

# BERTによる英文前置詞誤りの自動修正手法の提案

中谷 聡<sup>†</sup> 上野 史<sup>††</sup> 太田 学<sup>††</sup>

<sup>†</sup> 岡山大学大学院自然科学研究科 〒700-8530 岡山県岡山市北区津島中 3-1-1

<sup>††</sup> 岡山大学学術研究院自然科学学域 〒700-8530 岡山県岡山市北区津島中 3-1-1

E-mail: <sup>†</sup>p2x03pxw@s.okayama-u.ac.jp, <sup>††</sup>{uwano, ohta}@okayama-u.ac.jp

あらまし 本研究では、自然言語処理モデルである Bidirectional Encoder Representations from Transformers (BERT) を用いて、入力された英文に含まれる前置詞誤りを自動修正する手法を提案する。まず、英文中のマスクした前置詞を Masked Language Model (MLM) で予測する問題とマスクした箇所に前置詞が挿入されるかどうかの分類問題で、それぞれ BERT をファインチューニングする。次に、MLM を利用した単語の出現確率と前置詞の有無の分類問題を利用した前置詞が出現する確率を求め、二つの確率を利用して誤った前置詞を検出し、正しい前置詞を予測する。実験では、提案手法により前置詞誤りを自動で検出、修正し、それらの精度を評価した。またオンラインの英文修正サービスである Grammarly とこれらの結果を比較した。

キーワード BERT, 前置詞誤り, 誤り検出, 誤り修正

## 1 はじめに

前置詞は英語の中でも用法は複雑で、また日本語には前置詞が存在しないため日本人の英語学習者は前置詞の用法を誤ることがある。前置詞誤りは、複数の文法エラー訂正タスク [1] [2] で誤りタイプの一つとされている。また第二言語話者の前置詞の使い方や誤りの傾向について調査した研究がある。[3] では、多くの意味を持つ前置詞は誤りやすい傾向にあることが示された。[4] では、前置詞の目的語が使用頻度の低い単語の場合、誤りやすい傾向にあることが示された。また、前置詞などの誤りを訂正するとき、英語を母語とする人に添削してもらうことが最適だが、身近にいないことも多い。現在では、Grammarly [5] や Ginger [6] など英文修正サービスが複数存在しており、いつでも英文を修正できる。しかし適切でない前置詞を使用しても文法的に間違いではないため、英文修正サービスで修正されない場合がある。そのため期待とは異なる文意となり、読者の誤解を招くことがある。

本研究では、我々が [14] で定めた 3 種類の前置詞誤りである置換誤り、挿入誤り、欠落誤りを、ファインチューニングした BERT [7] により検出し、修正する。具体的には、まず複数の自然言語処理タスクで成果を上げている自然言語処理モデルの BERT を、英文中のマスクした前置詞を MLM で予測する問題とマスクした箇所に前置詞が挿入されるかどうかの分類問題でそれぞれファインチューニングする。次にファインチューニング後の二つのモデルを利用して、3 種類の前置詞誤りを検出し、自動修正する手法を提案する。

本稿の構成は以下のとおりである。2 節で関連研究について、3 節では BERT について述べる。4 節では、ファインチューニング後の BERT モデルを利用した前置詞誤りの検出と修正について説明する。5 節では、提案手法の評価実験とその結果を示し、6 節では、本稿のまとめと今後の課題を述べる。

## 2 関連研究

### 2.1 英文前置詞誤りの検出と修正

松本ら [11] は、疑似データの使用、リランキング、反復符号化の三つの手法を利用して、文法的な誤りを含む文を入力とし、正しい文を出力する文法誤り訂正タスクで再現率が向上することを示した。疑似データの使用は、疑似データを利用して、訓練データで前置詞誤りを含む文の割合を増やす。リランキングは、修正候補文で前置詞が修正されている数をスコアとして取り入れる。反復符号化は、1 回修正した文を再び入力し、再修正を試みる。また文法エラー訂正モデルとして transformer [10] を利用した。文法エラー訂正タスクのデータセットである BEA-2019 Shared Task で、三つの手法すべてを組み合わせることで、再現率がベースラインの 38.92 から 56.35 に向上した。

有富ら [12] は、検索エンジンの検索結果数を利用し、英文前置詞誤りを自動検出して修正候補を提示するシステムを提案した。前置詞を含む英文を入力として与えると、システムが前置詞をワイルドカードに置き換え、前置詞の前後の単語と組み合わせた句を自動生成し、検索フレーズとして検索する。そして検索結果からワイルドカードの箇所に入る前置詞の出現確率を計算し、結果を四段階でユーザに提示する。置換誤りで英文修正サービスの NativeChecker と比較した。誤り検出で F 値は 0.85 で 1.0 ポイント上回り、修正精度は 0.82 で 0.1 ポイント上回った。

### 2.2 BERT を用いた英文誤りの修正

Kaneko ら [13] は、文法誤りがある英語学習者コーパスでファインチューニングした BERT モデルを修正候補文のリランキングの要素として組み込むことで、あらゆる英文誤りを含む文のデータセットに対して、修正精度が向上することを示した。ま

ず BERT を文法誤りがあるかどうかのラベルが付与された文でファインチューニングした。これにより文単位で文法誤りがあるかどうかを区別することができる。次に transformer [10] を文法エラー訂正モデルとして利用し、入力文に対し仮の修正文を出力する。次に仮の修正文を入力とし、ファインチューニングした BERT モデルから求まる文法的な正しさのスコアを、ランキングの特徴量として利用し最終的な修正文を出力する。文法誤り訂正タスクのデータセット W&I+LOCNESS [2] で、BERT を利用することで英文誤り修正の再現率が 23.76 から 28.08 に、適合率に 2 倍の重みをかける F 値である  $F_{0.5}$  スコアが 34.41 から 35.35 に向上した。

