

名詞動詞句補充問題における錯乱肢の自動生成

小北 凌大† 中島 陽子‡ 本間 宏利‡ 山本 和英*

† 釧路工業高等専門学校 情報工学科

‡ 釧路工業高等専門学校 創造工学科

* 長岡技術科学大学 電気電子情報工学専攻

{honma, yoko}@kushiro-ct.ac.jp

1 はじめに

近年、国内外の多くの日本語教育機関で読解の学習支援が実施されている。コンピュータを日本語の読解教育に用いる試みは、1980年代から始められており、これまで、CAL(Computer-Assisted Learning) または CALL(Computer-Assisted Language Learning) の研究が進められてきた [1, 2, 3].

本研究は、テキストから日本語学習者向けの読解問題を自動生成するシステムの開発を目的とする。読解問題とは学習者の文章意味の理解度を測るための文章、設問と解選択肢のセットである。学習者は最初に与えられた文章を読み、その内容や意味に関する各設問に対して、提示された解選択肢から正解を示すことで文章理解度が評価される。

本研究では、高橋 [4] が解明した学童における読解過程のうち、語の意味の理解、それらの係り受けの理解に着目し、名詞・動詞句補充問題、会話文並び替え問題、高頻出語補充問題の3種類の読解問題を対象とする。従来研究 [5] では、名詞・動詞句補充問題で錯乱肢を生成する際に、問題文中に含まれない名詞、動詞句が多用されるといった問題点があった。会話文並び替え問題では会話の抜き出す発言によって、問題の難易度が上がりすぎるといった問題点があった。また、高頻出語補充問題では空欄数を多く設定することで問題文の内容が理解できない場合があることが問題点として残っていた。

本研究では、これらの問題点を改善し、自動生成する問題文と錯乱肢の質の向上をはかる。本システムを活用することにより、教育者の読解テスト開発支援、作成コストの削減、そして、日本語学習者の自律的な日本語学習支援や日本語学習者の学習モチベーションの維持が期待できる。

2 読解問題について

本研究では、日本語学習者向けの読解問題として名詞・動詞句補充問題、会話文並び替え問題、高頻出語補充問題を対象とする。以下に、これらの読解問題について解説する。

2.1 名詞・動詞句補充問題

名詞・動詞句補充問題とは、名詞や動詞句の部分が空欄化された文に対し、学習者が与えられた選択肢の中から最も意味的に適切と思われる解を答える形式の問題である。本研究では、学習者に提示する文は、任意に抽出した“名詞”+“助詞”+“動詞”の構文パターンを含む文の中の名詞または動詞句の1つを空欄に置き換えたものとする。また、選択肢は正答の選択肢1つと自動生成した錯乱肢3つからなる4択形式とする。名詞・動詞句補充問題のイメージを図1に示す。

問題文			
竹馬の友、セリヌンティウスは、深夜、王城に召された。暴君ディオニスの面前で、 <u>佳き友</u> と <u>佳き友</u> は二年ぶりに相逢うた。メロスは友に一切の事情を(?)。			
1.語った	2.訴えた	3.抱えた	4.訊いた

図1: 名詞・動詞句補充問題の例

2.2 会話文並び替え問題

会話文並び替え問題とは、テキストから抽出した会話領域の発言の一部を空欄化し、会話の流れに正しく沿うように、与えられた選択肢中の会話文を正しく並べる問題である。本研究では「」で囲われた文章を発言文とし、発言文間に非発言文が3文以上現れない領

域を1つの会話領域と定義する。また、会話領域内の空欄化させる発言は、3連続以上にならないようにする。会話文並び替え問題のイメージを図2に示す。

問題文
「1」 「市を暴君の手から救うのだ。」 「2」 「市に用事を残して来た。またすぐ市に行かねばならぬ。あす、お前の結婚式を挙げる。早いほうが良からう。」 「3」 「4」 「どっこい放さぬ。持ち物の全部を置いて行け。」
選択肢
a. 「この短刀で何をするつもりであったか。言え！」 b. 「何をやるのだ。私は陽の沈まぬうちに王城に行かねばならぬ。放せ。」 c. 「おまえがか？」 d. 「嬉しいか。綺麗な衣装も買って来た。さあ、これから行って、村の人たちに知らせて来い。結婚式は、あすだ。」

図 2: 会話文並び替え問題の例

2.3 高頻出語補充問題

高頻出語補充問題とは、テキスト中に高頻度で出現している単語(名詞)を空欄に置き換えられた文に対し、学習者が最も適切な語を選択肢から選んで答える問題である。本研究では、各段落の冒頭部分に出現度の高い語を高頻出語とする。また、2つの段落をまたいで出現する語を「重要語」と定義し、本問題の対象とする。会話文並び替え問題のイメージを図3に示す。

