

学修番号 18860611

修士論文

言語学習者向けの文法誤り検出機能付き 誤用例文検索システムの構築

新井 美桜

2020年2月21日

首都大学東京大学院
システムデザイン研究科 情報科学域

新井 美桜

審査委員：

小町 守 准教授 (主指導教員)

山口 亨 教授 (副指導教員)

高間 康史 教授 (副指導教員)

言語学習者向けの文法誤り検出機能付き 誤用例文検索システムの構築*

新井 美桜

修論要旨

第二言語学習者が作文を執筆する際に有用なツールとして、用例検索システムが挙げられる。用例検索システムとは、多数の用例を収録したコーパスから、文や語句などのユーザの検索クエリに応じて用例を提示するシステムであり、学習者は語句の使用例を確認することで作文執筆に役立てることができる。

学習者は作文を執筆する際、使い方が分からない語句を用例検索システムで検索する。このときクエリが正しければ検索できるが、そもそも学習者の想像している語句が間違っており、クエリが誤用である場合は、正しく検索できず用例を確認できない。また、誤用例の検索ができてそれに対する添削がなければ、それがどのような誤りであり、どのように直すべきか分からない。更に、誤用例と共に正用例を確認できたとしても、学習者自身が誤りを認識していない場合、どの部分を直すべきか分からず混乱を招く可能性がある。既存の学習者向けの用例検索システムでは、正用例のみあるいは誤用例のみの検索しかできない。また用例の総数が少なく、学習者が十分な情報を取得できない場合がある。更に、既存の用例検索システムは主に言語教育者向けに作られたものが多く、その言語の予備知識がない人は活用できないという問題もある。

学習者に文中の直すべき部分を知らせるために、文法誤り検出システムを用いることが考えられる。第二言語としての英語学習者のための文法誤り検出は広く研究されているが、第二言語としての日本語学習者のための文法誤り検出に関する研究は少なく、しかもそのほとんどは、特定の誤りタイプのみ焦点を当てている。関連する日本語の文法誤り訂正タスクでは、誤りタイプを限定していないものとして水本らの統計的機械翻訳を用いた研究がある。ただし、統計的機械翻訳ベースの文

*首都大学東京大学院 システムデザイン研究科 情報科学域 修士論文, 学修番号 18860611, 2020年2月21日.

法誤り訂正・検出は文字または単語 n -gram に依存しているため、長距離関係を考慮できず、長距離関係に依存する誤りは検出できない。

本研究では、誤用例を提示することの有用性に着目し、第二言語学習者のためのニューラルネットワークを用いた文法誤り検出機能付き大規模誤用例文検索システムを構築した。用例データには、相互添削型 SNS である Lang-8 のデータをもとに作られた、学習者の書いた誤用文に母語話者の書いた添削文が付与されているデータセットを用いた。このデータセットを利用することで、学習者は自身の文がどのように間違っているか分かり、どのように直すかの判断の参考にできる。提案システムは、文法誤り検出機能により入力文の誤りを自動で検出することで学習者に誤りを気づかせることができ、かつ、正用例付き誤用例を提示することで学習者自身がどのように自分の作文を訂正するかを判断できる検索システムとなっている。このシステムは、検出された誤りをクエリとして全文検索を行い、入力クエリと検索結果のベクトルのコサイン類似度を測り、その高い順に用例を表示する。これにより、学習者は自分の書いた文に近い用例を確認することができる。本研究の主な貢献は以下のとおりである：

- ニューラルネットワークを用いた文法誤り検出を日本語に適用した結果を報告する最初の研究である。比較手法である統計的機械翻訳を用いた手法と比べて高い検出精度を達成している。
- 提案システムは文法誤り検出機能を持つ最初の用例検索システムである。この機能により、ユーザはクエリのどの部分が間違っているのかを認識できる。
- 学習者の書いた誤り文である誤用例とそれに対応する添削文を対にして提示するため、学習者は用例を参考にして誤りを修正できる。
- 提案システムと比較システムを用いて評価実験を行い誤りタイプを分析した結果、提案システムは誤形成や脱落、余剰の誤りに対応できることを示した。
- 提案システムを使用した被験者実験で言語学習者のライティングスコアが向上することを示した。

本稿の構成は以下のようになっている。第 1 章では本研究の概要について述べる。第 2 章では第二言語学習者向けの用例検索システムと文法誤り検出の関連研究を説明する。第 3 章では構築した日本語および英語学習者向けの誤用例文検索シス

テムとそれに含まれる文法誤り検出システムについて説明する。第 4 章では 3 章で説明した日本語検索システムの評価実験について述べる。第 5 章では 3 章で説明した英語検索システムの評価実験について述べる。最後に、第 6 章で本研究のまとめを行う。

Incorrect Example Retrieval System with Grammatical Error Detection for Second Language Learners*

Mio Arai

Abstract

A standard method that supports the effort of learning a second language is the use of example sentences. Example retrieval systems in particular check for the appropriate use of words based on the context in which they are written. However, in such a system, if the query word is incorrect, to find appropriate examples is impossible using ordinary search engines, such as Google. Even if learners have access to an incorrect example retrieval system, they do not know how to search for the examples because they do not know whether their query includes errors. Moreover, they are often unable to modify composition in the absence of correct versions of the incorrect examples. These systems are primarily developed for use by language teachers. As such, they are not as helpful for learners who do not have a strong background.

The grammatical error detection system is one of the methods for making a learner know a part to be corrected in a sentence. Grammatical error detection for learners of English as a second language (ESL) is widely studied. However, there are few studies on grammatical error detection for learners of Japanese as a second language (JSL). Most studies on grammatical error detection in Japanese focus on a learner's particular error types, mainly with particles. Among others, there are studies using phrase-based statistical machine translation (PBSMT), which does not limit the types of grammatical errors made by a learner. How-

*Master's Thesis, Department of Computer Science, Graduate School of Systems Design, Tokyo Metropolitan University, Student ID 18860611, February 21, 2020.

ever, PBSMT-based grammatical error detection cannot consider long-distance relationships because it relies on either character or word n -grams. Neural networks can consider long-distance relationships because they can keep input histories as state variables.

Considering this, our study develops an example sentence retrieval system with neural grammatical error detection using the large-scale Lang-8 dataset for second language learners by focusing on the usability of automatic incorrect example retrieval. The main contributions of this work are as follows:

- This is the first study that tackles grammatical error detection in Japanese using a neural network. It shows the state-of-the-art F-score on the Lang-8 dataset and establishes a new baseline.
- To the best of our knowledge, our system is the first incorrect example sentence retrieval system using neural grammatical error detection. This function allows a user to recognize which part of the query is wrong.
- Our system seamlessly shows the incorrect sentences, and the corresponding sentences corrected by a native speaker. Thus, learners can rectify their mistakes while writing the composition.
- Our intrinsic evaluation shows that our system is good at correcting lexical choice and misformation errors in a learner's writing. Our extrinsic evaluation also shows that our example retrieval system improves the quality of a learner's writing.

