

# 和文英訳問題自動添削システムにおける意味内容の添削

松井颯汰<sup>1</sup> 綱川隆司<sup>1</sup> 西田昌史<sup>1</sup> 西村雅史<sup>1</sup>

<sup>1</sup> 静岡大学大学院総合科学技術研究科

matsui.sota.17@shizuoka.ac.jp

{tuna, nishida, nisimura}@inf.shizuoka.ac.jp

## 概要

本研究では、和文英訳問題の意味内容の誤りを自動添削する手法を提案する。BERTをベースとした手法で学習者訳中の意味内容の誤りを検出し、単語アライメントと英和辞典を用いて誤り単語と訂正候補の単語(熟語)を提示するシステムを開発した。実験では、提案システムの誤り検出および誤り訂正に関する精度を評価した。

## 1 はじめに

和文英訳問題は日本語の出題文を英語に翻訳する記述式問題である。これらは、高校・大学入学試験等で問われる問題であるため、和文英訳問題は入学試験に出題されるほど英語の学習において重要である。しかし、和文英訳問題は記述式問題であるため、学習者自身による正誤判定が困難である。そのため、習熟者による添削が必要となるが、家庭での学習時のような習熟者が常に存在する環境ばかりではない。このような背景から、和文英訳問題を含めた記述式問題を自動で採点・添削するシステムが注目され、実用化が求められている。

杉浦[1]によると、和文英訳の回答アプローチは「文法的手段」と「意味内容の表現手段」の2つに分かれているとされている。ここで述べられている「意味内容の表現手段」とは、日本語の意味を汲みとりその内容を英語的な発想を介して英語で表現するものであるということである。よって、本研究での和文英訳で発生する誤りは「文法の誤り」と「意味内容の誤り」とする。その中で、「文法の誤り」を自動添削するシステムはニューラルベースで開発したGrammarly[2]を始めとして盛んに研究が行われている[3]。一方で、「意味内容の誤り」の自動添削に関する研究は希少である、これらのことをふまえて本研究では、「意味内容の誤り」を自動添削することに焦点を当てる。

和文英訳問題を自動採点・添削する研究として、採点項目をあらかじめ用意し、その項目に応じて構築したシステム[4]や、学習者と出題文の機械翻訳文の尤度から採点する[5]がある。しかし、[4]を含めた現在実用化されている和文英訳問題自動添削システムは添削可能な問題に限られており、[5]は添削をせず採点のみにとどまっている。そこで本研究では、任意の問題に対して添削可能な和文英訳問題自動添削システムを開発する。提案システムによって、学習者が習熟者不在の環境下においても十分な添削を受けることを期待している。

自動添削を実現するにあたって、誤りを検出しその誤りを訂正する二つの工程に分ける。そして、検出・訂正ともにBERT[6]をベースにした手法を提案する。誤り検出では、BERTで構築した自然言語推論モデルとSentence-BERT(S-BERT)[7]を組み合わせた手法を用いる。誤り訂正では、BERTをfine-tuneした単語アライメントモデル[8]で学習者訳と出題文との単語アライメントを取り、英和辞典を活用した手法を用いる。具体的な手法は次章以降で述べる。

実験では、提案システムが大学受験レベルの和文英訳問題の答案において、70%以上で誤り文を検出し、その中で55%程度の訂正率であることを示した。

## 2 システム概要

「文法誤り」を添削する[2]によると、「文法誤り」の誤りタイプは「挿入」「置換」「削除」の3つに分けられている。「意味内容の誤り」の添削も誤りはこの三タイプからなるものとする。このうち、「削除」については、学習者は意味的に余分な単語を記述するケースが少ないため、提案システムでは「削除」する機能を実装していない。