問題文
「1」は激怒した。必ず、かの邪智暴虐の「2」を除かなければならぬと決意した。メロスには「3」がわからぬ。「1」は、村の牧人である。「4」を吹き、羊と遊んで暮して来た。けれども「5」に対しては、人一倍に敏感であった。きょう未明メロスは村を出発し、「6」を越え山越え、十里はなれた此のシラクスの「7」にやって来た。
選択肢
a. 邪悪 b. 政治 c. 野 d. メロス e. 市 f. 王 g. 笛 h. セリヌンティウス i. 妹

図 3: 高頻出語補充問題の例

3 錯乱肢について

本研究では、名詞・動詞句補充問題における4つの選択肢のうち、正解以外の選択肢を「錯乱肢」と呼ぶ。従来研究では、錯乱肢は青空文庫¹中の文から、問題文と同数の節と助詞、助動詞を有する動詞句を抽出し、それらからランダムに錯乱肢を選定するという

¹<https://www.aozora.gr.jp/>

手法であった。これに対し、本研究では、新たな錯乱肢獲得手法として、現代日本語書き言葉均衡コーパス(BCCWJ)²から文章パターンの頻度表を作成し、“名詞”+“助詞”の構造に続く可能性の高い“動詞句”からランダムに錯乱肢を獲得する方法を採用する。錯乱肢獲得のイメージを図4に示す。



図 4: 錯乱肢獲得手法のイメージ

3.1 同音錯乱肢の削除

同音錯乱肢を削除するため pykakasi³を用いて同じ読みを持つ錯乱肢を洗い出し、漢字が多い錯乱肢を検索用として残した。しかし、「行った」のように「いった」と「おこなった」など2つ以上の読みを持つ場合は、対処が難しくなることが難点である。

3.2 同義錯乱肢の削除

文章中に存在する名詞や動詞を用いた錯乱肢の生成が可能になったが、本手法では同義錯乱肢が生成されてしまう問題点がある。そこで、同義錯乱肢生成の対策として、連想語辞書を用いた同義錯乱肢の削除を行う。連想語辞書は本研究室で開発・編集されており、36,1430個の語句に対してそれぞれ連想語を格納している。1つの語句は平均42.5個の連想語を有する辞書となっている。図5に語句“みかん”に対する連想語のイメージを示す。

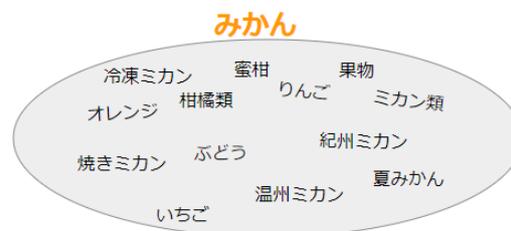


図 5: 連想類語辞書での検索例

²<https://pj.ninjal.ac.jp/>

³漢字, ひらがな, カタカナをローマ字に直すモジュール

同義錯乱肢は、「銀行を襲撃する」と「銀行を襲う」のように動詞部分の意味が同義である場合と、「手を上げる」と「手を出す」のように文章の意味が同義⁴である場合がある。いずれの場合でも連想語辞書による同義判定が可能であるため、生成した錯乱肢から類義・同義である錯乱肢の削除、新しい錯乱肢の追加操作を繰り返すことで同義錯乱肢の削除をおこなう。図??の例では、初期錯乱肢の中で「手を出す」と「手を入れる」が同義関係にあるため「手を入れる」を削除し、その後 BCCWJ より新たに錯乱肢となりうる動詞を追加し、再び同義判定を繰り返し行っている。

3.3 錯乱肢の再生成

錯乱肢作成、同音判定による同音錯乱肢の削除、類義・同義判定による同義錯乱肢の削除という工程を経て、錯乱肢3つと正解の選択肢が同じ読みでなく、類義・同義関係を持たなくなるまで類義・同義判定、同音判定、錯乱肢の再生成を繰り返す(図6参照)。以上の処理をもって作成した錯乱肢を、動詞句補充問題の選択肢として定義する。

