻訳において除外された等位接続詞を連結したものを入力とし、文単位の翻訳を出力するものである。その学習のためには対訳コーパスの原言語（英語）側をこの入力形式に変換したものが必要になるため、4.1 の方法で節に分割した英文の各節を文単位のベースライン英日 NMT に翻訳したものを元の英文の順序通りに等位接続詞を含めて連結したものを入力側とし、それを対訳コーパスの目的言語（日本語）文に翻訳するように学習したものを節翻訳統合モデルとする。

5　実験

提案方法の有効性を検証するため、以下の英日翻訳実験を行った。提案手法は長文の翻訳に焦点を当てたものであるため、その観点での検証を重視する。

5.1　実験設定

5.1.1　データセット

英日翻訳において比較的構造が複雑で長い文が多く、また 4.4.1 で述べたデータ拡張のための文脈情報が取得可能なデータセットとして、科学論文の要旨からなる対訳コーパスである ASPEC (Nakazawa et al. 2016) を用いた。ASPEC は学習データが 300 万文対、開発データが 1,790 文対、テストデータが 1,812 文対から構成されており、学習データは train1.txt、train2.txt、train3.txt に、それぞれ 100 万文ずつ分割されている。この学習データは対訳文対応付けの信頼度スコアに基づいてソートされているため、train2.txt、train3.txt 等後方のデータにはノイズの多い対訳文が含まれており、学習に悪影響を及ぼすとされている (Morishita et al. 2017)。そのため今回の実験では先頭の train1.txt のみを学習データに利用した。なお、予備実験では train2.txt の追加利用は BLEU の向上に寄与しなかった。

5.1.2　実装

対訳データ量が限られていることもあり、本実験では事前学習済みの多言語系列変換モデルである多言語 BART を利用し、実装には fairseq (Ott et al. 2019) を用いた。ベースラインの NMT モデルは多言語 BART を train1.txt を用いてファインチューニングしたものである。節翻訳モデルは、4.3 によりデータ拡張された文間コンテキストを有する対訳データと、4.2 による文内コンテキストを有する節単位対訳データを用いてファインチューニングした。ASPEC の学習データには文書 ID と文 ID が付与されているため、その情報をもとに文章内で連続する 2 文を用いて 4.4.1 のデータ拡張により約 20 万対のデータが得られた。ファインチューニングのハイパーパラメータ設定は、概ね英語 - ルーマニア語の多言語 BART ファインチューニングのデフォルト設定に従った[2]¥footnote。ただし、データ拡張により長文の占める割合が増加することを鑑み、ミニバッチ学習における 1 回の勾配計算に利用する対訳文を多くすることが良い影響を与える可能性を考慮するため、fairseq において勾配計算に利用する事例数を調整する update frequency を 2（デフォルト）とその 5 倍の 10 の 2 パターンでファインチューニングし、開発データでより良い BLEU スコアを達成した設定を選択した。

また、節同定のための統語解析には Stanza[3] の constituency 解析を用いた。

5.1.3　評価

評価指標には BLEU(Papineni et al. 2002)、BLEURT

2　https://github.com/facebookresearch/fairseq/blob/main/examples/mbart/README.md

3　https://github.com/stanfordnlp/stanza/

表２　単語長ごとのテストデータの文数

単語数	1-20	21-40	41-60	61-	すべて
全文	932	784	89	7	1812
分割された文	54	179	24	2	259

表３　入力単語数ごとの BLEU (テストセット全体)

単語数	1-20	21-40	41-60	61-	すべて
ベースライン	40.3	**42.1**	40.0	46.5	41.3
提案手法	40.2	**42.1**	**40.7**	**47.8**	**41.4**
単純連結	**40.4**	**42.1**	40.2	47.5	41.3
文数	932	784	89	7	1812

(Sellam et al. 2020)、COMET (Rei et al. 2022) を用いた。BLEU は単語 n-gram の表層的一致に基づく評価指標であるのに対し、BLEURT と COMET は文埋め込みに基づく回帰モデルを用いた文意の合致を評価することを志向した評価指標である。

BLEU の計算には sacrebleu (Post 2018) を用い、その際の日本語トークナイザには ja-mecab (Kudo 2006) を用いた。文の長さが翻訳に影響を調査するため、テストセットを入力側英文の単語数に基づき 1-20、21-40、41-60、61 以上の４つの長さベースのサブセットに分割した評価も行った。