また、「意味内容の誤り」にスペルミスも含むものとする。これは、「風呂(bath)」という単語に対し

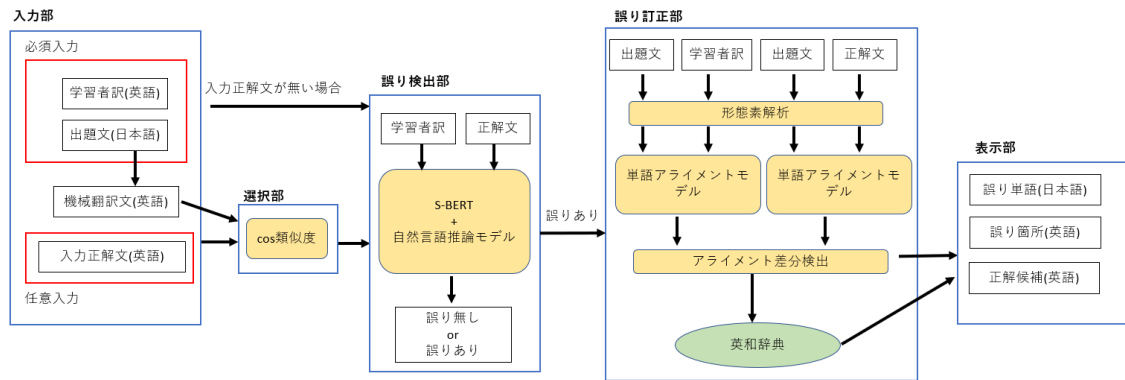


図1 提案システムの処理の流れ

て、“basu(人名)”や“bass(男声最低音)”等で訳した場合の誤りは「文法誤り」システムでは検出・訂正不可能であると考えたからである。

図1に提案システム全体の処理の流れを示す。まず、利用者は出題文(日本語)と学習者訳(英語)を入力する。また、用意可能であれば正解文(英語)も入力する。出題文を翻訳API(本研究ではDeepL APIを利用した)を用いて機械翻訳する。これにより、正解文が用意不可能な場合でも疑似的な正解文を作成する。入力された正解文がある場合、入力正解文、機械翻訳文のそれぞれと学習者訳とのcos類似度を求め、類似度が高い方を誤り検出・訂正用に用いる正解文とする。正解文と学習者訳それぞれの単語をベクトル化したものを $A_i$ 、 $B_i$ とすると、cos類似度は式1のように求める。

$$\text{cos 類似度} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (1)$$

そして、自然言語推論モデルとS-BERTで学習者訳中の意味内容の誤りの有無を検出する。誤りが検出された場合、単語アライメントモデルで出題文と学習者訳・正解文との対応関係をそれぞれ取る。正解文と学習者訳とで対応関係の取れていない単語と英和辞典に記載のない単語を誤りとする。最後に誤り箇所と正解文から抽出した訂正候補の単語を示すシステムとなっている。

### 3 誤り検出

#### 3.1 自然言語推論モデル

本研究では、意味内容の誤り検出に自然言語推論(NLI)を利用する。自然言語推論とは前提文に対して仮定文が「含意」となるのか「矛盾」となるのか

「中立」となるのかを判定するタスクである。

学習者訳と正解文との自然言語推論で「矛盾」がなければ意味内容に誤りがないとし、「矛盾」があれば意味内容に誤りがあるとした。自然言語推論モデルの構築には、約43万対あるラベル付けされたMultiNLIを学習データとし、事前学習モデルにBERT-LARGEモデルを用いてfine-tuneした。

#### 3.2 Sentence-BERT

前述した自然言語推論モデルに加え、本研究ではSentence-BERT(S-BERT)[7]を利用し誤り検出する。S-BERTは二つの類似したテキストペアを学習データとし、BERTなどの事前学習モデルでfine-tuneすることで、文ベクトルの生成を目的に考案されたモデルである。

天野ら[9]は意味内容の誤り検出にS-BERT利用の有効性を示している。本研究では、学習者訳と正解文二つの文ベクトルを取り、二つのベクトルのcos類似度を求めることで、意味内容の一致度を測る。S-BERTのモデルはMiniLM[10]でfine-tuneされたparaphrase-MiniLM-L3-v2を用いた。

