日本語文法誤り訂正におけるデータ増強および評価データ構築

加藤秀佳　岡部格明　北野道春　宿久洋
同志社大学
{khideyoshi2,mokab.0328,kitano.michiharu}@gmail.com,hyadohis@mail.doshisha.ac.jp

1　はじめに

文法誤り訂正 (Grammatical Error Correction: GEC) とは，入力したテキストの誤りを自動で検出して訂正することを目的とした自然言語処理タスクの 1 つである [1]．近年，校正業務におけるコスト削減が期待できることからビジネス分野での需要も高まり，そのための文章校正システムの開発が求められている．しかし，既存の文章校正システムは助詞の誤字，脱字，衍字などの誤り箇所の検出 (detect) のみに焦点が当てられており，文章の誤りを検出し，訂正 (correct) まで行う日本語 GEC システムはほとんど提案されていない．日本語 GEC 研究は英語や中国語といった他言語の GEC 研究と比較すると取り組みが少なく，それは日本語 GEC 研究の性質や環境などに様々な課題が存在することに起因する．そこで本研究では，日本語 GEC 研究が抱える**課題 1** 学習に利用できるデータ不足および**課題 2** 評価データの不十分性という 2 つの課題に対処することで GEC モデルの精度向上と頑健な評価に取り組む．

課題 1 について，近年の GEC 研究は文法的に誤りを含む文 (src) から文法的に正しい文 (tgt) への翻訳タスクとして取り組むことが一般的であり，その際に，src と tgt がペアとなった大量のデータ (対訳データ) が必要とされる [2]. 英語 GEC 研究において利用できる大規模な対訳データは，EFCAMDAT(約 250 万文)[3] など，合計して 450 万文以上用意されている一方で，日本語 GEC 研究において利用できる対訳データは，Lang8 コーパス [4] の約 160 万文のみであり，量および種類の観点からも対訳データが少ないことが問題視されている．この問題に対して，既存研究では用意したコーパスから src を生成し擬似的に対訳データを生成するデータ増強 (Data Augmentation: DA) が提案されている [5, 6]．しかし，新たな生成元コーパスを大量に準備することができない，または使用できるドメインのデータ量が限られている状況においては src を生成する既存研究では対応できない．そこで，本研究では src の生成だけではなく，tgt の生成も考慮に入れた DA(BERT-DA-tgt) を提案することで**課題 1** に対処することを考える．

課題 2 について，英語の GEC 研究において，モデルを評価するために (1) 誤りタイプごとの評価と (2) 複数ドメインごとの評価の 2 つの観点から行う研究やワークショップが数多く行われている [7, 8, 9]. (1) については，提案された GEC モデルが有効に作用する，または作用しないような誤りの性質を明らかにし，GEC システムを構築する際に，特定の誤り種類と相性の良いモデルを採用したいという意図がある [10]．(2) については，特定のドメインで高い性能を示すモデルが，異なるドメインで低い性能を示すような状況が生じた場合，構築されたモデルを正しく評価できたとは言えず，複数ドメインでの評価をしたいという意図がある [11]．近年の日本語 GEC モデルを評価するために広く使われているデータセットとして，NIST 誤用コーパス (NIL) [12, 13] と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 (TEC_JL) [14, 15] があるが，日本語 GEC 研究において GEC モデルを英語のように上記の 2 つの観点両方で評価した研究はない．そこで，本研究では (1)，(2) 両方の観点から評価できるような新たなデータセット (JGECM) の構築および構築方法の提案を行うことで，**課題 2** に対処し，さらに構築したデータを公開[1)]する．

2　BERT-DA-tgt

DA を使用しない GEC 手法は，図 1 における対訳データ $\mathcal{D} = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)|i = 1, 2, ..., n\}$ のみを用いてモデルを学習する．このとき，$\mathcal{D}$ は src の集合 $\mathcal{X} = \{\boldsymbol{x}_i|i = 1, 2, ..., n\}$ と，tgt の集合 $\mathcal{Y} = \{\boldsymbol{y}_i|i = 1, 2, ..., n\}$ の組から構成される．ここで，既存の DA は，図 1 における対訳データ $\mathcal{D}$ とは別の疑似データ $\mathcal{D}^a$ を生成し，モデルの学習に利用するデータ量を

1）https://github.com/hideyoshikato/JGECM

増やす手法のことである [2]．このとき，$\mathcal{D}^a$ は tgt の集合 $\mathcal{Y}^a = \{\boldsymbol{y}_j^a | j = 1, 2, ..., m\}$ から，src の集合 $\mathcal{X}^a = \{\boldsymbol{x}_j^a | j = 1, 2, ..., m\}$ を Direct Noise[16] などの手法によって生成される．