目次

図目次	viii
第 1 章 はじめに	1
第 2 章 関連研究	3
2.1 日本語学習者向けの用例検索システム	3
2.2 英語学習者向けの用例検索システム	4
2.3 文法誤り検出	4
第 3 章 第二言語学習者向けの文法誤り検出機能付き誤用例文検索システム	6
3.1 日本語学習者向けのシステム	6
3.1.1 データセット	6
3.1.2 ユーザインタフェース	6
3.1.3 使用例	7
3.1.4 日本語学習者作文の文法誤り検出	8
3.1.5 例文検索におけるリランキング	10
word2vec	10
BERT	10
3.1.6 用例検索アルゴリズム	10
3.1.7 事前実験: 文法誤り検出	11
実験設定	11
実験結果	11
3.2 英語学習者向けのシステム	13

	データ	13
	システム	13
第 4 章	日本語検索システムの評価実験	14
4.1	内的評価	14
4.2	外的評価	17
第 5 章	英語検索システムの評価実験	20
5.1	実験設定	20
5.2	結果	20
第 6 章	おわりに	23
	発表リスト	24
	謝辞	26
	参考文献	27

図目次

3.1	検索システムのユーザインタフェース。誤用検索された場合（左） と正用検索された場合（右）。	7
3.2	文法誤り検出モデルの構造。「私の元気です」という文が入力され た場合の例。*	9
3.3	英語学習者向けシステムのユーザインタフェース。	13

第1章 はじめに

近年、言語学習者向けの作文支援システムが多く開発されている。その中には、学習者作文の文法誤りを自動で訂正するシステムや、学習者作文を自動評価し点数をつけるシステムなどがある。このような作文支援システムにより、学習者及び教育者の負担を減らす事ができる。言語学習者が作文する際、書きたいことが思いついていてもその適切な表現が思いつかないことがある。そのため、学習者が単語などのクエリを入力し、そのクエリが使われている例文を確認することは、学習者が作文する際に有用である。我々はそのような有用性に着目し言語学習者向けの用例検索システム開発タスクに取り組んだ。

用例検索システムとは、多数の用例を収録したコーパスから、文や語句などのユーザの検索クエリに応じて用例を提示するシステムであり、学習者は語句の使用例を確認することで作文執筆に役立てることができる。学習者は作文を執筆する際、使い方が分からない語句を用例検索システムで検索する。このときクエリが正しければ検索できるが、そもそも学習者の想像している語句が間違っており、クエリが誤用である場合は、正しく検索できず用例を確認できない。また、誤用例の検索ができてそれに対する添削がなければ、それがどのような誤りであり、どのように直すべきか分からない。更に、誤用例と共に正用例を確認できたとしても、学習者自身が誤りを認識していない場合、どの部分を直すべきか分からず混乱を招く可能性がある。既存の学習者向けの用例検索システム [1, 2] では、正用例のみあるいは誤用例のみの検索しかできない。また用例の総数が少なく、学習者が十分な情報を取得できない場合がある。更に、既存の用例検索システムは主に言語教育者向けに作られたものが多く、その言語の予備知識がない人は活用できないという問題もある。

学習者に文中の直すべき部分を知らせるために、文法誤り検出システムを用いることが考えられる。第二言語としての英語学習者のための文法誤り検出は広く研究されているが、第二言語としての日本語学習者のための文法誤り検出に関する研究は少なく、しかもそのほとんどは、特定の誤りタイプのみ焦点を当てている [3, 4]。関連する日本語の文法誤り訂正タスクでは、誤りタイプを限定していないものとして水本らの統計的機械翻訳を用いた研究がある [5]。ただし、統計的機械翻

訳ベースの文法誤り訂正・検出は文字または単語 n -gram に依存しているため、長距離関係を考慮できず、長距離関係に依存する誤りは検出できない。

本研究では、誤用例を提示することの有用性に着目し、第二言語学習者のためのニューラルネットワークを用いた文法誤り検出機能付き大規模誤用例文検索システムを構築した。用例データには、相互添削型 SNS である Lang-8* のデータをもとに作られた、学習者の書いた誤用文に母語話者の書いた添削文が付与されているデータセット [5] を用いた。このデータセットを利用することで、学習者は自身の文がどのように間違っているか分かり、どのように直すかの判断の参考にできる。提案システムは、文法誤り検出機能により入力文の誤りを自動で検出することで学習者に誤りを気づかせることができ、かつ、正用例付き誤用例を提示することで学習者自身がどのように自分の作文を訂正するかを判断できる検索システムとなっている。このシステムは、誤りが検出された場合は誤り箇所を、されなかった場合はユーザが入力した文をクエリとして全文検索を行い、入力クエリと検索結果の文ベクトルのコサイン類似度を測り、高い順に用例を表示する。本研究の主な貢献は以下のとおりである：

- ニューラルネットワークを用いた文法誤り検出を日本語に適用した結果を報告する最初の研究である。比較手法である統計的機械翻訳を用いた手法と比べて高い検出精度を達成している。
- 提案システムは文法誤り検出機能を持つ最初の用例検索システムである。この機能により、学習者はクエリのどの部分が間違っているのかを認識できる。
- 学習者の書いた誤り文である誤用例とそれに対応する添削文を対にして提示するため、学習者は用例を参考にして誤りを修正できる。
- 提案システムと比較システムを用いて評価実験を行い誤りタイプを分析した結果、提案システムは誤形成や脱落、余剰の誤りに対応できることを示した。
- 提案システムを使用した被験者実験で言語学習者のライティングスコアが向上することを示した。

*<http://lang-8.com>

第2章 関連研究

2.1 日本語学習者向けの用例検索システム

今日、様々な日本語学習者向けの用例検索システムが開発されている。しかし、未だ学習者が実用的に利用することは難しい部分が多く存在する。既存の日本語学習者の作文用例検索システムには以下のようなものがある。

李ら [1] は日本学習者の発話データを書き起こした資料に言語情報を付与した『タグ付き KY コーパス』に用例検索機能を追加し、公開した。『タグ付き KY コーパス』は検索したい語を入力するとコーパスから学習者の正用例、誤用例を検索できる。しかし、『KY コーパス』に含まれる誤用例は数が非常に少なく、頻繁に使用するであろう語句にも誤用例が存在しない場合が多くある。また誤用例に対応する正用例が存在しないため、学習者が作文に一致する誤用例を見つけても、訂正の仕方が分からず、参考にすることが難しい。また、そもそも入力したクエリが誤用であると学習者自身が認識していないと誤用検索を行えないため、学習者の利用が困難であると考えられる。