#### 3.3 統合

誤り検出部は自然言語推論モデルとS-BERTを統合させて実装した。統合した誤り検出の流れを図2に示す。付録A.1に誤り検出におけるcos類似度閾値設定の予備実験を記載する。

### 4 誤り訂正

#### 4.1 形態素解析器

図1に示すように、誤り訂正では、まず、日本語の学習者訳と英語の出題文と学習者訳を形態素

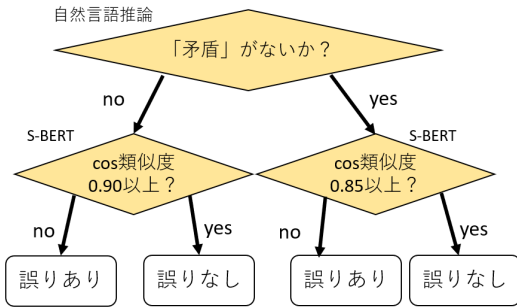


図2 誤り検出の流れ

解析する。本研究では、日本語の形態素解析器に MeCab[11] を用い、英語の形態素解析器に nltk の tokenizer[12] を用いた。形態素解析器を用いる狙いとしては、意味内容に誤りが起こらない「助詞」や「助動詞」などを訂正の候補から除外するためである。

## 4.2 単語アライメント

### 4.2.1 awesome-align

誤り訂正に利用する単語アライメントモデルとして、awesome-align[8] を利用する。これは、2021 年時点での日英間アライメントで最も低いエラー率を記録したアライメントモデルである。awesome-align の特徴として、アライメントが付与されていない対訳コーパスで学習可能であることがあげられる。多言語事前学習モデルである MBERT[6] を用いることで、別言語間で同一の単語の埋め込み表現を求め、埋め込み表現  $h_x$  と  $h_y$  ( $x$  と  $y$  はそれぞれの言語に対する文を示す) とすると式 2 のような類似行列を得る。

$$S = h_x h_y^T \quad (2)$$

双方向間の類似行列を取得してから閾値  $c$  でアライメントの可否を式 3 のように判定する。

$$A = (S_{x,y} > c) * (S_{y,x}^T > c) \quad (3)$$

アライメント関係にあるとされる単語を Mask し単語を予測するなど計 5 つの目的関数を定義し fine-tune することで、より高品質な単語埋め込み表現を得る。得られた埋め込み表現から式 2 と 3 から再度アライメントを求める。

[8] では京都フリータスク (KFTT)[13] を用いて日英間単語の fine-tune を実施した。KFTT は Wikipedia から京都に関係した文を取得したコーパスであり、専門単語の出現がわずかである和文英訳問題の出

題文に対しては不向きである。そこで、本研究では KFTT に加え、比較的簡潔な日本語で構成された田中コーパス [14] と様々な分野を網羅している JParaCrawl[15] の一部を加えた計 80 万文の対訳データで fine-tune した。

### 4.2.2 訂正への応用

出題文・学習者訳間の単語アライメントと出題文・正解文間の単語アライメントの二つを取得する。Mecab が「名詞」「動詞」「副詞」「形容詞」「接続詞」「感動詞」と判定した出題文中の日本語単語に対して、正解文のみにアライメントされている単語を訂正候補とした。また、上記品詞に属していても「い」や「て」のような意味を持たない単語を除外するルールを定めることで精度向上を測った。さらに、英語では複数単語で一つの日本語単語となる場合も考えられるので、フレーズ単位で対応させた。これらは前述した「挿入」の機能にあたる。

## 4.3 辞書を利用した訂正

awesome-align は、文構造を考慮してしまうので、正解文と類似した箇所にある学習者訳の単語がアライメントされてしまう。例えば、“私は野球をしました” という出題文に対して学習者が “I played soccer” と訳した場合、“soccer” と “baseball” の両単語がアライメントされてしまう。