5.1.4　比較手法

本実験では、提案手法とベースラインに加え、提案手法における節翻訳統合モデルの効果を検証するために、節翻訳結果を節翻訳統合モデルに通さず単純に連結した (以下、単純連結) しただけの手法も比較に加えた。なお、提案手法は、等位接続詞により節が連結されている場合にのみ適用されるため、その他の場合はベースラインの翻訳モデルを利用することになる。表２に、テストセット全体および提案手法が適用される文における、各サブセット毎の文数を示す。

5.2　結果

表３に示されたテストセット全体に対する BLEU の結果を示す。提案手法が適用された文は全テストデータ 1,812 文のうち約 14% の 259 文に留まっていることもあり、テストセット全体ではベースラインを BLEU で 0.1 ポイント上回るに留まったが、単語数 41-60、61 以上の範囲では、それぞれ 0.7 ポイント、1.3 ポイント改善が見られた。

表４　入力単語数ごとの BLEU (提案手法が適応された文のみ)

単語数	1-20	21-40	41-60	61-	すべて
ベースライン	49.7	44.8	40.1	31.4	44.7
提案手法	48.8	**44.9**	**42.5**	32.5	**45.0**
単純連結	**51.0**	44.6	40.8	**34.2**	44.8
文数	54	179	24	2	259

表５　入力単語数ごとの BLEURT (提案手法が適応された文のみ)

単語数	1-20	21-40	41-60	61-	すべて
ベースライン	0.792	0.729	0.726	**0.653**	0.742
提案手法	**0.797**	**0.738**	**0.732**	0.618	**0.749**
単純連結	0.796	0.734	0.720	0.617	0.745
文数	54	179	24	2	259

表６　入力単語数ごとの COMET (提案手法が適応された文のみ

単語数	1-20	21-40	41-60	61-	すべて
ベースライン	0.924	0.896	0.888	0.908	0.901
提案手法	**0.929**	**0.897**	0.888	**0.909**	**0.903**
単純連結	0.926	0.895	**0.889**	0.907	0.901
文数	54	179	24	2	259

節翻訳結果の単純連結による方法も単語数 41-60、61 以上の範囲ではベースラインを上回っているが、提案手法には及ばず、提案手法の節翻訳統合モデルの効果が確認できたと言える。節統合モデルの具体的な効果としては、節翻訳時に稀に発生する単語の繰り返しや文脈に沿わない単語の訳出などのエラーが修正されるケースが見られた。

続いて、表 4-6 に提案手法が適用された文のみを対象にした BLEU、BLEURT、COMET の結果をそれぞれ示す。BLEU に関しては、単語数 1-20 を除くすべての範囲で提案手法がベースラインを上回る BLEU を示し、特に単語数 41-60、61 以上のときにそれぞれ 2.4 ポイント、1.1 ポイントの向上を達成した。単純連結による手法は単語数 61 以上の範囲で提案手法を上回っているが、評価対象文数が２文と少なく、その中での Brevity Penalty の影響が大きいことが原因で、実験全体での影響は小さいと考えられる。一方、単語数 1-20 の文においては単純連結による手法が最も高い BLEU を示した一方で、提案手法は単