東京外国語大学のコーパスに基づく言語学教育研究拠点は『国際日本語学習者作文コーパス及び誤用辞典*』を構築した。このシステムは検索したい学習者の母語や年齢、性別などを選択するとその条件にあった文を検索ワードとともに検索でき、検索したキーワードに一致した文が KWIC (keyword in context) 形式で表示される。日本語教育者はこの検索システムを使うことで、学習者の誤りの特徴を得ることができる。この検索システムは教育者を主な対象にしたもので、学習者が扱いきやすいような用例の表示の仕方はしていない。誤用検索を行い誤り箇所とそれに対する訂正文を表示する機能もあるが、総文数が非常に少ないため、頻出単語の誤用であっても誤用例が存在しない場合が多い。また、文法誤り検出機能を持たないためユーザは自分の検索ワードに誤りが含まれているかどうかを認識できず、どのように検索すべきか分からない。

本研究では、学習者の書いた誤用文に母語話者の書いた添削文が付与されている Lang-8 データセット [5] を用いることで、添削文付きの誤用例検索を可能にし、か

* http://ngc2068.tufs.ac.jp/corpus_ja/

つ、誤り検索機能を追加し、学習者が自身の文に含まれる誤りを認識していない場合における誤用検索も可能にした。

2.2 英語学習者向けの用例検索システム

最も一般的な例文検索システムは、Web ベースの検索エンジンである。しかしながら、Web ベースの検索エンジンは言語学習者のために例文を検索するわけではないので、誤用検索や文法誤り検出の機能はない。以下に主要な英語学習者向けの用例検索システムを記す。いずれも作文執筆を支援する機能を持つが、学習者を支援するための文法誤り検出や誤用検索の機能は持たない。

FLOW [6] は、英語学習者が母語で文を書く際に、言い換えを利用してリアルタイムに英単語の候補を提示するシステムである。このシステムとは対照的に、我々のシステムは訂正済みの学習者コーパスに基づいて、学習者が書いた誤りを含む文とそれに対応する訂正文を対にして提示する。

WriteAhead [7] は英語学習者が英語の文を書く際にリアルタイムで用例を提示する。このシステムは、コロケーションと構文に基づくパターンを用いて用例を提案する。例えば、ユーザが “We discussed,” と書くと、システムは “discuss” という単語が使用されているパターンである “discuss in *something*” や “discuss with *something*” などを含む用例を表示する。我々のシステムは、単語 n -gram に基づいて用例を検索し、検索結果の用例のベクトルとクエリのベクトルのコサイン類似度を計算し、クエリに類似した用例をより上位に表示する。

Sketch Engine [8] は、単語に関連付く文法構成要素をソーラスの情報と共に表示する。前述のように、我々のシステムは、生コーパスから抽出された母語話者の書いた正しい例と共に、学習者コーパスを使用して学習者の書いた誤りを含む例を提示する。

2.3 文法誤り検出

英語の文法誤り検出タスクではニューラルネットワーク、特に Bi-directional Long Short-Term Memory (Bi-LSTM) [9] を用いた手法が盛んに研究されている

[10, 11, 12, 13]。また、関連する英語の文法誤り訂正タスクでは、統計的機械翻訳を用いた手法 [14] やニューラルネットワークを用いた手法 [15] が提案されている。日本語の文法誤り検出の研究は、学習者の誤り種類を限定したもの、主に助詞に限定したものがほとんどである [16, 3, 4, 17]。関連する日本語の誤り訂正タスクでは、学習者の誤り種類を限定していないものとして水本らの統計的機械翻訳を使用した研究がある [5]。その一方で、日本語ではニューラルネットワークを用いた文法誤り検出の研究は少ない。

本研究では、誤り種類を限定せず、ニューラルネットワークを用いての誤り検出器の構築に臨んだ。統計的機械翻訳は n -gram ベースであるため長距離の関係を考慮することができないが、ニューラルネットワークは入力履歴を状態変数として保持しておけるため長距離の関係を考慮することができる。近年、ニューラルネットワークベースの手法は、文法誤り訂正において統計的機械翻訳ベースの方法よりも優れた結果を出している [18, 15]。それらは文法誤り検出においても同様だと期待できる。

第 3 章 第二言語学習者向けの文法誤り検出機能付き誤用例文検索システム

3.1 日本語学習者向けのシステム

文法誤り検出機能と用例検索機能を組み合わせた日本語学習者向けの用例検索システムを構築した。本システムは、日本語学習者であるユーザの書いた文が誤りを含む可能性があるとして仮定し、ユーザの入力に文法誤り検出を施す。誤りが検出された場合、その誤りをクエリとして誤用例文検索を行う。それ以外の場合はユーザの入力全文がクエリとして正用例文検索によって処理される。

3.1.1 データセット

本研究では水本らによって作成された学習者コーパスである Lang-8 Learner Corpora [5] を使用した。Lang-8 Learner Corpora は学習者の作文であるエッセイとその添削、エッセイ ID、ユーザ ID、学習言語タグ、母語タグからなる。学習言語のタグが Japanese のエッセイ数は 185,991 編である。全ての学習者の文には 1 つ以上の添削文が付与されている。添削文が 1 つの場合は 1 つの学習者文と 1 つの添削文を文対とした。1 つの学習者文に 2 つ以上の添削文が付いている場合は、それぞれの添削文と学習者文を文対とした。本研究では、Lang-8 Learner Corpora に含まれる、日本語学習者が書いた添削前の文である誤用例と日本語母語話者が添削した文である正用例のペア 140 万文対を抽出し、この内誤用の長さが 100 文字以下であり、かつ、誤用と正用間の編集距離が 7 文字以下である 72 万文対を用いた。

3.1.2 ユーザインタフェース

図 3.1 は提案する日本語学習者向けの用例検索システムのユーザインタフェースを表している。ユーザインタフェースは 2 種類あり、図 3.1 の左に検索クエリに文法誤りが含まれていて誤用検索された場合、図 3.1 の右に検索クエリが正しい文であり正用検索された場合を示す。以下にそれぞれの要素の説明を記す。



図 3.1: 検索システムのユーザインタフェース。誤用検索された場合（左）と正用検索された場合（右）。

1. 検索窓 ここに入力された検索クエリを検索する。検索クエリは文または複数単語を想定する*。
2. 文法誤り検出結果 誤り検出を行った結果が表示される。検出された誤り部分は赤く表示される。
3. 検索結果 検索クエリに適合した検索結果が表示される。検索結果は訂正箇所が太字で強調された誤用例と正用例の文対である。文法誤り検出の結果検索されたクエリが書かれた緑色のボタンをクリックすると、そのクエリに適合する全ての用例が展開される。

