

LLM を用いた質問生成による児童の作文の詳述化支援の検討

横野 光¹ 成家 雅史² 宮城 信³

¹ 明星大学 ² 相模女子大学 ³ 富山大学

hikaru.yokono@meisei-u.ac.jp m-nariya@star.sagami-wu.ac.jp

miyagi@edu.u-toyama.ac.jp

概要

小学校の作文教育における内容に関する指導では出来事の具体的な描写を促すといったことが行われる。しかし、児童にとっては具体的なイメージがわからず詳述化が困難となることもある。これに対し、作文に内容に対する質問を児童に行うことで修正の指針を示すことができると考えられるが、多くの児童の作文に対し質問を考えることは教師にとって大きな負荷となる。そこで本研究では、内容の詳述化のための質問生成に大規模言語モデルが利用できるかについて、小学生に対して行った調査に基づき検討を行った。

1 はじめに

小学校など初等教育での作文指導では、“はじめ”、“なか”、“おわり”といった構成や全体の文章構成、助詞の使い方などの文法的な事項などについての指導だけでなく、内容についての指導も行う。内容に関する指導では、出来事の細かな描写やそのときの気持ちを詳しく書かせるといった具体的な描写を促すといったことが行われる。しかし、単純に“具体的に書いてみましょう”と指示しても、児童にとってはその具体的なイメージが付かず修正することが難しい可能性がある。これに対しては、実際にその作文の内容に関する具体的な質問をすることで、児童に作文の修正の指針を示すといったことが試みられている。内容の詳述化に関しては、作文を行う前の材料を集める段階で行うこともあるが、本研究では一度作文を書き終えた、推敲の段階での詳述化に焦点を当てる。

内容についての質問はその文章を読み手がどのように読んだかを間接的に表していると考えられる。もし読み手の質問が書き手が想定していなかったものであれば、書き手が伝えたかったことが読み手には正しく伝わらなかったとみなすことができる。読

み手を意識して書くということは作文において重要な要素であるが、それは児童にとっては難しいことである。児童同士がペアとなり互いに相手の作文を読んで質問や指摘を行うという共同推敲という作業も行われているが、児童同士での作業の場合、質問が局所的になる、時間がかかるなどといった課題もあり、教師の補助は必要となっている。教師が児童に対して質問するということが可能であるが、作文の内容に関する質問を作成するためには実際にその作文を読む必要があり、担当する児童の人数が多くなるほど教師の負担も大きくなる。そのため、現実的には、これらの支援は特に必要な児童に限られている。一方で、より多くの児童にこれらの支援が効率よく普及できれば、効果的な推敲作業が行われる可能性が高い。

そこで、本研究では作文の内容に関する質問を児童に提示することによる内容の詳述化において、大規模言語モデル (LLM) による質問生成が有用であるかについての検討を行う。

2 関連研究

小学校の作文教育支援に関する研究として、後藤田らは学習指導要領に基づいて文や構成の適切さの判定とその結果を用いた修正箇所の指摘による作文支援システムを構築している [1]。また、大村らは文を組み立てるという作業におけるゲーミフィケーションによるシステムを提案している [2]。

研究やモデルの構築に必要なコーパスに関しては、永田らは小学5年生の言語データからなる“こどもコーパス”を構築している [3]。また、宮城らは複数年の継続調査を含む小学生の作文コーパスを構築し [4]、その活用として漢字の使用実態の分析を行っている [5]。

初等教育における生成 AI の活用に関して、文部科学省はガイドラインを発表しており、その中で児童の生成 AI の活用では発達の段階などに留意しつ

つ、リスク対策を講じた上で利活用を検討すべきとしている [6]。また、実際に LLM を活用している事例もあり、小学校においては児童らが考えを深めたりするためのたたき台として用いられている [7]。

3 質問による内容の詳述化

内容の詳述化について、成家らは小学4年生が記述した学校紹介文に対して行った作文評価の研究で、作文の質的な評価として“相手意識”、“テーマの焦点化と一貫性”と並び重要な評価項目であることを論じている [8]。内容を詳述化するために質問することが有用であることについては、文章産出過程においてピアレビュー方式の共同推敲によって読み手を意識した記述が増えたり [9]、小学5年生の学習者同士が対話的交流による共同推敲によって読み手意識を働かせた記述になったり [10] と、読み手をより意識した記述への効果が認められている。

