

# 日本語作文の自動誤り訂正における 統計的機械翻訳とニューラル機械翻訳の性能評価

藤本 恭子<sup>†</sup> 水本 智也<sup>‡</sup> 井上 一成<sup>†</sup> 奥村 紀之<sup>§</sup>

<sup>†</sup> 明石工業高等専門学校 電気情報工学科 <sup>‡</sup> 理研 AIP <sup>§</sup> 大手前大学 現代社会学部

<sup>†</sup> {e1433@s., kinoue@}akashi.ac.jp, <sup>‡</sup> tomoya.mizumoto@riken.jp <sup>§</sup> noriyuki@otemae.ac.jp

## 1 はじめに

第二言語として日本語を習得しようとする外国人にとって、文法、綴り、語彙選択などの誤りを検出し、訂正することは容易ではない。また、日本語学習者のための日本語教師の数はおよそ6万人である一方で、学習者の数はおよそ365万人であり、学習者1人あたりの教師の数が不足しているのが現状である(2015年現在)。

本研究では、日本語教師が行う仕事の一つである日本語作文の添削を、自動誤り訂正システムを用いて代替することで日本語教師の負担を減らすことを目的とする。英語の自動誤り訂正は深層学習を使った手法を含め多くの手法が提案されて研究が進んでいる[3]一方で、日本語の自動誤り訂正は水本らの統計的機械翻訳(SMT)を使った誤り訂正以来ほとんど行われていない[1]。

本研究では、主として日本語学習者によって書かれた日本語作文に含まれる種々の誤りを検出し訂正する手法として、英語の自動誤り訂正で使われているニューラル機械翻訳(NMT)技術を利用した誤り訂正手法を構築する。一般に、日本語のような明示的な区切りを持たない言語を扱う場合、分かち書きの技術が必要となる。しかし、たとえ文法的に正しい文であっても分かち書きが正しく行われる保証はなく、誤りを含む文に対する分かち書きの頑健性は保証されない。そのため英語を対象とした誤り訂正技術を日本語に適用する際、分かち書きにかかる問題を解決する必要がある。また、別の問題として同じ表層の単語でも異なる用法になる曖昧性の問題がある。

これらの問題を解決するため、本稿では、形態素解析器を使用しない文字単位での対応によるモデル、分かち書きの結果に品詞情報を付与したモデルをそれぞれ構築し、日本語の自動誤り訂正の性能評価を行った。その結果、水本らのSMTによる誤り訂正手法に対し、GLEUSコアで13.65ポイントの改善が見られた。また、NMTに基づく誤り訂正手法に適したモデルは文字単位での対応によるモデルであることを確認している。

## 2 関連研究

本論文の評価対象である統計的機械翻訳およびニューラル機械翻訳を用いた自動誤り訂正システムは、学習者の文と人手による添削文を対にした大規模なテキストデータを必要とする。水本ら[1]は、大規模学習者コーパスとフレーズベース統計的機械翻訳(SMT)を用いた日本語作文の自動誤り訂正の手法を提案している。この研究では、日本語作文を対象とした自動誤り訂正における、相互添削型SNSから取得した大規模な学習者コーパスの有用性を示している。

Chollampattら[3]は多層畳み込みニューラルネットワーク(CNN)を用いることで、英語作文の自動誤り訂正を試みている。このCNNを用いた自動誤り訂正は、従来のSMTを用いた英語作文の自動誤り訂正に比べて $F_{0.5}$ 値が13.63ポイント上回っている。学習者の文は局所的な誤りが多く、文法誤りも近隣語にのみ依存することから、より小さな窓幅でたたみ込んだCNNは局所的な誤りの訂正に強いと報告されている。局所的な誤りは日本語の作文でも多く見られることから、本研究でもCNNの訂正システムを用いる。本研究では、NMTの中でもCNNを用いた自動誤り訂正を日本語作文を対象に実施した。

## 3 提案手法

本節では、大規模学習者コーパスとNMTを用いた自動誤り訂正システムを作成し、従来のSMTを用いた自動誤り訂正システムとその性能を比較評価する。NMTを用いた自動誤り訂正システムを作成する上で、よりシステムとして有用な誤り訂正を実現することが大きな目標となる。英語と異なり日本語の場合、単語に区切られていないため、訂正をする前に単語に分割する必要があるが、誤りを含んだ文では正しく分かち書きできない可能性がある。また、別の問題として、同じ表層の単語でも品詞が異なると訂正されやすい単語も変わると考える。

本研究ではこれらの問題に対して入力とするコーパス

の形式を変えることで対処する。まず最初に基本的な単語-単語対応コーパスの話をした後、文字-文字対応コーパスと品詞情報付き単語-単語対応コーパスについて説明する。なお、実験に使用した3つのコーパスは全て Chollampatt ら [3] のモデルを用いて学習する。

