

AI とアイデアユニットを活用した日本語学習者向け近代文学多読教材作成手法の研究：
樋口一葉「たけくらべ」を例とした言語的調整による汎用的変換アプローチ
A Universal Transformation Approach for Extensive Reading Material Creation for Japanese
Learners of Modern Literature Utilizing Generative AI-Type LLM and Idea Units:
A Case Study with Higuchi Ichiyo's "Takekurabe"

甘利 実乃[†]
Mino Amari

1. はじめに

本研究は、生成 AI 型大規模言語モデル (LLM: Large Language Models) を活用した日本語教育のための自動/半自動教材生成システムの開発を目的としている。特に、従来の教育手法では対応が難しかった多様な学習者のニーズに応じたカスタマイズ可能な教材作成に焦点を当てる。近年、自然言語処理 (NLP: Natural Language Processing) の分野で生成 AI が急速に進化し、その応用範囲は教育分野にも広がりつつある[1][2]。

従来の言語教育では、教材作成に多大な時間と労力が必要であり、学習者のレベルや興味に応じた教材を迅速に提供することが課題となっていた。しかし、大規模言語モデルを用いることで、自動的かつ効率的に多様な教材を生成することが可能となる。特に、自然言語処理技術を用いることで、既存の文学作品や学習資料を元に、学習者の理解度に応じた適切なレベルの教材を生成することができる[3]。

本研究では、樋口一葉の「たけくらべ」を題材にし、大規模言語モデルを用いて自動的にアイデアユニット (IU: Idea Unit) を抽出・分解し、それを元に教材を生成する手法を提案する。本研究では IU を「1 つ以上の主語と述語を含む、意味的にまとまりのある最小の文構造」と定義し、原文を IU に分割する。そして、学習者のレベルに応じて、IU 内の語彙・文法・文型などを言語的に調整し、原文を参照しながら文章を再構築することで、多読用教材を作成する[4]。

さらに、本手法の有効性を検証するため、具体的なケーススタディとして樋口一葉の「たけくらべ」を例に、IU の抽出と言語的調整を行う。また、文化的背景の理解を促進するため、オノマトペのイラスト化を補助的に活用し、教材の視覚的な魅力と理解度を高める。これにより、教材の質を保持しつつ、効率的な教材作成を実現する。

本研究は、AI を用いた日本語学習者向け教材作成の新たな可能性を提示し、教育現場における実践的な応用を目指す。

1.1 研究の背景と目的

日本語学習者向けの教材作成は、教育の質を高めるために非常に重要である。特に、近代文学を題材にした多読教材は、日本文化への理解を深めるための有効な手段となる。しかし、従来の教材作成方法では、以下のような課題が存在する。

- ① **教材作成の時間と労力:** 教材作成には多大な時間と労力が必要であり、教師や教材作成者にとって大

きな負担となっている[5]。

- ② **学習者の多様性:** 学習者のレベルや興味に応じた教材を迅速に提供することが難しく、画一的な教材では学習効果が限定的である場合がある[6]。
- ③ **文化的背景の理解:** 日本の近代文学は、言語だけでなく文化的背景も重要である。しかし、従来の教材では文化的背景の説明が不足しており、学習者の理解が深まりにくいという問題がある[7]。

本研究の目的は、これらの課題を解決するために、LLM と IU を活用した自動教材生成システムを提案することである。具体的には、以下の 3 点に焦点を当てる。

- ① **効率的な教材生成:** 大規模言語モデルを用いて、自動的に多様な教材を生成することで、教材作成の時間と労力を大幅に削減する[5][6]。
- ② **学習者レベルに応じた調整:** IU を用いて、学習者のレベルに応じた語彙・文法・文型の調整を行い、効果的な学習を支援する[7]。
- ③ **文化的背景の理解促進:** オノマトペのイラスト化を通じて、視覚的に文化的背景を説明し、学習者の理解を深める[8][9][10]。

本研究では、具体的なケーススタディとして樋口一葉の「たけくらべ」を題材にし、IU の抽出と言語的調整を行う。また、生成された教材の有効性を実験的に評価し、その結果を分析する。本研究が提案する手法は、日本語学習者向け教材作成の新たな可能性を提示し、教育現場における実践的な応用を目指す。

1.2 本研究の意義と革新性

本研究は、日本語学習者向けの教材作成における新たなアプローチを提示するものであり、以下の点で意義と革新性を持っている。

- ① **大規模言語モデルの活用:** LLM を活用することで、従来の手作業による教材作成の労力を大幅に削減し、効率的な教材生成を可能にする。このアプローチにより、教材作成の迅速化と多様化が実現される[5][6]。
- ② **アイデアユニットによる構造化:** IU を用いることで、文章を意味的にまとまりのある最小単位に分割し、学習者のレベルに応じて調整する。これにより、学習者個々のニーズに対応したパーソナライズド教材を提供できる[7]。
- ③ **多読教材の自動生成:** 近代文学作品を題材にした多読教材を自動生成することで、日本文化や歴史への理解を深めるとともに、学習者の興味を引き出す。具体的なケーススタディとして、樋口一葉の「たけくらべ」を題材にしている[4]。

[†] 東京外国語大学 Tokyo University of Foreign Studies

- ④ **文化的背景の視覚的補助:** オノマトペのイラスト化を通じて、文化的背景や感情表現を視覚的に補助する。これにより、学習者の理解が深まり、学習意欲が向上することが期待される[8][9][10]。

本研究は、LLM と IU の組み合わせにより、効率的かつ効果的な教材作成手法を提案する。この手法は、日本語学習者だけでなく、他の言語学習者向けの教材作成にも応用可能であり、教育分野における新たな可能性を開く。

2. 関連研究

本章では、LLM を用いた言語教材の自動作成に関する先行研究を包括的にレビューし、本研究の立ち位置と独自性を明確にする。また、IU を活用した言語処理技術および多読教材における視覚的補助の利用についても詳述し、関連する研究との関連性を考察する。

2.1 生成 AI 型 LLM を用いた言語教材の自動作成

2.1.1 自動生成の利点と課題

LLM を用いた言語教材の自動生成は、従来の手作業による教材作成と比較して多くの利点を有している。まず、LLM の利用により、教材作成の効率性が飛躍的に向上する点が挙げられる。従来の教材作成は、専門家の知識と労力を必要とし、多大な時間を要するものであった。LLM を活用することで、大量の教材を短時間で生成することが可能となり、教育現場の負担を軽減できる。

また、LLM は大量のテキストデータから言語の規則性やパターンを学習しているため、多様な文体やトピックに対応した教材の生成が可能である。これにより、学習者のニーズに応じたパーソナライズドな教材提供が実現される。特に、多読教材においては、学習者の興味やレベルに応じた多様なコンテンツを提供することが、学習意欲の向上と効果的な学習を促進する。