本研究では、前述した問題に対処するために英和辞典を用いる。正解文と学習者訳両者アライメントされた場合、学習者訳中の単語が出題文単語に近い意味が存在するのかを英和辞典で調べる。完全一致もしくは、出題文中の単語に漢字が含まれる場合は先頭漢字が一致しているものを正解とし、それ以外を不正解とした。使用した英和辞典は英辞郎 [16] および Web 上の Weblio 辞書 [17] である。これは前述した「置換」の機能にあたる。なお、誤りありと判定されたが添削単語が存在しない場合は、誤りなしとして表示することにした。訂正のより細かな流れの一例を図 3 に示す。

## 5 実験

### 5.1 実験条件

実験に利用するデータとして、大学受験レベルの和文英訳問題 [18],[19] の中からランダムに 51 問を抽出し、英語中級レベルの 5 名の解答者に英訳文作

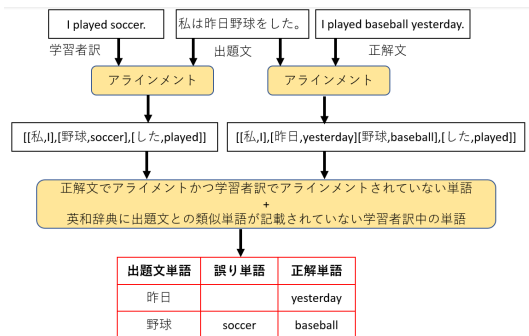


図3 誤り訂正の流れ

成を依頼した。

得られた回答延べ254文を著者らが「意味内容の誤り」に基づいて正解文と不正解文のラベル付けを実施し、このデータを用いて評価実験を行う。解答文データの内訳を表1に示す。

表1 解答文データの内訳

	正解文数	不正解文数
学習者訳	142	112

## 5.2 実験結果

本研究における実験として、どの程度誤りが検出可能かと、表1の誤答文に対して、提案システムが誤りと判定したものに対する訂正がどの程度適切であったのかを評価した。

まず、正解文・不正解文それぞれにおける誤り検出の適合率、再現率、F値を表2に示す。なお、本論文の結果は小数点第三位で四捨五入している。

表2 誤り検出結果

	適合率	再現率	F値
正解文	0.80	0.72	0.76
不正解文	0.70	0.76	0.73

また、112文の不正解文に対して、提案システムがどの程度単語(熟語)を添削できたのかを調べた。提案システムが不正解であると判定した個数は102文である。102文の添削結果の妥当性を著者らにより人手評価した。適切な添削を行った単語(熟語)数、不適切な添削を行った単語(熟語)数を表3に示す。

最後に、不正解文中の不適切な添削であったもののうち、適切な訳であるのに不適切な添削を実施してしまった単語(熟語)数と不適切な添削であるのに添削出来なかった単語(熟語)数を、それぞれ過剰添削、添削不足とし、その結果を表4に示す。

付録A.2に本実験における添削例を記載する。

表3 誤り訂正結果

適切な添削数	不適切な添削数
178	141

表4 不適切な添削

過剰添削	添削不足
120	21

## 5.3 考察

誤り検出では、正解文・不正解文において表2よりどの指標でも70%以上の精度を得ることができた。T5[20]などのBERTやminiLMより高度な言語モデルを基に自然言語推論モデルやS-BERTを構築することで、さらなる精度向上が期待できる。

誤り訂正では、半分程度の確率で適切な添削が可能であることを示した。その中でも表4より添削がなされない単語(熟語)より添削すべきでない単語(熟語)を添削する方がはるかに多い。これは、英和辞典に存在しない出題文とアライメント関係にあった学習者訳中の単語(熟語)が不正解判定されるためであると考えられる。それと同時に、101文中で添削されなかった単語が21個と少なかったため、作成した英和辞典が出題文中の単語をある程度網羅していることを示した。また、Dou and Neubig[8]によるもののような近年の単語アライメントモデルが、出題文と学習者訳・正解文それぞれの単語の対応関係が高精度で取れ、訂正タスクにも十分応用可能であることも示した。英和辞典やルールベースの拡張、類義語辞典を構築するなど過剰添削の削減に期待できる。