一方で，BERT-DA-tgt は，図 2 に示すとおり，対訳データ $\mathcal{D}$，疑似データ $\mathcal{D}^a$ のどちらとも異なる新たな疑似データ $\mathcal{D}^{a'} = \{(\boldsymbol{x}_k^{a'}, \boldsymbol{y}_k^{a'}) | k = 1, 2, ..., S \times L\}$ を生成する DA であり，大きく「$\boldsymbol{y}_k^{a'}$ の生成 ($generate\ \boldsymbol{y}_k^{a'}$)」と「$\boldsymbol{x}_k^{a'}$ の生成 ($generate\ \boldsymbol{x}_k^{a'}$)」という 2 つのアルゴリズムによって疑似データが生成される (Algotithm1)．S は $\mathcal{Y}^a$ からサンプリングする文の数を表すパラメータ，L は，BERT の Masked Language Model(Masked LM) における予測確率の順位を表し，DA を行う際のデータ量を調整するパラメータである．表 1 に生成した $\mathcal{D}^{a'}$ の具体例を示す．また，BERT-DA-tgt は様々な NLP タスクにおいて DA の有効性が示されている BERT の Masked LM を用いているため，以下の 3 つのメリットを得る．

1. BERT は文脈を考慮できるため，生成されたテキストは文法的 / 意味的な整合性を持つ特徴を得ることができる．
2. L の数を分析者が指定することで，学習に利用するデータ量を調整できる．
3. 豊富な語彙，多様な言い回しにより使用できるドメインや生成元コーパスの量が限られた状況での精度向上が期待できる．

Algorithm 1 BERT-DA-tgt による擬似データ生成アルゴリズム

Input: $\mathcal{Y}^a = \{\boldsymbol{y}_j^a | j = 1, 2, ..., m\}$, S, L
Output: $\mathcal{D}^{a'} = \{(\boldsymbol{x}_k^{a'}, \boldsymbol{y}_k^{a'}) | k = 1, 2, ..., S \times L\}$

$\mathcal{Y}^a$ から S 行サンプリングしたものを $\mathcal{Y}^{a'} = \{\boldsymbol{y}_k^a | k = 1, 2, ..., S\}$ とする
for $k = 1, ..., S$ **do**
  $generate\ \boldsymbol{y}_k^{a'}$
    $\boldsymbol{y}_k^a$ におけるマスクするトークンの Z を離散一様分布からランダムに指定
    指定した Z を BERT の Masked LM で L 番目までのトークンに置換したものをそれぞれ $\boldsymbol{y}_k^{a'(1)}, \boldsymbol{y}_k^{a'(2)}, \ldots, \boldsymbol{y}_k^{a'(L)}$ とする
  generate $\boldsymbol{x}_k^{a'}$
    生成された $\boldsymbol{y}_k^{a'(1)}, \boldsymbol{y}_k^{a'(2)}, \ldots, \boldsymbol{y}_k^{a'(L)}$ それぞれに対して Direct Noise を適用したものを $\boldsymbol{x}_k^{a'(1)}, \boldsymbol{x}_k^{a'(2)}, \ldots, \boldsymbol{x}_k^{a'(L)}$ とする
end for

表 1　BERT-DA-tgt の疑似データの具体例

$\boldsymbol{y}_k^a$	L	$\boldsymbol{y}_k^{a'}$
ライセンス契約の表示があります．	1	ライセンス契約の表示が現れます．
	2	ライセンス契約の表示が消えます．
	3	ライセンス契約の表示がなくなります．
	4	ライセンス契約の表示ができます．
ヘッケル自慢のレシピ	1	ヘッケル自慢のメニュー
	2	ヘッケル自慢の料理
	3	ヘッケル自慢のレストラン
	4	ヘッケル自慢のシェフ

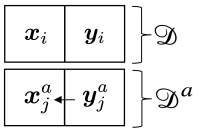

図 1　既存の DA

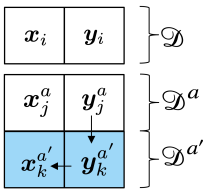

図 2　BERT-DA-tgt

3　JGECM

本節では，複数誤りタイプをもつ日本語文法誤り訂正のための評価用コーパス (JGECM) の構築手順および分析を行う．

3.1　構築手順

評価データは以下の 5 つの手順により構築する．

手順 1　BCCWJ [17] の ‘C-XML/VARIABLE/PN/’ に含まれる ‘.xml’ ファイルから `sentence` タグで囲まれたテキストをクローリング
手順 2　19 文字以上 50 文字未満の文を抽出
手順 3　「句点を含む文」を抽出
手順 4　ランダムに 500 文抽出
手順 5　指示書に基づき 10 名のアノテータが 1 文に対して 8 種類の誤りを付与