我々は [14] で、BERT を用いて前置詞誤りを自動検出し、修正候補を一つ以上提示する手法を提案した。この手法ではまず英文中の前置詞を検出し、前置詞を [MASK] トークンに置換、もしくは新たに [MASK] トークンを挿入する。次に MLM を利用し単語の出現確率を求め、それを利用して誤った前置詞を検出し、修正候補を提示する。置換誤りの検出で F 値は 0.908 であり、これは Grammarly を 22.3 ポイント上回った。置換誤りの修正精度は 0.877 であり、これは Grammarly を 2.5 ポイント上回った。本研究は [14] と同様に、入力英文から誤り検出のための [MASK] トークンを含む文を生成する。しかし、本研究が前置詞誤りの検出と修正にファインチューニングしたモデルを利用する点、前置詞の出現確率だけでなく、マスクした箇所に前置詞が入る確率を利用し誤りを修正する点、修正候補を複数提示するのではなく一つの修正候補で自動修正する点が異なる。

### 3 BERT

BERT [7] は自然言語処理における事前学習モデルの一つである。文の文法的正しさを判別するタスク (CoLA) [8] や、質問文と答えを含む文章の組を受け取り文章の答えを予測するタスク (SQuAD v1.1) [9] など複数の自然言語処理タスクで最も高い精度を出し、注目された。図 1 に BERT の概略図を示す。図 1 では  $E_i$  が入力を表し、 $T_i$  が出力を表している。Trm は、attention を用いた機械学習モデルの transformer [10] である。BERT に英文を入力すると単語がサブワードに分割される。例えば “Where do you live? I live in Okayama.” という 2 文では “Where”, “do”, “you”, “live”, “?”, “I”, “live”, “in”, “Okay”, “##ama”, “.” の 11 のサブワードに分割され、先頭に [CLS], 文と文の間に [SEP] という特別なトークンを挿入した 13 トークンになる。また, [MASK] トークンの位置を示すラベルや 1 文目か 2 文目かを識別するラベルが生成される。これらトークンとラベルの組が BERT への入力になる。サブワードは、出現頻度の高い単語はそのまま一つの単語、出現頻度の低い単語は分割することで別々の短い単語として扱う方法である。サブワードに分割することで未知語がなくなるメリットがある。従来の言語モデルでは、トークンを左から右の一方から読み込み、学習をしていた。一方 BERT ではトークンを右から左へ読み込む学習もする。これにより、文脈を前方向から

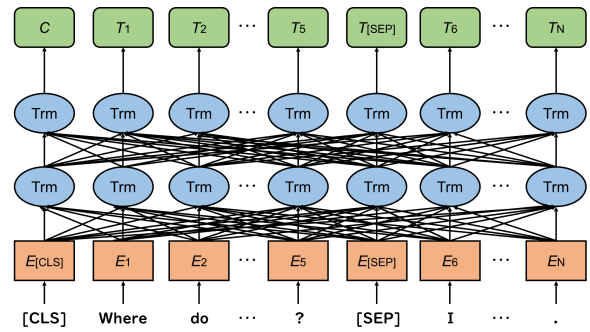


図 1 BERT の概略図 [7]

だけでなく後ろ方向からも学習し、前後の関係性を考慮することができる。BERT では事前学習の際に、一定確率でサブワードをマスクし、そのサブワードを予測する Masked Language Model (MLM) と、2 文を入力として受け取り、前の文の次に後ろの文が続くか否かを予測する Next Sentence Prediction (NSP) を用いて学習する。

MLM では、まずサブワードの 15% を [MASK] トークンに置き換える。次にマスクされなかったサブワードから得られる文脈に基づいて、[MASK] トークンに置き換えられたサブワードを予測する。例えば以下の文 (1) の場合、文中から 15% の確率でサブワードが選ばれ、“want” と “with” が選ばれた場合、それらをマスクした文 (2) を生成する。そして文 (2) の [MASK] トークンを推測する。推測結果として、正規化された値でサブワードごとの出現確率を出力する。

- (1) I want to play tennis with my friends.
- (2) I [MASK] to play tennis [MASK] my friends.

また BERT は様々なタスクに対応する汎用的なモデルであり、ファインチューニングにより感情分析や文書分類など様々な自然言語処理のタスクの問題を解くことができる。ファインチューニングは、事前学習済みのモデルにデータセットを追加し再学習することで、モデルのパラメータを更新する学習手法である。BERT ではタスクに応じた最終層を追加し、transformer [10] と最終層のパラメータを更新する。例えば感情分析のタスクでは、文と感情のラベルの組のデータセットでファインチューニングする。ファインチューニングにより、出力として最終層から感情のラベルとそのスコアを得る。

### 4 提案手法

ここでは、まず置換誤り、挿入誤り、欠落誤りの 3 種類の前置詞誤りについて説明する。次に BERT のファインチューニングで使用するデータセットについて述べる。次にマスクした前置詞を Masked Language Model (MLM) で予測する問題とマスクした箇所に前置詞が挿入されるかどうかの分類問題で、BERT をそれぞれファインチューニングする方法を説明する。次にこれらの二つの BERT による前置詞誤りの自動検出、自動修正の方法について述べる。

表 1 タグとして付与されている前置詞一覧

about	above	across	after	against	along	among
around	as	at	before	behind	beneath	beside
between	by	down	during	for	from	in
inside	into	like	of	off	on	onto
over	round	through	to	towards	with	

#### 4.1 前置詞誤りの種類

本研究では、以下の3種類の前置詞誤りを検出し修正する。

- 置換誤り：異なる前置詞が使用される誤り

誤り： I am a student **of** Okayama University.