## 6 おわりに

本研究では、「意味内容の誤り」を自動添削することに特化した和文英訳添削システムを構築した。提案システムとして、BERTをベースとした誤り検出と、単語アライメントモデルと英和辞典を利用した誤り訂正の二つを提案した。実験の結果、誤り検出では全ての指標において0.7以上を記録することができたが、誤り訂正では半分弱の添削が不適切であり、さらなる改善が必要である。

今後の課題として、単語アライメントモデルを訂正タスク向けに拡張したモデルの検討や過剰添削を防ぐアルゴリズムの検討を行うとともに、ユーザへの評価実験を行っていききたい。



## 謝辞

本研究は JSPS 科研費 JP20K00830 の助成を受けたものです。

## 参考文献

- [1] 杉浦正好. ライティング指導としての和文英訳. 愛知学院大学文学部紀要, No. 39, pp. 127–135, 2010.
- [2] Kostiantyn Omelianchuk, Vitaliy Atrasevych, Artem Chernodub, and Oleksandr Skurzhanyski. Gector-grammatical error correction: tag, not rewrite. **arXiv preprint arXiv:2005.12592**, 2020.
- [3] Yu Wang, Yuelin Wang, Jie Liu, and Zhuo Liu. A comprehensive survey of grammar error correction. **arXiv preprint arXiv:2005.06600**, 2020.
- [4] 菊地正弥, 尾中大介 and 舟山弘晃, 松林優一郎, 乾健太郎. 項目採点技術に基づいた和文英訳答案の自動採点. 言語処理学会, 2021.
- [5] 松岡保静, 中村俊允 and 村上聡一郎, 澤山熱気. 英作文採点・添削技術の開発. NTT DOCOMO テクニカル・ジャーナル, Vol. 27, No. 39, 2020.
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. **arXiv preprint arXiv:1810.04805**, 2018.
- [7] Nils Reimers and Iryna Gurevych. Sentence-bert: Sentence embeddings using siamese bert-networks. **arXiv preprint arXiv:1908.10084**, 2019.
- [8] Zi-Yi Dou and Graham Neubig. Word alignment by fine-tuning embeddings on parallel corpora. **arXiv preprint arXiv:2101.08231**, 2021.
- [9] 天野勇弥, 松井颯汰, 綱川隆司, 西田昌史, 西村雅史ほか. 和文英訳自動添削システムにおける内容誤り検出手法の検討. 第 84 回全国大会講演論文集, Vol. 2022, No. 1, pp. 807–808, 2022.
- [10] Wenhui Wang, Furu Wei, Li Dong, Hangbo Bao, Nan Yang, and Ming Zhou. Minilm: Deep self-attention distillation for task-agnostic compression of pre-trained transformers. **Advances in Neural Information Processing Systems**, Vol. 33, pp. 5776–5788, 2020.
- [11] (2022 年 12 月 22 日閲覧) . <https://github.com/neologd/mecab-ipadic-neologd>.
- [12] (2022 年 12 月 22 日閲覧) . <https://www.nltk.org/>.
- [13] (2022 年 12 月 22 日閲覧) . <http://www.phontron.com/kftt/index-ja.html>.
- [14] (2022 年 12 月 22 日閲覧) . [http://www.edrdg.org/wiki/index.php/Tanaka\\_Corpus](http://www.edrdg.org/wiki/index.php/Tanaka_Corpus).
- [15] Makoto Morishita, Jun Suzuki, and Masaaki Nagata. Jparacrawl: A large scale web-based english-japanese parallel corpus. **arXiv preprint arXiv:1911.10668**, 2019.
- [16] (2022 年 12 月 22 日閲覧) . <https://http://www.eijiro.jp/>.
- [17] (2022 年 12 月 22 日閲覧) . <https://ejje.weblio.jp/>.
- [18] 高橋響. 英語ライティングの鬼 100 則. 明日香出版社, 2021.
- [19] 竹岡広信. ドラゴン・イングリッシュ基本英文 100. 講談社, 2005.
- [20] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J Liu, et al. Exploring the limits of transfer learning with a unified text-to-text transformer. **J. Mach. Learn. Res.**, Vol. 21, No. 140, pp. 1–67, 2020.