手順 1 では，本研究では校正者が文法的な正しさが担保されていると考えられる新聞 (PN) レジスターを評価データとして設定し，`sentence` タグで囲まれた合計 51,977 文が抽出された．**手順 2** では，機械翻訳では構文の曖昧さから，長い文を翻訳することへの問題点が挙げられている [18]．また，長すぎる文は読点や区切り文字などであらかじめ分割しておくことで，問題に対処することができることから，本研究では，評価データの文字数を 19 文字以上 50 文字未満と設定し，合計 17,446 文が抽出された．**手順 3** では，評価対象を文とするために，句点を含まないタイトルや見出しなどを省くための手順である．その結果，合計 15,236 文が抽出された．**手順 4** では，訂正性能を評価するために一定のデータ数が必要であると判断したため，[4] を参考にラン

ダムに 500 文用意した．**手順 5** では，今回，[13] を参考に削除 (助詞)，挿入 (助詞)，置換 (助詞)，語彙選択，表記，動詞，削除 (助詞・動詞以外)，挿入 (助詞・動詞以外) という 8 種類の誤りを付与した．また，誤り箇所を明確にするために，誤りを付与した箇所を"["と"]"で挟むように指示を行った．実際に用いた指示書および構築された JGECM の具体例については，付録 A，B に示す．

3.2 分析

構築した評価データの欠損値の数については表 2 に示す．データに欠損値が含まれる理由は，誤りを付与する対象となる品詞の中に助詞や動詞が存在しない場合，セルの中身を空になるようにアノテータに指示しているためである．評価データの文字数の分布は図 3 の通りである．

表 2 欠損値の数

列名	欠損の数
削除 (助詞)	13
挿入 (助詞)	13
置換 (助詞)	13
語彙選択	0
表記	3
動詞	47
削除 (助詞・動詞以外)	0
挿入 (助詞・動詞以外)	4

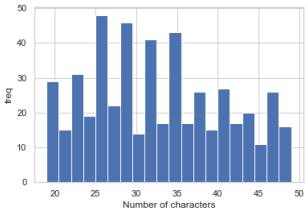

図 3 評価データの文字数の分布

4 実験

実験は，(i) 既存モデル (SMT に基づく手法および NMT に基づく手法) と提案モデル (BERT-DA-tgt を行った GEC モデル) の性能を比較し提案モデルの有効性を示すこと，(ii) **要素 (1)** データ生成元, **要素 (2)** データ生成量に関する比較実験を行い，提案した DA 手法 (BERT-DA-tgt) がより効果的に作用する知見を見つけ出すこと，の 2 つを目的として行う．ハイパーパラメータ，最適化，分割単位に関しては [5] の設定を用いる．

4.1 実験設定

データセット 学習データ，開発データには Lang8 コーパスを利用する．評価データには，NIL, TEC_JL，第 3 節で構築した JGECM の 3 種類を用いる．DA の生成元コーパスとして，BCCWJ [17] からクローリングした一般書籍 (BCCWJ-PB)，図書書籍 (BCCWJ-LB)，Yahoo!知恵袋 (BCCWJ-OC) という 3 種類の異なるドメインのデータを使用する．それぞれのデータの詳細については表 3 に示す．

表 3 各データの概要

データセット	カテゴリ	文 (ペア) の数	ドメイン
Lang8	学習 (train)	1,654,399	SNS
Lang8	開発 (valid)	5,000	SNS
NIL	評価 (test)	6,672	小論文
TEC_JL	評価 (test)	1,835	SNS
JGECM	評価 (test)	500	新聞
Lang8	疑似 (pseudo)	600,000	SNS
BCCWJ-LB	疑似 (pseudo)	891,179	一般書籍
BCCWJ-PB	疑似 (pseudo)	774,808	図書書籍
BCCWJ-OC	疑似 (pseudo)	478,010	Yahoo!知恵袋

モデル 既存の Encoder-Decoder モデルである Transformer に，Copy 機構を組むことで提案された TransformerCopy [16] を採用して実験を行う．

比較手法 比較手法には統計的機械翻訳 (SMT) ベースの GEC 手法として Moses [14, 19]，ニューラル機械翻訳 (NMT) ベースの GEC 手法として CNN [14, 20]，Bi-LSTM [21]，そして，最後に疑似データを考慮しない通常の TransfomerCopy [5, 16] をベースラインとする．