正解： I am a student **at** Okayama University.

- 挿入誤り：不要な前置詞が挿入される誤り

誤り： They discussed **about** food issues.

正解： They discussed food issues.

- 欠落誤り：必要な前置詞が欠落する誤り

誤り： He is slow to reply emails.

正解： He is slow to reply **to** emails.

#### 4.2 使用するデータセット

ファインチューニングには、The Preposition Project (TPP) データセットの SemEval コーパスを用いる。このコーパスは SemEval-2007 の Word-Sense Disambiguation of Prepositions [15] におけるデータセットに採用された。このコーパスは文法的に正しい 24,306 文で構成され、16,316 文が訓練用データ、7,990 文が検証用データとして分割されている。それぞれの文は少なくとも 34 種類の前置詞のうちいずれか一つ以上を含み、文中の前置詞のうち一つがタグとして付与されている。この前置詞の一覧を、表 1 に示す。

#### 4.3 MLM の前置詞予測問題によるファインチューニング

置換誤りは、本来使用すべきものと異なる前置詞を使用する誤りである。そこで前置詞を [MASK] トークンに置き換え、予測されるサブワードと出現確率を利用し、この置換誤りの検出と修正をする。BERT が出力する出現確率を基に検出と修正をするため、正しい前置詞の出現確率は高いほうが望ましい。そこで前置詞の箇所をマスクした文を使用し、MLM で前置詞を予測する問題で BERT をファインチューニングする。ファインチューニングのためのデータセットは、SemEval コーパスの訓練用データ 16,316 文から作成する。34 種類の前置詞いずれかのタグが付与されている文で、タグの単語と同じ単語を無作為に1か所選び、[MASK] トークンに置き換える。このように前置詞1か所を [MASK] トークンに置き換えた 16,316 文をファインチューニングに使用する。例えば “about” のタグが付与されている以下の文 (1) の場合、“about” を [MASK] トークンに置き換え文 (2) を生成する。

(1) Religion was the first vehicle for general thoughts about existence.

(2) Religion was the first vehicle for general thoughts [MASK] existence.

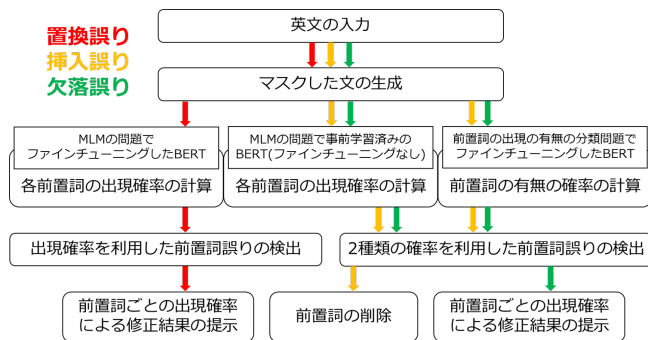


図 2 前置詞誤りの自動検出と自動修正の概要

[MASK] トークンを含む文を BERT に入力すると、サブワードとその出現確率の組で結果が出力される。例えばファインチューニング後のモデルに “The 15-month-old boy’s screams woke his mother [MASK] the early hours.” という文を入力する。この文では [MASK] トークンに “in” が入るのが正しい。出力の上位 2 件は、0.9981 の確率の “in” , 0.0013 の確率の “during” という組になる。“in” に対する出現確率は 1 に近いいため、[MASK] トークンに “in” が入る確率は高いと予測する。

#### 4.4 前置詞の有無の分類問題によるファインチューニング

挿入誤りと欠落誤りは、前置詞が必要か否かを取り違えた誤りである。そこで前置詞が入るかどうかを分類することで、挿入誤りと欠落誤りを検出する。そこで前置詞を無作為に1か所 [MASK] トークンに置き換えた文 (前置詞が入る文) と、[MASK] トークンを余分に1か所挿入した文 (前置詞が入らない文) を作成し、2 種類の文を分類する問題でファインチューニングする。これにより、[MASK] トークンの箇所に前置詞が入るかどうかが分類する。SemEval コーパスの訓練用データの文で、半分は 4.3 節と同様に前置詞を1か所 [MASK] トークンに置き換えて “0” のラベルを付与し、もう半分は1か所ランダムに [MASK] トークンを挿入し “1” のラベルを付与する。これらの操作により、ファインチューニングのためのデータセットを作成する。つまり “0” のラベルを付与された文は [MASK] トークンの箇所に前置詞が入る文、“1” のラベルを付与された文は [MASK] トークンの箇所に前置詞が入らない文になる。例えば “People have been very generous with their donations.” という文から、“0” のラベルが付与された文を作成すると以下の文 (1) になる。一方同じ文から、“1” のラベルが付与された文を生成する場合は以下の文 (2) になる。

(1) People have been very generous [MASK] their donations.

(2) People have been [MASK] very generous with their donations.

ファインチューニングした BERT に [MASK] トークンを含む文を入力することで、“0” のラベルに対する確率つまり [MASK] トークンに前置詞が入る文か否かを求められる。

#### 4.5 前置詞誤りの自動検出と自動修正

前置詞誤りの自動検出と自動修正の処理の流れを図 2 に示す。

まず誤り検出のためにマスクした文は, [14] で我々が提案した品詞分解による手法により生成する. 置換誤りと挿入誤りでは前置詞を一か所 [MASK] トークンに置換し, 欠落誤りでは一か所 [MASK] トークンを挿入する.