3.1.3 使用例

日本語学習者にとって助詞は非常に誤りやすい項目であるため、使用例として、学習者であるユーザが助詞を誤った文をクエリとして入力した場合を示す。

1. ユーザが「日本語が勉強する。」と検索窓に入力する。
2. システムは誤り検出を行い、「が」が誤りだと認識し、図 3.1 の左のように助

*適用する文法誤り検出システムの最低入力語数が 2 語であるため。

表 3.1: 正誤ラベルの例。ラベル c はターゲットの単語が正しいことを表し、ラベル i はターゲットの単語が誤っていることを表す。

文	いま	、	ぼく	は	がっ	く	が	と	も	い	そ	が	し	で	す	よ	。
ラベル	c	c	c	c	c	i	c	c	c	i	c	c	c	c	c	c	c

詞「が」の誤りを含み、「日本語が勉強する。」に類似した用例を提示する。

- 各誤用例には添削後の正用例が付いているので、ユーザは「日本語を勉強する。」が正しい使い方だと把握することができる。

3.1.4 日本語学習者作文の文法誤り検出

本研究では文法誤り検出を系列ラベリングタスクとして扱う。入力文の各単語に正誤ラベルを割り当てた。3.1.1 項で説明した誤用例文と正用例文の文対から動的計画法を用いて単語アラインメントをとり、一致しているところを c、一致していないところを i とした表 3.1 はラベル付けを行った例である。

Rei ら [11] と同様に、文法誤り検出のための Bi-LSTM を構築する。コードは GitHub にて公開している[†]。図 3.2 にモデルの構造を示す。システムは長さ T の単語系列 $[w_1..w_t...w_T]$ を入力として受け取り、最終的に各単語に対するラベルを予測する。まず、全ての単語 w_t は単語ベクトル e_t^w と文字ベクトル e_t^c に変換される。文字ベクトルは、各文字を文字分散表現に変換した後に Bi-LSTM に入力し、その先頭と文末の隠れ層を結合することで作成する。これらのベクトルは、学習時に更新される。また、この文字ベクトルを作成する LSTM を文字 LSTM と呼ぶ。そして、 e_t^w と e_t^c を結合することで t 番目の入力ベクトル \tilde{x}_t を作成する。入力ベクトルは Bi-LSTM を用いて以下のように隠れ層 h_t を計算する。

$$\vec{h}_t = \text{LSTM}(\tilde{x}_t, \vec{h}_{t-1}) \quad (3.11)$$

$$\overleftarrow{h}_t = \text{LSTM}(\tilde{x}_t, \overleftarrow{h}_{t+1}) \quad (3.12)$$

$$h_t = [\vec{h}_t; \overleftarrow{h}_t] \quad (3.13)$$

[†]https://github.com/kanekomasa/hiro/japanese_error_detection

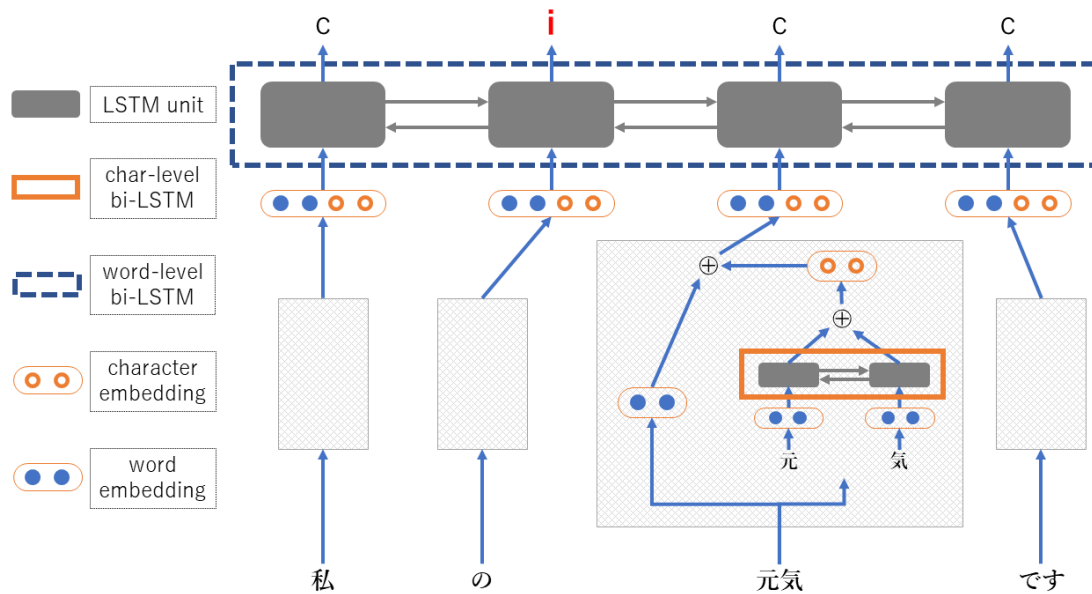


図 3.2: 文法誤り検出モデルの構造。「私の元気です」という文が入力された場合の例。[‡]

\vec{h}_t は順方向 LSTM、 \overleftarrow{h}_t は逆方向 LSTM、 h_t は両方向の隠れ層を結合したものである。また、 $h_0 \leftarrow = h_0$ 右 = 0 (ゼロベクトル) である。次に、全結合層を使用して Bi-LSTM と出力層の次元数の差を調整するための追加の隠れ層 d_t を計算する：

$$d_t = \tanh(W_d h_t) \quad (3.14)$$

ここで W_d は出力重み行列である。そして、出力層と softmax 関数を用いて予測を行う：

$$P(y_t | w_1 \dots w_T) = \text{softmax}(W_o d_t) \quad (3.15)$$

ここで W_o は出力重み行列であり、 y_t は予測ラベルである。

[‡]簡略化のため 3 番目の単語である「元気」にのみ単語ベクトル化と Bi-LSTM による文字ベクトル化、そしてその結合の処理を描いているが、実際には全ての単語でこの処理が行われる。

3.1.5 例文検索におけるリランキング

文法誤り検出結果による検索クエリによって抽出された例文のリランキング方法は、最初のクエリの文ベクトルと例文の文ベクトルのコサイン類似度の降順である。文ベクトルの生成方法はいくつか考えられるが、本研究では以下の代表的な2種類の手法を試みた。単語の意味を分散表現化することが可能な word2vec [19] を用いた手法と、近年自然言語処理で注目されている既存の様々なタスクに転移学習可能な大規模事前学習モデルである Bidirectional Encoder Representations from Transformers (BERT) [20] を用いた手法を使用した。

■word2vec 3.1.1 項と同様の前処理を行った Lang-8 データセットを用いて学習した word2vec モデルを使用して、各単語ベクトルの分散表現を算出し、その平均を文ベクトルとした。

■BERT 日本語 Wikipedia で事前学習済みのモデル [21] をファインチューニングを行わずそのまま使用する。このモデルの学習時の入力テキストは Juman++ (v2.0.0-rc2) [22] で形態素解析を行い、さらに Byte Pair Encoding (BPE) [23] を適用しサブワードに分割している。この設定に合わせて本研究においても入力テキストは Juman++ による分かち書きと BPE によるサブワード化を行い、トークンの系列に変換する。この系列を事前学習済みのモデルに入力し、最終層の全単語の隠れ層の平均を文ベクトルとした。