作文の推敲において内容に関する質問をすることによる詳述化の支援は有用であるが、内容の詳述化の質問を生成するためには、対象の作文のどの内容についての質問を行うか、どのようなことを質問するか、どのような形式で尋ねるか、といったことから、複数の質問を行う場合、どの順番で行うか、といったことまで考慮する必要がある。しかし、これらに関しては我々が知る限りまだ整理されていない。

本研究はこの問題に対して、質問生成モデルの構築を通して課題の整理や体系化を行う事を目指している。その準備として現在様々な状況で用いられている LLM がこの課題に対してどの程度有用であるかを検証する。

4 調査

都内の国立大学附属小学校の6年生のクラスの児童を対象にした調査を行った。過去に書いた作文について、その内容に関する質問を見ながら推敲を行ってもらい、その結果と作業に対する感想の分析から作文の推敲において内容に関する質問を提示することの効果を検証した。

4.1 設定

対象のクラスの児童は7月に自己を発見する学習の一環として自身についての作文を作成している。その作文に対して12月に推敲を行ってもらった。推敲は授業時間内に実施している。授業での流れを以下に示す。

1. **説明 (10分)** 以前に書いてもらった作文について、内容が読み手に伝わるかどうかという視点で修正するように説明する。
2. **修正 (15分)(作業A)** 自身の作文に対して、特に資料を提示せず修正をしてもらう。
3. **質問を見ながらの修正 (15分)(作業B)** 元の作文に対して、その作文に関する質問を提示し、その質問に答えながら修正をしてもらう。質問に対する回答を別途記録するような指示は与えない。
4. **事後アンケート (5分)** 修正をしてみた感想等を問うアンケートに回答してもらう。

作文は7月の時点では手書きであったが、それを Word ファイルにしたものを今回の修正作業で児童に提示している。児童は Microsoft Word で修正を行っている¹⁾。なお、作業に対する謝礼として児童にはタッチペンを進呈した。

作業Aと作業Bのそれぞれにおいて、作業開始時に推敲前の7月に執筆した作文を配布し、終了時に回収している。検証の内容で児童の学習状況に差が出ることを避けるため、クラスを半分に分けて作業2と作業3の順を変えるという事は行わなかった。

事後アンケートについては、質問を参照した修正についての感想、答えやすかった質問、答えにくかった質問を自由記述で回答してもらった。

4.2 質問の生成

児童に提示する質問には人手で作成したものと LLM によって生成したものを利用した。

人手による質問生成については、共著者2人が全作文を読み2つずつ作成した。LLM による質問生成には OpenAI²⁾ の GPT-4o を利用した。生成に用いたプロンプトを図1に示す。

あなたは小学校の国語の先生です。この後に入力する生徒の書いた「自分について」というテーマの作文に対して、より良い内容にするために生徒に対して行う作文の内容についての質問を3つ考えてください。

<本文>

図1 質問生成のプロンプト

1) 児童は他の授業でも Word 等を使用しており、文字入力等の問題はない

2) <https://openai.com/>

児童の作業時間は限られているため、作成した質問を全て提示するのではなく、2人の共著者による質問と GPT-4o で生成した質問からそれぞれ1つずつ合計3つを選んだ。似たような質問が提示されることを避けるために、それぞれの質問から互いの類似度の平均が最も小さくなる組み合わせを選んでいる。質問の類似度は東北大学自然言語処理研究グループが構築した日本語 BERT モデル³⁾を利用して得られた埋め込み表現のコサイン類似度で求めている。

児童の“自身は面倒くさがり屋であるが、他の人にはそう見られてない”といった内容の作文に対する共著者による質問と LLM による質問の例を表 1 に示す。

表 1 質問の例
質問

共著者	質問
共著者 A	なんでめんどくさがりの性格を直せないと思いますか？
共著者 B	なぜ「自分の本当のすがた」を題名にしたのですか？
LLM	「本当の自分を見つけない」という気持ちについて、どんなときに「本当の自分からない」と感じますか？

5 結果の分析

7月、12月の両方の授業に出席していた26名の児童の結果を得た。

5.1 統計的な分析

表 2 に各作文の文字数の平均を示す。作業 A、作

表 2 各作文の文字数の平均

作文	平均(標準偏差)
推敲前	323.12(122.84)
作業 A	353.65(129.26)
作業 B	421.81(135.87)