評価指標 GEC システムの評価には，出力文と src と tgt の 3 つを用いて行う参照有り評価手法と tgt を用いない参照無し評価手法が存在する [22]．本研究では，英語 Shared Task において最も用いられている前者の M^2 scorer [23] と GLEU [24] の結果を示す．

4.2 要素 (1)：生成元コーパス

生成元コーパスとして，出版書籍 (PB)，図書館書籍 (LB)，Yahoo!知恵袋 (OC) を用いた場合のモデルの性能を比較する．ベースラインは，通常の DA および事前学習を行わない TransformerCopy である．また，今回データ量の違いが性能に影響を与えることを防ぐために，それぞれのレジスターからランダムに 300,000 文を取得し，BERT-DA-tgt により DA したデータ量 $|\mathcal{D}^a \cup \mathcal{D}^{a'}| = 600,000$ に設定している．$\mathcal{D}^a \cup \mathcal{D}^{a'}$ は既存手法で用意できる $\mathcal{D}^a$ と提案手法で用意した $\mathcal{D}^{a'}$ を結合した事前学習に用いるデータである．各生成データごとのモデル性能比較に関する実験結果を表 4 に示す．実験の結果，TEC_JL は OC で，NIL，WGECM は PB で事前学習したモデルが最も性能で優れた．つまり，評価用データと近い性質を持つコーパスで事前学習を行うことで，精度向上が期待できる可能性を示唆している．

表 4　生成元コーパスごとの GLEU による性能比較

	TEC-JL	NIL	JGECM
ベースライン	78.67	56.86	85.63
Trans+BERT_OC	**81.04**	58.56	86.99
Trans+BERT_PB	78.31	**58.66**	**87.23**
Trans+BERT_LB	76.04	58.25	86.70

4.3　要素 (2)：疑似データ生成量

疑似データ生成量の違いによりモデル性能への影響について検証する．今回，$K = 300,000$ として生成元コーパスからサンプリングし，$L = 0, 1, 2, 3, 4$ の場合を報告している．$L = 0$ は DA を行わないモデルを表している．実験結果を表 5 に示す．$F_{0.5}$，GLEU は全ての評価データにおいて $L = 2$ のときに最も優れていたことが確認できる．これは，L の数を増やすほど予測確率が低いトークンに置換された文が生成されるため，tgt の文構造の崩壊を招きやすくなり性能が下がる可能性が示唆される．

表 5　データ生成量の違いによる性能比較

NIL

L	$\lvert\mathbb{D}^a \cup \mathbb{D}^{a'}\rvert$	Pre.	Rec.	$F_{0.5}$	GLEU
0	300,000	30.28	3.38	11.69	56.86
1	600,000	35.04	7.18	19.73	58.66
2	900,000	**37.33**	7.32	**20.52**	**58.91**
3	1,200,000	35.41	7.16	19.79	58.68
4	1,500,000	35.08	**7.49**	20.21	58.71

JGECM

L	$\lvert\mathbb{D}^a \cup \mathbb{D}^{a'}\rvert$	Pre.	Rec.	$F_{0.5}$	GLEU
0	300,000	41.02	5.78	18.49	85.63
1	600,000	55.84	19.32	40.52	87.23
2	900,000	**58.57**	20.19	**42.44**	**87.37**
3	1,200,000	56.51	19.78	41.20	87.25
4	1,500,000	57.05	**20.60**	42.14	87.34

4.4　既存システムとの性能比較

要素 (1),(2) で得られた知見の下，BERT-DA-tgt の生成元コーパスを BCCWJ-PB，$L = 2$ とした設定により構築したモデル (Trans+BERT_PB) と比較手法の性能比較を行う．DA を行わない通常の TransofrmerCopy によるモデル (ベースライン)，事前学習とファインチューニングを Lang8 で行うモデル (Trans+BERT_lang8) を比較手法とした結果についても記載する．実験結果を表 6 に示す．Trans+BERT_PB は，すべての評価データにおいて既存システムよりも性能で上回っている．加えて，提案手法を用いた結果，TEC_JL では 1,835 文のうち 541 文，NIL では 6,672 文のうち 1,605 文の訂正が行われた．また，疑似データを考慮したことで多様な語彙や言い回しに対応でき，TransfomerCopy では訂正ができなかった誤りに対しても訂正されている．また，Bi-LSTM のように誤り種類を特定して訂正を行う手法と比較しても，不適切な助詞の誤りや文字の挿入に対応することができ，誤りの種類を特定せずに訂正が行われていることも確認できた．

