

和文英訳問題自動添削システムにおける参照文拡張による性能改善手法の検討

城處裕也 綱川隆司 西田昌史

静岡大学

1. はじめに

日本語の出題文を英語に翻訳する和文英訳問題は高校、大学等の試験でよく出題されるなど、英語学習において重要な課題の一つである。しかし和文英訳問題は記述式問題であり、学習者自身での正誤判定が難しい。よって、和文英訳問題を自動で採点・添削するシステムは有用性がある。

松井ら(2023)の先行研究では、任意の和文英訳問題に対して添削できる自動添削システムの開発を目標とした。既存システムにおいては、誤り検出部と誤り訂正部の2つの処理をもとに自動添削を行う。検出部の処理を先にを行い、その後訂正の処理を行う。ここで、参照文は与えられている正解文(英語)と出題文(日本語)を機械翻訳した翻訳文(英語)の最大2つとしている。このシステムでは、50%程度の割合で適切な添削が可能であった。一方で、このシステムで出力される添削のうちおよそ半数が、適切な訳であるのに添削されてしまうことや、添削すべき部分が添削されないといった課題が残っている。

本研究では、既存システムの性能改善を目標とする。特に参照文が最大2つであった点に焦点をおく。和文英訳問題には多様な正答が存在する。よって、参照文を増やすことは正答とすべき英訳パターンを増やすことができ、不適切な添削の減少が期待できる。

2. 関連研究

和文英訳問題を自動採点・添削する研究として、石井ら(2022)による研究がある。この研究の提案システムでは、特定の問題に複数の明確な採点項目・基準を設定し、その項目に応じて自動採点・添削を行う。各設問に採点済み答案に加えて疑似誤り答案を利用し、項目ごとに個別の採点モデルを学習した。

この研究は、限られた事例で行われた研究であるため、一般化されていない。本研究では任意の問題に対して添削可能である自動添削システムの開発を目的とする。

3. 提案手法

先行研究のシステムの性能改善手法として、参照文拡張による性能改善を提案する。また参照文拡張に伴い、誤り訂正部を改良する。

3.1 参照文拡張の方法

参照文を拡張する方法として、ChatGPT¹を用いたテキスト生成方法を採用する。ChatGPTに役割は、「Translate Japanese sentences into English with

A Study on Performance Improvement Methods by Extending Reference Sentences in an Automatic Correction System for Japanese-English Translation Problems
Yuya Kidokoro, Takashi Tsunakawa, Masafumi Nishida

ten patterns.」(日本語文を10種類のパターンで英語に翻訳してください)というプロンプトを与える。その後、出題文を入力し、そこで生成された10個のパターンを参照文としてシステムに入力する。これらに先行研究で用いている正解文、機械翻訳文を加えた最大12文を参照文として用いる。

3.2 誤り訂正部の改良

先行研究においては、出題文(日本語文)においてMecab²による形態素解析結果が、「名詞」「動詞」「副詞」「形容詞」「接続詞」「感動詞」である単語のうち、参照文へのアライメントがある単語を訂正候補としていた。本研究では、これらの品詞に加えて「代名詞」を追加した。

また、先行研究において誤り検出部で誤りありと判定されたが、訂正候補の単語すべてについて最終的に添削結果が出力されない場合は誤りなしとする機能を拡張した。既存システムでは、2つの参照文と学習者訳とのcos類似度を求め、類似度が高いものを参照し、誤り訂正を行っていた。本研究では、最大12文の参照文全てと学習者訳とを比較することで誤り訂正を行い、先行研究と同様に訂正候補の単語の中から添削単語がなくなった場合は誤りなしと判定する。

4. 評価実験

4.1 実験データ

実験データとして、先行研究にて用いられたデータと同様のものを利用する。和文英訳問題51問に対して、英語中級レベルの5名の回答者から得られた回答文延べ254文を、正解文と不正解文のラベル付けを行い、このデータを用いて評価実験を行う。また、和文英訳問題の出題文をもとにChatGPTにて作成した10個の英訳パターンを用いる。回答文データのラベル付けの結果を表1に示す。

表1 回答文データのラベル付け結果

	正解文	不正解文
データ数	128	126

4.2 実験結果

本研究では、提案手法において参照文数を変更させ、性能評価を行った。はじめに、回答データに対してどの程度正誤判断が正しくされているかの分析を行った。次に、不正解文、126文に対して、どの程度単語(熟語)が添削できたかの調査を行った。

