

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

小山 碧海 喜友名 朝視 顕 小林 賢治 新井 美桜 小町 守

首都大学東京

{koyama-aomi@ed., kiyuna-tomoshige@ed., kobayashi-kenji1@ed., arai-mio@ed.,
komachi}@tmu.ac.jp

1 はじめに

文法誤り訂正とは、第二言語学習者文を入力として受け取り、それに含まれる誤りを訂正した文を出力するタスクである。また、文法誤り訂正システムの自動評価には、リファレンスを用いる手法が多い [1, 3, 10]。それらを用いる場合、信頼性の高い評価を行うためには質の高い評価データが必要となる [11]。

機械翻訳を利用した文法誤り訂正システムでは、大規模なコーパスを訓練データに用いるものが多い [4, 5, 8]。文法誤り訂正システムの訓練データとして使用されている大規模なコーパスの1つに NAIST Lang-8 Learner Corpora (Lang-8 コーパス) [9] がある。このコーパスには様々な言語における学習者文とその訂正文が収録されている。日本語学習者文は英語学習者文に次いで2番目に多く収録されており、その数は約130万文である。水本ら [9] は、Lang-8¹のデータの特徴として、訂正文の中に訂正者のコメントが含まれる場合があるという点を挙げている。コメントは学習者にとって有益な情報であるが、文法誤り訂正システムの評価データとしてはノイズとなる。また、彼らは訂正者が学習者文に含まれる誤りの一部だけしか訂正していない場合があるという点も挙げている。これらの理由により、Lang-8 コーパスの訂正文は文法誤り訂正システムの評価データとしては適していない。

そこで、本研究では、Lang-8 コーパスに収録されている日本語学習者文を用いて、日本語学習者の文法誤り訂正システムの評価データに適したコーパスを作成する。我々は、訂正基準を設定し、人手で文法誤り訂正を行う。また、学習者文に対して文法的に正しくなるような最低限の訂正を行う。Lang-8 コーパスでは、1つの学習者文に対して複数の訂正文が付与されている場合もあるが、1文のみ付与されている場合が多く、システム評価のためには1つの訂正文のみでは

不十分である。そのため、我々は、1つの学習者文につき複数の訂正文を付与する。

本研究の主要な貢献は、日本語学習者の文法誤り訂正システムのための評価用マルチリファレンスコーパスを作成し、公開した²ことである。

2 関連研究

「日本語学習者による日本語作文と、その母語訛との対訳データベース」(作文対訳 DB) [15] は学習者の手書き作文をもとに作成された日本語学習者コーパスである。各作文は日本語教師によって文法誤り訂正が行われている。また、NAIST 誤用コーパス [13] は作文対訳 DB に収録されている各作文に対して誤用タグを付与したコーパスである。日本語をタイプ入力する際、学習者は仮名漢字変換システムを使用するため、手書き文とタイプ入力文では、誤りの種類の傾向が異なる。そのため、作文対訳 DB 及び NAIST 誤用コーパスはタイプ入力文を対象とした文法誤り訂正システムの評価データに適しているとは言えない。我々は、タイプ入力文に対して文法誤り訂正を人手で行い、タイプ入力文を対象とした文法誤り訂正システムの評価データに適したコーパスを作成する。

劉ら [7] は、日本語機能表現を対象とした文法誤り訂正システムの研究のため、Lang-8 コーパスに収録されている日本語学習者文に対し、訂正対象を機能表現に限定して文法誤り訂正及び誤用タグ付けを人手で行っている。しかしながら、そのデータは公開されていない。我々は、訂正対象を限定せずに人手で文法誤り訂正を行い、コーパスとして公開する。

NUS Corpus of Learner English (NUCLE) [2] はシンガポールの大学生によって書かれた約1,400の作文からなる英語学習者コーパスである。各作文は英語を母語とする英語教師によって文法誤り訂正が行われ、

¹<http://lang-8.com>

²<https://forms.gle/roMnZdq1EKWSM2D9>

表 1: 設定した訂正基準をもとに学習者文を訂正した例.

