岐阜大学博士論文

日中・中日ニューラル機械翻訳のための文字特徴情報の利用とコーパス拡張手法

Additional Input Character Features and Corpus Augmentation for Neural Machine Translation between Japanese and Chinese

2020年3月

張 津一

Additional Input Character Features and Corpus Augmentation for Neural Machine Translation between Japanese and Chinese

Jinyi Zhang

Abstract

Machine translation is a subfield of artificial intelligence that investigates the transformation of text in the source language into its equivalent in the target language. Despite great progress in the field of Statistical Machine Translation (SMT) over the past two decades, translation quality has not yet satisfied users; at the same time, SMT systems have become increasingly complex with many different components built separately, rendering it extremely difficult to make further advancement.

Neural Machine Translation (NMT) is a recently-proposed framework for translation application based on sequence-to-sequence models: a large neural network is used to translate the source language sequence into the target language sequence. After years of development, NMT has produced richer translation results than ever over various language pairs, becoming a new machine translation model with great potential. NMT is powerful because it is an end-to-end deep-learning framework that is significantly better than SMT in capturing long-range dependencies in sentences and generalizing well to unseen texts.

One of the weaknesses of NMT is the limitation of vocabulary size due to its architecture. A usual practice is to construct a target vocabulary of the K most frequent words (a so-called shortlist), where K is often in the range of 30k to 80k. Any word not included in this vocabulary is mapped to a special token representing an unknown word (UNK). For Japanese-Chinese translation, Japanese and Chinese share Chinese characters (Kanji and Hanzi) which are logograms; it is difficult to divide a Chinese word into high-frequency subword units because many Chinese words are written with one or two Chinese character(s). Therefore, it is thought that the character-level modeling is suitable for NMT between Japanese and Chinese. The character-level NMT has also an advantage that errors and fluctuations do not occur in the word segmentation process.

Another weakness of NMT is that the NMT systems have a steeper learning curve with

respect to the amount of training data, resulting in worse quality in low-resource settings, but better performance in high-resource settings. In low-resource languages or domain-defined translation tasks, the parallel corpora is small. Therefore, studies of NMT under the condition of a low-resource language corpus have high practical value.

The contributions of this dissertation include 1) Improving Japanese-Chinese character-level NMT with radicals as an additional input feature, while some additional linguistic features of input words improve word-level NMT, any additional character features have not been used to improve character-level NMT so far. This research show that the radicals of Chinese characters (or kanji), as a character feature information, can be easily provide further improvements in the character-level NMT.

2) A corpus augmentation method for low-resource NMT, which is a solution to the poor-resource training data conditions for some language pairs like Japanese and Chinese. The method uses both source and target sentences of the existing parallel corpus and generates multiple pseudo-parallel sentence pairs from a long parallel sentence pair containing punctuation marks as follows: (a) split the sentence pair into parallel partial sentences; (b) back-translate the target partial sentences; and (c) replace each partial sentence in the source sentence with the back-translated target partial sentence to generate pseudo-source sentences. The word alignment information, which is used to determine the split points, is modified with "shared Chinese character rates" in segments of the sentence pairs. The experiment results of the Japanese-Chinese and Chinese-Japanese translation with ASPEC-JC (Asian Scientific Paper Excerpt Corpus, Japanese-Chinese) show that the method substantially improves translation performance.

This dissertation consists of the following six chapters.

Chapter 1 describes the background and purpose of this study.

Chapter 2 gives an outline of the problems of NMT, and then overviews the technical studies on NMT.

Chapter 3 describes the existing vanilla NMT models and the different types of NMT models according to the principles of classical NMT model, the common and shared problems of NMT model.

Chapter 4 proposes a method that improving Japanese-Chinese character-level NMT with radicals as an additional input feature. In experiments on WAT2016 Japanese-Chinese scientific paper excerpt corpus (ASPEC-JP), we find that the proposed method improves the translation quality according to two aspects: perplexity and BLEU. The character-level NMT with the radical input feature's model got a state-of-the-art result of 40.61 BLEU scores in the test set, which is an improvement of about 8.6 BLEU scores over the best system on the

WAT2016 Japanese-to-Chinese translation subtask with ASPEC-JC. The improvements over the character-level NMT with no additional input feature are up to about 1.5 and 1.4 BLEU scores in the development-test set and the test set of the corpus, respectively.

Chapter 5 presents a corpus augmentation method for NMT. The method has two variations: one is for all language pairs and the other for the Chinese-Japanese language pair. The method generates pseudo-parallel sentence pairs to extend the original parallel corpus. This dissertation describes the results obtained in the Japanese-Chinese and Chinese-Japanese translation with the ASPEC-JC corpus, which substantially improved the translation performance. The proposed method improvements over the character-level NMT are up to about 0.8 and 1.0 BLEU scores on the Japanese-Chinese direction, 2.4 and 2.2 BLEU scores on the Chinese-Japanese direction, in the development-test set and the test set of the corpus, respectively.

Chapter 6 summarizes the study and discusses the future work.

目次

図目次

表目次

第 1 章

緒言

1.1　研究の背景と目的

機械翻訳とは，コンピューターで自然言語を翻訳することであり，ある原言語の文を目的言語の意味的に同価な文に置き換えることである．現在の自然言語処理と人工知能の領域では，機械翻訳は極めて重要な研究として注目されている．機械翻訳の着想は 17 世紀まで遡る． 1629 年に，ルネ・デカルトは，普遍言語を提案した (Hutchins 2007)．世界中で機械翻訳は異なる言語間の障壁の解消に貢献できる技術として考えられ，1940 年代の後半から脚光を浴びるようになる，その研究は一時極めて盛んに行われた時期もあり，また低迷期もあり，70 年の年月を経て，様々な紆余曲折がありながら今日まで続けられている．

現在，インターネット技術の発展に伴い，Google 翻訳，Baidu 翻訳，Bing 翻訳などの多言語間のオンライン翻訳サービスが提供されている．機械翻訳とプロの翻訳者との間にはまだ大きなギャップがあるが，翻訳の品質がそれほど高くない領域や特定分野の翻訳作業では，機械翻訳の翻訳速度に明らかな利点がある．機械翻訳の複雑さと応用の見通しの観点から，学界や産業界は機械翻訳を重要な研究と位置付けており，自然言語処理において最も活発な研究分野の一つとなっている．

近年，日本をとりまく近隣諸国との間で，さまざまな技術的，文化的な交流および経済的な連携が盛んに行なわれてきた，これらの交流を円滑に進めるにあたって，その基礎となるさまざまな情報の導入と輸出が不可欠になっており，翻訳あるいは通訳に対する要求が高まっている．しかしながら現実として人手による翻訳は，時間的に，地域的，コスト的，数量的などの様々な面で，この要求を満たすことはとてもできない．この需給キャップを埋めるものとして，日本語と近隣諸国語の機械翻訳に対するニーズが高まってきている．

一方，日本と中国は一衣帯水の隣国で，古くから様々な交流が行われてきており，現代

では技術，経済の方面でも他の国と比較にならない緊密な関係を持ってきている．1972 年日中両国の国交正常化以来，日本は常に中国への最大投資国である．中国経済は，中国政府のいろいろな経済対策により景気の落ち込みが少なく，引き続き高い経済成長を遂げたことから，2009 年から日本の最大輸出相手になっていた．2013 年以降はアメリカが再び 1 位となったが，2018 年は中国が 6 年ぶりにトップへ返り咲いた．中国にとって，日本は米国に次ぐ第 2 位の貿易相手国である．2017 年 10 月時点の中国における日系企業拠点数は 3 万 2349 拠点（同一企業が複数事業所を有する場合は延べ数を計上）で拠点数として最多（第 2 位：米国，第 3 位：インド）である．また，2017 年の訪日中国人数は延べ 736 万人（前年比 15．4 %増）であり．訪日者数について，中国は第 1 位（第 2 位：韓国，第 3 位：中国台湾）となっている．

従って，日中の交流活動に生じる言葉の障壁の解消に寄与する日中機械翻訳スシテムは，文化交流，産業の連携活動に大きく貢献するものとして近年重視され，日本語と中国語の機械翻訳の研究は 90 年代前後から日本のいくつかの大学や研究機関で行われるようになってきている．

言語の問題は日中交流を邪魔する厳しい課題となっている．日中機械翻訳システムが実用化されれば，日中両国の科学技術の発展にとって有意義なことになると見込まれる．このような背景から高品質な日中機械翻訳システムに対する期待が高まっているといえる．

本論文ではまず，機械翻訳の歴史と研究状況について紹介する．そして，近年，注目すべき成果をあげているニューラル機械翻訳 (NMT) について述べる．

単語レベルの NMT における問題点として，語彙サイズが制限されることが挙げられる．日本語や中国語のように文中の単語の区切りが明示されない言語では，統一された正しい単語分割結果を得ることも容易ではない．文字レベルの NMT では，これらの問題を回避することができる．本研究では単語レベルの NMT において品詞などの単語の特徴情報を付加するすることで，翻訳精度の向上が図れる．何らかの文字特徴情報が有用ではないかと考え，漢字の部首および画数を入力特徴情報として加えて，文字レベルの NMT による日本語から中国語への機械翻訳を試みた．その結果，部首を特徴情報として加えることにより翻訳精度の向上が見られた．

また，NMT では学習データの量が翻訳結果の質を大きく左右する．英語を含む言語対や欧州の言語間の言語対などを除き，一般に十分な量の対訳データを入手するのは困難であることが NMT における問題点として挙げられる．日英に比べて日中の対訳コーパスは極端に少ない．そのため翻訳性能が大幅に低下する．本研究では，コーパス拡張をすることで，日本語から中国語へ，および中国語から日本語へのニューラル機械翻訳をさらに改善できないか，ASPEC-JC コーパスを用いて実験した．その結果，長文分割で逆翻訳してコーパスを拡張することにより，翻訳精度を向上させることができた．

1.2　論文の構成と概要

本論文の構成は以下のとおりである．

第 1 章　本章であり，本研究の背景，目的及び論文の構成を示す．

第 2 章　機械翻訳の研究の歴史と現状を概観する．

第 3 章　ニューラル機械翻訳 (NMT) について述べる．

第 4 章　文字レベルの NMT では，語彙サイズが制限されることの問題を回避することができる．本章では文字の特徴情報の一つとして漢字の部首を用いる．そこで，部首がもつ意味的な情報が翻訳精度の向上につながることを期待して，入力特徴情報に加えた．日本語から中国語への文字レベルのニューラル機械翻訳をさらに改善できないか，ASPEC-JC コーパスを用いて実験した．

第 5 章　NMT の一つの問題として，翻訳の品質が学習のための対訳データの量に強く依存する．本章では単語アラインメント情報を利用して長い対訳文から短い対訳文（あるいは句）を作成し，目的言語側の短文を NMT で逆翻訳して原言語短文を得た後，元の原言語文の一部を逆翻訳結果の短文と入れ替えて擬似的な原言語文を生成して対訳データを拡張する方法を提案する．

第 6 章　研究結果のまとめと今後の課題について述べる．

第 2 章

機械翻訳の歴史と現状

本章では機械翻訳の背景となる機械翻訳の歴史と方式をいくつか紹介するとともに，日中両言語の比較対照とニューラル機械翻訳の現状についても述べる．

2.1 機械翻訳の全般的な研究状況

2.1.1 機械翻訳の歴史

> 昔は世界の言語はひとつだったが，天まで届こうかという塔を築こうとする人間の尊大さをこらしめるため，神が人間の言葉を乱した．これにより，意思疎通が難しくなり，塔の建設も中止せざるをえなくなった．人々は散り散りになり，それにしがたい使う言葉も分かれていった．
> (大意)
>
> ---
>
> *The Tower of Babel.*
> 参照：旧約聖書創世記第 11 章.

現在，世界には 6,000 以上の言語がある．図 2.1 に示すように，異なる国や地域で話されている言語は非常に異なっていることがわかる[*1]． 異なる言語間のコミュニケーションには多くの問題があるが，これらの問題は通常「言語障壁」と呼ばれる．

機械翻訳は，言語障壁の問題を解決するための重要な技術である．

「機械翻訳」という概念の着想は，17 世紀にまでさかのぼる．フランスの数学者であったルネ・デカルトは，多言語間の同意語に単一の記号を割り当て普遍化する言語を提唱した．

[*1] 出典： https://en.wikipedia.org/wiki/Linguistic_map

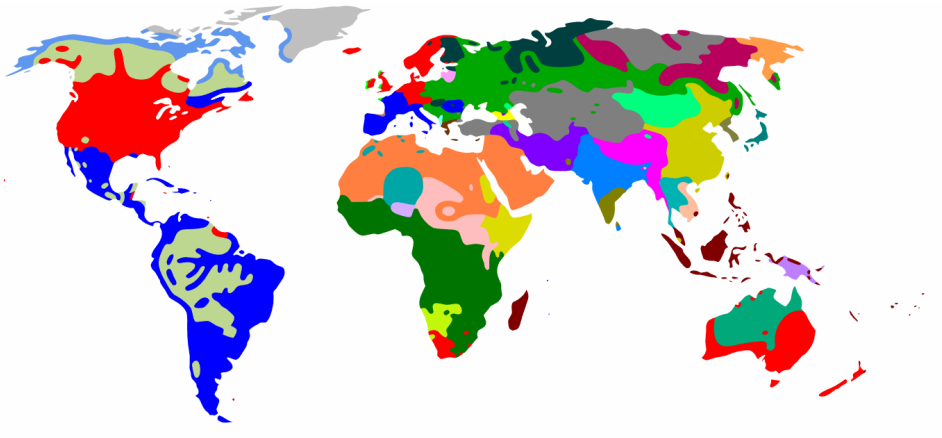

図 2.1　世界の言語地図，Wikipedia の “Linguistic_map” より転載

1947 年に Warren Weaver が送った手紙の中で，

> When I look at an article in Russian, I say “This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.”

と語ったことが最初だといわれている*2. そのときに考えたのは英語とフランス語の間の翻訳であった (Weaver 1949)．これを契機にして，機械翻訳という考え方は多くの研究者の興味を引き，1950 年代にはコンピューターを活用して実用化するための研究が始められ，1955 年にジョージタウン大学と IBM との共同で，ロシア語から英語への機械翻訳の最初のデモが行われ，学者たちの興味をひきつけるようになった (Sheridan 1955)．その後，機械翻訳研究はアメリカを中心に，ソ連，ヨーロッパ，日本に広がり，1960 年代の半ばまでは活発に行われた．日本では 1955 年あたりから九州大学と電気試験所（現在の産業技術総合研究所）で機械翻訳の研究が始まり，1959 年には電気試験所で「やまと」と称する英日機械翻訳システムが公開された*3.

1960 年前後に機械翻訳の開発のピークを迎えた. アメリカ政府は機械翻訳の研究にかなりの財政援助を行っており，近い将来に実用的な翻訳システムができると期待していたが，その時期の科学技術に対して非常に難しい課題であるために成功せず，1966 年，米国の自動言語処理諮問委員会 (Automatic Language Processing Advisory Committee, ALPAC) は，機械翻訳を「高価で，不正確，そして見込みがない」と報告し，機械翻訳に否定的な態度を示した. その後，機械翻訳の研究が長く停滞の時期を経て来た.

1970 年代には，ルールベースの機械翻訳（Rule-Based Machine Translation: RBMT）が次の主役の座をしめた．この技術は，ある言語を別の言語に切り替えるためにプログラ

*2 https://en.wikiquote.org/wiki/Warren_Weaver
*3 http://museum.ipsj.or.jp/computer/dawn/0027.html

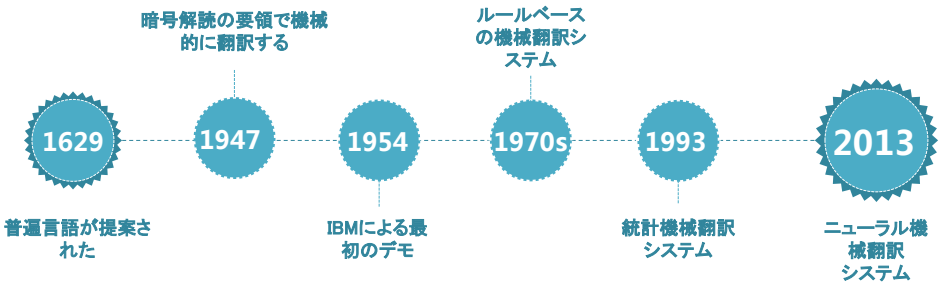

図 2.2　機械翻訳の歴史

マーによって入力され人手で整備されたルールに基づくものを適用した．しかし，微妙な言語間のニュアンスの違いが拾われることなく，抜け落ちてしまうという欠点がある．

1980 年代には一定の成果をあげ始めた．1984 年には京都大学の長尾氏がアナロジーに基づく翻訳を提案した．これは過去の翻訳用例（人手で整備された）を組み合わせることで新たな文の翻訳を実現するというもので，ルールに基づく方法とはまったく異なるアプローチであった．この方法は用例に基づく翻訳（Example-Based Machine Translation: EBMT）とも呼ばれる．

1980 年代後半には IBM の研究グループが統計的機械翻訳（Statistical Machine Translation: SMT）の研究を開始した (Brown et al. 1993)．これは単語の翻訳確率や並べ替えの確率などの翻訳に必要な知識を対訳コーパスから統計的な情報として学習するものであり，これを拡張したものが 2003 年に提案された句に基づく翻訳（Phrase-Based SMT: PBSMT）で，最近までスタンダードな機械翻訳手法として広く使われていた．2010 年前後は集積された膨大な「ビッグデータ」を活用した，「統計翻訳」の時代であった．統計翻訳自体は 1980 年頃から IBM によって研究が進められていたが，この時代になって成果が認められるようになった．

2013 年頃には，「統計翻訳」から人間の脳神経の活動方式を模倣したニューラルネットワークを利用した「ニューラル機械翻訳」（Neural Machine Translation: NMT）の時代へと移り変わった．Google が公開した機械翻訳システムは，統計翻訳以上に自然で正確な訳文の出力を実現した．

図 2.2 に以上の機械翻訳技術の変遷を示す．

2.1.2　機械翻訳の方式

コンピュータが生まれた 1940 年代から，機械による翻訳に関する技術の研究開発は始まり，半世紀以上の時を刻んでいる．全領域で実用的な翻訳はまだできていないと言わざるを得ない．従来の研究と最近の研究で現れた翻訳方式についていくつか紹介する．

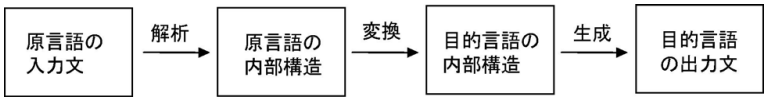

図 2.3　トランスファー方式

直接翻訳方式

直接翻訳方式は，計算言語学理論がまだ提案されていなかった初期の機械翻訳システムで用いられた方式である. 原言語文の形態素解析，対訳辞書による目的言語への単語変換，目的言語の形態素レベルの生成からなる．単語直接置き換え方式である．この方式は，表層的な解析だけを行うため，原言語の曖昧さを解消すること，目的言語に正確な表現で訳出すること，訳文の正確な語順規則を記述することができない．

トランスファー方式（変換方式）

トランスファー方式の翻訳システムでは，原言語の構文解析，意味解析を行う．それを目的言語の内部表現に変換して，そこから目的言語を生成するという流れで翻訳を行う．

まず，原言語が入力されると，原言語辞書を用いて，形態素解析，構文解析を行い，原言語の構文解析結果を得る．この時点でこの構文解析結果は原言語に依存した構文である．形態素解析は，ある文を単語や接辞などの文法上最小単位となる要素の列に分解し，その要素の品詞属性や活用形などを定める．また，この段階で未知語の確定や複合語の分割という問題も処理し，後の処理負担を軽減する．構文解析は，形態素解析で得られた情報を基に文の構文構造を決める．原言語の構文解析結果を変換処理によって，目標言語に依存した構造に変換する．この時点で対訳辞書と変換規則を利用する．

対訳辞書には，類似単語の曖昧性を解消するために，終端記号としての原言語の単語を目標言語の単語，即ち訳語に置き換える．実翻訳システムは対訳辞書の条件を検査し，条件に最も一致する訳語を選ぶ．表現構造の対応づけは，一般に両言語の木構造やグラフ構造間で対応をつけ，各ノードや部分構造を変換規則データベース上で検索する．格文法や結合価パターンがよく利用される．

生成の段階では，目的言語の構文表現より文を生成する．目的言語の特徴によって，様相変形，語順調整，省略補完などの処理を行う．

全部の流れは図 2.3 の通りである．

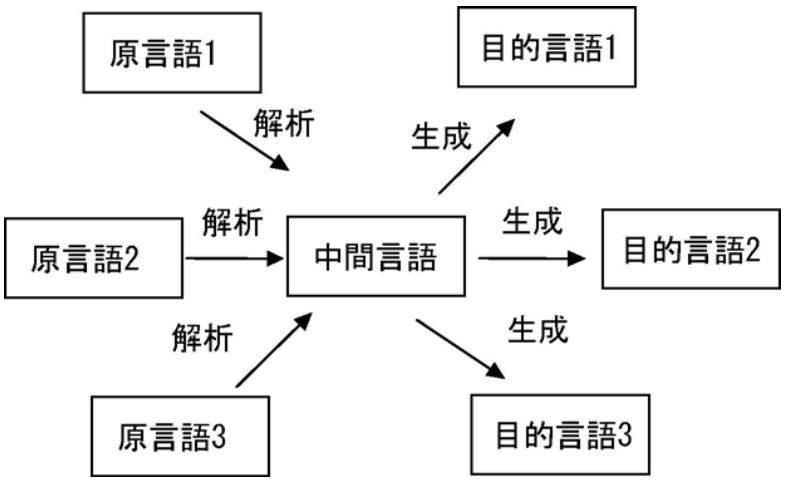

図 2.4　中間言語方式

中間言語方式

中間言語翻訳方式は一度言語に依存しない意味表現である中間言語というものに変換してから目的言語の訳文を生成するというものである．しかし言語に依存しない抽象的な意味表現が存在するかという問題もあるため，広く商用開発で利用されるまでには至っていない．

図 2.4 のように, 中間言語方式によれば，まず原言語の文が入力されると，原言語辞書を用いて解析が行われ，中間言語に変換される．この中間言語より目的言語辞書を用いて目的言語の文が生成される．しかし，すべての言語表現を網羅するような中間言語を人工的に設計されるのは容易でない．

用例翻訳方式

用例翻訳の最初の枠組みは 1984 年に京都大学の長尾らによって提案され，90 年代の翻訳システムに大きな影響を与えた (Nagao 1984)．用例翻訳方式では，原言語の文の翻訳を，それとよく似た文の翻訳例（用例）を見つけ，それを模倣することによって行う．用例翻訳方式では翻訳の例文を記憶した「用例辞書 (対訳用例データベース)」と単語対応を記憶した「単語辞書 (シーソーラス)」を使用する．システムの流れは以下の通りである (図 2.5)．1. システムに原文 f が与えられる．2. 用例辞書から f と似た文 f’とそのペア (f, f’) を検索する．3. システムは，f と f’の差分を取る (類似度判定のため)．4. 訳文内の適切な単語を単語辞書により置き換える．5. 置き換え結果を訳文として出力する．この方法ではトランスファー方式によく使われる文法規則も利用し，翻訳過程もトランスファー方式と似ているので，用例トランスファー翻訳方式とも呼ばれる．

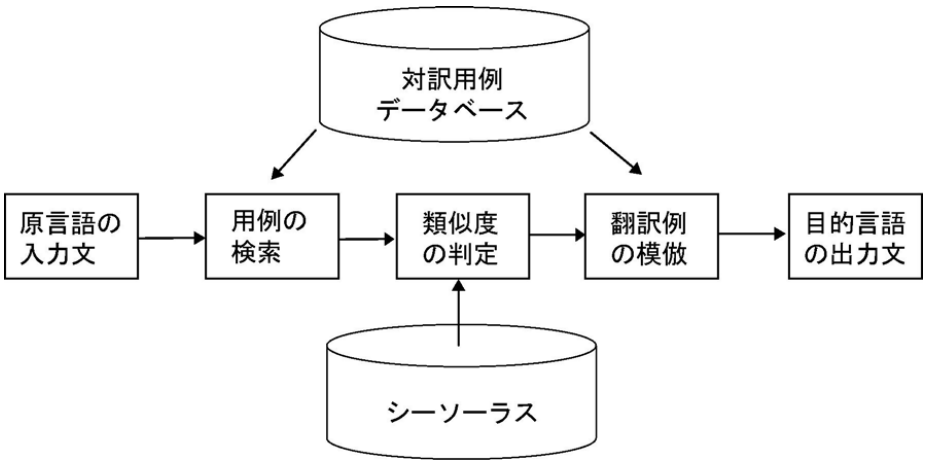

図 2.5　用例翻訳方式

統計翻訳方式

計算機の発達によって 1990 年代以降研究が盛んになったのは，統計的な手法を用いた機械翻訳である．統計に基づく翻訳方式では，パラレルコーパスと呼ばれる複数の言語で文同士の対応が付いたコーパスを利用し，翻訳のルールを自動的に獲得し，各ルールの重要度を統計的に推定する．パラレルコーパスには自前のデータを利用することもあるが，最近では各言語に翻訳された特許や，Web ページのクローリングデータなどを利用することもある．この理論は本来，従来音声認識の分野で用いられていた雑音チャネルモデルを応用したもので，これを原言語から目的言語への翻訳に適用する．

原言語の文 S が雑音のある通信路を通って，目標言語の文 T になったと考え，T から元の S を推測する．S の推定値 S' として，T が与えられた時に S の条件付き確率 $P(S|T)$ を最大にする S を求めれば，誤りを最小にできる．$P(S)$ を言語モデル，$P(T|S)$ を翻訳モデルという．この二つのモデルの要素は大規模言語データベースを用いて自動推定される．

ニューラル機械翻訳方式

新しい手法として 2014 年に登場したのが，ニューラルネットワークを用いた機械翻訳，いわゆるニューラル機械翻訳（Neural Machine Translation: NMT）である．NMT は原言語文に対する目的言語文の条件付き確率を計算する．NMT では 1 つのニューラルネットワークを用意するだけで訓練も翻訳もすべて同じ枠組みで行うことができる．対訳コーパスを与えるだけで，ニューラルネットワークが翻訳に必要な何らかの情報を自動的に学習するのである．

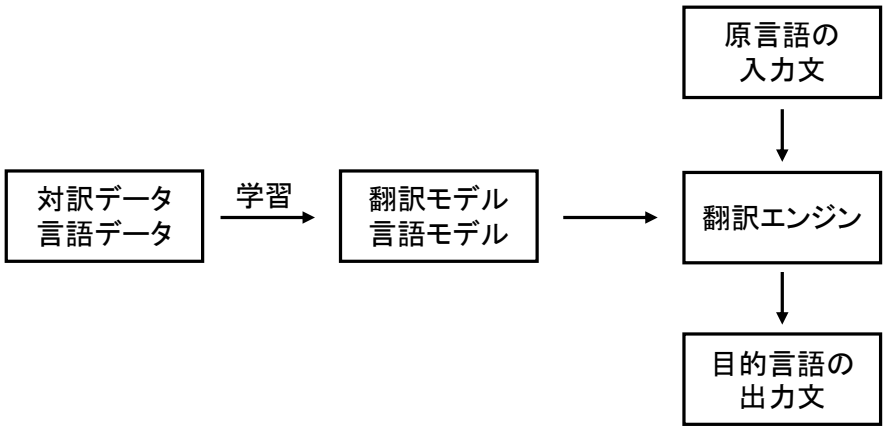

図 2.6　統計翻訳方式

具体的な内容は第三章で紹介する.

2.1.3　Web 上の翻訳サイト

インターネットの普及に伴い，無料で使える翻訳サイトが多く現れてきた．日中/中日翻訳サイトもいくつが挙げられる．これらの翻訳サイトはテキスト翻訳だけでなく，Web サイトの翻訳にも対応する.

A）　Infoseek マルチ翻訳 https://translation.infoseek.co.jp/
B）　Excite 翻訳 https://excite.co.jp/world/
C）　Yahoo!翻訳 https://honyaku.yahoo.co.jp/
D）　Youdao 翻訳 https://fanyi.youdao.com/
E）　Google 翻訳 https://translate.google.co.jp/
F）　Baidu 翻訳 https://fanyi.baidu.com/
G）　Bing 翻訳　 https://bing.com/translator

これらの商用システムの翻訳結果を見ると，簡単な文や文の基本的構造部分は大体うまく翻訳できている．しかし，多義語の選択や，語順，慣用句への対応が不十分であるように見られる．助詞，とりたて表現のような両言語における対応が複雑な部分の翻訳と長文の翻訳にも多くの問題が見られる.

2.1.4　機械翻訳の翻訳結果に対する主な評価尺度

ここで機械翻訳の翻訳結果における主な評価尺度を紹介する.