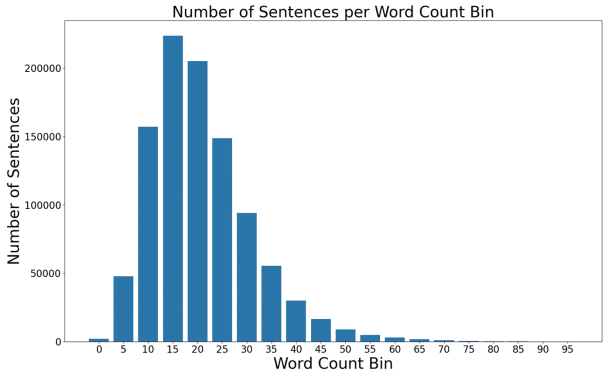

図 6　ASPEC コーパスの単語数ビンごとの文数概略

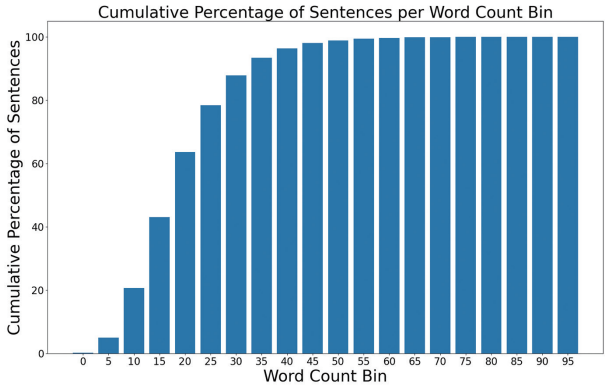

図 7　ASPEC コーパスの単語数のビンごとの文数の累積割合

純連結やベースラインに劣る結果となった。実際の翻訳結果を見ると、単純連結と提案手法で大きな翻訳結果の違いはないものの、参照訳との長さ比が提案手法では 0.963 であったのに対し単純連結では 0.991 であり、Brevity Penalty の影響の大きさが伺える。提案手法では節統合の際に訳文中の繰り返し要素を修正するような効果が得られるため、これにより訳出長が短くなったと考えられる。ただし、この BLEU の差は BLEURT や COMET の結果も勘案すると翻訳品質における大きな差を示すものではないと考えられる。BLEURT についても全体として BLEU に類似した傾向が見られ、提案手法がベースラインを上回ることが確認された。単語数 61 以上のサブセットにおけるベースラインとの大きなスコア差の要因は不明であるが、BLEURT では訳語の変化で大きなスコアの変化が起こることがあり、その影響が考えられる。COMET スコアの比較では、BLEU、BLEURT に見られる同様の傾向が認められるものの、提案手法とベースライン間のスコアの差はやや小さくなっている。これは翻訳の品質がある程度高い（BLEU で 40 を大きく上回る）ことで COMET の意味的評価ではあまり大きな差に結びつかなかったことが考えられる。

表 7　テストデータの節と文の平均単語数

テストデータ	1-20	21-40	41-60	61-
節の平均単語数	8.6	13.4	21.6	24.4
文の平均単語数	14.8	27.7	46.8	68.1

5.3　分析

提案手法による BLEU スコアの改善が特に見られた単語数 41-60、61 以上の範囲における改善の要因の一つとして、節単位の分割により翻訳モデルにとって扱いやすい単位に変換されたことが考えられる。図 6、7 は、それぞれ学習データにおける単語数の範囲（5 刻み）ごとの文数とその累積割合を示しており、単語数 10-30 の範囲に学習データが集中していることがわかる。これにより、単語数 10-30 の範囲が翻訳モデルにとって扱いやすい文長であることが示唆される。

また、表 7 は、テストデータにおける単語長の範囲ごとの節と文の平均単語数を示しており、単語数 41-60、61 以上の範囲の節が単語数 10-30 の範囲に収まっていることが確認できる。これは、翻訳モデルが扱いやすい単位に文を分割して処理することで、より効率的に入力文の情報を処理し、結果として翻訳精度を向上させることに繋がったことを示唆している。

5.4　アブレーションテスト

提案手法における節翻訳時の文内コンテキスト利用の有効性を確認するため、文内コンテキスト情報を活用する場合（提案手法による結果）と活用しない場合の比較実験を実施した。BLEU、BLEURT、COMET による評価結果を表 8-10 にそれぞれ示す。表より、節翻訳時に文内コンテキストを活用しない場合、文内コンテキストを利用する場合と比較して、BLEU、BLEURT、COMET の全スコアが低下することが確認された。これは、提案手法による節翻訳時に節の前後の文内コンテキスト情報を利用することで、分割により失われがちなコンテキスト情報の損失を抑え、節翻訳の品質を向上させることが可能であることを示している。

6　おわりに

本論文では、長文の翻訳品質の向上を目指し、節単位での文分割による分割統治的な NMT において、節単位

表 8　文内コンテキスト利用の有無による BLEU の比較
(提案手法が適応された文のみ)

単語数	1-20	21-40	41-60	61-	すべて
ベースライン	49.7	44.8	40.1	31.4	44.7
提案手法	48.8	44.9	42.5	32.5	45.0
提案手法 (コンテキストなし)	48.9	43.7	41.3	29.3	43.9
文数	54	179	24	2	259

表 9　文内コンテキスト利用の有無による BLEURT の比較
(提案手法が適応された文のみ)

単語数	1-20	21-40	41-60	61-	すべて
ベースライン	0.792	0.729	0.726	**0.653**	0.742
提案手法	**0.797**	**0.738**	**0.732**	0.618	**0.749**
提案手法 (コンテキストなし)	0.784	0.732	0.719	0.642	0.745
文数	54	179	24	2	259

表 10　文内コンテキスト利用の有無による COMET の比較
(提案手法が適応された文のみ)

単語数	1-20	21-40	41-60	61-	すべて
ベースライン	0.924	0.896	0.888	0.908	0.901
提案手法	**0.929**	**0.897**	0.888	**0.909**	**0.903**
提案手法 (コンテキストなし)	0.915	0.878	**0.872**	0.897	0.885
文数	54	179	24	2	259