訂正基準	学習者文	訂正文
L1	デザートを食べながら、チャンペンを飲みました。	デザートを食べながら、シャンパンを飲みました。
L2	そのあと、スポットライトと言うクラフトの店を見に行きました。	そのあと、スポットライトというクラフトの店を見に行きました。
G1	私のうちは四人家族です、僕と母と兄と婆ちゃんです、ちょっと狭いアパートに住んでいる。	私のうちは四人家族です。僕と母と兄と婆ちゃんです。ちょっと狭いアパートに住んでいます。
G2	昨日の成績みたら、失敗しました。	昨日の成績をみたら、失敗しました。
G3	ただちょっと心配してるね。。	ただちょっと心配してるの。。

誤用タグが付与されている。また、学習者文に対して文法的に正しくなるような最低限の訂正が行われている。我々も、日本語学習者文に対して文法的に正しくなるような最低限の訂正を人手で行う。

JHU FLuency-Extended GUG corpus (JFLEG) [12] は母語や習熟度の異なる学習者が書いた約 750 文からなる英語学習者コーパスである。NUCLE とは異なり、文全体の流暢さを考慮しつつ文法誤り訂正が行われている。また、クラウドソーシングを用いて、1 つの学習者文につき 4 つの訂正文が付与されている。我々は、クラウドソーシングを用いず、1 つの学習者文に対して 2 人以上が訂正を行い、1 つの学習者文につき 2 つ以上の訂正文を付与する。

3 アノテーション

3.1 データ

本研究では、Lang-8 コーパスに収録されている日本語学習者文を用いて、日本語学習者の文法誤り訂正システムの評価データに適したコーパスを作成した。まず、Lang-8 コーパスから日本語学習者文 192,673 文章 (1,296,114 文) を抜き出した。その後、コーパスを作成するため、抜き出した文章から 139 文章 (2,042 文) をランダムに抽出した。次に、訂正基準を設定するため、ランダムに抽出した文章のうち 16 文章 (207 文) に対して、日本語を母語とする情報系の大学生 3 人が人手で文法誤り訂正を行った。これ以降、ここで 3 人が訂正を行った 16 文章をコアデータと呼ぶこととする。訂正基準を決定した後、1 つの文章につき 2 人が訂正を行うように残りの 123 文章 (1,835 文) を 3 人に割り振った。訂正が一致しなかった場合、訂正者間で議論をした後、最終決定を行った。

3.2 訂正基準

本研究では、文単位ではなく文章単位で文法誤り訂正を行うこととした。また、学習者文が文法的に正しくなるような最低限の訂正を行った。1 つの文章に対して複数の解釈が可能な場合には、それぞれの訂正者の解釈で訂正を行った。

設定した訂正基準は以下の通りである。表 1 に設定した訂正基準をもとに学習者文を訂正した例を示す。Lang-8 コーパスに収録されている文章の多くは学習者が日記を書くような感覚で書いた日記風の文章である。そのため、本研究では、日記風の文章であることを考慮して、訂正基準の G2 と G3 を設定した。

文単位の訂正基準

- L1 日本語以外の単語が本来の発音のまま表記してある場合には、日本で一般的に使われている表記³に訂正する。
- L2 補助用言が漢字で書かれている場合には、ひらがなに訂正する。

文章単位の訂正基準

- G1 1 つの文章内では常体・敬体を揃えるように訂正する。
- G2 格助詞が抜けている場合でも不自然でない場合⁴には訂正しない。
- G3 終助詞は不自然でない場合⁴には訂正しない。

³検索エンジンを用いて、Web 上に使用例があるかどうかを調査し、一般的に使われている表記かどうかの判断をした。

⁴不自然かどうかの判断はそれぞれの訂正者が行った。

表 2: コアデータ中の学習者文に対する各訂正者の訂正文及び元データの文字単位の編集距離の平均と分散.

	訂正者 A	訂正者 B	訂正者 C	元データ
平均	1.58	1.70	1.66	3.88
分散	5.12	6.83	6.32	75.3

3.3 分析

定量評価 文単位の文法誤り検出の訂正者間一致率を測るため、コアデータに対して Fleiss' kappa を計算した結果、その値は 0.72 であった。したがって、文単位の文法誤り検出の訂正者間一致率は高い [6]。

表 2 にコアデータに対する各訂正文及び元データ⁵の文字単位の編集距離の平均と分散を示す。3 人の訂正者の編集距離の平均と分散は同程度である。また、元データの編集距離の平均と分散はいずれの訂正者のものよりも大きい。これは、元データには訂正者のコメントが含まれる場合があるためである。本研究の訂正者は学習者文に対して文法的に正しくなるような最低限の訂正を行っているため、それぞれの訂正者の編集距離の平均と分散は元データに比べて小さい。