自動評価尺度 BLEU

BLEU(BiLingual Evaluation Understudy，BLEU) は，機械翻訳システムの自動評価において，現在主流の評価法である (Papineni et al. 2002)．BLEU は，$N-gram$ 適合率で評価を行う．一般的には $4-gram$ を用いる．BLEU は 0 から 1 のスコアを算出し，スコアが大きい方が良い評価である．BLEU の計算式を以下に示す.

$$BLEU \quad = \quad B \cdot \exp\left(\frac{1}{N}\sum_{n=1}^{N}\log P_n\right)$$

$$\beta = \begin{cases} 1 & (c > r) \\ e^{1-\frac{r}{c}} & (c \leq r) \end{cases}$$

$$W_n = \frac{1}{N}$$

$$P_n = \frac{\sum_i \text{出力文中 } i \text{ と参照文中 } i \text{ で一致した } N-gram \text{ 数}}{\sum_i \text{出力文中 } i \text{ の中の全 } N-gram \text{ 数}}$$

ここで，B は短い翻訳文が高い評価にならないように補正を行うパラメータである．c は出力文の長さ，r は参照文（正解文）の長さを表す．また W_n は $N-gram$ の重みである．

単語誤り率 WER

音声認識の評価などで広く用いられる尺度として単語誤り率 (Word Error Rate，WER) がある (Klakow & Peters 2002)．単語誤り率は，まず「挿入 (I)」「削除 (D)」「置換 (S)」という 3 種類の編集操作を定義し，システム出力を参照文へと変更するのに必要な編集操作を参照文の長さ R で割ったものとして求められる．スコアが小さい方が良い評価である．

$$WER = \frac{I+D+S}{R}$$

WER は BLEU が開発される前から，音声認識などの評価で広く使用されていた．

翻訳編集率 TER

TER (Translation Edit Rate，TER) は数値化することにより，翻訳結果の後編集を行った際のコストに着目した評価尺度である (Snover et al. 2006)．すなわち，入力の変化に対して，出力がどの程度変化するかを数値化するということである．TER は，機械翻訳の出力を参照訳に近づくよう編集した際の編集距離を評価するものである．単語の削除，挿入といった編集操作だけでなく単語のシフト操作も考慮するため，単純な編集距離よりも適切な編集操作を評価できる．スコアが小さい方が良い評価である．

具体的には，通常の WER で対象となる挿入，削除，置換以外に，「並べ替え」操作も加える．これにより，「white bird」を「bird white」に変更するために，2 回の置換ではなく，white を bird の後へ並べ替えるという 1 回の操作だけで済む．

2.2　日中両言語の比較対照

日本語と中国語は漢字圏の言語であるので，漢字を見ると似通っていて，理解やすいように思う人が多いかもしれない．しかし，共に漢字を使うが，読み方と意味は全く異なっており，また両言語は構造上かなり違う言語である．ここで日中両言語の異同を紹介する (張 2016; 頼 2015; 陳 2013; 方 2013; 池田 2009; 王 2008; ト 2004; 謝 2003)．

2.2.1　言語の類型

形態的類型論では，基本的に，言語は孤立語・膠着語・屈折語の 3 つの類型に分類される．抱合語または複統合語を第 4 の類型として加えることもある．Wikipedia ではこれらの類型を以下のように説明している．

孤立語は各語が概念のみを表し，語の文法的な役割は語順により表され，また語形変化がない．中国語，タイ語，ベトナム語などはその類である．

屈折語は単語の文法的な機能が屈折（動詞や名詞の活用）により表され，文中の位置によって左右されることが少ない．ラテン語やアラビア語がこれに属する．

膠着語は実質的な意味を持つ語や語幹に接辞（助詞，助動詞）などの付属的形式が比較的に緩やかに接合することによって，その文法的機能が示される．日本語や朝鮮語などはその類である．

抱合語は目的語などを表す諸要素が動詞に付着して一語文を形成するもので，アイヌ語や多くのアメリカ･インディアン語がある．

英語は普通屈折語に分類されるが，実際には孤立語や膠着語両方の特徴も持っているといえる．日本語も膠着語であるが，用言の活用は屈折語的特徴も示す．

2.2.2　日中両言語の異同比較

日本語と中国語はそれぞれ膠着語と孤立語に属し，構文構造や，言葉の持つ意味に対する捉え方などに違いが現れる．一方では，日本と中国はは古くから今日に至って文化における交流が盛んに行われてきて，言語にその交流の痕跡を反映する点も多く見られる．以下に日本語と中国語のいくかの基本的な異同を簡単にまとめてみる．

相違点

a 語順：日本語文は基本 SOV の順番で並べるが中国語文は基本 SVO の順である．これは基本的語順の異なりであるが，その他，部分的な言語現象にも，語順の違いが目立つ．例えば否定辞の語順，取立て詞の語順において両言語の違いは大きい．従って日中機械翻訳においてはこれらの部分の翻訳に，語順に関する情報が必要である．

b 助詞の使い方：中国語では構文機能を果たす助詞はまれであり，単語の文中での役割は主にその語順で決まる．一方日本語では「が，を」などの大量の助詞が使われ，単語の構文上の機能もその付属する助詞により相当に明瞭である．

c 語形変化：日本語の用言は語形変化が多様であり，その語形変化あるいは語形変化＋助詞/助動詞によって，テンス，アスペクト，様相，可能，仮定などを表現するのに対して，中国語の謂詞は語形変化がないため，助詞，介詞（前置詞），副詞などの付着により上記の文法機能を表す．

d 敬語の使い方：日本語は複雑な敬語の使用があり，第二人称の省略が頻繁にある．中国語の敬語表現は日本語より少ない上，使われる場合も限られている．さらに日本語の授受動詞と敬語を結びつけて使う表現は日中機械翻訳では厄介なことになる．

e 補語構造：述語の動詞，形容詞を修飾する成分は，日本語では連用修飾語があるが，中国語では状語と補語という二種類に分けられる．状語の位置は連用修飾語と基本的に同様であり，述語の前にある．補語は述語の後に位置し，日本語では，連用修飾語や，連用修飾節，あるいは複合動詞の一部になったりする．

f 特殊文型：中国語の「把」字文，連動文などの特殊文型は日本語との対応が複雑である．

g 表記のゆれ：日本語の表記の形として，漢字，平仮名及び送り仮名などがある．このため，多くの場合一つの語に複数の表記が可能である．例えば「言い替える」は「言い換える」,「言換える」,「言替える」,「言いかえる」などの表記の仕方がある．一方，中国語では単語は全て漢字で表示し，日本語のような表記の多様性はないが，同形多音字と異形同音字がある．例えば,「重（chong）复」と「重（zhong）要」は同形多音字である．異形同音字も数多くある．例えば「tong」という発音の二声調のものでも「铜」,「同」,「筒」,「桐」,「童」,「瞳」などがある．声調の制限がないと，同音の漢字の数はその 4 倍以上増えていく可能性がある．これは音声翻訳においては問題となりやすい．

h 外来語：日本語における借用語のうち，漢語とそれ以前の借用語を除いたものである．洋語（ようご）とも呼ばれる．英語などの音訳に漢字を当てたものは一般に外来語と見なされない．中国語には全くない．

相似（類似）：

a 語彙の類似性：日本語でも中国語でも使用される共同の漢字が 2000 以上あり（例え

ば日，学，国など），中国語の常用の名詞，動詞及び形容詞の中で，形も意味も日本語と同様のもの或いは日本人がその意味を推測できるものは約全体の 50 %を占めているという．

b 数字，日時の表記はほぼ同様：例えば，2016 年 8 月 1 日→ 2016 年 8 月 1 日．二千五百→二千五百

c 連用修飾語の語順：中国語にも日本語の連用修飾語に相当する成分（状語）があり，且つその語順も同じく述語の前に位置する．

d 連体修飾語における類似点：両言語とも連体修飾語＋中心語の語順である．日本語の連体修飾語の中の「N1 の N2」という構成は中国語の「N1 的 N2」の構成に対応できるものが多い．

e 発音：「ん（n，ng）」以外の音節が全部開音節（母音で終わる）であることが似ている．

f 文法における類似点：動詞と目的語の位置関係は逆だが，それ以外の語順は比較的似ていて，特に日本語「いつ」「どこで」中国語「何时」「哪里」などの副詞の位置はよく似ている．また，平叙文の文末に助詞（日本語「か」，中国語「吗」）をつけると，そのまま疑問文になるところや，人称代名詞などの後ろに接尾語（日本語「たち」，中国語「们」）をつけると，そのまま複数形になるところも似ている．

日本語と中国語には多くの違いがあるが，ニューラル機械翻訳の登場により，日中対訳コーパスがあれば，翻訳モデルを訓練するだけで従来の翻訳方式より良い翻訳結果を得ることができるようになった (Wang et al. 2017; Zhang & Matsumoto 2017; Meng et al. 2019a).

2.3　ニューラル機械翻訳の現状

1980 年代以降，Back Propagation（BP）が多層階層型ニューラルネットワークの学習方法として Multilayer Perceptron（MLP）に導入された．入力層へ或る情報が与えられると，出力層はそれに対応した或る情報を出力の学習方法となる．出力結果を元にニューラルネットワーク全体の修正をその都度を行っていく仕組みである．それ以来，Hinton，LeCun，Bengio などの研究者たちの推進力のもと，ニューラルネットワークは世界の研究者の注目を集めた．

2006 年に，Hinton ら (Hinton et al. 2006) は，階層ごとの事前訓練方法によってニューラルネットワーク訓練の問題を解決した．後に，並列計算，グラフィックス処理装置（GPU）などの計算能力の増大によって，ニューラルネットワークは学界および産業界において高く評価されてきた．

ニューラルネットワークは人間の神経細胞における情報伝達の仕組みを模した計算モデルであり，数年で画像，音声，人工知能，自動運転などさまざまな分野において大きな成果を上げていた．機械翻訳を含む自然言語処理も，その恩恵を受け，それまでの成果を大

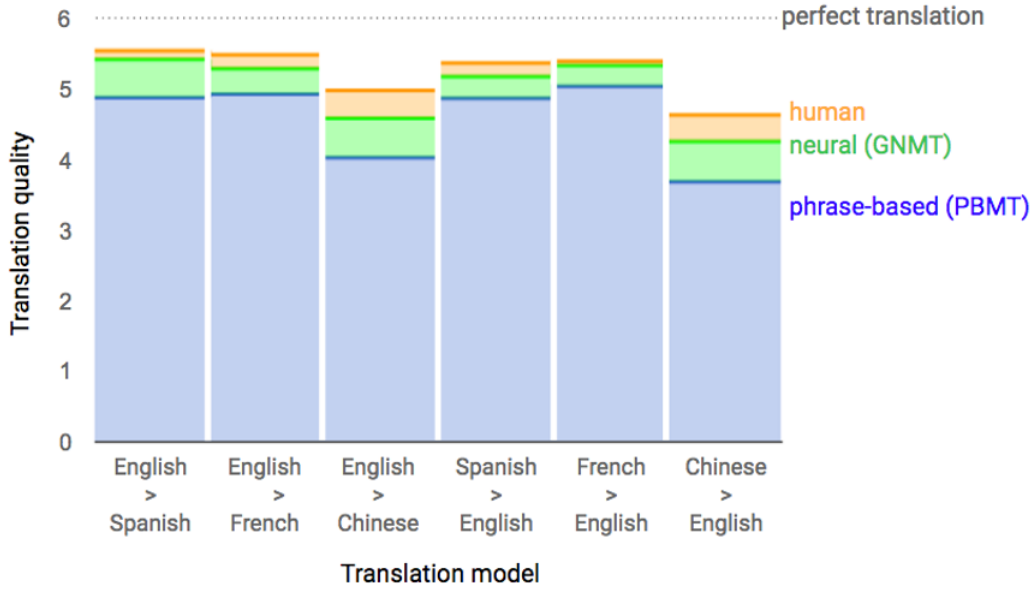

図 2.7　Google のニューラル機械翻訳のパフォーマンス，Google Research Blog より転載

きく上回る結果を残していた (Nakazawa 2017).

2014 年以降，Sutskever ら (Sutskever et al. 2014) と Jean ら (Jean et al. 2015b) はニューラルネットワークに基づいた機械翻訳モデルを実装した．スタンフォード大学の自然言語処理研究室もニューラル機械翻訳システムを開発した (Luong & Manning 2015).

2016 年，Junczys-dowmunt ら (Junczys-dowmunt et al. 2016) は，United Nations Parallel Corpus v1.0 を使用して，30 言語ペアでニューラル機械翻訳と統計機械翻訳を比較した．統計翻訳方法 (中国語-英語，中国語-ロシア語，中国語-フランス語) の翻訳タスクについて，ニューラル機械翻訳は BLEU 値で統計機械翻訳より 6~9% 向上した．大規模な GPU と並列計算に支えられて，Baidu はディープニューラルネットワークアーキテクチャを利用し，機械翻訳ワークショップ (WMT2014) の英仏翻訳タスクで初めて統計機械翻訳を上回り，その時点に最良の結果を達成した (Zhou et al. 2016).

さらに，2017 年の Proceedings of the Conference on Machine Translation(WMT) では，エジンバラ大学で開発されたニューラル機械翻訳システムが，英語からドイツ語への翻訳タスクにおいて統計翻訳を超えた (Sennrich et al. 2017).

業界では，NMT が提案されてしばらくすると Baidu，Google，Microsoft などの大手 IT 企業も NMT の実用化を始めた．中でも 2016 年 11 月に Google 翻訳が自社開発の NMT を採用したときには大きな話題となった (Johnson et al. 2017)．大規模対訳コーパス，巨大な NMT モデル，大量の GPU を生かして高精度な機械翻訳を実現していた．図 2.7 は Google 翻訳のパフォーマンスを示す.

よく知られている商用の機械翻訳会社 SYSTRAN も，12 種類の言語から 32 言語ペアをカバーするニューラル機械翻訳システムを開発した (Crego et al. 2016).

自然言語の多様性と複雑さのために，ある言語を別の言語に翻訳することは依然として困難である．現在，大規模なコーパスと計算能力の条件下で，ニューラル機械翻訳は大きな可能性を示し，新しい機械翻訳方法へ発展してきた．この方法は大規模な翻訳モデルを訓練するのに対訳コーパスだけを必要とし，それは高い研究価値を有するだけではなく，強力な工業化能力を有している．

2.3.1　統計翻訳との比較研究

ニューラル機械翻訳 (NMT) は，ニューラルネットワークを使用して，原言語から目的言語への直接翻訳を実装する．全体として，この方法はブラックボックス構造に似ており，単語のアライメント，言語モデル，翻訳モデルなどの統計機械翻訳 (SMT) の必要な部分に使用でき，暗黙的な方法で実装される．

表 2.1　NMT と SMT の差異

評価方法	NMT	SMT
数学表示	連続	離散
モデル	非線形	対数線形
モデルのパラメーターの数	少	多
訓練時間	長	短
モデルの可解釈性	弱	強
メモリ使用量	小	大
GPU	必要	不要

ニューラル機械翻訳と統計機械翻訳の違いは次のとおりである．

1. 単語アライメント：原言語と目的言語の単語間の対応をモデリングする単語アライメントは，統計的機械翻訳の重要な部分である．ニューラル機械翻訳モデルでは，単語のアライメントは不要であり，Attention メカニズムに基づいて，デコード中に生成された単語に関連するソース言語の単語情報を自動的に取得できる．Attention メカニズムを使用して単語のアライメント情報を取得できるが，単語のアライメントは統計的な機械翻訳の単語のアライメントよりも少ない情報をしか持っていない．
2. 翻訳効果の比較：ニューラル機械翻訳は，ソース言語情報と生成された翻訳情報を

使用して翻訳を生成する．これは，複数のモジュールをシームレスに統合するのと同等である．

実験により，ニューラル機械翻訳の翻訳結果の流暢さは統計的機械翻訳の翻訳結果よりも優れていることが示されており，統計的機械翻訳を処理するのが難しい，複雑な構造順序付けおよび長距離順序付け問題も処理できる (Junczys-dowmunt et al. 2016).

上記に加えて，ニューラル機械翻訳と統計的機械翻訳の違いを表 2.1 に示す．NMT と SMT は，それぞれニューラル機械翻訳と統計的機械翻訳を表す．

第 3 章

ニューラル機械翻訳について

本章では，ニューラル機械翻訳について全般的に紹介するとともに，ニューラル機械翻訳のいくつかの問題点にも触れる．

3.1　基本的なニューラル機械翻訳モデル

統計機械翻訳では，翻訳問題を確率問題と同等と見なす．つまり，原言語 s が与えられた場合，目的言語 t の条件付き確率 p を求める． 翻訳モデルを選択した後，これらのモデルのパラメータは，コーパスから学習される． 原言語が入力されると，学習されたモデルによって上記の条件付き確率 p が最大化され，最適な翻訳結果が得られる．上記の考え方に基づいて，ニューラル機械翻訳はニューラルネットワークを使用して，原言語から目的言語への直接翻訳を実現する．

1990 年代，一部の学者たちは小規模なコーパスを使用してニューラルネットワークベースの翻訳方法を実装した (Neco & Forcada 1997; Castaño & Casacuberta 1997)．しかし，コーパスと計算能力の制限により，注目を集めなかった．深層学習のブームの到来後，統計翻訳の翻訳モデルを計算するためにニューラルネットワークがよく使用される．単語の整列，翻訳ルールの抽出など (Zhang & Zong 2015)． 2013 年，ニューラルネットワークベースの変換方法は Kalchbrenner ら (Kalchbrenner & Blunsom 2013) によって再提案され，応用に大きな可能性が示された．その後，Sutskever(Sutskever et al. 2014), Jean ら (Jean et al. 2015a,b) と Cho ら (Cho et al. 2014b,a) などがそれぞれ，完全なニューラルネットワークベースの機械翻訳モデルを提案し，実装した．これらのモデルは，機械翻訳だけでなく，質問応答システムやテキスト要約などの他の自然言語処理タスクにも適用できる，本質的に系列から系列へのモデルである基本的なニューラル機械翻訳モデルに属す．

3.1.1　フィードフォワードニューラルネットワーク

フィードフォワードニューラルネットワークとはニューラルネットワークの最初に考案された単純な構造で，データの流れが 1 方向だけで，入力ノード→中間ノード→出力ノードというように，データが行き来したり，ループしないような構造になっているものである．順伝播型ニューラルネットワークとも呼ばれる．

具体例で説明する．図 3.1 のように，一番左側は入力層，真ん中が中間層，一番右が出力層と言う．入力層は，入力されたデータをそのまま出力するだけである．中間層には，入力されたデータの各成分に適当な重みを付けて和を取ったものが入力される．出力層には，中間層から変換を行い値を出力する．次に図の各青い丸は，ノードと言う．中間層の一番上のノードに対しては入力層の値が

$$(x_1, x_2, x_3)^T$$

であれば，

$$w_1 x_1 + w_2 x_2 + w_3 x_3$$

が入力される．このときの w を重みと言う．中間層の一番上のノードは

$$a_1 = w_{11} x_1 + w_{12} x_2 + w_{13} x_3 + b_1$$

という値になる．このときの b_1 をバイアスと言う．

次に中間層は、入力された値に対して $h\left(a_j\right)$ を行う．中間層の一番上のノードは

$$h\left(a_1\right) = h\left(w_{11} x_1 + w_{12} x_2 + w_{13} x_3 + b_1\right)$$

を計算する．このときの変換 $h(\cdot)$ を活性化関数と言う．中間層はこのようして変換を行った値を，出力層に渡す．出力層は中間層から受け取る値に対して重み付き和を得て，その値からバイアスを加えて，活性化関数によって変換をして何らかの値を得る．

ニューラルネットワークの学習は出力誤差（ニューラルネットワークの出力値と真値として訓練データの誤差）を最小化する最適化問題を解くこと．最適化問題の解法は誤差逆伝播法などの最適化アルゴリズムを使うのが一般的である．誤差は二乗和誤差を使うのが一般的である．汎化能力を高めるために，誤差に正則化項を加算することが多い．

活性化関数

ニューラルネットワークのモデルにおいて，シナプスに相当するもの．ある値を超えると急に出力値が大きくなる（発火する）関数を利用する．シグモイド関数，ReLU 関数などが用いられる．

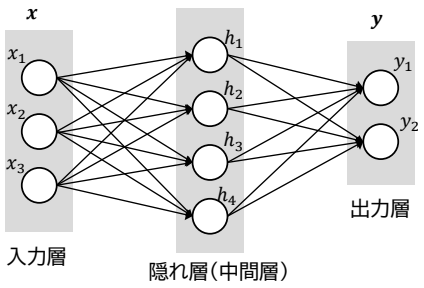

図 3.1　フィードフォワードニューラルネットワークの構造図

損失関数

損失関数は出力誤差を返す，重みとバイアスを調整する関数．クラス分類の場合は，クロスエントロピー関数（交差エントロピー誤差）を用いる場合が多い．回帰問題の場合は二乗誤差関数などを用いる．損失関数で求めた値をできるだけ小さくなるように，重み，バイアスを調整することが最適化アルゴリズムという対策になる．以下に最適化アルゴリズムを紹介する．

1. 誤差逆伝播法：多層パーセプトロンにおいて，入力データを伝達させて値が出力された時，正解値である教師データとの誤差の最小二乗和を計算し，各層の結合重みを調整することで，誤差を減らす方法．この方法によって，パーセプトロンの多層化が可能となる．
2. 勾配降下法：ある関数の最小値を求める方法．ある変数の値に応じた関数の勾配を求め，変数を勾配の方向に動かし勾配計算を繰り返すことで最小値を得ること．
3. 確率的勾配降下法（SGD）：大量のデータがある場合，勾配降下法の計算量を減らすため訓練データの一部をランダムに選ぶ勾配降下法の手法である．

3.1.2　再帰型ニューラルネットワーク（RNN）と長短期記憶ニューラルネットワーク (LSTM)

本節では Olah 氏の解説記事 (Olah 2015) を参考に，RNN と LSTM の概略を述べる．本節の図は同記事から転載したものである．

RNN とは，自己回帰型の構造をもつニューラルネットワークの総称であり図 3.2 のような構造をもつ．A は，入力 X_t を見て，値 h_t を出力する．ループは，情報をネットワークの一つステップから次のステップに渡すことを可能にする．

RNN の R は Reccurent（再帰）という意味で，直前の計算に左右されずに，連続的な要素ごとに同じ作業を行わせることができる．言い方を変えると，RNN は以前に計算され

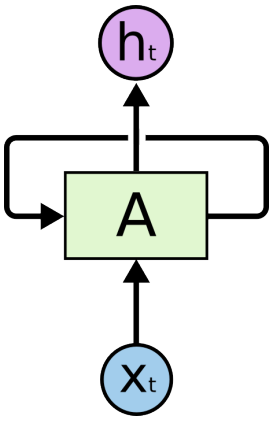

図 3.2　一般的な RNN 構造図

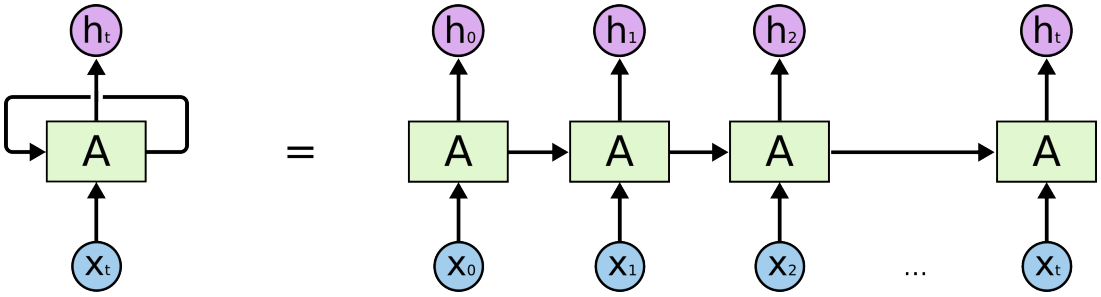

図 3.3　展開された RNN 構造図

た情報を覚えるための記憶力を持っている．理論的には RNN はとても長い文章の情報を利用することが可能である．従来のニューラルネットワークが，前の出来事についての推論を後のものに教えるために，どのように使用できるかは不明である．RNN は，この問題に対処する．

この鎖状の性質は図 3.3 に示すように，リカレントニューラルネットワークが配列やリストに密接に関連している．それは，このようなデータに使用するための自然なアーキテクチャである．

RNN の特長の 1 つは，前のビデオ・フレームの使用が現在のフレームの理解を助けるように，前の情報を現在のタスクに関係づけることができるというアイデアである．例えば，これまでに現れた単語列に基づいて，次の単語の予測を試みる言語モデルについて考える．「the clouds are in the sky」の最後の単語を予測する場合，次の単語が sky になることはかなり明白であり，これ以外のコンテキストは必要ない．このように関連する情報とそれを必要とする場所の隔たりが小さい場合，RNN は過去の情報を利用することを学習することができる．しかし，関連する情報とそれを必要とする場所のギャップが非常に大きくなることもある．残念ながら，ギャップが大きくなるに従い，RNN は情報を関連づけて学習することができなくなる．

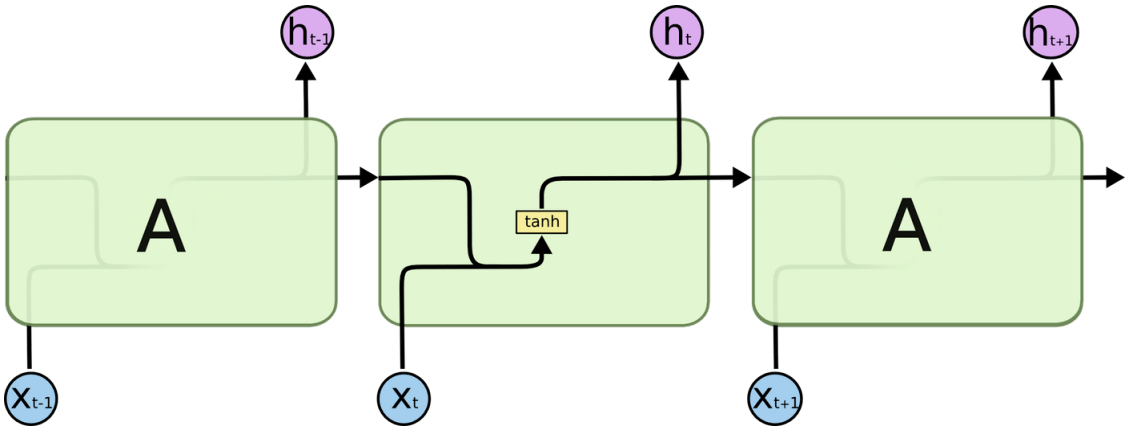

図 3.4　標準 RNN の繰り返しモジュール

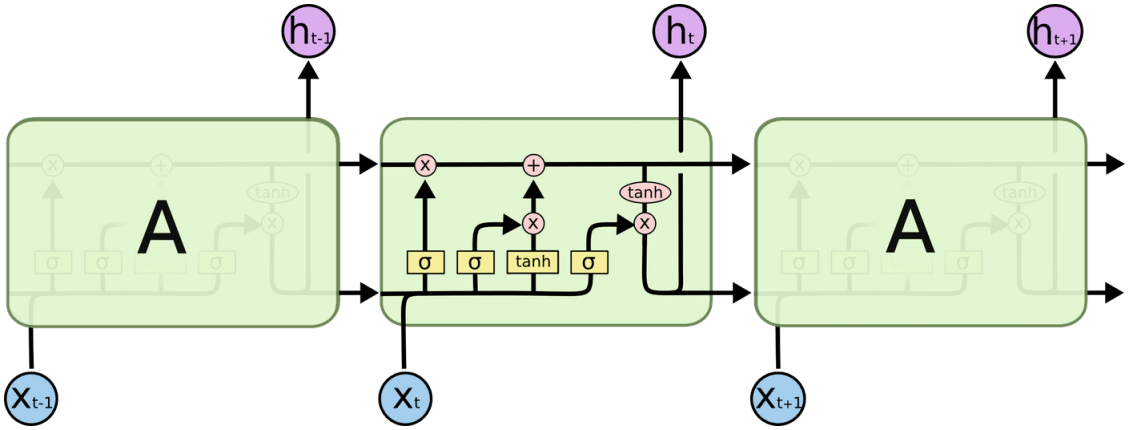

図 3.5　LSTM の繰り返しモジュール

Long Short Term Memory ネットワークは，通常は「LSTM」と呼ばれ，長期的な依存関係を学習することのできる，RNN の特別な一種である．これらは Hochreiter ら (Hochreiter & Schmidhuber 1997) により導入され，後続の研究で多くの人々によって洗練され，広められた．それは多種多様な問題に非常にうまく機能し，現在では広く使用される．

すべてのリカレントニューラルネットワークは，ニューラルネットワークのモジュールを繰り返す，鎖状をしている．標準の RNN では，この繰り返しモジュールは，図 3.4 のように，単一の tanh 層などの非常に単純な構造を持つ．LSTM もまたこの鎖のような構造を持つが，繰り返しモジュールは異なる構造を持つ．単一のニューラルネットワーク層ではなく，非常に特別な方法で相互作用する 4 つの層を持つ．

図 3.6 に示すように，図図 3.5 の各線は 1 つのノードの出力から別のノードの入力へ，ベクトル全体を伝える．ピンクの円は，ベクトルの加算のような，要素ごとの演算を表し，黄色のボックスは学習されるニューラルネットワークの層を表す．合流している線は連結を意味し，分岐している線は内容がコピーされ，そのコピーが別の場所に行くことを