一方で、LLM を用いた教材自動生成にはいくつかの課題も存在する。まず、生成された教材の品質保証の問題が挙げられる。LLM は膨大なデータを基に学習しているが、その生成物が常に適切で正確な内容を含むとは限らない。誤情報や不適切な表現が含まれる可能性があり、それらを検出し修正するプロセスが必要である。

さらに、LLM は文化的背景や文脈の理解が不十分な場合がある。特に、言語教育においては、単に正確な言語表現を生成するだけでなく、文化的なニュアンスや背景情報を正確に伝えることが重要である。この点において、LLM の生成物がどの程度これらの要素を含んでいるかを評価し、必要に応じて補完することが求められる。

2.1.2 現行の手法とその評価

LLM を用いた言語教材自動生成の研究は近年活発に行われている。特に、BERT (Bidirectional Encoder Representations from Transformers) や GPT (Generative Pre-trained Transformer) などのモデルは、その高度な言語生成能力により多くの注目を集めている。これらのモデルは、大量のテキストデータを基に事前学習を行い、与えられた入力に対して適切な応答や文章を生成する能力を有している。

BERT は、双方向のトランスフォーマーアーキテクチャを使用しており、文脈の両側から情報を取得することで、より精緻な言語理解を実現している。これにより、文脈に

依存した言語生成が可能となり、教材生成においても高い適用性を持つ。一方、GPT は、事前学習と微調整のプロセスを経ることで、多様なタスクに対応可能な言語生成モデルとして機能する。特に、GPT-3 は、1750 億のパラメータを持ち、高度な文章生成能力を持つことが示されている。

これらのモデルを用いた教材生成の試みとして、英語教育や他の言語教育において成功事例が報告されている。例えば、Zhu ら[2]は、LLM を用いて英語の読解教材を自動生成するシステムを提案し、その有効性を実証している。このシステムでは、学習者のレベルに応じた適切な語彙や文法構造を選択し、パーソナライズドな教材を提供することが可能である。

日本語教育においても、同様の手法が適用され始めている。特に、近代文学作品を題材とした多読教材の自動生成は、日本文化への理解を深めるための有効な手段として注目されている。しかし、日本語特有の文法構造や表現、文化的背景を正確に反映するためには、さらなる研究と改良が必要である。

これらの現行手法は、LLM の能力を最大限に活用しつつも、その限界を認識し、適切な補完手段を講じることで、より高品質な教材生成を目指している。本研究においても、これらの先行研究を踏まえ、LLM の強みを活かした自動教材生成システムを提案し、その有効性を検証する。

2.2 アイデアユニットを活用した言語処理技術

2.2.1 アイデアユニットの定義と応用

IU は、文章を意味的にまとまりのある最小単位に分割するための手法であり、「1 つ以上の主語と述語を含む意味的に独立した文構造」を指す。IU の概念は、言語学や認知科学の分野で発展してきたものであり、特に文章の理解や生成において重要な役割を果たしている。

IU を活用することにより、文章をより細かい単位で解析・生成することが可能となる。これにより、学習者の理解度や興味に応じて適切に調整された教材を提供することができる。例えば、複雑な文構造を持つ文章を IU に分割し、それぞれの IU を独立して扱うことで、学習者がより理解しやすい形で教材を提供することが可能である。

IU の応用範囲は広く、教育のみならず、機械翻訳や要約生成などの分野でもその有用性が示されている。特に、NLP の分野において、IU はテキストの意味構造をより正確に捉えるための有効な手段となっている。

2.2.2 言語処理におけるメリット

IU を活用した言語処理技術には多くのメリットがある。まず、IU は文章の意味的なまとまりを保持しつつ、細かい単位での解析を可能にするため、より精緻な言語処理が可能となる。これにより、文脈に依存した意味解析や生成がより正確に行われる。

次に、IU は文章の再構築や要約生成においても効果を発揮する。文章を IU に分割することで、重要な情報を含む IU のみを選択し、要約を生成することが可能である。また、IU を用いることで、異なる文脈やトピックを含む文章を統合し、一貫性のあるテキストを生成することができる。

さらに、IU は教育現場においても重要な役割を果たす。IU を用いることで、教材の難易度や内容を学習者のレベルに応じて柔軟に調整することが可能である。例えば、初心者向けの教材では、簡単な IU を中心に構成し、上級者向

の教材では、より複雑な IU を含めることで、学習者のニーズに応じたカスタマイズが可能となる。

IU の利点は、以下の点で特に顕著である：

- ① **意味的な一貫性の保持:** IU は意味的に独立した単位であるため、文章全体の意味を損なうことなく再構築や要約が可能である。
- ② **柔軟な調整:** IU を用いることで、教材の内容や難易度を学習者のニーズに応じて調整できる。
- ③ **精緻な解析:** IU に分割することで、文章の細かい意味構造を捉え、精緻な解析が可能となる。

2.2.3 IU を用いた具体的な手法

IU を用いた具体的な手法として、以下のステップが考えられる：

- ① **文章の分割:** 文章を意味的に独立した IU に分割する。これにより、文章全体をより細かい単位で扱うことができる。
- ② **IU の解析:** 分割された IU を解析し、それぞれの意味構造を理解する。このステップでは、各 IU がどのような情報を含んでいるかを詳細に解析する。
- ③ **IU の調整:** 学習者のレベルや興味に応じて、IU の内容や難易度を調整する。例えば、初心者向けには簡単な単語や文法構造を使用し、上級者向けには複雑な表現を用いる。
- ④ **教材の生成:** 調整された IU を組み合わせて、新たな教材を生成する。このステップでは、IU が自然に連結され、一貫性のある文章となるように注意する。

このように、IU を用いた言語処理技術は、教育現場において大きな可能性を秘めている。本研究では、IU を活用して日本語学習者向けの多読教材を自動生成する手法を提案し、その有効性を検証する。IU を用いることで、学習者のニーズに応じたカスタマイズ可能な教材を効率的に提供することが可能となり、教育の質を向上させることが期待される。

2.3 日本語学習者向け多読教材の設計と評価

2.3.1 多読教材の特性と効果

多読は、学習者が大量のテキストを読むことによって言語能力を向上させる手法である。この手法は、特に第二言語学習において効果的であるとされ、語彙力、読解力、文法知識の向上に寄与することが多くの研究で示されている [5][6]。多読教材の特性として、以下の点が挙げられる。

- ① **大量の読書量:** 学習者が大量のテキストに触れることによって、言語に対する感覚を自然に養うことができる。
- ② **学習者のレベルに応じた調整:** 多読教材は、学習者の言語レベルに応じて選定されるため、適切な難易度のテキストが提供される。
- ③ **自主性の尊重:** 学習者が興味のあるテキストを選んで読むことができるため、学習の自主性が尊重され、動機付けが高まる。
- ④ **文脈を通じた学習:** テキストを通じて文脈の中で語彙や文法を学習するため、実用的な言語運用能力が向上する。

