

## 日本語学習者の作文におけるエラータイプの自動分類へ向けて

大山 浩美 (奈良先端科学技術大学院大学 情報科学研究科)

小町 守 (奈良先端科学技術大学院大学 情報科学研究科)

藤野 拓也 (奈良先端科学技術大学院大学 情報科学研究科)

松本 裕治 (奈良先端科学技術大学院大学 情報科学研究科)

## Towards Automatic Error Type Classification on Japanese Language Learners' Writings

Hiromi Oyama, Mamoru Komachi, Takuya Fujino, Yuji Matsumoto

(Graduate School of Information Science, Nara Institute of Science and Technology)

### 1 はじめに

現在、様々な種類の学習者コーパスが収集され、学生へのフィードバック、教材研究、教授法の見直しへ役立てるための言語教育の調査研究に利用されている。学習者コーパスは、大学等の日本語教育機関などで収集されるのみならず、SNSを利用したウェブ上での言語学習者のための添削サービス (Lang-8<sup>1</sup>) もあり、膨大な数の学習者の作文を集めることが可能である。しかし、学習者が生み出した文は新聞や書籍などのコーパスと異なり正用だけでなく誤用も含まれており、そのまま分析をするのは難しく、調査分析のための前処理が必要となってくる。その処理には、半自動の誤り検出・訂正とエラータイプの分類、実際の人手によるエラータグ付与作業などが含まれる。

言語学習者にとってどの部分が誤用かを知ることが重要であるが、なぜ誤用なのか、どんな誤用なのか、という理由を知ることがさらに重要である。そのためエラータイプを分類するというタスクは学習者コーパス整備のための一つの重要なタスクであると考えられる。

現在のコーパスは、それぞれの機関でそれぞれの研究目的のために収集されているため、エラータグの設計方法は様々である。また、人手で付与されるため、エラータグ付きのコーパスは少なく、タグ判定も不安定になりがちになる。エラータイプ自動分類は、人手でしか行えなかったエラー判定を一部自動化し、人手の負担、コストを軽減しようというものである。しかし、そのみならず、現在の様々なエラータグの再現性、妥当性を検証するのにも役立てる。そういった点が検証できれば、これからのエラータグ設計に役立てることもできる。

上記のような理由から、本研究は、大規模なエラータグつけコーパスがないことやエラータグの付け方がコーパスごとに異なるなどのエラータグ付与作業の困難さを一部支援することを目的し、誤用コーパスから機械的にエラータグ付与のための自動的な処理を行う手法を検討する。

具体的にはアノテータが修正した箇所のエラータイプを当てるタスクに取り組み、エラータグが付与された NAIST 誤用コーパス [17] のタグ付与基準に基づき、機械学習を用いた多クラス分類によるエラータイプ分類実験を行った。

---

<sup>†</sup>{hiromi-o,komachi,takuya-fu,matsu}@is.naist.jp

<sup>1</sup><http://www.lang-8.com>

実験は、2種類に分けられる。1つ目は、NAIST 誤用コーパス内でトレーニングとテストを行い、どれくらいエラータイプを当てられるかというタスク（トレーニングデータ：NAIST 誤用コーパス、テストデータ：NAIST 誤用コーパス）と、もう1つは、まったく別のエラータグの付与されていない日本語学習者コーパス（Lang-8）（トレーニングデータ：NAIST 誤用コーパス、テストデータ：Lang-8）でどれだけ当てられるかというタスクである。

NAIST 誤用コーパス内で行った実験では、正解率は77.1%であった。Lang-8をテストデータとして行った実験では83.3%であった。「動詞」や「表記」の誤りに比べて「形容詞」や「名詞」のエラータイプの自動分類が難しいことが分かり、エラータイプが正しく判定されなかった理由は、実験素性の選択の問題、形態素解析による問題、タグ範囲の設定による問題と分析した。