表 6　既存システムとの GLEU による性能比較

Models	TEC_JL	NIL	JGECM
Moses [14, 19]	73.9	-	-
CNN [14, 20]	73.5	-	-
Bi-LSTM [21]	78.8	44.9	81.8
ベースライン [5, 16]	78.7	56.9	85.6
Trans+BERT_lang8	80.1	58.1	86.4
Trans+BERT_PB	**81.9**	**59.6**	**87.5**

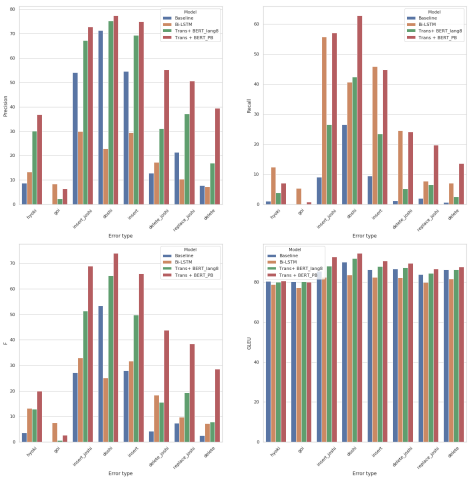

図 4　JGECM における誤りタイプごとの評価[2)]

5　おわりに

日本語 GEC 研究が抱える 2 つの課題に対して，tgt の生成も考慮に入れた DA(BERT-DA-tgt) を提案し，2 つの要素について性能比較を行った．その結果，(1) 生成元コーパスは PB(出版書籍ドメイン) を用いる．(2) BERT の Masked LM における候補は 2 番目まで増やす．という 2 つの知見を獲得した．これらの設定に基づき，事前学習された TransformerCopy によって GEC モデルを構築し性能比較を行った結果，複数ドメインの評価データにおいて既存の GEC システムを上回る性能を達成し，BERT-DA-tgt の有効性を示した．今後の課題として，学習データを逐次的に収集し，モデルを再学習できる GEC システムの構築，入力に用いる分割単位に違いによる性能比較，開発 (valid) データの違いによる性能比較などが挙げられる．

2)　'hyoki'，'goi'，'insert_joshi'，'doshi'，'insert'，'delete_joshi'，'replace_joshi'，'delete'，はそれぞれ表記，語彙選択，挿入 (助詞)，動詞，挿入 (助詞・動詞以外)，削除 (助詞)，置換 (助詞)，削除 (助詞・動詞以外) を表す．

参考文献

[1] Yu Wang, Yuelin Wang, Jie Liu, and Zhuo Liu. A comprehensive survey of grammar error correction. *arXiv preprint arXiv:2005.06600*, 2020.

[2] Shun Kiyono, Jun Suzuki, Masato Mita, Tomoya Mizumoto, and Kentaro Inui. An empirical study of incorporating pseudo data into grammatical error correction. pp. 1236–1242. Association for Computational Linguistics, November 2019.

[3] Jeroen Geertzen, Theodora Alexopoulou, and Anna Korhonen. Automatic linguistic annotation of large scale l2 databases: The ef-cambridge open language database (efcamdat). In *Proceedings of the 31st Second Language Research Forum. Somerville, MA: Cascadilla Proceedings Project*, pp. 240–254, 2013.

[4] 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 5th International Joint Conference on Natural Language Processing*, pp. 147–155, 2011.

[5] 小川耀一朗, 山本和英. 日本語文法誤り訂正における誤り傾向を考慮した擬似誤り生成. 2020.

[6] 白井稔久, 萩行正嗣, 小町守. 擬似誤りコーパスを用いた天気予報原稿のニューラル誤り検出. 人工知能学会全国大会論文集 一般社団法人 人工知能学会, pp. 3Rin242–3Rin242. 一般社団法人 人工知能学会, 2019.

[7] Christopher Bryant, Mariano Felice, and Ted Briscoe. Automatic annotation and evaluation of error types for grammatical error correction. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 793–805, Vancouver, Canada, July 2017. Association for Computational Linguistics.

[8] Hwee Tou Ng, Siew Mei Wu, Ted Briscoe, Christian Hadiwinoto, 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*, pp. 1–14, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[9] Hwee Tou Ng, Siew Mei Wu, Yuanbin Wu, Christian Hadiwinoto, and Joel Tetreault. The CoNLL-2013 shared task on grammatical error correction. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning: Shared Task*, pp. 1–12, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.

[10] 清野舜, 鈴木潤, 三田雅人, 水本智也, 乾健太郎. 大規模疑似データを用いた高性能文法誤り訂正モデルの構築. 2020.

[11] 三田雅人, 水本智也, 金子正弘, 永田亮, 乾健太郎. 文法誤り訂正のコーパス横断評価:単一コーパス評価で十分か？ 2020.