次に図 2 の「各前置詞の出現確率の計算」では, 置換誤りは MLM でファインチューニングした BERT, 挿入誤りと欠落誤りはファインチューニングなしで MLM で事前学習済み BERT で, [MASK] トークンに対する表 1 の 34 種類の前置詞の出現確率を求める. 「前置詞の有無の確率の計算」では, 挿入誤りは前置詞を一か所 [MASK] トークンに置き換えた文で, 欠落誤りは一か所 [MASK] トークンを挿入した文で, [MASK] トークンに前置詞が出現する確率を, 前置詞の有無の分類問題でファインチューニングした BERT モデルで求める.

次に二つのファインチューニングした BERT を利用して確率を得る. この確率を利用し, 前置詞誤りを検出する. 3 種類の誤りの検出の判定基準は以下の通りである.

- 置換誤りの検出

対象の前置詞の出現確率が  $10^{-3}$  以下かつ対象の前置詞とは異なる前置詞で最も出現確率の高い前置詞の出現確率が 0.25 以上の場合, 対象の前置詞を置換誤りとして検出する.

- 挿入誤りの検出

表 1 の 34 種類の前置詞に対する出現確率上位五つの前置詞の出現確率の合計と, 前置詞の有無の分類問題で [MASK] トークンに前置詞が入る確率を乗算する. 乗算した値が  $10^{-3}$  以下の場合, 対象の前置詞を挿入誤りとして検出する.

- 欠落誤りの検出

表 1 の 34 種類の前置詞に対する出現確率上位五つの前置詞の出現確率の合計と, 前置詞の有無の分類問題で [MASK] トークンに前置詞が入る確率を乗算する. 乗算した値が値が 0.75 以上の場合, その [MASK] トークンの箇所欠落誤りがあるとする.

置換誤りの検出と挿入誤りの検出の両方の条件満たした場合は, 挿入誤りとして検出し, 置換誤りは検出しない. 置換誤りでは, MLM の問題でファインチューニングした BERT を利用する. MLM を利用し得られる [MASK] トークンに入るサブワードとその出現確率を基に前置詞誤りを検出する. 挿入誤りと欠落誤りについては, ファインチューニングなしで MLM で事前学習済み BERT, 前置詞の有無の分類の問題でファインチューニングした BERT を利用する. MLM を利用し得られる [MASK] トークンに入るサブワードとその出現確率と [MASK] トークンに前置詞が出現する確率を基に前置詞誤りを検出する.

検出した前置詞誤りの自動修正は以下のように行う. 置換誤りと欠落誤りについては, 検出で得る前置詞とその出現確率の組のうち, 最も出現確率が高い前置詞で修正する. なお挿入誤りは検出すれば, その前置詞の削除を促すことが修正になるため, 修正候補は提示しない.

## 5 評価実験

まず, 評価指標として用いる Recall(再現率), Precision(適

合率), F 値を説明する. 次に, 前置詞誤りの検出と修正の精度を評価する. 次に, 提案手法と比較する英文修正サービスの Grammarly を説明し, 提案手法と Grammarly との比較をする. 次に, 実験結果に基づいて, 提案手法について考察する.

### 5.1 評価指標

前置詞誤りの検出の評価指標として, Recall(再現率), Precision(適合率), 再現率と適合率の調和平均の F 値を用いる. これらの指標は以下の式で求められる.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{F 値} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}} \quad (3)$$

誤り検出では, TP, FN, FP は以下の通りである.

- TP: 検出できた前置詞誤り
- FN: 検出し損なった前置詞誤り
- FP: 誤検出した誤りを含まない前置詞

また, 置換誤りと欠落誤りの修正の評価指標として, 以下の修正精度を用いる.

$$\text{修正精度} = \frac{\text{正しく誤りを修正できた前置詞の数}}{\text{誤りとして正しく検出された前置詞の数}} \quad (4)$$

なお挿入誤りは, 正しく検出すれば修正となるため, 評価は検出のみで修正の評価しない.

### 5.2 SemEval コーパス [15] による前置詞誤りの検出と修正の精度評価

テストデータは, SemEval コーパスの検証用データから無作為に抜き出した 50 文を用いる. ここでは, この 50 文へ 3 種類の前置詞誤りを加えた以下の各 50 文, つまり 3 種類の前置詞誤りをそれぞれ含む計 150 文を作成する. この 150 文には, 正しいものと誤っているもの合わせて 510 の前置詞が含まれている.

- 置換誤りを含む 50 文

各文の無作為に選んだ前置詞一か所を表 1 の 34 種類いずれかの前置詞に変更した 50 文

- 挿入誤りを含む 50 文

各文の無作為な位置に表 1 の 34 種類いずれか前置詞を一つ挿入した 50 文

- 欠落誤りを含む 50 文

各文の無作為に選んだ前置詞一か所を削除した 50 文

提案手法を利用した英文前置詞誤りの検出と修正の正誤判定は以下である.

- 検出の正誤判定

提案手法が入力文の誤り箇所, 誤りを検出し修正をしたとき, 前置詞誤りを検出したとする. 置換誤りを挿入誤りとして検出するとき, 挿入誤りを置換誤りとして検出するときも検出したと判定する. 欠落誤りの検出では, 欠落誤り以外の箇所欠落誤りを検出したときは誤検出となる.