3.1.6 用例検索アルゴリズム

用例のデータベースには 3.1.1 項で構築したデータを用いる。ユーザの検索クエリを受け取り、データベースからその文に関連する誤用例と正用例の文対を抽出し、表示する。検索クエリは文または複数単語を想定している。

1. 3.1.7 項で説明した文法誤り検索システムを用いて検索クエリの誤りを検出する。
2. 誤りが検出された場合 その単語を誤用例に含む文を抽出する。誤りが複数箇所ある場合は全ての誤りに対して個別に抽出する。3.1.5 項で説明し

た通りに誤用例をリランキングし、正用例と共に表示する。

誤りが検出されなかった場合 クエリの単語 n -gram に基づくパターンを用いて正用例を抽出する。この時一致する単語 n -gram が多いパターンを含む用例ほど上位に表示する。特定のパターンを含む用例が複数存在する場合は、3.1.5 項で説明した通りにリランキングを行う。

3.1.7 事前実験: 文法誤り検出

文法誤り検出の事前実験には 3.1.1 項で説明したデータを用いた。このデータを学習データ 72,000 文と開発データ 1,000 文、評価データ 1,000 文の 3 つに分割して利用した。文法誤り検出の正解ラベルとして、自動で作成した誤りタグを使用した。評価は単語単位で行う。正確なフィードバックのほうがカバレッジの高い誤り検出より学習効果のあることが知られているため [24]、Precision を重視する $F_{0.5}$ で評価した。

■**実験設定** 形態素解析器には MeCab (ver. 0.996)[§]を使用し、辞書は UniDic (ver. 2.2.0) を使用した。文ベクトルを作るために gensim[¶]を使用した。ハイパーパラメータの設定は単語分散表現と単語 LSTM の次元数は 300、文字分散表現と文字 LSTM の次元数は 100 とした。adadelta (学習率: 1.0) でバッチサイズを 64 文として最適化した。

比較システムとして水本ら [5] の統計的機械翻訳 (SMT) システムを再実装した。単語-単語対応モデルを使用した。訂正結果から本文法誤り検出モデルと同様に動的計画法を用いて誤り検出モデルの学習のためのラベル付けを行った。Minimum Error Rate Training (MERT) [25] を用いて $F_{0.5}$ で最適化した。

■**実験結果** 各モデルにおける日本語文法誤り検出精度を表 3.2 に示す。Precision と Recall、 $F_{0.5}$ の全てにおいて提案モデルのほうが従来手法より上回っており、提案モデルのほうが優れていることがわかる。

表 3.3 に誤りタイプごとの True Positives (TP) と False Negatives (FN)、False

[§]<https://github.com/taku910/mecab>

[¶]<https://github.com/RaRe-Technologies/gensim>

表 3.2: 日本語文法誤り検出モデルの精度の比較。

モデル	Precision	Recall	F _{0.5}
SMT モデル	0.599	0.121	0.333
提案モデル	0.615	0.304	0.511

表 3.3: 提案モデルにおける True Positives (TP) と False Negatives (FN)、False Positives (FP) の数。

誤りタイプ	TP	FN	FP
全て	294	263	106
助詞選択	75	60	
混合	16	38	
語彙選択	22	77	
脱落	33	18	
誤形成	53	14	
余剰	40	27	
発音	55	25	
その他	0	4	

Positives (FP) の数を示す。助詞選択、発音、誤形成の誤りは検出率が高く、語彙選択と混合の誤りは検出率が低かった。助詞選択は全ての誤りの大部分を占めるため、FN の数も多くなっている。

表 3.4 に提案モデルにおける TP と FN、FP の例を示す。提案手法はニューラルネットワークを用いているため、表 3.4 の“TP”の例のような長距離を考慮した誤りを検出することができる。この例は文頭に「お願い」とあるので、未来形ではなく願望を表す文であり、「なる」ではなく「なりたい」が適切な表現となる。一方で、表 3.4 の“FN”の例のような誤りはデータ不足のため、検出することができない。また表 3.4 の“FP”のような例もデータ不足により誤って検出してしまふ。「名刺」

表 3.4: TP と FN、FP の例。

TP	誤用	おねがい、しあわせになる！
	正用	おねがい、しあわせになりたい！
FN	誤用	定刻になると、徳川家安が出てきます。
	正用	定刻になると、徳川家康が出てきます。
FP	正用	これ、私の名刺でございます。

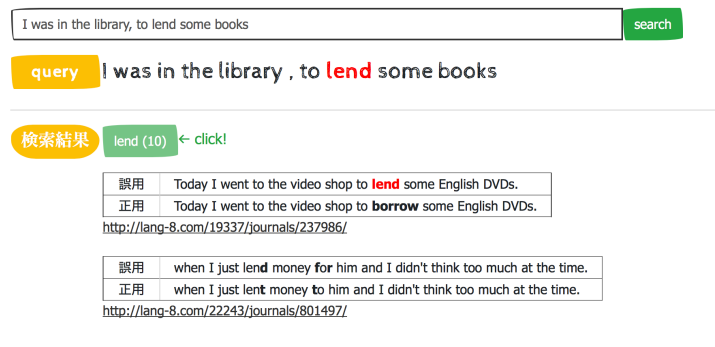


図 3.3: 英語学習者向けシステムのユーザインタフェース。

を使った用例はコーパス中に 2 件しかなく、どちらも「で」ではなく「に」を使ったものであった。そのためシステムは誤って検出したものと考えられる。

3.2 英語学習者向けのシステム

提案システムに使用した Lang-8 Learner Corpora には日本語以外の言語も含まれており、同じ形式で別の言語のシステムも構築できるのではないかと考え、英語学習者のデータを用いて、英語学習者用のシステムを作成した。

■データ 3.1.1 項で説明した Lang-8 Learner Corpora の学習者タグが英語の文対を抽出し、誤用の長さ 100 単語以下、編集距離 7 以下の約 39 万文対を使用した。

■システム 図 3.3 は英語学習者向けシステムのユーザインタフェースである。検索のアルゴリズムは日本語のものと同じである。

第 4 章 日本語検索システムの評価実験

誤用例表示と誤り検出の有用性を確かめるため、比較システムを用いて、内的評価と外的評価の 2 つの方法で評価実験を行った。

4.1 内的評価

検索システムのコーパスに使用していない Lang-8 の学習者作文データから、フレーズをランダムに抽出し、比較システムと提案システムで検索した。実際にユーザが使用する時は、誤りと正用どちらも検索するため、誤りフレーズを 55 フレーズ、正しいフレーズを 55 フレーズ、合計 110 フレーズ抽出し、使用した。誤りフレーズの場合は、検索結果の上位 10 件以内に正解のフレーズを認識することができた数を、正しいフレーズの場合はそのフレーズを含む用例が上位 10 件以内に出てきた数を数えその $\text{precision}@10^*$ を測った。誤りフレーズの誤りタイプを混同 (A)、語彙選択 (L)、脱落 (O)、誤形成 (M)、余剰 (R)、発音 (P)、その他 (X) の 7 種類に分類した。表 4.1 は誤りタイプ別のフレーズ例である。