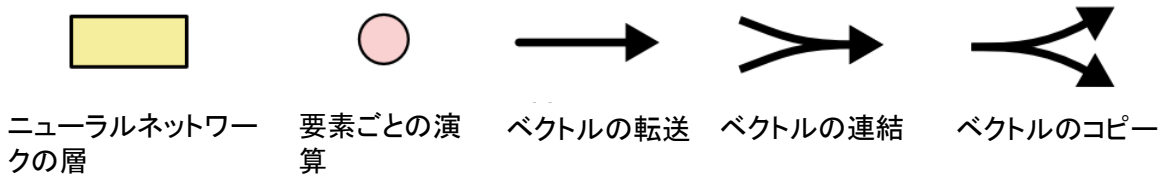

図 3.6　図中の記号

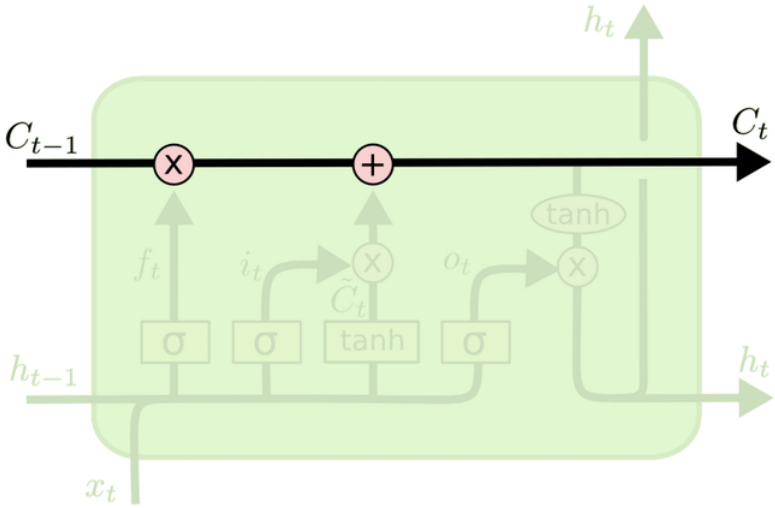

図 3.7　LSTM のセル状態

意味する．

LSTM の鍵は，セル状態（図 3.7 の上部を通る水平線）である．セル状態は一種のコンベア・ベルトのようなものである．それはいくつかのちょっとした線形相互作用のみを伴い，鎖全体をまっすぐに走る．情報を変化させずにセル状態（水平線）に沿って流すのは容易である．

LSTM は，セル状態に対し情報を削除したり追加する機能を持っている．この操作はゲートと呼ばれる構造により注意深く制御される．ゲートは選択的に情報を通す通路である．これはシグモイド・ニューラルネット層と要素ごとの乗算演算子により構成される．シグモイド層は 0 から 1 までの数値を出力する．この数値は各成分をどの程度通すべきかを表す．0 は「何も通さない」ことを，1 は「全てを通す」ころを意味する．LSTM は，セル状態を保護し，制御するために，このようなゲートを 3 つ持つ．

LSTM の最初のステップは，セル状態から捨てる情報を決定することである．この判定は「忘却ゲート層」と呼ばれるシグモイド層によって行われる．図 3.9 に示すように，忘却ゲート層は，h_{t-1} と x_t を見て，セル状態 C_{t-1} の中の各数値のために 0 と 1 の間の数値を出力する．1 は「完全に維持する」ことを表し，0 は「完全に取り除く」ことを表す．式 3.1 は図 3.9 の過程を表す．

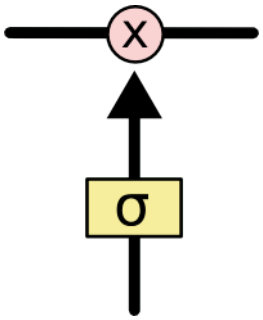

図 3.8　LSTM のゲート

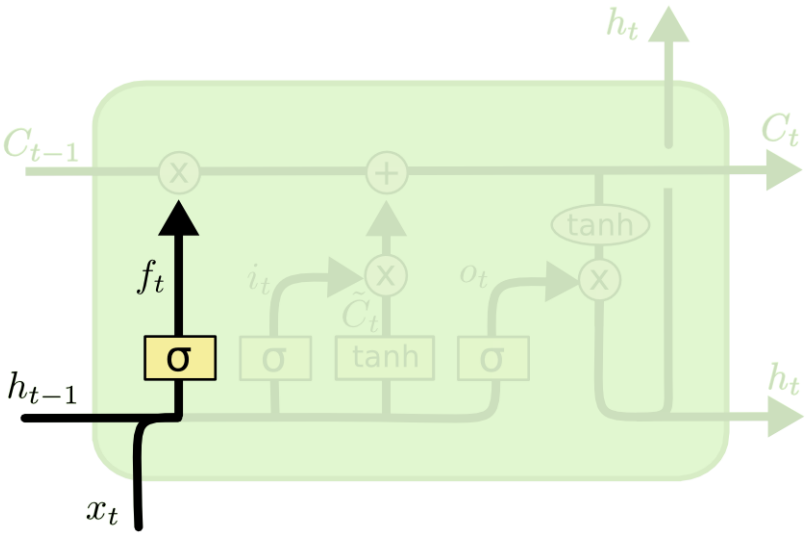

図 3.9　LSTM の最初のステップ (忘却する情報の決定)

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3.1}$$

次のステップは，セル状態に保存する新たな情報を決定することである．これは 2 つの部分から成る．まず，「入力ゲート層」と呼ばれるシグモイド層 (図 3.10) は，どの値を更新するかを判定する．次に，tanh 層は，セル状態に加えられる新たな候補値のベクトル $\tilde{C}_t$ を作成する．次のステップでは，状態を更新するために，これら 2 つを組み合わせる．式 3.2 は図 3.10 の過程を表す．

$$\begin{aligned} i_t &= \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \\ \tilde{C}_t &= \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \end{aligned} \tag{3.2}$$

そして，古いセル状態 C_{t-1} から新しいセル状態 C_t に更新する．何をするべきかについては前のステップですでに決定した．古い状態に f_t を掛け，さきほど忘れると決定され

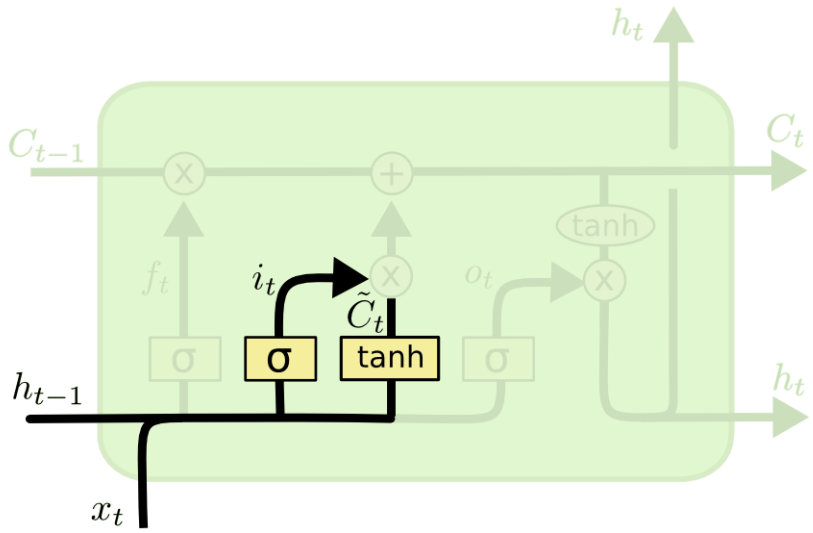

図 3.10　LSTM の第二のステップ (新たに記憶する情報の決定)

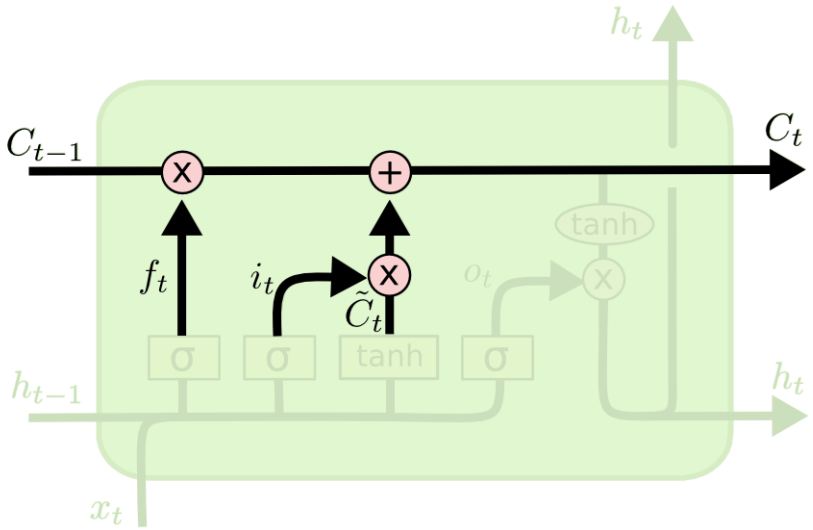

図 3.11　LSTM の第三のステップ (セル状態の更新)

たものを忘れる．そして，その積に $i_t * \tilde{C}_t$ を加える (図 3.11)．これは，各状態値を更新すると決定した割合で増減された，新たな候補値である．式 3.3 は図 3.11 の過程を表す．

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3.3}$$

最後に，出力するものを決定する．この出力は，セル状態に対してフィルタリングを施したになる．まず，図 3.12 のシグモイド層を実行する．この層は，セル状態のどの部分を出力するかを判定する．その後，決定された部分のみ出力するため，セル状態に（値を –1 と 1 の間に圧縮するために）tanh を適用し，それにシグモイド・ゲートの出力を掛け

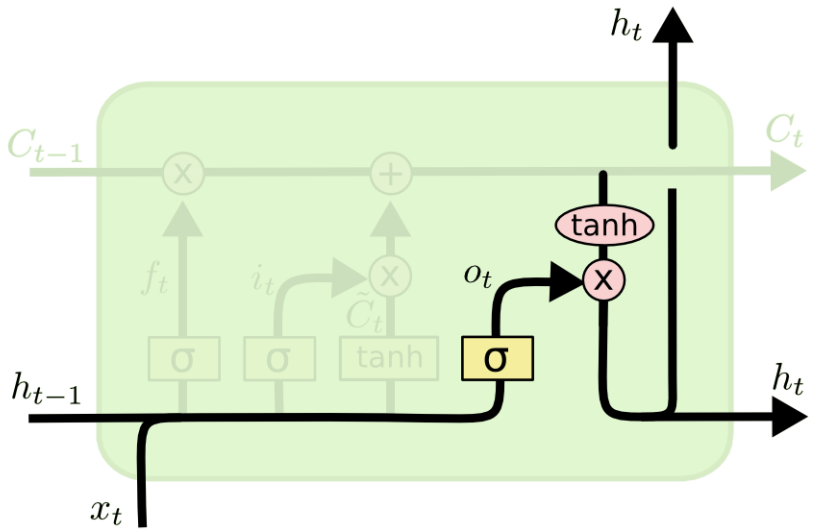

図 3.12　LSTM の最後のステップ (出力の決定)

る．式 3.4 は図 3.12 の過程を表す．

$$
\begin{aligned}
o_t &= \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \\
h_t &= o_t * \tanh\left(C_t\right)
\end{aligned}
\tag{3.4}
$$

3.1.3　Sequence-to-Sequence モデル

Sequence-to-Sequence モデルは Sutskever ら (Sutskever et al. 2014) によって提案され，「Seq2Seq モデル」「Encoder-Decoder モデル」「系列変換モデル」といった名前でも呼ばれる．

特徴は系列を入力として系列を出力する機構である．文章を単語の系列として捉えれば，Sequence-to-Sequence モデルを使うことで文章を入力として文章を出力するようなモデルを作れることになる．例えば英独機械翻訳で使用されている Sequence-to-Sequence モデルは，英語の単語の系列を受け取りその翻訳に対応するドイツ語の単語の系列を出力する．

Sequence-to-Sequence モデルは Encoder と Decoder の 2 つの RNN セルで構成されている．Encoder の RNN で入力系列をベクトルに圧縮し，そのベクトルを Decoder に渡し出力系列を生成する．

図 3.13 に示すように，翻訳元の単語列 A，B，C を多層からなる RNN で読み取っていき，隠れ状態ベクトルを求める．End of Sentence (EOS) が来たら翻訳先の単語列を多層からなる RNN を出力するような学習をさせる．この時，Encoder と Decoder として利用する RNN は別々に学習させる必要がある．

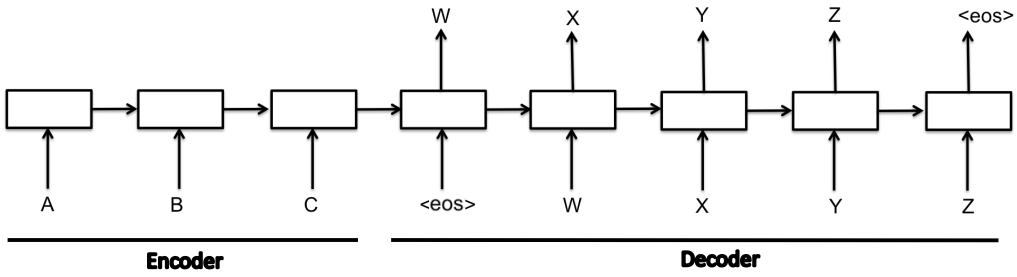

図 3.13　Sequence-to-Sequence モデルの構造

Dropout

Dropout は階層の深いニューラルネットを精度よく最適化するために Hinton らによって提案された手法である (Srivastava et al. 2014)．ニューラルネットワークを学習する際に，ある更新で層の中のノードのうちのいくつかを無効にして（そもそも存在しないかのように扱って）学習を行い，次の更新では別のノードを無効にして学習を行うことを繰り返す。これにより学習時にネットワークの自由度を強制的に小さくして汎化性能を上げ，過学習を避けることができる．

現在，Dropout はニューラルネットワークのモデリングの中に一般的な手段である。

Beam Search

学習済みのモデルを用いて翻訳を実際に行う場合，ある入力文 X から最適な文 Y を出力する処理は次のように記述できる．

$$\hat{Y} = \underset{Y}{\operatorname{argmax}} P(Y|X)$$

ここで Y の系列長は無制限であるため，最適な文章の候補は無限に存在する．そこで，近い解を見つけるために，貪欲法と Beam Search がよく用いられる（貪欲法は Beam Search の特殊な場合）(Medress et al. 1976).

貪欲法はある時刻に得られた確率分布の中で確率最大の単語をその時刻の出力単語として確定させていく手法である．時刻 t で選択される単語は時刻 t の確率分布 p_t の中で確率最大のものを選ぶので，時刻 t においては最適な選択をする．貪欲法が各時刻において最適なもの，つまり評価値上位 1 個を選択するのに対し，Beam Search は上位 n 個を選択する手法である．

図 3.14 では確率の累積値を評価値として単語を選択する[*1]．最終評価値が，Beam

[*1] http://renom.jp/ja/notebooks/tutorial/time_series/jp-en_nmt_seq2seq/notebook.html

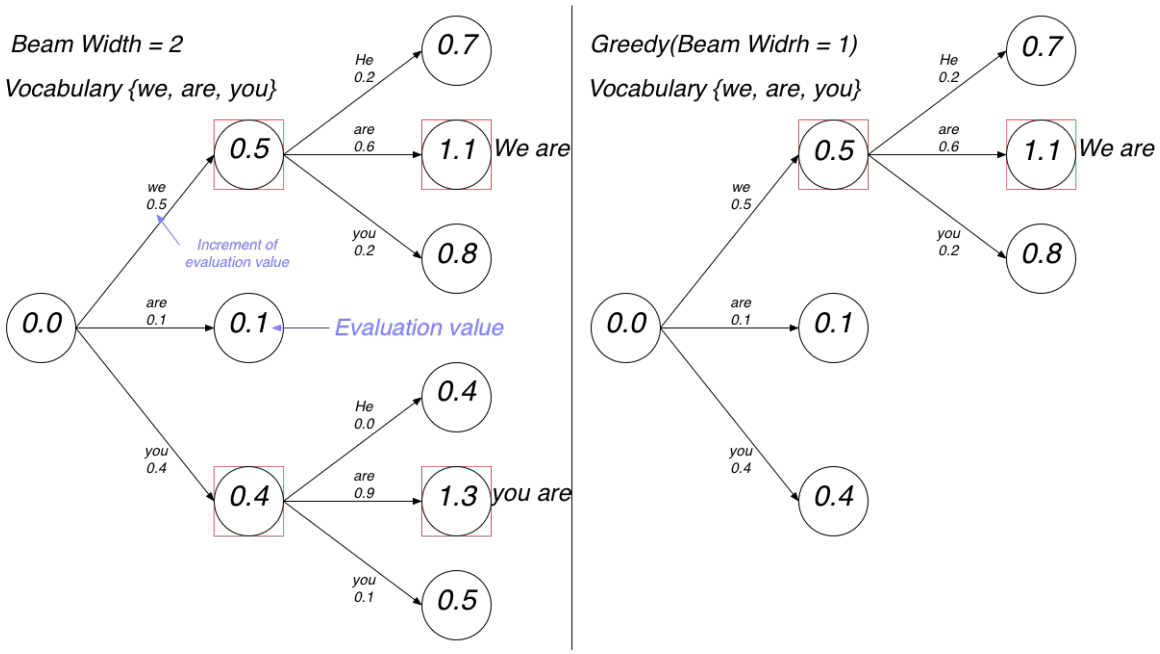

図 3.14　Beam Search と貪欲法 (Greedy)，ReNom 社の Tutorial より転載

Search の場合は 1.3，貪欲法の場合は 1.1 であること．貪欲法の場合は最初に評価値が最大となる ‘we’ を選択し，それ以外は無視している．Beam Search の場合は ‘we’ と ‘you’ の上位 2 つを選択する．この時点では ‘we’ を選んだ方が評価値が 0.5 となり最適なのであるが，‘you’ の次に ‘are’ を選択することで評価値が 1.3 となる．結果として最初に ‘we’ を選択するよりも評価値の大きい文を生成できたことになる．

3.1.4　Attention メカニズム付きエンコーダ・デコーダモデル

Sequence-to-Sequence モデルでは入力系列の情報を Encoder で圧縮したベクトルとしてしか Decoder に伝えることができないため，入力系列が長いと入力系列の情報を Decoder にしっかりと伝えることが難しくなる．そこで Decode 時に入力系列の情報を直接参照できるようにする仕組みが Attention メカニズムである．

ここでは Luong ら (Luong et al. 2015) によるグローバル Attention メカニズム付きエンコーダ・デコーダモデルを紹介する．本研究ではこれを文字レベルで使用した．

エンコーダは，双方向 LSTM リカレントニューラルネットワークであり，入力系列 $\boldsymbol{x} = (x_1, \cdots, x_m)$ を読み取って，順方向の隠れ状態列 $(\overrightarrow{h_1}, \cdots, \overrightarrow{h_m})$ と逆方向の隠れ状態列 $(\overleftarrow{h_1}, \cdots, \overleftarrow{h_m})$ を求める．隠れ状態 $\overrightarrow{h_j}$ と $\overleftarrow{h_j}$ は連結され，アノテーションベクトルが作られる．

デコーダは，目的言語文 $\boldsymbol{y} = (y_1, \cdots, y_n)$ を予測する LSTM リカレントニューラルネットワークである．各単語（文字レベルの場合，各文字）y_i は，リカレント隠れ状態 s_i

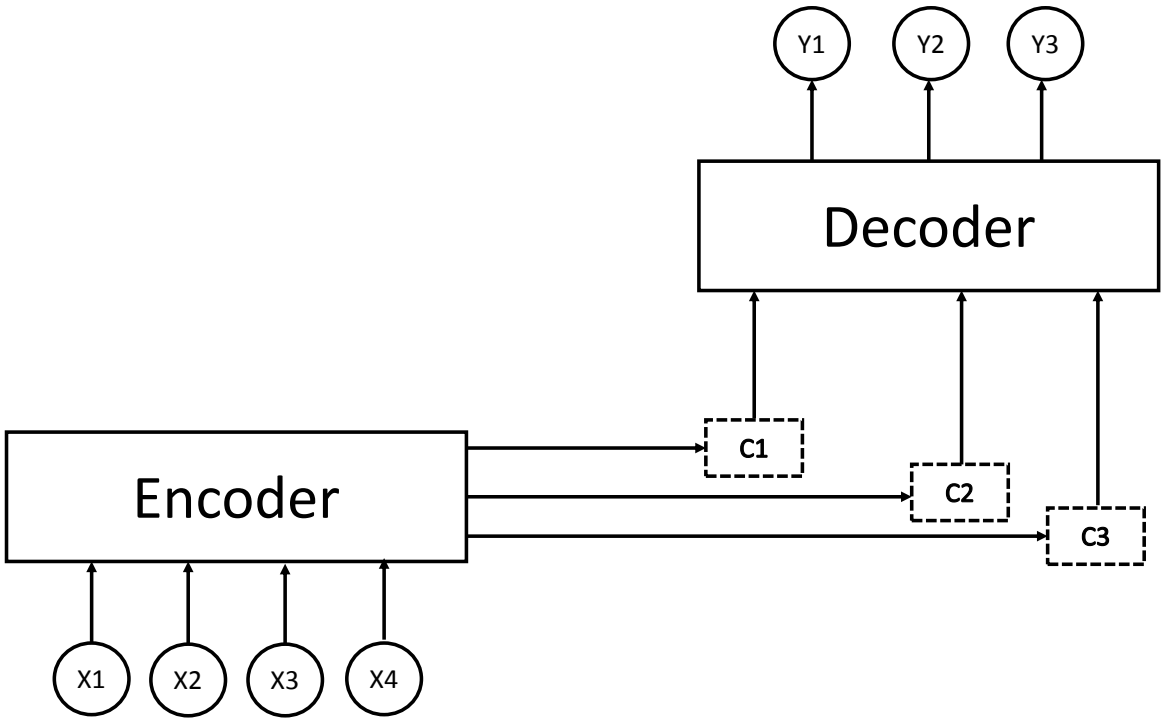

図 3.15　Attention メカニズムの構造

と，前回予測された単語（または文字）y_{i-1}，文脈ベクトル c_i を元に予測される．文脈ベクトル c_i は，アノテーション h_j の加重和として計算される．各 h_j の重みは，y_i と x_j のアラインメントについての情報を表すモデル α_{ij} を通じて決められる．

エンコーダの順方向状態は以下のように表される．

$$\vec{h_j} = \tanh(\vec{W}Ex_j + \vec{U}\vec{h}_{j-1}) \tag{3.5}$$

ここで，$E \in \mathbb{R}^{p \times V_x}$ は単語埋め込み行列であり，$W \in \mathbb{R}^{q \times p}$ と $U \in \mathbb{R}^{q \times q}$ は重み行列である．p，q，V_x はそれぞれ，単語ベクトルのサイズ，隠れユニットの数，原言語の語彙サイズである．

図 3.15 に Attention メカニズムの概略を簡単に示す．図 3.16 は日中対訳文における Attention メカニズムの中の文脈ベクトル c_i を可視化したものである．文字間の対応が確認できる．

3.2　ニューラル機械翻訳の研究動向

ここではニューラル機械翻訳モデルに関する研究を紹介する．

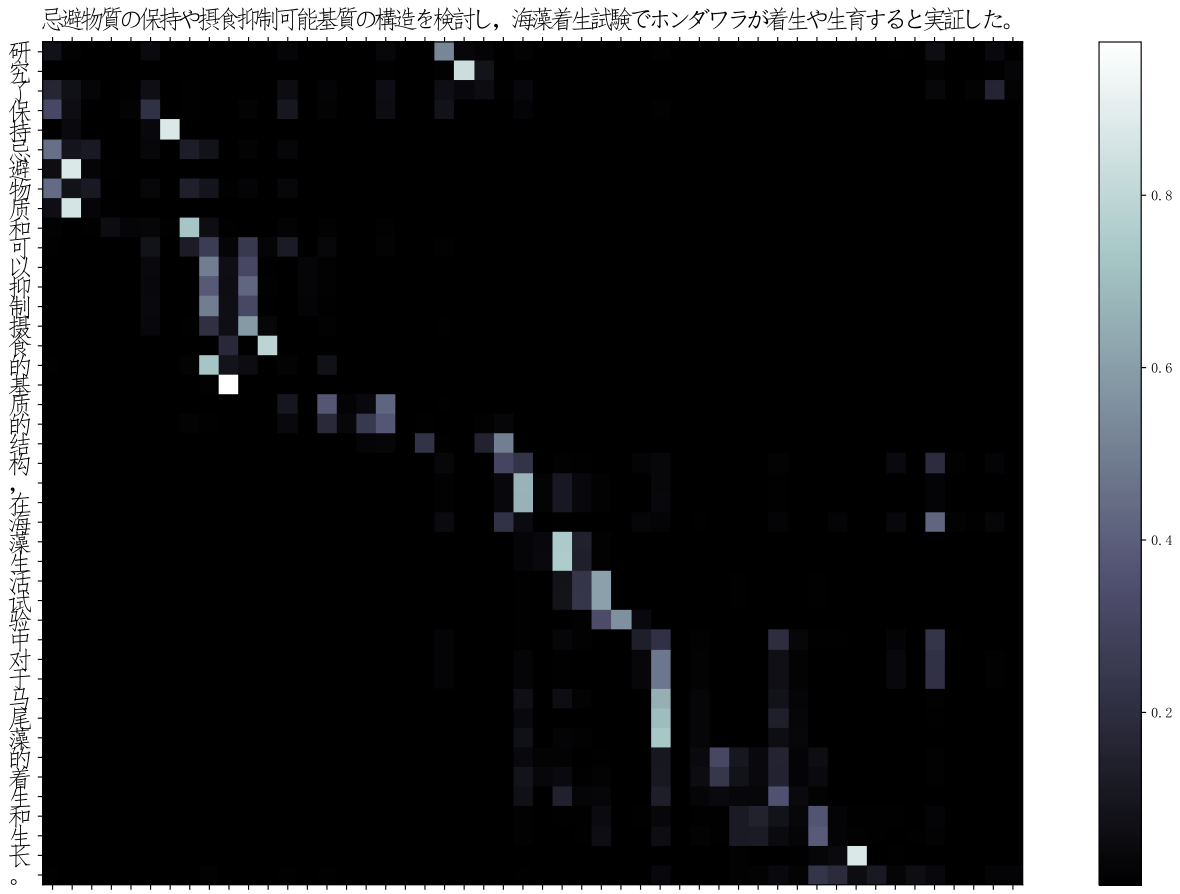

図 3.16　文脈ベクトル c_i の可視化

3.2.1　Attention メカニズムに関する研究

Attention メカニズムは，原言語と目的言語の言語要素間の関係性，注意箇所を学習する機構であり，翻訳精度が大幅に向上することから，現在ではこれを使用した NMT が主流となっている．

Attention メカニズム付きのニューラル機械翻訳システムは，原言語の文を固定ベクトルではなくベクトル系列にエンコードをする．目的言語の単語を生成するとき，生成された単語に関連する原言語の単語情報を利用できる．

Attention メカニズムは教師なしモデルである．異なるアテンション間に明示的な制約はない．また，重みを割り当てる場合，原言語の文のすべての単語の重みを計算する必要がある．これは非常に多くの計算リソースを必要とする．より完璧な Attention メカニズムの設計は，現在の研究のホットなトピックになっており，重要な成果が得られている．

Attention メカニズムの計算量削減

Attention メカニズムは大量の計算を必要とする．計算量を削減するために，Xu ら (Xu et al. 2015) は，画像記述生成タスクに対する Attention をソフトアテンション (Soft Attention) とハードアテンション (Hard Attention) に分割した．後者は元の画像領域の一部のみに Attention を注目し，計算量を減らすことができる．

上記の考えに基づいて，Luong ら (Luong et al. 2015) は Local Attention モデルを提案した．これは，従来の Global Attention の改善であり，計算量を削減できる．Global Attention は文脈ベクトル c_i を計算する際，原言語のすべてのコーディング系列を考慮する．これは，計算量が多い Bahdanau ら (Bahdanau et al. 2014) によって提案された Attention メカニズムと同様である．

Local Attention は，原言語エンコーディングの小さな文脈窓 (window) のみに焦点を合わせるため，計算量を大幅に削減できる．この方法は，原言語から生成された単語に関連するアライメント位置を見つける．文脈ベクトル c_i を計算するとき，アライメントポイントを中心に固定サイズで計算する．Local Attention は，文脈ベクトルを生成する際に原言語のごく一部に焦点を合わせ，文全体に次々注目し，長い文の翻訳に適する．WMT2014 の英語からドイツ語への翻訳タスクでは，Local Attention が Global Attention と比較して 0.9 BLEU 増加した．長い文の翻訳実験では，文の長さが増加しても，Local Attention では BLEU 値が減少しなかった．

教師あり Attention メカニズム

Attention メカニズムは，目的言語の単語に対応する原言語の単語を予測するときに単語自身の情報を利用しない教師なし学習モデルである．そのため，単語アラインメントの品質は低い．

この問題は統計機械翻訳では十分に対処されており，単語アライメントの品質は非常に高くなっている．Chen ら (Chen et al. 2016) は上記の考えに基づいて，Attention メカニズムをガイドするための事前知識として単語アライメント情報を使用する方法を提案した．

基本的な考え方は以下のとおりである．最初に Och ら (Och & Ney 2003) が提案した GIZA++ というアライメントツールを使用し訓練コーパスの単語アライメント情報を取得する．次に単語ラインメント情報を先験的知識として，Attention メカニズムの単語アラインメントが可能な限り整列する．結果として，製品タイトル翻訳タスクで NMT システムの BLEU スコアが 18.6 から 21.3 に改善された．

Attention メカニズムに関する分析

Vaswani ら (Vaswani et al. 2017) は，リカレントニューラルネットワークと畳み込みニューラルネットワークを放棄し，Attention メカニズムのみを使用して Sequence-to-Sequence モデルを実装した．このモデルは強力な並列処理を備えており，翻訳の品質も向上する．

Raganato ら (Raganato & Tiedemann 2018) は，Transformer (Vaswani et al. 2017) の Attention がどこに向いているのかを分析した．結果として，低次の層では表面的な文法に，高次の層では文の持つ意味に対して Attention が向く傾向にあることがわかった．

Domhan ら (Domhan 2018) は，NMT モデルを構成要素に分解，それらを組み合わせて NMT モデルを自体表現する Architecture Definition Language (ADL) を導入した．そしてこの ADL を用いて，各構成要素がどんな働きをするのかを様々な NMT タスクで分析した．

過剰翻訳と不十分な翻訳

過剰翻訳とは，一部の単語またはフレーズが繰り返し翻訳されることを意味し，不十分な翻訳とは，一部の単語またはフレーズが完全に翻訳されないことを意味する．この問題は，Attention メカニズム付きニューラル機械翻訳を含む，ニューラル機械翻訳で広く見られる．