[12] Hiromi Oyama, Mamoru Komachi, and Yuji Matsumoto. Towards automatic error type classification of japanese language learners' writings. In *Proceedings of the 27th Pacific Asia Conference on Language, Information, and Computation (PACLIC 27)*, pp. 163–172, 2013.

[13] 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イプの階層的アノテーションに基づく機械学習による自動分類. 自然言語処理, Vol. 23, No. 2, pp. 195–225, 2016.

[14] Aomi Koyama, Tomoshige Kiyuna, Kenji Kobayashi, Mio Arai, and Mamoru Komachi. Construction of an evaluation corpus for grammatical error correction for learners of Japanese as a second language. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pp. 204–211, Marseille, France, May 2020. European Language Resources Association.

[15] 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. 2020.

[16] Wei Zhao, Liang Wang, Kewei Shen, Ruoyu Jia, and Jingming Liu. Improving grammatical error correction via pre-training a copy-augmented architecture with unlabeled data.

[17] Kikuo Maekawa, Makoto Yamazaki, Toshinobu Ogiso, Takehiko Maruyama, Hideki Ogura, Wakako Kashino, Hanae Koiso, Masaya Yamaguchi, Makiro Tanaka, and Yasuharu Den. Balanced corpus of contemporary written japanese. *Language resources and evaluation*, Vol. 48, No. 2, pp. 345–371, 2014.

[18] Yoon-Hyung Roh, Young-Ae Seo, Ki-Young Lee, and Sung-Kwon Choi. Long sentence partitioning using structure analysis for machine translation. In *NLPRS*, 2001.

[19] Philipp Koehn, Franz J Och, and Daniel Marcu. Statistical phrase-based translation. Technical report, UNIVERSITY OF SOUTHERN CALIFORNIA MARINA DEL REY INFORMATION SCIENCES INST, 2003.

[20] Shamil Chollampatt and Hwee Tou Ng. A multilayer convolutional encoder-decoder neural network for grammatical error correction. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, February 2018.

[21] 高橋諒, 蓑田和麻, 舛田明寛, 石川信行. Bidirectional LSTM を用いた誤字脱字検出システム. 人工知能学会全国大会論文集 一般社団法人 人工知能学会, pp. 3C4J903–3C4J903. 一般社団法人 人工知能学会, 2019.

[22] 浅野広樹, 水本智也, 乾健太郎. 文法性・流暢性・意味保存性に基づく文法誤り訂正の参照無し評価. 自然言語処理, Vol. 25, No. 5, pp. 555–576, 2018.

[23] Daniel Dahlmeier and Hwee Tou Ng. Better evaluation for grammatical error correction. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pp. 568–572, Montréal, Canada, June 2012. Association for Computational Linguistics.

[24] Courtney Napoles, Keisuke Sakaguchi, Matt Post, and Joel Tetreault. Ground truth for grammatical error correction metrics. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pp. 588–593, Beijing, China, July 2015. Association for Computational Linguistics.

A　評価データ構築におけるアノテータに対する指示書

下記の誤りの種類に対する作業指示に従って誤りを付与してください．

[削除 (助詞)]
文に含まれる助詞を削除してください．
助詞とは「が」「を」「に」「の」「と」などのことです．
特にあなたが最も間違いやすいと思うものを削除してください．
削除した箇所に [] を加えてください．
元の文に助詞が無い場合，セルの中身を削除してください．
(例)「英語がわかる」→「英語 [] わかる」

[挿入 (助詞)]
文に含まれる助詞の前後に助詞を挿入してください．
助詞とは「が」「を」「に」「の」「と」などのことです．
特にあなたが最も間違いやすいと思うものを挿入してください．
挿入した文字に [] を加えてください．
元の文に助詞が無い場合，セルの中身を削除してください．
(例)「英語がわかる」→「英語が [を] わかる」

[置換 (助詞)]
文に含まれる助詞を他の助詞に置換してください．
助詞とは「が」，「を」，「に」，「の」，「と」などのことです．
特にあなたが最も間違えやすいと思うものを付け加えてください．
置換した文字に [] を加えてください．
元の文に助詞が無い場合，セルの中身を削除してください．
(例)「英語がわかる」→「英語 [に] わかる」

[語彙選択]
文の中に含まれる 1 つの単語をふさわしくない文字または単語に置換してください．
特にあなたが最も間違いやすいと思うものに置換してください．
置換した文字に [] を加えてください．
(例 1)「権利を尊重する」→「権利を尊 [敬] する」
(例 2)「常に外国の援助がいる」→「[まいにち] 外国の援助がいる」