表 2 評価用 150 文に対する置換誤りの検出結果

置換した前置詞 (50)		その他の前置詞 (460)		
検出		非検出	非検出	誤検出
置換誤り	挿入誤り			
41	0	9	450	10

表 3 評価用 150 文に対する挿入誤りの検出結果

挿入した前置詞 (50)		その他の前置詞 (460)		
検出		非検出	非検出	誤検出
挿入誤り	置換誤り			
45	2	3	449	11

表 4 評価用 150 文に対する欠落誤りの検出結果

削除した前置詞の箇所 (50)		前置詞が不要な箇所 (2426)	
検出	非検出	非検出	誤検出
43	7	2419	7

表 5 前置詞誤りの検出性能

	再現率	適合率	F 値
置換誤り	0.820	0.804	0.812
挿入誤り	0.940	0.810	0.870
欠落誤り	0.860	0.860	0.860

表 6 検出した置換誤りと欠落誤りに対する修正性能

	修正できた	修正できなかった	修正精度
置換誤り	33	8	0.805
欠落誤り	35	8	0.814

#### ● 修正の正誤判定

置換誤りと欠落誤りの修正の場合は、提案手法が正しい前置詞に修正した場合に修正したと判定する。挿入誤りの修正については、検出すれば、修正もできたとする。

表 2 に、510 の前置詞を含む評価用 150 文を使った実験における、置換誤りの 50 の前置詞とその他の 460 の前置詞に対する置換および挿入誤りの自動検出結果を示す。

表 3 に、510 の前置詞を含む評価用 150 文を使った実験における、挿入誤りの 50 の前置詞とその他の 460 の前置詞に対する挿入および置換誤りの自動検出結果を示す。

表 4 に、評価用 150 文を使った実験における、欠落誤り 50 箇所と誤りのない 2426 箇所に対する欠落誤り自動検出結果を示す。評価する計 2476 箇所は、[14] の方法で各文に [MASK] トークンが挿入され、生成された。

表 5 に、3 種類の前置詞誤りの検出性能をまとめる。表 5 に示すように、すべての種類の前置詞誤り検出の結果で、再現率と適合率ともに 0.8 以上だった。

表 6 に、前置詞誤りとして検出できた前置詞に対する自動修正の結果を示す。置換誤りと欠落誤りで、修正精度は 0.8 以上を示した。

前置詞誤りの検出は、MLM の問題から求まる前置詞の出現確率と前置詞の有無の分類問題から求まる確率を利用する。そのため誤り検出の閾値を変化させ、これらの確率による検出の F 値の変化を示す。評価のため使用する文は、同じ評価用 150 文である。

置換誤りでは、MLM の問題から求まる対象の前置詞の出現確率に対して、誤り検出の閾値を変化させる。ファインチューニングなしで MLM で事前学習済み BERT と、MLM の問題でファインチューニングした BERT を対象にする。二つのモデルを比較することで、ファインチューニングの効果を検証する。また置換誤りの検出は、対象の前置詞の出現確率と、異なる前置詞で最も出現確率の高い前置詞の出現確率の二つを基にしていた。そこで対象の前置詞の出現確率のみを利用する場合と、対象の前置詞の出現確率と異なる前置詞の出現確率の二つを利用する場合に対して、誤り検出の閾値を変化させる。これにより、置換誤りの検出で異なる前置詞の出現確率を利用する効果を検証する。なお、ここでの置換誤りの検出は、前置詞一か所に対して置換誤りと挿入誤り両方を同時に検出することはない。そのため表 3 で挿入誤りと検出した前置詞の一部が、置換誤りとしても検出している。検出結果の一部の変化により、F 値が表 5 と図 4 で異なっている。

図 3 は、置換誤りの検出で、対象の前置詞の出現確率に対して閾値を変化させたときの F 値の変化を示す。対象の前置詞の出現確率が閾値以下のとき、置換誤りを検出する。F 値の最大値は、MLM の問題でファインチューニングした BERT で 0.681、ファインチューニングなしで MLM で事前学習済み BERT で 0.534 だった。

図 4 は、置換誤りの検出で、対象の前置詞とは異なる前置詞の出現確率が一定以上のとき検出するという条件も付けて、対象の前置詞の出現確率の閾値を変化させたときの検出の F 値の変化を示す。対象の前置詞とは異なる前置詞の出現確率の条件は、MLM の問題でファインチューニングした BERT モデルで 0.25 以上、ファインチューニングなしで MLM で事前学習済み BERT で 0.1 以上である。なお事前学習済み BERT で 0.1 以上という条件は、[14] の置換誤りの検出での、異なる前置詞に対する条件である。対象の前置詞の出現確率が閾値以下のとき、置換誤りを検出する。F 値の最大値は、MLM の問題でファインチューニングした BERT で 0.789、ファインチューニングなしで MLM で事前学習済み BERT で 0.832 だった。

挿入誤りと欠落誤りでは、前置詞の有無の分類問題と MLM の問題の両方を利用し、誤りを検出する。そこで、前置詞の有無の分類問題で得る [MASK] トークンに前置詞が出現する確率、MLM で得る出現確率上位五つの前置詞の出現確率の和、この二つの確率を乗算した値の三つの閾値を変化させる。

図 5 は、挿入誤りの検出で、三つの閾値をそれぞれ変化させたときの F 値の変化を示す。三つの値それぞれが閾値以下のとき、挿入誤りを検出する。F 値の最大値は、前置詞の有無の分類問題で得る [MASK] トークンに前置詞が出現する確率で 0.765、MLM で得る出現確率上位五つの前置詞の出現確率の和で 0.811、前置詞の有無の分類問題と MLM で得る二つの確率を乗算した値で 0.874 だった。

図 6 は、欠落誤りの検出で、三つの閾値をそれぞれ変化させたときの F 値の変化を示す。三つの値それぞれで閾値以上のとき、欠落誤りを検出する。F 値の最大値は、前置詞の有無の分類問題で得る [MASK] トークンに前置詞が出現する確率で