ニューラル機械翻訳には，翻訳済みの単語情報や未翻訳の単語情報などの履歴情報を記憶するための優れたメカニズムがなかったが，Tu らは Coverage 機構を提案した (Tu et al. 2016)．これは，統計機械翻訳の Coverage メカニズムを Attention メカニズム付きニューラル機械翻訳に導入したものである．Coverage ベクトルは，翻訳プロセスの過去のアテンション情報を記録するように設計されている．これにより，Attention メカニズムは未翻訳の単語により注意を向け，既に翻訳済みの単語の重みを減らすことができる．Coverage メカニズムは，翻訳の整合性を確保するための統計機械翻訳の一般的な方法である．

ニューラル機械翻訳では，Coverage メカニズムを直接モデル化することは非常に困難である．Tu らは，Attention の履歴を保持するための Coverage ベクトルを導入し，Attention モデルによる以降の Attention の調整を補助をする．これにより NMT システムは，原言語文中の未翻訳の単語をより重視するようになり，過剰翻訳が抑制される．この方法は，過剰翻訳の問題を軽減することができ，効果は明らかである．

別の解決策は，翻訳結果に対する原言語情報と目的言語情報の割合を制御することである．この考え方は直感的であり，翻訳中に原言語コンテキストと目的言語コンテキストがそれぞれ翻訳の忠誠度と流暢さに影響する．したがって，実際の単語を生成するときは，原言語のコンテキストに注意を向け必要があり，単語を生成するときは，目的言語のコン

テキストに依存する必要がある．これには，2 種類の情報が翻訳結果に与える影響を制御する動的な手段が必要である．これは，ニューラル機械翻訳には欠けている．この点での対策は，Tu らによって提案されたコンテキストゲート方式であり，これは翻訳の忠実度を保証しながら翻訳の流暢さも保証する (Tu et al. 2017)．Coverage メカニズムとコンテキストゲートを組み合わせて，相互に補完することができる．Coverage メカニズムは，翻訳の十分性に焦点を合わせて，より優れた原言語コンテキストベクトルを生成できる．コンテキストゲートは，2 種類の情報の影響を動的に制御し，原言語と目的言語コンテキストの重要性に従って目的言語単語を生成できる．

外部メモリの使用

Wang らは外部メモリの使用によって，Sequence-to-Sequence モデルのデコーダを改良した (Wang et al. 2016)．メモリはエンコーダの隠れ層のどこに注目するか決定し，デコーダの隠れ層でメモリを更新する，Attention を改良する．この方法は，メモリ内の後続のメモリで使用できる中間状態情報を選択的に保存する．これにより，Attention メカニズムの不十分さをある程度補償し，ニューラル機械翻訳モデルの表現能力をより適切に拡張し，長距離依存効果を強化できた．

3.2.2　文字レベルのニューラル機械翻訳に関する研究

文字レベルの NMT は，登録されていない単語，単語の分割などの問題を解決するために提案されたニューラル機械翻訳モデルで，主な特徴は入力および出力の単位を単語から文字に小さくすることである．

単語コーディング

ほとんどのニューラル機械翻訳は，単語を翻訳の基本単位として使用する．中国語や日本語などの言語では，未登録の単語，スパースデータ，単語分割の問題がある．さらに，英語やフランス語などの形態の変化の多い言語では，単語を基本単位として使用すると，単語間の形態の変化と意味情報が失われる．たとえば，英語の単語「run」,「runs」,「ran」,「running」は，同じ接頭辞「run」を持つことを無視して，4 つの異なる単語と見なされる．上記の問題を解決するために，さまざまな単語コーディング方式が提案されており，それは入力単位に応じて次の 2 つのタイプに分類できる．

1. 文字エンコード方式: 英語やフランス語などの表音文字の場合，文字は単語の基本単位であり，文字でモデル化できる (Kim et al. 2016)．この方式には，単位サイズが小さすぎて，英語やフランス語などの語彙サイズが同じである言語間の翻訳ににしか適していないなどの欠点もある．

2.　サブワードコーディング方式: サブワードコーディング方式によって選択される翻訳の基本単位は文字と単語の中間であり，2 つの単位の共通の利点が得られる．形態素の単位も文字と単語の中間にあるが，欠点は特定の言語に依存しているため，適用の汎用性が制限されることである．したがって，サブワードは通常，BPE（Byte pair encoding）によって取得される (Sennrich et al. 2016b)．例えば単語「dreamworks interactive」は，「dre + am + wo + rks / in + te + ra + cti + ve」という系列に分割できる．BPE はシンプルで効果的で適応性がある．
Kudo が (Kudo 2018)，サブワード分割の曖昧性を使い，ニューラル機械翻訳モデルの正則化をかける「サブワード正則化」を提案した．原言語，目的言語両方に対する分割パターンをノイズとして扱い機械翻訳モデルへ正則化をかける．また，ニューラル機械翻訳にかぎらず適用することを可能にした．従来のサブワード法よりも良い結果を示した．

文字レベルのニューラル機械翻訳

文字レベルのニューラル機械翻訳では，入力と出力の両方が文字に基づいている．エンコーダとデコーダに文字から単語へのマッピングメカニズムを追加することにより，文字列の入出力を実現する．

Ling ら (Ling et al. 2015) は，エンコーダに文字から単語へのマッピングを追加して，文字レベルの入力を実装し，デコード時に目的言語文字列を生成し，原言語単語列に注目する Attention メカニズムを追加した．このメソッドは，文の開始と終了，それぞれ「SOS」(Start of Sentence) と「EOS」(End of Sentence) を人為的に追加し，「SOW」(Start of Word) と「EOW」(End of Word) もそれぞれ追加して，単語と文の開始と終了を含む文字レベルの出力を実現する．「EOS」を生成すると，完全な文を生成し，「EOW」を生成すると，完全な単語を生成することを意味する．このようにして，文字レベルの入出力が実現される．

Lee ら (Lee et al. 2016) は，文字ベクトル（Character Embeddings）系列を畳み込みニューラルネットワークに入力し，出力を固定長の系列に分割する．最大プーリング（Max-pooling）操作を各固定長の系列に適用し，セグメンテーションコーディングを取得する．セグメンテーションコーディングは，セマンティックユニットとして使用され，エンコーダに入力される．デコーダでは，Attention メカニズムが原言語に焦点を合わせて，コーディング系列をセグメント化し，目的言語文字の系列を生成する．

2 つの方法の主な違いは，原言語の意味の基本単位にあり，2 番目の方法は長さは固定である．1 番目の方法では，意味単位は単語である．このタイプの方法の主な特徴は，原言語で文字から単語へのマッピングを実装するためにニューラルネットワークを使用する

ことである．したがって，文字レベルの入力を実現し，未登録の単語の問題を解決する．目的言語では，単語と文の境界が分割マーカーによって判断される．

ドキュメントレベルのニューラル機械翻訳

従来の機械翻訳は，文を単位に処理を行う．統計機械翻訳以降の機械学習に基づく機械翻訳では，文を単位とした対訳コーパスを学習データとし，原言語の一文を入力として目的言語の一文を出力する翻訳モデルを学習する．

Wang ら (Wang et al. 2017) は階層的な RNN を使用し，ドキュメントレベルで過去 3 文を 1 つの文脈ベクトルとしてマッピングし，それを文レベルのエンコーダの初期値とする NMT システムを提案した．従来の NMT と比較して，BLEU スコアが改善した．Wang らの研究では，文脈情報を NMT に利用することで曖昧な単語の訳し分けをすることに成功していた．

Maruf ら (Maruf & Haffari 2018) は文レベル NMT とメモリネットワークを組合せたドキュメントレベル NMT を提案した．また，ドキュメントレベルで翻訳を行なうため，より広範な文脈情報を利用することが可能である．原言語と目的言語の文脈情報の両方を利用した．そのため，Maruf らの研究では文脈情報を NMT が利用し，語彙の曖昧性の訳し分け，名詞と代名詞の照応関係などの点でベースラインの NMT から改善した．

3.2.3　多言語のニューラル機械翻訳に関する研究

ある言語から別の言語への 1 対 1 の翻訳とは異なる多言語機械翻訳は，複数の言語間で翻訳できる．ニューラルネットワークに基づく多言語機械翻訳は，系列から系列への学習とマルチタスクの学習から派生し，単言語から多言語への翻訳と多言語から単言語への翻訳，多言語から多言語への翻訳に分けることができる．

単言語から多言語への翻訳

単言語から多言語への翻訳では，原言語は 1 つしかなく，目的言語は複数の機械翻訳方法である．Dong ら (Dong et al. 2015) は，マルチタスク学習を系列間学習に初めて導入し，単言語から多言語へのニューラル機械翻訳を実現した．マルチタスク学習とは，マルチタスク学習モデルをエンコーダとデコーダに追加し，原言語はエンコーダを使用し，各目的言語は単一のデコーダを使用する．各デコーダは独自の Attention メカニズムを持つが，同じエンコーダを共有する．

実験では EuroParl Corpus を使用し，原言語は英語，目的言語はフランス語，スペイン語，オランダ語，ポルトガル語である．実験結果は，単言語から多言語への機械翻訳効果が英語と他の言語間の個々の BLEU 値が増加することを示している．この方法は，原言

語エンコーダを共有し，リソースの少ない言語間の翻訳の品質を向上させることができる．欠点は，各デコーダに個別の Attention メカニズムがあり，計算が複雑で，大規模な言語ペアに応用することが制限される点である．

多言語から単言語への翻訳

多言語から単言語への翻訳は，複数の原言語と 1 つの目的言語のみを使用した機械翻訳方法である．典型的な研究は，Zoph ら (Zoph & Knight 2016) によって提案された多言語から単言語への翻訳方法である．この方法には，エンコーダごとに 1 つの原言語が対応あり，マルチ原言語の Attention メカニズムが採用されている．これは Luong ら (Luong et al. 2015) によって提案された Local Attention の改善である．文脈ベクトルは，2 つの原言語からの合計を取得され，同時にデコードに適用される．

この実験では，WMT 2014 コーパスを使用する．原言語がフランス語とドイツ語で，目的言語が英語の場合は 1 対 1 の翻訳に比べて 4.8 BLEU 値が向上する．原言語が英語，フランス語，目的言語がドイツ語の場合，1.1 BLEU 値が増える．この結果により，原言語の違いが手法に大きな影響を与えていることがわかる．さらに，原言語ごとに個別の Attention メカニズムが使用されるが，計算が複雑になる．

多言語から多言語への翻訳

多言語から多言語への翻訳は，複数の原言語と目的言語を使用した機械翻訳方法であり，複数の言語間で翻訳できる．Firat ら (Firat et al. 2016) は，WMT 2015 にある英語からフランス語，チェコ語，ドイツ語，ロシア語，フィンランド語の翻訳，合計 10 言語ペアを使用した多言語ニューラル機械翻訳方法を提案した．多言語から多言語への翻訳では，1 対 1 の翻訳と比較して翻訳が大幅に改善されることはない．この方法は，原言語と目的言語ごとに別々のエンコーダとデコーダを適用するが，Attention メカニズムを共有する．これにより，計算の複雑さが軽減され，モデルのパラメータが言語の数に比例して増加する．

多言語機械翻訳はより多いパラメータを使用しているが，Google は既存のニューラル機械翻訳モデルを変更することなく，多言語から多言語への翻訳方法を提案している (Johnson et al. 2017)．この手法は，パラメータを追加せずに GNMT(Johnson et al. 2017) というシステムに実装される．対訳コーパスの原言語にマーカーを追加して，翻訳する目的言語を示し，処理済みの多言語対訳コーパスを組み合わせて訓練する．この方法は，単言語から多言語への翻訳，多言語から単言語への翻訳，および多言語から多言語への翻訳に効果があり，コーパス内の直接対訳対応言語なしで翻訳を達成できる．この方法は，既存のニューラル機械翻訳モデルを変更せず，実装は簡単で効果的であり，大規模な実用的なアプリケーションに便利である．

3.2.4　制限された語彙サイズの問題に関する研究

訓練時間が膨大になるのを防ぐため，ニューラル機械翻訳では語彙サイズと文の長さを一定の範囲に制限する．たとえば，辞書はコーパス内のより高い頻度の単語で構成され，その数は通常 30,000〜80,000 に制限される．その他の低頻度語は〈unk〉などの特殊記号によって，文の長さは 50 単語に制限される (Jean et al. 2015b)．この制限は未登録語（Out-of-vocabulary, OOV）の問題を悪化させ，低頻度の語の学習を困難になる．

未登録単語の問題

未登録語の問題は，コーパス内の一部の単語が辞書の範囲を超えているため，単語が正確に翻訳されないことである．辞書のサイズが制限されている場合，登録されていない単語の数が増えるとニューラル機械の翻訳品質が大幅に低下する (Cho et al. 2014b)．実際には，言語は動的に変化するものであり，語彙サイズを修正するのは困難である．人，場所，施設の名前などの典型的な固有表現，および新しい単語やホットワードが常に作成されている．したがって，未登録語の問題はニューラル機械翻訳の基本的な研究テーマであり，解決策は大まかに次の 3 つのカテゴリに分類される．

1. 未登録の単語を処理する間接的な方法．ニューラルネットワーク構造を最適化し，大規模な翻訳辞書またはオープン辞書を実装して未登録語の問題を解決する．もう 1 つは，未登録の単語の問題を回避するために，翻訳の基本単位として文字やサブワードを使用するなど，原言語と目的言語の翻訳単位を小さくすることである．どちらの方法も未登録の単語をある程度扱うことができるが，前者は形態的な変化のある言語には効果的な方法ではないという欠点がある．
2. 文脈情報による未登録語の予測方法．この方法の基本的な考え方は，目的言語の未登録語に対応する原言語がわかっている場合，原言語に対応する語を検索辞書によって目的言語の翻訳語に変換するか，文脈に従って未登録語を予測することができるというものである．既存の研究のほとんどは，この考えに基づいている．置換方法は最も基本的な処理方法であり，未登録の単語を生成する場合，その単語に対応する原言語の単語が Attention メカニズムによって検出され，一致する可能性が最も高い原言語の単語が目的言語の単語としてコピーされる (Gulcehre et al. 2016)．統計機械翻訳の単語アラインメントモデルなど，他の単語アラインメント手法により，対応する原言語の単語の対応する翻訳を見つけ，未登録の単語を翻訳する (Jean et al. 2015a)．この方法はシンプルで直感的で，一定の効果があるが，言語の複雑な変更や 1 対多の特殊なケースは無視される．Luong ら (Luong et al. 2015) は，未登録語を処理するための未登録語ラベリング方法を提案した．この方

法では，未登録語をより正確に処理するために原言語と目的言語の相対位置情報を使用する．Li ら (Li et al. 2016) は，コーパス内の低頻度の単語を類似の単語に置き換える「置換-翻訳-復元」モデルを提案した．翻訳および復元では，低頻度の単語の置換後にコーパスを使用して翻訳モデルが取得される．低頻度の単語は翻訳中に置き換えられ，置き換えられた単語を翻訳される．3 つの方法はすべて，未登録語の問題をある程度まで処理できる．違いは，最初の 2 つの方法は訓練コーパス外の未登録語を処理できず，3 番目の方法はこの問題を処理できることである．

3. 文字またはサブワードを翻訳の基本単位として使用する方法．この方法は通常，前処理または後処理として使用され，ニューラル機械翻訳モデルには変更が加えられない．この研究には，主に Hirschmann ら (Hirschmann et al. 2016) が提案した複合語分割法と，Sennrich ら (Sennrich et al. 2016b) が提案したサブワード表現法がある．このタイプの方法では，低頻度の単語および一部の単語は，単語よりも小さい単位で翻訳できると見なされる．この方法は，前処理および後処理としてのみ使用され，ニューラル機械翻訳モデルを変更せず，未登録語の問題をより適切に処理できる．欠点は，入力系列と出力系列の長さが大幅に増加し，それに応じて計算量が増加することである．

大規模な翻訳辞書を実装する方法

大規模翻訳辞書とは，より大きな辞書（30,000～80,000 と比較）または無制限のサイズを指し，一般に目的言語辞書と呼ばれる．ニューラル機械翻訳モデルの訓練の難しさの 1 つは，目的言語の単語の確率を計算することである．大規模な辞書の応用では，既存のソリューションを大きく 3 つのカテゴリに分類できる．

1. 目的言語の単語の確率計算を最適化．Jean ら (Jean et al. 2015a) は，重要度サンプリングに基づく重要度計算方法を提案した．この方法では，モデルの更新ごとに辞書の一部のみが使用される．翻訳するとき，辞書の全部を使用するか，一部を使用するかを選択できる．大規模辞書を使用する場合，サイズは 500,000 であり，訓練の複雑さはそれほど増加しないが，訓練で使用される目的言語辞書が 30,000 であり，計算の複雑さが依然として高いという欠点がある．Mi ら (Mi et al. 2016) は，3000 サイズの文レベルの辞書を使用した．この方法は，各原言語文について，単語レベル，フレーズベースの統計機械翻訳モデルを介して各原言語文に対応する目的言語単語を取得し，2000 個の目的言語共通単語を追加して，文レベルの辞書を構築する．

   WMT 2015 の英語からフランス語への翻訳では，比較して BLEU 値が 1.0 増加している．この方法には，速度と翻訳品質の両方で大きな利点がある．

2. 単語レベルのモデルと文字レベルのモデルを組み合わせて，登録されていない単語の文字レベルのモデリング．Luong ら (Luong & Manning 2016) は，主に単語レベルのニューラル機械翻訳モデルを使用し，原言語の未登録単語に文字レベルの表現方法を採用し，目的言語の未登録単語に別の文字レベルの未登録単語処理モデルを使用するハイブリッドモデルを提案した．この方法には，高速な単語レベルの訓練という利点があり，文字レベルの系列が長くなりすぎるという欠点が回避される．オープン辞書のニューラル機械翻訳は，両方の利点を組み合わせることによって実現される．
3. 辞書の符号化．符号化方法を使用して，ニューラル機械翻訳が辞書サイズの制限あり条件下でより多くの原言語および目的言語の単語を処理できるようにする．この方法は，V がコーパス内のすべての単語を含むより大きな辞書である場合，W はより小さな辞書である．V と W の辞書リストのマッピングが競合や可逆性のない符号化で実装されている場合，既存の翻訳モデルを変更せずに大規模な翻訳辞書を実装できる．上記の考えに基づいて，Chitnis ら (Chitnis & DeNero 2015) は，ハフマン符号化に基づく方法を提案した．低頻度の単語は，ハフマン符号化によって 2 つの疑似単語系列にエンコードされ，合計辞書サイズは，共通の単語と疑似単語の数の合計である．この手法は，翻訳モデル自体を変更せず，追加のパラメーターも追加せず，変換の前後に前処理と後処理のみが必要である．

長い文への対応

ニューラル機械翻訳は，約 20 単語までの短い文で良好な結果を達成しており，翻訳の効果は文の長さが長くなるにつれて減少する (Cho et al. 2014a)．RNN の長期記憶の問題のため，長文の翻訳が不十分となるが主な理由である．この問題の処理は，次の 2 つのカテゴリに分類される．

1. 長い文の分割方法．長い文は，直接翻訳できる長さのセグメントに分割される．セグメントの翻訳結果が結合されて，最終的に完全な文の翻訳結果が得られる．Abadie ら (Pouget-Abadie et al. 2014) の研究は，類似した語順を持つ言語間ではうまく機能する．不利な点は，セグメント間の長距離順序付け能力がないことである．
2. 主に Attention メカニズムを強化し，外部メモリ (Wang et al. 2016) およびその他の情報を追加するために，長距離依存の能力を強化する．

3.2.5　事前知識の利用に関する研究

事前知識は，事前に準備された単言語，バイリンガル，注釈付きデータなどであり，ニューラル機械翻訳の訓練を導くことができる．ほとんどのニューラル機械翻訳モデルは，文レベルの単語情報のみに依存しており，構文やテキスト情報などの十分な言語構造の知識を学習することはできない．ニューラル機械翻訳に外部の事前知識を統合する方法は，次のカテゴリに分類される．

統計機械翻訳の利用

統計機械翻訳を使用してニューラル機械翻訳の翻訳精度を改善することは，事前知識を統合する方法の 1 つである．He ら (He et al. 2016) は，対数線形 NMT 法を提案した．これは，目的言語の単語を生成するときに，追加の翻訳テーブルと言語モデルを追加する．翻訳テーブルは，低頻度単語の翻訳を改善でき，言語モデルは翻訳結果の流暢さを改善できる．2 つモデルは個別に訓練され，対数線形モデルによって統合される．この方法は浅い統合方法であり，ニューラル機械翻訳の利点を十分に活用していない．対数線形モデルとは、分割表の各セルにおける期待値を対数変換し，それを各属性の主効果およびそれらの交互作用で説明するモデルである。Wang ら (Wang et al. 2016) によって提案した深い統合法もある．基本的な考え方は次のとおりである．目的言語の単語を生成するときに統計機械翻訳によって目的言語の候補単語リストを生成し，これを目的言語の単語生成品質を向上させるために使用する．候補単語リストとニューラル機械翻訳のデコーダは，ゲートメカニズムによって結合される．これら 2 つの部分は，両方の翻訳モデルを活用するために一緒に訓練できる．

上記の作業に加えて，Zhou ら (Zhou et al. 2017) は，ニューラル機械翻訳と統計機械翻訳の翻訳結果をフレームワークに入力する，ニューラルネットワークベースの統合フレームワークを提案した．デコードでは，さまざまなシステムの翻訳結果が複数の Attention メカニズムを介して取得され，ニューラル機械翻訳と統計機械翻訳の共通の利点がこの方法得られる．Stahlberg ら (Stahlberg et al. 2017) は，統計機械翻訳のベイジアンリスク情報をニューラル機械翻訳のデコードに融合し，複数の言語ペアの翻訳品質を大幅に改善した．統計機械翻訳の研究には数十年歴史があるので，その利点を最大限に活用してニューラル機械翻訳モデルの欠陥を補う方法は，さらに研究する価値がある．

言語知識の追加

言語知識は，統計機械翻訳やその他の自然言語処理タスクの効果を改善できる．たとえば，接辞処理により，同じ単語の異なる形式を 1 つの表現にできる．これは，データが疎

らな分布を減らすのに役立ち．さらに，品詞タグ付けおよび構文依存のタグ付け情報により，翻訳効果をある程度向上させることができる．

エンコーダとデコーダによる言語知識の追加は，言語知識を使用する 1 つの方法である．Sennrich ら (Sennrich & Haddow 2016) は，さまざまな特徴ベクトルを合わせるために使用される特徴情報の組み合わせとしてエンコーダを更新した．実験では見出語情報（Lemma），サブワードタグ情報（Subword Tags），形態学的情報，品詞タグ情報などを試した．より良い翻訳精度が得られた．さらに，Chen ら (Chen et al. 2017) は，原言語の依存関係情報をニューラル機械翻訳に追加した．Li ら (Li et al. 2017) および Bastings ら (Bastings et al. 2017) は，エンコーダで原言語の構文情報を合成する．Wu ら (Wu et al. 2017) はソースツリーの単語のグローバルな依存性を強化するために依存ツリー情報を使用し，Chen ら (Chen et al. 2017) はエンコーダとデコーダで原言語と目的言語の構文情報を同時に追加した．Zhang ら (Zhang & Matsumoto 2017) は日中ニューラル機械翻訳で日本語文字ごとに部首などの情報を追加した．このタイプの方法は，エンコーダとデコーダの構造を拡張し，より多くの言語機能を組み込み．原言語表現の品質を改善し，目的言語生成の品質も改善する．

さらに，Niehues ら (Niehues & Cho 2017) はマルチタスク学習法を使用して，品詞タグ付け機能と名前付き情報をニューラル機械翻訳に統合している．マルチタスク学習とは、複数の関連するタスク同士の情報を共有することにより、予測精度を上げることができるという手法である．Zhang ら (Zhang et al. 2017) は，事前知識（バイリンガル辞書，フレーズリストなど）をニューラル機械翻訳に統合する一般的な対数線形フレームワークを提案した．

構文木には豊富な言語構造情報が含まれており，ニューラル機械翻訳モデルをから系列から構文木ベースの形式に拡張する研究もある．Eriguchi ら (Eriguchi et al. 2016) は，構文木ベースのエンコーダを使用して原言語文のフレーズ構造情報をボトムアップで取得する，構文木から系列へのニューラル機械翻訳モデルを提案した．原言語には 2 つのエンコーダがあり，1 つは単語系列情報をエンコードし，もう 1 つは構文構造情報をエンコードし，Attention メカニズムを介して 2 種類のコーディング情報を融合し，デコード時に 2 つの情報の構造形式を同時に考慮することができる．Aharoni ら (Aharoni & Goldberg 2017) は，系列から構文木へのニューラル機械翻訳モデルを提案し，目的言語は線形化された構文木の形式で生成され，系列から系列へのモデルの利点を維持し，デコーダを強化できる．Wu ら (Wu et al. 2018) は，目的言語の単語系列と単語間の依存関係を同時にモデル化して，目的言語の品質を改善できる系列依存性ニューラル機械翻訳モデルを提案した．

単言語コーパスの知識を追加

単言語コーパスは，大量に存在し，入手が容易であるという利点を持つ非常に重要なリソースである．統計機械翻訳では，大規模な目的言語の単言語コーパスにより高品質の言語モデルを得ることができる．これは，翻訳の流暢さの向上において重要な役割を果たす．

ニューラル機械翻訳で使用できる単言語コーパスは，主に目的言語の単言語コーパスと原言語の単言語コーパスに分けられる．目的言語に対する単言語コーパスの 1 つ応用は言語モデルである．Glehre ら (Gülçehre et al. 2015) は，大規模な単言語コーパスを使用してニューラル機械翻訳の翻訳結果を改善する方法を提案した．言語モデルは，単言語コーパスによって訓練され，ニューラル機械翻訳に統合される．統合方法は，浅い統合と深い統合に分けられる．浅い統合方法は，言語モデルを追加情報として使用して，デコード中に候補語を生成する．深い統合方法は，ニューラル機械翻訳モデルの隠れ状態と言語モデルを統合し，制御メカニズムを介してデコードに対する 2 つのモデルの効果を動的にバランスさせ，デコード中に言語モデル情報をキャプチャできる．これらの統合方法はどちらも翻訳効果を改善でき，深い統合方法はより効果的である．さらに，Domhan ら (Domhan & Hieber 2017) は，ニューラル機械翻訳へのマルチタスク学習アプローチを提案した．目的言語の翻訳モデルと言語モデルは，大規模な目的言語の単言語コーパスを利用するために一緒に訓練する．

目的言語の単言語コーパスの別の使用法として，Sennrich ら (Sennrich et al. 2016a) が提案した逆翻訳 (Back-translation method) によって提案された訓練データの拡張手法が挙げられる．疑似データは，目的言語の単言語コーパスを使用して構築され，訓練コーパスに追加される．この手法は，翻訳モデルを変更せず，シンプルで効果的であり，翻訳効果はある程度改善されるが，改善の効果は生成されたデータの品質に依存する．

上記の研究はすべて目的言語の単言語コーパスを使用しているが，Zhang ら (Zhang & Zong 2016) は原言語の単言語コーパスをニューラル機械翻訳に適用する方法を提案した．これを実現するには 2 つの方法があるが，最初の方法では，自己学習法によってバイリンガル訓練コーパスのサイズを拡大する．もう一つの方法では，マルチタスキングにより原言語に対するエンコーダの表現品質を向上させる．これらの方法はどちらも，翻訳効果を大幅に改善できる．欠点は，原言語の単言語コーパスのサイズと内容が翻訳モデルのパフォーマンスに影響を与える可能性がある点である．

原言語と目的言語の単言語コーパスが使用され，主に Cheng ら (Cheng et al. 2016) によって提案された半教師あり学習法がある．基本的な考え方は，原言語と目的言語の単言語コーパスを使用して翻訳効果を向上させるために，ソースからターゲットへの翻訳モデルとターゲットからソースへの言語翻訳モデルのオートエンコード（Auto encoder）を原

言語に導入し，半教師ありの方法で双方向ニューラル機械翻訳を訓練するというものである．この方法の明らかな利点は，原言語と目的言語の単言語コーパスを同時に使用できることであり，欠点は，単言語コーパスの未登録単語を処理できないことがある．さらに，Ramachandran ら (Ramachandran et al. 2017) は，Sequence-to-Sequence モデルを 2 つの言語モデルと見なし，原言語と目的言語の言語モデルを大規模な単言語コーパスを通じて個別に訓練する，従来より簡単な方法を提案した．ニューラル機械翻訳モデルのエンコーダおよびデコーダのパラメータは，それぞれ 2 つの言語モデルのパラメータによって初期化される．次に，バイリンガルコーパスを使用して，訓練中に言語モデルのパラメータが同時に調整される．

3.2.6　ニューラル機械翻訳のドメイン適応に関する研究

十分なサイズの高品質の対訳コーパスは英語などのヨーロッパ言語ペアでしか使用できない．言語ペアごとに，ドメイン固有のコーパスと使用可能なドメインの数が制限されている．大部分の言語ペアとドメインでは，利用可能な対訳コーパスはほとんどない．汎用の翻訳システムはパフォーマンスが低いため，特定のドメイン適応の翻訳システムを開発することが重要である．

ChineaRios らは (Chinea-Ríos et al. 2017)，合成対訳データを活用することにより，一般的な NMT システムを特定のタスクに適応させる新しい手法を提案した．この手法は，原言語文の大きな単言語プールから，特定のテストセットに関連する実例を選択することで合成対訳データを構成する．

Gu らは (Gu et al. 2019)，ドメインの共通の特徴と特定のドメインの固有の特徴，これら 2 種類の情報を使用する Sequence-to-Sequence モデルを提案した．さらに，識別機能を追加して，翻訳のパフォーマンスを改善した．実験の結果は，マルチドメインデータで提案されたモデルの性能が最高の性能に達した．