業 B ともに推敲前よりも平均は増加しており、両作業ともにウィルコクソンの符号付順位和検定で有意水準 5%において有意差が見られた。

また、推敲によってどのくらい語彙が増えたのかについて、作文中の内容語(名詞、動詞、形容詞、副詞)の異なり数に関して、作業 A と推敲前、作業 B と推敲前、作業 B と作業 A のそれぞれについての変化を調査した。表 3 にそれぞれの平均と標準偏差を示す。

表 3 語彙の差分の平均

作文	平均(標準偏差)
作業 A-推敲前	6.62(5.50)
作業 B-推敲前	15.58(16.05)
作業 B-作業 A	13.27(16.26)

補助資料無しで行った作業 A に比べて、質問を提示した作業 B の方が修正後の文字数、語彙数ともに多くなっている。また、作業 B と作業 A の語彙数に関しても作業 B の方が増加しており、作業 B の方がより様々な表現が用いられていることが分かる。しかし、4.1 節で述べたように作業 A と作業 B の順は変えていないため、どちらも推敲前の作文からの修正ではあるが、後に行った作業 B の方がより考える時間が長かったため多くの修正が行われた可能性がある。

5.2 推敲結果の定性的な分析

書き手の書きたいことがよく分かるかどうかという観点からの評価基準において、修正前の作文とそれぞれの作業で修正された作文の3つのうちどれが最も良いかという評価を、作文指導の経験がある現職教員3名に依頼した。その結果を表 4 に示す。

表 4 人手による評価結果

作文	良いと回答した数
修正前	11
作業 A	13
作業 B	54

多くの場合、内省による推敲によって、さらには補助資料などの外部からの支援を受けることによって、作文はよりよく書き換えられる傾向があることが分かった。一方、表 4 の結果から、推敲すると文章が良くなるとは必ずしも言い切れないということが示されている。今回の調査の推敲では多くの例で手を入れるほど詳述化されるという傾向が見られた。しかしながら、一部の結果では詳述化するほどよく分かるということにはなっていない。作業者のコメントの一つに“(著者注：修正前の方が)2つの性格を1段落にまとめてしまっはいるが、簡潔でよい”とあることがこの事実を端的に表している。

5.3 LLM が生成した質問の分析

GPT-4o によって生成された質問の多くは作文中のフレーズを引用し、そのフレーズが性格に関するものであれば具体的なエピソードを、出来事であれ

3) <https://huggingface.co/tohoku-nlp/bert-base-japanese-v3>

ば詳細を求めるものであった。そのような問いは、考えの本質にとって明らかにする必要があるかどうかの検討が慎重になされないのであれば、児童らにはなぜ問われているのか理解できず、同時に回答することが難しいという評価につながる可能性がある。また、質問を生成するように指示したが“～について詳しく教えてください”といった依頼形式の文が生成されることも多く、安易な設問設定に感じられることもあった。

これに対して、人が作成した質問は基本的には詳細を求めるという点では LLM の生成した質問と共通しているが、LLM が生成した質問が抽象的な問い方になっているのに対し、人が作成した質問は表 1 に示したように、記述に対してその理由や目的を問うて書き手の児童の考えの本質を問うものとなっており、個々の作文に応じた内容になっている傾向が見られた。

今回調査に用いた質問の生成では図 1 に示すように具体的な指示を与えてはいない。しかしながら、人が生成した質問と比較すると課題は見受けられるものの、ある程度は妥当な質問が生成されているという点において、有用な質問の要件を整理することで、より望ましい質問が生成される可能性はある。

5.4 アンケート結果の分析

作業後に行ったアンケートでは 25 件の回答を得ることができた。

作業に関する感想からは、質問が提示されたことにより、書きやすくなった、直す箇所が分かりやすかったなどといったポジティブな意見が多く見られたが、質問の回答を考えること自体が難しかったといったネガティブな意見も見られた。質問を難しいと感じる理由については、一次的な思い込みで書いてしまい理由を問われても説明できない、複数の要因があり簡潔に説明することができない、そこで理由を問われることなど考えもしなかった等、様々な要因が推察される。この問題は児童らがどこまで自律的に推敲することができるか、またはどの程度の具体的な支援が必要となるかという本稿における問いの本質と直結しており、推敲における今後の検討課題の一つとしたい。