[表記]
文に含まれる 1 つの単語に対して
漢字変換，平仮名，カタカナ，英字・日本語変換ミス，
適切でない言い換えなどに関する誤りを付与してください．
特にあなたが最も間違いやすいと思うものを付け加えてください．
誤りを付与した文字または単語に [] を加えてください．
(例 1：漢字変換ミス)「異議を申し立てる」→「[意義] を申し立てる」
(例 2：平仮名)「ねんばいの人」→「[ねんば] の人」
(例 3：カタカナ)「レストランを離れる」→「[レストラ] を離れる」
(例 4：英字・日本語変換ミス)「8 %から 10 %へ引き上げる」→「8 %[kara]10 %へ引き上げる」
(例 5：適切でない言い換え)「父を車に乗せる」→「[おやじ] を車に乗せる」

[動詞]
文に含まれる動詞に対して動詞の一部の削除，
または，動詞の一部への文字の挿入，
または，動詞の一部の文字の置換を行ってください．
元の文に助詞が無い場合，セルの中身を削除してください．
削除の場合，削除した箇所に [] を加えてください．
挿入の場合，挿入した文字に [] を加えてください．
置換の場合，置換した文字に [] を加えてください．
(例 1：削除)「レストランを離れる．」→「レストランを離れ [].」
(例 2：挿入)「手紙を書かない」→「手紙を書か [か] ない」
(例 3：置換)「大きくなりました」→「大きく [され] ました」

[削除 (助詞・動詞以外)]
文に含まれる助詞・動詞以外の 1 つの文字または，単語を削除してください．
特にあなたが最も間違いやすいと思うものを削除してください．
削除した箇所に [] を加えてください．
(例 1)「タバコを吸うことは悪いことです」→「タバコを吸う [] は悪いことです」
(例 2)「同志社大学を離れる」→「同志社 [] 学を離れる」

[挿入 (助詞・動詞以外)]
文に含まれる助詞・動詞以外の文字または，単語を挿入してください．
特にあなたが最も間違いやすいと思うものを挿入してください．
挿入した文字に [] を加えてください．
(例 1)「犬は可愛い」→「犬は可愛い [い]」
(例 2)「ご飯を食べます」→「[お] ご飯を食べます」

B　JGECM の具体例および既存システムと提案モデルの出力例の比較

表 7　JGECM の具体例

original sentennce	削除 (助詞)	挿入 (助詞)
関連企業に資材を供給するＳＧ会がある．	関連企業 [を] 資材を供給するＳＧ会がある．	関連企業 [の] に資材を供給するＳＧ会がある．
大企業より，中小企業の数は圧倒的に多い．	大企業 []，中小企業の数は圧倒的に多い．	大企業より，中小企業の数は圧倒的に [を] 多い．
それで掃除の範囲を決めていたこともある．	それで掃除 [] 範囲を決めていたこともある．	それで掃除の範囲 [に] を決めていたこともある．
軍国主義を美化するとの批判は当たらない．	軍国主義 [] 美化するとの批判は当たらない．	軍国主義を [が] 美化するとの批判は当たらない．
これを破った勢いで準々決勝にまで進んだ．	これを破った勢いで準々決勝 [] まで進んだ．	これを [が] 破った勢いで準々決勝にまで進んだ．

置換 (助詞)	語彙選択	表記
関連企業に資材を供給するＳＧ会 [と] ある．	[間柄] 企業に資材を供給するＳＧ会がある．	[間] 連企業に資材を供給するＳＧ会がある．
大企業より，中小企業 [が] 数は圧倒的に多い．	大企業より，中小企業の [点] は圧倒的に多い．	大企業より，[重症] 企業の数は圧倒的に多い．
それで掃除の範囲 [は] 決めていたこともある．	それで掃除の範囲を [定] めていたこともある．	それで掃除の [犯意] を決めていたこともある．
軍国主義を美化するとの批判 [が] 当たらない．	軍国主義を美 [観] するとの批判は当たらない．	軍国主義を美化するとの批判は [充] たらない．
これを破った勢い [の] 準々決勝にまで進んだ．	これを破った [スピード] で準々決勝にまで進んだ．	これを [裁った] 勢いで準々決勝にまで進んだ．

動詞	削除 (助詞・動詞以外)	挿入 (助詞・動詞以外)
関連企業に資材を供給するＳＧ会があ [ら] る．	関連企 [] に資材を供給するＳＧ会がある．	関連企業に資材 [い] を供給するＳＧ会がある．
	大企業 []，中小企業の数は圧倒的に多い．	大企業より，中小企業の数は圧倒的に多い [い]．
それで掃除の範囲を決め [て] ていたこともある．	それで掃除の範囲を決めていた [] もある．	それで掃除 [機] の範囲を決めていたこともある．
軍国主義を美化す [] との批判は当たらない．	軍国主 [] を美化するとの批判は当たらない．	軍国主義を美化するとの批判 [ん] は当たらない．
これを破った勢いで準々決勝にまで進 [す] んだ．	これを破った勢 [] で準々決勝にまで進んだ．	これを破った勢いで準々決勝 [う] にまで進んだ．