3.2.7　言語資源不足の言語への対応に関する研究

ニューラル機械翻訳は，大規模な対訳コーパスの条件下でより良い翻訳品質を達成している．一部のリソース不足の言語または特定の領域の翻訳タスクでは，対訳コーパスの規模が比較的小さいめ，翻訳効果が大幅に低下する (Koehn & Knowles 2017)．したがって，リソース不足の条件下でのニューラル機械翻訳の研究は，実用的な価値が高い．

より多くの外部知識を統合することは，リソースの少ない言語のニューラル機械翻訳の翻訳効果を改善する方法の 1 つである．たとえば，バイリンガル辞書，モノリンガルのコーパスの追加，多言語ニューラル機械翻訳，マルチタスキングなど．これらの方法

は，基本的により多くの外部知識を組み合わせて，ニューラル機械翻訳の翻訳能力を向上させ，単語の意味情報とバイリンガル単語間の対応をモデル化する．対訳コーパスの数を増やすことは，リソース不足の言語でのニューラル機械翻訳の品質を改善する効果的な方法である．たとえば，逆翻訳メソッドを使用して，対訳コーパスをすばやく構築する (Sennrich et al. 2016a)．また，目的言語の文を原言語の文として使用するコピー法 (Currey et al. 2017)．さらに，低頻度単語に対する Data Augmentation 手法も，対訳コーパスを拡張するための効果的な方法である (Fadaee et al. 2017)．

Wang らは (Wang et al. 2018)，対訳コーパスの拡張方法について研究を行った．コーパス拡張の最適化問題を定式化し，一般的な分析ソリューションを導き出す．原言語文と目的言語文の両方の単語を，対応する語彙からの他のランダムな単語にランダムに置き換える．異なるスケールの 3 つの翻訳データセットでの実験は，BLEU スコアが約 0.5 ポイントの一貫した改善をもたらした．Gao らは (Gao et al. 2019)，文脈を考慮したソフトなデータ拡張手法を提案した．文内の単語をランダムにドロップ，スワップ，または置換する従来の拡張方法とは異なり，複数の関連する単語の文脈混合によって，文内のランダムに選択された単語をソフトに拡張する．単語のワンホット表現を，語彙上の分布（言語モデルによって提供される）に置き換える．小規模および大規模の機械翻訳データセットの両方での実験結果は，従来の拡張手法より良い結果を取得した．

転移学習を利用して，十分な量の対訳データを持つ言語対の NMT モデルのパラメータを少ない量の対訳データを持つ言語対の NMT モデルへ転移するアプローチは，この問題の解決策の一つである．Firat ら (Firat et al. 2016) はこの考え方に基づいて，訓練された豊富な量の対訳データを持つ言語対（フランス語-英語）NMT モデルを親モデル，少ない量の対訳データを持つ言語対（スペイン語-英語）の NMT モデルを子モデルとし，親モデルのいくつかのパラメータを子モデルに転送することによって，少ない量の対訳データを持つ言語間の翻訳精度を大幅に改善した．しかし，この方法には，親モデルと子モデルは類似の言語構造を持たなければならないという制約がある．

少ない量の対訳データを持つ言語間の翻訳に豊富な量の対訳データを持つ言語対を利用することも効果的である．通常，中間言語を介して実装される．例えば，A，B，C の 3 つの言語の場合，A と C の間に対訳コーパスは存在しないが，A と B，B と C の間に何らかの対訳コーパスがあれば，B を中間言語としてて A と C が繋げる．Chen ら (Chen et al. 2017) が提案したの‘教師-学生’フレームワーク，Zheng ら (Zheng et al. 2017) が提案したの最尤推定法，Cheng ら (Cheng et al. 2017) が提案した Joint training 法などがある．基礎の方法はより少ないパラメータしか必要としないが，効果はより弱い．他の方法は，多量のパラメータを有するが，翻訳効果を大幅に改善することができる．

3.2.8　超特大言語資源下のニューラル機械翻訳に関する研究

Google の研究者たちは (Aharoni et al. 2019)，250 億を超える文対で NMT モデルを訓練した．これは，500 億を超えるパラメータを持つ，100 以上の言語から英語への単一の NMT モデルで翻訳される．結果により，言語資源が豊富と不足の両方で翻訳パフォーマンスが大幅に向上し，単一のドメイン/言語にも簡単に適応できた．

Meng らは (Meng et al. 2019b)，400 億を超える多言語文対から成る，これまでで最大規模のコーパスで NMT モデルを訓練した．このような状況では，データのノイズや非常に長い訓練時間など，以前の NMT 作業と比較して前例のない課題が生じる．これらの問題に対処するための実践的な解決策を提案し，大規模な事前訓練によって NMT の性能が大幅に向上することを実証した．WMT17 タスクで中英翻訳の BLEU スコアを 32.3 に上げることができ，既存の最先端の結果に対して +3.2 の大幅なパフォーマンス向上を遂げた．

3.2.9　ニューラル機械翻訳の頑健性に関する研究

ニューラル機械翻訳は大きな成功を収めているものの，入力データの微修正に対して非常に敏感だという弱点を持つ．入力文の中の 1 つ単語を同義語に入れ替えただけで，翻訳の出力文が全く違うものとなってしまう可能性が高い．

Liu らは (Liu et al. 2019)，NMT の頑健性，特に同音異義語のノイズを改善した．翻訳の入力時に，単語の特徴情報として，入力語の発音情報を単語ベクトルに追加する．実験結果は，ノイズ下での翻訳システムの頑健性を大幅に改善するだけでなく，ノイズなし条件下での翻訳システムのパフォーマンスも大幅に改善した．

Vaibhav らは (Vaibhav et al. 2019)，ノイズがあるテキストデータを活用し，クリーンデータの自然発生ノイズを模倣・合成することにより，NMT システムの頑健性を強化していた．このようにノイズを合成することで，最終的には，NMT システムを自然に発生するノイズに強くし，そこから生じる精度の損失を部分的に軽減することができる．

Cheng らは (Cheng et al. 2019)，人間が識別できない程度のノイズを画像にのせることで翻訳モデルを混乱させる「Adversarial Examples」というアルゴリズムを取り入れた．この手法は敵対的生成ネットワーク (GAN) に触発されているが，真偽を判定する Discriminator に頼るのではなく，Adversarial Examples を学習に取り入れて訓練データを多様化・拡張したものとなった．「中国語-英語」「英語-ドイツ語」という組みあわせの翻訳タスクでベンチマークを行ったところ，既存の Transformer モデルと比べ，BLEU スコアがそれぞれ 2.8 ポイントと 1.6 ポイントの向上がみられた．

3.2.10　新しいモデルと新しいアーキテクチャ

ニューラル機械翻訳の研究は急速に発展しており，従来のニューラル機械翻訳モデルに加えて，いくつかの新しいモデルと新しいアーキテクチャが主に次のように提案されている．

マルチモーダルニューラル機械翻訳

マルチモーダルニューラル機械翻訳で利用されるリソースはテキストに限定されない．現在の研究は画像情報を使用してニューラル機械翻訳の翻訳を改善することに焦点を当てている (Calixto et al. 2017; Delbrouck & Dupont 2017; Calixto & Liu 2017; Caglayan et al. 2016).

このタイプの方法は通常，2 つのエンコーダを使用する．エンコーダは通常のニューラル機械翻訳と同じ方法でテキスト情報をエンコードし，別のエンコーダは画像情報をエンコードする．デコード時には，Attention メカニズムを介して異なるモーダル情報が翻訳に適用される．

非リカレントニューラルネットワークのニューラル機械翻訳モデル

ほとんどのニューラル機械翻訳モデルは，リカレントニューラルネットワークによって実装されるが，モデルのタイミング依存性により，並列処理が困難であるため，訓練とデコードの速度が遅くなる．

Gehring ら (Gehring et al. 2017) は，完全に畳み込みニューラルネットワークに基づく Sequence-to-Sequence モデルを提案した．従来のニューラル機械翻訳モデルと比較して，速度が約 10 倍向上し，翻訳品質も大幅に向上した．また，前述したの Transformer モデル (Vaswani et al. 2017) はリカレントニューラルネットワークと畳み込みニューラルネットワークを放棄し，Attention メカニズムのみを使用して Sequence-to-Sequence モデルを実装した．

事前学習の言語モデル **BERT**，**XLM** および **XLNet**

2017 年に機械翻訳のために Google によって提案された Transformer (Vaswani et al. 2017) は，異なる単語またはサブワード間の文脈を学習しながらテキスト入力全体を処理するために Attention メカニズムを使用する． Transformer には，エンコーダとデコーダが含まれる．エンコーダは，入力テキストを（単語ベクトルなどの）特徴表現に変換する．デコーダは，前の特徴表現を通じて翻訳された結果を生成する．

しかし，Transformer はテキストを処理するときに限られた文脈情報しか利用できない．Google が 2018 年に Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al. 2019) を提案するまで，この状況は改善しなかった．BERT は Transformer の Encoder を使用して，単語の一部をランダムに Mask してからマスクされた単語を予測することで，言語モデルを学習する．この学習過程は，マスクされた単語の前と後を含む完全な文脈情報を利用することができる．BERT は，テキスト分類と機械翻訳の 2 つのタスクでより良い結果を更新した．

BERT は 100 種類を超える言語を学習するが，BERT はクロスランゲージモデルとしては最適ではなく，異なる言語間の多数の語彙を共有していないため，共有される知識は限られる．この問題を解決するために，Cross-lingual Language Model (XLM) はいくつかの方法で BERT を改善する (Lample & Conneau 2019)．BERT に基づいて，2 つのアップグレードが行われた．XLM の訓練サンプルは，内容は同じだが言語が異なるの 2 つテキストで構成されるが，BERT の訓練サンプルは単一言語である．BERT の目的はマスクされた単語を予測することだが，XLM モデルの目的はそれだけではない．ある言語の文脈を使用して別の言語のトークンを予測できる．さらに，各言語はランダムに Mask される．XLM は各言語の言語 ID とトークンの位置情報も入力する．BERT と比べて，これらの新しいデータは異なる言語に関連付けられたトークン間の関係情報をよりよく学習するのに役立つ．

このようなのクロスリンガル言語モデルは事前学習と言語横断でサブワード語彙を共有する．多言語対応の文表現を得る際，どんなタスクが良いのか検証した研究である．ベースは言語モデルで，通常通り次の単語を予測する Causal LM，単語を Mask した箇所を予測する Masked LM および翻訳データがある場合に，並べた文で Masked LM を行う Translation LM の計 3 つを提案した．MLM は CLM より良いであるが，TLM を使用すれば，CLM と MLM を強化できるという結果になった．

他には，BERT の弱点を修正した Generalized Autoregressive Pretraining for Language (XLNet) も提案された (Yang et al. 2019)．BERT では Mask 箇所を予測するが，‘Mask’ は通常発生しないためノイズになる．そこで単語の予測時に使用する Context の順序を変える手法を提案した．Self を含まない Context から予測する一方，Context 自体は通常の Self を含む Attention で作成する，自己回帰モデルによる学習を可能にした NLP モデルになった，20 種類の言語処理タスクで BERT を上回る成果を得た．

教師なしのニューラル機械翻訳

教師あり機械翻訳の問題点の一つは，大量の対訳文が必要なことである．Artetxe らは (Artetxe et al. 2018)，機械翻訳で初めて本格的に教師なし学習手法を提案した．それは，共通のエンコーダに通し得られた表現ベクトルを元にターゲット言語に翻訳する．元の文

にノイズを入れ，ノイズ除去を行うことで言語知識を取得する．また，学習途中のモデルを使って疑似対訳コーパスを生成し，逆翻訳された文と元の文が同じになるように学習する．この研究は，教師なし学習に貢献が非常に大きいが，性能的には改善の余地がある．

Facebook の研究者たちは (Lample et al. 2018)，フレーズベースの教師あり学習手法 Phrase Based Statistical Machine Translation（PBSMT）によって，両言語の翻訳ペアデータから，フレーズごとに言語変換テーブルを作成し，翻訳時は変換スコアの最大化問題を解く手法を提案した．英仏翻訳タスクで Artetxe らの手法より，BLEU スコアを +13 の大幅なパフォーマンス向上を遂げた．この論文は EMNLP2018 の Best Paper であった．また，事前学習の言語モデルを利用して，教師なし NMT をより良い結果を更新した (Lample & Conneau 2019).

新しい学習パラダイム

現在，一部の研究者は，ニューラル機械翻訳に新しい学習パラダイムを適用しようと考えている．たとえば，デュアル学習 (Dual Learning) を使用して，対訳コーパスの使用量を大幅に削減している (He et al. 2016)．強化学習 (Reinforcement Learning) を通じて人工的なフィードバック結果を適用するニューラル機械翻訳 (Nguyen et al. 2017); Yang ら (Yang et al. 2018) および Wu ら (Wu et al. 2018) は，Generative Adversarial Network(GAN) を独立ニューラル機械翻訳に適用し，翻訳効果を大幅に改善した．これらの探索的研究は，ニューラル機械翻訳に新しい視点を提供する．

第 4 章

文字レベルの日中ニューラル機械翻訳における文字特徴情報の利用

4.1 はじめに

近年，ニューラル機械翻訳（NMT）は注目すべき成果をあげている (Bahdanau et al. 2014; Luong et al. 2015)．単語レベルの NMT における問題点として，語彙サイズが制限されることが挙げられる．日本語や中国語のように文中の単語の区切りが明示されない言語では，統一された正しい単語分割結果を得ることも容易ではない．文字レベルの NMT では，これらの問題を回避することができる．

一方，R. Sennrich & B. Haddow (2016) は，通常の単語レベルの NMT において，POS（品詞）タグなどの単語の特徴情報が翻訳精度の向上に有効であることを示した．本章では文字レベルの NMT においても何らかの文字特徴情報が有用ではないかと考え，漢字の部首を入力特徴情報として加えて，文字レベルの NMT による日本語から中国語への機械翻訳を試みた．その結果，部首を特徴情報として加えることにより翻訳精度の向上が見られた．NMT システムは Minh Thang Luong et al. (2015) のものをベースとして用い，実験には WAT2017 の学術論文サブタスクでも用いられた ASPEC-JC コーパス (Nakazawa et al. 2016) を文字ごとに分割して使用した．

本研究では文字の特徴情報の一つとして漢字の部首を用いる．六書では漢字の造字法・用字法を，象形・指事・形声（形聲）・会意・転注・仮借の 6 つに分類しているが，漢字の 80% 以上は，意符（意味成分，物事の類型を表す）と音符（発音を表す）を組み合わせて作られた形声文字であると言われている．例えば，「銅」の部首「釒」（かねへん）は金属という意味カテゴリを表し，「同」は音を表す．そこで，部首がもつ意味的な情報が翻訳精度の向上につながることを期待して，入力特徴情報に加えた．

4.2　関連研究

どの言語においても単語の数に比べて文字の数は遥かに少ない．文字レベルでの自然言語処理の利点は，言語モデル (Kim et al. 2016)，POS タグ付け (Santos & Zadrozny 2014)，固有表現抽出 (Santos & Guimarães 2015)，構文解析 (Ballesteros et al. 2015)，学習表現 (Chen et al. 2015) など，これまでにもいくつか示されてきた．

欧米の言語を対象とした NMT では，単語を文字ではなく部分文字列（サブワード）に分割することで語彙サイズの制約に対処する方法も提案されている (Sennrich et al. 2016b)．しかし，単語が比較的多くの文字で表現される欧米の言語に比べ，表語文字である漢字を使用する日本語や中国語では単語の文字数が少なく，特に中国語では一文字の単語も多い．そのような単語をサブワードに分割することは困難である．

日中両言語間の NMT に対して，サブ文字の情報にも翻訳品質を改善できる．Zhang らは，原言語側または目的言語側でサブ文字の系列を使用する手法を提案した (Zhang & Komachi 2018)．Du と Way は原言語側で「ピンイン」系列を使用し、原言語側の中国語文字が分解された NMT モデルを訓練した (Du & Way 2017)．ピンインは、中国語で音節を音素文字に分け、ラテン文字化して表記する発音表記体系を指す．Wang らは，NMT モデルを構築する前に，BPE アルゴリズムを文字系列に直接適用した (Wang et al. 2017)．

最近，Meng ら (Meng et al. 2019a) は，中国語を含む NLP モデルの中に，文字レベルのモデルがそれ以外のサブワードおよび単語レベルのモデルより優れていることを発見した．本研究では，文字レベルの日中ニューラル機械翻訳における文字の特徴情報の一つとして漢字の部首を追加することに，より良い翻訳結果を得た．

4.3　NMT と特徴情報の追加

NMT は原言語文に対する目的言語文の条件付き確率を計算する．ここでは本研究で使用する Luong et al. (2015) による NMT システムについて，文献 Sennrich & Haddow (2016) を元に簡単に説明する．この NMT システムは，リカレントニューラルネットワークを用いたグローバルな注意機構付きのエンコーダ・デコーダモデルを実装したものであるが，本研究ではこれを文字レベルで利用する．エンコーダは，双方向 LSTM リカレントニューラルネットワークであり，入力系列 $\boldsymbol{x} = (x_1, \cdots, x_m)$ を読み取って，順方向の隠れ状態列 $(\overrightarrow{h_1}, \cdots, \overrightarrow{h_m})$ と逆方向の隠れ状態列 $(\overleftarrow{h_1}, \cdots, \overleftarrow{h_m})$ を求める．隠れ状態 $\overrightarrow{h_j}$ と $\overleftarrow{h_j}$ は連結され，アノテーションベクトルが作られる．

デコーダは，目的言語文 $\boldsymbol{y} = (y_1, \cdots, y_n)$ を予測する LSTM リカレントニューラルネットワークである．各単語（文字レベルの場合，各文字）y_i は，リカレント隠れ状態 s_i と，

前回予測された単語（または文字）y_{i-1}，文脈ベクトル c_i を元に予測される．文脈ベクトル c_i は，アノテーション h_j の加重和として計算される．各 h_j の重みは，y_i と x_j のアラインメントについての情報を表すモデル α_{ij} を通じて決められる．

エンコーダの順方向状態は以下のように表される．

$$\vec{h_j} = \tanh(\vec{W}Ex_j + \vec{U}\vec{h}_{j-1}) \tag{4.1}$$

ここで，$E \in \mathbb{R}^{p \times V_x}$ は単語埋め込み行列であり，$W \in \mathbb{R}^{q \times p}$ と $U \in \mathbb{R}^{q \times q}$ は重み行列である．p，q，V_x はそれぞれ，単語ベクトルのサイズ，隠れユニットの数，原言語の語彙サイズである．

入力特徴情報の数を $|F|$ とすると，式 (5.1) は次のように一般化することができる．

$$\vec{h_j} = \tanh(\vec{W}(\mathop{\Big\|}_{k=1}^{|F|} E_k x_{jk}) + \vec{U}\vec{h}_{j-1}) \tag{4.2}$$

ここで，演算子 || はベクトルの連結を表す．$E_k \in \mathbb{R}^{p_k \times V_k}$ は特徴埋め込み行列であり，$\sum_{k=1}^{|F|} p_k = p$ である．V_k は k 番目の特徴情報の語彙サイズ（種類数）である．特徴情報の埋め込みベクトルは，単語埋め込みベクトルと同じ方法で別々に求められ，最後に単語埋め込みベクトルと連結される．

4.4 ASPEC-JC コーパス

本研究では日中対訳コーパスとして，Asian Scientific Paper Excerpt Corpus（ASPEC）の日中学術論文抜粋コーパス（ASPEC-JC）を使用した (Nakazawa et al. 2016)．これは文献データベース JDream II 搭載の和文妙録と，電子ジャーナルサイト J-STAGE の情報処理学会，言語処理学会，人工知能学会論文誌の和文論文の妙録を人手で中国語に翻訳して構築された対訳コーパスであり，医学，情報，生物，環境，化学，材料，農業，エネルギー分野の論文の和文妙録とその中国語訳から成る．コーパスは表 5.3 に示すように，train（672,315 文対），dev（2,090 文対），dev-test（2,148 文対），test（2,107 文対）の 4 つデータセットで構成されており[*1]，互いに同じ論文に属する文は含まれていない．本研究では train を学習データ，dev をバリデーションデータ，test をテストデータとして使用した．

[*1] 中国訳の内容が“..”のみの文が，train データと test データに少量含まれるが，本研究ではそれらを除外して使用した．

表 4.1　ASPEC-JC コーパスの対訳文対数

Data Type	File Name	Number of sentences
TRAIN	train.txt	672,315
DEV	dev.txt	2,090
DEVTEST	devtest.txt	2,148
TEST	test.txt	2,107

4.5　日本語文字の特徴情報

本研究では，日本語から中国語への文字レベルのニューラル機械翻訳において，文字の埋め込みベクトルに，文字の部首を入力特徴情報として加える．

4.5.1　部首

部首は漢字の構成要素の一つであり，漢字を字画の構成で分類・配列する際に基準として用いられる．部首によって漢字を分類した辞典を（狭義の）字書というが，ある漢字がどの部首に分類されるかは字書による．意味カテゴリを表す意符（形符）と音声を表す音符（声符）で構成される形声文字では，意符が部首として用いられることが多い．例えば，「江」「河」の部首「氵」（さんずい）は水を表し，残り部分「工」「可」は音を表す．康熙字典では漢字が 214 の部首に分けられ，画数順に記載されている．康熙部首を図 4.1 に示す．康熙部首はすべて Unicode に収録されている（U+2F00～2FD5）．

現代の日本語の標準的な文章には漢字と仮名，数字，英字，句読点や括弧などの記号が混在する．漢字以外の文字には部首は存在しないが，本研究では日本語入力文中のすべての文字に特徴情報を付与するため，漢字以外の文字に対しても以下のように部首を設定した．

仮名は，日本語を表記するために漢字の音を借用して用いられた万葉仮名（借字）が元になっており，平仮名は万葉仮名の草書化が進められて独立した字体になったもの（図 4.2），片仮名は漢文を和読するための訓点として万葉仮名の一部を省略して付記したものが始まりと考えられている．

ひらがなは，漢字ではなく，文章の表記に用いる場合と音を示すことを目的とする場合に用いられる．カタカナは，主に外来語，植物および動物の名前のような固有名詞のために使用される．ひらがなやカタカナには部首が存在しない．しかし，これらは漢字から派生しているため，便宜的に元の漢字の部首をそれらの部首として使用する．図 4.2 は，漢字からひらがなへの変化を示す．上の部分は元の漢字を示し，中央の部分は草体化した漢

一	丨	丶	丿	乙	亅	二	亠	人	儿	入	八	冂	冖	冫	几
凵	刀	力	勹	匕	匚	匸	十	卜	卩	厂	厶	又	口	囗	土
士	夂	夊	夕	大	女	子	宀	寸	小	尢	尸	屮	山	巛	工
己	巾	干	幺	广	廴	廾	弋	弓	彐	彡	彳	心	戈	戸	手
支	攴	文	斗	斤	方	无	日	曰	月	木	欠	止	歹	殳	毋
比	毛	氏	气	水	火	爪	父	爻	爿	片	牙	牛	犬	玄	玉
瓜	瓦	甘	生	用	田	疋	疒	癶	白	皮	皿	目	矛	矢	石
示	内	禾	穴	立	竹	米	糸	缶	网	羊	羽	老	而	耒	耳
聿	肉	臣	自	至	臼	舌	舛	舟	艮	色	艸	虍	虫	血	行
衣	襾	見	角	言	谷	豆	豕	豸	貝	赤	走	足	身	車	辛
辰	辵	邑	酉	釆	里	金	長	門	阜	隶	隹	雨	青	非	面
革	韋	韭	音	頁	風	飛	食	首	香	馬	骨	高	髟	鬥	鬯
鬲	鬼	魚	鳥	鹵	鹿	麥	麻	黃	黍	黑	黹	黽	鼎	鼓	鼠
鼻	齊	齒	龍	龜	龠										

図 4.1　214 康熙部首

字の字形を示し，下は等価のひらがなを示す．上部から，ひらがなの元の漢字を見つけ，元の漢字の部首を得ることができる．カタカナの部首も同じように得ることができる．

アラビア数字は，対応する漢数字の部首を使用する．英字には一律に「英」の部首（くさかんむり）を割り当て，記号には「符」の部首（たけかんむり）を割り当てる．

4.5.2　部首の取得

表 4.2 は日本語の原文の一部と，その各文字に対応する康熙部首を示している．この文には，漢字，平仮名，数字，英字，記号が含まれている．

入力文中の各文字の部首を取得するために，本研究の実験では cjklib[*2]を使用した．cjklib は中国，日本，韓国で使われる漢字の発音，部首，グリフの構成部品，筆画，異体字などの情報を得るための Python ライブラリである．現時点では Python3 に対応していないため，Python3 で使用できるように一部手を加えて使用した．

[*2] https://github.com/cburgmer/cjklib

図 4.2　漢字から平仮名への変化（Wikipedia「平仮名」より転載）

表 4.2　日本語入力文字列と各文字の特徴情報の例

日本語文	溝幅は１０ｍｍ以上が必要と推定した．
康熙部首	水巾水一雨艹艹人一力心襾止手宀丿大竹

4.6　翻訳実験

実験には，ASPEC（Asian Scientific Paper Excerpt Corpus）の日中学術論文抜粋コーパスを使用した．モデル中のパラメータは [− 0.1，0.1] を範囲とする一様分布の乱数により初期化を行い，バイアス項は 0 とした．各パラメータの学習には確率的勾配降下法（初期学習率は 1.0）を用い，ミニバッチサイズを 10 とした．勾配ノルムは 1 でクリップした．また，単語ベクトル，隠れ層の次元は全て 512 とした．過学習を避けるため，dropout 確率は 0.8 に設定し，デコード時に行うビームサーチのビームサイズは 5 とした．

文字ベースでの翻訳のため，日本語テキスト・中国語テキストともに文字ごとに空白文字を挿入して分割するが，単語ベースの翻訳システムと同じ条件で BLEU スコアを計算するために，出力テキスト中の空白文字をいったん取り除いた後，中国語文は Python モジュール Jieba[*3]を使って，日本語文は Mecab を使って[*4]単語に分割した．

翻訳システムの実装には OpenNMT を用いた (Klein et al. 2017)．訓練には NVIDIA 社

[*3] http://github.com/fxsjy/jieba
[*4] http://taku910.github.io/mecab

の GeForce GTX 1080Ti を使用したところ，テストデータの翻訳の処理時間は 1 秒あたり約 3 千文であり，モデルの訓練には 3〜4 日かかった．

実験の結果を表 4.3 と表 4.4 に示す．表中の「ppl」は perplexity を表す．これはモデルが与えられた原文の参照翻訳をどの程度うまく予測できるかを示すのに有効な指標である．文字の特徴情報として部首のみを追加した文字レベルの日中 NMT システムでは，devtest データと test データでそれぞれ BLEU 値 39.62 と 39.65 を得た．特徴情報を何も追加しない文字レベルの翻訳結果と比べて 0.4〜0.6 向上した．さらに，dropout を 0.3 に調整したとき，perplexity が 3.07，devtest データと test データで BLEU 値がそれぞれ 40.58 及び 40.61 となり，最も良い結果が得られた．文字の特徴情報として部首のみを追加した文字レベルの中日 NMT システムでは，devtest データと test データで BLEU 値はそれぞれ 39.68 及び 39.53 となった．特徴情報を何も追加しない文字レベルの翻訳結果よりも BLEU 値がそれぞれ 0.03 及び 0.25 向上した．さらに，dropout を 0.3 に調整したとき，perplexity が 2.32，devtest データと test データで BLEU 値がそれぞれ 41.39 及び 41.22 となり，最も良い結果が得られた．この実験により，部首を特徴情報として加えることにより文字レベルの NMT システムは日中両言語の機械翻訳において，さらに良い結果が得られることがわかった．

同時に，表 4.5 に示す翻訳結果の一部を観察することによって，提案した NMT が，単語レベルの NMT と比較して，単語の翻訳精度を向上させることができることがわかった．表中の「src」は入力文，「ref」は人手による翻訳結果，「best」は翻訳モデルから得られた最も良い翻訳結果，「base」は何も追加しない文字レベルの NMT での翻訳結果を表す．

日中翻訳結果の文を観察したところ，部首を特徴情報として追加した提案手法では，特徴情報を何も追加しない文字レベルの NMT と比較して，単語の翻訳精度が向上している例が見られた．例えば，表 4.5 の文では「正常な状態」を「漂白状態」と正しく翻訳し，「界面活性剤」を「界面活性剤」と正しく翻訳したが，「ヘキサデシルトリメチルアンモニウムブロミド」のような翻訳が困難な単語は NMT によって正しく翻訳されなかった．中日翻訳結果では, 例文 (1) において，「过滤中」，「法罗群岛」などの単語は提案手法により正しく翻訳できるようになった．「法罗群岛」については，ベースライン（「法羅群島」）では簡体字から日本の漢字への置き換えだけが行われたが，提案手法では「ファロー諸島」と正しく翻訳された．ただし，「塞舌尔群岛」の翻訳については，提案手法とベースラインはともに誤訳となった．

4.7　おわりに

本章では，部首を文字の特徴情報として追加することで，日中両言語の文字レベルのニューラル機械翻訳をさらに改善できないか検討し，ASCPEC-JC コーパスを用いた実験

表 4.3　日中実験結果

システム	ppl (↓)	BLEU (↑)	
	dev	devtest	test
文字レベル（追加特徴情報なし）	3.73	39.03	39.25
文字レベル + 部首	3.64	**39.62**	**39.65**
（同上，dropout 調整時）	**3.07**	**40.58**	**40.61**

表 4.4　中日実験結果

システム	ppl (↓)	BLEU (↑)	
	dev	devtest	test
文字レベル（追加特徴情報なし）	2.59	39.65	38.78
文字レベル + 部首	2.58	**39.68**	**39.53**
（同上，dropout 調整時）	**2.32**	**41.39**	**41.22**