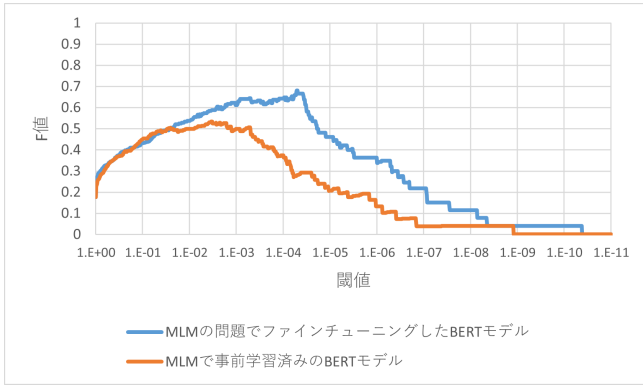


図 3 置換誤りの検出における出現確率の閾値と検出の F 値（対象前置詞の出現確率のみによる検出）

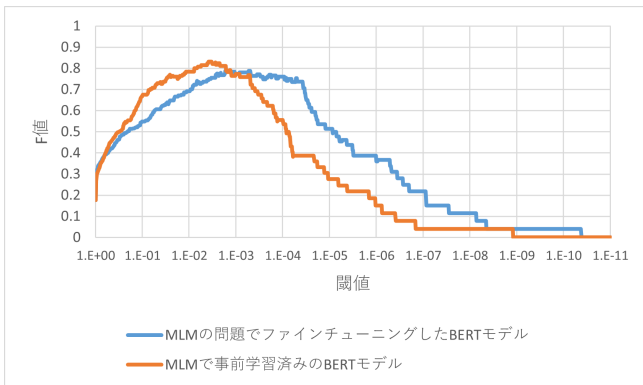


図 4 置換誤りの検出における出現確率の閾値と検出の F 値（出現確率が一定以上の他の前置詞がある場合）

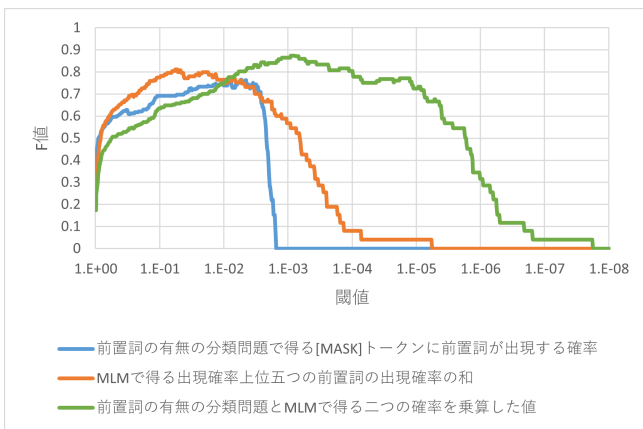


図 5 挿入誤り検出における出現確率の閾値と検出の F 値

0.880, MLM で得る出現確率上位五つの前置詞の出現確率の和で 0.595, 前置詞の有無の分類問題と MLM で得る二つの確率を乗算した値で 0.889 だった。

### 5.3 Grammarly との前置詞誤りの検出と修正の性能比較

Grammarly [5] は、入力した英文の文法やスペルの修正、ワードチョイスの提案を受けられるオンラインサービスである。英文を入力すると、修正が必要と予想される箇所やよりよい表現に書き換えられる箇所が、修正の種類によって 4 色に色分けされて表示される。例えば、赤色の箇所は文法やスペルの間違

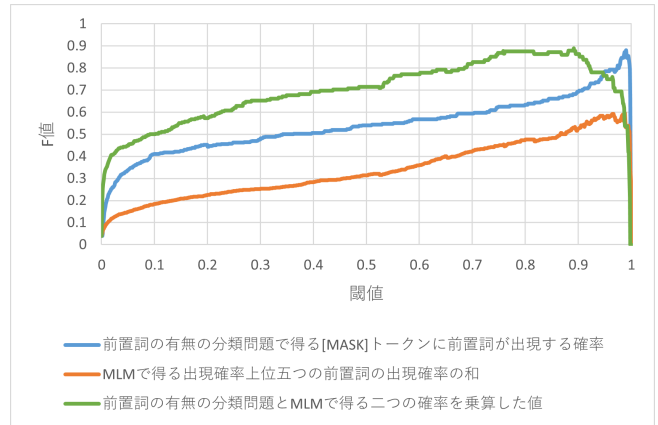


図 6 欠落誤り検出における出現確率の閾値と検出の F 値

い、青色の箇所は理解しにくい表現である。色の付いた箇所をクリックすると表示される修正候補を選択することで、文章を修正できる。無料版とプレミアム版でサービス内容に違いがある。Grammarly の無料版では文法、スペル、句読点の位置などを修正できる。プレミアム版ではそれらに加えて可読性の高い表現の提案、修正理由の解説、盗用チェックなどのサービスがある。他のアプリとの連携も可能であり、ブラウザの拡張機能やワープロソフトのアドインとして利用することで、英文を書くとき修正の提案を受けることができる。

Grammarly と提案手法で、前置詞誤りの検出と修正の精度を比較する。Grammarly は、機能が多くのプレミアム版を使用する。テストデータは 5.2 節で使用した 3 種類の前置詞誤りをそれぞれ 50 文ずつ含む 150 文である。Grammarly による前置詞誤りの自動検出、自動修正の正誤判定は以下の通りとする。

- 自動検出の正誤判定

置換誤りと欠落誤りでは、入力文の前置詞の誤り箇所で修正候補が提示されたとき、前置詞誤りを検出したとする。挿入誤りでは、挿入した前置詞の削除を提示されたとき、検出したとする。なお、誤りのない箇所では前置詞の修正が提示されたときは誤検出となる。

- 自動修正の正誤判定

置換誤りと欠落誤りでは、Grammarly の提示する誤り箇所の修正候補に正解前置詞がある場合、修正したとする。挿入誤りでは、挿入した前置詞の削除を提示されたとき、修正したとする。

