論文 / 著書情報
Article / Book Information

題目(和文)	学習時と推論時における入力データの特徴の違いを考慮したニューラル機械翻訳モデルの学習手法
Title(English)	
著者(和文)	美野秀弥
Author(English)	Hideya Mino
出典(和文)	学位:博士（工学）, 学位授与機関:東京工業大学, 報告番号:甲第11827号, 授与年月日:2022年3月26日, 学位の種別:課程博士, 審査員:徳永 健伸,岡崎 直観,村田 剛志,宮崎 純,下坂 正倫
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11827号, Conferred date:2022/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

TOKYO INSTITUTE OF TECHNOLOGY
SCHOOL OF COMPUTING

博士論文

学習時と推論時における入力データの特徴の違いを考慮したニューラル機械翻訳モデルの学習手法

指導教員　徳永　健伸　教授

令和4年3月

提出者

情報理工学院 情報工学系 知能情報コース

美野　秀弥

目　次

第1章

序論

1.1 機械翻訳

機械翻訳は，ある言語で書かれた文（原言語文）を入力としてこれを異なる言語の文（目的言語文）に変換する技術であり，異なる言語を扱う国や地域の人々とのコミュニケーションや異なる言語で書かれた情報の理解などの効率化のための重要な技術として注目されている．近年は，深層学習を用いたニューラル機械翻訳により旧来の統計的機械翻訳と比較して翻訳精度が飛躍的に向上し，盛んに研究が行われている．

ニューラル機械翻訳は，Long short-term memory (LSTM) や Gated recurrent unit (GRU) などの回帰型ニューラルネットワーク（Recurrent neural network (RNN)）を用いたモデル (Sutskever et al. 2014; Bahdanau et al. 2015; Luong and Manning 2015) や，注意機構に着目したトランスフォーマーを用いたモデル (Vaswani et al. 2017) などによって構成されている．どちらもエンコーダとデコーダの２つのサブネットワークを用いたエンコーダ・デコーダモデルと呼ばれる系列変換モデルである．エンコーダは，入力文の各トークン（単語などに分割された情報）を入力とし，入力文の情報を有する文ベクトルを中間表現として出力する．デコーダは，エンコーダが出力した文ベクトルを入力とし，翻訳結果を出力する．ニューラル機械翻訳は，旧来の統計的機械翻訳と同様に対訳データに基づく手法であり，学習のための対訳データのコーパス（対訳コーパス）の構築が必須である．学習に用いる対訳データは，一般的には多ければ多いほど翻訳精度が高くなるため，ウェブなどから大量の対訳データを自動で収集したり，単言語データを機械翻訳して擬似的にデータを増やすデータ拡張などを用いたりして対訳コーパスを構築する手法がある．しかし，自動で収集したり機械翻訳を用いたりして構築した対訳コーパスの中には，誤訳などを含んだ，翻訳品質の悪い対訳データが混在していることがある．

1.2 ニューラル機械翻訳の課題

誤訳などのノイズが混在している対訳データをそのまま用いて翻訳モデルを学習すると，ニューラル機械翻訳器の翻訳精度は低くなってしまう．この問題は，翻訳時と学習時におけるデータの特徴の違いによるものである．翻訳時は翻訳元の入力データと翻訳先の出力データが綺麗に翻訳されたデータであるべきなのに対し，学習時は出力データにノイズが含まれたデータとなっており，データの特徴が異なっている．

ノイズを含まない綺麗なデータであったとしても，学習時と翻訳時におけるデータの特徴の違いにより翻訳精度が低下する場合もある．例えば，口語的な表現を翻訳する日英機械翻訳器の学習に特許の日英対訳データを用いるなど，翻訳時の入力データとして想定していないジャンルの対訳データを用いて学習したニューラル機械翻訳の翻訳精度は向上しない (Chu and Wang 2018)．また，入力データとして文脈情報や単語の対応関係などの翻訳対象文以外の外部情報を用いて機械翻訳する手法 (Bawden et al. 2018; Garg et al. 2019) において，学習時と翻訳時とで用いる外部情報のデータの特徴が違うと翻訳精度が低下すると考えられる．

上記のように，理想的なニューラル機械翻訳器を作るためには，翻訳時に求める要件（ジャンルや翻訳品質など）を満たした学習データを大量に構築する必要がある．しかし，人手や計算機などの学習データ構築に必要なコストや学習時と翻訳時との状況の違いなどの理由から要件を満たしていない学習データが使われることが多く課題となっている．

1.3 目的と手法

本論文では，上記の課題に着目し，学習データと翻訳時の要件を満たした入出力データとの間の特徴が違う場合におけるニューラル機械翻訳の翻訳精度の向上を目的とする．学習データと翻訳時の入出力データとの間の特徴は下記の４つの観点で分類でき，それぞれで特徴の違いが生じる場合がある．

(a) 翻訳対象（入力データ）

(b) 翻訳結果（出力データ）

(c) 翻訳対象と翻訳結果との関係

(d) 外部情報（翻訳対象とは別の入力データ）

(a) の翻訳対象の特徴の違いは，ニュース文や特許文書など，翻訳対象の入力データの語彙，体裁、内容などの違いなどが該当する．(b) の翻訳結果の特徴

の違いは,「平易な語彙で翻訳する」,「専門用語などは固定訳を用いる」など翻訳結果の出力データのスタイルの違いが該当する．(c) の翻訳対象と翻訳結果との関係の違いは，(b) の翻訳結果の特徴の違いにも関連するが,「翻訳対象のデータをコンパクトに翻訳する」,「詳細に翻訳する」など翻訳スタイルの違いが該当する．翻訳品質の違いも (c) に該当する．(d) の外部情報は,（翻訳対象，翻訳結果）のペアである対訳データとは別に翻訳を補助する入力データのことである．外部情報には翻訳対象の周辺の文の情報や翻訳時に利用可能な辞書情報など考えられ，これらの外部情報のスタイルの違いや品質の違いなどが該当する．

本論文では，上記４種類のデータの特徴の違いに起因する翻訳精度の低下の問題を解くために，下記の２つのニューラル機械翻訳のタスクを対象にして新たな手法を提案する．

１つ目は，学習時と翻訳時との間の翻訳対象文の特徴の違いに着目した，ドメインタグを用いたニューラル機械翻訳のタスクである．このタスクでは，上記の (a)，(b)，(c) の特徴の違いを扱う．ニューラル機械翻訳は，翻訳モデルの学習に対訳データを必要とする．しかし，対訳データの構築にはコストがかかることから，翻訳誤りなどのノイズを含むデータ（(c) に該当）や翻訳対象とは違う分野のデータ（(a)，(b) に該当）など，実際に翻訳するデータの特徴とは違う特徴を有する対訳コーパスのデータを学習データとして利用することが多い．このようなドメイン*の違う対訳コーパスを混在して学習すると翻訳精度は低下することが知られている．本論文では，この問題を解決するために，対訳データの原言語側が有する特徴，目的言語側が有する特徴，および原言語側と目的言語側との対訳関係に関する特徴の３つを明示的に区別して翻訳モデルを学習するドメインタグを用いた手法を提案する．

２つ目は，文脈情報考慮型ニューラル機械翻訳を扱う．１つ目のタスクとは異なり，学習時と翻訳時との間の外部情報の特徴（(d) に該当）の違いに着目する．本論文が扱う外部情報は文脈情報である．ニューラル機械翻訳は原言語一文を入力して目的言語への翻訳結果を出力するシステムが多いが，さらに翻訳精度を上げるために翻訳対象の周辺の文を文脈情報として利用する手法がある (Maruf et al. 2019; Voita et al. 2019; Agrawal et al. 2018; Bawden et al. 2018; Kuang et al. 2018; Läubli et al. 2018; Miculicich et al. 2018; Tiedemann and Scherrer 2017; Wang et al. 2017)．しかし，従来手法では，文脈情報に原言語側の周辺の文を用いることで翻訳精度が向上するが，文脈情報に目的言語側の周辺の文を用いると翻訳精度が低下することが報告されている (Bawden et al. 2018)．従来手法は，目的言語側の周辺の文として学習時は誤訳のない綺麗な正解訳を，翻訳時は機械翻訳誤りを含む機械翻訳結果を用いている．本論文では，目的言語側の周辺の文を用いる従来手法の問題が翻訳モデルの学習時と翻訳時で

*本論文では対訳コーパスに共通した特徴をドメインと定義する．

用いる周辺の文のデータの特徴の違いにあると仮定し，学習時と翻訳時で用いる目的言語側の文の特徴の違いによる影響を緩和するための学習データ制御方法を提案する．

本論文では，上記の２つのニューラル機械翻訳のタスクが有する問題に対して翻訳精度が向上する有効な手法を提案し，翻訳実験により双方の手法の有効性を示す．

1.4 論文構成

本論文の構成は以下の通りである．第1章（本章）では，本論文の概要，目的，提案について述べた．第2章では，ニューラル機械翻訳の技術について述べる．第3章では，ドメインタグを用いたニューラル機械翻訳と文脈情報考慮型ニューラル機械翻訳の従来手法の問題点を述べる．第4章では，ドメインタグを用いたニューラル機械翻訳の新たな手法を提案する．第5章では，文脈情報考慮型ニューラル機械翻訳の新たな手法を提案する．最後に，第6章では，本論文のまとめを述べる．

第2章

ニューラル機械翻訳

本章では，本論文で用いているトランスフォーマーモデル (Vaswani et al. 2017) を含めた 3 つの代表的なニューラル機械翻訳について概要を説明し，トランスフォーマーモデルを用いた際のニューラル機械翻訳の学習，翻訳時の手続きを説明する．また，翻訳結果を評価する際に用いられる自動評価尺度について説明する．

2.1 ニューラル機械翻訳の概要

ニューラル機械翻訳は，トークン長 K の原言語の文 $x = \{x_1, x_2, ..., x_K\}$ をトークン長 J の目的言語の文 $y = \{y_1, y_2, ..., y_J\}$ に変換する．各トークンは原言語側，目的言語側のトークンともに語彙数次元のベクトル（ワンホット（one-hot）ベクトル）で表現する．Seq2seq モデル (Sutskever et al. 2014)，注意機構付き Seq2seq モデル (Bahdanau et al. 2015; Luong and Manning 2015)，トランスフォーマーモデル (Vaswani et al. 2017) などの代表的なニューラル機械翻訳はエンコーダ・デコーダモデルを用いており，図 2.1 に示すように，原言語の文からベクトルを出力するエンコーダと，エンコーダが出力したベクトルを入力として目的言語の翻訳結果を 1 トークンごとに出力するデコーダの 2 つのサブネットワークから構成され，エンコーダとデコーダの構成はモデルによって異なる．

2.1.1 Seq2seq モデル

Seq2seq モデル (Sutskever et al. 2014) は，可変長の入出力を扱うことができる Long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) や Gated recurrent unit (GRU) (Cho et al. 2014) などの回帰型ニューラルネットワーク（Recurrent neural netword (RNN)）を用いたモデルであり，自身が出力した結果を再度自身の入力に用いることで過去の情報と現在の情報の両方を利用することが可能となり，旧来の統計的機械翻訳を上回る翻訳精度を達成した．図 2.2 に Seq2seq モデルの概要図を示す．エンコーダは，以下の式によりワンホットベク

図 2.1: エンコーダ・デコーダモデルの概要図.

トルで各トークンを表現した入力文 x から隠れベクトル $c = (c_1, c_2, ..., c_K)$ を計算する.

$$c_k = RNN(c_{k-1}, E_x(x_k)) \tag{2.1.1}$$

RNN は回帰型ニューラルネットワーク層，E_x は原言語側の単語埋め込み層を示す．デコーダは，エンコーダが出力した c を用いて，以下の式により j 番目の翻訳結果 y_j の確率分布をソフトマックス関数（softmax）を用いて出力する.

$$p(y_j|y_{1:j-1}, c) = softmax(Linear(RNN(h_{j-1}, E_y(y_{j-1})))) \tag{2.1.2}$$

$Linear$ は線形変換層，h はデコーダの隠れベクトル，E_y は目的言語側の単語埋め込み層を示す．デコーダは，目的言語側の語彙と同じ次元数を持つ確率分布を出力し，その中で最も大きい値を持つ次元に対応する単語が選択され翻訳結果となる.

図 2.2 では，入力単語列（トークン列）である，x_1 :“He”，x_2 :“come”，x_3 :“.” をそれぞれエンコーダに入力してエンコーダは入力文 x の文ベクトルとして c_3 を出力する．デコーダは隠れベクトル h_{j-1} と 1 つ前に出力したトークン（最初の出力時は，c_3 と文頭を表す特殊記号トークン（<BOS>）のベクトル）を入力して隠れベクトル h_j，出力ベクトル y_j を計算して翻訳結果を 1 トークンずつ出力している．文末を表す特殊記号トークン（<EOS>）が出力されると翻訳が終了となる.

図 2.2 から分かるように，回帰型ニューラルネットワークは可変長の入出力が可能なネットワークであるが，入力長が長いと初期の入力情報（例えば，表 2.2 の x_1 :“He”）の消失の問題があり，LSTM や GRU はこの問題を緩和するために提案された回帰型ニューラルネットワークである.

2.1.2 注意機構付き Seq2seq モデル

LSTM や GRU を用いることで入力長が長い場合に翻訳精度が大きく低下する問題は解決されたが，これらのモデルには，全ての入出力の隠れベクトルを保持し続ける必要があるという課題がある (Cho et al. 2014)．保持すべき情報が増加すればするほど，処理が複雑となり翻訳精度は低下する．注意機構付き

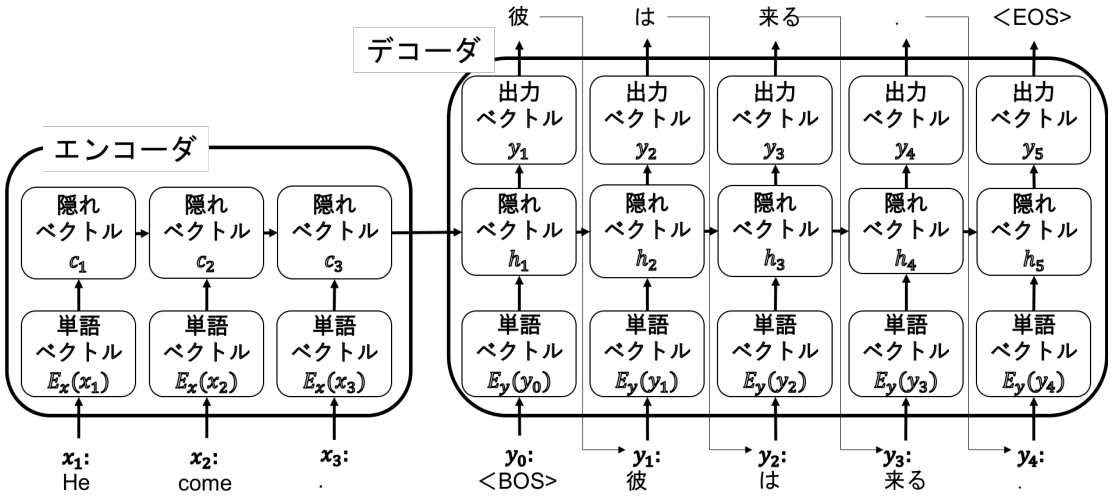

図 **2.2:** Seq2seq モデルの概要図.

Seq2seq モデル (Bahdanau et al. 2015; Luong and Manning 2015) は，回帰型ニューラルネットワークが持つ長文を入力した際に入力の情報を十分に保持できなくなるという問題を解決するために，注意機構を加えたモデルである．注意機構は，デコーダが翻訳結果を生成する際に，エンコーダのどの時点の情報を利用するかを重み（注意重み）を用いて制御する仕組みである．

図 2.3 に注意機構付き Seq2seq モデルの概要図を示す．エンコーダは，Seq2seq モデルと同様に，以下の式により入力文のベクトル x から隠れベクトル $c = (c_1, c_2, ..., c_K)$ を計算する．

$$c_k = RNN(c_{k-1}, E_x(x_k)) \tag{2.1.3}$$

RNN は回帰型ニューラルネットワーク層，E_x は原言語側の単語埋め込み層を示す．デコーダは，エンコーダの出力を用い，以下の式により j 番目の翻訳結果 y_j を出力する．

$$h_j = RNN(h_{j-1}, E_y(y_{j-1})) \tag{2.1.4}$$

$$f_j = Attn(h_j, c) \tag{2.1.5}$$

$$\hat{h}_j = tanh(W_f[f_j; h_j]) \tag{2.1.6}$$

$$p(y_j|y_{1;j-1}, c) = softmax(Linear(\hat{h}_j)) \tag{2.1.7}$$

h_j は j 番目の目的言語側の隠れベクトル，E_y は目的言語側の単語埋め込み層を示す．$Attn$ は注意機構であり，エンコーダの隠れベクトルの重みつけ和を計算する．f_j は文脈ベクトルを示し，$\hat{h}_j$ はデコーダの隠れベクトルと文脈ベクトルを用いて計算する．

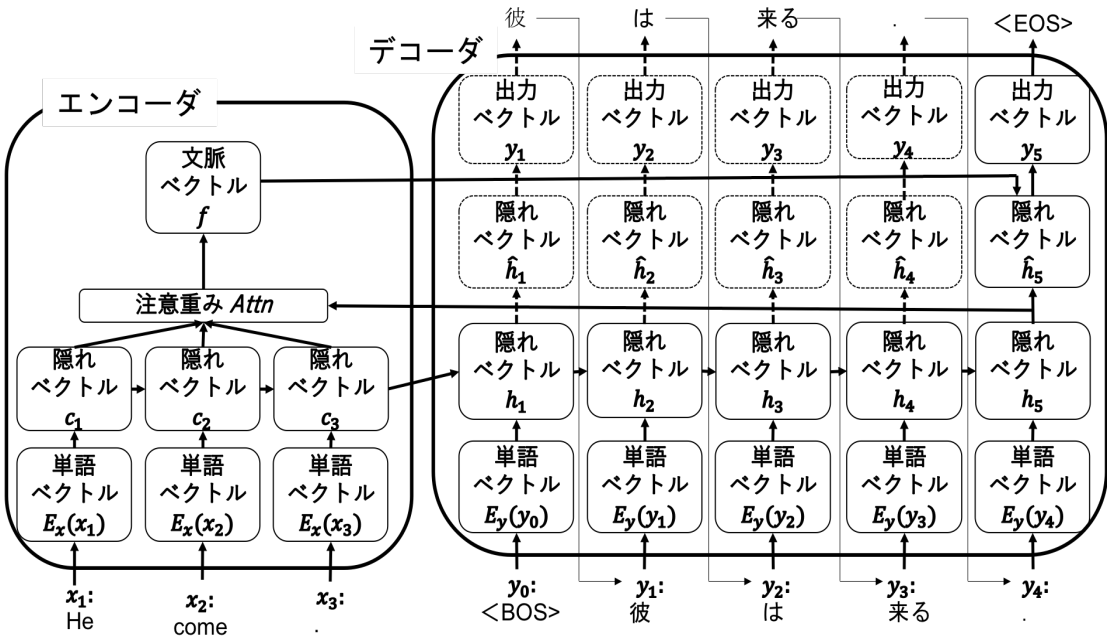

図 2.3: 注意機構付き Seq2seq モデルの概要図.