でその効果を確認した．その結果，漢字，仮名やアラビア数字などの文字にも部首を設定し，文字の特徴情報として加えることにより，翻訳精度を向上させることができた．日本語から中国語への翻訳について，特徴情報を追加しない文字レベルの NMT と比較すると，パープレキシティは約 0.1，BLEU 値は devtest データと test データでそれぞれ約 0.5 および 0.4 向上した．中国語から日本語への翻訳について，特徴情報を追加しない文字レベルの NMT と比較すると，パープレキシティは約 0.1，BLEU 値は devtest データと test データでそれぞれ約 0.03 および 0.7 向上した．

中国語から日本語，あるいは，日本語から他の言語への文字レベルの翻訳においても，部首やその他の特徴情報が翻訳精度の向上に役立つ可能性があると考えられる．

表 4.5　翻訳実験結果の一部

日中翻訳結果 (1)	
src	製造工程の作業性や着生状況を解析し，摂食抑制用の溝幅は１０ｍｍ以上が必要と推定した．
ref	解析了制造工程的工作状况及着生状况，推断了抑制摄食时的水沟宽度为１０ｍｍ以上．
best	对制造工序的作业性和附着状况进行分析，推测用于抑制摄食的沟宽需要１０ｍｍ以上．
base	分析了制造工程的作业性和着生状况，进食抑制用的沟幅在１０ｍｍ以上．
日中翻訳結果 (2)	
src	硫酸ジルコニウムメソ多孔質構造体（ＺＳ）は，Ｚｒ（ＳＯ４）２・４Ｈ２Ｏと界面活性剤ヘキサデシルトリメチルアンモニウムブロミドを用いて，１００℃で４８時間水熱反応して合成した．
ref	硫酸锆介多孔质构造体（ＺＳ）是使用Ｚｒ（ＳＯ４）２・４Ｈ２Ｏ和界面活性剂溴化十六烷基三甲铵，在１００℃下经过４８小时水热反应合成的．
best	硫酸锆膜多孔结构体（ＺＳ）使用Ｚｒ（ＳＯ４）２・４Ｈ２Ｏ和表面活性剂十六烷基三甲基溴铵，在１００℃下进行４８小时的水热反应合成．
base	硫酸锆的多孔质结构体（ＺＳ）使用Ｚｒ（ＳＯ４）２・４Ｈ２Ｏ和界面活性剂己烷基三甲基铵，在１００℃下进行４８小时水热反应合成．
中日翻訳結果 (1)	
src	过滤中使用的纤维材料很多，不仅仅是提高了材料特性，还改良了纤维的形状．
ref	ろ過に利用する繊維材料は多様であり，材料特性を向上しただけでなく繊維の形状も改良した．
best	ろ過に使用した繊維材料は多く，材料特性の向上だけでなく，繊維の形状を改良した．
base	フィルタリングに用いた繊維材は多く，材料特性を高めるだけでなく，繊維の形状を改良した．
中日翻訳結果 (2)	
src	在本次的风险评价中通过重新审核厚生劳动省公布的注意事项，本评价以《法罗群岛的前瞻研究》与《塞舌尔群岛的儿童成长研究》为基础．
ref	今回のリスク評価は厚生労働省が公表した注意事項の見直しの検討にあたり，「フェロー諸島前向き研究」と「セイシェル小児発達研究」を基としている．
best	今回のリスク評価では，厚生労働省の公表する注意事項を見直すことにより，「ファロー諸島の前向き研究」と「セイヨウ群島の児童成長研究」を基礎とした．
base	今回のリスク評価では，厚生労働省における注意事項を見直すことにより，本評価は「法羅群島の展望研究」と「塞舌ル群島の児童成長研究」に基づいている．

第 5 章

ニューラル機械翻訳における長文分割によるコーパスの拡張

5.1 はじめに

近年，ニューラル機械翻訳（Neural Machine Translation， NMT）の登場によって流暢で精度の高い翻訳が可能となってきた．NMT では翻訳の品質が学習のための対訳データの量に強く依存し，質の高い翻訳結果を得るには大量の対訳データを必要とする．しかし，英語を含む言語対や欧州の言語間の言語対などを除き，一般に十分な量の対訳データを入手するのは困難である．これは NMT の大きな問題点の一つであり，その解決策がいくつか提案されてきた．

転移学習を利用して，十分な量の対訳データを持つ言語対の NMT モデルのパラメータを対訳データの少ない言語対の NMT モデルへ転移するアプローチは，この問題の解決策の一つである．Firat ら (Firat et al. 2016) はこの考え方に基づいて，訓練された豊富な量の対訳データを持つ言語対（フランス語-英語）の NMT モデルを親モデル，対訳データの少ない言語対（スペイン語-英語）の NMT モデルを子モデルとし，親モデルのいくつかのパラメータを子モデルに転送することによって，少ない量の対訳データを持つ言語間の翻訳精度を大幅に改善した．しかしこの方法には，親モデルと子モデルが類似の言語構造を持たなければならないという制約がある．

Zero-shot 翻訳は，単一の NMT エンジンを使用して複数の言語間の翻訳を行うメカニズムである．対訳データが提供されない言語資源が少ない言語ペアにも対応する．このようなの研究は，主に Google によって提案された (Johnson et al. 2017)．Lakew らは Zero-shot 翻訳に対して，翻訳モデルによって生成された逆翻訳を活用する単純な反復訓練方法を提案した (Lakew et al. 2018)．Mattoni らは，訓練データが少ない言語ペアに焦点を合わせた (Mattoni et al. 2017).

言語 A，B，C において，言語対（A, C）の対訳コーパスは存在しないが，（A, B）と（B, C）の言語資源は豊富にある状況において，言語 B を介して（A, C）間の翻訳を行う中間翻訳方式がある．これをベースとして，Chen ら (Chen et al. 2017) の「教師-学生」フレームワーク，Zheng ら (Zheng et al. 2017) の最尤推定法，Cheng ら (Cheng et al. 2017) の合同訓練法などの改良法も提案されている．

他の解決策としてデータ拡張（水増し）の手法がいくつか提案されている．Fadaee ら (Fadaee et al. 2017) は，対訳文中の低頻度語を別の単語に置換して得られた文を学習データに加えることで翻訳性能が向上することを示した．Sennrich ら (Sennrich et al. 2016a) は，目的言語の単言語コーパスを機械翻訳によって原言語へ逆翻訳することで擬似的な対訳文を生成し，対訳コーパスと混合して訓練する方法を提案している．Currey ら (Currey et al. 2017) は，目的言語の単言語コーパスの文をそのままコピーして原言語側のデータとして用いるだけでも翻訳精度が向上することを示した．

本章では，対訳コーパスの原言語文と目的言語文の双方を利用して，コーパスを拡張する方法を提案する．本手法では，既存の対訳文のうち比較的長い文（読点などを含む文）を対訳部分文に分割してデータ拡張に利用する．NMT では文が長くなると翻訳精度が低下する．とくに文字レベルの学習データは単語レベルのものよりさらに長くなるため，訳抜けや訳語の重複が発生しやすくなる．本研究では，読点などの記号を含む対訳文対（長い文が多い）をその記号位置で分割し，各セグメントに含まれる単語のアラインメント情報と漢字共有率を利用して，長い対訳文から対訳部分文を作成し，目的言語側の部分文を NMT で逆翻訳して原言語部分文を得た後，元の原言語文の一部を逆翻訳結果の部分文と入れ替えて擬似的な原言語文を生成して対訳データを拡張する方法を提案する．また，日中両言語の文には漢字が多く含まれ，字体は異なるものの同じ意味で使われることが多いため，セグメント間の漢字の共有率を単語アラインメント情報の補正に利用する．

NMT モデルとして Luong ら (Luong et al. 2015) のものを，実験用データとしてアジア学術論文抜粋コーパス（ASPEC-JC）を用いて評価実験を行った結果，単純に対訳データをコピーして水増しした場合よりも高い BLUE スコアが得られた．

5.2　関連研究

これまでに対訳コーパスを拡張するためのいくつかの方法が提案された．

対訳コーパスは，目的言語の単言語データを使用した逆翻訳法により短時間で構築できる (Sennrich et al. 2016a)．Sennrich らは，目的言語側の単言語コーパスを単に複製してソース側のデータとして使用する手法を実現した (Sennrich et al. 2017)．疑似対訳コーパスは，このコピー方法を使用して構築できる．つまり，ターゲット側言語の文が対応するソース側言語の文としてコピーされる (Currey et al. 2017)．翻訳結果が不十分であっても

有益であることが実証された．低頻度単語の data augmentation も効果的な方法であることが証明されている (Fadaee et al. 2017).

逆翻訳の概念は統計機械翻訳にまで遡る，逆翻訳方法は半教師付き学習に使用した (Bojar & Tamchyna 2011)．Gwinnup らは，逆翻訳を繰り返し適用することにより，疑似対訳コーパスを構築する手法を実装した (Gwinnup et al. 2017)．Lample らは，ターゲット側で訓練された言語モデルで疑似データのノイズを除去することにより，生成された逆翻訳データ（疑似データ）を使用した結果を調査した (Lample et al. 2018)．また，高言語資源と低言語資源の両方で反復逆翻訳を行うことで，翻訳のパフォーマンスを改善できた (Poncelas et al. 2018)．逆翻訳のより洗練されたアイデアは，He らの二重学習アプローチである．これは，対訳データの訓練と単言語データの訓練を循環取引で統合する (He et al. 2016).

Park らは疑似データのみで訓練されたモデルを提案した．彼らの研究では，以下の手法で構成された対訳コーパスで NMT モデルを訓練した．（1）ソース側のみの疑似データ，（2）ターゲット側のみの疑似データ，（3）ソース側またはターゲット側のいずれかが合成された対訳文の混合 (Park et al. 2017).

Karakanta らは逆翻訳データを使用して，低言語資源を持つ NMT を改善した．彼らは高言語資源と低言語資源の類似性を利用した．文字変換を使用して，高言語資源のデータを低言語資源に似たデータに変換する．翻訳モデルは，Wikipedia の記事タイトルから抽出された高言語資源ペアとして訓練された．次に，自動的に逆翻訳された単言語の低言語資源データを，文字変換された高言語資源データから訓練されたモデルで変換し，結果として対訳コーパスを使用して翻訳モデルを最終に訓練した (Karakanta et al. 2018).

5.3 NMT システム

本研究では第 4 章と同じく Luong ら (Luong et al. 2015) によるグローバル注意機構付きエンコーダ・デコーダモデルを実装した NMT システムを文字レベルで使用する．エンコーダは，双方向 LSTM リカレントニューラルネットワークであり，入力系列 $\boldsymbol{x} = (x_1, \cdots, x_m)$ を読み取って，順方向の隠れ状態列 $(\overrightarrow{h_1}, \cdots, \overrightarrow{h_m})$ と逆方向の隠れ状態列 $(\overleftarrow{h_1}, \cdots, \overleftarrow{h_m})$ を求める．隠れ状態 $\overrightarrow{h_j}$ と $\overleftarrow{h_j}$ は連結され，アノテーションベクトルが作られる．

デコーダは，目的言語文 $\boldsymbol{y} = (y_1, \cdots, y_n)$ を予測する LSTM リカレントニューラルネットワークである．各単語（文字レベルの場合，各文字）y_i は，リカレント隠れ状態 s_i と，前回予測された単語（または文字）y_{i-1}，文脈ベクトル c_i を元に予測される．文脈ベクトル c_i は，アノテーション h_j の加重和として計算される．各 h_j の重みは，y_i と x_j のアラインメントについての情報を表すモデル α_{ij} を通じて決められる．

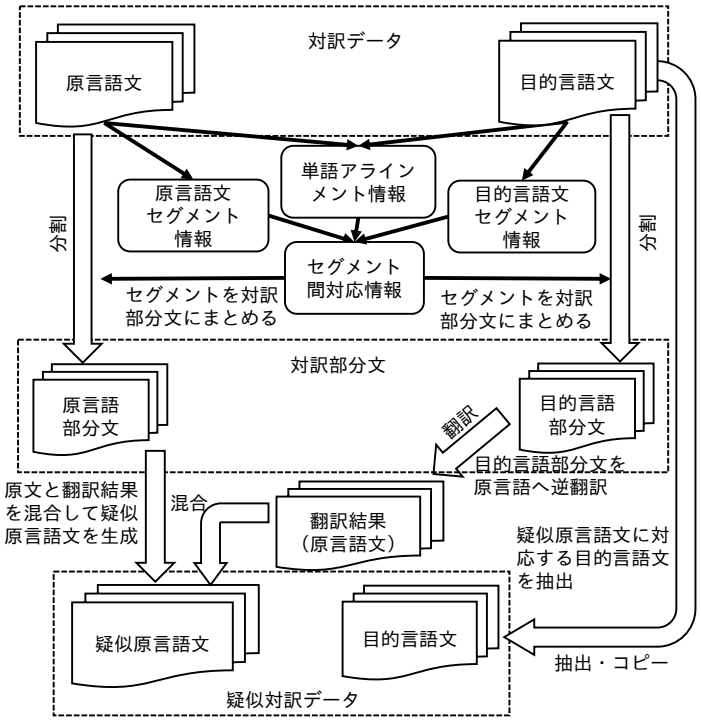

図 5.1　長文分割によるコーパスの拡張の流れ

エンコーダの順方向状態は以下のように表される．

$$\vec{h}_j = \tanh(\vec{W}Ex_j + \vec{U}\vec{h}_{j-1}) \tag{5.1}$$

ここで，$E \in \mathbb{R}^{p \times V_x}$ は単語埋め込み行列であり，$W \in \mathbb{R}^{q \times p}$ と $U \in \mathbb{R}^{q \times q}$ は重み行列である．p，q，V_x はそれぞれ，単語ベクトルのサイズ，隠れユニットの数，原言語の語彙サイズである．

NMT システムの実装としては OpenNMT を用いた (Klein et al. 2017).

5.4　長文の分割によるコーパスの拡張

Sennrich ら (Sennrich & Haddow 2016) は既存の対訳データとは別に，目的言語の単言語コーパスを用意し，それを原言語へ逆翻訳することで対訳データを増やす方法を提案した．本研究で提案するコーパス拡張の処理は図 5.1 に示すように，(1) 既存の対訳データ中の読点等を含む比較的長い文から対訳部分文を生成するフェーズと，(2) 目的言語側の部分文の逆翻訳結果と原言語側の部分文を組み合わせて，一つの対訳文対から複数の擬似対訳データを生成するフェーズからなる．

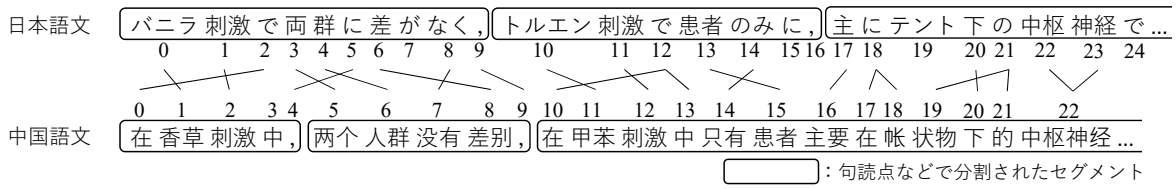

図 5.2　文のセグメント分割と単語アラインメント情報の例

5.4.1　対訳部分文の生成

以下の手順により，学習データに含まれる長い対訳文（読点などを含む対訳文）を適切に分割して対訳部分文を生成する（図 5.1 前半部分）.

1. 対訳文を単語に分割した後，単語アラインメント情報を取得する．後述の実験では，日本語文と中国語文の単語分割にそれぞれ MeCab[*1] と jieba[*2] を用い，単語アラインメント情報の取得には fast_align[*3] を用いた.
2. 対訳文を「,」「；」「、」「：」などの区切り記号の位置で複数のセグメントに分割する.
3. 次のようにして原言語セグメントから目的言語セグメント間への対応関係を求める．原言語側の各セグメント s-seg$_i$ と目的言語側の各セグメント t-seg$_j$ に対して，s-seg$_i$ 内の単語のうち t-seg$_j$ 内の単語に対応するものの数を，単語アラインメント情報をもとにカウントする．例えば図 5.2 の例では，原言語（日本語）文の最初のセグメント s-seg$_0$ に含まれる単語のうち，4 つが目的言語（中国語）文の最初のセグメント t-seg$_0$ 内の単語と対応しており，残りの 5 つが目的言語文の 2 番目のセグメント t-seg$_1$ 内の単語と対応している．この場合，図 5.3 に示すように s-seg$_0$ から t-seg$_0$ への対応の割合は $4/9 \approx 0.44$，s-seg$_0$ のから t-seg$_1$ への対応の割合は $5/9 \approx 0.56$ となる．s-seg$_i$ のから t-seg$_j$ への対応の割合が閾値 θ_1 以上のとき，s-seg$_i$ のから t-seg$_j$ への対応関係があると判断する.
4. 同様にして，目的言語セグメントから原言語セグメントへの対応関係を求める.
5. セグメント間の対応関係が 1 対多または多対多となる場合には，1 対 1 になるように複数側のセグメントを 1 つにまとめる．図 5.3 の例では，原言語文，目的言語文ともに 3 つのセグメントに分割され，2 つの対訳部分文が生成される.

[*1] http://taku910.github.io/mecab/

[*2] https://github.com/fxsjy/jieba（jieba では全角英単語が文字ごとに分割されてしまうため，分割を防ぐ処理を加えた.）

[*3] https://github.com/clab/fast_align

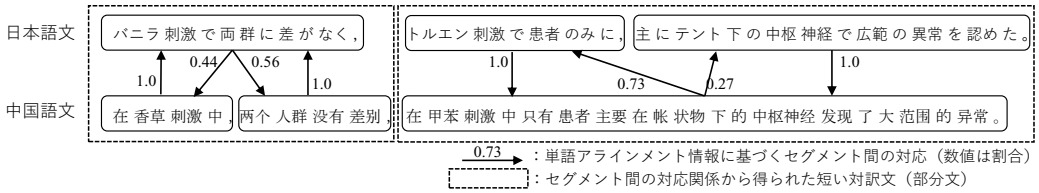

図 5.3　セグメント間の対応関係と対訳部分文の生成

漢字共有率によるセグメント対応情報の補正

上述の手順で生成される対訳部分文の分割位置の評価を，ASPEC-JC の dev データ（2,090 文対）を用いて人手で行ったところ，生成された 4,381 の対訳部分文対のうち，その多くは適切に分割されていたが，誤りも見られた．誤りの原因としては，両言語の基本語順の違い（日本語は S-O-V，中国語は S-V-O）のほか，単語アラインメントの誤りがあげらる．その単語アラインメントの誤りには，日本語と中国語の漢字の対応を考慮すれば防げるものが含まれていた．そこで，対訳部分文の生成ステップ 3 および 4 で得られるセグメント間の対応の割合を，セグメント内の漢字の共有率を元に補正することを考えた．以下，その補正方法について述べる．

漢字の字体には，日本の新字体・旧字体，中国本土で使われる簡体字，台湾や香港で使われる繁体字などがある．Unicode の CJK 統合漢字では，各国の類似した漢字に同じコードが割り当てられており，例えば，日本の「写」と簡体字の「写」は U+5199 に統一されている．一方，由来が同じでも形が異なる「見」と「见」，「発」と「发」，「広」と「广」などには異なるコードが割り当てられており，コードからは元来同じ漢字かどうか判断できない．

Chu ら (Chu et al. 2011, 2012; 中澤ほか 2012) は，日本の漢字（JIS 第 1，第 2 水準漢字）と簡体字，繁体字との対応表を作成し，それを用いて日中両言語の文書中の漢字の共通化を行うことで，単語アラインメントの精度を向上させ，用例ベース機械翻訳システムの翻訳精度向上に利用できることを示した．本研究では，Chu らの漢字対応表を用いて，対訳文対における日本の漢字を簡体字に置き換えることにより，セグメント間の漢字の共有率を求め，セグメント間の対応情報の補正に利用する．

日中の一対のセグメントにおいて，日本語セグメントに含まれる漢字の数を n_j，中国語セグメントに含まれる漢字の数を n_c，両方のセグメントに共通に現れる漢字の数を n_s としたとき，このセグメント対の漢字共有率 σ を次のように定義する．

$$\sigma = \frac{2n_s}{n_j + n_c} \tag{5.2}$$

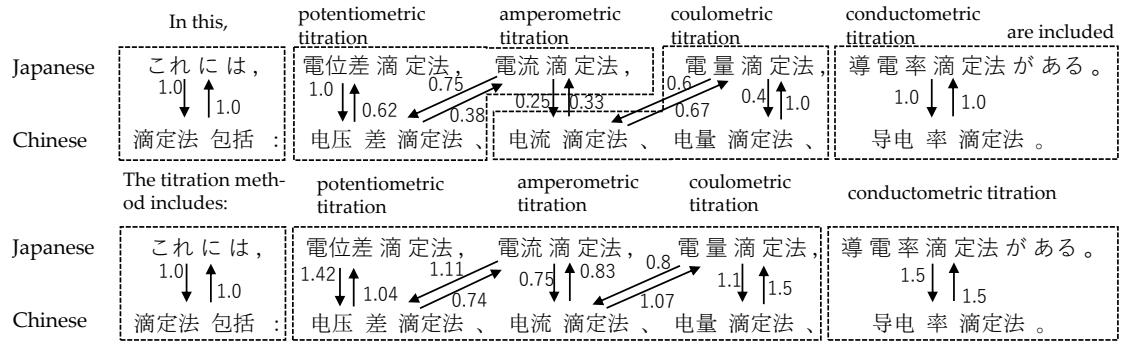

図 5.4　漢字共有率によるセグメント対応情報の補正（上：補正前，下：補正後）

漢字共有率 σ を用いて，セグメント間の対応割合 ρ の値を次のように補正する．

$$\rho' = \begin{cases} \rho + \sigma \cdot w & (\sigma \geq \theta_2) \\ \rho & (\sigma < \theta_2) \end{cases} \tag{5.3}$$

ここで，θ_2 と w はそれぞれ閾値と重みを表す（後述の実験により値を決める）．なお，補正後の値 ρ' は割合にはなっていない．

図 5.4 に補正の例を示す．この例では，「電流滴定法,」を含む部分文の対応が修正されている．

5.4.2　対訳データの拡張

生成した対訳部分文を用いて，擬似的な対訳文を以下の手順で構成する（図 5.1 後半部分）．なお，対応するセグメントを持たないセグメントが存在する文や，原言語と目的言語でセグメントの順序が異なり，対応が交差する箇所のある文からは，正しい対訳文が得られない可能性が高いと考え，データ拡張には使用しない．

1. 生成した目的言語の部分文を NMT により原言語に逆翻訳する．逆翻訳のためのモデルは，拡張前の対訳データにより構築したものを使用する．
2. 対訳データの各原言語文の一部を，逆翻訳によって得られた部分文で置き換えて，元の原言語文と一部異なる擬似原言語文を作る．これにより文の分割数と同じ数の擬似原言語文のバリエーションが生成できる．図 5.3 の日本語文から生成された擬似原言語文を表 5.1 に示す．
3. 作成された擬似原言語文に対応する目的言語文を用意（コピー）して，擬似的な対訳文とする．

以上の手順で生成した疑似対訳文対を元の対訳データに加えることでコーパスを拡張する．

表 5.1　生成された擬似原言語文の例（//は分割点）

原文	バニラ刺激で両群に差がなく，//トルエン刺激で患者のみに，主にテント下の中枢神経で広範の異常を認めた．
擬似原言語文 1	香草刺激では，両群に差はなかった // トルエン刺激で患者のみに，主にテント下の中枢神経で広範の異常を認めた．
擬似原言語文 2	バニラ刺激で両群に差がなく，//トルエン刺激には主に帳票物下の中枢神経で広範囲の異常が認められた．

5.4.3　部分文に分割されない文の利用

読点などの記号を含むものの，5.4.2 節の対訳部分文生成過程で，対訳部分文対に分割できない文対が存在する．漢字共有率を利用したセグメント間の対応関係の補正によっても（セグメントが併合され），分割できる文対の数は減少する．

対訳部分文対に分割できず，データ拡張に利用できない文対を対象に，その目的言語文のみを使って以下のように対訳文対の生成を試みる．

1. 部分文対に分割できなかった文対のうち，目的言語文が複数セグメントからなる文を抽出する．
2. 抽出した目的言語文を原言語に逆翻訳する．
3. 目的言語文とその逆翻訳結果の文を，「,」「；」「、」「：」などの区切り記号の位置で複数のセグメントに分割する．
4. 目的言語文とその逆翻訳結果をそれぞれ t および $\bar{t}$ とし，t と $\bar{t}$ のセグメント数をそれぞれ n および m とする．また, $t = (s_1, s_2, \ldots, s_n), \bar{t} = (s'_1, s'_2, \ldots, s'_m)$ とする．このとき，$(n = m)$ かつ $(n \geq 2)$ となる文対 $(t, \bar{t})$ を抽出する．
5. t の各セグメント s_i の逆翻訳結果 $\overline{s_i}$ を取得する．
6. $\bar{t}$ のセグメント s_i（$1 \leq i \leq n$）のうちの一つだけを $\overline{s_i}$ で置き換えて，n とおりの擬似原言語文を生成する．
7. 得られた n とおりの擬似原言語文のそれぞれと，目的言語文 t を組みにして，n 個の擬似対訳文対を作る．

5.5　翻訳実験

提案したデータ拡張方法の評価のため，条件を変えて翻訳実験を行った．

5.5.1　実験方法

翻訳システムの実装には OpenNMT を用いた．モデルのパラメータは (− 0.1, 0.1) の範囲の一様乱数で初期化し，最適化にはデフォルトの確率的勾配降下法を用いた．学習率はエポック 6 までは 1.0 とし，それ以降はエポックごとに 0.5 倍する．最大勾配ノルムは 1，最大バッチサイズは 100 とした．また，LSTM リカレント層は 1 層で，単語ベクトルと隠れ層の次元は 512 とした．dropout 確率は 0.5 に設定し，デコード時のビームサイズは 5 とした．文の最大長は，デフォルトでは 250 だが，文字レベルでは長くなるため 500 に設定した．

学習と翻訳は文字レベルで行うが，評価は単語レベルのシステムと同じ条件で行うため，日本語文は MeCab，中国語文は jieba で単語ごとに分割した後，OpenNMT 付属の multi-bleu.perl で BLEU スコアを算出した．

多くの場合，エポック 10 前後で validation perplexity（dev データでの perplexity）が下げ止まった．その時点からエポック 16 までの BLEU スコアの平均を評価値とした．ベースラインは何も加工しない訓練データによる文字レベルの翻訳である．

実験には，ASPEC（Asian Scientific Paper Excerpt Corpus）(Nakazawa et al. 2016) の日中学術論文抜粋コーパスを利用し，ASPEC-JC の train データからランダムに抽出した 30 万文対を訓練データとして使用した．

5.5.2　閾値 θ_1，θ_2 と重み w の選択

ここでは，5.4 節で述べた閾値 θ_1，θ_2 と重み w の選択について述べる．

閾値 θ_1（5.4.1 節のセグメント間のアラインメントの閾値）を決定するために，θ_1 の変化による対訳文 30 万文から生成される対訳部分文の数の変化を実験により調べた．その結果，図 5.5 に示すように，θ_1 が 0.5 のとき，生成される対訳部分文の数は最大となった．

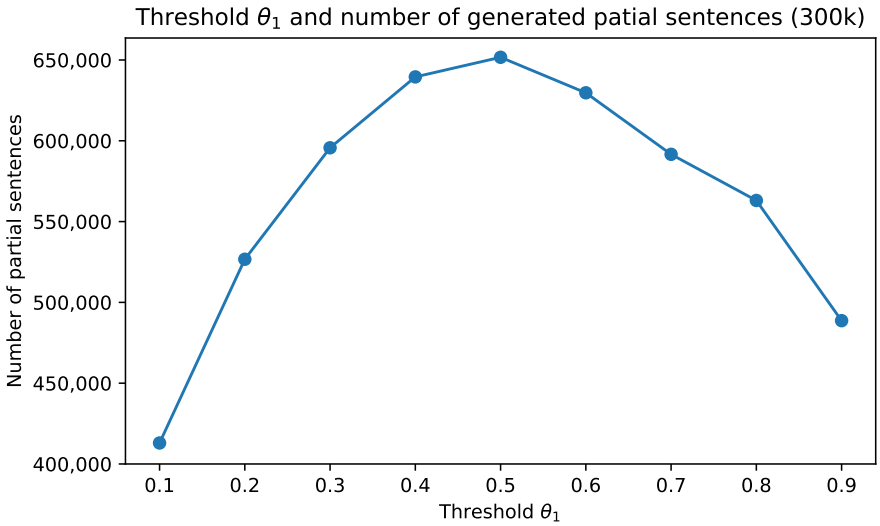

図 5.5　閾値 θ_1 と生成された部分文の数 (30 万文).