各システムの誤りタイプ別の頻度と精度を表 4.2 に示す。検索結果は、自動でタグ付けされたデータの正解と一致する場合正解であり、そうでない場合不正解と判断した。先行研究の誤用例文検索システムは入力したクエリを正用例で検索するか誤用例で検索するかをユーザが選択するものであるため、比較システムは正用検索を選択した場合と誤用検索を選択した場合の両方を比較実験に使用した。提案システムはそのフレーズを含む文単位で、比較システムはフレーズ単位で検索した。クエリをそれぞれのシステム（比較システム（正用）、比較システム（誤用）、提案システム）で検索し、その正解のフレーズに完全に一致する用例が得られたフレーズの数数を数え、 $\text{precision}@10$ を測った。

比較システム（正用）は比較システムの正用検索バージョンである。正しいフレーズでは最も高い precision となったが、誤りフレーズの検索クエリは誤り文字列であるため、正用例で検索をかけても一致するものがなく、適切な回答を得られ

* ランキングの上位 10 件における適合率。

表 4.1: 誤りタイプ別テストフレーズ例。

誤り文字列	正解文字列	比較システム (正用)	比較システム (誤用)	提案システム	誤りタイプ
おねさん	おね <u>え</u> さん	×	×	✓	O
ニュージ ランド	ニュー <u>ジ</u> ランド	×	✓	✓	O
みんなさん	みなさん	×	✓	✓	R
大体 <u>に</u>	大体	×	×	✓	R
疑問 <u>を</u> して	疑問 <u>に</u> 思っ <u>て</u>	×	×	✓	M
驚 <u>い</u>	驚 <u>き</u>	×	✓	×	M
が <u>ら</u> え <u>る</u>	し <u>か</u> ら <u>え</u> な <u>い</u>	×	×	×	A
稼 <u>ぐ</u>	稼 <u>い</u> だ	×	✓	×	A
ち <u>さ</u> い	少 <u>な</u> い	×	✓	×	L
助 <u>け</u> ら <u>れ</u> る	で <u>き</u> る	×	×	✓	L
し <u>ま</u> し <u>だ</u>	い <u>ま</u> し <u>た</u>	×	×	×	P
死 <u>ん</u> ち <u>ゃ</u> う	死 <u>ん</u> じ <u>ゃ</u> う	×	✓	✓	P
ハ <u>ウ</u> ス	家	×	✓	✓	X

ることができなかつたため、precision は 0.00 となった。

比較システム（誤用）は比較システムの誤用検索バージョンである。誤用例で検索をかけるため、正しいフレーズで検索した場合には用例はほとんど出てこなかった。誤りフレーズ検索では高い precision を得られた。検索クエリに一致した誤り文字列を含む誤用例と、対になる正用例を検索できる。

一番右のシステムが提案システムである。誤り全体の precision は比較システム（誤用）よりも少し低いが、比較システムはユーザが手動で誤用検索と正用検索を切り替えなければならない。しかし、どのクエリが誤りでどのクエリが正しいかわからないため、その負担が大きい。提案システムはユーザがその切り替えをする必要なく、誤り検出機能により入力したクエリが正しいか誤りであるかを判断して、自動で適切な方を検索するため、誤りフレーズ検索と正しいフレーズ検索の両方を含めた全体の平均 precision は提案手法であるこのシステムが最も高かった。

表 4.2: 内的評価実験結果。「比較・正」は比較システムの正用検索、「比較・誤」は比較システムの誤用検索、「提案」は提案システムを表す。

誤りタイプ	頻度	precision@10		
		比較・正	比較・誤	提案
誤り全て	55	0.00	0.45	0.44
混同	19	0.00	0.37	0.32
語彙選択	16	0.00	0.38	0.19
脱落	8	0.00	0.75	0.75
誤形成	6	0.00	0.40	0.67
余剰	3	0.00	0.67	1.00
発音	2	0.00	0.50	0.50
その他	1	0.00	1.00	1.00
正例	55	0.90	0.15	0.85
全ての平均	110	0.45	0.30	0.65

比較システムと比べて、提案システムは誤形成の precision が高かった。このタイプは誤り箇所が明示的であるため、誤り検出機能の精度が良く、さらに類似度でソートしているため、入力クエリに似た文が上位に来るため、それを正しく直された文が表示できるからだと考える。また、脱落や余剰などの分かりやすい誤りは、問題なく誤り検出が機能し、高い precision で同定することができた。

一方で、語彙選択が検索しづらかった。提案システムは検索結果を類似度でソートしているため、入力のクエリと似たような語を使った文が上位にくる。そのため入力のクエリと全く異なる語彙が正解だった場合、その類似度は低くなり、提案システムは適切な答えを出すことができない。また、誤り検出の精度が十分でないため、誤っている入力クエリが正しいと判断されることも多く、その場合自動的に正用検索になるため、正しく直された文が表示されないこともある。今後誤り検出の精度が上がることで用例検索システムの精度も上がるのが期待される。

表 4.3: 外的評価のための日本語作文の質問。

No.	質問
1	あなたの住んでいる町を紹介。
2	夏と冬とどちらが好きですか。理由も。
3	日本語の難しいところ。
4	テレビニュースと新聞は何が違うのか。
5	海外で経験してみたいこと。
6	自由な時間があったら、何をしてみたいか。
7	自分の国の魅力を紹介。
8	嘘をつくのはいいことか。
9	あなたが健康のためにしていること。
10	大学生活で一番楽しかったこと。

4.2 外的評価

被験者実験では、提案システムを以下の3つの設定で使用する。システム3が提案手法である。例文のリランキングに word2vec を使用したシステムと BERT を使用したシステムの両方で内的評価を行った結果、全く同じ結果となったため、検索スピードの速い word2vec を使用したシステムを外的評価で使用する。全てのシステムで 3.1.1 項で構築したデータを用い、比較箇所以外は全て提案システムの設定を適用する。

- システム 1: 誤り検出なし、正用例のみ表示
- システム 2: 誤り検出なし、正用例と誤用例を表示
- システム 3: 誤り検出あり、正用例と誤用例を表示

まず、システム1を使った作文とシステム2を使った作文の点数を比較して、誤用例と正用例の文対表示の有用性を確認する。次に、システム2を使った作文とその作文をシステム3を使って直した後の点数を比較して、誤り検出機能の実用性を確かめる。