### 3.1 単語-単語対応コーパス

一般的に機械翻訳で日本語から他言語に変換する場合、単語に分割してから翻訳を行う。本研究において、日本語学習の自動誤り訂正に向けた学習に用いるテキストデータ(学習者の文と人手による添削文の対)に含まれる各文を分かち書きしている。分かち書きを行ったテキストデータを単語-単語対応コーパスとする。なお分かち書きには、mecab-ipadic を辞書として用いた MeCab0.996<sup>1</sup> を使用する。以下に単語-単語対応コーパスに含まれるテキストデータの例を示す。Learner は学習者の文、Correct は人手による添削文である。

Learner: また寝ようとしましょう。  
Correct: また寝てみるとしましょう。

水本ら [1] は学習者の文に分かち書きを行う際の問題点として、学習者の文に対して正しい分かち書きができない点を挙げている。表1は、学習者の文と人手による添削文のそれぞれを分かち書きした例である。

表1: 分かち書きの例

	原文	分かち書きした文
Learner	こんいち は!!	こん いち は !!
Correct	こんにちは!!	こんにちは !!

分かち書きされた学習者の文と人手による添削文の分かち書き結果を比較すると、学習者の文には形態素解析誤りが含まれており、人手による添削文と対応を取ることができない。また、単語-単語対応コーパスは、同じ単語でも場合によっては品詞や用法が異なる可能性がある。例えば、「そうよ」という文と「このよはむじょう(この世は無常)」という文のそれぞれに形態素解析を行うと、それぞれの文に含まれる「よ」という単語は、「そうよ」の場合は助詞、「このよはむじょう」の場合は名詞となる。コーパスの性質上、これらの「よ」は同じ語彙とされるが、日本語作文内に現れた時、品詞の異なる同じ単語が同じ文脈で用いられることはない。これらの問題を解決するのが、次の文字-文字対応コーパスおよび品詞情報付き単語-単語対応コーパスである。

### 3.2 文字-文字対応コーパス

文字-文字対応コーパスは、学習に用いるテキストデータに文字分割を行ったものである。以下は文字-文字対

応コーパスに含まれるテキストデータの例である。

Learner: また寝ようとしましょう。  
Correct: また寝てみるとしましょう。

文字-文字対応コーパスは、形態素解析器を用いないため、単語-単語対応コーパスに現れるような解析誤りの影響を受けない。

### 3.3 品詞情報付き単語-単語対応コーパス

品詞情報付き単語-単語対応コーパスは、分かち書きにおける品詞の違いを考慮したものである。以下は、品詞情報付き単語-単語コーパスに含まれるテキストデータの例である。

Learner: こん<名詞> いち<名詞> は<助詞> !!<名詞>  
Correct: こんにちは<感動詞> !!<名詞>

品詞情報付き単語-単語コーパスでは、分かち書きをしたテキストデータに MeCab の出力結果から得た品詞情報を利用してタグを付与している。

## 4 実験

実験では、3つのコーパスを使って日本語作文の自動誤り訂正システムを作成し、それぞれのシステムを定量的、定性的に検証する。また、SMT と NMT のそれぞれ用いた自動誤り訂正システムの性能を評価するために、ベースラインとして水本らの研究の再現実験を実施した。実験には既存の統計的機械翻訳システム Moses4.0<sup>2</sup> と単語アライメントツールとして GIZA++<sup>3</sup> を利用した。

### 4.1 データセット

本研究では、水本らが発表した相互添削型 SNS である Lang-8 から抽出した大規模学習者コーパス (Lang-8 Learner Corpora) を用いた。また、日本語作文を自動誤り訂正の対象とするという観点から、全作文データのうち、学習者の母語に関わらず日本語の作文データ 28,978 エッセイ分 (2,345,274 文対) のみを抽出した。Lang-8 Learner Corpora の特徴として、添削文にコメントが付与されることが挙げられる。添削文にコメントが存在する場合、本来必要のない文が添削文に挿入されていることとなり、学習者の文とのレーベンシュタイン距離は大きくなる。レーベンシュタイン距離とは、1文字の挿入・削除・置換によって、一方の文字列をもう一方の文字列に変形するのに必要な手順の最小回数として定義される。