図 2.3 は，デコーダがすでに y_1 :「彼」，y_2 :「は」，y_3 :「来る」，y_4 :「.」を出力している状態で次の翻訳結果 y_5 :「<EOS>」を出力する図である．エンコーダが文ベクトル c を各入力トークンの隠れベクトルの重み（注意重み）付き和で計算しており，注意重みはデコーダが翻訳結果を出力するごとに再計算するため毎回変化する．これにより，翻訳結果を出力する際に必要なエンコーダの情報を選択的に利用することができる．注意機構を組み込むことで，エンコーダの全ての情報をデコーダ側で保持する必要がなくなり，より多くの必要な情報を保持できるようになり翻訳精度が向上した (Bahdanau et al. 2015).

2.1.3 トランスフォーマーモデル

注意機構付き Seq2seq モデルは回帰型ニューラルネットワークを用いているために入出力共に逐次的に処理する必要があり，文長が長いと処理が長くなる他，文内の処理の並列化が困難であるという課題がある．この問題に対処したニューラル機械翻訳モデルが図 2.4 に示すトランスフォーマーモデル (Vaswani et al. 2017) である．トランスフォーマーモデルはエンコーダ，デコーダともに回帰型ニューラルネットワークを使わず，エンコーダでは入力の各トークンを並列に処理することが可能となり，デコーダにおいても学習時はデコーダの予測結果ではなく正解トークンを用いることができるため並列処理が可能となる．

注意機構付きSeq2seqモデルにはないトランスフォーマーモデルの特徴の1つに，入出力ともに各トークン間の文脈情報を考慮できるように，自己注意と呼ばれる注意機構を導入していることがある．注意機構付きSeq2seqで用いられた注意機構（クロス注意）はデコーダが自身の入力トークンではなくエンコーダのトークンの情報を選択的に参照するのに対し，自己注意はエンコーダ間およびデコーダ間のトークン間の情報を選択的に参照する，という違いがある．すなわち，注意機構付きSeq2seqモデルは自己注意を持っていないため，系列データが持つ前後の文脈情報は回帰型ニューラルネットワークの隠れ状態の遷移のみに依存しているのに対し，トランスフォーマーモデルは自己注意によりエンコーダについては入力全体，デコーダについては出力済みの情報を参照することが可能である．トランスフォーマーモデルは，このクロス注意と自己注意をN層にわたり繰り返し適用すること依存関係を潜在的に利用可能にしている．

ここで，トランスフォーマーモデルのエンコーダのn層目の出力$H_{enc}^{(n)}$は以下で計算する．

$$C_{enc}^{(n)} = LN(H_{enc}^{(j-1)} + SelfAttn(H_{enc}^{(n-1)})), \tag{2.1.8}$$

$$H_{enc}^{(n)} = LN(C_{enc}^{(n)} + FF(C_{enc}^{(n)})). \tag{2.1.9}$$

LNはレイヤーにおける正規化処理，$SelfAttn$は自己注意，FFはReLU関数を用いたフィードフォワードネットワークを示し，$C_{enc}^{(0)}$はエンコーダに入力される単語ベクトルを示す．トランスフォーマーモデルのデコーダのn層目の出力$S_{dec}(n)$，およびj番目の翻訳結果y_jは以下で計算する．

$$C_{dec}^{(n)} = LN(H_{dec}^{(n-1)} + SelfAttn(H_{dec}^{(n-1)})), \tag{2.1.10}$$

$$G_{dec}^{(n)} = LN(C_{dec}^{(n)} + CrossAttn(C_{dec}^{(n)}, H_{enc}^{N})), \tag{2.1.11}$$

$$H_{dec}^{(n)} = LN(G_{dec}^{(n)} + FF(G_{dec}^{(n)})), \tag{2.1.12}$$

$$p(y_j|y_{1;j-1}, H_{enc}^{N}) = softmax(Linear(H_{dec}^{n})). \tag{2.1.13}$$

$CrossAttn$はクロス注意を示す．さらに，トランスフォーマーモデルで用いる自己注意とクロス注意は，ともに複数ヘッド機構を用いることにより複数の観点で隠れベクトルを重視することを決めることができ，大域的な文脈情報を扱うことが可能となる．複数ヘッド機構を用いた自己注意およびクロス注意は以下で計算する．

$$Attention(H) = softmax(QK^T / \sqrt{d_k})V. \tag{2.1.14}$$

Qはクエリー，Kはキー，Vはバリューであり，それぞれd_k次元の部分空間に射影された行列である．d_kはトークンの埋め込み次元数dとヘッド数n_{head}によって決まる定数($d_k = d/n_{head}$)である．$SelfAttn$では，直前のエンコーダ層または

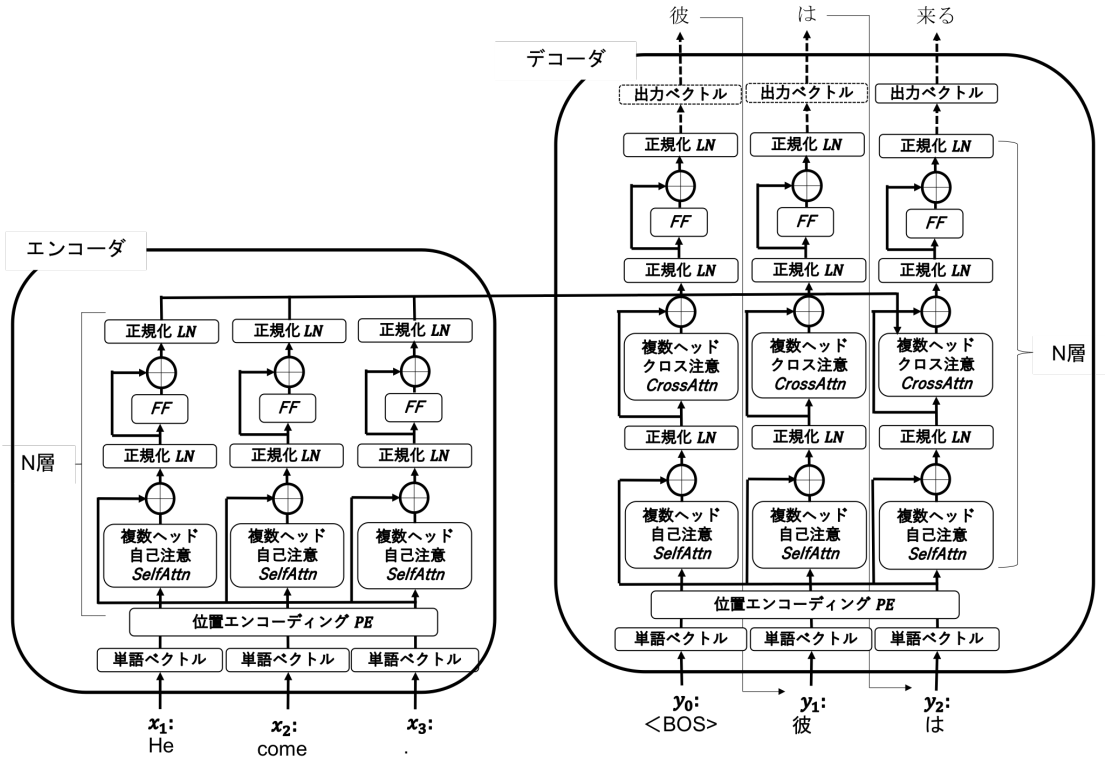

図 2.4: トランスフォーマーモデルの概要図.

デコーダ層の出力を重みパラメータ $W^{Q^j} \in \mathbb{R}^{d\times d_{head}}$，$W^{K^j} \in \mathbb{R}^{d\times d_{head}}$，$W^{V^j} \in \mathbb{R}^{d\times d_{head}}$ を用いて Q，K，V に射影する．*CrossAttn* では，直前のデコーダ層の出力を Q に，エンコーダの最終層の出力を K，V に，それぞれ重みパラメータを用いて射影する．デコーダの最終層の出力 H^N_{dec} を目的言語側の語彙数次元のベクトルに線形変換し，ソフトマックス関数を適用して確率値として出力トークンを生成する．なお，トランスフォーマーモデルは，位置エンコーディングと呼ばれる処理により絶対的な位置情報を単語ベクトルに埋め込んでいる．位置エンコーディングは，以下に示す周期関数を用いて位置情報を定数で表現して埋め込む処理を行なっており，相対的な位置情報を扱う回帰型ニューラルネットワークとは異なる．

$$PE_{(pos,2i)} = \sin(pos/10000^{2i/d}), \tag{2.1.15}$$

$$PE_{(pos,2i+1)} = \cos(pos/10000^{2i/d}). \tag{2.1.16}$$

pos は単語の位置，i は埋め込み成分の次元，d はトークンの埋め込み次元を示す.

図 2.4 は x_1 :“He”，x_2 :“come”，x_3 :“.” をそれぞれエンコーダに入力し，デコー

ダがすでに「彼」,「は」を出力した状態で,「来る」を出力する状態を表している．各トークンが複数ヘッド注意とフィードフォワードニューラルネットワーク（FF）を連結し，レイヤーにおける正規化処理を行なっている．

2.2 学習

ニューラル機械翻訳のモデルの学習は，正解トークンのワンホットベクトルとソフトマックス関数で正規化された出力ベクトルとの交差エントロピーを損失として，誤差逆伝搬法を用いてパラメータの更新に必要な勾配情報を計算し，パラメータの更新を行う．交差エントロピーを用いた損失$Loss$は以下で計算する．

$$Loss = -\sum_{j=1}^{J} \mathbf{t_j} \log p(\mathbf{y_j}|\mathbf{x}, \mathbf{y_{<j}}) \tag{2.2.1}$$

$\mathbf{x}$は入力文ベクトル，$\mathbf{y_j}, \mathbf{t_j}$は翻訳出力文ベクトル，および参照訳の$j$番目の出力ベクトルに対応するワンホットベクトルである．パラメータは，誤差逆伝搬法により計算される損失の勾配情報を用い，確率的勾配降下法（SGD）(Robbins and Monro 1951)，Adagrad (Duchi et al. 2011)，RMSprop (Tieleman and Hinton 2012)，Adam (Kingma and Ba 2015)などの最適化アルゴリズムにより更新される．本論文ではAdamを用いる．

2.3 機械翻訳結果の評価尺度

機械翻訳モデルの出力の評価には，人手による評価と自動評価がある．人手による評価は，出力の品質を内容の伝達レベルや流暢性などの定められた基準に従って人間が判断する評価方法である．機械翻訳モデルの翻訳性能を正しく評価できることが期待される反面，コストがかかるため大規模に実施することが困難である．一方，自動評価は，人手によって作成された正解訳（参照訳）と機械翻訳モデルの出力との類似度を自動で算出する評価方法であり，BiLingual Evaluation Understudy (BLEU) (Papineni et al. 2002)，Rank-based Intuitive Bilingual Evaluation Score (RIBES) (Isozaki et al. 2010)，Metric for Evaluation of Translation with Explicit ORdering (METEOR) (Denkowski and Lavie 2014)などがある．本論文では，自動評価尺度の中で，特に標準的な評価尺度として広く用いられているBLEUを用いて評価する．

BLEU は n-gram 精度に基づく評価尺度であり，以下の式で計算する．

$$\mathrm{BLEU} = \mathrm{BP}_{\mathrm{BLEU}} \cdot \exp\left(\sum_{n=1}^{N} \frac{1}{N} \log p_n\right) \tag{2.3.1}$$

$$p_n = \frac{\sum_i \text{出力文} i \text{と参照訳} i \text{で一致した n-gram 数}}{\sum_i \text{出力文} i \text{中の全 n-gram 数}} \tag{2.3.2}$$

p_n は評価対象全体について，出力文と参照訳とを比較し，n-gram の一致率を算出する．N は通常 4 が用いられ，その場合の BLEU スコアを BLEU-4 と呼ぶ．$n = 1$ の場合（1-gram）は単語訳の正しさを表す指標，高次の n-gram は翻訳の流暢さを表す指標となっており，BLEU スコアは両者を組み合わせた指標といえる．$\mathrm{BP}_{\mathrm{BLEU}}$ (brevity penalty) は，出力文が参照訳よりも長い場合は $BP = 1$ を与え，出力文が参照訳よりも短い場合には $BP < 1$ を与えるペナルティ項であり，以下の式で計算する．

$$\mathrm{BP}_{\mathrm{BLEU}} = \begin{cases} 1 & (c \geq 1) \\ \exp(1 - r/c) & (c < r) \end{cases} \tag{2.3.3}$$

c は出力長，r は参照訳長を表す．

BLEU スコアは 0 から 1 の実数で表され，スコアが高いほど翻訳精度が高いと判断される．

第3章

関連研究

本論文では，1.3節で示したようにニューラル機械翻訳における学習時と翻訳時の間の特徴の違いを「(a) 翻訳対象（入力データ）」,「(b) 翻訳結果（出力データ）」,「(c) 翻訳対象と翻訳結果との関係」,「(d) 外部情報（翻訳対象とは別の入力データ）」の４つの観点で分類し，これらの特徴の違いに起因する翻訳精度の低下の問題を解くことを目的としている．４つの観点のうち,「(a) 翻訳対象（入力データ）」,「(b) 翻訳結果（出力データ）」,「(c) 翻訳対象と翻訳結果との関係」の特徴の違いを考慮すべきニューラル機械翻訳のタスクとしてドメインタグを用いたニューラル機械翻訳に着目し,「(d) 外部情報（翻訳対象とは別の入力データ）」の特徴の違いを考慮すべきニューラル機械翻訳のタスクとして目的言語側の前文を用いた文脈考慮型ニューラル機械翻訳に着目した．上記２つのタスクにおいて，従来手法と提案手法が着目している特徴の項目を表3.1に表す．本章では，本論文の目的に合致した２つのニューラル機械翻訳のタスクの関連研究を説明し，双方の従来手法の問題点を示す．

3.1 ドメインタグを用いたニューラル機械翻訳

ニューラル機械翻訳の学習は，利用可能な対訳コーパスの全てのデータを一括りに扱う場合が多いが，翻訳対象のドメイン以外のデータやノイズを含むデータで学習すると翻訳精度が低下することが知られている (Luong and Manning 2015; Belinkov and Bisk 2018)．この問題を解決するために，転移学習の一種であるドメイン適応 (Pan and Yang 2010) を用いたニューラル機械翻訳がある (Chu and Wang 2018)．本論文では，ニューラル機械翻訳で用いられているドメイン適応手法のうち，翻訳対象のドメインのデータと翻訳対象ではないドメインの違いを考慮して学習するドメインタグ (Kobus et al. 2017; Chu et al. 2017; Berard et al. 2019a, 2019b; Caswell et al. 2019; Vaibhav et al. 2019) に着目した．

Kobus et al. (2017) と Chu et al. (2017) は，複数の対訳コーパスのドメインを区別して学習するニューラル機械翻訳の手法として，対訳データ中の原言語側の文の先頭に対訳コーパスのドメインを表すタグ（コーパスタグ）を挿入して学

手法	データの特徴の分類項目 (a) 翻訳対象	(b) 翻訳結果	(c) 翻訳対象と翻訳結果との関係	(d) 外部情報
ドメインタグを用いたニューラル機械翻訳				
コーパスタグ (Kobus et al. 2017)	✓	(✓)	(✓)	-
ノイズタグ (Vaibhav et al. 2019)			✓	-
２つのタグ (Berard et al. 2019a)	✓		✓	-
提案手法	✓	✓	✓	-
目的言語側の前文を用いた文脈情報考慮型ニューラル機械翻訳				
従来手法 (Bawden et al. 2018)	-	-	-	
	-	-	-	
提案手法	-	-	-	✓

表 3.1: 着目している学習時と翻訳時の間の特徴.

習することを提案した*.

Berard et al. (2019b)，Caswell et al. (2019)，Vaibhav et al. (2019)，は，逆翻訳を用いて構築したコーパスを追加して学習する際に，逆翻訳のデータに対してタグ（ノイズタグ）を付与して学習することで翻訳精度が向上することを示した.

Berard et al. (2019a) は，コーパスタグとノイズタグの２つのドメインタグを付与することを提案した.

ドメインタグを用いた上記の従来手法は，コーパスの持つ特徴を表すためのドメインタグ（コーパスタグ）やコーパスの翻訳精度を表すドメインタグ（ノイズタグ）を付与し，これらの特徴を区別して学習するための手法である．対訳データにコーパスタグやノイズタグを付与して学習したニューラル機械翻訳の翻訳モデルは，対訳データが持つ特有の文体，言い回し，ノイズなどの特徴を学習する．翻訳時は，原言語側の文の先頭に翻訳時の要件を満たすドメイン

*コーパスタグを用いた手法は対訳コーパスの持っている１つの特徴に対してタグを付与する手法であり，広義では表 3.1 のデータの特徴の分類項目のうちの１つに着目した手法である．しかし，対訳コーパスの多くは特許コーパスや科学技術論文コーパスなど翻訳対象の特徴に着目して構築されていると考え，表 3.1 ではコーパスタグを用いた手法は (a) 翻訳対象の特徴の違いに着目した手法とした.