式 5.3 の閾値 θ_2 と重み w を決定するため，θ_2 と w をそれぞれ 0.3，0.5，0.7 に設定して対訳文 30 万文から対訳部分文を生成した．それらの中から 9000（=500×18）の対訳部分文をランダムに抽出して，それらが実際に対訳になっているかどうかを人手で評価した．その結果，誤り率（対訳になっていない誤った対訳部分文の割合）は表 5.2 のようになった．表中の「cc」は，漢字共有率を考慮した場合，「cc なし」は考慮しない場合の誤り率を表している．閾値 θ_2，重み w ともに 0.5 のとき，誤り率は最小となった．また，漢字共有率を考慮することで誤り率を低減させられることが確認できた．

表 5.2　異なる閾値 θ_2 および重み w を使用し，30 万文から生成された対訳部分文の対応のエラー率．「cc なし」は，漢字共有率による補正方法を使用しないことを示す．「cc」は，漢字共有率による補正方法を使用することを示す．

アライメントのエラー率 (%)	$\theta_2 = 0.3$		$\theta_2 = 0.5$		$\theta_2 = 0.7$	
	cc なし	cc	cc なし	cc	cc なし	cc
$w = 0.3$	8.8	6.8	7.6	5.6	9.2	9.0
$w = 0.5$	2.1	1.7	1.7	0.8	3.3	2.2
$w = 0.7$	7.0	3.0	6.0	3.8	7.6	7.2

以上の結果より，以下の実験では θ_1，θ_2，w のいずれも 0.5 に設定した．

5.5.3　長文分解による学習データの拡張

実験 1

翻訳実験に使用した ASPEC-JC コーパスの対訳文数を表 5.3 に示す．実験では，ASPEC-JC コーパスの TRAIN データ約 67 万文から 30 万文をランダムに抽出して訓練データとして使用した．また，テストデータとして TEST および DEVTEST データを使用し，検証データとして DEV データ（いずれも約 2 千文）を使用した．

表 5.3　ASPEC-JC コーパスの対訳文対数

Data Type	**File Name**	**Number of sentences**
TRAIN	train.txt	672,315
DEV	dev.txt	2,090
DEVTEST	devtest.txt	2,148
TEST	test.txt	2,107

TEST データの翻訳結果の BLEU 値と TER 値のエポックごとの変化をそれぞれ図 5.6 と図 5.7 に示す．また，検証データの perplexity 値（表中の ppl）が下げ止まった時点での perplexity 値，その時点でのモデルを用いてテストデータ（TEST と DEVTEST）を翻訳した際の BLEU 値および TER 値を表 5.4 および表 5.5 に示す．

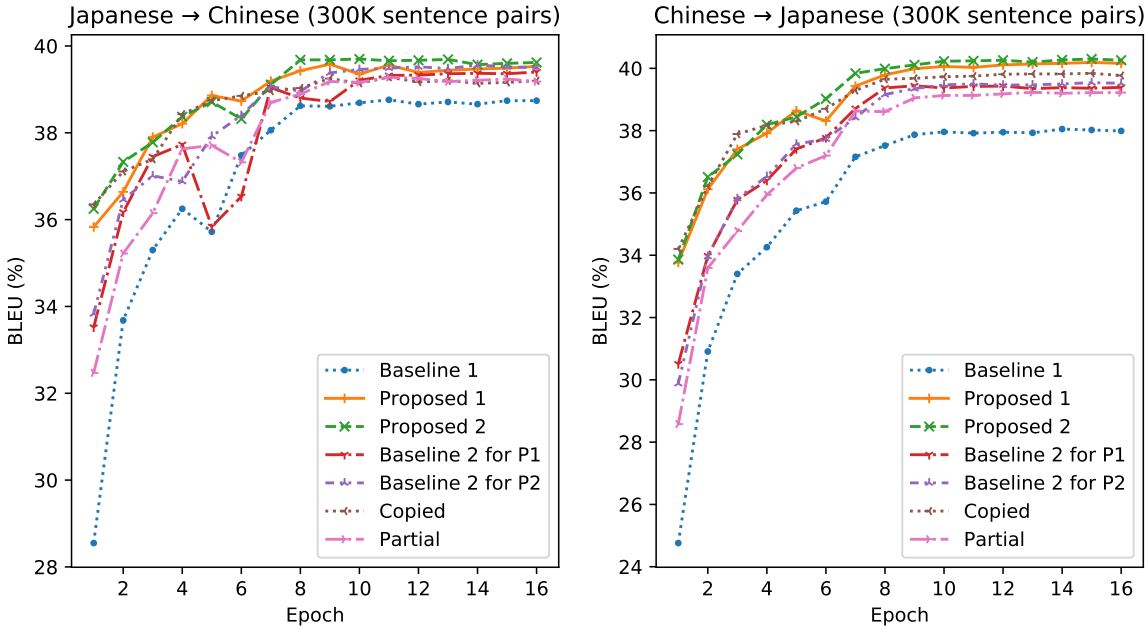

図 5.6　テストデータでの BLEU スコアの変化 (30 万文の訓練データ).「P1」は提案手法 Proposed 1 を示し,「P2」は提案手法 Proposed 2 を示す.

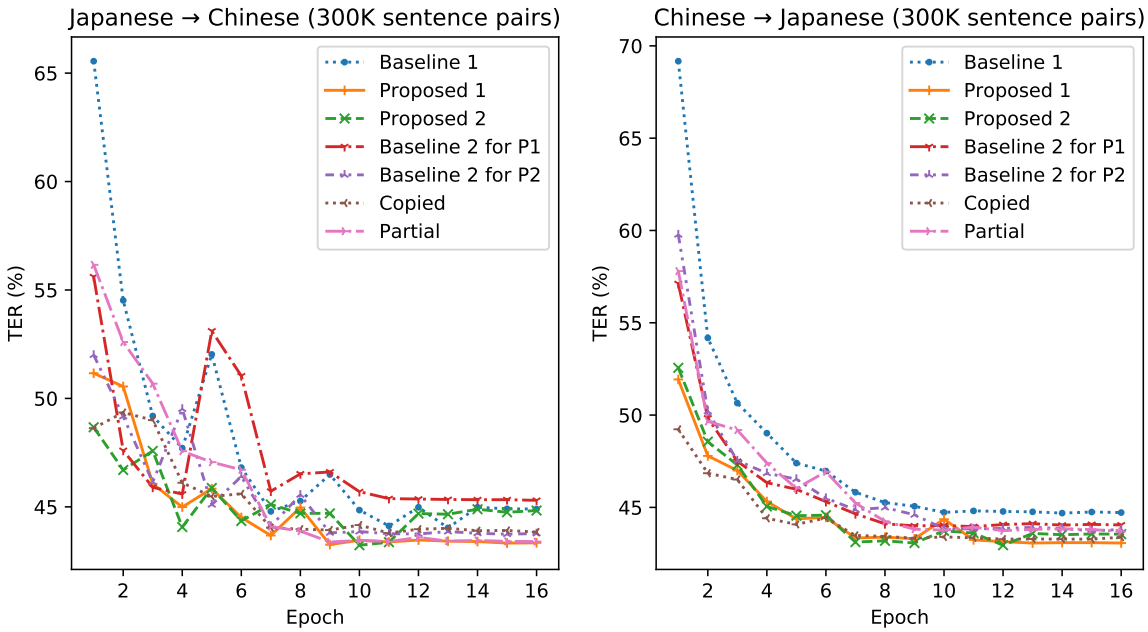

図 5.7　テストデータでの TER スコアの変化 (30 万文の訓練データ).「P1」は提案手法 Proposed 1 を示し,「P2」は提案手法 Proposed 2 を示す.

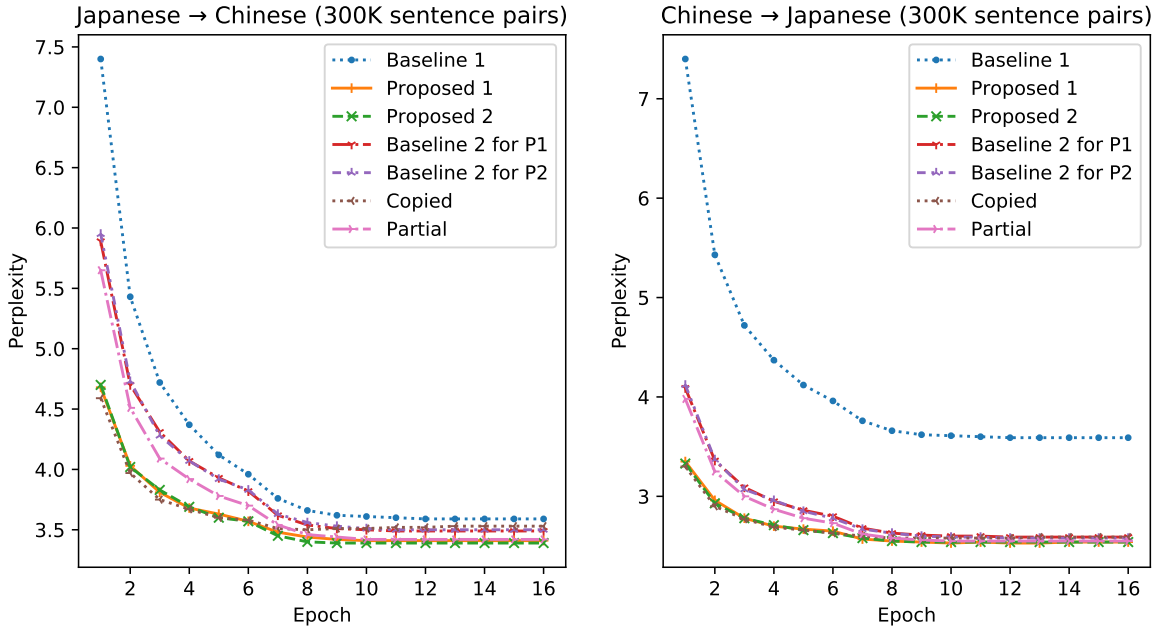

図 5.8　開発データ（dev）での perplexity 値の変化 (30 万文の訓練データ).「P1」は提案手法 Proposed 1 を示し，「P2」は提案手法 Proposed 2 を示す.

開発（dev）データで止まった後の最低の perplexity 値，および各手法によるテストデータおよび開発テストデータでの翻訳結果の BLUE および TER スコアを，表 5.4 と 5.5 に示す.

表 5.4　30 万文の訓練データを使用した日中 NMT の実験結果.「ppl」は perplexity を示す.「Dev」は開発データを示す.「Dev-test」は、開発テストデータを示す.

Method	**Japanese→Chinese**							
	# Sentences		**# BT**	**ppl**	**BLEU (%)**		**TER (%)**	
	Raw	**Used**		**Dev**	**Dev-Test**	**Test**	**Dev-Test**	**Test**
Baseline 1	300k	300k	0	3.6	38.5	38.7	44.0	44.8
Baseline 2 for P1	518k	518k	218k	3.5	39.1	39.4	43.9	45.3
Baseline 2 for P2	531k	531k	231k	3.5	39.2	39.5	43.0	43.8
Copied	977k	977k	0	3.5	38.9	39.2	43.7	43.9
Partial	984k	984k	0	3.5	38.9	39.2	43.0	43.4
Proposed 1	952k	923k	218k	3.4	39.2	39.5	44.2	43.4
Proposed 2	977k	945k	231k	3.4	39.3	39.7	42.9	43.4

表 5.5　30 万文の訓練データを使用した中日 NMT の実験結果．「ppl」は perplexity を示す．「Dev」は開発データを示す．「Dev-test」は、開発テストデータを示す．

Method	Chinese→Japanese							
	# Sentences		# BT	ppl	BLEU (%)		TER (%)	
	Raw	Used		Dev	Dev-Test	Test	Dev-Test	Test
Baseline 1	300k	300k	0	2.7	38.1	38.0	44.9	44.8
Baseline 2 for P1	518k	518k	218k	2.6	40.0	39.4	44.5	44.1
Baseline 2 for P2	529k	529k	229k	2.6	40.1	39.5	43.6	43.8
Copied	972k	972k	0	2.6	39.7	39.8	43.5	43.3
Partial	984k	984k	0	2.6	39.0	39.2	43.9	43.8
Proposed 1	952k	947k	218k	2.5	40.2	40.1	42.9	43.3
Proposed 2	972k	967k	229k	2.5	40.5	40.2	43.6	43.4

ここで，Baseline 1 は訓練データ 30 万文を拡張せずにそのまま文字レベルで訓練した結果である．逆翻訳にはこの訓練で構築された翻訳モデルを使用した．

Proposed 1（P1）と Proposed 2（P2）は，本研究の提案手法により 30 万文のデータを拡張したコーパスで訓練した結果を表す．このうち Proposed 1 では，漢字共有率によるセグメント間のアラインメントの補正処理と 5.4.3 節の処理をせず，Proposed 2 ではそれらの処理を実施した．

Baseline 2 は従来の（長文分割を行わない）逆翻訳手法である．Baseline 2 for P1 と Baseline2 for P2 は，それぞれ Proposed 1 と Proposed 2 で逆翻訳対象となったのと同じ文を逆翻訳してデータ拡張を行ったものである．従来の逆翻訳手法では，1 つの目的言語文から 1 つの擬似対訳文対しか生成されないため，生成された擬似対訳文対を複製して，提案手法と同数になるようにした．

Copied は逆翻訳を用いず，訓練データをコピーして増やす手法である．Proposed 2 の場合と同じ文を同じ回数コピーして追加した．NMT では学習不足の場合，訓練データをコピーして増やすだけでも翻訳精度が向上するため，それとの比較を行った．

Partial は，5.4.1 節の手順で生成された対訳部分文をそのまま訓練データに加えてコーパスを拡張する手法である．5.4.2 節の処理（目的言語部分文の逆翻訳と原言語部分文との混合）の必要性を確認するために加えた．

表中の # Sentences のうち，Raw は各手法によって得られた訓練用対訳文対の数，Used は得られた対訳文対のうち訓練に用いられた対訳文対の数を表す．提案手法では Raw と Used の数が異なっているが，これは，目的言語部分文の逆翻訳時の誤訳により，同じ単語や句が繰り返し出現して，訓練データの文の長さの上限を超えたために利用されなかった

文が存在するためである．また，# BT は，各手法で逆翻訳の対象となった文の数である．

実験の結果，提案手法 Proposed 1 および Proposed 2 は，テストデータ TEST と DEVTEST の双方に対して，日-中，中-日いずれの方向に対してもほとんどの場合，Baseline 1，Baseline 2，Copied，Partial のいずれよりもよい BLEU 値と TER 値を得ることができた（BLEU 値は大きいほどよく，TER 値は小さいほどよい）．

提案手法によって生成された擬似原言語文には，翻訳誤りや不自然な表現が含まれていたが，元の（正しい）対訳文対をコピーして増やす Copied よりも良い結果が得られている．また，生成した対訳部分文をそのまま訓練データに加える手法 Partial よりも提案手法の方が良い結果が得られたことから，対訳部分文から擬似対訳文を生成する過程（5.4.2 節のステップ 2）の必要性・有効性が示された．

実験 2

従来の逆翻訳手法では，目的言語の単言語データを逆翻訳して擬似原言語文を生成し，それを元の目的言語文と対にすることで対訳データを生成する．それに対して，本研究の提案手法は，追加の単言語コーパスを使用せず，与えられた対訳コーパスだけを元にコーパスを拡張する．したがって，他のコーパス拡張手法によって拡張された対訳データに対して，さらに本手法を適用することができる．そこで，Sennrich らの逆翻訳手法 (Sennrich et al. 2016a) で拡張された対訳データを，本研究の提案手法でさらに増やすことにより，翻訳精度が向上するか実験を行った．

ASPEC-JC コーパスの TRAIN データ約 67 万文を，15 万文（拡張対象となる対訳データ）＋残りの約 52 万文（追加の単言語データ），および，30 万文（拡張対象となる対訳データ）＋残りの約 37 万文（追加の単言語データ）にランダムに分割してこの実験に使用した．また，実験 1 の結果から，提案手法 Proposed 2 はほとんどの場合で Proposed 1 より優れていたため，この実験では Proposed 2 のみを実験対象とした．

テストデータとして，TEST データと DEVTEST データを使用したときの実験結果を表 5.6~ 表 5.9 に示す．

表 5.6 および表 5.7 は，対訳データ 15 万文，残りの約 52 万文を追加の単言語データとして行った実験の結果である．ここで，Baseline 1 はコーパス拡張を行わない 15 万文の対訳データに対する文字レベル翻訳である．逆翻訳にはこの訓練で得られたモデルを使用した．

“Baseline 2+mono (522k)” は，追加の単言語コーパス約 52 万文を用いた Sennrich らの逆翻訳手法である．元の 15 万文の対訳データに，約 52 万文の擬似対訳データが追加され，対訳データは約 67 万文に拡張された．

“150k+mono (522k)+P2” は，“Baseline 2+mono (522k)” で得られた約 67 万文に対して提案手法 Proposed 2 を適用したものである．これにより対訳データは 231 万文（日-

中）及び 224 万文（中-日）に拡張された．翻訳精度については，BLEU 値，TER 値ともに改善され，とくに中日翻訳のときに BLEU 値が 1.0〜1.3 向上し，TER 値は 0.9〜1.4 向上しており，拡張されたデータに対して提案手法を適用することの有効性が確認できた．これに対して，日中翻訳では BLEU 値の向上は 0.0〜0.1，TER 値の向上は 0.3〜0.7 で，中日翻訳と比較すると効果は小さかった．

表 5.8 および表 5.9 は，対訳データ 30 万文，残りの約 37 万文を追加の単言語コーパスとして行った実験の結果である．中日翻訳では BLEU 値が 1.3〜1.5 向上し，TER 値は 0.8〜1.1 向上しており，提案手法の効果が確認できた．一方，日中翻訳でも翻訳精度の改善が見られたが，BLEU 値で 0.2〜0.4，TER 値で 0.2〜0.9 と中日翻訳に比べて小幅な向上であった．

日中翻訳では対訳部分文を中国語から日本語に逆翻訳するが，この過程で日中逆翻訳に比べて NMT 特有の誤訳（同じ語や句が繰り返し現れる）が目立って発生しており，それが中日翻訳での効果が低かった要因の一つと考えられる．

この実験から，他のデータ拡張手法と組み合わせて使用してさらに対訳データを増やすことができ，翻訳精度の向上も見込めることがわかった．

表 5.6　15 万文の訓練データと約 52 万文の単言語データを使用した日中 NMT の実験結果．「ppl」は perplexity を示す．「Dev」は開発データを示す．「Dev-test」は、開発テストデータを示す．

Method	**Japanese→Chinese**							
	# Sentences		**# BT**	**ppl**	**BLEU (%)**		**TER (%)**	
	Raw	**Used**		**Dev**	**Dev-Test**	**Test**	**Dev-Test**	**Test**
Baseline 1	150k	150k	0	4.3	36.5	36.5	48.5	50.1
Baseline 2 + mono (522k)	672k	672k	522k	3.8	38.8	39.1	44.6	44.7
150k + mono (522k) + P2	2313k	2201k	525k	3.7	38.9	39.1	43.9	44.4

表 5.7　15 万文の訓練データと約 52 万文の単言語データを使用した中日 NMT の実験結果．「ppl」は perplexity を示す．「Dev」は開発データを示す．「Dev-test」は、開発テストデータを示す．

Method	**Chinese→Japanese**							
	# Sentences		**# BT**	**ppl**	**BLEU (%)**		**TER (%)**	
	Raw	**Used**		**Dev**	**Dev-Test**	**Test**	**Dev-Test**	**Test**
Baseline 1	150k	150k	0	3.1	35.4	35.5	48.4	47.5
Baseline 2 + mono (522k)	672k	672k	522k	2.8	39.3	39.1	45.1	44.6
150k + mono (522k) + P2	2239k	2134k	515k	2.7	40.6	40.1	43.7	43.7

表 5.8 30 万文の訓練データと約 37 万文の単言語データを使用した日中 NMT の実験結果.「ppl」は perplexity を示す.「Dev」は開発データを示す.「Dev-test」は、開発テストデータを示す.

Method	Japanese→Chinese							
	# Sentences		# BT	ppl	BLEU (%)		TER (%)	
	Raw	Used		Dev	Dev-Test	Test	Dev-Test	Test
Baseline 1	300k	300k	0	3.6	38.5	38.7	44.0	44.8
Baseline 2 + mono (372k)	672k	672k	372k	3.4	39.6	39.7	42.8	44.2
300k + mono (372k) + P2	2287k	2234k	522k	3.4	39.8	40.1	42.6	43.3

表 5.9 30 万文の訓練データと約 37 万文の単言語データを使用した中日 NMT の実験結果.「ppl」は perplexity を示す.「Dev」は開発データを示す.「Dev-test」は、開発テストデータを示す.

Method	Chinese→Japanese							
	# Sentences		# BT	ppl	BLEU (%)		TER (%)	
	Raw	Used		Dev	Dev-Test	Test	Dev-Test	Test
Baseline 1	300k	300k	0	2.7	38.1	38.0	44.9	44.8
Baseline 2 + mono (372k)	672k	672k	372k	2.6	40.5	39.9	43.3	43.3
300k + mono (372k) + P2	2223k	2213k	501k	2.5	41.8	41.4	42.2	42.5

5.6 おわりに

本章では，対訳コーパス中の読点等を含む比較的長い文対を，単語アラインメント情報を利用して対訳部分文対に分割し，逆翻訳した目的言語部分文と原言語部分文を組み合わせることで対訳データを拡張する手法を提案した．また，対訳部分文への分割の際，日中両言語に共通して現わる漢字の割合を元に，単語アラインメント情報を補正する手法も提案した．ASPEC-JC コーパスを使用した翻訳実験の結果，逆翻訳手法と同等以上の翻訳性能が得られた．また，ベースラインシステムに対する改善も報告した．中国語から他の言語へ，あるいは，日本語から他の言語への文字レベルの翻訳においても，翻訳精度の向上に役立つ可能性があると考えられる．

将来的には，提案方法を他の拡張手法と組み合わせることを計画している．これは，ただの逆翻訳よりも有益である可能性が示唆されている．一方、生成された各疑似文には疑似部分文が 1 つしかないため，疑似部分文のさまざまな組み合わせを考慮して，より多くの疑似文を生成する必要がある．さらに，他の拡張方法と組み合わせることも調査したい．

第 6 章

結言

6.1　研究結果の概要

機械翻訳の精度はニューラル機械翻訳（NMT）の登場によって，従来の統計機械翻訳（SMT）より大きく向上した．しかし，ハードウェアの制限のため，またはモデルの訓練時間が膨大になるのを防ぐため，単語レベルの NMT では一般に語彙サイズを一定の範囲に制限するが，それにより未知語が増え，翻訳精度が低下するという問題がある．表語文字である漢字を共有している日中間の NMT では，文字レベルで訓練・翻訳を行うことでその問題が回避できる．また，NMT の問題点として，良い翻訳結果を得るには，SMT 以上に大規模な対訳コーパスを必要とし，そのような大量の対訳コーパスが存在する言語対は限られることがあげられる．日中対訳コーパスも，日英，英中の対訳コーパスに比べるとその規模は小さく，それが翻訳結果にも影響を与える．

本研究では，文字レベルの日中両言語間の NMT の翻訳精度をさらに改善できないかと考え，文字の特徴情報の一つとして漢字の部首を入力特徴情報に追加した．また，訓練用対訳コーパス不足を補うための対訳コーパス拡張手法を提案し，実験によりそれらの有効性を確認した．

第 2 章では機械翻訳の歴史と現状，日中両言語の比較対照について述べた後，NMT の現状を概観した．

第 3 章では，NMT の原理と研究動向について述べた．

第 4 章では，漢字の部首に着目した日中間の文字レベル NMT に対する改善手法を提案した．文字レベルの NMT では，単語レベルの NMT における語彙サイズの問題を回避できる．とくに表語文字である漢字を共有する日中間の翻訳では，単語やサブワード（単語の部分文字列）単位の NMT よりも文字単位の NMT の方が優れているとされる (Meng et al. 2019)．単語レベルの NMT では単語の品詞情報などを入力特徴情報として加えることで翻訳精度が向上することが知られているが，本章では 文字レベルの NMT においても，

文字に関する何らかの情報を加えることで翻訳精度を改善できる可能性があると考え，漢字の部首（主に漢字の意味カテゴリを表す）の情報を追加する方法を提案した．漢字以外の文字にも便宜的に部首を設定している．ASPEC-JC コーパス（対訳データセット）を用いて翻訳実験を行ったところ，とくに日本語から中国語への翻訳に対して安定して改善の効果が見られた．翻訳結果の評価尺度である BLEU スコアの値は，部首情報を加えない場合と比較して，日中翻訳の場合で 0.4〜0.5，中日翻訳の場合で 0.03〜0.7 向上した．訓練時の perplexity は，日中翻訳では 0.1，中日翻訳では 0.01 改善された．

第 5 章では，NMT における訓練データである対訳コーパス拡張する手法を提案した．NMT では翻訳の品質が対訳データの量に強く依存し，高品質な翻訳結果を得るには大規模な対訳データが必要となる．しかし，日英や中英の対訳コーパスと比較すると，日中の対訳コーパスの量は未だ小さい．本章では単語アラインメント情報を利用して，長い（読点などを含む）対訳文対を，対訳部分文対に分割して，目的言語部分文の逆翻訳結果と原言語部分文を組み合わせることで，既存の対訳コーパスを拡張する方法を提案した．また，日中両言語では漢字を共有していることを利用して，日中のフレーズ内で同じ（由来の）漢字が含まれる割合を元に，単語アラインメント情報を補正する手法も提案した．ASPEC-JC コーパスの訓練用対訳データに対して提案手法を適用した結果，拡張前に比べて対訳コーパスの量は約 3.2 倍に増え，BLEU スコアの値は，日中翻訳の場合で 0.7〜1.2 向上し，中日翻訳の場合で 2.1〜2.2 向上した．Sennrich らの逆翻訳手法と比べても，概ね同等以上の結果が得られた．

6.2　ニューラル機械翻訳の今後の研究方向

現在，ニューラル機械翻訳は大きな成功を収めており，統計機械翻訳後の新しい機械翻訳方法と呼ばれる新しい研究結果が現れている．厳密に言えば，2014 年以降，ニューラル機械翻訳が広く注目され (Sutskever et al. 2014; Cho et al. 2014c,a)，多数の関連する結果が公開されている．研究時間が短いため，翻訳モデルにはさらなる調査に値する多くの問題が残っており，以下の点が今後の研究の焦点になる可能性がある (Koehn & Knowles 2017).

1. 言語の解釈可能性の改善：エンコーダとデコーダに基づくニューラル機械翻訳は原言語から目的言語への直接翻訳が利用可能になるが，適切な言語解釈を得るための翻訳プロセスは困難である．暗黙の構文構造情報は単語レベルのニューラル機械翻訳エンコーダから抽出できることが証明されており (Shi et al. 2016)，ニューラル機械翻訳の翻訳プロセスがある程度説明され分析されている (Ding et al. 2017). ニューラル機械翻訳モデルから対応する言語知識を抽出して，翻訳プロセスを説明

し，翻訳モデルを改善することは，将来のニューラル機械翻訳の重要な研究方向である．

2. 外部の事前知識の追加：構文表記，品詞タグ付け，バイリンガル辞書など，個別のシンボルで表される外部リソースは非常に重要な事前知識ではあるが，ニューラル機械翻訳の翻訳プロセスで完全に活用することは困難である．豊富な事前知識の追加は，ニューラル機械翻訳の重要な研究内容であり，翻訳効果を改善するためさらなる研究が必要である．
3. 構文ベースのニューラル機械翻訳：ニューラル機械翻訳は，ほとんどの場合，構文情報が少ない単語レベルの系列間モデルである．構文は文構造の重要な情報であり，構文木から系列 (Eriguchi et al. 2016)，系列から構文木，構文木から構文木などの翻訳モデルを構文ベースの翻訳モデルに拡張する．これは，ニューラル機械翻訳モデルのアーキテクチャの重要な革新である．
4. 多言語機械翻訳：連続空間表現は効果的な多言語意味表現法であり (Zhang & Zong 2015)，Attention メカニズムは異なる言語間で共有されることが実験的に証明されている (Firat et al. 2016)．これらは多言語機械翻訳研究の優れた基盤を提供する．多言語対訳コーパスに基づくニューラル機械翻訳の研究は，学術的価値があるだけでなく，実用的な価値も高く，将来の重要な開発方向である．
5. マルチモーダル翻訳：ニューラルネットワークは，テキスト，画像，音声などのさまざまなモーダルデータを統一された形式で表すことができる．現在，テキストと画像 (Reed et al. 2016) および画像情報間の End-to-End の直接翻訳 (Calixto & Liu 2017) は，ニューラル機械翻訳にも適用されている．マルチモーダル翻訳を構築するために，音声，画像，ビデオなど，テキストそのもの以外の情報を効率的に使用することが，ニューラル機械翻訳をより実用的なものにする可能性がある．

謝辞

本論文は筆者が岐阜大学大学院工学研究科電子情報システム工学専攻博士後期課程に在籍中の研究成果をまとめたものです．

同専攻准教授松本忠博先生は指導教官として本研究の実施の機会を与えて頂き，研究のための最新設備を購入して頂き，その遂行にあたって終始，熱心なご指導いただき，暖かい激励を賜りました．松本忠博先生は，修士課程からの 5 年間にわたり，公私問わず筆者を支えてくださり，留学生活を安心して過ごすことができました．心より深謝の意を表します．

同専攻教授草刈圭一朗先生，並びに，同専攻教授山口忠先生には主査と副査としてご助言を頂くとともに本論文の細部にわたりご指導を戴いた．ここに感謝の意を表します．

研究を進めてきた松本研メンバーの皆さんよりも，明るい雰囲気で常に勇気づけて頂きました．感謝致します．いつもお世話になりました，精神的にも支えられた工学部グローバル化推進室の川瀬真弓特任助教，留学支援室と学務系の皆さんに，この場を借りて厚く御礼申し上げます．

筆者は中国政府留学基金委員会の奨学金（No.201708050078）の助成を受けました．また，日本国岐阜県国際交流センターの奨学金，日本国文部科学省外国人留学生の奨学金と岐阜大学工学部より国際会議にて発表を行う学生のための奨学金も受けました．ここに感謝の意を表します．

最後に，大なる心配をかけながら渡日して以来，研究者の道を志さんとする筆者に理解を示し，一人で支えてくれた母親に心より深謝致します．

参考文献

Aharoni, R. & Y. Goldberg (2017). "Towards String-To-Tree Neural Machine Translation." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* 132–140. Association for Computational Linguistics.