そのため、水本ら [1] の手法と同様に、学習者の文と人手による添削文のレーベンシュタイン距離が5以上である場合は、その文対を学習コーパスから除去した。

また、Yoon-Hyung ら [2] の研究では、機械翻訳におい

<sup>1</sup><http://taku910.github.io/mecab/>

<sup>2</sup><http://www.statmt.org/ Moses/>

<sup>3</sup><https://github.com/moses-smt/giza-pp/>

て長い文を翻訳することへの問題点を挙げている。SMTを用いた自動誤り訂正では、学習者の文と人手による添削文で単語同士の対応関係を解析する単語アライメントという処理を行うため、構文が曖昧であると解析に問題が生じる。本研究では、この問題を解決するため、読みやすい文の限界を80文字と定め、80文字以上で構成される文を長い文と仮定した。そして、80文字以上の長い文を除去した。これらの処理を行った結果、全体の13.6%(312,415文対)が除去された。本節以降、残りの2,027,459文対を用いて学習を行うこととする。訓練データと開発データにはLang-8 Learner Corporaを使用し、テストデータには日本語学習者による日本語/母語発話の対照言語データベース<sup>4</sup>[4]を用いた。訓練データは2,022,059文対、開発データは5,400文対、テストデータは4,952文対である。訓練は開発セットで最良のBLEUスコア[6]を示したEpoch数で終了している。

## 4.2 定量評価

実験結果はGLEU[5]を使って評価する。評価の際、全てのシステムを同じ基準で評価するため、システムからの出力文は一度スペースの無い日本語作文に加工し、その後分かち書きを行った。GLEUでは、評価システムの出力と人手による添削文の4-gram一致率から、学習者の文と評価システムの出力には存在し正解文に存在しない4-gramが多いほど減点することによって評価する。表2に各自動誤り訂正システムのGLEUスコアを示す。

表 2: 各自動誤り訂正システムのGLEUスコア

入力コーパス	MTの種類	GLEU
学習者の文	-	42.11
単語-単語	NMT	48.70
文字-文字	NMT	51.35
品詞情報付き単語-単語	NMT	48.60
単語-単語 (Baseline)	SMT	37.70

SMTを用いた自動誤り訂正システムとNMTを用いた自動誤り訂正システムを比較すると、文字-文字対応コーパスで学習したNMTを用いたシステムの方が13.65ポイントGLEUスコアが高い。また、学習者の文におけるGLEUスコアと各自動誤り訂正システムのGLEUスコアを比較してみても、SMTを用いたシステムが学習者の文のGLEUスコアを約5ポイント下回ったのに対して、NMTを用いたシステムは学習者の文のGLEUスコアを18ポイント以上上回った。

次に、NMTを用いた自動誤り訂正システムにおける入力コーパスに着目する。提案した3種類のコーパス

のうち、文字-文字対応コーパスが最も高いGLEUスコアを記録し、単語分割を用いたコーパスに比べて平均で2.6ポイント以上GLEUスコアが高くなった。

## 4.3 定性評価

表3から表8に、日本語作文の自動誤り訂正における各システムの実際の出力例を示す。表3は評価対象である4つ全てのシステムで訂正ができた例である。

表 3: 全てのシステムで訂正できた例

学習者の文	そのよるを祭りがおわります。
システムの出力	そのよる祭りがおわります。
人手による添削	そのよる祭りがおわります。

どのシステムも1文字から3文字程度の短い単語の挿入や削除を行う訂正は、いくつか訂正ミスがあるもののほとんど問題なく行えた。

単語-単語対応コーパスを用いたシステムの正しい出力例を表4に示す。

表 4: 単語-単語対応コーパスの出力例

学習者の文	子どもに親しくしられません。
システムの出力	子どもに親しくできません。
人手による添削	子どもと親しくできません。

単語-単語対応コーパスを用いたシステムでは、入力されたテストデータに対して、“しられません”を“できません”に変換できた他、“びっくりした”を“びっくりしました”、“味があります”を“味がします”といったより人間らしい表現に置換訂正をすることも可能であった。このように、単語-単語対応コーパスを用いたシステムは、添削の際に頻繁に注意される語彙の置換に秀でる一方、“やっぱり”を“やぱり”とする、入力ミスなどが原因の文法と関わりのない誤りはあまり訂正されなかった。

表5は文字-文字対応コーパスを用いたシステムの正しい出力例である。

表 5: 文字-文字対応コーパスの出力例

学習者の文	私のために、煙草を吸あない。
システムの出力	私のために、煙草を吸わない。
人手による添削	私のために、煙草を吸わない。

文字-文字対応コーパスを用いたシステムは、このように“吸あない”を“吸わない”、“結婦式”を“結婚式”といった1文字の挿入、置換、削除など、局所的な誤りの訂正に強いことが分かった。