## 2 関連研究

現在、存在している日本語学習者コーパスは、大阪大学の寺村コーパス [18]（データ総数4,601文中3,131文がエラータグ付け済み）、名古屋大学の学習者コーパス [21]（756ファイル）、東京外国語大学の「オンライン日本語誤用コーパス辞典」<sup>2</sup>（エラータグつき作文40ファイル）、筑波大学の「日本語学習者作文コーパス」 [22]<sup>3</sup>（540ファイル）などがある。本研究もこれらの研究と同じく学習者コーパスにエラータイプを付与することを目的としているが、これらの研究がすべて人手でエラータイプを付与しているのに対し、我々は半自動化で付与することを目的としている点が異なる。

学習者コーパスにおける整備作業には、データ収集、タグ設計、誤り検出・訂正、エラータイプ分類が主に考えられ、それらに関する研究が進められている。

その中で誤り判定においては、英語のスペルミスの誤り訂正 ([13])、英語の名詞の可算性（数えられる名詞）、不可算性（数えられない名詞）の誤り検出・訂正研究 ([2], [23])、前置詞の誤り検出・訂正に関する研究 ([3], [4], [5], [12], [6])、冠詞誤り検出・訂正に関する研究 ([7], [5], [6], [14]) などがある。日本語を対象とする研究では、格助詞を対象とした研究が多い ([15], [9], [16], [24], [11])。さらに、エラータイプに特に着目せずに文を誤用文と正用文とに分類する研究もある ([10], [19])。

エラータイプ分類においては、[1]が英語学習者の作文において実験を行っている。データは Cambridge Learner Corpus (CLC)<sup>4</sup>で、75の母語数、2千万語から成る。そのうち、5百万語分にエラータグが付与されている [8]。多クラス分類モデルで、最大エントロピーを用い、15クラスのエラータイプ（余剰、不足、変更、スペルミス、動詞の時制など）に分ける実験を行っている。本稿とは、学習者の書いた誤用文と正用文を多クラス分類モデルで分類するという点が共通している。しかし、日本語学習者の作文におけるエラータイプ分類実験はまだ見られない。

## 3 NAIST 誤用コーパスにおけるエラータグのアノテーション

本稿におけるエラータイプ分類実験は、我々が作成した NAIST 誤用コーパス [17] におけるエラータイプの自動分類に取り組んだので、まず NAIST 誤用コーパスの概要について述

<sup>2</sup>[http://cblle.tufts.ac.jp/llc/ja\\_wrong/index.php?m=default](http://cblle.tufts.ac.jp/llc/ja_wrong/index.php?m=default)

<sup>3</sup><http://www34.atwiki.jp/jccorpus/>

<sup>4</sup><http://www.cambridge.org/elt/corpus/clc.htm>

表 1: NAIST 誤用コーパスのエラータグ集計結果（上位 10 位まで）

VN=ベトナム/TH=タイ/CN=中国/ML=マレーシア/MN=モンゴル/KH=カンボジア/KR=韓国/SG=シンガポール

	VN	TH	CN	ML	MN	KH	KR	SG	全体的な割合
語彙選択	35.0	27.0	17.2	22.8	29.2	12.8	25.2	23.8	24.1
助詞	21.8	23.1	20.6	24.2	22.1	17.4	17.3	30.6	22.1
表記	9.8	10.1	19.8	16.9	12.7	33.6	15.5	6.8	15.7
動詞	13.8	15.3	16.8	12.1	14.2	15.9	14.6	10.2	14.1
成句	6.2	7.0	2.6	7.3	5.2	1.7	3.4	4.9	4.8
文体	1.7	1.2	2.3	6.0	4.1	6.1	3.1	6.3	3.9
名詞	2.5	2.6	3.5	1.4	3.4	2.0	4.4	2.9	2.8
文全体	2.0	2.6	1.2	3.4	0.7	1.4	2.4	2.4	2.0
形容詞	2.0	0.9	2.6	1.5	1.9	1.7	1.5	1.5	1.7
語順	1.0	1.3	1.2	0.3	0.4	1.2	0.6	0.0	0.8