Aharoni, R., M. Johnson, & O. Firat (2019). "Massively Multilingual Neural Machine Translation." *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)* 3874–3884. Association for Computational Linguistics.

Artetxe, M., G. Labaka, E. Agirre, & K. Cho (2018). "Unsupervised neural machine translation." *Proceedings of the Sixth International Conference on Learning Representations*.

Bahdanau, D., K. Cho, & Y. Bengio (2014). "Neural Machine Translation by Jointly Learning to Align and Translate." *The International Conference on Learning Representations (ICLR)*, 1–15.

Ballesteros, M., C. Dyer, & N. A. Smith (2015). "Improved Transition-Based Parsing by Modeling Characters instead of Words with LSTMs." *Proc. 2015 Conf. on Empirical Methods in Natural Language Processing* 349–359.

Bastings, J., I. Titov, W. Aziz, D. Marcheggiani, & K. Sima'an (2017). "Graph Convolutional Encoders for Syntax-aware Neural Machine Translation." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* 1957–1967. Association for Computational Linguistics.

Bojar, O. & A. Tamchyna (2011). "Improving Translation Model by Monolingual Data." *Proceedings of the Sixth Workshop on Statistical Machine Translation* 330–336. Association for Computational Linguistics.

Brown, P. F., S. D. Pietra, V. J. D. Pietra, & R. L. Mercer (1993). "The Mathematics of Statistical Machine Translation: Parameter Estimation." *Computational Linguistics*Vol. 19 263–311.

Gülçehre,c, O. Firat, K. Xu, K. Cho, L. Barrault, H.-C. Lin, F. Bougares, H. Schwenk, &

Y. Bengio (2015). "On Using Monolingual Corpora in Neural Machine Translation." Vol. abs/1503.03535.

Caglayan, O., W. Aransa, Y. Wang, M. Masana, M. García-Martínez, F. Bougares, L. Barrault, & J. v. d Weijer (2016). "Does Multimodality Help Human and Machine for Translation and Image Captioning?." *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers* 627–633. Association for Computational Linguistics.

Calixto, I. & Q. Liu (2017). "Incorporating Global Visual Features into Attention-based Neural Machine Translation.." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* 992–1003. Association for Computational Linguistics.

Calixto, I., Q. Liu, & N. Campbell (2017). "Doubly-Attentive Decoder for Multi-modal Neural Machine Translation." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1913–1924. Association for Computational Linguistics.

Castaño, M. A. & F. Casacuberta (1997). "A connectionist approach to machine translation." *EUROSPEECH* 91–94.

卜朝暉 (2004). "日中機械翻訳に関する研究～とりたて表現，否定表現の翻訳規則を中心に～." 岐阜大学大学院工学研究科　博士論文.

Chen, H., S. Huang, D. Chiang, & J. Chen (2017). "Improved Neural Machine Translation with a Syntax-Aware Encoder and Decoder." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1936–1945. Association for Computational Linguistics.

Chen, K., R. Wang, M. Utiyama, L. Liu, A. Tamura, E. Sumita, & T. Zhao (2017). "Neural Machine Translation with Source Dependency Representation." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* 2846–2852. Association for Computational Linguistics.

Chen, W., E. Matusov, S. Khadivi, & J.-T. Peter (2016). "Guided Alignment Training for Topic-Aware Neural Machine Translation." Vol. abs/1607.01628.

Chen, X., L. Xu, Z. Liu, M. Sun, & H. Luan (2015). "Joint Learning of Character and Word Embeddings." *Proc. 24th Int. Conf. on Artificial Inteligence (IJCAI'15)* 1236–1242. AAAI Press.

Chen, Y., Y. Liu, Y. Cheng, & V. O. Li (2017). "A Teacher-Student Framework for Zero-Resource Neural Machine Translation." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1925–1935. Association for Computational Linguistics.

Cheng, Y., L. Jiang, & W. Macherey (2019). "Robust Neural Machine Translation with

Doubly Adversarial Inputs." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 4324–4333. Association for Computational Linguistics.

Cheng, Y., W. Xu, Z. He, W. He, H. Wu, M. Sun, & Y. Liu (2016). "Semi-Supervised Learning for Neural Machine Translation." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1965–1974. Association for Computational Linguistics.

Cheng, Y., Q. Yang, Y. Liu, M. Sun, & W. Xu (2017). "Joint Training for Pivot-based Neural Machine Translation." *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17* 3974–3980.

Chinea-Ríos, M., Á. Peris, & F. Casacuberta (2017). "Adapting Neural Machine Translation with Parallel Synthetic Data." *Proceedings of the Second Conference on Machine Translation* 138–147. Association for Computational Linguistics.

Chitnis, R. & J. DeNero (2015). "Variable-Length Word Encodings for Neural Translation Models." *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing* 2088–2093. Association for Computational Linguistics.

Cho, K., B. v Merriënboer, D. Bahdanau, & Y. Bengio (2014a). "On the Properties of Neural Machine Translation: Encoder–Decoder Approaches." *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation* 103–111. Association for Computational Linguistics.

Cho, K., B. v Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, & Y. Bengio (2014b). "Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation." *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* 1724–1734. Association for Computational Linguistics.

Cho, K., B. v Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, & Y. Bengio (2014c). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation." *EMNLP*.

Chu, C., T. Nakazawa, & S. Kurohashi (2012). "Chinese Characters Mapping Table of Japanese , Traditional Chinese and Simplified Chinese." *Proc. 8th Conf. on Int. Language Resources and Evaluation (LREC'12)* 2149–2152.

Chu, C., T. Nakazawa, & S. Kurohasi (2011). "Japanese-Chinese phrase alignment using common Chinese characters information." *Proc. Machine Translation Summit XIII* 475–482.

Crego, J. M., J. Kim, G. Klein, A. Rebollo, K. Yang, J. Senellart, E. Akhanov, P. Brunelle, A. Coquard, Y. Deng, S. Enoue, C. Geiss, J. Johanson, A. Khalsa, R. Khiari, B. Ko,

C. Kobus, J. Lorieux, L. Martins, D. Nguyen, A. Priori, T. Riccardi, N. Segal, C. Servan, C. Tiquet, B. Wang, J. Yang, D. Zhang, J. Zhou, & P. Zoldan (2016). "SYSTRAN's Pure Neural Machine Translation Systems." Vol. abs/1610.05540.

Currey, A., A. V. Miceli Barone, & K. Heafield (2017). "Copied Monolingual Data Improves Low-Resource Neural Machine Translation." *Proceedings of the Second Conference on Machine Translation* 148–156. Association for Computational Linguistics.

方丹 (2013). "日中機械翻訳に関する研究－使役表現及び受身表現の翻訳処理を中心に－." 岐阜大学大学院工学研究科　修士論文.

Delbrouck, J.-B. & S. Dupont (2017). "An empirical study on the effectiveness of images in Multimodal Neural Machine Translation." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* 910–919. Association for Computational Linguistics.

Devlin, J., M.-W. Chang, K. Lee, & K. Toutanova (2019). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *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)* 4171–4186. Association for Computational Linguistics.

Ding, Y., Y. Liu, H. Luan, & M. Sun (2017). "Visualizing and Understanding Neural Machine Translation." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1150–1159. Association for Computational Linguistics.

Domhan, T. (2018). "How Much Attention Do You Need? A Granular Analysis of Neural Machine Translation Architectures." *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1799–1808. Association for Computational Linguistics.

Domhan, T. & F. Hieber (2017). "Using Target-side Monolingual Data for Neural Machine Translation through Multi-task Learning." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* 1500–1505. Association for Computational Linguistics.

Dong, D., H. Wu, W. He, D. Yu, & H. Wang (2015). "Multi-Task Learning for Multiple Language Translation." *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* 1723–1732Beijing, ChinaAssociation for Computational Linguistics.

Santos, C. N.d & V. Guimarães (2015). "Boosting Named Entity Recognition with Neural

Character Embeddings." *Proc. 5th Named Entity Workshop, joint with 53rd ACL and the 7th IJCNLP* 25–33.

Du, J. & A. Way (2017). "Pinyin as Subword Unit for Chinese-Sourced Neural Machine Translation,." *Proceedings of the 25th Irish Conference on Artificial Intelligence and Cognitive Science*.

Eriguchi, A., K. Hashimoto, & Y. Tsuruoka (2016). "Tree-to-Sequence Attentional Neural Machine Translation." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 823–833. Association for Computational Linguistics.

Fadaee, M., A. Bisazza, & C. Monz (2017). "Data Augmentation for Low-Resource Neural Machine Translation." *Proc. 55th Annual Meeting of the Assoc. for Computational Linguistics (Volume 2: Short Papers)* 567–573.

Firat, O., K. Cho, & Y. Bengio (2016). "Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism." *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* 866–875.

Gao, F., J. Zhu, L. Wu, Y. Xia, T. Qin, X. Cheng, W. Zhou, & T.-Y. Liu (2019). "Soft Contextual Data Augmentation for Neural Machine Translation." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 5539–5544. Association for Computational Linguistics.

Gehring, J., M. Auli, D. Grangier, D. Yarats, & Y. N. Dauphin (2017). "Convolutional Sequence to Sequence Learning." *Proceedings of the 34th International Conference on Machine Learning - Volume 70* 1243–1252. JMLR.org.

Gu, S., Y. Feng, & Q. Liu (2019). "Improving Domain Adaptation Translation with Domain Invariant and Specific Information." *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)* 3081–3091. Association for Computational Linguistics.

Gulcehre, C., S. Ahn, R. Nallapati, B. Zhou, & Y. Bengio (2016). "Pointing the Unknown Words." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 140–149Berlin, GermanyAssociation for Computational Linguistics.

Gwinnup, J., T. Anderson, G. Erdmann, K. Young, M. Kazi, E. Salesky, B. Thompson, & J. Taylor (2017). "The AFRL-MITLL WMT17 Systems: Old, New, Borrowed, BLEU." *Proceedings of the Second Conference on Machine Translation* 303–309. Association

for Computational Linguistics.

He, D., Y. Xia, T. Qin, L. Wang, N. Yu, T.-Y. Liu, & W.-Y. Ma (2016). "Dual Learning for Machine Translation." *Proceedings of the 30th International Conference on Neural Information Processing Systems* 820–828. Curran Associates Inc.

He, W., Z. He, H. Wu, & H. Wang (2016). "Improved Neural Machine Translation with SMT Features." *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* 151–157. AAAI Press.

Hinton, G. E., S. Osindero, & Y.-W. Teh (2006). "A Fast Learning Algorithm for Deep Belief Nets." *Neural Comput.*, **18**(7), 1527–1554.

Hirschmann, F., J. Nam, & J. Fürnkranz (2016). "What Makes Word-level Neural Machine Translation Hard: A Case Study on English-German Translation." *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers* 3199–3208. The COLING 2016 Organizing Committee.

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

Hutchins, W. J. (2007). "Machine translation: A concise history." *Computer Aided Translation: Theory and Practice*, 29–70.

池田尚志 (2009). "日本語からアジア諸言語への機械翻訳システムの構築奮闘記." 日本語学，28(12)， 62-71.

Jean, S., K. Cho, R. Memisevic, & Y. Bengio (2015a). "On Using Very Large Target Vocabulary for Neural Machine Translation." *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* 1–10. Association for Computational Linguistics.

Jean, S., O. Firat, K. Cho, R. Memisevic, & Y. Bengio (2015b). "Montreal Neural Machine Translation Systems for WMT'15." *Proceedings of the Tenth Workshop on Statistical Machine Translation* 134–140Lisbon, PortugalAssociation for Computational Linguistics.

張津一 (2016). "日中機械翻訳システム jaw/Chinese に関する研究—助詞「で」および多義語「流す」「乗る」の翻訳処理を中心に—." 岐阜大学大学院工学研究科　修士論文.

Johnson, M., M. Schuster, Q. V. Le, M. Krikun, Y. Wu, Z. Chen, N. Thorat, F. Viégas, M. Wattenberg, G. Corrado, M. Hughes, & J. Dean (2017). "Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation." *Transactions of the Association for Computational Linguistics*, **5**, 339–351.

謝軍 (2003). "日中機械翻訳システムに関する研究－変換・生成の方式およびテンス・モ

ダリティの翻訳処理－.” 岐阜大学大学院工学研究科　博士論文.

Junczys-dowmunt, M., T. Dwojak, H. Hoang, & C. Science (2016). “Is Neural Machine Translation Ready for Deployment? A Case Study on 30 Translation Directions.” *Proc. 9th IntẆorkshop on Spoken Language Translation (IWSLT)* 1–8.

Kalchbrenner, N. & P. Blunsom (2013). “Recurrent Continuous Translation Models.” *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* 1700–1709. Association for Computational Linguistics.

Karakanta, A., J. Dehdari, & J. v Genabith (2018). “Neural machine translation for low-resource languages without parallel corpora.” *Machine Translation*, **32**(1), 167–189.

Kim, Y., Y. Jernite, D. Sontag, & A. M. Rush (2016). “Character-aware Neural Language Models.” *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence* 2741–2749. AAAI Press.

Klakow, D. & J. Peters (2002). “Testing the correlation of word error rate and perplexity.” *Speech Communication*, **38**, 19–28.

Klein, G., Y. Kim, Y. Deng, J. Senellart, & A. Rush (2017). “OpenNMT: Open-Source Toolkit for Neural Machine Translation.” *Proceedings of ACL 2017, System Demonstrations* 67–72Vancouver, CanadaAssociation for Computational Linguistics.

Koehn, P. & R. Knowles (2017). “Six Challenges for Neural Machine Translation.” *Proceedings of the First Workshop on Neural Machine Translation* 28–39. Association for Computational Linguistics.

Kudo, T. (2018). “Subword Regularization: Improving Neural Network Translation Models with Multiple Subword Candidates.” *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 66–75. Association for Computational Linguistics.

賴坤竜 (2015). “日中機械翻訳システムに関する研究―助詞「と」及び「に」の翻訳処理を中心に―.” 岐阜大学大学院工学研究科　修士論文.

Lakew, S. M., Q. Lotito, M. Negri, M. Turchi, & M. Federico (2018). “Improving Zero-Shot Translation of Low-Resource Languages.” *Proceedings of the 14th International Workshop on Spoken Language Translation*Vol. abs/1811.01389.

Lample, G. & A. Conneau (2019). “Cross-lingual Language Model Pretraining.” Vol. abs/1901.07291.

Lample, G., M. Ott, A. Conneau, L. Denoyer, & M. Ranzato (2018). “Phrase-Based & Neural Unsupervised Machine Translation.” *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* 5039–5049. Association for Computational Linguistics.

Lee, J. D., K. Cho, & T. Hofmann (2016). "Fully Character-Level Neural Machine Translation without Explicit Segmentation." *Transactions of the Association for Computational Linguistics*, **5**, 365–378.

Li, J., D. Xiong, Z. Tu, M. Zhu, M. Zhang, & G. Zhou (2017). "Modeling Source Syntax for Neural Machine Translation." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 688–697. Association for Computational Linguistics.

Li, X., J. Zhang, & C. Zong (2016). "Towards Zero Unknown Word in Neural Machine Translation." *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* 2852–2858. AAAI Press.

Ling, W., I. Trancoso, C. Dyer, & A. Black (2015). "Character-based Neural Machine Translation." *arXiv preprint arXiv:1511.04586*.

Liu, H., M. Ma, L. Huang, H. Xiong, & Z. He (2019). "Robust Neural Machine Translation with Joint Textual and Phonetic Embedding." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* 3044–3049. Association for Computational Linguistics.

Luong, M.-T. & C. D. Manning (2015). "Stanford Neural Machine Translation Systems for Spoken Language Domain." *The International Workshop on Spoken Language Translation (IWSLT)*.

Luong, M. T. & C. D. Manning (2016). "Achieving Open Vocabulary Neural Machine Translation with Hybrid Word-Character Models." *Proc. 54th Annual Meeting of the Assoc. for Computational Linguistics*, 1054–1063.

Luong, M. T., H. Pham, & C. D. Manning (2015). "Effective Approaches to Attention-based Neural Machine Translation." *Proc. 2015 Conf. on Empirical Methods in Natural Language Processing* 1412–1421. ACL.

Luong, T., I. Sutskever, Q. Le, O. Vinyals, & W. Zaremba (2015). "Addressing the Rare Word Problem in Neural Machine Translation." *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)* 11–19Beijing, ChinaAssociation for Computational Linguistics.

Maruf, S. & G. Haffari (2018). "Document Context Neural Machine Translation with Memory Networks." *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1275–1284. Association for Computational Linguistics.

Mattoni, G., P. Nagle, C. Collantes, & D. Shterionov (2017). "Zero-Shot Translation for

Indian Languages with Sparse Data." *Proc. Machine Translation Summit 2017*.

Medress, M. F., F. S. Cooper, J. W. Forgie, C. C. Green, D. H. Klatt, M. H. O'Malley, E. P. Neuburg, A. Newell, R. Reddy, H. B. Ritea, J. E. Shoup-Hummel, D. E. Walker, & W. A. Woods (1976). "Speech Understanding Systems." *Artificial Intelligence*Vol. 9 307–316.

Meng, Y., X. Li, X. Sun, Q. Han, A. Yuan, & J. Li (2019a). "Is Word Segmentation Necessary for Deep Learning of Chinese Representations?." *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, **abs/1905.05526**.

Meng, Y., X. Ren, Z. Sun, X. Li, A. Yuan, F. Wu, & J. Li (2019b). "Large-scale Pre-training for Neural Machine Translation with Tens of Billions of Sentence Pairs." Vol. abs/1909.11861.

Mi, H., Z. Wang, & A. Ittycheriah (2016). "Vocabulary Manipulation for Neural Machine Translation." , 124–129.

Nagao, M. (1984). "A Framework of a Mechanical Translation Between Japanese and English by Analogy Principle." *Proc. Of the International NATO Symposium on Artificial and Human Intelligence* 173–180. Elsevier North-Holland, Inc.

Nakazawa, T. (2017). "New paradigm for machine translation: How the neural machine translation works." Vol. 60 299–306. Japan Science and Technology Agency.

Nakazawa, T., M. Yaguchi, K. Uchimoto, M. Utiyama, E. Sumita, S. Kurohashi, & H. Isahara (2016). "ASPEC: Asian Scientific Paper Excerpt Corpus." *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.

Neco, R. P. & M. L. Forcada (1997). "Asynchronous translations with recurrent neural nets." *Proceedings of International Conference on Neural Networks (ICNN'97)*Vol. 4 2535–2540 vol.4.

Nguyen, K., H. Daumé III, & J. Boyd-Graber (2017). "Reinforcement Learning for Bandit Neural Machine Translation with Simulated Human Feedback." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* 1464–1474. Association for Computational Linguistics.

Niehues, J. & E. Cho (2017). "Exploiting Linguistic Resources for Neural Machine Translation Using Multi-task Learning." *Proceedings of the Second Conference on Machine Translation* 80–89. Association for Computational Linguistics.

Och, F. J. & H. Ney (2003). "A Systematic Comparison of Various Statistical Alignment Models." *Computational Linguistics*, **29**(1), 19–51.

Olah, C. (2015). "Christopher Olah 氏のブログ記事." http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Papineni, K., S. Roukos, T. Ward, & W. j Zhu (2002). "BLEU: a Method for Automatic

Evaluation of Machine Translation." *In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics* 311–318.

Park, J., J. Song, & S. Yoon (2017). "Building a Neural Machine Translation System Using Only Synthetic Parallel Data." *CoRR*, **abs/1704.00253**.

Poncelas, A., D. Shterionov, A. Way, G. M. d Buy Wenniger, & P. Passban (2018). "Investigating Backtranslation in Neural Machine Translation." *Proceedings of the 21st Annual Conference of the European Association for Machine Translation*, **abs/1804.06189**, 249–258.

Pouget-Abadie, J., D. Bahdanau, B. v Merriënboer, K. Cho, & Y. Bengio (2014). "Overcoming the Curse of Sentence Length for Neural Machine Translation using Automatic Segmentation." *Proceedings of SSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation* 78–85. Association for Computational Linguistics.

Raganato, A. & J. Tiedemann (2018). "An Analysis of Encoder Representations in Transformer-Based Machine Translation." *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP* 287–297. Association for Computational Linguistics.

Ramachandran, P., P. Liu, & Q. Le (2017). "Unsupervised Pretraining for Sequence to Sequence Learning." *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* 383–391. Association for Computational Linguistics.

Reed, S., Z. Akata, X. Yan, L. Logeswaran, B. Schiele, & H. Lee (2016). "Generative Adversarial Text to Image Synthesis." *Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48* 1060–1069. JMLR.org.

Santos, C. & B. Zadrozny (2014). "Learning Character-level Representations for Part-of-Speech Tagging." *Proc. 31st Int. Conf. on Machine Learning (ICML '14)* 1818–1826.

Sennrich, R., A. Birch, A. Currey, U. Germann, B. Haddow, K. Heafield, A. V. M. Barone, & P. Williams (2017). "The University of Edinburgh's Neural MT Systems for WMT17." *Proceedings of the Second Conference on Machine Translation*, **1708.00726**, 389–399.

Sennrich, R. & B. Haddow (2016). "Linguistic Input Features Improve Neural Machine Translation." *Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers* 83–91Berlin, GermanyAssociation for Computational Linguistics.

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

Sennrich, R., B. Haddow, & A. Birch (2016b). "Neural Machine Translation of Rare Words

with Subword Units." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1715–1725Berlin, GermanyAssociation for Computational Linguistics.

陳劭毓 (2013). "日中機械翻訳システム jaw/Chinese に関する研究ー"把" 字文，助数詞，取り立て詞「も」の処理を中心に一." 岐阜大学大学院工学研究科　修士論文.

Sheridan, P. (1955). "Research in language translation on the IBM type 701." *IBM Technical Newsletter*No. 9，.

Shi, X., I. Padhi, & K. Knight (2016). "Does String-Based Neural MT Learn Source Syntax?." *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* 1526–1534. Association for Computational Linguistics.

Snover, M., B. Dorr, R. Schwartz, L. Micciulla, & J. Makhoul (2006). "A study of translation edit rate with targeted human annotation." *In Proceedings of Association for Machine Translation in the Americas* 223–231.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, & R. Salakhutdinov (2014). "Dropout: A Simple Way to Prevent Neural Networks from Overfitting." *Journal of Machine Learning Research*Vol. 15 1929–1958.

Stahlberg, F., A. d Gispert, E. Hasler, & B. Byrne (2017). "Neural Machine Translation by Minimising the Bayes-risk with Respect to Syntactic Translation Lattices." *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* 362–368. Association for Computational Linguistics.

Sutskever, I., O. Vinyals, & Q. V. Le (2014). "Sequence to Sequence Learning with Neural Networks." Ghahramani, Z., M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger eds., *Advances in Neural Information Processing Systems 27* 3104–3112. Curran Associates, Inc.

Tu, Z., Y. Liu, Z. Lu, X. Liu, & H. Li (2017). "Context Gates for Neural Machine Translation." Vol. 5 87–99.

Tu, Z., Z. Lu, Y. Liu, X. Liu, & H. Li (2016). "Modeling Coverage for Neural Machine Translation." *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 76–85. Association for Computational Linguistics.

Vaibhav, V., S. Singh, C. Stewart, & G. Neubig (2019). "Improving Robustness of Machine Translation with Synthetic Noise." *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)* 1916–1920. Association for Computational Linguistics.

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

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

Wang, M., Z. Lu, H. Li, & Q. Liu (2016). "Memory-enhanced Decoder for Neural Machine Translation." *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* 278–286. Association for Computational Linguistics.

Wang, X., Z. Lu, Z. Tu, H. Li, D. Xiong, & M. Zhang (2016). "Neural Machine Translation Advised by Statistical Machine Translation." *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)* 3330–3336.

Wang, X., H. Pham, Z. Dai, & G. Neubig (2018). "SwitchOut: an Efficient Data Augmentation Algorithm for Neural Machine Translation." *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* 856–861. Association for Computational Linguistics.

Wang, Y., L. Zhou, J. Zhang, & C. Zong (2017). "Word, Subword or Character? An Empirical Study of Granularity in Chinese-English NMT." *The 13th China Workshop on Machine Translation*, **abs/1711.04457**.

Weaver, W. (1949). "Translation." W. N. Locke & A. D. Boothe eds., *Machine Translation of Languages*MIT Press, 15–23.Reprinted from a memorandum written by Weaver in 1949.

Wu, L., Y. Xia, F. Tian, L. Zhao, T. Qin, J. Lai, & T.-Y. Liu (2018). "Adversarial Neural Machine Translation." Zhu, J. & I. Takeuchi eds., *Proceedings of The 10th Asian Conference on Machine Learning*Vol. 95 534–549. PMLR.

Wu, S., D. Zhang, Z. Zhang, N. Yang, M. Li, & M. Zhou (2018). "Dependency-to-Dependency Neural Machine Translation." Vol. 26 2132–2141.

Wu, S., M. Zhou, & D. Zhang (2017). "Improved Neural Machine Translation with Source Syntax." *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17* 4179–4185.

Xu, K., J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, & Y. Bengio (2015). "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention." Bach, F. & D. Blei eds., *Proceedings of the 32nd International Conference on Machine Learning*Vol. 37 of *Proceedings of Machine Learning Research* 2048–2057Lille,

FrancePMLR.

Yang, Z., W. Chen, F. Wang, & B. Xu (2018). "Improving Neural Machine Translation with Conditional Sequence Generative Adversarial Nets." *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)* 1346–1355. Association for Computational Linguistics.

Yang, Z., Z. Dai, Y. Yang, J. G. Carbonell, R. Salakhutdinov, & Q. V. Le (2019). "XLNet: Generalized Autoregressive Pretraining for Language Understanding." *ArXiv*.

王軼謳 (2008). "日中機器翻訳システムに関する研究－存在表現及び軽動詞構造に関する翻訳処理を中心に－." 岐阜大学大学院工学研究科　博士論文.

Zhang, J., Y. Liu, H. Luan, J. Xu, & M. Sun (2017). "Prior Knowledge Integration for Neural Machine Translation using Posterior Regularization." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* 1514–1523. Association for Computational Linguistics.

Zhang, J. & C. Zong (2015). "Deep Neural Networks in Machine Translation: An Overview." *IEEE Intelligent Systems*Vol. 30 16–25.

Zhang, J. & C. Zong (2016). "Exploiting Source-side Monolingual Data in Neural Machine Translation." *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* 1535–1545. Association for Computational Linguistics.

Zhang, J. & T. Matsumoto (2017). "Improving Character-level Japanese-Chinese Neural Machine Translation with Radicals as an Additional Input Feature." *Proc. 21st Int. Conf. on Asian Language Processing* 172–175.

Zhang, L. & M. Komachi (2018). "Neural Machine Translation of Logographic Language Using Sub-character Level Information." *Proceedings of the Third Conference on Machine Translation: Research Papers* 17–25. Association for Computational Linguistics.

Zheng, H., Y. Cheng, & Y. Liu (2017). "Maximum Expected Likelihood Estimation for Zero-resource Neural Machine Translation." *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17* 4251–4257.

Zhou, J., Y. Cao, X. Wang, P. Li, & W. Xu (2016). "Deep Recurrent Models with Fast-Forward Connections for Neural Machine Translation.." *Transactions of the Association for Computational Linguistics*Vol. 4 371–383.

Zhou, L., W. Hu, J. Zhang, & C. Zong (2017). "Neural System Combination for Machine Translation." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)* 378–384. Association for Computational Linguistics.

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

中澤敏明・C. Chu・黒橋禎夫 (2012). "日中共通漢字の整理とこれを利用した日中機械翻訳の高度化." Japio Year Book 2012, 258–261.

研究業績

学術論文

1. Jinyi Zhang，Tadahiro Matsumoto : “Corpus Augmentation for Neural Machine Translation with Chinese-Japanese Parallel Corpora,” *Applied Sciences*, vol.9, no.10, 2036, pp.1-16; https://doi.org/10.3390/app9102036, (2019).

国際会議

1. Jinyi Zhang，Tadahiro Matsumoto : “Character Decomposition for Japanese-Chinese Character-Level Neural Machine Translation,” *Proceedings of the 2019 International Conference on Asian Language Processing (IALP)*, pp.35-40, Shanghai (East China Normal University), China, (15-17 Nov, 2019).
2. Jinyi Zhang，Tadahiro Matsumoto : “Improving character level Japanese-Chinese neural machine translation with radicals as an additional input feature,” *Proceedings of the 2017 International Conference on Asian Language Processing (IALP)*, pp.172-175, Singapore (National University of Singapore), (5-7 Dec, 2017).
3. Jinyi Zhang，Tadahiro Matsumoto : “Japanese-Chinese machine translation for the Japanese case particle "de", ” *Proceedings of the 2017 International Conference on Asian Language Processing (IALP)*, pp.330-333, Singapore (National University of Singapore), (5-7 Dec, 2017).

講演（研究会・大会等）

1. 張津一，松本忠博：文字レベルの日中・中日ニューラル機械翻訳における文字分解による低頻度文字の削減，言語処理学会第 25 回年次大会（NLP2019）発表論文集，P1-7, pp.332-335，名古屋，(2019).
2. 張津一，松本忠博：ニューラル機械翻訳における長文分割によるコーパスの拡張，

言語処理学会第 25 回年次大会（NLP2019）発表論文集，P4-6，pp.683-686，名古屋，(2019).

3. 張津一，松本忠博：文字レベルの日中ニューラル機械翻訳における付加的特徴情報の検討，言語処理学会第 24 回年次大会（NLP2018）発表論文集，P10-21，pp.1069-1072, 岡山，(2018).
4. 張津一，松本忠博：日中機械翻訳システム jaw/Chinese における助詞「で」の翻訳処理，言語処理学会第 22 回年次大会（NLP2016）発表論文集，P13-7，pp.541-544，仙台，(2016).