日本語を母語としない日本語学習者6人に日本語作文の問題を1人につき10問解答してもらった。全員、情報工学を専攻する大学院生で、母語は中国語である。

表 4.4: 外的評価結果。

被験者	1	2	3
A	14	20	21
B	26	27	29
C	15	16	16
D	28	25	26
E	22	25	25
F	20	23	28
平均	20.8	22.7	24.2

日本語のレベルは、5人が日本語能力試験 N1、1人が N2 である。A テーマと B テーマに 5 問ずつわけ、A テーマにはシステム 1 を、B テーマにはシステム 2 を使って書いてもらった。その後、システム 2 で書いた作文をシステム 3 を通し校正してもらった。評価基準を統一するため、1 問につき 3 文の制限を付けた。表 4.3 は実験に使用した日本語作文の質問である。被験者に書いてもらった文章に減点法で点数をつけた。持ち点 1 人 30 点で、1 つの文法誤りにつき 1 点引いていき、各システムごとに合計し、最終的な点数を計算した。その結果を表 4.4 に示す。点数の平均はシステム 1 が 20.8、システム 2 が 22.7、システム 3 が **24.2** と提案手法が最も高く、システム 2 と同じ点数も含め、6 人中 5 人が提案手法で最も高い点数を出した。

表 4.5 はシステム 2 を使って書いた作文をシステム 3 を用いて校正した時に誤りを訂正できた割合を表す[†]。内的評価と同じように、誤形成が最も高い割合で訂正できていた。内的評価とは異なって語彙選択もよく訂正できていたが、内訳を見ると内容語が 0.17、機能語が 0.38 と、機能語が高い割合で訂正できていることが分かる。内的評価では助詞などの機能語は評価の対象ではないので分からなかったが助詞などの機能語の選択誤りには対応できることが分かる。

[†]この時、誤り数と誤りの訂正数は人手で数えた。

表 4.5: 外的評価での誤り種類別の頻度と正解率。

誤りタイプ	頻度	#	正解率
誤り全て	44	9	0.20
混同	6	1	0.17
語彙選択	14	4	0.29
脱落	7	1	0.14
誤形成	6	2	0.33
余剰	9	1	0.11
発音	2	0	0.00
その他	0	0	0.00

第 5 章 英語検索システムの評価実験

5.1 実験設定

英語検索システムの有用性を確かめるため、日本語学習者向けの検索システムで行った実験と同じように、検索システムに使用していない Lang-8 のデータから、誤用と正用をそれぞれ 50 フレーズずつランダムに抽出し、提案システムと比較システムで検索し、正しい用例を提示できた数を比較した。英語学習者向け検索システムの評価実験では、日本語母語話者の大学生が書いた英文エッセイを集めた KJ データからも誤用と正用をそれぞれ 25 フレーズずつ抽出し各システムで検索し、比較した。

比較システムには、英単語を打ち込むとその単語を使用した表現パターンを表示する WriteAhead [7]*を使用した。誤用フレーズの場合は、検索結果の上位 10 位以内に正解のフレーズを認識することができた数を、正しいフレーズの場合はそのフレーズを含む用例が上位 10 位以内に出てきた数を数えその正解率を測った。

5.2 結果

結果を表 5.1 に示す。ほとんどの結果で提案手法が比較システムを上回ることができた。Lang-8 の正用フレーズは頻繁に使われやすいパターンが比較的多く、そのような形はクエリが正しければ WriteAhead でも検索しやすいため、提案システムと同等の正解率となった。

表 5.2 及び表 5.3 は誤りタイプ別の各システムの正解率である。Lang-8 データセットを使用した実験では、提案システムは、タイプミスにおいて非常に高い結果を出している。この誤りタイプは表層的な誤りなので、誤り検出の精度が良く、更に誤りが分かりやすいので用例の検索もしやすかった。大文字と小文字の区別が高いのも同じ理由である。単数複数や時制は、比較システムよりも低い正解率となった。これらの誤りタイプの使用例を見てみると決まったパターンで使われるものが多く、比較システムは、その単語に用いられやすいパターンを提示するため、適切

*<http://writeahead.nlpweb.org/>

表 5.1: 英語の提案システムと比較システム (WriteAhead) の評価実験結果。

システム	正解率			
	Lang-8 (誤)	Lang-8 (正)	KJ (誤)	KJ (正)
WriteAhead	0.16	0.42	0.08	0.44
提案システム	0.28	0.42	0.36	0.52

表 5.2: Lang-8 コーパスでの評価詳細。

誤りタイプ	頻度	正解率	
		提案システム	WriteAhead
誤り全て	50	0.28	0.16
語彙選択	13	0.08	0.08
冠詞	12	0.33	0.25
前置詞	6	0.33	0.17
タイプミス	6	0.83	0.00
単数複数	5	0.00	0.40
大文字小文字	3	0.67	0.00
時制	2	0.00	0.50
その他	3	0.00	0.00

な用例を提示できたと考えられる。語彙選択はどちらのシステムでも正解の語彙を提示することが難しかった。

KJ コーパスを使用した実験においても、提案システムはタイプミスに強いことが分かる。KJ コーパスの英文は Lang-8 コーパスの英文よりも比較的簡単で分かりやすい英語が使われているため、提案システムは Lang-8 コーパスよりも KJ コーパスの方が高い精度で用例を提示することができた。

表 5.3: KJ コーパスでの評価詳細。

誤りタイプ	頻度	正解率	
		提案システム	WriteAhead
誤り全て	25	0.36	0.08
語彙選択	6	0.00	0.00
冠詞	5	0.40	0.00
時制	5	0.40	0.00
前置詞	4	0.25	0.25
タイプミス	3	1.00	0.00
単数複数	2	0.50	0.50

第 6 章 おわりに

本研究では、誤用例文検索の有用性に着目し、正用例付き誤用例文の表示を行った。さらに、本システムでは誤り検出を自動で行うことによって言語学習者が気づいていない誤りを認識し、誤用例を提示する。この機能により、学習者は誤りを把握して解釈し、訂正するか否かを選択できる。

日本語学習者向けの文法誤り検出機能付き誤用例文検索システム実験により、内的評価と外的評価で共に提案手法が最も優れていることを示した。また、英語学習者向けの誤用例文検索システムも作成し、用例検索システムに文法誤り検出を入れると言語に依存せずに誤用検索の正解率が上がることが確認できた。

発表リスト

国際会議

1. Mio Arai, Tomonori Kodaira and Mamoru Komachi. **Sakura: Large-scale Incorrect Example Retrieval System for Learners of Japanese as a Second Language**. In The 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019): System Demonstrations. July, 2019.
2. Mio Arai, Masahiro Kaneko and Mamoru Komachi. **Grammatical-Error-Aware Incorrect Example Retrieval System for Learners of Japanese as a Second Language**. In 14th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 14). August 2, 2019.