タグを付与することで，付与したドメインタグが持つ特徴を反映するように機械翻訳するため，翻訳精度が向上する．しかし，従来手法が付与するドメインタグではコーパスの特徴を区別するには不十分である場合がある．例えば，従来手法はコーパスのドメインが同じ翻訳対象文が複数の翻訳スタイルで書かれているデータに対して，複数の翻訳スタイルを区別して学習するようなドメインタグを付与することはできない．

本論文では，対訳データを区別して学習するためには，1.3 節で示した「(a) 翻訳対象（入力データ）」,「(b) 翻訳結果（出力データ）」,「(c) 翻訳対象と翻訳結果との関係」の全ての特徴を考慮したドメインタグを付与する必要があると考え，従来手法はこの違いを考慮していないことが問題であると考えた．

3.2 目的言語側の前文を用いた文脈情報考慮型ニューラル機械翻訳

ニューラル機械翻訳において文脈の情報は，語彙，時制，指示詞，省略などの文間の表現の一貫性を保持する上で重要である (Bawden et al. 2018; Läubli et al. 2018; Müller et al. 2018; Voita et al. 2018)．加えて，翻訳する上で対応する必要が出てくるゼロ照応解析 (Okumura and Tamura 1996; Iida et al. 2016) や語義曖昧性の解消などにも文脈は有効な情報となる．

文脈考慮型ニューラル機械翻訳の実装方法には，結合ベース文脈考慮型ニューラル機械翻訳 (Tiedemann and Scherrer 2017; Bawden et al. 2018) とマルチエンコーダ文脈考慮型ニューラル機械翻訳 (Bawden et al. 2018; Kim et al. 2019) がある．Tiedemann and Scherrer (2017) は，前文と翻訳対象の原文の間に特殊なトークン（*_BREAK_*）を挿入して結合して単体のエンコーダに入力する結合ベース文脈情報考慮型ニューラル機械翻訳を提案した．Bawden et al. (2018) は，マルチエンコーダニューラル機械翻訳 (Zoph and Knight 2016; Libovický and Helcl 2017; Wang et al. 2017) を拡張し，原言語側の前文と翻訳対象の文とを別々のエンコーダに入力するマルチエンコーダ文脈情報考慮型ニューラル機械翻訳を提案した．マルチエンコーダ文脈考慮型ニューラル機械翻訳は，各エンコーダの出力ベクトルを連結（attention concatination），重み付き和（attention gate），階層（hierarchical attention）のいずれかにより結合する．Kim et al. (2019) は，各エンコーダの出力ベクトルの結合をデコーダの前段，またはデコーダ内で行うことができることを示してその違いを調査した．

これらの文脈考慮型ニューラル機械翻訳は，翻訳対象の文だけでなく原言語側の文脈，目的言語側の文脈，またはその両方を入力できる (Tiedemann and Scherrer 2017; Agrawal et al. 2018; Bawden et al. 2018; Kim et al. 2019)．Bawden et al. (2018) は，原言語側と目的言語側の前文を文脈情報として用いた手法を提案

してそれぞれの効果を比較した．Agrawal et al. (2018) は，原言語側と目的言語側の複数の前文と後文を文脈情報として用いる手法を提案した．Bawden et al. (2018) は，目的言語側の前文を用いる手法で学習時に前文の参照訳を用いて翻訳時には機械翻訳結果を用いた．実験の結果，目的言語側の文脈情報の効果が低かったことを報告している．Agrawal et al. (2018) は，目的言語側の文脈を使うと誤差が伝搬することで翻訳性能が低下することを報告している．また，複数の文脈情報を使うことで翻訳品質が向上する場合があるが，膨大なメモリと計算量が必要となる傾向があることを報告している．

上記のように，従来手法の多くは翻訳精度が向上することから原言語側の文脈を用いた文脈考慮型ニューラル機械翻訳を用いており，目的言語側の文脈を使った従来手法は翻訳性能が低下することを報告している．しかし，例えば、日英翻訳において「ですます調」と「である調」の訳語の統一のためには目的言語側の前文がどのように翻訳しているかが重要な情報となり、目的言語側の文脈情報が有益な場合もあると考えられる．

本論文では，従来手法で目的言語側の前文を使うと翻訳性能が劣化した原因は，目的言語側の前文自体の問題ではなく，学習時に前文の参照訳を用い，翻訳時には機械翻訳結果を用いていることにあると考えた．すなわち，目的言語側の前文を用いた文脈考慮型ニューラル機械翻訳は，1.3 節で示した学習時と翻訳時との間の特徴の違いを分類した 4 つの観点のうち「(d) 翻訳対象とは別の入力データ（外部データ）」の特徴の違いを考慮しなければならず，従来手法はこの違いを考慮していないことが問題であると考えた．

第4章

複数のドメインタグを用いたニューラル機械翻訳

本章では，ドメインタグを用いたニューラル機械翻訳を扱う．ドメインタグを用いたニューラル機械翻訳はコーパスの特徴を表すドメインタグを用いてコーパスの特徴を区別して翻訳モデルを効果的に学習する手法として用いられている (Kobus et al. 2017; Vaibhav et al. 2019; Berard et al. 2019b)．例えば，ニュースの日英機械翻訳のモデルの構築にニュースではないジャンルの対訳データを用いる際などに用いられる．ニュースの日英機械翻訳を対象のタスクとすると，ニュースのジャンルをインドメインと呼び，ニュースではないジャンルをアウトオブドメインと呼ぶ．インドメインの機械翻訳モデルを構築する際にドメインタグを用いずにアウトオブドメインのデータを用いて学習すると，インドメインとアウトオブドメインのデータ間の特徴の違いにより，翻訳精度が低下することがある．本章では，ドメインの違いを区別して学習するドメインタグを用いた手法に着目し，従来手法では扱えなかった複数の特徴を区別して学習するための複数のドメインタグを用いた手法を提案する*．提案手法の効果を確認するために，翻訳元の言語（原言語）側のデータを 1 文ごとに内容が等価になるように翻訳先の言語（目的言語）に翻訳した対訳コーパス（内容等価コーパス），原言語側と目的言語側のデータ同士を文単位で自動対応づけした対訳コーパス（自動アライメントコーパス），目的言語側のデータを機械翻訳で翻訳した対訳コーパス（逆翻訳コーパス）の 3 種類のニュースの対訳コーパスを用いて翻訳実験を行った．実験の結果，提案手法が従来手法よりも複数の特徴を持つコーパスを用いたニューラル機械翻訳の学習に効果的であることを確認した．

*本論文では，翻訳対象と翻訳先を含んだ学習データ全体の体裁や内容などの特徴をまとめてドメインと定義する．

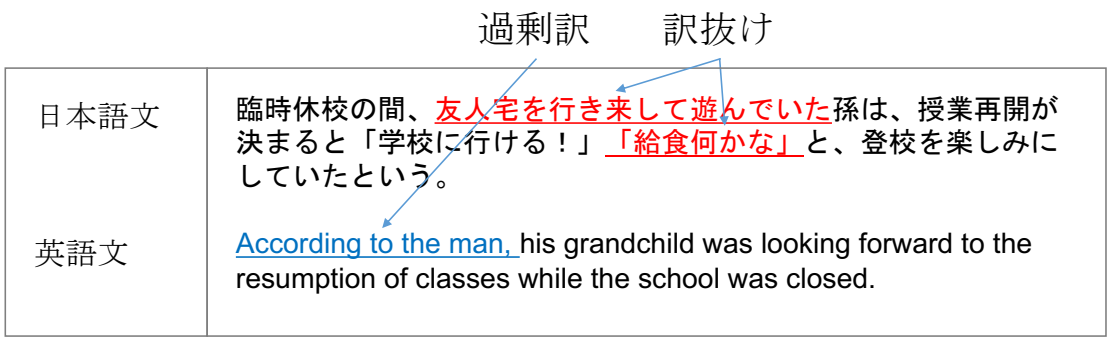

図 4.1: 自動で文単位で対応づけれらた日英対訳データの例．一部の単語やフレーズが翻訳されていない「訳抜け」や翻訳元にない情報が翻訳結果に含まれる「過剰訳」を含む．

4.1 はじめに

ニューラル機械翻訳のモデルの学習には対訳データが必要であり，膨大な量の翻訳対象のドメインの対訳データを学習することで翻訳精度が向上する．しかし，コストの面などから翻訳対象のドメインの適切な対訳データ（インドメインのデータ)[†]を大量に用意することは困難である．学習データ[‡]を増やすために，翻訳対象ではないドメインの対訳データや自動で獲得したノイズを含んだ対訳データなど（アウトオブドメインのデータ）を用いることが考えられるが，アウトオブドメインのデータを追加して学習することでインドメインの翻訳精度が低下する場合がある．

アウトオブドメインのデータとして用いられる自動で獲得可能なデータには，文単位で自動対応づけられた対訳データや，機械翻訳で翻訳した対訳データなどがある．自動で対応づけられた対訳データは，様々な理由から 1 文ごとに対応づけられていないデータ[§]の中から対訳データである可能性の高いデータ対を自動で抽出することにより構築されたものである．図 4.1 の例のように，一部の単語やフレーズが翻訳されていなかったり（訳抜け），翻訳元にない情報が翻訳結果に含まれていたり（過剰訳）することがある．また，機械翻訳で翻訳した対訳データは，事前に用意した機械翻訳器を用いて翻訳したデータから構築されたもので，機械翻訳器の性能に依存して機械翻訳結果に訳抜け，過剰訳，誤訳が含まれることがある．機械翻訳で翻訳する際には，翻訳元の言語

[†]本論文では，翻訳元データと翻訳元の情報をそのまま翻訳した翻訳結果の対をインドメインのデータと定義する．

[‡]学習に用いる対訳データを学習データと呼ぶ．

[§]例えば，ニュース翻訳は 1 文単位ではなく記事単位で行われることが多く，翻訳元の複数文をまとめて 1 つのデータとして翻訳したり，翻訳元の 1 文を複数文に分割して翻訳したりする．さらに，翻訳元の文の順序が変わることもある．

（原言語）のデータを翻訳先の言語（目的言語）に翻訳（順翻訳）する場合と目的言語のデータを原言語に翻訳（逆翻訳）する場合が考えられるが，目的言語側のデータがノイズのない正しい文となるように，逆翻訳を用いることが多い (Sennrich et al. 2016a).

これらの違う特徴を持つデータを用いて効果的に機械翻訳のモデルを学習する技術として，ドメインタグを用いる手法 (Chu et al. 2017; Kobus et al. 2017) があるが，それぞれのデータに対して 1 種類のタグを付与する手法であるため，それぞれのデータの特徴を十分に表現できない場合がある．例えば，自動で対応づけられた対訳データと機械翻訳で逆翻訳した対訳データは，翻訳先である目的言語側のデータが同じ特徴を持っているにも関わらず，従来手法ではその情報をタグで表すことができない．目的言語側の文の特徴は出力文を制御する上で重要な特徴であり，この情報を活用することで翻訳精度の向上が期待できる．

本章では，この問題を解決するために，1.3 節で示した「(a) 翻訳対象（入力データ）」,「(b) 翻訳結果（出力データ）」,「(c) 翻訳対象と翻訳結果との関係」の３つの観点で対訳データの特徴を適切に区別して学習するための複数のドメインタグを用いた手法を提案する．近年重要性が高まっている日英ニュース機械翻訳を対象に翻訳実験を行い，提案手法が従来手法のドメインタグ手法と比較して翻訳精度を向上させることを確認した．

本章の構成は以下の通りである．4.2 節では本章が用いる日英ニュースコーパスについて説明する．4.3 節ではドメインタグを用いたニューラル機械翻訳の従来手法について説明する．4.4 節では提案手法を説明する．4.5 節では翻訳実験の詳細を示し，4.6 節で翻訳実験の結果を示して分析を行う．最後に，4.7 節で本章をまとめる．

4.2 日英ニュースコーパス

本節では，実験で用いる日英ニュースコーパスについて説明する．時事通信社のニュースを基とする特徴の違う 3 種類の日英ニュースコーパスを用いた．日本語側のデータを 1 文ごとに内容が等価になるように英語に翻訳したコーパス（内容等価コーパス），既存の日本語側と英語側のデータ同士を文単位で自動対応づけしたコーパス（自動アライメントコーパス），目的言語である英語のデータを機械翻訳で日本語に翻訳したコーパス（逆翻訳コーパス），である．表 4.1 に，用いた 3 種類のコーパスの特徴を示し，表 4.2 に各コーパスの統計情報を示す．本章では，内容等価コーパスを，適切な翻訳が行われたインドメインの対訳データとした．

コーパス	日本語側	英語側	日本語側と英語側のデータ間の関係
内容等価コーパス	元のニュース	内容が等価になるように翻訳	内容が等価である
自動アライメントコーパス	元のニュース	元のニュース	訳抜けや過剰訳などが含まれる
逆翻訳コーパス	機械翻訳出力	元のニュース	機械翻訳特有の訳抜け，過剰訳，誤訳が含まれる

表 4.1: 各日英ニュースコーパスの特徴

コーパス	データ数	日本語側		英語側	
		文数	文字数	文数	単語数
内容等価コーパス	220,180	220,180	10,427,732	235,407	6,052,647
自動アライメントコーパス	286,247	286,247	17,521,620	286,247	9,090,966
逆翻訳コーパス（内容等価）	533,581	533,581	22,798,841	533,581	14,348,304
逆翻訳コーパス（アライメント）	533,581	533,581	21,861,647	533,581	14,348,304

表 4.2: 各コーパスの統計情報．逆翻訳コーパスについては，機械翻訳モデルの学習時に内容等価コーパスと自動アライメントコーパスを用いて内容等価コーパスのコーパスの特徴に合わせて翻訳するモデルと，自動アライメントコーパスに合わせて翻訳するモデルの 2 つのモデルでそれぞれ翻訳した．

4.2.1 内容等価コーパス

内容等価コーパスは，日本語のデータを内容が同等となるように人手で英語に翻訳して構築されたコーパスである．本章では，この翻訳を適切な翻訳と定義し，本コーパスの対訳データをインドメインのデータとした．

内容等価コーパスは，次節で説明する自動アライメントコーパスと同様に，田中英輝 他 (2021) のデータ[¶]を用いた．データ数は 220,180 である．

[¶]田中英輝 他 (2021) は，自動アライメントコーパスがノイズを含んでいることに着目し，ノイズを含まないコーパスとして，時事通信社の日本語ニュースおよび英語ニュースをそれぞ

4.2.2 自動アライメントコーパス

自動アライメントコーパスは，日本語側のコーパスと英語側のコーパスの中から翻訳の対応関係にあるかどうかを自動で判定して構築されたコーパスである．公開されている自動アライメントコーパスには，時事コーパス‖や読売新聞コーパス**などがある．本章では，自動アライメントコーパスとして，田中英輝 他 (2021) が構築した時事通信社の自動アライメントデータを用いた．時事通信社の多くの英語ニュースは日本語ニュースを基に翻訳されているが，単なる翻訳ではなく内容の大きな編集が伴う場合があるため，英語側の文は情報が省略されていたり，追加されていたりすることがあり，自動アライメントコーパスにはこれらのノイズが含まれる．さらに，自動アライメントする際のアライメントエラーによるノイズも含まれる．自動アライメントは，動的計画法を用いて原言語側と目的言語側の文の類似度スコアを計算してアライメントを行う Utiyama and Isahara (2007) の手法を用いている．アライメントの精度の問題などにより内容が同等ではない対訳データが生成されることがあり，これをアライメントエラーと呼んでいる．これらのノイズはニューラル機械翻訳の翻訳精度の低下の要因となる．用いた自動アライメントコーパスのデータ数は，アライメントされた 582,572 文対の対訳データ候補のうち，類似度スコアが 0.3 以上であった 286, 247 文対である．

4.2.3 逆翻訳コーパス

ニューラル機械翻訳の精度を向上させる手法の 1 つに、単言語データを用いたデータ拡張がある。Sennrich et al. (2016a) は、既存の対訳データで学習した目的言語から原言語へ翻訳するニューラル機械翻訳器を用いて目的言語の単言語データを原言語に翻訳した擬似的な対訳データを生成し、それを既存の対訳データに追加して学習することで翻訳精度が向上することを示した。目的言語から原言語に翻訳することは「逆翻訳」と呼ばれ、逆翻訳によって生成された擬似対訳データは、目的言語側のデータの流暢性が担保されていることから、訳文の流暢性が向上することが知られている。本章では，時事通信社の英語ニュース記事の英語文のデータと，そのデータを日本語に逆翻訳したデータとを逆翻訳コーパスとして用いた．逆翻訳に用いた英語文は自動アライメントコーパスに含まれていない時事通信社のニュース記事の英語文とした．また，逆翻訳に用いた英日機械翻訳モデルには，内容等価コーパスと自動アライメン

れ人手で 1 文ごとに内容が等価になるように翻訳した内容等価コーパスを構築した．本章では，このうち時事通信社の日本語ニュースを人手で翻訳した対訳データのみを内容等価コーパスとして用いた．

‖http://lotus.kuee.kyoto-u.ac.jp/WAT/jiji-corpus/

**https://database.yomiuri.co.jp/about/glossary/

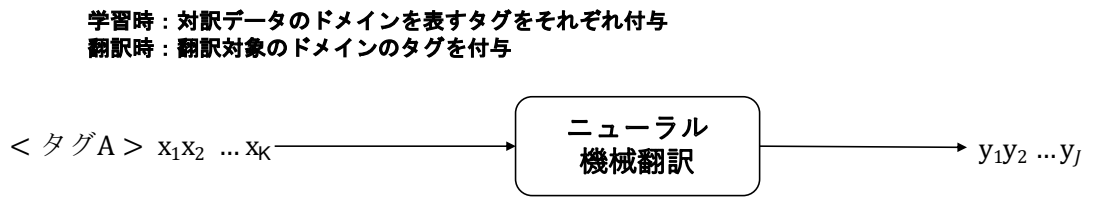

図 4.2: ドメインタグを用いたニューラル機械翻訳．各データのドメインのタグ（<タグA>，<タグB>，...）を原言語側のデータに付与して学習し，翻訳時は翻訳対象のドメインのタグを付与して翻訳する．原言語のデータを $\boldsymbol{x} = (x_1, x_2, ..., x_K)$，目的言語側のデータを $\boldsymbol{y} = (y_1, y_2, ..., y_J)$ で示す．K，J は原言語側，目的言語側のデータ長を示す．

トコーパスの両方を用い，ドメインタグを用いた手法（次節にて詳細を説明）で学習したモデルを用いた．翻訳時には内容等価コーパス，自動アライメントコーパスのいずれかに合わせた翻訳結果を出力可能であり，両方の出力を用いた．両者を区別するために，内容等価コーパスに合わせた翻訳結果を用いた対訳データを逆翻訳コーパス（内容等価），自動アライメントコーパスに合わせた翻訳結果を用いた対訳データを逆翻訳コーパス（アライメント）と呼ぶ．

4.3 ドメインタグを用いた従来手法

4.2節で示した対訳コーパスを用いて翻訳モデルを効果的に学習するために，ドメインタグを用いたニューラル機械翻訳を用いる．ドメインタグを用いたニューラル機械翻訳は，翻訳対象のドメイン（インドメイン）のデータと翻訳対象ではないドメイン（アウトオブドメイン）の違いを考慮して学習することで翻訳精度を向上させる手法であり，本章ではドメインタグを用いた手法を用いる．ドメインタグを用いた手法は，図4.2に示すように，各コーパスのデータにコーパスの特徴を表すドメインタグを原言語側の入力データに付与して学習，翻訳を行う．

4.3.1 従来手法が付与するドメインタグ

従来手法のドメインタグを用いたニューラル機械翻訳で学習時に付与するタグは下記の通りである．

- 単一ドメインタグを付与する手法 (Kobus et al. 2017)（コーパスタグ）：
  <CE> : 内容同等コーパスと逆翻訳コーパス（内容等価）に付与

<AA> : 自動アライメントコーパスと逆翻訳コーパス（アライメント）に付与

- 逆翻訳であることを示すドメインタグを付与する手法 (Vaibhav et al. 2019)（ノイズタグ）:
  <CE> : 内容同等コーパスに付与
  <AA> : 自動アライメントコーパスに付与
  <BT-CE> : 逆翻訳コーパス（内容等価）に付与
  <BT-AA> : 逆翻訳コーパス（アライメント）に付与

- 2 種類のドメインタグを付与する手法 (Berard et al. 2019b)（ 2 つのタグ）:
  <BT> : 逆翻訳コーパス（内容等価）と逆翻訳コーパス（アライメント）に付与
  <CE> : 内容同等コーパスに付与
  <AA> : 自動アライメントコーパスに付与

表 4.3 の 1 ～ 3 行目に，従来手法により日英ニュースコーパスに付与されるドメインタグを示す.