表 7 に、前置詞誤り評価用 150 文に対する、提案手法と Grammarly の前置詞誤りの自動検出結果を示す。提案手法と Grammarly を比較すると、置換誤りデータで置換誤りを検出した文の数は、提案手法が 41, Grammarly が 24 であり、誤検出した文の数は、提案手法が 2, Grammarly が 1 である。3 種類の前置詞誤りすべてで、提案手法のほうが誤りを正しく検出した文の数は多い。一方、提案手法のほうが誤検出した文の数も多い。

表 8 に、提案手法と Grammarly の置換誤りと欠落誤りの修正性能を示す。表 8 に示すように、提案手法が、置換誤りの修正精度で 1.3 ポイント、欠落誤りの修正精度で 14.7 ポイント

表 7 前置詞誤り評価用 150 文に対する前置詞誤りの検出結果

		置換誤りを含む 50 文		挿入誤りを含む 50 文		欠落誤りを含む 50 文	
		提案手法	Grammarly	提案手法	Grammarly	提案手法	Grammarly
誤りの検出に成功した文	誤検出なし	39	23	33	39	40	29
	誤検出あり	2	1	14	2	3	1
誤りの検出に失敗した文	誤検出なし	9	26	3	8	5	19
	誤検出あり	0	0	0	1	2	1

表 8 検出した置換誤りと欠落誤りに対する修正性能

		修正できた	修正できない	修正精度
置換誤り	提案手法	33	8	0.805
	Grammarly	19	5	0.792
欠落誤り	提案手法	35	8	0.814
	Grammarly	20	10	0.667

表 9 評価用 150 文で誤検出なく正しく修正した文の数

	提案手法	Grammarly
置換誤りを含む 50 文	32	18
挿入誤りを含む 50 文	32	39
欠落誤りを含む 50 文	32	20

Grammarly を上回っている。

誤りを含む文に対して、誤検出せずに誤りのみ正しく修正した文が英文修正で最も望まれる結果である。表 9 に、提案手法と Grammarly の、誤りを正しく検出して正しく修正し、かつ誤検出もない文の数を示す。置換誤りでは、提案手法が 50 文中 32 文、Grammarly が 50 文中 18 文がこれに該当した。置換誤りと欠落誤りで、提案手法のほうが誤検出せずに正しく修正した文の数は多い。

#### 5.4 CoNLL-2014 Shared Task データセット [1] における Grammarly との検出と修正の精度比較

ここでは文法誤り訂正のベンチマークデータセットである CoNLL-2014 Shared Task データセット [1] の問題を用いて、Grammarly と提案手法で前置詞誤りの検出と修正の性能を比較する。このデータセットは、英語を第二言語とする大学生が書いた文章と、英語を専門とする大学教員がそれを手作業で修正した文章で構成されている。このデータセットには 28 種類の文法誤りが含まれており、誤りの種類ごとにタイプが付与されている。前置詞誤りは、“Prep” のタイプになる。実験に使用するデータは、“Prep” のタイプの英文から、3 種類の誤りを含む文を無作為にそれぞれ 50 文ずつ抜き出したものである。各種類の誤りを含む英文を入力とし、5.2 節と 5.3 節と同様に 3 種類の誤りに対する自動検出と修正の性能を評価する。

表 10 に、前置詞誤り評価用の 150 文に対する、提案手法と Grammarly の前置詞誤りの自動検出結果を示す。提案手法と Grammarly を比較すると、3 種類の前置詞誤りすべてで、Grammarly のほうが誤検出した文の数は少ない。誤りを検出した文の数も Grammarly のほうが多い。

表 11 に、提案手法と Grammarly の置換誤りと欠落誤りの修正性能の比較を示す。表 11 に示すように、Grammarly が、置換誤りの修正精度で 2.6 ポイント、欠落誤りの修正精度で 0.3

ポイント提案手法を上回っている。

表 12 に、提案手法と Grammarly の、誤りを正しく検出して正しく修正し、かつ誤検出もない文の数を示す。3 種類の前置詞誤りすべてで、Grammarly のほうがこれに該当する文の数は多かった。

#### 5.5 考 察

図 3 から、MLM の問題でファインチューニングする前後のモデルを比較する。ファインチューニング後のモデルの F 値が最大になる閾値は、ファインチューニング前より小さい値である。MLM の問題によるファインチューニングにより、誤っている前置詞の出現確率が小さくなる傾向があることが確認できる。図 5 と図 6 から、前置詞の有無の分類によるファインチューニングのみで、[MASK] トークンに前置詞が入るかどうかが分類できることを確認できる。また、前置詞の有無の分類問題の結果だけでなく、MLM の出現確率という別の基準と組み合わせることで、F 値の向上を確認できる。図 4 では、対象の前置詞の出現確率のみでなく、異なる前置詞の出現確率を組み合わせることで F 値の向上を確認できる。しかし、MLM の問題でファインチューニングしたモデルのほうが F 値の最大値は小さい。そのため、異なる前置詞の出現確率ではない別の基準を組み合わせることで、F 値が向上する可能性がある。

表 9 から、ファインチューニングに使用した SemEval コーパスの検証用データセットでは提案手法が Grammarly を一部上回った結果を確認できる。しかし、表 12 の CoNLL-2014 Shared Task データセットでは提案手法の性能は Grammarly に完敗した。ファインチューニング用データの前置詞誤りはランダムに選んでいた。そのため、例えば置換誤り用の文として、“on” を “among” に置換した “I play tennis among Sunday.” のような、人であればまずしない誤りを含む文も生成されていた。一方、CoNLL-2014 Shared Task データセットの前置詞誤りは第二言語話者の実際の誤りである。そのためファインチューニング用に作成した文が、実在する誤りに即していない可能性がある。ファインチューニングのデータセットを、現実の前置詞誤りデータに変えることを検討したい。

