

日本語テキスト内容理解問題の自動生成システム

後藤大明(長岡技術科学大学大学院生)・山本和英(長岡技術科学大学)

1. 研究目的

言うまでもなく、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(岡本, 1983; 畑佐, 2003; Higgins, 1982)。先行研究では、コンピュータを用いた教材提供やインターネットを通じた教材選択が主な目的であり、読解の評価にコンピュータを活用することに焦点を当てた研究は、管見の限り見られない。テキスト読解を評価する際は、テキストの内容理解問題等によって構成された読解テストを実施するケースが多い。内容理解問題の作成にあたっては、授業で扱った問題をそのまま使用する、類似した新たなテキストから内容理解問題を作成する等、教師は多くの時間と労力を費やしている。

そこで、本研究では、任意のテキストを入力することで、内容理解問題を自動的に生成するシステムを開発した。このシステムを活用することにより、教師による読解テスト開発支援だけでなく、学習者自身による自律的な読みの学習支援が可能となる。

2. 内容理解問題

本研究では、高橋(2001)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」の自動生成システムを開発した。

2.1. 単語の選定

本研究では、形態素解析で獲得できないテキスト中の名詞を獲得するために横野(2008)で開発された手法に基づきn-gram($2 \sim 8$)で分割した文字列に形態素解析を行い、単語の選定を行った。そして、小文字や記号で始まる文字列、連体詞、助詞、接続詞を含む文字列、テキスト中「は」または「に」を計3回以上後接しない文字列を削除することにより『走れメロス』から「メロス」、「セリヌンティウス」や『ごんぎつね』から「ごん」、「兵十」などの名詞を獲得した。

2.2. 問題文の選定

問題文の選定は、入力されたテキストの文のうち、疑問系、あるいは穴埋め形式に変換することができる文を抽出するという方法を用いた。そこで、横野(2008)が開発した物語要約システムに基づき、テキスト中の全ての文に対して主題(名詞)を割り振った。そして、連続する同じ主題の文章を1つのブロックとして捉え、ブロック中で初めに登場した文を問題に変換することができる文として抽出した。

2.3. 各問題の生成

「名詞・動詞句補充問題」とは前節で抽出された文から名詞、もしくは動詞句を抜き出した、または「誰が」、「どうした」に置き換えた文を表示し、当てはまるものを2

つ、または4つの選択肢から選ぶ問題である。本研究では、テキスト中でより高頻出な名詞と、述語と述語にかかる節の内、「の」、「を」、「に」、「へ」、「と」を後接する節を動詞句として抜き出した。錯乱肢は文章中の名詞、青空文庫中の同じ数の節と助詞、助動詞を持つ動詞句の中からランダムに選択される。

図1 『走れメロス』より生成された「名詞・動詞句補充問題」

問題文：誰が落ち着いて呟き、ほっと溜息をつきましたか。			
(1)身代わり	(2)メロス	(3)妹	(4)暴君

「会話文並び替え問題」とは、テキストから抽出した会話の一部を抜き出し、会話の流れに当てはまるように並べ替える問題である。本研究では、鉤括弧で囲われた文を発言とし、間に5文以上文を挟まない発言を1つの会話として定義した。そして、ランダムに選ばれた4つの各会話から先頭2文(計8つの発言)を抽出し、その中から4文を抜き出すことで、問題文を生成した。「高頻出語補充問題」とは、テキスト中の段落の冒頭のうち、高頻度で出現した語(名詞、形容詞、副詞)を空欄に置き換え、当てはまる語を選択する問題である。以下に生成された問題の例を示す。「会話文並び替え問題」と「高頻出語補充問題」の具体例については、紙面の都合上、割愛する。

3.まとめ

本研究で開発したシステムにより、教師による読解アセスメントを支援することができる。また、学習者が読みたいと思ったテキストに内容理解問題を付与した読解教材として扱うこともできるため、学習意欲の向上や自律的な読みの学習につながる。

今回は、限られた問題形式にとどまったが、将来的には、全ての問題を自動生成できるシステムをめざす。このシステムが実現すれば、テキストを用意するだけで、効率的なテスト開発や学習ニーズにあった学習教材の開発が可能となる。

付記

本研究の一部は(本マテリアルの開発は)、2013年～2016年度JSPS科研費25284096基盤研究(B)『アーティキュレーションを保証する言語能力アセスメント実施支援システムの構築』(研究代表者：渡部倫子)の助成を受けた。

引用文献

- [1]岡本敏雄, “「CAIからCALへ」”, CAI学会誌, Vol. 3, No2-3, pp21-23, 1983
- [2]畑佐一味, “日本語CALLの現状と今後”, 畑佐由紀子(編)第二言語習得への招待, くろしお出版, pp. 113-130, 2003.
- [3]Higgins John, “Computer Assisted Language Learning”, Language Teaching, Vol. 16, No. 2, pp. 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-”,
- [5]横野 光, “整合性を考慮した物語要約システムの構築”, 自然言語処理, Vol. 15, No. 5, pp. 45-71, 2008.