表 3 にコアデータ中のそれぞれの誤りの種類の個数を示す。括弧内は文章単位では誤りになるが文単位では誤りにならないものの個数である。1 つの文章内では常体・敬体を揃える必要があるが文単位ではその必要がないため、常体・敬体の誤りは文単位では誤りにならない。また、時制は文間文脈によって変える必要があるため [14]、文章単位では誤りになるが文単位では誤りにならなかった。

表 4 の 2 カラム目に 3 人の訂正が一致した例を示す。この例では、助詞「が」の誤り訂正が一致している。助詞の誤りは構文上の誤りであるため、助詞の誤り訂正の訂正者間一致率は 55.6% と高かった⁶。

定性評価 表 4 の 3 カラム目に 3 人の訂正が一致しなかった例を示す。「結果がない恋」という表現は一般的でないため、3 人の訂正者はこのフレーズを訂正している。このフレーズには誤りと考えられる箇所が 2 箇所ある。1 つ目は助詞「が」の誤りである。2 つ目は「結果」と「恋」のコロケーションの誤りである。1 つ目の誤りは 3 人とも訂正を行った。一方で、2 つ目の

⁵元データとは、Lang-8 コーパスに収録されている訂正文を指す。1 つの学習者文に対して 2 つ以上の訂正文が付与されている場合は 1 つ目の訂正文を元データに用いた。

⁶助詞の誤り検出の訂正者間一致率は 69.4% であった。

表 3: コアデータ中の誤りの種類とその個数。括弧内は文章単位では誤りになるが文単位では誤りにならないものの個数。

大分類	小分類	誤り数
語彙選択	助詞	36 (10)
	動詞	13 (4)
	名詞	6 (1)
	形容詞	5 (0)
	副詞	4 (0)
	助動詞	3 (0)
	接続詞	3 (1)
	連体詞	2 (0)
	指示詞	1 (1)
	その他	2 (0)
過不足	脱字	22 (2)
	挿入	10 (0)
	省略	8 (3)
表記	誤入力	9 (0)
	不適切な表記	8 (0)
	翻字	3 (0)
運用	アスペクト	6 (0)
	時制	4 (4)
	活用	2 (0)
	音便	1 (0)
文全体	常体・敬体	23 (23)
	接続	18 (0)
	句読点	14 (0)
	語順	5 (1)
	その他	2 (0)
	合計	210 (50)

誤りは訂正者 A と訂正者 C は訂正を行ったが、訂正者 B は訂正を行わなかった。また、「結果がない恋」を訂正者 A は「実らない恋」、訂正者 C は「成就しない恋」とそれぞれ訂正している。このように、訂正者ごとに誤りとするかどうかの判断が異なる場合や異なる単語を用いて訂正を行う場合があった。

3.4 NAIST 誤用コーパスとの比較

助詞の誤りの割合は NAIST 誤用コーパスとコアデータの両方で高く、前者での割合は 22.1% であり [13]、後者での割合は 17.1% である。これは、日本語学習者にとって助詞を正しく使うことが難しいためである。

表記や文体の誤り⁷は NAIST 誤用コーパスとコアデータで割合が大きく異なる。文字の書き間違いは手書き作文をもとにしている NAIST 誤用コーパス特有

⁷大山ら [13] では、誤用タグの「表記」を文字の書き間違い、「文体」を常体・敬体の誤りとしている。

表 4: 3 人の訂正が一致した例と一致しなかった例.

	訂正が一致した例	訂正が一致しなかった例
学習者文	夜、ラベンダーが室内に置きます。	このストーリーは、結果がない恋についてのちょっと悲しいストーリーです。
訂正者 A	夜、ラベンダーを室内に置きました。	このストーリーは、実らない恋についてのちょっと悲しいストーリーです。
訂正者 B	夜、ラベンダーを室内に置きます。	このストーリーは、結果のない恋についてのちょっと悲しいストーリーです。
訂正者 C	夜、ラベンダーを室内に置きます。	このストーリーは、成就しない恋についてのちょっと悲しいストーリーです。