表 8　既存システムと提案モデルの出力例の比較

src	**やて**，アナコンダがテリーたちの前に現れ，犠牲者が出る．
tgt	**やがて**，アナコンダがテリーたちの前に現れ，犠牲者が出る．
Bi-LSTM	**しは**，アナコンダがテリーたちの前に現れ，犠牲者が出る．
Trans+BERT_lang8	**やて**，アナコンダがテリーたちの前に現れ，犠牲者が出る．
Trans+BERT_PB	**やがて**，アナコンダがテリーたちの前に現れ，犠牲者が出る．

src	確かに今回の自衛隊の活動には**危**が伴うといわれている．
tgt	確かに今回の自衛隊の活動には**危険**が伴うといわれている．
Bi-LSTM	確かに今回の自衛隊の活動には**力**が伴うといわれている．
Trans+BERT_lang8	確かに今回の自衛隊の活動には**危**が伴うといわれている．
Trans+BERT_PB	確かに今回の自衛隊の活動には**危険**が伴うといわれている．

src	今週中に全五十州の知事に配布される**定**．
tgt	今週中に全五十州の知事に配布される**予定**．
Bi-LSTM	今週中には五十州の知事に配布される**定**．
Trans+BERT_lang8	今週中に全五十州の知事に配布される**定**．
Trans+BERT_PB	今週中に全五十州の知事に配布される**予定**．

src	右太ももの肉離れでリハビリ中の森岡が練習に参加，**順な**回復ぶりを見せた．
tgt	右太ももの肉離れでリハビリ中の森岡が練習に参加，**順調な**回復ぶりを見せた．
Bi-LSTM	右太ももの肉離れでリハビリ中の森岡が練習に参加，**順調**回復ぶりを見せた．
Trans+BERT_lang8	右太ももの肉離れでリハビリ中の森岡が練習に参加，**順な**回復ぶりを見せた．
Trans+BERT_PB	右太ももの肉離れでリハビリ中の森岡が練習に参加，**順調な**回復ぶりを見せた．

src	子どもの体や健康，子育ての悩**や質問**をお寄せください．
tgt	子どもの体や健康，子育ての悩**みや質問**をお寄せください．
Bi-LSTM	子どもの体や健康，子育ての悩**み質問**をお寄せください．
Trans+BERT_lang8	子どもの体や健康，子育ての悩**や質問**をお寄せください．
Trans+BERT_PB	子どもの体や健康，子育ての悩**みや質問**をお寄せください．

src	そのメディアとしてのスポーツ**も**っと組みたい．
tgt	そのメディアとしてのスポーツ**を**もっと組みたい．
Bi-LSTM	そのメディアとしてのスポーツもっと組みたい．
Trans+BERT_lang8	そのメディアとしてのスポーツ**を**もっと組みたい．
Trans+BERT_PB	そのメディアとしてのスポーツ**を**もっと組みたい．

src	地籍調査で（不動産の）権利関係をはっきりさせ，固定資産税きちっと徴収しなければならない．
tgt	地籍調査で（不動産の）権利関係をはっきりさせ，固定資産税をきちっと徴収しなければならない．
Bi-LSTM	地震調査で（不動産の）権利関係をはっきりさせ，固定資産税をもずっと徴収しなければならない．
Trans+BERT_lang8	地籍調査で（不動産の）権利関係をはっきりさせ，固定資産税きちっと徴収しなければならない．
Trans+BERT_PB	地籍調査で（不動産の）権利関係をはっきりさせ，固定資産税をきちっと徴収しなければならない．

src	トカラ列島の悪石島北西沖で米潜水艦ボーフィンの攻撃受け，沈んだ．
tgt	トカラ列島の悪石島北西沖で米潜水艦ボーフィンの攻撃を受け，沈んだ．
Bi-LSTM	トカラ列島の悪石島北西沖で米潜水艦ボーフィンの攻をだけ，沈んだ．
Trans+BERT_lang8	トカラ列島の悪石島北西沖で米潜水艦ボーフィンの攻撃受け，沈んだ．
Trans+BERT_PB	トカラ列島の悪石島北西沖で米潜水艦ボーフィンの攻撃を受け，沈んだ．