これらの特性により、多読は言語学習者にとって非常に効果的な手法であるとされている。特に、日本語学習者に

としては、漢字や語彙の学習において多読が有効であり、多様な表現や文体に触れることができる。

2.4 多読教材における視覚的補助の利用

視覚的補助は、言語学習において重要な役割を果たす。特に多読教材において、視覚的補助は学習者の理解を深め、学習意欲を高めるための効果的な手段である。本節では、オノマトペの視覚化と文脈に基づくプロンプト生成と画像生成について詳細に考察する。

2.4.1 オノマトペの視覚化

オノマトペは、日本語における擬音語や擬態語の総称であり、音や動きを模倣する言葉である。これらは日本語学習者にとって非常に魅力的でありながら、難解な部分でもある。オノマトペの視覚化は、これらの表現を直感的に理解させるための有効な手段である。

まず、オノマトペの定義とその重要性について述べる。オノマトペは、音や動き、感情を直接的に表現するための言葉であり、日常会話や文学作品において頻繁に使用される。これにより、オノマトペは日本語のリズムや音感を学ぶ上で重要な要素となる。

次に、オノマトペを視覚化する手法について説明する。視覚化の手法として、以下のような方法が考えられる。

- ① **イラストやアニメーション:** オノマトペが表す音や動きを視覚的に表現するために、イラストやアニメーションを用いる。この方法は、学習者がオノマトペの意味を直感的に理解するのに役立つ。
- ② **音声と連動した視覚化:** オノマトペの発音と連動した視覚的な表現を用いることで、音と視覚の両方から理解を深める。この方法は、聴覚と視覚を同時に刺激し、記憶に残りやすい。
- ③ **インタラクティブな教材:** 学習者がオノマトペの意味を自ら探索できるようなインタラクティブな教材を作成する。これにより、学習者の自主的な学習を促進する。

オノマトペの視覚化は、日本語の豊かな表現力を学ぶ上で非常に有効である。視覚的な補助を活用することで、学習者はオノマトペの持つ微妙なニュアンスを理解しやすくなり、言語習得の過程がより楽しく効果的なものとなる。

2.4.2 文脈に基づくプロンプト生成と画像生成

文脈に基づくプロンプト生成と画像生成は、LLM の応用により実現可能な技術であり、多読教材の質を高めるための革新的な手法である。この技術は、テキストの文脈を解析し、適切な視覚的補助を自動的に生成することを目的としている。

まず、文脈解析の手法について述べる。文脈解析は、テキストの意味や意図を理解し、関連する情報を抽出するプロセスである。これには、以下のような技術が用いられる。

- ① **NLP:** NLP 技術を用いて、テキストの文脈を解析し、主要なキーワードやフレーズを抽出する。これにより、テキストの主題や重要な概念を特定することができる。
- ② **セマンティック解析:** セマンティック解析により、テキストの意味的な関係性を把握し、文脈を理解する。この手法は、同義語や関連語を含むテキストの意味をより深く理解するのに役立つ。

次に、AI によるプロンプト生成について説明する。プロンプト生成は、文脈解析の結果を基に、視覚的な補助に適したプロンプトを作成するプロセスである。これには、以下のステップが含まれる。

- ① **キーワード抽出:** 文脈解析により抽出されたキーワードを基に、プロンプトを生成する。このプロンプトは、視覚的な補助に適した内容を含むものである。
- ② **プロンプトの調整:** 生成されたプロンプトを、特定の視覚的補助 (例えば、イラストや画像生成 AI) に適用するために調整する。この調整は、プロンプトが具体的かつ明確な内容を持つように行う。

最後に、プロンプトからの画像生成の技術について述べる。画像生成 AI (例えば、DALL-E や StyleGAN) を用いて、生成されたプロンプトに基づく視覚的補助を作成する。この技術は、以下のようなステップを含む。

- ① **プロンプトの入力:** 生成されたプロンプトを画像生成 AI に入力し、視覚的な補助を作成する。AI はプロンプトに基づいて適切な画像を生成する。
- ② **画像の調整と最適化:** 生成された画像を教材に適用するために調整し、最適化する。このステップでは、画像の品質や適切性を確認し、必要に応じて修正を行う。

文脈に基づくプロンプト生成と画像生成の技術は、多読教材における視覚的補助の質を飛躍的に向上させる。これにより、学習者はテキストの内容をより深く理解し、視覚とテキストの相互作用を通じて効果的に学習することができる。本研究では、これらの技術を活用し、多読教材の新たな可能性を探求する。

3. 提案手法

3.1 システムアーキテクチャ

本章では、本研究で提案する自動教材生成システムのアーキテクチャについて詳細に述べる。システムアーキテクチャは、データ収集と前処理、モデルの設計と訓練の各要素から構成される。

3.1.1 データ収集と前処理

データ収集は、システムの性能を左右する重要なプロセスである。本研究では、日本語教育に特化した高品質なテキストデータを収集することに重点を置いた。収集対象とするデータは、近代文学作品、学術論文、新聞記事、教育用テキストなど多岐にわたる。特に、樋口一葉の「たけくらべ」のような近代文学作品は、文化的背景や言語の豊かさを学習者に提供するために重要である。

データ収集の具体的な手法としては、以下のステップを踏む。

- ① **データソースの選定:** 高品質なデータソースを選定する。具体的には、青空文庫や国立国会図書館のデジタルコレクションなど、信頼性の高いデータベースからデータを収集する。
- ② **データの収集:** Web スクレイピングや API を使用してデータを収集する。収集されたデータは、生データとして保存され、後続の前処理プロセスに渡される。
- ③ **データのフィルタリング:** 不要な情報やノイズを除去するために、正規表現やフィルタリングアルゴ

リズムを用いてデータをクレンジングする。この過程で、重複データや誤った情報が排除される。

前処理は、収集されたデータをモデルの訓練に適した形式に変換するための重要なステップである。前処理の具体的な手法としては、以下のプロセスを含む。

- ① **トークン化:** テキストデータを単語やフレーズ単位に分割する。形態素解析ツール (例えば、MeCab や Sudachi) を用いて、日本語テキストを正確にトークン化する。
- ② **正規化:** トークン化されたデータを統一的形式に変換する。例えば、ひらがな、カタカナ、漢字の統一や、全角半角の統一などが含まれる。
- ③ **ストップワードの除去:** 学習に不要な一般的な語彙 (ストップワード) を除去する。これにより、モデルが有用な情報に集中して学習できるようになる。
- ④ **文脈情報の付与:** テキストの文脈情報を付与するために、依存関係解析や品詞タグ付けを行う。これにより、モデルがテキストの意味構造をより深く理解できるようになる。