べる<sup>5</sup>。「作文対訳 DB」の中で添削が施してある 313 名の作文中の誤用部分にタグを付与し、様々な情報を補完した [17]。ファイル数は 313、総文字数は 191,994 字となっている。本稿は「作文対訳 DB」に対しアノテーションを行なっているため、データは共通しているが、「作文対訳 DB」には誤用の種類がアノテーションされていない。

NAIST 誤用コーパスのエラータグの上位 10 位までを表 1 に示す<sup>6</sup>。誤用で多いのは、「語彙選択」、「助詞」、「動詞」、「表記」、「成句」、「名詞」、「形容詞」に関する誤用である。「語彙選択」は、ふさわしい語彙を選べなかった（「国民」を「人民」）誤りである。「助詞」は、助詞が抜けている（不足）、不必要な助詞がつけられている（余剰）、誤った助詞が入れられている（変更）などがサブカテゴリーに含まれる。「動詞」カテゴリーの誤用には、サブカテゴリーとして動詞の活用、自動詞か他動詞か、受け身の誤り、テンスアスペクトの誤りなどが含まれる。「表記」には、ひらがな、かたかな、漢字に関する誤りなどが入る。「成句」は、きまったフレーズ（「～たり～たり」など）がうまく使えなかった誤りを含む。「名詞」には文を名詞化するときの「の」と「こと」を使い誤ったものや品詞選択誤りなど、「形容詞」には活用など、詳細に分類されている。「文全体」は、文がすべて書き換えられている誤り事例である。「文体」は、「です・ます体」に関する誤りで、「語順」は、語彙の出現する位置に関する誤りである。

#### 4 多クラス分類によるエラータイプ分類

実験の概要を図 1 に示す。まず、正用文と誤用文が対になっており、エラータイプに分類されているコーパスを利用し、正用文と誤用文とを取り出す。上記の 2 種類の文から、対応する誤用箇所 ( $x$ ) と、正用箇所 ( $y$ )、エラータイプ ( $t$ ) の 3 つ組からなる事例 ( $x, y, t$ ) を取り出す。

それらを取る際に、正用文と誤用文において動的計画法によるマッチングを用いて置換対

<sup>5</sup>アノテーションについて詳しくは [17] を参照されたい

<sup>6</sup>その他の誤用には、「接続」「モダリティ」「コロケーション」、「助動詞」、「指示詞」、「体言修飾」、「否定」、「副詞」、「仮定、条件」、「代名詞」、「その他 文法事項に関する誤り」などがある。詳しくは [17] を参照されたい

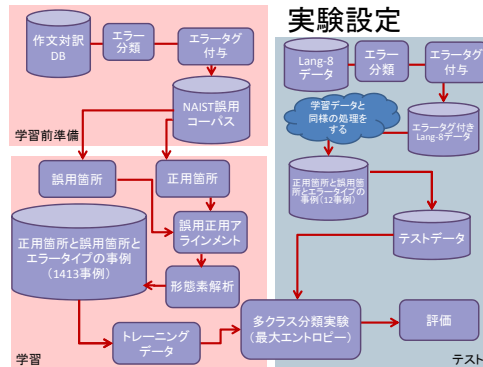


図 1: 実験概要

の抽出を行った。誤用文と正用文をそれぞれ文頭から 1 文字ずつ読み進めていき、誤用文には存在するが正用文には存在しない文字列を挿入箇所としてみなした。さらに、正用文には存在するが誤用文には存在しない文字列を削除箇所としてみなした。今回は挿入箇所と削除箇所が連続した部分を置換のペアとして抽出した。

例えば、「誰に (誤) → で (正) もたばこを吸う権利がある」という文から事例をとりだす。ラベルとして「助詞」の誤用 (t) と、さらに、誤用箇所「に」 (x) と、正用箇所「で」 (y) という事例が 1,413 個取り出される。素性には、誤用箇所そのものに加えて誤用箇所周辺の文脈、正用箇所そのものに加えて正用箇所周辺の文脈、およびそれらの組み合わせなどを用いることができる。1つの文に別々の誤用が 2 つ以上ある場合は、1 誤用につき 1 事例として取り出した。それらの事例をトレーニングし、多クラス分類器として機械学習の「最大エントロピー法<sup>7</sup>」を利用し、5 つの多クラス分類を試みた。