なお、答えるのが難しかった質問に関しては、多数の児童が少なくとも 1 つの質問を挙げており、中でも具体的なエピソードを尋ねられる質問を挙げた児童が多く見られた。この事実は、児童らにとって

必要感のある内容で書くことの重要性を示唆しており、作文指導の現場において、“体験したことを具体的に書きなさい”と指示して安易に詳述化を求める指導への問題提起ともなる。

6 おわりに

本研究では、小学校の作文教育において、質問することによる内容の詳述化の支援に関して、教師の負担を低減するために LLM による質問生成の有用性について検証した。どのような質問が適切かを具体的に指示として与えなくとも、比較的有用な質問が生成されることが小学生を対象とした調査によって明らかになった。一方で、具体的 (例えば、本文を引用してそのエピソードを問うような質問) であれば、即推敲時の参考になるわけではないといった課題も浮き彫りになった。

また、本稿で提案したような質問による支援が自己に対する理解の深まりや記述の詳述化を推進するとすれば、質問の質的な問題以外に提示する順番も重要となると考えられる。例えば、簡単で答えやすいものから徐々に内容に深入りするようなものとなるような順番で提示することで同じ質問でもより多くのことを児童から引き出せる可能性がある。今後は、どのような質問がより効果的であるかについての具体的な分析、さらには提示の順番も考慮した支援のあり方についても検討する。

附記 本研究は、相模女子大学ヒトを対象とする研究に関する倫理委員会の承認を得ている (受理番号:23022 号)。

謝辞

本稿における実験授業に際して、児童の皆さん及び担任教諭にご協力頂いた。また、作文評価に際して、中野区立令和小学校清水絵里教諭、東京学芸大学附属小金井小学校橋浦龍彦教諭にご協力を頂いた。厚く御礼申し上げます。

参考文献

- [1] 後藤田結, 田村直良. 学習指導要領に基づく小学生作文支援システムの構築. 言語処理学会 第 22 回年次大会, 2016.
- [2] 大村和正, 久保圭, 黒橋禎夫. ことばつなぎゲーム: ゲーミフィケーションによる小学生の作文教育. 言語処理学会第 27 回年次大会, pp. 895–899, 2021.
- [3] 永田亮, 河合綾子, 須田幸次, 掛川淳一, 森広浩一郎. 作文履歴をトレース可能な子供コーパスの構築. 自然言語処理, Vol. 17, No. 2, pp. 51–65, 2010.
- [4] 宮城信, 今田水穂. 『児童・生徒作文コーパス』の設計. 第 7 回コーパス日本語学ワークショップ, pp. 223–232, 2015.
- [5] 宮城信, 今田水穂. 『児童・生徒作文コーパス』を用いた漢字使用能力の発達過程の分析. 計量国語学, Vol. 31, No. 5, pp. 352–369, 2018.
- [6] 文部科学省. 初等中等教育段階における生成 ai の利活用に関するガイドライン (ver.2.0). https://www.mext.go.jp/content/20241226-mxt_shuukyo02-000030823_001.pdf, 2024.
- [7] 藤川大祐. 初等中等教育実践における生成 AI の活用のあり方—文部科学省「暫定的なガイドライン」前後の状況—. 千葉大学教育学部研究紀要, 第 72 巻, pp. 83–90, 2024.
- [8] 成家雅史, 土屋晴裕, 安部真治, 大澤千恵子, 白勢彩子, 中村和弘, 齋藤ひろみ. 児童作文の評価項目に関する予備的検討-紹介文に見る内容的観点と計量的観点-. 第 137 回全国大学国語教育学会, pp. 311–314, 2019.
- [9] 深谷優子, 市川洋子. 継続したピアレビュー方式の協同推敲活動が小論の作成および推敲に対する書き手の思考態度に与える効果. 東北大学大学院教育学研究科研究年報, Vol. 67, No. 2, pp. 1–16, 2019.
- [10] 堀口史哲. 共同推敲時に働く読み手意識の育成過程に関する研究-共同推敲者の批判的思考に着目して-. 国語科教育, Vol. 95, pp. 53–60, 2024.