これらの前処理を経たデータは、次のモデル訓練ステップに進む。

3.1.2 モデルの設計と訓練

モデルの設計と訓練は、システムアーキテクチャの中核である。本研究では、LLM を用いた言語生成モデルを採用し、特に日本語テキストの生成に適したモデルを設計した。

モデル設計の具体的な要素としては、以下の点が挙げられる。

- ① **モデルの選定:** 言語モデルとして、BERT や GPT-4 などの最新のトランスフォーマーベースのモデルを採用する。これらのモデルは、双方向の文脈理解や高精度の生成能力を有しており、教材生成に適している。
- ② **アーキテクチャのカスタマイズ:** 日本語特有の文法構造や表現を学習するために、モデルのアーキテクチャをカスタマイズする。具体的には、日本語の形態素や依存関係を考慮したアテンション機構を導入する。
- ③ **ハイパーパラメータの調整:** モデルの性能を最適化するために、学習率、バッチサイズ、エポック数などのハイパーパラメータを調整する。これにより、過学習を防ぎつつ、高精度の生成結果を得ることができる。

モデル訓練の具体的な手法としては、以下のプロセスを含む。

- ① **データの分割:** 前処理されたデータを訓練データ、検証データ、テストデータに分割する。通常、80% を訓練データ、10% を検証データ、10% をテストデータとして使用する。
- ② **訓練プロセス:** トランスフォーマーモデルを使用して、訓練データを用いてモデルを訓練する。このプロセスでは、損失関数 (例えば、クロスエントロピー損失) を最小化するように勾配降下法を適用する。

- ③ **検証とチューニング:** 訓練中に検証データを用いてモデルの性能を評価し、ハイパーパラメータを調整する。これにより、モデルの汎化性能を向上させる。

- ④ **テストと評価:** 訓練完了後、テストデータを用いてモデルの最終的な性能を評価する。評価指標としては、精度、再現率、F1 スコアなどが用いられる。

モデルの設計と訓練を通じて、LLM を活用した高精度な日本語教材生成モデルを構築する。本研究のモデルは、最終的には日本語学習者向けの多読教材の生成に特化することを目標とし、学習者のレベルや興味に応じたパーソナライズドな教材提供を目指している。

ただし、一方で、どの日本語教育関係者でも身近な AI 環境を用いて実践できる手法の開発にも力点を置いており、本研究においては、後者を実現させるように、実際の設計は、既存の LLM 上のプロンプトベースで行う。

3.2 アイデアユニットの抽出と調整

IU は、文章を意味的にまとまりのある最小単位に分割するための手法であり、本研究では自動教材生成において重要な役割を果たす。本節では、IU の抽出方法とレベル調整について詳細に述べる。

3.2.1 IU の抽出方法

IU の抽出は、テキストを意味的に独立した単位に分割するプロセスである。これにより、教材の各部分が明確な意味を持つ最小単位として扱われ、学習者の理解を助けることができる。本研究では、甘利[11]の手法を参照しながら IU 分割を行う。

IU 分割手法には以下のプロンプトを使用する。

多言語対応アイデアユニット近似評価法

本プロンプトは、文章の理解と分析を深化させるためのガイドラインです。このプロンプトは、文の構成要素の詳細な分析、アイデアユニット間の論理的関連性の明確化、および多様な言語文化に対する適応性の向上を目的としています。

アイデアユニットの分割ルール

1. 文は、主語と述語の関係に基づき、一つまたは複数のアイデアユニットに分割される。
2. 主語が文中に明示されていない場合、文脈や共参照を通じて推測し、アイデアユニットを構成する。
3. 複文は、それぞれの独立した節を基に、複数のアイデアユニットに分割する。ただし、これらのアイデアユニット間の論理的な関連性を明確に示す。

アイデアユニットの評価ルール

1. 各アイデアユニットが意味的に完結しているかどうかを評価する。
2. アイデアユニット間の論理的つながりを確認し、その系統的な関連性を明確にする。
3. アイデアユニットの並び替えや統合を通じて、文の意味をより明確かつ論理的に展開させる。

多言語適応性の向上

1. さまざまな言語における文法構造と表現の特性を考慮し、アイデアユニットの分割と評価ルールを適用する。
2. 言語によって異なる表現形式（たとえば、能動態と受動態、暗示的表現など）を特定し、それぞれの言語文化の文脈において意味がどのように構成されるかを理解する。

使用例

日本語

入力文: 彼女は図書館に行き、本を借りた。

アイデアユニット分割:

- 彼女は図書館に行った。
- 彼女は本を借りた。

英語

入力文: She went to the library and borrowed a book.

アイデアユニット分割:

- She went to the library.
- She borrowed a book.

IU の抽出には以下のステップが含まれる。

- ① **文のトークン化:** テキストを文単位に分割する。文単位の分割は、NLP ツール（例えば、形態素解析ツールや文分割アルゴリズム）を用いて行う。文単位の分割は、文末の句読点や接続詞を基に行われる。
- ② **文内の主語・述語の識別:** 各文において、主語と述語を識別する。これには、依存関係解析や構文解析ツールを使用する。主語と述語の関係を明確にすることで、IU の核となる意味構造を抽出する。
- ③ **意味的にまとまりのある単位の抽出:** 主語と述語を含む意味的にまとまりのある単位を抽出する。このプロセスでは、各文の主語と述語が一貫した意味を持つかどうかを判断し、必要に応じて文を複数の IU に分割する。

IU の境界設定: 抽出された IU の境界を設定する。IU の境界は、文法的な構造や意味的な切れ目に基づいて決定される。これにより、各 IU が独立した意味を持つ最小単位として機能する。

具体的な例を挙げると、以下ようになる。

原文: “樋口一葉は、日本の近代文学を代表する作家であり、多くの名作を残している。彼女の作品には、深い心理描写と豊かな文化的背景が含まれている。”

IU の抽出結果:

- ① “樋口一葉は、日本の近代文学を代表する作家である。”
- ② “樋口一葉は、多くの名作を残している。”
- ③ “彼女の作品には、深い心理描写と豊かな文化的背景が含まれている。”

このように、IU を抽出することで、文章を意味的に独立した単位に分割し、学習者が理解しやすい形で教材を提供することが可能となる。

3.2.2 IU のレベル調整

IU のレベル調整は、抽出された IU を学習者のレベルに応じて調整するプロセスである。これにより、各学習者に

適した教材を提供することができる。レベル調整には以下のステップが含まれる。

- ① **語彙の難易度調整:** IU 内の語彙を学習者のレベルに応じて調整する。具体的には、難解な語彙を簡単な同義語に置き換えたり、注釈を付け加えるなどの方法がある。例えば、「心理描写」を「心の動きを描くこと」に置き換えるなどである。
- ② **文法構造の調整:** 学習者の文法知識に応じて、IU の文法構造を調整する。例えば、複雑な文を単純な文に分割したり、文法的に簡単な表現を用いるなどの方法がある。これにより、学習者が文の意味を理解しやすくなる。
- ③ **文型の調整:** 学習者のレベルに応じて、IU の文型を調整する。例えば、敬語を簡単な形に変換したり、主語を明示するなどの方法がある。これにより、文の構造が明確になり、学習者が文の意味を把握しやすくなる。
- ④ **文化的背景の補足:** IU に含まれる文化的背景や歴史的な情報を補足する。特に日本語学習者にとって、文化的な背景が理解の助けとなる場合が多いため、必要に応じて注釈や解説を付け加える。具体的な例を挙げると、以下のようになる。