#### 4.1 データ

NAIST 誤用コーパスから正用文と誤用文のペアを任意に取り出し、トレーニングデータとして使用した。テストデータは、NAIST 誤用コーパスから任意に取り出した文 (実験 1) と Lang-8 から 12 文を取り出し、エラータイプに分類したもの (実験 2) である。Lang-8 はもともとエラータグ付与されていないため、任意に 12 文を取り出し、エラータイプを付与した。

実験に使うエラータイプは、表 1 の「形容詞」、「動詞」、「助詞」、「名詞」、「表記」のサブカテゴリーを含んだ大カテゴリー 5 つに限定した。それぞれの例を表 2 に示す。

#### 4.2 素性

素性は、誤用箇所の表層、正用箇所の表層、正用箇所の形態素解析結果を使用した。さらに、unicd 辞書を使った MeCab 形態素解析器<sup>8</sup> を利用し、形態素情報を素性として取り入れた [20]。上記の例文から、誤用箇所の「に」、それが対応する箇所の「で」、その形態素解析

<sup>7</sup>[http://homepages.inf.ed.ac.uk/lzhang10/maxent\\_toolkit.html](http://homepages.inf.ed.ac.uk/lzhang10/maxent_toolkit.html)

<sup>8</sup><http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

表 2: 誤用文例

エラータイプ	誤用文例
形容詞	発音が大変 <u>  </u> だけではない (誤) →大変 <u>な</u> だけではない (正)
名詞	趣味は映画を見る <u>の</u> です (誤) →見る <u>こと</u> です (正)
表記	<u>年ば</u> の人 (誤) → <u>年配</u> の人 (正)
助詞	英語を <u>を</u> わかる (誤) →英語 <u>が</u> わかる (正)
動詞	手紙を書 <u>き</u> ない (誤) →書 <u>か</u> ない (正)

結果である「助詞」が素性となる。

### 4.3 評価尺度

評価尺度として再現率、適合率、F 値、正解率を利用した。再現率は、対象とする各エラータイプの中で、正しく分類されたエラータイプを指し、適合率は、システムがあるエラータイプだと分類したもののうち正解を当てた率である。F 値はそれらの調和平均を表している。正解率とは、すべての事例の中で、正しいエラータイプに属すると判定されたもの (True Positives) とこの誤用文はエラータイプに属しないと正しく判定されたもの (True Negatives) との割合である。

$$\text{再現率} = \frac{\text{正しく分類された事例数}}{\text{各エラータイプの全事例数}} \times 100 \quad (1)$$

$$\text{適合率} = \frac{\text{正しく分類された事例数}}{\text{システムがあるエラータイプだと分類した事例数}} \times 100 \quad (2)$$

$$F \text{ 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}} \quad (3)$$

$$\text{正解率} = \frac{\text{正しく分類された事例数}}{\text{すべての事例}} \times 100 \quad (4)$$

## 5 NAIST 誤用コーパスに対するエラータイプ分類実験結果

まず、一つ目の実験、NAIST 誤用コーパス内でエラータイプ分類はどのくらいできるのかを検証するために分割交差検定 (10 分割) を行った。正解率は、全体で 77.14%であった。各エラータイプにおける適合率、再現率、F 値を表 3 に示す。これを見ると、「表記の誤用」また、「動詞の誤用」が安定して分類されていることがわかる。最も分類されにくいのが「形容詞の誤用」である。

表 4 には、ある分割内での誤り事例を取り出し、エラー分析を行った結果を示している。システムが誤って分類した理由は、「素性の問題」、「プログラムミス」、「形態素解析による問題」、「タグつけ時のミス」、「タグ範囲における問題」などに分けられた。以下に数の多い 3 つの問題について詳しく述べる。

### 5.1 素性の問題

表記の間違いに多く見られたのは、現在の素性で分類するのが難しい問題であった。表記の間違いは一般的に書き間違い (spelling errors) を意味している。しかし、例文「結婚式は、