4.3.2 従来手法の課題

ドメインタグを用いた従来手法は，コーパスの持つ特徴に着目したコーパスタグや，コーパスの翻訳精度に着目したノイズタグなどの特徴を区別することで翻訳モデルを効果的に学習するための手法である．しかし，前節で示した従来手法では，表 4.1 で示した本章で用いる日英ニュースコーパスが持つ特徴を十分に区別することができない．例えば，逆翻訳コーパス (内容等価) に着目すると，表 4.3 のコーパスタグ (Kobus et al. 2017) では <CE> タグを付与しているが，逆翻訳による機械翻訳特有のノイズが含まれているという特徴を反映できていない．ノイズタグ (Vaibhav et al. 2019) では，<BT-CE> タグを付与することで機械翻訳特有のノイズが含まれていることが分かるようになっているが，逆翻訳に内容同等コーパスのドメインに合わせて翻訳する機械翻訳モデルを用いたという特徴が反映できていない．2 つのタグ (Berard et al. 2019a) では，<BT> と <CE> の 2 つのタグを付与することで，ノイズが含まれており，かつ，逆翻訳に内容同等コーパスのドメインに合わせて翻訳する機械翻訳モデルを用いたという特徴が反映されているが，目的言語側の英語文が時事通信社の元の英語文であるという特徴が反映できていない．逆翻訳コーパス (アライメント) に着目した場合も同様に，従来手法で付与するタグではコーパス間の特徴の違いを十分に区別できないことが分かる.

ドメインタグ	内容等価 コーパス	自動アライメント コーパス	逆翻訳 コーパス （内容等価）	逆翻訳 コーパス （アライメント）
コーパスタグ (Kobus et al. 2017)	<CE>	<AA>	<CE>	<AA>
ノイズタグ (Vaibhav et al. 2019)	<CE>	<AA>	<BT-CE>	<BT-AA>
２つのタグ (Berard et al. 2019a)	<CE>	<AA>	<BT> <CE>	<BT> <AA>
複数タグ （提案手法）	<NS-S> <CE-T> <NO-BT> <NO-AN>	<NS-S> <NS-T> <NO-BT> <AN>	<CE-S> <NS-T> <BT> <NO-AN>	<NS-S> <NS-T> <BT> <AN>

表 4.3: 各手法がコーパスに付与するタグ.

4.4 提案手法

従来手法の問題点を解決するには，表 4.1 の各コーパスの特徴を全て表現できるドメインタグの設計が必要となる．そこで，対訳データに対して，1.3 節で示した「(a) 翻訳対象（入力データ）」,「(b) 翻訳結果（出力データ）」,「(c) 翻訳対象と翻訳結果との関係」の特徴を表現するための複数タグを用いたドメインタグの手法を提案する．提案手法では，下記に示す４つのタグを用いる．

- 原言語側のデータの特徴を表すタグ（翻訳対象）:
  <NS-S>: 原言語側のデータが元のニュースであることを示す.
  <CE-S>: 原言語側のデータが逆翻訳により生成された機械翻訳結果であることを示す.

- 目的言語側の特徴を表すタグ（翻訳結果）:
  <CE-T>: 目的言語側のデータが内容が等価になるように元のニュースから人手で翻訳されたことを示す.
  <NS-T>: 目的言語側のデータが元のニュースであることを示す.

- 逆翻訳による機械翻訳特有のノイズの有無を表すタグ（翻訳対象と翻訳結果との関係）:

<NO-BT> : 機械翻訳を用いた対訳データではないこと示す.
<BT> : 機械翻訳を用いた対訳データであることを示す.

- 自動アライメントによるノイズの有無を表すタグ（翻訳対象と翻訳結果との関係）:
  <NO-AN> : 自動アライメントによるノイズがないことを示す.
  <AN> : 自動アライメントによるノイズが含まれることを示す.

表 4.3 の 4 行目に，本章で用いる日英ニュースコーパスに提案手法が付与するドメインタグを示す.

4.4.1 比較手法

提案手法と比較するドメインタグを用いた手法には，4.3 節で従来手法として示した「単一ドメインタグを付与する手法 (Kobus et al. 2017)（コーパスタグ）」,「逆翻訳であることを示すドメインタグを付与する手法 (Vaibhav et al. 2019)（ノイズタグ）」,「２つのドメインタグを付与する手法 (Berard et al. 2019b)（２つのタグ）」の３つの手法を用いた．ドメインタグを用いた従来手法との比較に加え，ドメインタグを用いない場合とも比較を行う.

4.5 評価実験

提案手法の有効性を評価するために，4.2 節で示した日英ニュースコーパスを用いた日英ニュース機械翻訳実験を行った.

4.5.1 詳細設定

学習データには，4.2 節で示した日英ニュースコーパスの合計 1,573,589 文対を用いた．テストセットには，内容等価コーパスからあらかじめ学習データとは別に取っておいた 2,000 文対を用いた．全てのデータは，下記の前処理を実施して用いた．英語のデータのクリーニング，および単語分割には Moses toolkit[*] の clean-corpus-n.perl[†]，tokenizer.perl[‡] を用いた．日本語の形態素解析には，KyTea (Neubig et al. 2011) を利用した．日本語，英語ともに，バイトペア符号化（byte pair encoding）を用いたサブワード (Sennrich et al. 2016b) を用いた．語彙数は 32,000 とし，頻度が 35 以下の語彙はサブワードに分割した．ニューラル

[*] https://github.com/moses-smt/mosesdecoder
[†] https://github.com/moses-smt/mosesdecoder/blob/master/scripts/training/clean-corpus-n.perl
[‡] https://github.com/moses-smt/mosesdecoder/blob/master/scripts/tokenizer/tokenizer.perl

機械翻訳には，トランスフォーマーモデル (Vaswani et al. 2017) のエンコーダとデコーダを用いた．ニューラル機械翻訳の実装には Sockeye toolkit (Hieber et al. 2018) を用いた．学習，翻訳ともに，Nvidia P100 Tesla GPU を用いて計算を行った．本章のモデルの学習には Adam (Kingma and Ba 2015) を optimizer として用い，学習率は 0.0002 とした．バッチ内のデータは 5,000 トークン以内になるように構築した．GPU メモリに制限があるため，学習時の各エンコーダへの入力の最大長は 200 トークンに制限した．翻訳時の各エンコーダへの入力の最大長は設定しなかった．その他のハイパーパラメータの設定は Sockeye toolkit のデフォルト値に従った．早期終了（Early stopping）を設定し，patience の値を 32 とした．翻訳時は，ビームサーチを用い，ビーム幅を 5 とした．式 5.3.1 のハイパーパラメータ k は 2 とした．翻訳性能の比較には，ランダムシードの異なる 5 つのモデルを学習してそれぞれのモデルが出力した翻訳結果の BLEU スコア (Papineni et al. 2002) の中央値を用いた．BLEU スコアには，case-insensitive BLEU-4 を用い，multi-bleu.perl[§] で計算した．サンプル数を 10,000 に設定した paired-bootstrap.py[¶] を用い，BLEU スコアにおける提案手法と他手法との有意差（有意水準 $\alpha = 0.05$）を調べた．

4.6 実験結果

4.6.1 提案手法の効果

表 4.4 に実験結果を示す．実験で用いた手法の中で最も翻訳精度が高かったのは提案手法を用いた場合であった．ドメインタグを用いないニューラル機械翻訳の結果との比較では，ドメインタグを用いた手法は，従来手法を含めいずれも有意に翻訳精度が向上した[*]．ドメインタグを用いた手法間の比較では，ドメインタグの種類を増やすことで翻訳精度が向上した．コーパスタグのみを用いた結果と比較して，ノイズタグ，２つのタグ，提案手法のタグを用いた結果が有意に翻訳精度が向上した．ノイズタグ，２つのタグを用いた結果と比較して提案手法の結果の翻訳精度は向上したが，有意ではなかった．

4.6.2 ドメインタグの効果の比較

本節では，ドメインタグを用いた手法間において付与したドメインタグの効果の比較を，コーパスの特徴の違いを基に考察する．表 4.4 より，タグ無しと比較して，コーパスタグの付与により BLEU スコア 2.05 の向上，ノイズタグの

[§] https://github.com/moses-smt/mosesdecoder/blob/master/scripts/generic/multi-bleu-detok.perl
[¶] https://github.com/neubig/util-scripts/blob/master/paired-bootstrap.py
[*] 有意水準 $\alpha = 0.05$．本章では全て同じ p 値を利用した．

ドメインタグ	BLEUスコア	Δ
タグ無し	20.36	-
コーパスタグ (Kobus et al. 2017)	22.41	+2.05
ノイズタグ (Vaibhav et al. 2019)	24.19	+3.83
２つのタグ (Berard et al. 2019a)	24.25	+3.89
提案手法	**24.56**	+4.20
７つのタグ	24.26	+3.90

表 4.4: ドメインタグを用いた日英ニュース翻訳の実験結果．学習データは全手法ともに 1,573,589 文対．Δ はタグ無しの手法との BLEU スコアの差分を示す．

付与により 3.83 の向上を確認した．表 4.3 より，両者のドメインタグの違いは逆翻訳コーパスに対して別のタグを付与しているかどうかであり，逆翻訳コーパス（内容等価）と内容等価コーパス，および逆翻訳コーパス（アライメント）と自動アライメントコーパスがそれぞれ別のドメインであるという情報（機械翻訳結果を用いているかどうかの情報）を持ったドメインタグを付与して明示的に学習させることで翻訳精度が向上したと考えられる．

２つのタグを用いた手法は，ノイズタグで逆翻訳コーパスに付与したドメインタグ（<BT-CE>，<BT-AA>）を，２つのタグ（<BT> <CE>，<BT> <AA>）に置き換えている．２つのタグに置き換えたことにより，逆翻訳コーパス（内容等価）と逆翻訳コーパス（アライメント）が機械翻訳されたコーパスであるという情報に加え，それぞれ内容等価コーパス，自動アライメントコーパスに合わせて機械翻訳されたコーパスであるという情報を持っている．２つのタグを用いた手法は，ノイズタグを用いた手法と比較して，コーパスの持つ情報をより多く表現したドメインタグといえるが，翻訳精度は BLEU スコア 3.89 の向上となっており，ノイズタグを用いた手法と比較して 0.06 の向上にとどまった．

提案手法は，２種類のタグを用いた手法と比較して，各コーパスに４つのタグを付与することで逆翻訳コーパス（内容等価）と逆翻訳コーパス（アライメント）の英語側のデータが元のニュースの英語文であるという情報が追加されている．翻訳精度は BLEU スコア 4.20 の向上となっており，ノイズタグを用いた手法との比較では 0.37，２種類のタグを用いた手法との比較では 0.31 となった．

上記の実験結果より，ドメインタグの数を増やせば増やすほど翻訳精度が向上したが，コーパスに付与するタグの違いによる翻訳精度の向上の程度は異なっており，翻訳精度の向上に寄与するタグの種類は限定的であると考えられ

る．そこで，コーパスの特徴を考慮せずに，用いた４つのコーパスの全ての組み合わせのタグを付与して同様の翻訳実験を行った．各コーパスに付与したタグは７つである†．表 4.4 の実験結果より，７つのタグを用いた場合の BLEU スコアは 24.26 となり，提案手法と比較して翻訳精度は低下した．

以上より，翻訳精度を向上させるためにはコーパスの特徴を考慮したタグを付与する必要があることが示唆される．

4.6.3 ドメインタグを行わない場合における特徴の違うコーパスの追加の効果

4.6.1 節と 4.6.2 節では，ドメインタグを用いない手法の学習データとして，翻訳対象のドメインであるインドメインと翻訳対象のドメインではないアウトオブドメインの両方の対訳データを用いたが，アウトオブドメインのデータを追加したことにより翻訳精度が低下している可能性がある．そこで，本節では，ドメインタグを用いない手法を用い，インドメインである内容等価コーパスをベースにアウトオブドメインのコーパスを学習データに追加して学習することで翻訳精度が向上するかどうかを確認する．表 4.2 にある４つのコーパスから，インドメインの内容等価コーパスのみ，インドメインの内容等価コーパスにアウトオブドメインの自動アライメントコーパスを追加したコーパス，インドメインの内容等価コーパスにアウトオブドメインの自動アライメントコーパスと逆翻訳コーパス（内容等価）,（アライメント）を追加したコーパス，の３種類のコーパスをそれぞれ構築して翻訳モデルを学習し，翻訳精度を確認した．参考のため，アウトオブドメインの自動アライメントコーパスのみで翻訳モデルを学習した場合の翻訳精度も確認した．

表 4.5 に結果を示す．内容等価コーパスのみを用いた場合の翻訳精度は 20.93 の BLEU スコアであったのに対して，自動アライメントコーパス，逆翻訳コーパスを追加することで翻訳精度が低下することを確認した．自動アライメントコーパスのみを用いた場合の BLEU スコアは 10.30 であり，内容等価コーパスのみを用いた場合と比較して，10 以上の BLEU スコアの低下を確認した．

以上より，ドメインタグを用いない場合は，同じジャンルの対訳データ（本章では日英ニュースの対訳データを利用）であっても，特徴の違うデータを加えることで翻訳精度が低下する場合があることを確認した．

† １４（${}_4C_1 + {}_4C_2 + {}_4C_3$）のタグから７つのタグを付与する．

学習に用いたコーパス	データ数	BLEU スコア
内容等価コーパス	0.22M	**20.93**
内容等価コーパス＋自動アライメントコーパス	0.51M	20.68
内容等価コーパス＋自動アライメントコーパス＋逆翻訳コーパス（内容等価）,（アライメント）	1.57M	20.36
自動アライメントコーパス	0.29M	10.30

表 4.5: ドメインタグを用いなかった日英ニュース翻訳の実験結果．インドメインのコーパスは内容等価コーパス．アウトオブドメインのコーパスは自動アライメントコーパスと逆翻訳コーパス．

4.6.4 インドメインのデータ数の違いによる提案手法の効果

4.6.2 節で示したように，本実験では，コーパスの特徴を考慮してタグを増やす提案手法により翻訳精度は向上したが，7 つのタグを用いた場合は翻訳精度が向上しなかった．これは，タグを増やすことでデータスパースネスの問題が発生したためだと考えられる．そこで，データスパースネスによる翻訳精度の低下が提案手法でも起こり得るかどうかを確認するために，インドメインのデータ数を減らすことでデータスパースネスが発生しやすい状況を作り，タグ無しと提案手法の実験を行った．実験には，表 4.2 のインドメインの内容等価コーパスのデータ（総データ数は 220,180）からランダムに $100K, 50K, 10K, 5K, 1K$ ずつを抽出したデータを用いた．

表 4.6 に実験結果を示す．インドメインのデータを減らしていくにつれて，翻訳精度は下がったが，タグ無しと比較して翻訳精度が低くなることはなかった．また，BLEU スコアの差分値は，インドメインのデータ数を $220K$ から $5K$ まで少なくしても大きく変わらなかった．インドメインのデータを $1K$ にした場合の差分は +2.57 となり，翻訳精度の向上幅は小さくなった．日英ニュースコーパスを用いた本実験下ではタグを付与することでタグ無しと比較して翻訳精度が低下することはなく，提案手法ではデータスパースネスによる翻訳精度の低下は起こりにくいと考えられるが，アウトオブドメインのデータの質によってこの傾向は変わる可能性がある．

4.6.5 インドメインとアウトオブドメインのデータの分析

4.2 節で示した通り，アウトオブドメインのデータにはノイズが含まれているが，どの程度のノイズが含まれているかは明らかではない．そこで，本節で

	データ数					
ドメインタグ	220K	100K	50K	10K	5K	1K
タグ無し	20.36	18.30	17.01	14.75	14.34	14.31
提案手法	24.56	22.20	21.33	19.03	18.22	16.88
Δ	+4.20	+3.90	+4.32	+4.28	+3.88	+2.57

表 4.6: 学習に用いたインドメイン（内容等価コーパス）のデータ数の違いによる提案手法の日英ニュース翻訳の実験結果．220*K* は元のインドメインのデータ全てを用いた場合の実験結果．Δはタグ無しとの差分を示す．

は，アウトオブドメインの中の 1 ドメインに合わせて機械翻訳した結果を用いてアウトオブドメインのデータの質を分析する[‡]．ドメインタグを用いた手法には，最も翻訳精度の高かった提案手法を用いた．本章では内容等価コーパスをインドメインのデータ，自動アライメントコーパスと逆翻訳コーパスをアウトオブドメインのデータとしたが，本節ではアウトオブドメインのデータのうち人手で編集して翻訳されたデータが含まれる自動アライメントコーパスに合わせて機械翻訳した[§]．学習データ，およびテストセットには，4.5.1 節で示したデータを用いた．評価尺度には，BLEU スコアに加え，人手評価を実施した．人手評価は，テストセット 2,000 文対のうちランダムに抽出した 300 文に対して訳抜けと過剰訳が含まれる単語数を評価者 1 名がカウントした．

表 4.7 に実験結果を示す．アウトオブドメイン（自動アライメントコーパス）の特徴に合わせて翻訳した場合の BLEU スコアは 23.21 となり，インドメイン（内容等価コーパス）の特徴に合わせて翻訳した場合の BLEU スコアとの差分は 1.35 となった．また，人手評価の結果より，訳抜けと過剰訳の数がそれぞれ自動アライメントコーパスの特徴に合わせて翻訳することで増加しており，この増加が BLEU スコアの差に影響していると考えられる．以上より，自動アライメントコーパスは，訳抜けと過剰訳を含んだコーパスであることが確認できた．この特徴は自動アライメントのアライメントエラーによる可能性もあるが，人手で編集して翻訳する際には一定の情報の追加や削除があることが示唆される．