原文の IU: “彼女の作品には、深い心理描写と豊かな文化的背景が含まれている。”

レベル調整後の IU: “彼女の作品には、人々の心の動きを描いた部分と、日本の文化についての情報がたくさん含まれている。”

このように、IU のレベル調整を行うことで、学習者の理解度や興味に応じたパーソナライズされた教材を提供することが可能となる。本研究では、これらの手法を用いて、日本語学習者向けの多読教材を自動生成し、その有効性を検証する。

3.3 多読教材の生成

本研究では、IU を活用した自動教材生成システムを提案する。生成された教材は、多読に最適化され、学習者のレベルや興味に応じてカスタマイズ可能である。本節では、教材生成のフローとカスタマイズについて詳細に述べる。

3.3.1 教材生成のフロー

多読教材の生成には、以下のフローを採用する。各ステップは、効率的かつ効果的な教材生成を実現するために設計されている。

- ① **入力データの準備:** 多読教材の生成には、高品質なテキストデータが必要である。まず、入力データとして利用するテキストを選定する。選定基準として、言語の豊かさ、文化的背景の重要性、学習者の興味を引く内容であることが挙げられる。例えば、近代文学作品や新聞記事などが適している。
- ② **IU の抽出:** 前述の方法を用いて、選定されたテキストから IU を抽出する。IU は、意味的に独立した最小単位であり、学習者が理解しやすい形で分割される。
- ③ **IU のレベル調整:** 抽出された IU を学習者のレベルに応じて調整する。具体的には、語彙の難易度を

調整し、文法構造を簡略化する。また、必要に応じて文化的背景の説明を追加する。

- ④ **教材の構築:** 調整された IU を組み合わせて教材を構築する。この際、IU 間の論理的な関連性を考慮し、一貫性のあるテキストを作成する。教材の構造は、導入部、本文、まとめといった一般的な構成を持つことが望ましい。
- ⑤ **視覚的補助の追加:** オノマトペの視覚化や文脈に基づく画像生成を活用し、視覚的補助を追加する。これにより、学習者の理解を深め、興味を引く教材を作成する。
- ⑥ **最終調整と検証:** 生成された教材を最終調整し、学習者や専門家による検証を行う。必要に応じてフィードバックを反映し、教材の品質を向上させる。

3.3.2 教材のカスタマイズ

教材のカスタマイズは、学習者のニーズやレベルに応じた個別対応を可能にするための重要な要素である。本研究では、以下の方法で教材のカスタマイズを行う。

- ① **学習者レベルに応じた調整:** 学習者の日本語能力に応じて、IU の難易度を調整する。具体的には、初級者向けには簡単な語彙や文法構造を使用し、上級者向けには複雑な表現を取り入れる。また、注釈や翻訳を追加し、理解を助ける工夫を行う。
- ② **学習者の興味に応じた内容選定:** 学習者の興味を引くトピックを中心に教材を構成する。例えば、歴史、文化、科学など、学習者が関心を持つ分野のテキストを選定する。これにより、学習意欲を高め、学習効果を向上させることができる。
- ③ **視覚的補助のカスタマイズ:** 学習者の理解を助けるために、視覚的補助を適切にカスタマイズする。オノマトペの視覚化や文脈に基づく画像生成を活用し、テキストの内容を視覚的に補完する。これにより、学習者の理解を深めるとともに、学習意欲を高めることができる。
- ④ **インタラクティブ要素の追加:** 学習者が積極的に学習に参加できるように、インタラクティブな要素を追加する。具体的には、クイズ形式の問題やディスカッションを促す問いかけなどを組み込む。これにより、学習者の理解を確認し、学習効果を高めることができる。
- ⑤ **フィードバックの提供:** 学習者が自身の理解度を自己評価できるように、フィードバック機能を提供する。具体的には、AI を活用して学習者の回答を評価し、適切なフィードバックを提供するシステムを構築する。これにより、学習者は自分の弱点を把握し、効率的に学習を進めることができる。

以上の手法を用いることで、学習者一人ひとりに最適化された多読教材を生成し、効率的かつ効果的な日本語学習を支援することが可能となる。本研究では、これらのカスタマイズ手法を組み合わせることで、学習者のニーズに応じた高品質な教材を提供することを目指す。

4. 実践と評価

4.1 樋口一葉「たけくらべ」の選定理由

本研究において、樋口一葉の「たけくらべ」を実践の対象作品として選定した理由は以下の通りである。

まず、「たけくらべ」は日本近代文学の中でも特に重要な位置を占める作品であり、その文学的価値と歴史的意義が非常に高い。樋口一葉は明治時代を代表する女性作家であり、その作品群は当時の社会状況や人々の生活、心情を精緻に描写している。「たけくらべ」もその一つであり、東京の下町に生きる人々の日常や心の葛藤を、繊細な筆致で描き出している。

次に、「たけくらべ」は現代の日本語話者にとっても理解が困難な古典的な日本語で書かれている点が挙げられる。明治時代の言葉遣いや文体は、現代の日本語とは大きく異なり、多くの現代日本語話者にとっては読解が難しい。このような難解なテキストを多読教材として適切に変換できるかどうかを検証することは、AI と人間の協働による教材生成手法の有効性を評価する上で極めて重要である。

さらに、「たけくらべ」を教材化することで得られる教育効果も見逃せない。明治時代の東京の風俗や人々の生活様式、価値観を学習者に伝えることができる点は、日本文化の理解を深める上で非常に有益である。特に、日本語を学ぶ外国人学習者にとって、日本の近代史や文化に対する理解を深めることができる教材は価値が高い。

また、「たけくらべ」のように難解な古典的テキストを AI と人間が協働して多読教材化できるならば、およそほとんどの近代文学作品の多読教材化が可能であると考えられる。これは、AI の自然言語処理技術と人間の専門知識を組み合わせることで、従来の手法では対応が難しかった教材作成の課題を克服できることを示唆している。

以上の理由から、本研究では樋口一葉の「たけくらべ」を対象作品として選定し、その多読教材化のプロセスを通じて、LLM と日本語教育専門家の協働による教材生成手法の有効性を実証することを目指す。