表 3: NAIST 誤用コーパスの 10 分割交差検定結果

エラータイプ	適合率	再現率	F 値
助詞	65.7%	64.0%	62.1%
動詞	76.0%	85.2%	80.3%
表記	84.3%	82.8%	83.5%
名詞	53.6%	36.2%	42.2%
形容詞	31.7%	24.1%	26.5%

主に結婚式場やホテルやおおきいビルがあります」で見られるように、漢字で書くべきところを書いていない誤りも含んでいる。さらに、送り仮名(嫌らう(誤)→嫌う(正))や漢字ひらがな混合(かぶル(誤)→かぶる(正))などの事例も「表記」の誤用なのであるが、システムは「表記」だと正しく分類しなかった。今回のエラータイプ分類実験には、サブカテゴリーを含んだ大分類のタグを使用している。分類が大きければ大きいほど、エラータイプ分類の揺れがないと考えたためと、サブカテゴリーまで自動分類するのは難しいと考えたためである。しかし、例えば「表記」のように主に「書き間違い」があるグループに「送り仮名」の問題や「漢字ひらがな混合」の問題など、様々な問題が混在している場合、それぞれの事例数が足りないのが問題であることは十分に考えられる。

## 5.2 形態素解析による問題

今回の実験では、学習者の作文は誤りを含むため形態素解析が難しいと考え、誤用文の品詞情報は利用しなかった。そのため、例えば、「私が入学できた大学は学期の 始めの (誤) → 始まる (正) 前の3日間入学式を行った」という文において、「始まる - 動詞」の結果は利用できるが、「始めの」の形態素解析結果を素性に組み込まなかった。誤りが含まれていたとしても、「始め-名詞」という素性が利用できれば、「名詞」の誤りと分類される可能性はある。誤用文の形態素解析結果を利用すればこのような誤分類は解決できる可能性がある。

## 5.3 タグ範囲における問題

タグ付与作業を行う時、正用文と誤用文で、単語同士が1対1に一致しない場合、どこまでにタグを振るのかということが問題となる。例えば、「おいのりのあとで、家で私の家族はいっしょに朝食を食べるものです(誤)→ます(正)」という文の場合、「食べ」までは正用文でも誤用文でも同じなのでそこにはタグを振らずに、異なる部分だけに振るとする。そうすると、エラータイプを分類するための問題となる箇所は「もの」であるが、その前後をどうするのかという問題がある。現在は、異なる部分をすべてタグで囲んでいる。そうしなければ変更されている箇所がわからない。しかし、そのすべてにタグを振ると誤用の理由となるべき箇所がわからなくなるジレンマがある。学習者は自由に誤用を犯す。教師によるその添削方法も様々である。そのため、エラータグは、単語間の対応になっていない場合が多い。エラータグを設定するときのエラーの範囲の決定の仕方には、議論の余地が十分にある。

表 4: NAIST 誤用コーパスでのエラー分析結果

	システムの 問題	タグ範囲に おける問題	形態素解析 による問題	プログラム ミス	タグつけ時 のミス	エラー全事 例
表記	11	0	1	0	2	14
動詞	2	2	2	4	1	11
名詞	0	7	1	0	0	8
形容詞	0	1	2	1	1	5
助詞	2	0	0	0	0	2
計	15	10	6	5	4	40

## 6 Lang-8 に対するエラータイプ分類実験結果

Lang-8 をテストデータとして使用した場合、システムはテスト文 12 文において 10 文の分類に成功した。

表 5 は、テスト文中のどのような文においてシステムが分類に成功し、どんな文において分類ミスをしているかを表している。傾向は NAIST 誤用コーパスを用いた実験と同様で、「表記」、「助詞」は、すべての文で分類に成功しており、判定しやすい事が分かる。「形容詞」や「名詞」で分類に失敗している。テスト文 1 において、「形容詞」の誤用を「動詞」と判定している。文 1 のように、「形容詞」や「名詞」は、訂正されるときに「小さい—形容詞」「の—助詞」から「少し—副詞」「の—助詞」というように、別の品詞になってしまう。文 6 においても「名詞」を「動詞」と判定している。この際にも「その場合—名詞」から「そう—副詞 なれ—動詞-非自立可能ば—助詞-接続助詞」となり、「表記」のようにほとんどが名詞から名詞への変更となる場合とは異なる。