より詳細な分析のため，インドメイン（内容等価コーパス）とアウトオブド

[‡]本来は用いた対訳データ自身を人手などで比較・分析する必要があるが，コーパス間で翻訳元のデータが違っており，厳密な比較が難しいため，同一のニュースデータ（テストセット）をインドメイン，アウトオブドメインそれぞれに合わせて機械翻訳した結果を比較して分析する．

[§]表 4.3 の自動アライメントコーパスの提案手法のタグ（<NS-S> <NS-T> <NO-BT> <AN>）を付与して翻訳した．

		人手評価	
	BLEU	訳抜け	過剰訳
インドメイン	24.56	4.27	1.44
（内容等価コーパスに合わせて翻訳）			
アウトオブドメイン	23.21	18.40	9.39
（自動アライメントコーパスに合わせて翻訳）			

表 4.7: 提案手法を用いてインドメイン（内容等価コーパス），およびアウトオブドメイン（自動アライメントコーパス）の特徴にそれぞれ合わせて機械翻訳した実験結果．学習データはともに 1,573,589 文対．BLEU スコアはテストセット 2,000 文，人手評価はテストセットからランダムに抽出した 300 文を対象に評価した．人手評価の結果は 100 単語ごとの平均値を示す．

メイン（自動アライメントコーパス），それぞれの特徴に合わせて機械翻訳した結果例を表 4.8 に示す．#1 と #2 は自動アライメントコーパスの特徴に合わせて機械翻訳した結果に訳抜けが生じた例，#3 は自動アライメントコーパスの特徴に合わせて機械翻訳した結果に過剰訳が生じた例である．#1 では，組織名の略称と地域の国名が訳抜けし，#2 では，日本語文の前半部分が訳抜けしている．#3 では，北海道，東北，近畿などの地域名に補足情報が過剰訳として翻訳結果に湧き出ている．これらの訳抜けや過剰訳は人手で編集して翻訳する際の特徴であり，これらの特徴を反映した翻訳結果が生成されていると考えられる．

4.7 まとめ

本章では，翻訳対象ではないドメイン（アウトオブドメイン）のコーパスを活用してニューラル機械翻訳の翻訳精度を向上させるために複数のドメインタグを用いた手法を提案した．原言語側（日本語）のニュースデータを 1 文ごとに内容が等価になるように翻訳した対訳データのコーパス（内容等価コーパス），原言語側（日本語）と目的言語側（英語）のニュースデータ同士を自動で文単位で対応づけした対訳データのコーパス（自動アライメントコーパス），目的言語側（英語）のニュースデータを機械翻訳で原言語（日本語）に翻訳した対訳データのコーパス（逆翻訳コーパス（内容等価）と逆翻訳コーパス（アライメント）の 2 種類のコーパス），の合計 4 種類の日英ニュースコーパスを用いて日英ニュース機械翻訳の実験を行い，コーパスの特徴を考慮して複数の

#1	日本語文	アジア開発銀行（ＡＤＢ）の年次総会など一連の国際会議が５月２日、フィリピンのマニラで始まる。
	翻訳結果（インドメイン）	A series of international meetings including the annual meeting of the Asian Development Bank **(ADB)** will start on May 2 in Manila, **Philippines**.
	翻訳結果（アウトオブドメイン）	A series of international meetings, including an annual meeting of the Asian Development Bank, will start in Manila on May 2.
#2	日本語文	新幹線は車体軽量化の効果もあり、高架橋や鉄橋で深刻な損傷は今のところ見つかっていない。
	翻訳結果（インドメイン）	**The Shinkansen has the effect of reducing body weight,** and so far no serious damage has been found on elevated bridges and iron bridges.
	翻訳結果（アウトオブドメイン）	So far, no serious damage has been found on elevated tracks and bridges.
#3	日本語文	東海・中京圏と北海道、東北はいずれも２桁増、近畿圏は微増となった。
	翻訳結果（インドメイン）	The Tokai and Chukyo regions, Hokkaido and Tohoku saw double-digit growth, and the Kinki region saw a slight increase.
	翻訳結果（アウトオブドメイン）	The Tokai-Chukyo central region and **the northernmost prefecture of** Hokkaido and the Tohoku **northeastern** region saw double-digit growth, while the Kinki **western** region saw a slight increase.

表 4.8: 翻訳結果例．インドメインは内容等価コーパスの特徴に，アウトオブドメインは自動アライメントコーパスの特徴に，それぞれ合わせて機械翻訳した結果を示す．

ドメインタグを付与する提案手法の効果を確認した．今後の課題として，翻訳精度向上に寄与するドメインタグの選定手法の検討がある．また，他のデータセットを用いた実験を行い，アウトオブドメインのデータの質と提案手法の効果との関連性を明らかにする必要もある．

第5章

目的言語側の前文を用いた文脈情報考慮型ニューラル機械翻訳

本章では，文脈情報考慮型ニューラル機械翻訳を扱う．文脈情報考慮型ニューラル機械翻訳とは，翻訳対象の文に加えて文脈の情報を用いて翻訳を行うニューラル機械翻訳のことである．前章のニューラル機械翻訳は翻訳対象の文のみを用いたモデルであったのに対し，本章が扱う文脈情報考慮型ニューラル機械翻訳は翻訳対象以外の外部情報を活用したモデルである．文脈の情報には，翻訳元の言語（原言語）や翻訳先の言語（目的言語側）の翻訳対象の周辺の文を用いることが多く，目的言語側の周辺の文を用いる手法は翻訳精度が下がることが報告されている．しかし，翻訳結果の文末表現，訳語の統一，翻訳結果のゼロ代名詞の推定などを実現するためには，目的言語側の周辺の文の情報の利用が有効であると考えられる．そこで，目的言語側の前文を文脈情報として用いてニューラル機械翻訳の精度を向上する手法を提案する．目的言語側の前文を文脈情報として利用した従来の文脈考慮型ニューラル機械翻訳は，学習時は目的言語側の前文の参照訳を用い，翻訳時は目的言語側の前文の機械翻訳結果を用いており，学習時と翻訳時とで特徴が違う外部データを用いていることが翻訳精度の低下の原因であると考えられる．本章では，学習時と翻訳時とで用いる目的言語側の文脈情報の特徴の違いによる影響を緩和するために，学習時に用いる目的言語側の文脈情報を学習の進行に応じて参照訳から機械翻訳結果へ段階的に切り替えていく手法を提案する．時事通信社のニュースコーパスを用いた英日・日英機械翻訳タスクと，IWSLT2017 の TED トークコーパスを用いた英日・日英，および英独・独英機械翻訳タスクの評価実験により，従来の目的言語側の文脈を利用した機械翻訳モデルと比較して，翻訳精度が向上することを確認した．

5.1 はじめに

ニューラル機械翻訳は通常，原言語一文を入力して目的言語への翻訳結果を出力するが，さらに翻訳精度を上げるために翻訳対象の周辺の文を文脈情報として活用する手法がある (Maruf et al. 2019; Voita et al. 2019; Agrawal et al. 2018; Bawden et al. 2018; Kuang et al. 2018; Läubli et al. 2018; Miculicich et al. 2018; Tiedemann and Scherrer 2017; Wang et al. 2017)．本章では，特に目的言語側の前文を活用したニューラル機械翻訳の改善手法を提案する．文脈情報を用いる手法には原言語側や目的言語側の周辺の文を用いる手法があるが，文脈情報に目的言語側の周辺の文を用いる手法は翻訳精度が下がることが報告されている (Bawden et al. 2018)．この翻訳精度の低下は，翻訳モデルの学習時と翻訳時で用いる周辺の文の特徴の違いが原因の 1 つと考えられる．従来研究では，学習時は目的言語側の周辺の文の参照訳を使用し，翻訳時は翻訳モデルによって生成された機械翻訳結果を用いているが，機械翻訳結果には，参照訳には含まれない訳抜け，過剰訳，誤訳などの機械翻訳特有の誤り（機械翻訳誤り）が含まれる可能性がある．また，機械翻訳誤りが無い場合でも，機械翻訳結果は，翻訳調（Translationese）と呼ばれる偏りのある文となっている．Toral (2019) は，機械翻訳結果が人手翻訳で作られた翻訳文と比較して単純で標準的な翻訳になる傾向があることを示している．学習時と翻訳時とで用いる周辺の文特徴が違うことは，1.3 節で示した学習時と翻訳時との間の特徴の違いを分類した 4 つの観点のうち「(d) 外部情報（翻訳対象とは別の入力データ）」の特徴の違いに該当している．学習時と翻訳時で特徴が違うデータを用いることで生じるモデルの偏りは，露出バイアス（exposure bias）と呼ばれており (Ranzato et al. 2016)，翻訳品質の低下に繋がることが報告されている (Zhang et al. 2019)．目的言語側の周辺の文の利用において，学習時に参照訳を用い，翻訳時に機械翻訳誤りを含み，翻訳調の特徴を有する機械翻訳結果を用いることは，露出バイアス（exposure bias）による翻訳低下を引き起こすと考えられる．下記の (a)，(c)，(d) は，IWSLT2017 (Cettolo et al. 2017) で提供されている日英対訳データセット (Cettolo et al. 2012) から抽出したデータ例である．翻訳対象の原言語文 (c)，翻訳対象の参照訳 (d)，翻訳対象の機械翻訳結果 (e) に加え，文脈情報として，前文の参照訳 (a) と機械翻訳結果 (b) を示している．機械翻訳結果は，日英対訳データセットをトランスフォーマーモデル (Vaswani et al. 2017) で学習した機械翻訳器を用いて翻訳した．

(a) 前文の参照訳： **She** lays eggs, **she** feeds the larvae – so an ant starts as an egg, then it's a larva.
(b) 前文の機械翻訳結果： And when they get their eggs, they get their eggs, and the **queen** is there.
(c) 翻訳対象の原言語文： 脂肪を吐き出して幼虫を育てます

(d) 翻訳対象の参照訳： **She** feeds the larvae by regurgitating from her fat reserves.
(e) 翻訳対象の機械翻訳結果： **They** take fat and they raise the larvae of their larvae.

翻訳対象の原言語文 (c) では主語が省略されているため，この文のみで主語を推定することができず参照訳 (d) を生成することは困難である．翻訳対象の機械翻訳結果 (e) では，“she” ではなく誤った主語 “They” を用いて翻訳されている．しかし，前文の参照訳 (a)，または機械翻訳結果 (b) には主語を推定する情報（参照訳には “she”，機械翻訳結果には “queen”）が含まれており，目的言語側の前文を使うことで正しい翻訳ができる可能性がある．一方で，前文の参照訳 (a) と機械翻訳結果 (b) を比較すると機械翻訳結果には誤訳が含まれており，参照訳と機械翻訳結果との間に特徴の違いがあることが分かる．誤訳のない参照訳のみを文脈情報として学習した翻訳モデルは文脈に含まれる誤訳に頑健でないと考えられ，参照訳と機械翻訳を学習時と翻訳時で別々に用いる手法は露出バイアス（exposure bias）による翻訳精度低下を引き起こす可能性がある．

本章では，目的言語側の文脈から目的言語側の前文を用いた文脈情報考慮型ニューラル機械翻訳を対象にする*．上記の翻訳精度低下の問題に対し，スケジュールドサンプリング法 (Bengio et al. 2015) を参考にして学習時と翻訳時の目的言語側の前文の特徴の違いの影響を緩和するための学習データ制御手法を提案する．具体的には，初期の学習では従来手法と同様に文脈情報として目的言語側の前文に参照訳のみを用い，学習が進行するにつれて，段階的に参照訳から機械翻訳結果へ切り替える．この処理により，翻訳モデルが機械翻訳誤りに頑健になり，翻訳調の特徴に対応できるようになると期待できる．機械翻訳結果は，対訳データをトランスフォーマーモデルで学習した機械翻訳器で生成する．実験では，提案手法を結合ベース文脈考慮型ニューラル機械翻訳 (Bawden et al. 2018; Tiedemann and Scherrer 2017) とマルチエンコーダ文脈考慮型ニューラル機械翻訳 (Kim et al. 2019; Bawden et al. 2018) で実装し，ニュースコーパス (田中英輝 他 2021) を用いた英日・日英機械翻訳タスク，および IWSLT2017 データセット (Cettolo et al. 2012) を用いた英日・日英，および英独・独英機械翻訳タスクで評価した．その結果，提案手法が従来手法と比較して BLEU スコア (Papineni et al. 2002) により翻訳精度が改善していることを確認した．

本章の構成は以下の通りである．まず，5.2 節で提案手法が前提とする文脈考慮型ニューラル機械翻訳について述べ，5.3 節で提案手法を説明する．5.4 節で提案手法の効果を確認するための翻訳実験の詳細を示し，5.5 節で翻訳実験の結果を示して提案手法の効果を分析する．最後に 5.6 節で本章をまとめる．

*文脈情報として翻訳対象以外の文（翻訳対象の前後の複数文）の利用も考えられるが，用いる文脈情報が増えると翻訳モデルが大きくなるという課題がある．計算機の性能を考慮し，本章では目的言語側の文脈を目的言語側の前文に限定した．

5.2 文脈考慮型ニューラル機械翻訳

本節では，文脈考慮型ニューラル機械翻訳モデルの従来手法である，結合ベース文脈考慮型ニューラル機械翻訳と，マルチエンコーダ文脈考慮型ニューラル機械翻訳を説明する．

5.2.1 結合ベース文脈考慮型ニューラル機械翻訳

結合ベース文脈考慮型ニューラル機械翻訳 (Tiedemann and Scherrer 2017; Bawden et al. 2018) は，ニューラル機械翻訳モデルの構造を変えずに，翻訳対象の文と文脈情報との間に特殊なトークン（*_BREAK_*）を挿入して結合してエンコーダに入力する．翻訳時に，翻訳対象の文の翻訳結果のみを出力するモデルを 2-to-1 モデル，翻訳対象の文に加えて文脈情報を併せて翻訳するモデルを 2-to-2 モデル[†]と呼ぶ．2-to-2 モデルでは，翻訳対象の文の翻訳結果と文脈情報の翻訳結果の間に特殊なトークンを併せて出力させることで，出力から翻訳対象の文の翻訳結果のみを抽出する．李凌寒 他 (2019) は，新聞データを用いた日英・英日機械翻訳実験で 2-to-2 モデルと 2-to-1 モデルの比較を行い，2-to-1 モデルの方が翻訳精度が向上したことを報告しており，本章でも事前実験の結果も踏まえて 2-to-1 モデルを用いた．

5.2.2 マルチエンコーダ文脈考慮型ニューラル機械翻訳

マルチエンコーダ文脈考慮型ニューラル機械翻訳 (Bawden et al. 2018; Kim et al. 2019) は，翻訳対象の文を入力とするエンコーダと，文脈情報を入力とするエンコーダを持つ．複数のエンコーダの出力ベクトルは結合する必要があり，複数の出力ベクトルを「どこで結合させるか」と,「どのように結合させるか」の違いでいくつかのバリエーションがある.「どこで結合させるか」は，エンコーダの出力ベクトルをデコーダの内側と外側のどちらで結合するかを表す．Kim et al. (2019) は，エンコーダの出力ベクトルをデコーダの内側で結合したモデルの方が翻訳精度が高いことを示しており，本章でもデコーダの内側で結合する手法を用いた.「どのように結合させるか」は，連結（attention concatination），重み付き和（attention gate），階層（hierarchical attention）の 3 種類がある (Bawden et al. 2018)．本章では，事前実験で BLEU スコアの高かった重み付き和を用いた．重み付き和は，翻訳対象のベクトル $\mathbf{c_s}$ と文脈のベクトル $\mathbf{c_r}$

[†]2-to-1 は片言語の文脈があれば適用可能であるのに対し，2-to-2 は対訳の文脈が必要となり，適用範囲が異なる．

を用いて下記の式によりベクトル $\mathbf{c}$ を得る (Bawden et al. 2018).

$$\mathbf{g} = \tanh(W_r\mathbf{c_r} + W_s\mathbf{c_s}) + \mathbf{b}, \tag{5.2.1}$$

$$\mathbf{c} = \mathbf{g} \odot (W_t\mathbf{c_r}) + (\vec{1} - \mathbf{g}) \odot (W_u\mathbf{c_s}), \tag{5.2.2}$$

$\mathbf{g}$ は各ベクトルの要素の重要度の違いを表すベクトルである．$\mathbf{b}$ はバイアスベクトル，W_r, W_s, W_t, W_u は重み行列である．$\odot$ はアダマール積を計算する．

5.3 提案手法

本節では，文脈を考慮した機械翻訳手法の一手法である目的言語側の文脈を考慮する機械翻訳における，学習時と翻訳時に用いる目的言語側の文脈情報の特徴の違いによる影響を低減することを目的としたサンプリング手法について説明する．

5.3.1 利用する文脈情報

学習時にニューラル機械翻訳のエンコーダに入力する目的言語側の文脈情報として，前文の参照訳と機械翻訳結果の両方を用い，前節で示した 2 種類のアーキテクチャに適用する．翻訳対象が文書の先頭のときは，前文の情報がないため，特殊なトークン（*_BOS_*）を文脈情報として用いる．図 5.1（a）は結合ベース文脈考慮型ニューラル機械翻訳に適用したモデルを示し，図 5.1（b）はマルチエンコーダ文脈考慮型ニューラル機械翻訳に適用したモデルを示している．翻訳対象である，文書中の j 番目の原文のトークン列は X^j で表す．文脈情報として用いる，文書中の $j-1$ 番目の参照訳のトークン列は $\hat{Y}^{j-1}$ で表し，$j-1$ 番目の機械翻訳結果のトークン列は $\bar{Y}^{j-1}$ で表す．文脈情報として用いる機械翻訳結果は，学習時と翻訳時ともに文脈を考慮しないニューラル機械翻訳で生成する．図 5.1 の p は 1 エポックの学習データに含まれる参照訳の割合を示し，$1-p$ が機械翻訳結果が含まれる割合となる．$p = 0$ の場合は入力される文脈情報データの全てが機械翻訳，$p = 1$ の場合は文脈情報データの全てが参照訳で学習されることになる．翻訳時は $p = 0$，すなわち常に機械翻訳結果が文脈情報として用いられる．結合ベース文脈考慮型ニューラル機械翻訳（図 5.1（a））では，エンコーダが出力したベクトルがデコーダに入力される．マルチエンコーダ文脈考慮型ニューラル機械翻訳（図 5.1（b））では，原文エンコーダと文脈エンコーダが出力した 2 つのベクトルを 1 つのベクトルに変換してからデコーダに入力し，デコーダを通して目的言語側の文のトークン列 Y^j を得る．