4.2 IU 抽出と言語的調整の詳細

本節では、樋口一葉の「たけくらべ」第 1 章を対象に、IU の抽出および語彙・文法・文型の調整について詳細に述べる。IU の抽出は、文章を意味的に独立した最小単位に分割するプロセスである。本研究では、甘利[11]に基づく前掲のプロンプトを使用し、Gemini 1.5 Pro による LLM としての力のみを用いて IU を抽出した。

以下の表は、Gemini 1.5 Pro によって前述のプロトコルを実行し、「原文の文脈を詳しく分析して、最適な主語を付けて IU 分割後の文章を完成させること」という条件に基づいて自動的に生成された IU 分割結果である。Python などのプログラミング言語は使用せず、Gemini 1.5 Pro 自身の能力により実施された。

4.2.1 原文の IU 分割

以下の表 1 は、上記のプロセスを経て Gemini 1.5 Pro が自動分割して表として出力した、「たけくらべ」第 1 章の全 114 の IU である。この時点ではプロンプトからの指示だけで生成されており、人手による手動修正は一切加えていない。

表 1 「たけくらべ」第 1 章 IU 分割

IU 番号	アイデアユニット
IU1	大門の見返り柳は、とても長い。
IU2	お歯ぐる溝に燈火が映っている。
IU3	三階の騒ぎも手に取るようだ。
IU4	明け暮れなしに車の往来がある。

IU5	車の往来から、計り知れない全盛をうらない知る。
IU6	大音寺前は、名は広くさい。
IU7	人は、大音寺前をさりととは陽気な町と申している。
IU8	三嶋神社の角を曲がってから、これぞと思うような大きな建物はない。
IU9	かたぶいた軒端の干軒長屋、二十軒長屋がある。
IU10	商売は、あまり利益にならないところである。
IU11	雨戸は半開きになっている。
IU12	雨戸の外には、不思議な形に紙を切り抜いて胡粉を塗り、彩色をしたものが貼ってある。
IU13	それは、田楽のようだ。
IU14	裏に貼ってある串の様子も面白い。
IU15	一軒だけでなく、二軒、三軒と続いている。
IU16	人々は、朝日に干して夕日にしまう。
IU17	手入れは、どの家でも同じだ。
IU18	一家内がこの作業にかかっている。
IU19	ある人が、それは何だと尋ねた。
IU20	別の人は、知らないのかと聞き返した。
IU21	別の人は、霜月西の日に、神社に欲深い様にかつぐものと言った。
IU22	別の人は、それは、熊手の下ごしらえだと言った。
IU23	熊手を作る人々は、正月の門松をとりつける頃から取り掛かる。
IU24	一年を通して、それは誠の商売になる。
IU25	熊手を作る人々は、片手間仕事でも、夏から手足を彩色する。
IU26	熊手を作る人々は、新年着の支度も、これをあてにする。
IU27	人々は、「南無や大鳥大明神、買う人にさえ大福を与えてください、製造元の我々には万倍の利益を」とロクに言う。
IU28	しかし、熊手を作る人々は、そうは思っていない。
IU29	このあたりに、大金持ちのうわさも聞かない。
IU30	任んでいる人の多くは廓者だ。
IU31	廓者の良人は小格子の中にいる。
IU32	下足札をそろえて、がらんがらんと音がする。
IU33	夕暮れから羽織を引っかけて、廓者が外出する。
IU34	廓者は、後ろで切火を打つ女房の顔を見るのが最後かもしれないと思う。
IU35	廓者は、十人斬りの側杖で無理心中をし損ねた。
IU36	廓者は、恨みは、この身の果てまで続くと思う。
IU37	廓者は、死ぬと言ったら命がけの勤めだと思う。
IU38	廓者は、遊山のように見えるのも面白い。
IU39	大離の下には、娘がいる。
IU40	娘は、新造と呼ばれている。
IU41	娘は、七軒の何屋かの客回りだ。
IU42	娘は、提灯を下げてちょこちょこ走り回っている。
IU43	それは、修業だ。
IU44	人々は、娘が卒業して何になるのかと思う。
IU45	とにかく、檜舞台と見たのは面白い。
IU46	垢抜けした三十歳過ぎの女性がいる。
IU47	小さっぱりした唐棧のぞろいだ。
IU48	女性は、紺足袋をはいている。
IU49	女性は、雪駄をちゃらちゃら鳴らしている。
IU50	女性は、忙しそうだ。
IU51	女性は、横抱きの小包を持っている。
IU52	茶屋が棧橋ととんと沙汰している。
IU53	茶屋は、ここから料理を運ぶ。
IU54	このあたりでは、茶屋を仕立て物屋と呼ぶ。

IU55	このあたりの風俗は、他の場所と違っている。
IU56	女は、きちんと後帯をしている人が少ない。
IU57	女は、粋がって市広の巻帯をしている。
IU58	年増はまだいい。
IU59	十五六の小娘が酸漿をふくんでも、この格好だ。
IU60	人によっては、目を覆いたくなるだろう。
IU61	所柄、仕方がない。
IU62	昨日、私は河岸の店で何紫という源氏名を聞いた。
IU63	今日、地回りの吉という者が、手慣れない焼き鳥の夜店を出している。
IU64	地回りの吉は、身代を叩き、骨になったら、再び古巣に帰るつもりだ。
IU65	地回りの吉は、その時は、内儀の格好をするだろう。
IU66	地回りの吉は、どこの素人よりも、それらしく見えるだろう。
IU67	このあたりに染まらない子供はいない。
IU68	秋の九月の仁和賀の頃の大路を見てほしい。
IU69	子供たちは、露八の物真似をしている。
IU70	子供たちは、榮喜の作ったものを着ている。
IU71	子供たちは、孟子の母も驚くような上達ぶりだ。
IU72	子供たちは、うまいと褒められて、今日も張り切っている。
IU73	生意気な子供は、七つ八つから威張っている。
IU74	生意気な子供は、やがて、肩に手を置き、鼻歌を歌うようになる。
IU75	十五の少年は、ませている。
IU76	十五の少年は、学校の唱歌にも「ぎっちゃんちゃん」と拍子をとる。
IU77	十五の少年は、運動会で木遣り音頭を踊るかもしれない。
IU78	それでも、教育は難しい。
IU79	教師は苦勞しているだろう。
IU80	入谷の近くに育英舎という学校がある。
IU81	育英舎は私立だが、生徒数は千人近くいる。
IU82	狭い校舎に生徒がぎゅうぎゅう詰めた。
IU83	教師は人望がある。
IU84	このあたりでは、学校といえば育英舎のことだ。
IU85	通っている子供たちは様々だ。
IU86	子供たちの父親は、火消し、鳶、人足などだ。
IU87	ある子供は、父親が刎橋の番屋にいることを知っている。
IU88	子供たちは、梯子登りの真似をして忍び返しを折ってしまう。
IU89	子供たちは、訴えられても言い訳をする。
IU90	三百代言の子供もいるだろう。
IU91	ある子供は、父親が馬丁だからかわれて、恥ずかしそうにしている。
IU92	出入りの貸座敷の隠し子が、寮に住んでいる。
IU93	隠し子は、華族の息子を気取っている。
IU94	隠し子は、ふさ付き帽子をかぶって、洋服を着て、花々敷を歩いている。
IU95	坊ちゃん坊ちゃんと呼ばれて、子供たちが追従している。
IU96	龍華寺の信如という少年がいる。
IU97	信如は、千筋と称される黒髪をしている。
IU98	信如は、十五歳で並背だ。
IU99	信如のいが栗の頭髪も、やがて、墨染めに変わるだろう。
IU100	信如は、袖の色も変わるだろう。
IU101	人々は、信如は、本当に発心したのだろうかと思う。
IU102	信如は、親譲りの勉強家だ。
IU103	友達は、信如がおとなしいのを心配してい