今回は、素性として単語の表層と形態素解析結果しか使っていないため、形態素を超えた修正がないエラータイプのほうが分類されやすかったのではないかと考えられる。

## 7 おわりに

本稿では、エラータイプの自動分類に向けた実験を試みた。現在、収集されている様々な学習者コーパスのエラータグ付与支援ができるように NAIST 誤用コーパスとは別コーパスである Lang-8 でのテストを行った。結果、NAIST 誤用コーパスにおいても Lang-8 コーパスにおいても 80%前後の正確率を得ることができた。「助詞」、「動詞」、「表記」の誤用などは、事例数が多いことからか、十分な判定結果が観測された。しかし、「名詞」や「形容詞」などのエラータイプは分類するのが難しかった。理由として、タグの範囲が必ずしも単語対単語の 1 対 1 の対応をせず、形態素解析の情報が素性に取り込めない点などが考えられる。これは、今後の素性選択に活かしたい。

また、事例数が少ないことも自動分類を難しくした要因である。次回の実験ではトレーニング、テストともにデータを増やすつもりである。

表 5: Lang-8 のテスト文結果

結果	エラー タイプ	システム 出力結果	誤用文例
×	形容詞	動詞	私たちはうちに小さい時間いました (誤) →少しの時間 (正)
○	表記	表記	こんな宴会・飲み会は日本語で話す <u>たみ</u> にいいチャンスだ (誤) →ため (正)
○	表記	表記	<u>ゆべ 23:00</u> に友達が到着しました (誤) →ゆうべ (正)
○	表記	表記	座の手前にきれいな <u>ちゅーりつぷ</u> を見ました (誤) →チューリップ (正)
○	表記	表記	<u>せのもん</u> はなんですか。 (誤) →せんもん (専門) (正)
×	名詞	動詞	<u>その場合は</u> いいと思う (誤) →そうならば (正)
○	名詞	名詞	私のお気に入りの趣味は読み <u>__</u> です (誤) →読むこと (正)
○	助詞	助詞	今日最後 <u>までに</u> できたらはいいと思います (誤) →最後まで (正)
○	助詞	助詞	もし面接のときこのような質問 <u>を</u> 出て来るかもしれません (誤) → (正) 質問が
○	助詞	助詞	わけ <u>が</u> なく、微笑みながら雪が降るのを眺めていました (誤) →わけもなく (正)
○	助詞	助詞	普通 <u>に</u> 学校はあまり日本語で喋らないです。 (誤) →普通は (正)
○	動詞	動詞	今晚のパーティは本当の宴会と <u>思よ</u> (誤) →思います (正)

参考文献

- [1] Swanson B. and Yamangil E. Correction detection and error type selection as an esl educational aid. In *Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 357–361, 2012.
- [2] C. Brockett, W.B. Dolan, and M. Gamon. Correting ESL Errors Using Phrasal SMT Techniques. *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL*, pp. 249–256, 2006.
- [3] M. Chodorow, J. Tetreault, and N-R. Han. Detection of grammatical errors involving prepositions. *Proceedings of the 4th ACL–SIGSEM Workshop on Prepositions*, pp. 45–50, 2007.
- [4] R. De Felice and S.G. Pulman. Automatically acquiring models of prepositional use. *Proceedings of the 4th ACL–SIGSEM Workshop on Prepositions*, pp. 45–50, 2007.
- [5] R. De Felice and S.G. Pulman. A classifier-based approach to preposition and determiner error correction in L2. *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, pp. 169–176, 2008.