## 6 ま と め

本稿では、BERT を利用して、英文前置詞誤りを自動で検出して修正する手法を提案した。提案手法ではまず、マスクした前置詞を Masked Language Model (MLM) で予測する問題とマスクした箇所に前置詞が挿入されるかどうかの分類問題それぞれで、BERT をファインチューニングする。次に [MASK]

表 10 CoNLL-2014 Shared Task データセットにおける前置詞誤り評価用 150 文に対する前置詞誤りの検出結果

		置換誤りを含む 50 文		挿入誤りを含む 50 文		欠落誤りを含む 50 文	
		提案手法	Grammarly	提案手法	Grammarly	提案手法	Grammarly
誤りの検出に成功した文	誤検出なし	20	32	24	42	26	30
	誤検出あり	1	1	3	2	7	4
誤りの検出に失敗した文	誤検出なし	23	16	19	6	12	15
	誤検出あり	6	1	4	0	5	1

表 11 CoNLL-2014 Shared Task データセットにおける検出した置換誤りと欠落誤りに対する修正性能

		修正できた	修正できない	修正精度
置換誤り	提案手法	16	5	0.762
	Grammarly	26	7	0.788
欠落誤り	提案手法	30	3	0.909
	Grammarly	31	3	0.912

表 12 CoNLL-2014 Shared Task データセットにおける評価用 150 文で誤検出なく正しく修正した文の数

	提案手法	Grammarly
置換誤りを含む 50 文	16	25
挿入誤りを含む 50 文	23	42
欠落誤りを含む 50 文	23	28

トークンを含む文を入力し、MLM による [MASK] トークンに対応する単語の出現確率と、前置詞の有無の分類問題による [MASK] トークンに前置詞が入る確率を求める。そして二つの確率を利用して前置詞誤りを自動で検出して修正する。

実験では、提案手法による前置詞誤りの検出と修正の性能を確かめた。さらに、提案手法と英文修正サービスの Grammarly と検出と修正の性能を比較した。SemEval コーパスの検証用データセットでは、誤検出なく誤った前置詞のみを正しく修正した文は、置換誤りを含む文で提案手法は 50 文中 32 文、Grammarly は 50 文中 18 文であり、欠落誤りを含む文で提案手法は 50 文中 32 文、Grammarly は 50 文中 20 文で、提案手法が Grammarly を上回った。一方、挿入誤りを含む文では、提案手法は 50 文中 32 文、Grammarly は 50 文中 39 文となり、提案手法が Grammarly を下回った。CoNLL-2014 Shared Task データセットに対する同様の実験では、3 種類の前置詞誤りすべてで提案手法が Grammarly を下回った。

今後の課題として、英語学習者の実際の前置詞誤りを含むデータセットを利用した検出精度の向上が挙げられる。また前置詞以外の品詞誤りの検出と修正も検討したい。具体的には、前置詞と同様に種類が少なく第二言語話者の間違いが多い冠詞に対して、提案手法のアプローチが有効かどうか評価することを考えている。

## 文 献

[1] Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadwinoto, Raymond Hendy Susanto, and Christopher Bryant, The CoNLL-2014 Shared Task on Grammatical Error Correction. In Proceedings of the Eighteenth Conference on

Computational Natural Language Learning: Shared Task, pages 1–14, 2014.

[2] Christopher Bryant, Mariano Felice, Øistein E. Andersen, and Ted Briscoe. The BEA-2019 Shared Task on Grammatical Error Correction. In Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications, pages 52–75, 2019.

[3] 中西淳, 学習者コーパスを用いた日本人英語学習者の前置詞誤用パターンの抽出. 統計数理研究所共同研究レポート, 400 83-96, 2018.

[4] 須田考司, 岡村明夢, 高等学校における英語教育が日本人英語学習者の前置詞の習得に与える影響. 国際関係・比較文化研究 17 (2), 1-11, 2019.

[5] Grammarly, <https://app.grammarly.com/>

[6] Ginger, <https://www.getginger.jp/.com/>

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” arXiv:1810.04805, 2018.

[8] Alex Warstadt, Amanpreet Singh, Samuel R. Bowman, “Neural Network Acceptability Judgments,” arXiv:1805.12471, 2018.

[9] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, Percy Liang, “SQuAD: 100,000+ Questions for Machine Comprehension of Text,” arXiv:1606.05250, 2016.

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser and Illia Polosukhin, “Attention Is All You Need,” In Advances in Neural Information Processing Systems, Vol. 30, pp. 5998–6008, 2017.

[11] 松本悠太, 清野舜, 乾健太郎, 高再現率な文法誤り訂正システムの実現に向けて. 言語処理学会第 27 回年次大会, pp. 1475-1480, 2021.

[12] 有富隼, 太田学, 検索エンジンによる英文前置詞誤り修正支援, DBSJ Journal Vol. 9, No. 1, pp. 70-75, 2010.

[13] Masahiro Kaneko, Kengo Hotate, Satoru Katsumata, and Mamoru Komachi, TMU Transformer System Using BERT for Re-ranking at BEA 2019 Grammatical Error Correction on Restricted Track. In Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications, pages 207–212, 2019.

[14] 中谷聡, 上野史, 太田学, BERT による英文前置詞誤り修正支援の一手法, DEIM2022, H41-5, 2022.

[15] Kenneth C. Litkowski and Orin Hargraves, SemEval-2007 Task 06: Word-Sense Disambiguation of Prepositions. In Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007), pages 24–29, 2007.