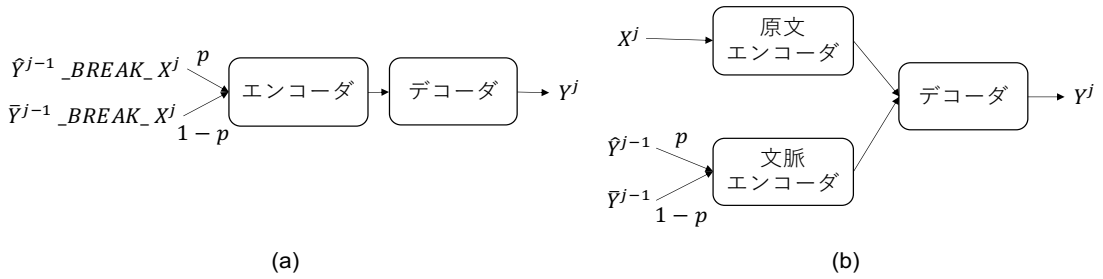

図 5.1:（a）結合ベース文脈考慮型ニューラル機械翻訳,（b）マルチエンコーダ文脈考慮型ニューラル機械翻訳

5.3.2 サンプリング制御

学習データ内の目的言語側の文脈情報として，参照訳と機械翻訳結果のどちらを用いるかをサンプリングレート p により制御するが，タスクや学習時の翻訳モデルの性能によって適切なサンプリングレートは異なる．学習時において，目的言語側の文脈情報に機械翻訳結果を使う場合ではエンコードされた文脈情報ベクトルは訳抜け，過剰訳，誤訳など機械翻訳特有の誤り（機械翻訳誤り）を含み，翻訳モデルはこれらの機械翻訳誤りを考慮して学習される．一方，参照訳のみを使う場合では機械翻訳誤りがないので，文脈情報を効果的に利用するように翻訳モデルが学習されることが期待できる．翻訳時において，目的言語側の文脈情報に機械翻訳誤りのない参照訳を使うことができる場合では，学習時も同様に参照訳を使うことが望ましい．しかし，翻訳時には参照訳ではなく機械翻訳結果を用いるため，翻訳モデルは訳抜け，過剰訳，誤訳などが含まれた文脈を利用して翻訳することになる．さらに，機械翻訳結果は，機械翻訳誤りがない場合でも，翻訳調（Translationese）と呼ばれる偏りのある文であることが知られている．Toral (Toral 2019) は，機械翻訳は人手翻訳よりもシンプルかつ標準化された結果になる傾向にあると示しており，参照訳と機械翻訳結果は多様性の側面からも特徴が違うと考えられる．

翻訳時と同じ特徴を有する目的言語側の文脈情報を使うという観点では機械翻訳結果を用いることが好ましいが，効果的な文脈の利用という観点では参照訳を用いることが好ましい．そこで，双方の利点を活かすために，特徴の違う参照訳と機械翻訳結果の両方を用いて学習することを提案する．提案手法は，カリキュラム学習 (Bengio et al. 2009) とスケジュールドサンプリング (Bengio et al. 2015) の手法を参考にしてサンプリングを制御する．Bengio et al. (2015) は，学習時に，回帰型ニューラルネットワーク（Recurrent neural network）に入力する直前のトークンとしてモデルの出力と正解のどちらを用いるかをサンプリング制御したのに対し，本章ではこの手法を，目的言語側の文脈情報として機械

コーパス	言語対	訓練データセット		開発データセット		テストセット	
		データ数	記事数	データ数	記事数	データ数	記事数
ニュース	日英	220,180	28,043	2,000	253	2,000	256
コーパス							
TED トーク	英日	194,170	1,863	879	8	1,285	15
コーパス	英独	203,998	1,705	888	8	1,080	12

表 5.1: コーパスの統計情報.

翻訳の出力と正解のどちらを用いるかのサンプリング制御に応用する．この手法は，下記の手順により学習エポックごとに，入力に用いる文脈情報を変更することで，翻訳モデルを段階的に機械翻訳誤り，および翻訳調に対応させる．学習エポック数lに対応したサンプリングレートをp_lとする．

1. 1エポック目（$l = 1$）の学習はサンプリングレートを$p_1 = 1$とし，目的言語側の文脈情報には全て参照訳を用いる．
2. 2エポック目（$l \geq 2$）以降の学習はサンプリングレートp_lを下記の逆シグモイド関数により減少させる．

$$p_l = \frac{k}{k + \exp(l - 2/k)}, \tag{5.3.1}$$

$k(\geq 1)$はpの減少スピードを制御するハイパーパラメータである．1エポック終了時の翻訳モデルの性能が高いのであればkの値を1に近づけ，2エポック目以降の学習データ内の機械翻訳の割合を増やす．上記の手順により，翻訳誤りに頑健で翻訳調の特徴に対応した翻訳モデルの学習を実現する．

5.4 評価実験

本節では，提案手法の効果を確認するための機械翻訳実験について述べる．

5.4.1 データ

実験では，下記の2種類のコーパスを用いる．表5.1に各データセットのデータ数と記事数を示す.

ニュースコーパス
: 時事通信社のニュース記事を利用して開発された日英ニュース対訳コーパス (田中英輝 他 2021) の一部を用いた．このコーパスは日本語のニュース記事の一部を文単位で人手翻訳している．本実験では，ニュースタイトルを除きニュース本文を利用した．コーパス中の 1 記事あたりの平均文数は 8 文であった．

TED トークコーパス
: IWSLT2017 (Cettolo et al. 2012) で公開されている TED の字幕対訳コーパスを用いた．開発データセットは “dev2010” を用いた．テストセットは英日・日英機械翻訳タスクについては “tst2014” を用い，英独・独英機械翻訳タスクについては “tst2015” を用いた．コーパス中の 1 記事あたりの平均文数は 119 文であった．

5.4.2 文脈の利用

5.2 節で述べた結合ベース文脈考慮型ニューラル機械翻訳（図 5.1（a））とマルチエンコーダ文脈考慮型ニューラル機械翻訳（図 5.1（b））の 2 種類の文脈考慮型ニューラル機械翻訳を用いた．目的言語側の文脈情報には翻訳対象の文の 1 つ前の文を利用した．

5.4.3 詳細設定

英語とドイツ語の単語分割には Moses toolkit[*] の tokenizer.perl[†] を用いた．日本語の形態素解析には，KyTea (Neubig et al. 2011) を利用した．データのクリーニングには，Moses toolkit の clean-corpus-n.perl[‡] を用い，データの形態素長が 2 語以上 150 語以下かつ対訳データ間の形態素長比が 9 以下のデータのみを抽出した．各タスクで語彙数を制限するため，バイトペア符号化（byte pair encoding）を用いたサブワード (Sennrich et al. 2016b) を用いた．語彙数は 48,000 とし，頻度が 35 以下の語彙はサブワードに分割した．データ内の長大語は特別な処理を行わなかった[§]．結合ベース文脈考慮型ニューラル機械翻訳には，トランスフォーマーモデル (Vaswani et al. 2017) のエンコーダとデコーダを用いた．原言語文と目的言語側の文脈情報が結合された単語列をエンコーダが中間表現であるベクトルに変換し，デコーダがそのベクトルを用いて翻訳結果となる目的言語文を生成する．マルチエンコーダ文脈考慮型ニューラル機械翻訳について

[*] https://github.com/moses-smt/mosesdecoder
[†] https://github.com/moses-smt/mosesdecoder/blob/master/scripts/tokenizer/tokenizer.perl
[‡] https://github.com/moses-smt/mosesdecoder/blob/master/scripts/training/clean-corpus-n.perl
[§] 頻度などの設定条件に従い，バイトペア符号化によりサブワード分割される．

比較手法	用いる文脈情報	ニューラル機械翻訳
Single-Sentence	文脈情報を用いない	トランスフォーマーモデル
Src-Context (Agrawal et al. 2018)	翻訳対象の前文（原言語文）	結合ベース文脈考慮型
		マルチエンコーダ文脈考慮型
Trg-Context GT (Bawden et al. 2018)	翻訳対象の前文の参照訳（目的言語文）	結合ベース文脈考慮型
		マルチエンコーダ文脈考慮型
Trg-Context Pred.	翻訳対象の前文の機械翻訳結果（目的言語文）	結合ベース文脈考慮型
		マルチエンコーダ文脈考慮型

表 5.2: 比較手法．

も，トランスフォーマーモデルを用い，Littell et al. (2019) のアーキテクチャを基に実装した．結合ベース文脈考慮型ニューラル機械翻訳とマルチエンコーダ文脈考慮型ニューラル機械翻訳のいずれも Sockeye toolkit (Hieber et al. 2018) を用いて実装した．

本章のモデルの学習には Adam (Kingma and Ba 2015) を最適化アルゴリズムとして用い，学習率は 0.0002 とした．バッチ内のデータは 5,000 トークン以内になるように構築した．学習時の各エンコーダへの入力の最大長は，結合ベース文脈考慮型ニューラル機械翻訳では 200 トークン，マルチエンコーダ文脈考慮型ニューラル機械翻訳では原文エンコーダ，文脈エンコーダともに 100 トークンに制限した[¶]．翻訳時の各エンコーダへの入力の最大長は設定しなかった．その他のハイパーパラメータの設定は Sockeye toolkit のデフォルト値に従った．早期終了（Early stopping）を設定し，patience の値を 32 とした．翻訳時は，ビームサーチを用い，ビーム幅を 5 とした．式 5.3.1 のハイパーパラメータ k は 2 とした．学習，翻訳ともに，Nvidia P100 Tesla GPU を用いて計算を行った．翻訳性能の比較には，ランダムシードの異なる 5 つのモデルを学習してそれぞれのモデルが出力した翻訳結果の BLEU スコア (Papineni et al. 2002) の中央値を用いた．BLEU スコアには，case-insensitive BLEU-4 を用い，multi-bleu.perl [‖] で計算した．また，サンプル数を 10,000 に設定した paired-bootstrap.py[**] を用い，BLEU スコアにおける提案手法と他手法との有意差（有意水準 $\alpha = 0.05$）を調べた．

[¶]結合ベース文脈考慮型ニューラル機械翻訳は，原文と文脈の合計が 200 トークン以内であれば学習データとして用いるため，マルチエンコーダ文脈考慮型ニューラル機械翻訳よりも学習データの数は 6〜153 データ多くなった．

[‖]https://github.com/moses-smt/mosesdecoder/blob/master/scripts/generic/multi-bleu-detok.perl

[**]https://github.com/neubig/util-scripts/blob/master/paired-bootstrap.py

5.4.4 比較手法

提案手法との比較を行うために本章で用いた比較手法（表 5.2）を説明する．

文レベル機械翻訳（Single-Sentence）
: 文脈として目的言語側の前文を考慮することによる効果を知るために，文脈を考慮しない機械翻訳手法を比較手法とした．翻訳対象のみをエンコーダに入力してデコーダが翻訳結果を出力する．トランスフォーマーモデルのニューラル機械翻訳を用いた．

文脈考慮型機械翻訳（Src-Context，Trg-Context）
: 文脈を考慮する機械翻訳手法として，原言語側の情報を用いる手法 (Agrawal et al. 2018) と目的言語側の情報を用いる手法 (Bawden et al. 2018) がある．両方の情報を併用することもできるが，本章ではどちらの情報が有効であるかを比較するために，原言語側の情報を用いた場合（Src-Context）と目的言語側の情報を用いた場合（Trg-Context）の双方を比較手法とした．目的言語側の情報を用いる手法では，従来手法は参照訳を用いていたが機械翻訳結果を用いることもできるため，双方を用いた手法（Trg-Context GT，Trg-Context Pred.）を比較した．文脈情報は，提案手法と同様に翻訳対象の文の 1 つ前の文を利用した．Src-Context，Trg-Context GT，Trg-Context Pred. ともに，結合ベース文脈考慮型ニューラル機械翻訳とマルチエンコーダ文脈考慮型ニューラル機械翻訳の 2 種類のシステムに適用した．

5.5 実験結果

ニュースコーパスを用いた英日・日英機械翻訳実験の結果，および TED トークコーパスを用いた英日・日英，英独・独英機械翻訳実験の結果を表 5.3, 5.4 に示す．文脈を用いない手法（Single-Sentence），原言語側の前文を用いた手法（Src-Context），目的言語側の前文として，参照訳を用いて学習した手法（Trg-Context GT），機械翻訳結果を用いて学習した手法（Trg-Context Pred.），参照訳と機械翻訳結果の両方を用いて学習した提案手法，それぞれの BLEU スコアを示している．

5.5.1 提案手法の有効性の検証

提案手法と，文脈を考慮しないニューラル機械翻訳，目的言語側の前文を用いた手法，および原言語側の前文を用いた手法とを比較することで提案手法の有効性を検証する．

手法	ニューラル機械翻訳	学習時文脈情報 Src	学習時文脈情報 Trg GT	学習時文脈情報 Trg MT	ニュースコーパス 英日	ニュースコーパス 日英
Single-Sentence	vanilla				41.99	24.23
Src-Context	concat.	✓			42.87	24.40
Trg-Context GT	concat.		✓		41.45	24.40
Trg-Context Pred.	concat.			✓	42.07	23.83
提案手法	concat.		✓	✓	42.89$^{\dagger\diamond\star}$	24.62*
Src-Context	multi-enc.	✓			42.06	24.37
Trg-Context GT	multi-enc.		✓		42.40	24.80
Trg-Context Pred.	multi-enc.			✓	42.45	24.31
提案手法	multi-enc.		✓	✓	42.79†	24.86†

表 5.3: 原言語側，または目的言語側の文脈情報を用いた，ニュースコーパスの英日・日英機械翻訳の実験結果．vanilla は文脈を用いないニューラル機械翻訳，concat. は結合ベース文脈考慮型ニューラル機械翻訳，multi-enc. はマルチエンコーダ文脈考慮型ニューラル機械翻訳を表している．Src，Trg GT，Trg MT は，学習時に用いた文脈情報として，原言語側の前文，目的言語側の参照訳，目的言語側の機械翻訳結果を用いたことを表している．値は BLEU スコアを表している．†，‡，⋄，⋆ は，提案手法が Single-Sentence，Src-Context，Trg-Context GT，Trg-Context Pred. と比較して有意に BLEU スコアが高いことを示している（有意水準 $\alpha = 0.05$）．Src-Context，Trg-Context GT，Trg-Context Pred については同じシステム内（concat. または multi-enc.）での比較とした．

先ず，提案手法と，文脈を用いないニューラル機械翻訳（Single-Sentence）とを比較する．表 5.3, 5.4 より，結合ベース文脈考慮型ニューラル機械翻訳（concat.）を用いた場合は，ニュースコーパス日英機械翻訳と TED トークコーパス英日・日英機械翻訳以外のタスクにおいて提案手法の BLEU スコアが有意に高かった[††]．マルチエンコーダ文脈考慮型ニューラル機械翻訳（multi-enc.）を用いた場合は，全てのタスクにおいて提案手法の BLEU スコアが有意に高かった．

次に，目的言語側の前文を用いた手法間の比較として，提案手法と，参照訳のみを用いて学習した手法（Trg-Context GT），および機械翻訳結果のみを用いて学習した手法（Trg-Context Pred.）とを比較する．表 5.3, 5.4 より，結合ベー

[††]有意水準 $\alpha = 0.05$．本章では全て同じ p 値を利用した．

手法	ニューラル機械翻訳	学習時文脈情報 Src	学習時文脈情報 Trg GT	学習時文脈情報 Trg MT	TED トークコーパス 英日	日英	英独	独英
Single-Sentence	vanilla				15.69	8.48	24.04	28.46
Src-Context	concat.	✓			15.17	8.79	25.96	30.03
Trg-Context GT	concat.		✓		14.72	7.55	25.09	29.11
Trg-Context Pred.	concat.			✓	15.32	7.37	23.99	29.05
提案手法	concat.		✓	✓	15.33$^{\diamond}$	8.36$^{\diamond\star}$	25.09†*	29.74†*
Src-Context	multi-enc.	✓			15.58	9.33	25.97	29.55
Trg-Context GT	multi-enc.		✓		16.15	8.98	25.96	30.25
Trg-Context Pred.	multi-enc.			✓	16.27	9.15	25.98	30.23
提案手法	multi-enc.		✓	✓	16.37†‡	9.66$^{\dagger\diamond}$	26.01†	30.60†‡

表 5.4: 原言語側，または目的言語側の文脈情報を用いた，TED トークコーパスの英日・日英機械翻訳，英独・独英機械翻訳の実験結果．vanilla は文脈を用いないニューラル機械翻訳，concat. は結合ベース文脈考慮型ニューラル機械翻訳，multi-enc. はマルチエンコーダ文脈考慮型ニューラル機械翻訳を表している．Src，Trg GT，Trg MT は，学習時に用いた文脈情報として，原言語側の前文，目的言語側の参照訳，目的言語側の機械翻訳結果を用いたことを表している．値は BLEU スコアを表している．†，‡，$\diamond$，$\star$ は，提案手法が Single-Sentence，Src-Context，Trg-Context GT，Trg-Context Pred. と比較して有意に BLEU スコアが高いことを示している（有意水準 $\alpha = 0.05$）．Src-Context，Trg-Context GT，Trg-Context Pred については同じシステム内（concat. または multi-enc.）での比較とした．

ス文脈考慮型ニューラル機械翻訳（concat.）を用いた場合は，TED トークコーパスの英独機械翻訳の Trg-Context GT との比較では同じ BLEU スコアだったが，それ以外では提案手法の BLEU スコアが高く，一部のタスクでは有意に高かった．マルチエンコーダ文脈考慮型ニューラル機械翻訳（multi-enc.）を用いた場合は，全タスクにおいて提案手法の BLEU スコアが高く，一部のタスクでは有意に高かった．

最後に，提案手法と，原言語側の前文を用いた手法（Src-Context）とを比較する．表 5.3, 5.4 より，マルチエンコーダ文脈考慮型ニューラル機械翻訳

手法	翻訳時文脈情報	ニュースコーパス		TEDトークコーパス			
		英日	日英	英日	日英	英独	独英
Single-Sentence	N/A	41.99	24.23	15.69	8.48	24.04	28.46
Trg-Context GT	機械翻訳	42.40	24.80†	16.15	8.98	25.96†	30.25†
	参照訳	42.50†	25.09†	16.39†	9.51†	25.98†	30.33†