- [6] M. Gamon, J. Gao, C. Brockett, A. Klementiev, W.B. Dolan, D. Belenko, and L. Vanderwende. Using contextual speller techniques and language modelling for ESL error correction. *Proceedings of the 3rd International Joint Conference on Computational Linguistics (IJCNLP 2008)*, 2008.
- [7] N. R. Han, M. Chodorow, and C. Leacock. Detection errors in english article usage by non-native speakers. *Natural Language Engineering*, Vol. 12(2), pp. 115–129, 2006.
- [8] D. Nicholls. The Cambridge Learner Corpus—error coding and analysis for lexicography and ELT. In Archer et al., editor, *Proceedings of the Corpus Linguistics 2003 Conference (CL2003)*, pp. 572–581. 2003.
- [9] H. Oyama, Y. Matsumoto, M. Asahara, and K. Sakata. Construction of an error information tagged corpus of japanese language learners and automatic error detection. *Proceedings of the Computer Assisted Language Instruction Consortium*, 2008.
- [10] G. Sun, X. Liu, G. Cong, M. Zhou, Z. Xiong, J. Lee, and C.Y. Lin. Detecting Erroneous Sentences using Automatically Mined Sequential Patterns. *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 81–88, 2007.
- [11] H. Suzuki and K. Toutanova. Learning to Predict Case Makers in Japanese. *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1049–1056, 2006.
- [12] J. Tetreault and M. Chodorow. The Ups and Downs of Preposition Error Detection in ESL Writing. *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, pp. 865–872, 2008.
- [13] A. Wilcox-O’Hearn, G. Hirst, and A. Budanitsky. Real-word spelling correction with trigrams: A reconsideration of the Mays, Damerau, and Mercer model. *Proceedings of 9th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2008) (Lecture Notes in Computer Science Vol.4919)*, pp. 605–616, 2008.
- [14] X. Yi, J. Gao, and W.B. Dolan. A web based English proofing system for ESL users. *Proceedings of the 3rd International Joint Conference on Computational Linguistics (IJCNLP 2008)*, pp. 619–624, 2008.
- [15] 大木環美, 大山浩美, 北内啓, 末永高志, 松本裕治. 非日本語母国話者の作成するシステム開発文書を対象とした助詞の誤用判定. 言語処理学会第17回年次大会, pp. 1047–1050, 2011.
- [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研究会報告, pp. 39–46, 2003.
- [17] 大山浩美, 小町守, 松本裕治. 日本語学習者の作文における誤用タグつきコーパスの構築について—NAIST 誤用コーパスの開発—. 第一回テキストアノテーションワークショップ, 2012.

- [18] 寺村秀夫. 外国人学習者の日本語誤用例集接続詞・副詞. 大阪大学：データベース版、国立国語研究所, 1990.
- [19] 水本智也, 小町守, 松本裕治. 大規模添削コーパスを用いた統計的機械翻訳手法による日本語誤り訂正. 言語処理学会第 17 回年次大会, pp. 1095–1098, 2011.
- [20] 藤野拓也, 水本智也, 小町守, 永田昌明, 松本裕治. 日本語学習者の作文の誤り訂正に向けた単語分割. 言語処理学会第 18 回年次大会, pp. 26–29, 2012.
- [21] 大曾美恵子, 杉浦正利, 市川保子, 奥村学, 小森早江子, 白井英俊, 滝沢直宏, 外池俊幸. 日本語学習者の作文コーパス：電子化による共有資源化. 言語処理学会第 3 回年次大会論文集, pp. 131–145, 1997.
- [22] 李在鎬, 林● (火偏に玄) 情, 宮岡弥生, 柴崎秀子. 言語処理の技術を利用したタグ付き日本語学習者コーパスの構築. 2012 年度日本語教育学会春季大会予稿集, 2012.
- [23] 永田亮, 若菜崇宏, 河合敦夫, 森広浩一郎, 榊井文人, 井須尚紀. 可算／不可算名詞の判定に基づいた英文の誤り検出. 電子情報通信学会論文誌 J89-D(8), pp. 1777–1790, 2006.
- [24] 南保亮太, 乙武北斗, 荒木健治. 文節内の特徴を用いた日本語助詞誤りの自動検出・校正. 情報処理学会研究報告 自然言語処理研究報告, pp. 107–112, 2007.