IU104	る。 友達は、信如に様々な悪戯を仕掛ける。
IU105	ある時、友達は猫の死骸を縄にくくりつけて、信如に投げつけた。
IU106	友達は、お役目だから引導を頼むと言った。
IU107	それは昔の話だ。
IU108	今は、信如は校内一の人だ。
IU109	信如を侮る者はいない。
IU110	信如は十五歳で、並背だ。
IU111	信如は、いが栗色の頭髪をしている。
IU112	信如の髪は、思いの外俗っぽくない。
IU113	信如は、藤本信如と名乗っている。
IU114	信如は、どこかしらお坊さんのようだ。

4.2.2 語彙・文法・文型の調整

IU への分割後、各 IU を学習者のレベルに応じて調整するプロセスである。語彙の難易度調整、文法構造の簡略化、文型の調整が行われる。以下にその具体的な手順を示す。

具体的な手順:

- ① **語彙の難易度調整:** 抽出された IU 内の語彙を学習者のレベルに応じて調整する。具体的には、難解な語彙を簡単な同義語に置き換えたり、必要に応じて注釈を付け加える。
例: 「心理描写」を「心の動きを描くこと」に置き換える。
- ② **文法構造の調整:** 学習者の文法知識に応じて、IU の文法構造を調整する。例えば、複雑な文を単純な文に分割したり、文法的に簡単な表現を用いることで、学習者が文の意味を理解しやすくする。
例: 「多くの名作を残している。」を「彼女はたくさんの有名な作品を書いた。」に変える。
- ③ **文型の調整:** 学習者のレベルに応じて、IU の文型を調整する。例えば、敬語を簡単な形に変換したり、主語を明示することで、文の構造が明確になり、学習者が文の意味を把握しやすくする。
例: 「彼女の作品には、深い心理描写と豊かな文化的背景が含まれている。」を「彼女の作品は、人々の心の動きを描き、日本の文化についても詳しい。」に変える。
- ④ **文化的背景の補足:** IU に含まれる文化的背景や歴史的な情報を補足する。特に日本語学習者にとって、文化的な背景が理解の助けとなる場合が多いため、必要に応じて注釈や解説を付け加える。
例: 「彼女の作品には、深い心理描写と豊かな文化的背景が含まれている。」に対して、「彼女の作品は、人々の心の動きを描き、日本の文化についても詳しい。例えば、明治時代の東京の生活や風俗が詳しく描かれている。」を加える。

これらの調整を通じて、学習者のレベルや興味に応じたパーソナライズドな多読教材を提供することが可能となる。本研究では、これらの手法を用いて生成された教材の有効性を検証し、最適な日本語教育の実現を目指す。

4.3 視覚的補助の適用

本節では、生成された多読教材における視覚的補助の適用について述べる。特に、オノマトペの視覚化と生成された画像の具体的な適用例を詳述する。

4.3.1 オノマトペの視覚化

オノマトペは日本語の特徴的な表現手法であり、音や動きを言語化することで、読者の理解を深める役割を果たす。本研究では、オノマトペを視覚化することで、学習者の理解を補助する手法を提案する。

オノマトペの視覚化は以下のプロセスを経て行われる：

- ① **オノマトペの抽出:** ChatGPT-4o がテキストからオノマトペを自動的に抽出する。例えば、「たけくらべ」第1章の中で使用されている「ちょこちょこ」などが対象となる。
- ② **画像生成用プロンプトの生成:** 抽出されたオノマトペに基づき、画像生成用のプロンプトを作成する。このプロセスも ChatGPT-4o によって原文と原文の文脈も参照しながら自動的に行われる。例えば、「ちょこちょこ」に対しては、「A lively scene of a young girl running lightly on a narrow street at night, holding a glowing paper lantern. She is dressed in a traditional kimono, and her feet move quickly in small steps, creating a sense of “choko choko” movement. The background shows other lantern-lit houses and shops, adding to the bustling night-time atmosphere. Motion lines and a slightly blurred effect on her legs enhance the sense of rapid, light-footed movement.」といった具体的な情景を描写するプロンプトが生成される。画像生成用のプロンプトは日本語で書いても良いが、ChatGPT-4o から内部でプロンプトを渡される画像生成 AI である DALL-E 3 は、基本的には英語で動いており、細かなニュアンスは、現時点では英語の方が伝わりやすい。
- ③ **画像生成:** 生成されたプロンプトを用いて、DALL-E 3 を使用して画像を生成する。AI による生成画像は多様なバリエーションを持つが、その中から日本語教育専門家が最も適切な画像を選定する。場合によっては、プロンプトの書き換えや、部分指定しての大幅な修正が必要となることもある。
- ④ **選定と適用:** 選定された画像を教材に適用し、視覚的な補助として使用する。これにより、オノマトペの表現が視覚的に補完され、学習者の理解を促進する。

4.3.2 生成された画像の適用例

具体的な適用例として、「たけくらべ」第1章の中で使用されている「ちょこちょこ」のオノマトペを取り上げる。以下に、生成された画像例を図1に示す。なお、今回は、ChatGPT-4o が自動生成したプロンプトに、日本のマンガ風の白黒のイラストで画像生成するようにとの指示を付けて加えて DALL-E 3 で実行された。



図1 「ちょこちょこ」

この画像は、AI がオノマトペを自動的に抽出し、プロンプトを生成して作成されたものである。日本語教育専門家は、この画像は完璧であるとは言えないものの、教材に使用可能であると判断し、挿し絵にすることに決定した。

キャプションは日本語教師による手動で「少女がちょうちんを持って夜の町をちょこちょこ走っています。いそがしそうです。」と付けられた。キャプションを付けることにより、具体的な状況が視覚と共に伝わると考えられる。