表 5.5: 学習時と翻訳時両方ともに目的言語側の前文の参照訳を用いた実験結果（BLEUスコア）との比較．マルチエンコーダ文脈考慮型ニューラル機械翻訳を用いた．目的言語側の前文の機械翻訳の出力にはSingle-Sentenceの出力を用いた．†はSingle-Sentenceと比較して有意にBLEUスコアが高いことを示している（有意水準$\alpha = 0.05$）．

（multi-enc.）を用いた場合は，全タスクにおいて提案手法のBLEUスコアの方が高く，TEDトークコーパスの英日，独英機械翻訳タスクでは有意に高かった．従来手法では目的言語側の文脈の利用により翻訳精度が低下することが報告されていたが (Bawden et al. 2018)，multi-enc. を用いた提案手法では，文脈を用いていない手法と比較して有意にBLEUスコアが高かっただけでなく，Src-Contextと比較しても全タスクでBLEUスコアが高くなっており，Src-Contextと同等以上の効果が得られた．原言語側の前文を用いる手法と目的言語側の前文を用いる手法は異なる情報を用いており，競合するものではなく，条件によって一方のみが利用可能である場合や，併用が可能である場合が考えられる．

上記の比較により，提案手法の有効性を確認した．

5.5.2 露出バイアス（exposure bias）の影響の考察

提案手法は露出バイアス（exposure bias）の問題を緩和するための手法であるが，学習時，翻訳時ともに目的言語側の文脈に参照訳を用いれば，露出バイアス（exposure bias）の問題は起こらないといえる．そこで，露出バイアス（exposure bias）の影響を考察するために，翻訳時に目的言語側の前文として，露出バイアス（exposure bias）の影響がない参照訳を用いた場合の結果を表5.5に示す．機械翻訳器は，マルチエンコーダ文脈考慮型ニューラル機械翻訳を用いた．翻訳時の文脈情報に参照訳を用いたTrg-Context GTは，露出バイアス（exposure bias）の問題が起こると考えられる機械翻訳の出力（Single-Sentenceの出力）を用いたTrg-Context GTと比較して，全てのタスクでBLEUスコアが

高かった．また，Single-Sentenceとの比較では，翻訳時の文脈情報に機械翻訳結果を用いたTrg-Context GTはニュースコーパスの日英機械翻訳タスクとTEDトークコーパスの英独・独英機械翻訳タスクのみで有意にBLEUスコアが高かったが，参照訳を用いたTrg-Context GTは全てのタスクで有意にBLEUスコアが高いことを確認できた．以上の結果より，翻訳時に目的言語側の前文として利用した機械翻訳結果と参照訳の違いが翻訳精度に影響していると言え，これが露出バイアス（exposure bias）の影響と考えることができる．

5.5.3 目的言語側の前文に機械翻訳結果のみを用いる影響の考察

表5.3, 5.4より，目的言語側の前文に機械翻訳結果のみを用いた手法（Trg-Context Pred.）のBLEUスコアは提案手法よりも低かった．また，結合ベース文脈考慮型ニューラル機械翻訳（concat.）を用いたニュースコーパスの日英機械翻訳タスクとTEDトークコーパスの英日・日英，英独機械翻訳タスクでは，Trg-Context Pred.のBLEUスコアは文脈を考慮しないトランスフォーマーモデル（Single-Sentence）よりも低かった．

Trg-Context Pred.は，学習時，翻訳時ともに機械翻訳結果を用いており露出バイアス（exposure bias）による影響を回避できているが，BLEUスコアが低かった理由について考察する．表5.3, 5.4の参照訳のみを用いて学習した手法（Trg-Context GT）は，学習時に参照訳，翻訳時に機械翻訳結果を用いているが，Trg-Context GTの翻訳時に参照訳を用いることで，Trg-Context Pred.と同様に露出バイアス（exposure bias）による影響を回避することができる．Trg-Context GTの翻訳時に目的言語側の前文として参照訳を用いた結果は，表5.5（一番下の行）で示したように，Trg-Context Pred.の結果と比較してBLEUスコアが高い．両者の違いは，翻訳時に目的言語側の前文として機械翻訳結果を用いたか，参照訳を用いたかの違いであり，これらの訳質の違いであるといえる．以上より，Trg-Context Pred.のBLEUスコアが低かった理由は，文脈として利用している目的言語側の前文の機械翻訳結果の訳質が低いために，文脈情報を効果的に利用するように学習することが困難であったためと考えられる．

提案手法は，学習の初期に訳質の高い参照訳を用いて文脈情報を効果的に利用し，段階的に機械翻訳の出力に置き換えることで，文脈情報の効果的な利用と露出バイアス（exposure bias）の影響の回避を両立したと考えられる．提案手法の学習終了時の学習データ内に含まれる参照訳の割合は9%（学習終了時のエポック数が8），または6%（学習終了時のエポック数が9）であり‡‡，学習終了時には露出バイアス（exposure bias）の影響がほぼ無い学習を行っている．

‡‡ニュースコーパスの英日・日英機械翻訳タスクとTEDトークコーパスの英日機械翻訳タスクの学習終了時のエポック数は8であり，TEDトークコーパスの日英，英独・独英機械翻訳タスクの学習終了時のエポック数は9であった．

手法	学習時文脈情報	ニュースコーパス		TED トークコーパス			
		英日	日英	英日	日英	英独	独英
Single-Sentence	N/A	41.99	24.23	15.69	8.48	24.04	28.46
Trg-Context GT	参照訳	42.40	24.80	16.15	8.98	25.96	30.25
Trg-Context Pred.	機械翻訳	42.45	24.31	16.27	9.15	25.98	30.23
Trg-Context GT&Pred.	固定	42.44	24.30	16.28	9.05	25.99	30.06
提案手法	変化	42.79	24.86†	16.37	9.66†	26.01	30.60

表 5.6: 学習時に目的言語側の前文の参照訳と機械翻訳の出力とを固定の割合（50 : 50）で用いた場合（Trg-Context GT&Pred.）との比較．マルチエンコーダ文脈考慮型ニューラル機械翻訳を用いた．目的言語側の前文の機械翻訳の出力には Single-Sentence の出力を用いた．翻訳時の目的言語側の前文の機械翻訳結果には Single-Sentence の出力を用いた．† は Trg-Context GT&Pred. と比較して提案手法が有意に BLEU スコアが高いことを示している（有意水準 $\alpha = 0.05$）．

5.5.4 カリキュラムラーニングの効果の考察

提案手法は露出バイアス（exposure bias）の影響の緩和を目的とし，学習時の目的言語側の前文として参照訳と事前に用意した機械翻訳の出力をサンプリング制御しながら用いて段階的に機械翻訳の出力に切り替えるカリキュラムラーニングの手法を用いている．学習時の目的言語側の前文として参照訳のみを用いた手法は Trg-Context GT，機械翻訳結果のみを用いた手法は Trg-Context Pred. として表 5.3, 5.4 に実験結果を示した．他の選択肢として，参照訳と事前に用意した機械翻訳の出力を段階的にサンプリング制御するのではなく，固定の割合でサンプリングすることも考えられる．そこで，一定の割合でサンプリングした場合（Trg-Context GT&Pred.）と提案手法とを比較する．本章では，学習時の目的言語側の前文として参照訳と機械翻訳の出力を同じ割合ずつサンプリングした．表 5.6 に実験結果を示す．表 5.3, 5.4 から Single-Sentence，Trg-Context GT, Trg-Context Pred. の結果も転載する．機械翻訳器は，マルチエンコーダ文脈考慮型ニューラル機械翻訳を用いた．Trg-Context GT&Pred. と比較し，提案手法は全てのタスクで BLEU スコアが高く，ニュースコーパスの日英機械翻訳タスクと TED トークコーパスの日英機械翻訳タスクでは有意に BLEU スコアが高いことを確認した．以上の結果より，学習時の目的言語側の前文として参照訳と機械翻訳の出力をサンプリング制御することが翻訳精度向上に寄与しており，この差がカリキュラムラーニングの効果であると考えることができる．

手法	翻訳時文脈情報（機械翻訳結果）	ニュースコーパス		TED トークコーパス			
		英日	日英	英日	日英	英独	独英
Trg-Context Pred.	Single-Sentence の出力	42.45	24.31	16.27	9.15	25.98	30.23
	Trg-Context Pred. の出力	42.98	24.50	16.28	9.76†	25.98	30.30†
提案手法	Single-Sentence の出力	42.79	24.86	16.37	9.66	26.01	30.60
	提案手法の出力	43.45†	24.98	16.47	9.76	26.06	30.62

表 5.7: 目的言語側の前文の自身の機械翻訳結果を用いた実験結果（BLEU スコア）との比較．マルチエンコーダ文脈考慮型ニューラル機械翻訳を用いた．† は翻訳時に目的言語側の前文として Single-Sentence の出力を用いた場合と比較して有意に BLEU スコアが高いことを示している（有意水準 $\alpha = 0.05$）．

5.5.5 翻訳時に自身の機械翻訳結果を用いた場合の効果の考察

提案手法は，学習時の目的言語側の前文として参照訳と事前に用意した機械翻訳の出力（Single-Sentence の出力）をサンプリング制御しながら用い，翻訳時の目的言語側の前文としても同様の Single-Sentence の出力を用いることを前提とした．他の選択肢として，目的言語側の前文に学習時と翻訳時ともに Single-Sentence の出力の代わりに自身の機械翻訳結果を用いることも考えられる．自身の出力を用いる場合には，文書単位で 1 文ずつ順番に機械翻訳する必要があり，文書内の文を並列に翻訳できないという問題がある．さらに，学習時にはモデルの更新ごとに機械翻訳する必要がある．一方で,「目的言語側の前文に機械翻訳結果のみを用いる影響の考察」で示したように文脈情報として用いる機械翻訳結果の訳質も翻訳精度に影響することから，目的言語側の前文として，Single-Sentence の出力の代わりに，訳質の高い自身の出力を用いることで，翻訳精度の向上が期待できる．目的側の前文に学習時の自身の出力を用いることは困難であるが，翻訳時であれば自身の出力を用いることが可能である．そこで，Trg-Context Pred. の翻訳時には目的言語側の前文として Trg-Context Pred. の出力を用い，提案手法の翻訳時には目的言語側の前文として提案手法の出力を用いた場合の結果を表 5.7 に示す．機械翻訳器は，マルチエンコーダ文脈考慮型ニューラル機械翻訳を用いた．Trg-Context Pred.，および提案手法のいずれにおいても，目的言語側の前文に自身の出力を利用することで，一部の

手法	主語の省略なし	主語の省略あり	
		正解訳数	誤訳数
Single-Sentence	201	49	50
提案手法	201	63	36

表 5.8: 人手評価結果.

タスクで有意にBLEUスコアが高く，かつBLEUスコアが低下したタスクはなかった．これらは目的言語側の前文の訳質の向上によるものと考えられる．以上の結果より，翻訳速度を優先する場合はSingle-Sentenceの出力を用い，翻訳品質を優先する場合は自身の出力を用いればよい．

5.5.6 人手評価による考察

表5.3, 5.4の翻訳実験の結果より，目的言語側の前文を用いた提案手法の文脈考慮型ニューラル機械翻訳が，文脈情報を用いないニューラル機械翻訳と比較してBLEUスコアにおいて翻訳精度が高いことを確認したが，目的言語側の前文を用いたことで翻訳結果に具体的にどのような効果があるかを確認するために，TEDトークコーパスの日英タスクのテストセットからランダムに抽出した300文対を対象にして人手評価を行った．人手評価は，文脈を考慮することで効果があると考えられる翻訳結果の主語の補完に着目した (Voita et al. 2018). 評価者1名が日本語側で主語が省略されているかを確認し，日本語側の主語が省略されている場合は，英語側で正しく主語を補完して翻訳しているかどうかを確認した．表5.8に人手評価の結果を示す．TEDトークコーパスの日英タスクの日本語側のデータで主語の省略のあった文はおよそ1/3であった．そのうち，主語を正しく補完して翻訳できた文は，Single-Sentenceでは49であったのに対して，提案手法では63となり，提案手法を用いることで正しく主語が補完できたことを確認した．

5.5.7 事例を用いた考察

表5.9に，TEDトークコーパスを用いた日英タスクの出力例を示す．上段（点線の上）には，目的言語側の前文の参照訳と機械翻訳結果（MT-OUT），翻訳対象の日本語文と参照訳を示し，下段（点線の下）には，各手法による翻訳結果を示す．機械翻訳器は，マルチエンコーダ文脈考慮型ニューラル機械翻訳を用いた．例文の翻訳対象の日本語文には，ゼロ代名詞が含まれ，かつ，目

的言語側の文から代名詞を推定できると考えられるものを選んだ．最初の例（#1）は，同じ文（“We choose to go to the moon.”）が２度続くケースであり，代名詞（“We ＝私たち”）が省略されている。翻訳対象文では，文の主語である“We=私たち”が省略されており，文脈を考慮しない機械翻訳（Single-Sentence）は主語を誤って翻訳した．一方，文脈を考慮した機械翻訳では４つとも主語を適切に補完して翻訳した． ２番目の例（#2）は、前文が長文の例であり，代名詞（“him ＝彼”）が省略されている．文脈を考慮しない機械翻訳では代名詞を補完せずに翻訳した．文脈考慮型機械翻訳では，原言語側の前文を用いた手法（Src-Context）と提案手法は代名詞を正しく補完して翻訳しているが，目的言語側の前文を用いたその他の手法（Trg-Context GT，Trg-Context Pred.）では，誤った代名詞を補完して翻訳した．前文が長文になるとエンコーダが目的言語側の文脈を適切に処理することが困難となる．この例は，Trg-Context GT，Trg-Context Pred. において，エンコーダが適切に処理できなかった例であると考えられる． ３番目の例（#3）は、代名詞（“his ＝彼”）が省略されている．提案手法のみが代名詞を適切に補完して翻訳している．

5.6 まとめ

本章では，文脈考慮型ニューラル機械翻訳において，目的言語側の文脈を効果的に利用する手法を提案した．提案手法では，学習時と翻訳時の間にある目的言語側の前文のデータの特徴の違い（露出バイアス（exposure bias））の問題の解消を目的とし，目的言語側の参照訳と機械翻訳結果の両方を用いて，スケジュールドサンプリング法で学習データを制御した．結合ベース文脈考慮型ニューラル機械翻訳とマルチエンコーダ文脈考慮型ニューラル機械翻訳を用いて提案手法を実装し，英日・日英，および英独・独英の機械翻訳タスクで評価した．実験により，BLEU スコアにおいて提案手法の有効性を確認した．今後の課題として，提案手法の各タスクにおける最適なハイパーパラメーター値の調査がある．また，本章では計算機の性能を考慮して目的言語側の文脈情報を前文に限定したが，目的言語側と原言語側の周辺の複数文を併用した手法の検討も必要である．

	入力・参照訳（上段） 手法（下段）	入力データ（上段） 翻訳結果（下段）
#1	目的言語側の前文 （参照訳）	**We** choose to go to the moon.
	目的言語側の前文 （MT-OUT）	**We** decided to go to the moon.
	翻訳対象文	月に行くことにしたんです
	参照訳	**We** choose to go to the moon.
	Single-Sentence	**I** went to the moon.
	Src-Context	**We** decided to go to the moon.
	Trg-Context GT	**We** decided to go to the moon.
	Trg-Context Pred.	**We** decided to go to the moon.
	提案手法	**We** decided to go to the moon.
#2	目的言語側の前文 （参照訳）	On the day after September 11 in 2001, I heard the growl of a sanitation truck on the street, and I grabbed my infant son and I ran downstairs and there was a man doing his paper recycling route like **he** did every Wednesday.
	目的言語側の前文 （MT-OUT）	On September 11, 2001, I heard a young man flying in a city, and he went down, and a man came down the stairs, he was going down the stairs, he was going down a piece of paper, and **he** was doing it every Wednesday.
	翻訳対象文	あの日に一日中仕事をしてくれていることに感謝しようとしました でも涙が出てきたのです
	参照訳	And I tried to thank **him** for doing his work on that day of all days, but I started to cry.
	Single-Sentence	And I was trying to appreciate the fact that I was doing my work every day, but tears were coming out.
	Src-Context	And on that day, I decided to thank **him** for doing my work, but I came up with tears.
	Trg-Context GT	And I was going to thank you for that day, and I was going to thank you for doing my work every day, but I got to tears.
	Trg-Context Pred.	And on that day, I thank you very much for doing everything in that day, but tears come up and tears come up.
	提案手法	And I was grateful for that day, and I tried to thank **him** for doing work all the day, but I had tears in my tears.

#3	目的言語側の前文（参照訳）	I would like to talk to you about a story about a small town **kid**.
	目的言語側の前文（MT-OUT）	I'm going to talk about **a boy** in a village.
	翻訳対象文	**名前**は知りませんが彼の話はできます
	参照訳	I don't know **his name**, but I do know his story.
	Single-Sentence	I don't know **the name**, but I can tell you his story.
	Src-Context	You don't know **the name**, but you can talk to him.
	Trg-Context GT	I don't know **the name**, but I can tell him.
	Trg-Context Pred.	He doesn't know **the names**, but he can tell the story.
	提案手法	I don't know **his name**, but I can tell you his story.

表 5.9: TED トークコーパスを用いた日英機械翻訳実験の翻訳結果例（上段：入力と参照訳，下段：各手法の結果）.