## A 付録

### A.1 誤り検出における cos 類似度評価

予備実験として、S-BERT の意味内容の誤り検出における適切な cos 類似度の数値を調査した。実験には、本論文で用いたデータとは別のクラウドソーシングで収集した。その後著者らで正解文と不正解文に分け、それぞれ 50 文ずつ計 100 文をランダムに抽出した。正解文・不正解文と事前に用意した問題集の解答とで S-BERT が示した cos 類似度の平均値を表 5 に示す。

表 5 cos 類似度の平均値

	正解文	不正解文
cos 類似度	0.83	0.63

表 5 より、正解文に対する cos 類似度の平均値が 0.83 であることから、0.8 周辺で閾値設定すればよいと仮説を立てた。

次に、それぞれ 50 文の正解文・不正解文と事前に用意した問題集の解答とで S-BERT が示す cos 類似度の値をそれぞれ箱ひげ図で表した。その箱ひげ図を図 4 に示す。

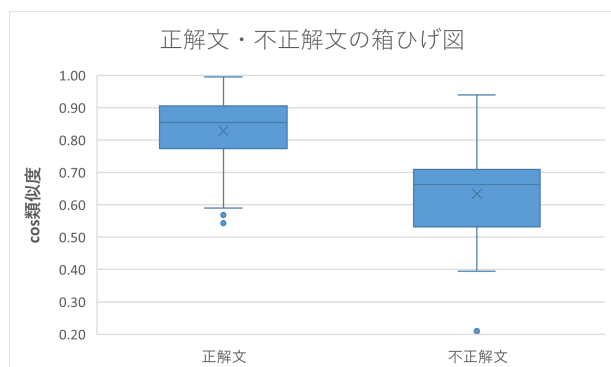


図 4 正解文・不正解文に対する S-BERT が示す cos 類似度の箱ひげ図

箱ひげ図 4 より、正解文に対する cos 類似度が 0.8 弱から 0.9 強に収まっている。また、不正解文に対する cos 類似度が 0.9 以上のものがほぼない。さらに、正解文に対する cos 類似度はいくつかの外れ値を含んでいる。それらの結果と訂正候補が無い場合は誤りなしとすることをふまえて、正解文の cos 類似度が 0.85 以上となればその文は正解文であろうと仮定し、閾値設定した。

### A.2 添削例

ここでは、提案システムにおける添削の一例を示す。

#### A.2.1 添削が適切な例

出題文が“もし宝くじにあたったら、彼らは一生懸命働かなくなるでしょう。”の時に学習者が“If they get big amount of money, they won’t work forever.”と訳した場合の例を示す。この時の誤り単語(熟語)は“宝くじ”，“あたる”，“一生懸命”である。その時の提案システムの出力を表 6 に示す。

表 6 適切な添削例

誤り単語(熟語)	システム出力
宝くじ	big~money ⇒ lottery
あたら	get ⇒ won
一生懸命	forever ⇒ hard

表 6 中の“get ⇒ won”は“get”から“won”への変更を促している。

#### A.2.2 添削が不適切な例

出題文が“医療の向上のおかげで、多くの国で平均寿命が高くなりました。”の時に学習者が“Thanks to the medical improvement, people live longer in many countries.”と訳した場合の例を示す。この時の誤り単語(熟語)は“平均寿命”である。その時の提案システムの出力を表 7 に示す。

表 7 不適切な添削例

誤り単語(熟語)	システム出力
向上	longer ⇒ higher

“longer”の原型である“long”において、英和辞典では“向上”という意味の記載がないため不正解判定されている。また、英和辞典において“live”に“平生使う”という記載があるため、“平均寿命”とマッチしてしまった。