の対訳データを活用した、文内コンテキストを参照できる形で翻訳を行う節翻訳モデルの学習方法を提案した。実験により、提案手法は単語数 41 以上の長文の翻訳において特に有効であり、ベースラインと比較して BLEU スコアを向上させること、また文内コンテキストの活用が有効であることが確認された。

今後の課題として、提案手法の適用範囲を広げることによる全体としての翻訳品質のさらなる向上が挙げられる。

そのため、等位接続詞以外での新しい節分割パターンの追加を検討し、接続詞を持たない長文にも対応できるように提案手法を拡張することを目指す。

謝辞

本研究の一部は JSPS 科研費 23K21697 と 21H05054 の助成を受けたものです。

参考文献

・Bahdanau, D., Cho, K., and Bengio, Y. (2014). “Neural Machine Translation by Jointly Learning to Align and Translate.” In Proc. ICLR.
・Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014). “On the Properties of Neural Machine Translation: Encoder–Decoder Approaches.” In Proc. SSST, pp. 103–111.
・Kano, Y. (2022). “Improving Neural Machine Translation by Syntax-Based Segmentation.” 奈良先端科学技術大学院大学 修士論文 .
・Kano, Y., Sudoh, K., and Nakamura, S. (2022). “Simultaneous Neural Machine Translation with
・Prefix Alignment.” In Proc. IWSLT, pp. 22–31.
・Koehn, P. and Knowles, R. (2017). “Six Challenges for Neural Machine Translation.” In Proc.WNMT, pp. 28–39.
・Kudo, T. (2006). “Mecab : Yet another part-of-speech and morphological analyzer.” https://taku910.github.io/mecab/.
・Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., and Zettlemoyer, L.(2020). “Multilingual Denoising Pre-training for Neural Machine Translation.” Transactions of the ACL, Vol. 8, pp. 726–742.
・Luong, T., Pham, H., and Manning, C. D. (2015). “Effective Approaches to Attention-based Neural Machine Translation.” In Proc. EMNLP, pp. 1412–1421.
・Morishita, M., Suzuki, J., and Nagata, M. (2017). “NTT Neural Machine Translation Systems at WAT 2017.” In Proc. WAT, pp. 89–94.
・Nakazawa, T., Yaguchi, M., Uchimoto, K., Utiyama, M., Sumita, E., Kurohashi, S., and Isahara, H. (2016). “ASPEC: Asian Scientific Paper Excerpt Corpus.” In
・Proc. LREC, pp. 2204–2208.
・Neishi, M. and Yoshinaga, N. (2019). “On the Relation between Position Information and Sentence Length in Neural Machine

Translation." In Proc. CoNLL, pp. 328–338.
· Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., Grangier, D., and Auli, M. (2019).
· "fairseq: A Fast, Extensible Toolkit for Sequence Modeling." In Proc. NAACL (Demos), pp. 48–53.
· Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). "Bleu: a Method for Automatic
· Evaluation of Machine Translation." In Proc. ACL, pp. 311–318.
· Post, M. (2018). "A Call for Clarity in Reporting BLEU Scores." In Proc. WMT, pp. 186–191.
· Pouget-Abadie, J., Bahdanau, D., van Merrienboer, B., Cho, K., and Bengio., Y. (2014). "Overcoming the curse of sentence length for neural machine translation using automatic segmentation." In Proc. SSST, pp. 78–85.
· Rei, R., C. de Souza, J. G., Alves, D., Zerva, C., Farinha, A. C., Glushkova, T., Lavie, A., Coheur, L., and Martins, A. F. T. (2022). "COMET-22: Unbabel-IST 2022 Submission for the Metrics Shared Task." In Proc.
· WMT, pp. 578–585.
· Sellam, T., Das, D., and Parikh, A. (2020). "BLEURT: Learning Robust Metrics for Text Generation." In Proc. ACL, pp. 7881–7892.
· Sudoh, K., Duh, K., Tsukada, H., Hirao, T., and Nagata., M. (2010). "Divide and translate:
· improving long distance reordering in statistical machine translation." In Proc. WMT-MetricsMATR, pp. 418–427.
· Sutskever, I., Vinyals, O., and Le, Q. V. (2014). "Sequence to Sequence Learning with Neural Networks." In Proc. NeurIPS, pp. 3104–3112.
· Tiedemann, J. and Scherrer, Y. (2017). "Neural Machine Translation with Extended Context." In Proc. DiscoMT, pp. 82–92.
· Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and
· Polosukhin, I. (2017). "Attention is all you need." In Proc. NeurIPS, pp. 5998–6008.