第6章

結論

本論文では，ニューラル機械翻訳の学習時と翻訳時において特徴が違うデータを用いることで翻訳精度が低下する課題に着目した．データ構築に必要な人手や計算機のコストなどを考慮すると，学習時と翻訳時において特徴が違うデータを用いるニューラル機械翻訳の精度を向上させることは実用的に重要である．本論文では，学習時と翻訳時におけるデータの特徴を，翻訳対象である入力データ，翻訳結果である出力データ，入出力データ間の関係，翻訳対象とは別の外部情報の４種類に分類し，これらの特徴の違いにより翻訳精度が低下する２つのニューラル機械翻訳のタスクに取り組んだ．

１つ目は，翻訳対象である入力データ，翻訳結果である出力データ，入出力データ間の関係，の特徴の違いを扱うドメインタグを用いたニューラル機械翻訳のタスクである．このタスクは，翻訳対象とはドメインが違う対訳データを活用して翻訳を行う．従来手法であるタグを用いてドメインの違いを区別して学習する手法に着目し，従来手法では扱えなかった複数の特徴を区別して学習できる複数のドメインタグを用いた手法を提案した．４つのニュースの対訳コーパスを用いた日英機械翻訳の実験で，従来手法と比較して提案手法の翻訳精度が向上することを確認した．

２つ目は，翻訳対象とは別の外部情報として目的言語側の前文を用いる文脈考慮型ニューラル機械翻訳のタスクである．このタスクは，翻訳対象の文に目的言語側の前文を入力データに追加して翻訳を行う．従来手法は，学習時と翻訳時とで特徴が違う目的言語側の前文を用いていたために翻訳精度が低下していた．本論文では，学習時と翻訳時で用いるデータの特徴が違うことによる影響を緩和するために，学習時に用いる目的言語側の前文を学習の進行に応じて参照訳から機械翻訳結果へ段階的に切り替える手法を提案した．ニュースコーパスを用いた英日・日英機械翻訳と，IWSLT2017 の TED トークコーパスを用いた英日・日英，および英独・独英機械翻訳の実験で，従来手法と比較して提案手法の翻訳精度が向上することを確認した．

6.1 貢献

本論文全体，および個々のタスクの貢献をまとめる．

全体の貢献

本論文の主な貢献は，学習時と翻訳時におけるデータの特徴を 4 種類に分類し，これらの特徴の違いに起因する翻訳精度の低下の問題を低減する手法を提案したことである．学習時と翻訳時におけるデータの特徴を体系的に分類し，データの特徴の違いを詳細に区別して対応することで翻訳精度の向上が期待でき，2 つのニューラル機械翻訳のタスクで効果を確認した．学習時と翻訳時におけるデータの特徴の違いはニューラル機械翻訳にとって一般的な課題であり，本論文が扱った 2 つのニューラル機械翻訳以外のタスクにも応用が可能である．

ドメインタグを用いたニューラル機械翻訳

従来手法はデータの特徴の違いを対訳データ全体で捉えずに個々の文の特徴に着目して分類していた．それに対して提案手法は，従来手法の分類が十分でないことを示し，データの特徴の違いを，原言語側の特徴，目的言語側の特徴，原言語側と目的言語側のデータ間の特徴に分類して学習することで，翻訳精度が向上することを示した．

目的言語の前文を用いた文脈考慮型ニューラル機械翻訳

従来手法は翻訳精度が低下するという問題があったが，問題の原因が明らかではなかった．本論文では，この原因が学習時と翻訳時で特徴が違う目的言語側の前文を用いていたためであることを明らかにし，学習時に用いる目的言語側の文脈情報を学習の進行に応じて参照訳から機械翻訳結果へ段階的に切り替える手法を提案した．これにより，目的言語側の周辺の文を用いた場合でも翻訳精度が向上することを示した．

6.2 今後の課題

まず，ニューラル機械翻訳における将来的な課題を 2 つ示す．

- 本論文が扱った 2 つのニューラル機械翻訳以外のタスクにおいて，学習時と翻訳時との特徴の違いにより翻訳精度が低下しているかどうか検証する必要がある．学習時と翻訳時におけるデータの特徴の違いは，大量の学習データを必要とするタスクで起こりやすい現象であると考える．例えば，マルチモーダル翻訳や要約翻訳などで検証を行い，従来手法が学習時と翻訳時におけるデータの特徴の違いを考慮した手法となっているか検証する必要がある．

- ドメインが不明なコーパスに対応する必要がある．提案手法は用いるコーパスの特徴が分かっていることが前提となっているが，例えばウェブから収集したデータなどデータごとに異なる特徴を持つコーパスには対応できない．このようなコーパスに提案手法を適用するにはクラスタリング手法などを用いてコーパスを分類する必要があり，有効な分類手法を検討する必要がある．

次に，個々のタスクにおける今後の課題を示す．

ドメインタグを用いたニューラル機械翻訳

翻訳精度の向上に寄与するドメインタグの選定手法を検討する必要がある．本論文では，翻訳元の言語側（原言語側）の特徴，翻訳先の言語側（目的言語側）の特徴，および原言語側と目的言語側のデータ間の特徴に分けてドメインタグを付与したが，適切なドメインタグの付与は利用する対訳コーパスに依存すると考えられる．例えば，原言語側と目的言語側のデータ間の特徴の違いは，訳抜け，過剰訳，誤訳，意訳など，コーパスにより複数の特徴の組み合わせで区別できる場合がある．その場合，データ間の特徴の違いによって複数のタグを付与した方がよい可能性があり，他のデータセットを用いた実験を行う必要がある．また，翻訳品質の低いアウトオブドメインのデータセットを用いた実験を行い，アウトオブドメインのデータの翻訳品質と提案手法の効果との関連性を明らかにする必要がある．

目的言語の前文を用いた文脈考慮型ニューラル機械翻訳

用いるデータを変えた場合に，提案手法の各タスクにおける最適なハイパーパラメーター値がどのように変わるかを調査し，提案手法の効果の範囲を確認する必要がある．また，本論文では計算機資源の関係で目的言語側の前の 1 文のみを対象として文脈考慮型ニューラル機械翻訳を用いたが，提案手法は目的言語側の周辺の複数文を用いることも可能である他，原言語側の周辺の複数文を併用することも可能である．文脈情報の適用範囲を広げた際の効果を明らかにする必要があると考える．

謝辞

本研究を進め，論文をまとめるにあたり、ご指導とご助言を賜りました多くの方々に対し、感謝の言葉を述べさせていただきます。主指導教官である徳永 健伸教授には，本研究だけでなく，自然言語処理全般の研究に関して，日頃より多くのご指導とご助言を賜りました．深く感謝いたします．審査委員である，岡崎 直観教授，宮崎 純教授，村田 剛志教授，下坂 正倫准教授には，本研究に関して貴重なご指導とご助言を賜りました．深く感謝いたします．また，本研究について様々なご意見を賜りました徳永研究室のみなさまにお礼申し上げます．NHK 放送技術研究所の後藤功雄氏，山田一郎氏には，本研究をまとめる上でのご配慮を賜りました．深く感謝いたします．国立研究開発法人情報通信研究機構 (NICT) の田中英輝氏には，本研究で用いたコーパスに関して貴重なご意見を賜りました．深く感謝いたします．本研究成果の一部は国立研究開発法人情報通信研究機構 (NICT) の委託研究「多言語音声翻訳高度化のためのディープラーニング技術の研究開発」により得られたものです．ありがとうございました．

参考文献

[Agrawal et al.] Agrawal, R., Turchi, M., and Negri, M. (2018). "Contextual handling in neural machine translation: Look behind, ahead and on both sides." In *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, pp. 11—20, Alacant, Spain. European Association for Machine Translation.

[Bahdanau et al.] Bahdanau, D., Cho, K., and Bengio, Y. (2015). "Neural Machine Translation by Jointly Learning to Align and Translate." In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

[Bawden et al.] Bawden, R., Sennrich, R., Birch, A., and Haddow, B. (2018). "Evaluating Discourse Phenomena in Neural Machine Translation." In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 1304–1313, New Orleans, Louisiana. Association for Computational Linguistics.

[Belinkov and Bisk] Belinkov, Y. and Bisk, Y. (2018). "Synthetic and Natural Noise Both Break Neural Machine Translation." In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*.

[Bengio et al.] Bengio, S., Vinyals, O., Jaitly, N., and Shazeer, N. (2015). "Scheduled Sampling for Sequence Prediction with Recurrent Neural Networks." In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R. (Eds.), *Advances in Neural Information Processing Systems 28*, pp. 1171–1179. Curran Associates, Inc.

[Bengio et al.] Bengio, Y., Louradour, J., Collobert, R., and Weston, J. (2009). "Curriculum learning." In Danyluk, A. P., Bottou, L., and Littman, M. L. (Eds.), *ICML*, Vol. 382 of *ACM International Conference Proceeding Series*, pp. 41–48. ACM.

[Berard et al.] Berard, A., Calapodescu, I., Dymetman, M., Roux, C., Meunier, J.-L., and Nikoulina, V. (2019a). "Machine Translation of Restaurant Reviews: New Corpus for Domain Adaptation and Robustness." In *Proceedings of the 3rd Workshop on Neural*

Generation and Translation, pp. 168–176, Hong Kong. Association for Computational Linguistics.

[Berard et al.] Berard, A., Calapodescu, I., and Roux, C. (2019b). "Naver Labs Europe's Systems for the WMT19 Machine Translation Robustness Task." In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pp. 526–532, Florence, Italy. Association for Computational Linguistics.

[李凌寒 他] 李凌寒，中澤敏明，鶴岡慶雅 (2019). 文脈情報を考慮した日英ニューラル機械翻訳. 言語処理学会第 25 回年次大会, pp. 101–104.

[Caswell et al.] Caswell, I., Chelba, C., and Grangier, D. (2019). "Tagged Back-Translation." In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pp. 53–63, Florence, Italy. Association for Computational Linguistics.

[Cettolo et al.] Cettolo, M., Federico, M., Bentivogli, L., Niehues, J., Stüker, S., Sudoh, K., Yoshino, K., and Federmann, C. (2017). "The IWSLT 2017 Evaluation Campaign." In *Proceedings of the 14th International Workshop on Spoken Language Translation: Evaluation Campaign*, pp. 2–14, Tokyo, Japan.

[Cettolo et al.] Cettolo, M., Girardi, C., and Federico, M. (2012). "WIT^3: Web Inventory of Transcribed and Translated Talks." In *Proceedings of the 16th Conference of the European Association for Machine Translation (EAMT)*, pp. 261–268, Trento, Italy.

[Cho et al.] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). "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)*, pp. 1724–1734, Doha, Qatar. Association for Computational Linguistics.

[Chu et al.] Chu, C., Dabre, R., and Kurohashi, S. (2017). "An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation." In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 385–391, Vancouver, Canada. Association for Computational Linguistics.

[Chu and Wang] Chu, C. and Wang, R. (2018). "A Survey of Domain Adaptation for Neural Machine Translation." In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 1304–1319, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

[Denkowski and Lavie] Denkowski, M. and Lavie, A. (2014). "Meteor Universal: Language Specific Translation Evaluation for Any Target Language." In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pp. 376–380, Baltimore, Maryland, USA. Association for Computational Linguistics.

[Duchi et al.] Duchi, J., Hazan, E., and Singer, Y. (2011). "Adaptive Subgradient Methods for Online Learning and Stochastic Optimization." *Journal of Machine Learning Research*, **12** (61), pp. 2121–2159.

[Garg et al.] Garg, S., Peitz, S., Nallasamy, U., and Paulik, M. (2019). "Jointly Learning to Align and Translate with Transformer Models." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pp. 4453–4462, Hong Kong, China. Association for Computational Linguistics.

[Hieber et al.] Hieber, F., Domhan, T., Denkowski, M., Vilar, D., Sokolov, A., Clifton, A., and Post, M. (2018). "The Sockeye Neural Machine Translation Toolkit at AMTA 2018." In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Papers)*, pp. 200–207, Boston, MA. Association for Machine Translation in the Americas.

[Hochreiter and Schmidhuber] Hochreiter, S. and Schmidhuber, J. (1997). "Long Short-Term Memory." *Neural Computation*, **9** (8), pp. 1735–1780.

[Iida et al.] Iida, R., Torisawa, K., Oh, J.-H., Kruengkrai, C., and Kloetzer, J. (2016). "Intra-Sentential Subject Zero Anaphora Resolution using Multi-Column Convolutional Neural Network." In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 1244–1254, Austin, Texas. Association for Computational Linguistics.

[Isozaki et al.] Isozaki, H., Hirao, T., Duh, K., Sudoh, K., and Tsukada, H. (2010). "Automatic Evaluation of Translation Quality for Distant Language Pairs." In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 944–952, Cambridge, MA. Association for Computational Linguistics.

[田中英輝 他] 田中英輝，中澤敏明，美野秀弥，伊藤均，後藤功雄，山田一郎，川上貴之，大嶋聖一，朝賀英裕 (2021). 時事通信社ニュースの日英対訳コーパスの構築–第 3 報. 言語処理学会第 27 回年次大会, pp. 228–231.

[Kim et al.] Kim, Y., Tran, D. T., and Ney, H. (2019). "When and Why is Document-level Context Useful in Neural Machine Translation?" In *Proceedings of the Fourth Work-*

shop on Discourse in Machine Translation, DiscoMT@EMNLP 2019, Hong Kong, China, November 3, 2019, pp. 24–34.

[Kingma and Ba] Kingma, D. P. and Ba, J. (2015). "Adam: A method for stochastic optimization." In *International Conference on Learning Representations (ICLR)*.

[Kobus et al.] Kobus, C., Crego, J., and Senellart, J. (2017). "Domain Control for Neural Machine Translation." In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, pp. 372–378, Varna, Bulgaria. INCOMA Ltd.

[Kuang et al.] Kuang, S., Xiong, D., Luo, W., and Zhou, G. (2018). "Modeling Coherence for Neural Machine Translation with Dynamic and Topic Caches." In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 596–606, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

[Läubli et al.] Läubli, S., Sennrich, R., and Volk, M. (2018). "Has Machine Translation Achieved Human Parity? A Case for Document-level Evaluation." In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 4791–4796, Brussels, Belgium. Association for Computational Linguistics.

[Libovický and Helcl] Libovický, J. and Helcl, J. (2017). "Attention Strategies for Multi-Source Sequence-to-Sequence Learning." In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 196–202, Vancouver, Canada. Association for Computational Linguistics.

[Littell et al.] Littell, P., Lo, C.-k., Larkin, S., and Stewart, D. (2019). "Multi-Source Transformer for Kazakh-Russian-English Neural Machine Translation." In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pp. 267–274, Florence, Italy. Association for Computational Linguistics.

[Luong and Manning] Luong, M.-T. and Manning, C. D. (2015). "Stanford Neural Machine Translation Systems for Spoken Language Domain." In *International Workshop on Spoken Language Translation*, Da Nang, Vietnam.

[Maruf et al.] Maruf, S., Martins, A. F. T., and Haffari, G. (2019). "Selective Attention for Context-aware Neural Machine Translation." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 3092–3102, Minneapolis, Minnesota. Association for Computational Linguistics.

[Miculicich et al.] Miculicich, L., Ram, D., Pappas, N., and Henderson, J. (2018). "Document-Level Neural Machine Translation with Hierarchical Attention Networks." In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2947–2954, Brussels, Belgium. Association for Computational Linguistics.

[Müller et al.] Müller, M., Rios, A., Voita, E., and Sennrich, R. (2018). "A Large-Scale Test Set for the Evaluation of Context-Aware Pronoun Translation in Neural Machine Translation." In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pp. 61–72, Brussels, Belgium. Association for Computational Linguistics.

[Neubig et al.] Neubig, G., Nakata, Y., and Mori, S. (2011). "Pointwise Prediction for Robust, Adaptable Japanese Morphological Analysis." In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 529–533, Portland, Oregon, USA. Association for Computational Linguistics.

[Okumura and Tamura] Okumura, M. and Tamura, K. (1996). "Zero Pronoun Resolution in Japanese Discourse Based on Centering Theory." In *COLING 1996 Volume 2: The 16th International Conference on Computational Linguistics*.

[Pan and Yang] Pan, S. and Yang, Q. (2010). "A Survey on Transfer Learning." *IEEE Transactions on Knowledge and Data Engineering*, **22** (10), pp. 1345–1359.

[Papineni et al.] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). "Bleu: a Method for Automatic Evaluation of Machine Translation." In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pp. 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

[Ranzato et al.] Ranzato, M., Chopra, S., Auli, M., and Zaremba, W. (2016). "Sequence Level Training with Recurrent Neural Networks." In Bengio, Y. and LeCun, Y. (Eds.), *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

[Robbins and Monro] Robbins, H. and Monro, S. (1951). "A stochastic approximation method." *Annals of Mathematical Statistics*, **22**, pp. 400–407.

[Sennrich et al.] Sennrich, R., Haddow, B., and Birch, A. (2016a). "Improving Neural Machine Translation Models with Monolingual Data." In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 86–96, Berlin, Germany. Association for Computational Linguistics.

[Sennrich et al.] Sennrich, R., Haddow, B., and Birch, A. (2016b). "Neural Machine Translation of Rare Words with Subword Units." In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1715–1725, Berlin, Germany. Association for Computational Linguistics.

[Sutskever et al.] Sutskever, I., Vinyals, O., and Le, Q. V. (2014). "Sequence to Sequence Learning with Neural Networks." In *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, pp. 3104–3112, Cambridge, MA, USA. MIT Press.

[Tiedemann and Scherrer] Tiedemann, J. and Scherrer, Y. (2017). "Neural Machine Translation with Extended Context." In *Proceedings of the Third Workshop on Discourse in Machine Translation*, pp. 82–92, Copenhagen, Denmark. Association for Computational Linguistics.

[Tieleman and Hinton] Tieleman, T. and Hinton, G. (2012). "Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude." COURSERA: Neural Networks for Machine Learning.

[Toral] Toral, A. (2019). "Post-editese: an Exacerbated Translationese." In *Proceedings of Machine Translation Summit XVII Volume 1: Research Track*, pp. 273–281, Dublin, Ireland. European Association for Machine Translation.

[Utiyama and Isahara] Utiyama, M. and Isahara, H. (2007). "A Japanese-English patent parallel corpus." In *Proceedings of Machine Translation Summit XI, Copenhagen, Denmark, 2007*, pp. 457–482.

[Vaibhav et al.] Vaibhav, V., Singh, S., Stewart, C., and Neubig, G. (2019). "Improving Robustness of Machine Translation with Synthetic Noise." In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 1916–1920, Minneapolis, Minnesota. Association for Computational Linguistics.

[Vaswani et al.] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. u., and Polosukhin, I. (2017). "Attention is All you Need." In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R. (Eds.), *Advances in Neural Information Processing Systems 30*, pp. 5998–6008. Curran Associates, Inc.

[Voita et al.] Voita, E., Sennrich, R., and Titov, I. (2019). "Context-Aware Monolingual Repair for Neural Machine Translation." In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint*

Conference on Natural Language Processing (EMNLP-IJCNLP), pp. 877–886, Hong Kong, China. Association for Computational Linguistics.

[Voita et al.] Voita, E., Serdyukov, P., Sennrich, R., and Titov, I. (2018). "Context-Aware Neural Machine Translation Learns Anaphora Resolution." In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1264–1274, Melbourne, Australia. Association for Computational Linguistics.

[Wang et al.] Wang, L., Tu, Z., Way, A., and Liu, Q. (2017). "Exploiting Cross-Sentence Context for Neural Machine Translation." In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2826–2831, Copenhagen, Denmark. Association for Computational Linguistics.

[Zhang et al.] Zhang, W., Feng, Y., Meng, F., You, D., and Liu, Q. (2019). "Bridging the Gap between Training and Inference for Neural Machine Translation." In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4334–4343, Florence, Italy. Association for Computational Linguistics.

[Zoph and Knight] Zoph, B. and Knight, K. (2016). "Multi-Source Neural Translation." In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 30–34, San Diego, California. Association for Computational Linguistics.