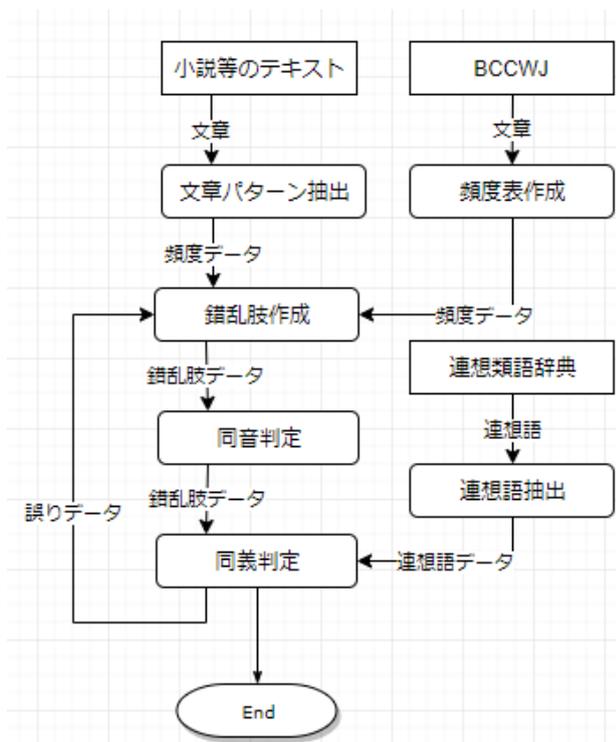


図 6: 錯乱肢生成の流れ

⁴両方「暴力をふるう」という意味

3.4 名詞句の錯乱肢生成における問題点

名詞句補充問題の錯乱肢は動詞句補充問題の錯乱肢生成の方法と同じ方法で生成を行った。しかし、名詞の錯乱肢をコーパスで補おうとした場合、文章に含まれていない名詞が多く使用されてしまう。そこで文章中にある名詞を優先的に錯乱肢に選定することで精度の改善が見込まれると考えたが、しかし文章中にある名詞だけでは“助詞”+“動詞”の構文パターンに一致するものが少なく、適した選択肢を作成するのは困難である。名詞錯乱肢を文章中の語句で作成しようとした場合、文が不自然になるという問題点がある。

4 錯乱肢生成の精度比較

錯乱肢生成の精度検証のために、図6の手順で同音、および同義錯乱肢の削除処理を実行した場合と、何も処理をしなかった場合で検証実験を行った。

実験は、最初に任意の100個の設問に対応する錯乱肢を図4の方法で100セット(4語/1セット)自動生成する。1セット中の4つの錯乱肢の中で、同音、同義のものが1組でも存在すれば、そのセットは同音錯乱肢、同義錯乱肢を含む不適切なものとして判定される。錯乱肢削除処理を実行しなかった場合と図6の手順で錯乱肢削除および追加生成処理を行って取得した100セットの錯乱肢を対象に精度検証実験を行った。

削除処理を行わなかった場合の実験結果を表1に、同音および同義錯乱肢の削除処理を行なった場合の実験結果を表2に示す。

表 1: 削除処理を行わなかった場合

	同義錯乱肢	同音錯乱肢
不適合割合	68/100	31/100

表 2: 同音・同義錯乱肢の削除処理後

	同義錯乱肢	同音錯乱肢
不適合割合	18/100	2/100

5 まとめ・今後の課題

本研究では主に名詞・動詞句補充問題における錯乱肢の自動生成について研究を行い、提案手法によって

同音および同義錯乱肢を効果的に削除できることを示した。

紙面の都合上、会話文並び替え問題と高頻出語補充問題の改良については割愛したが、簡単にこれらの取り組みについて記述する。

会話文並び替え問題での会話の抜き出す発言によって、問題の難易度が上がりすぎるという問題の対策として、空欄の設定条件の他に、空欄にする文の長さの条件を設定することで、より会話文並び替え問題の選択肢に適した発言を抜き出すことが可能となった.. 実験では空欄にする発言の長さは15~20文字の発言に制限することで、「うん」や「そう」などの返答や、情報量の多い長い文章を選択肢から省くことができた。さらなる改良として、発言者の判別処理の実現がある。会話文並び替え問題は複数人の発言の集合で構築された会話領域を使い、問題を作成するという主旨であるが、発言者が判別できないと、独り言のように同じ発言者が連続して発言している会話も問題に使用されてしまう。この問題を解決するためには、明確に発言者が書かれてない場合、発言前後の文脈から発言者を予測するといった処理が必要である。

高頻出語補充問題では空欄数を多く設定することで問題文の内容が理解できない場合があることが問題点があった。この対策として、一文中に空ける括弧の個数を制限することで、問題の質の向上を実現した.. 実際に問題を作成する際には、一文の主語、述語を同時に含まないこと、一文中に3つ以上括弧を空けないこと、という2つの条件を追加し作問を行った。さらなる改良として、各段落をまたいで出現する重要語の特定をし、優先的に重要語の箇所を空欄に設定する処理の実現化がある。

参考文献

- [1] 岡本敏雄, “「CAI から CAL へ」”, CAI 学会誌, Vol.3, No.2-3, p.21-23, 1983.
- [2] 畑佐一味, “日本語 CALL の現状と今後”, くろしお出版, p.113-130, 2003.
- [3] J. Higgins, “Computer Assisted Language Learning”, Language Teaching, Vol.16, No.2, p.102-114, 1982.
- [4] 高橋登, “学童期における読解能力の発達過程 1-5 年生の縦断的な分析”, 教育心理学研究 Vol.49, No.1, p.1-10, 2001.
- [5] 後藤大明, 山本和英, “日本語テキスト内容理解問題の自動生成システム”, 長岡技科大研究報告, 2017.