の誤りであり、その割合は 15.7% である。一方で、コアデータ中の文はタイプ入力文であるため、文字の書き間違いは起きないが誤入力が起き、その割合は 4.29% である。日本語をタイプ入力する際、学習者は仮名漢字変換システムを使用するため、コアデータでの誤入力の割合は NAIST 誤用コーパスでの表記の誤りの割合に比べて低い。文体の誤りの割合は、NAIST 誤用コーパスでは 3.9% であるが、コアデータでは 11.0% と比較的高い。これは、NAIST 誤用コーパスの文章は小論文をもとにしているのに対して、Lang-8 コーパスの文章は日記風の文章であるためだと考えられる。

4 おわりに

本研究では、日本語学習者の文法誤り訂正システムのための評価コーパスを作成し、公開した。本コーパスでは、学習者文に対して文法的に正しくなるような最低限の訂正を手で行っている。そして、1つの学習者文につき複数の訂正文を付与している。したがって、本コーパスは日本語学習者の文法誤り訂正システムの評価データに適している。また、今後の課題として、誤用タグ付け [13] が挙げられる。本研究では、コアデータ中の誤りの種類とその個数を調べた。これを発展させ、本コーパスの全文に対して誤りの種類とその個数を調べ、誤用タグを付与することに取り組みたい。

謝辞

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。

参考文献

- [1] Daniel Dahlmeier and Hwee Tou Ng. Better evaluation for grammatical error correction. In *Proceedings of NAACL*, pp. 568–572, 2012.
- [2] Daniel Dahlmeier, Hwee Tou Ng, and Siew Mei Wu. Building a large annotated corpus of learner English: The NUS corpus of learner English. In *Proceedings of BEA*, pp. 22–31, 2013.
- [3] Mariano Felice and Ted Briscoe. Towards a standard evaluation method for grammatical error detection and correction. In *Proceedings of NAACL*, pp. 578–587, 2015.
- [4] Tao Ge, Furu Wei, and Ming Zhou. Fluency boost learning and inference for neural grammatical error correction. In *Proceedings of ACL*, pp. 1055–1065, 2018.
- [5] Roman Grundkiewicz, Marcin Junczys-Dowmunt, and Kenneth Heafield. Neural grammatical error correction systems with unsupervised pre-training on synthetic data. In *Proceedings of BEA*, pp. 252–263, 2019.
- [6] J. Richard Landis and Gary G. Koch. The measurement of observer agreement for categorical data. In *Proceedings of Biometrics*, pp. 159–174, 1977.
- [7] Jun Liu, Fei Cheng, Yiran Wang, Hiroyuki Shindo, and Yuji Matsumoto. Automatic error correction on Japanese functional expressions using character-based neural machine translation. In *Proceedings of PACLIC*, pp. 394–403, 2018.
- [8] Yu-Chun Lo, Jih-Jie Chen, Chingyu Yang, and Jason Chang. Cool English: a grammatical error correction system based on large learner corpora. In *Proceedings of COLING*, pp. 82–85, 2018.
- [9] Tomoya Mizumoto, Mamoru Komachi, Masaaki Nagata, and Yuji Matsumoto. Mining revision log of language learning SNS for automated Japanese error correction of second language learners. In *Proceedings of IJCNLP*, pp. 147–155, 2011.
- [10] Courtney Napoles, Keisuke Sakaguchi, Matt Post, and Joel Tetreault. Ground truth for grammatical error correction metrics. In *Proceedings of ACL-IJCNLP*, pp. 588–593, 2015.
- [11] Courtney Napoles, Keisuke Sakaguchi, and Joel Tetreault. There’s no comparison: Reference-less evaluation metrics in grammatical error correction. In *Proceedings of EMNLP*, pp. 2109–2115, 2016.
- [12] Courtney Napoles, Keisuke Sakaguchi, and Joel Tetreault. JFLEG: A fluency corpus and benchmark for grammatical error correction. In *Proceedings of EACL*, pp. 229–234, 2017.
- [13] Hiromi Oyama, Mamoru Komachi, and Yuji Matsumoto. Towards automatic error type classification of Japanese language learners’ writings. In *Proceedings of PACLIC*, pp. 163–172, 2013.
- [14] Toshikazu Tajiri, Mamoru Komachi, and Yuji Matsumoto. Tense and aspect error correction for ESL learners using global context. In *Proceedings of ACL*, pp. 198–202, 2012.
- [15] 井上優, 宇佐美洋, 成田高宏, 鎌水兼貴. 作文対訳データベースの多様な利用のために. 独立行政法人国立国語研究所, 第 1 版, 2006.