まず、提案手法について正誤判断が正しくされている

¹ <https://chat.openai.com/>

² <https://github.com/neologd/>

かについて、参照文数を変更した結果を図 1 に示す。また誤った判断について内訳を図 2 に示す。

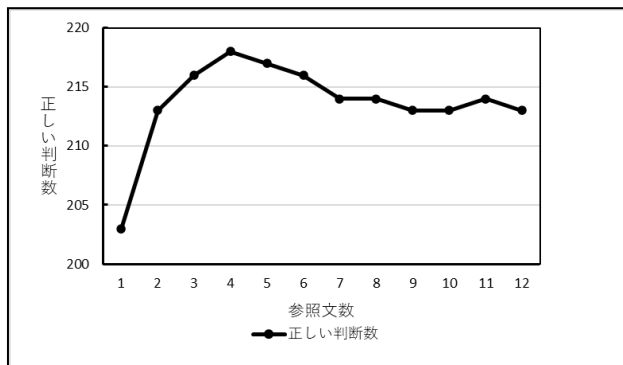


図 1 正誤判断の結果

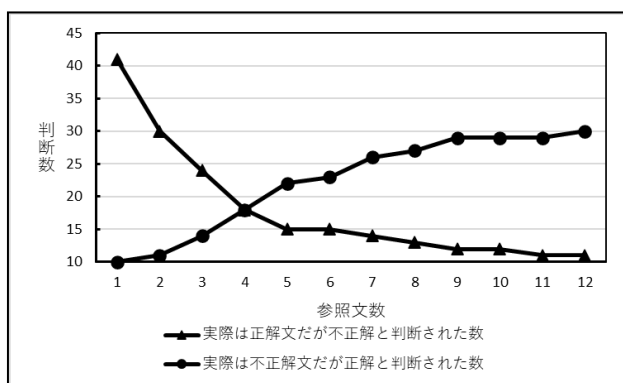


図 2 誤った判断の内訳

次に、不正解文に対して、どの程度単語(熟語)が添削できたかについて調査した。不正解文 126 文に対して、添削結果の妥当性を著者によって人手評価した。適切・不適切な添削を行った単語(熟語)数を提案手法について、参照文数を変更した結果を図 3 に示す。また、不適切な添削の中から、適切な訳であるのに添削されたもの、不適切な訳であるのに添削されなかったものをそれぞれ過剰添削、過剰不足とし、不適切な添削における内訳を図 4 に示す。

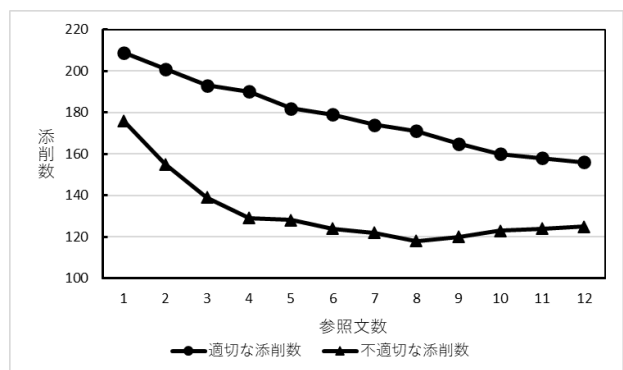


図 3 誤り訂正結果

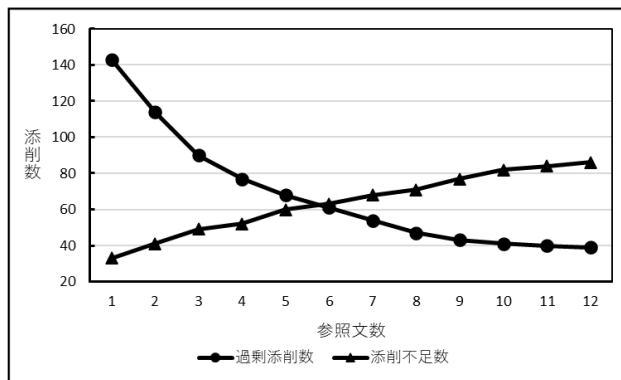


図 4 不適切な添削の内訳

4.3 考察

図 1 より、参照文が 4 つである場合に正しい判断数が一番多くなる結果となった。また、図 2 より誤った判断のうち、不正解と誤った場合と正解と誤った場合は参照文数が増えるに従いそれぞれ減少、増加しトレードオフの関係にあることが示された。図 3 より、不適切な添削が一番少なくなるのは参照文が 8 つである場合であった。また、図 4 より不適切な添削のうち、過剰添削と添削不足もトレードオフの関係がみられた。

参照文が 2 文であったときと比べ、参照文を増やすことで正誤判断・誤り訂正ともに精度が上がる結果が得られた。特に誤り訂正における過剰添削数を大きく減らすことができた。しかし、参照文を増やしすぎると、不正解文を正解と判断してしまうことや、添削不足が増えてしまう結果となった。これは参照文を増やすことで、正答とすべき英訳パターンを増やすことができたと同時に、増やした参照文の中には意識的なものも存在し、それが訂正すべき単語を正答、もしくは訳に必要な単語として判断されることにつながっていると考えられる。

5. おわりに

本研究では、先行研究において参照文を拡張することで和文英訳自動添削システムの性能改善を目指した。提案システムとして、ChatGPT を利用した参照文拡張、添削時に、訂正候補の単語がなくなるまで参照文全てと学習者訳とを比較することの 2 つを提案した。実験の結果、先行研究同様参照文が 2 文であるときと比べ、参照文拡張により正誤判断・誤り訂正ともに精度が上がった。

今後の課題として、参照文を増やすことで生じた添削不足を防ぐために、単語アライメントモデルの検討・および生成した参照文を選択する方法について検討していきたい。

参考文献

- 松井颯汰, 綱川隆司, 西田昌史, 西村雅史. 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. 言語処理学会第 29 回年次大会. 2023
- 石井雄隆, 菊地正弥, 舟山弘晃, 松林優一郎, 乾健太郎. 説明可能な AI を指向した和文英訳自動採点システムの開発と評価. 日本教育工学会研究報告書, Vol. 2022, No. 4, pp1-7, 2022