国内会議

1. 新井美桜, 小平知範, 小町守. **Lang-8 を用いた日本語学習者向けの誤用検索システムの構築**. 言語処理学会第 24 回年次大会. pp.204-207. March 13, 2018.
2. 新井美桜, 金子正広, 小町守. **日本語学習者向けの文法誤り検出機能付き誤用例文検索システム**. 言語処理学会第 25 回年次大会. pp.1097-1100. March 15, 2019.
3. 平尾礼央, 新井美桜, 嶋中宏希, 勝又智, 小町守. **ニューラルネットワークを利用した日本語学習者の複数項目作文能力推定**. NLP 若手の会第 14 回シンポジウム (YANS 2019). August 27, 2019.

4. 平尾礼央, 新井美桜, 嶋中宏希, 勝又智, 小町守. 複数項目の採点を行う日本語学習者の作文自動評価システム 言語処理学会第 26 回年次大会発表予定. March 16-19, 2020.
5. 小山碧海, 喜友名朝視顕, 小林賢治, 新井美桜, 小町守. 日本語学習者の文法誤り訂正のための評価コーパス構築 言語処理学会第 26 回年次大会発表予定. March 16-19, 2020.

謝辞

本論文の執筆に際して、研究室に配属されてから 3 年間、ご指導ご鞭撻を頂きました小町守准教授に深く感謝いたします。文系出身で外部からの進学である私を快く受け入れてくださりありがとうございました。また、日々の研究生活の中で、研究室の先輩、同期、後輩には多くのアドバイスやご協力を頂きました。分からないことばかりの私に優しく懇切丁寧に教えてくださり本当にありがとうございました。また、副査を引き受けてくださった山口教授、高間教授に心より感謝いたします。加えて、Lang-8 のデータ使用に際して、快諾してくださった株式会社 Lang-8 社長喜洋洋氏に感謝申し上げます。皆様へ心から感謝の気持ちと御礼を申し上げます。謝辞と代えさせていただきます。

参考文献

- [1] 李 在鎬, 淺尾仁彦, 濱野寛子, 佐野香織, 井佐原均, “タグ付き日本語学習者コーパスの開発,” 言語処理学会第 14 回年次大会, pp.658–661, 2008.
- [2] K. Nishina, B. Hodošček, Y. Yagi, and T. Abekawa, “Construction of a learner corpus for Japanese language learners: Natane and Nutmeg,” *Acta Linguistica Asiatica*, vol.4, no.2, pp.37–51, 2014.
- [3] H. Suzuki and K. Toutanova, “Learning to predict case markers in Japanese,” *Proceedings of ACL*, pp.1049–1056, 2006.
- [4] K. Imamura, K. Saito, K. Sadamitsu, and H. Nishikawa, “Grammar error correction using pseudo-error sentences and domain adaptation,” *Proceedings of ACL*, pp.388–392, 2012.
- [5] 水本智也, 小町 守, 永田昌明, 松本裕治, “日本語学習者の作文自動誤り訂正のための語学学習 SNS の添削ログからの知識獲得,” *人工知能学会論文誌*, vol.28, no.5, pp.420–432, 2013.
- [6] M.-H. Chen, S.-T. Huang, H.-T. Hsieh, T.-H. Kao, and J.S. Chang, “FLOW: A first-language-oriented writing assistant system,” *Proceedings of the ACL 2012 System Demonstrations*, pp.157–162, 2012.
- [7] T.-H. Yen, J.-C. Wu, J. Chang, J. Boisson, and J. Chang, “WriteAhead: Mining grammar patterns in corpora for assisted writing,” *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, pp.139–144, 2015.
- [8] A. Kilgariff, P. Rychly, P. Smrž, and D. Tugwell, “The sketch engine,” *Proceedings of EURALEX*, pp.105–116, 2004.
- [9] M. Schuster and K.K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Transactions on Signal Processing*, vol.45, pp.2673–2681, 1997.
- [10] M. Rei, G.K. Crichton, and S. Pyysalo, “Attending to characters in neural sequence labeling models,” *Proceedings of COLING*, pp.309–318, 2016.
- [11] M. Rei and H. Yannakoudakis, “Compositional sequence labeling models for error detection in learner writing,” *Proceedings of ACL*, pp.1181–1191, 2016.
- [12] M. Kaneko, Y. Sakaizawa, and M. Komachi, “Grammatical error detection using error- and grammaticality-specific word embeddings,” *Proceedings of IJCNLP*, pp.40–48, 2017.
- [13] S. Kasewa, P. Stenetorp, and S. Riedel, “Wronging a right: Generating better errors to improve grammatical error detection,” *Proceedings of EMNLP*, pp.4977–4983, 2018.
- [14] M. Junczys-Dowmunt and R. Grundkiewicz, “Phrase-based machine translation is state-of-the-art for automatic grammatical error correction,” *EMNLP*, pp.1546–

- 1556, 2016.
- [15] S. Chollampatt and H.T. Ng, “A multilayer convolutional encoder-decoder neural network for grammatical error correction,” *Proceedings of AAAI*, pp.5755–5762, 2018.
 - [16] 今枝恒治, 河合敦夫, 石川裕司, 永田亮, 榊井文人, “日本語学習者の作文における 格助詞の誤り検出と訂正,” *情報処理学会研究報告コンピュータと教育*, pp.39–46, 2003.
 - [17] H. Oyama, M. Komachi, and Y. Mastumoto, “Towards automatic error type classification of Japanese language learners’ writing,” *Proceedings of PACLIC*, pp.163–172, 2013.
 - [18] M. Junczys-Dowmunt, R. Grundkiewicz, S. Guha, and K. Heafield, “Approaching neural grammatical error correction as a low-resource machine translation task,” *Proceedings of ACL*, pp.595–606, 2018.
 - [19] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” *Proceedings of ICLR*, 2013.
 - [20] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *Proceedings of NAACL*, pp.4171–4186, 2019.
 - [21] 柴田知秀, 河原大輔, 黒橋禎夫, “BERT による日本語構文解析の精度向上,” *言語処理学会 第 25 回年次大会*, pp.205–208, 2019.
 - [22] H. Morita, D. Kawahara, and S. Kurohashi, “Morphological analysis for unsegmented languages using recurrent neural network language model,” *Proceedings of EMNLP*, pp.2292–2297, 2015.
 - [23] R. Sennrich, B. Haddow, and A. Birch, “Neural machine translation of rare words with subword units,” *Proceedings of ACL*, pp.1715–1725, 2016.
 - [24] R. Nagata and K. Nakatani, “Evaluating performance of grammatical error detection to maximize learning effect,” *Proceedings of COLING*, pp.894–900, 2010.
 - [25] F.J. Och, “Minimum error rate training in statistical machine translation,” *Proceedings of ACL*, pp.160–167, 2003.