品詞情報付き単語-単語対応コーパスを用いたシステムの誤り訂正は、学習者の文をより人間らしい表現に置換訂正できる点で単語-単語対応コーパスを使ったシス

<sup>4</sup><http://db3.ninjal.ac.jp/contr-db/>

テムと誤り訂正の傾向が似ていた。しかし、表 6 に示すような名詞を組み合わせた文が全く違う文章に置き換わってしまうといった特徴が見られた。

表 6: 品詞情報付き単語-単語コーパスの出力例

学習者の文	作文課題 2
システムの出力	よろしくお願ひします。
人手による添削	作文課題 2

また、表 7 のように単語-単語対応コーパスと品詞情報付き単語-単語対応コーパスに共通して、学習データに含まれない固有名詞が未知語 (<unk>) として訂正文に現れてしまうという特徴があった。

表 7: 分かち書きを用いるシステム共通の出力例

学習者の文	ナークは仏になります。
システムの出力	<unk>は仏になりました。
人手による添削	ナークは僧になります。

表 8 に示すように、SMT を用いた自動誤り訂正における特徴として、文章内に新たな語彙を追加したり、語彙を代替するような訂正に優れていることが分かった。

表 8: 単語-単語対応モデル (SMT) の出力例

学習者の文	一度見るつもりです。
システムの出力	一度この目で見るつもりです。
人手による添削	一度見てみたいです。

しかしながら、この特徴は以下の例のように不要な情報を文章に入れてしまう原因にもなりやすい。Correct は人手による添削文、Output はシステムの出力文である。

Correct: 祭礼は朝から始まります。  
Output : 祭礼は朝 7時から始まります。

## 5 考察

本実験で用いた学習データは相互添削型 SNS の添削ログ、テストデータは日本語を母語としない日本語学習者による作文課題とその添削結果である。以下に示すように、相互添削型 SNS の添削ログ (添削ログ) は、コミュニケーションを行うツールを利用したものであるため、砕けた表現が多い一方、日本語学習者による作文課題とその添削結果 (作文課題) には、堅い表現が多い。

添削ログの文例: 遠藤さんカワ (・▽・) イイ!!  
作文課題の文例: 今年、「ハリヤヤ」が二回あります。

つまり、それぞれの文章は用いられる場面や日本語の習熟度が違う為、誤りの傾向にも違いがあると予想される。例えば、添削ログに含まれる文には助詞誤りの訂正が多く、作文課題含まれる文には語彙や表現の訂正が

多い。水本らは、SMT を用いた日本語作文の自動誤り訂正において、学習データとテストデータのドメインが一致しているほど、再現率が高くなったと述べており、NMT を用いた日本語作文の自動誤り訂正に関しても、学習データとテストデータのドメインが一致していれば、さらに人手の添削に近い誤り訂正が可能だと考えられる。

## 6 おわりに

本研究では、大規模学習者コーパスと NMT を用いた自動誤り訂正システムを作成し、従来の SMT を用いた自動誤り訂正システムとその性能の比較評価を行った。英語作文の自動誤り訂正の分野で高い成果を挙げた手法を、日本語作文にも適応することで、従来の SMT を用いた自動誤り訂正システムよりも GLEU スコアで 13.65 ポイント優れた自動誤り訂正システムを構築することができた。

その一方で、未知語問題や学習や評価に用いたテキストデータのドメインの違いなど、自動誤り訂正システムとして不完全なところも多く確認された。今後は、未知語問題への取り組みやテキストデータを精査し、システム性能の向上を目指す。

## 参考文献

- [1] 水本 智也・小町 守・永田 昌明・松本 裕治 (2013) 「日本語学習者の作文自動誤り訂正のための語学学習者 SNS の添削ログからの知識獲得」, 『人工知能学会論文誌』 28, p420-432, 人工知能学会.
- [2] Yoon-Hyung Roh, Young-Ae Seo, Ki-Young Lee, Sung-Kwon Choi(2001) 「Long Sentence Partitioning using Structure Analysis for Machine Translation」 NLPRS, November 27th, 2001.
- [3] Shamil Chollampatt and Hwee Tou Ng(2018) 「Multilayer Convolutional Encoder-Decoder Neural Network for Grammatical Error Correction」 AAAI-18, February 2nd-7th, 2018.
- [4] 宇佐美 洋・籠宮 隆之・梶本 総子 (2004) 「『日本語学習者による日本語/母語発話の対照言語データベース』の設計」, 『聴覚研究会資料』 34, p253-p258, 日本音響学会.
- [5] Napoles Courtney, Sakaguchi Keisuke, Post Matt, Tetreault Joel(2015) 「Ground Truth for Grammatical Error Correction Metrics」, 『Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing』, p.588-p.593, Association for Computational Linguistics.
- [6] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu(2002) 「BLEU: a method for Automatic Evaluation of Machine Translation」, 『Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics』, p01-p1040, ACL.