視覚的補助を加えることで、学習者はオノマトペの具体的な使い方やニュアンスを直感的に理解することができるが、このイラストは印刷を念頭に作られているため実際の音や動きといったものはない。マンガ風ではあるものの、マンガ絵ほどの説得力があるかは未知数ではある。現状の LLM と画像生成 AI により、自動生成されたという点に、今後の発展と応用の広がり期待される。

本研究では、視覚的補助を通じて学習者の理解を深めることを目指しているが、全てのオノマトペに対して画像生成することは、量が多い場合にはあまり現実的ではなく、また、必ずしも画像に頼った説明が最適であるとも限らないため、言葉による解説も並行検証していく必要がある。

4.4 教材生成過程の詳細

本節では、生成された教材が学習者に適した形となるように行われたレベル変換と最終調整の詳細、およびその評価方法について述べる。特に、Gemini 1.5 Pro による自動レベル変換と日本語教育専門家による手動調整について詳述する。

4.4.1 レベル変換と最終調整

教材生成において重要なのは、学習者のレベルに適した内容に調整することである。本研究では、樋口一葉の「たけくらべ」第1章を JLPT N4 レベルに合わせて調整を行った。この過程は以下のステップで実施された。

- ① **Gemini 1.5 Pro による自動レベル変換:**
 - Gemini 1.5 Pro を用いて、原文の各 IU を自動的に JLPT（日本語能力試験）N4 レベルの語彙・文法・文型に変換する。具体的な変換ルールとして、以下の点をプロンプト経由で指示した：
 - N4 レベルの語彙・文法を使用する（例：「俗っぽい」→「普通」）。
 - 読みやすさを考慮して、表現を調整する。
 - 固有名詞は変更しない。
 - 旧仮名遣いは新仮名遣いに変換する。
- ② **日本語教育専門家による手動調整:**
 - 自動変換されたテキストを日本語教育の専門家がレビューし、必要に応じて手動で修正を加える。このプロセスでは、学習者の理解を助けるための注釈追加や文脈の補足が行われる。

以下に、Gemini 1.5 Pro による自動レベル変換の結果と、日本語教育専門家による手動調整後の文章を対照させた表を示す。

表2 LLM 文と日本語教師調整文との対照

LLM (Gemini 1.5 Pro)	日本語教師 (人間)
大門をぐるりと回ると、とても長い柳の木が見えます。歯を黒く染めるための溝には、灯の	大門をぐるりと回ると、とても長いやなぎの木が見えます。歯を黒くそめるためのみぞに

明かりが映り、三階からは騒いでいる声が聞こえてくるようです。朝から晩まで、たくさんの車が行き来しています。

車の多さを見ると、この町のとても栄えている様子がうかがえます。大音寺前は、お寺の名前ですが、仏教らしい感じはしません。人々は、大音寺前をにぎやかな町だと言っています。三嶋神社の角を曲がると、これだと思えるような大きな建物はありません。その代わり、片側が傾いた屋根の十軒長屋、二十軒長屋が並んでいます。商売であまり儲からない場所なので、雨戸は半分閉まっています。

雨戸の外には、不思議な形に紙を切り抜いて、白い粉を塗ったり色を塗ったりしたものが貼ってあります。それは、田楽のようです。裏に貼ってある串の様子が面白いです。軒だけではなく、二軒、三軒と続いています。人々は、朝日に干して、夕方になると片付けます。どこの方でも同じように手入れをしています。家族みんなでこの作業をしています。ある人が、それは何かと尋ねました。別の人が、知らないのかと聞き返しました。別の人が、11月の酉の日に、神社に欲張りな人がかつぐものだと言いました。別の人が、それは、熊手の下ごしらえだと言いました。

熊手を作る人たちは、正月の門松を取り付ける頃から準備を始めます。一年中、それは本当の商売になります。熊手を作る人たちは、他の仕事をしていても、夏から手足を色付けします。お正月に着る新しい服を買うお金も、熊手を売って得るつもりです。人々は、「大鳥大明神様、熊手を買う人に福を与えてください。私たちが作る者には、その何倍もの利益をください」とロ々に言います。でも、熊手を作る人たちは、そんなふうには思っていない。このあたりに、お金持ちがいるという話は聞いたことがありません。住んでいる人の多くは、遊郭で働く女性と暮らしている男です。

遊郭で働く女性と暮らしている男は、小さな格子の中に入ります。下駄を預ける札を並べて、がらんがらんと音がします。夕方になると、遊郭で働く女性と暮らしている男は、羽織を着て出かけます。男は、後ろで火を起している女房の顔を見るのが最後かもしれないと思います。男は、十人を一度に斬る刀を持って無理心中をしようとして失敗しました。男は、恨みはこのまま一生続くのだと思います。「死ぬ」と言ったら、本当に命をかけて働くつもりです。男は、遊びに行くように見えるのも面白いです。

大籬の下には、娘がいます。娘は、新造と呼ばれています。

は、ひの明かりがうつり、三階からはさわいでいる声が聞こえてくるようです。朝から晩まで、たくさんの車が行き来しています。

車の多さを見ると、この町のとても栄えているようすがわかります。大音寺前は、お寺の名前ですが、仏教らしい感じはしません。人々は、大音寺前をにぎやかな町だと言っています。三嶋神社のかどをまがると、すごいと思えるような大きな建物はあります。そのかわり、片側がかたむいた屋根の十軒長屋、二十軒長屋がならんでいます。商売であまりもうからない場所なので、あまどは半分閉まっています。

あまどの外には、ふしぎな形に紙を切りぬいて、白いこなをぬったり色をぬったりしたものがはってあります。それは、とうふのおでんのようなものです。うらにはあってあるくしのようすもおもしろいです。一軒だけでなく、二軒、三軒と続いています。人々は、朝日にして、夕方になるとかたづけします。どこの方でも同じように手入れをしています。家族みんなでこの作業をしています。ある人が、それは何かとたずねました。別の人が、知らないのかと聞き返しました。別の人が、11月の熊手の下ごしらえの日、神社によくばりな人がかつぐものだと言いました。別の人が、それは、熊手の下ごしらえだと言いました。

熊手を作る人たちは、正月のかどまつを取り付けるころから準備を始めます。一年中、それは本当の商売になります。熊手を作る人たちは、ほかの仕事をしていても、夏から手足を色付けします。お正月に着る新しい服を買うお金も、熊手を売って手に入れるつもりです。人々は、「大鳥大明神様、熊手を買う人に福を与えてください。私たちが作る者には、そのなん倍ものりえきをください」とロ々に言います。でも、熊手を作る人たちは、そんなふうには思っていない。このあたりに、お金持ちがいるという話は聞いたことがありません。住んでいる人の多くは、ゆうかくで働く女性とくらしている男です。

ゆうかくで働く女性とくらしている男は、小さなこうしの中に入ります。げたをあずけるふだを並べて、がらんがらんと音がします。夕方になると、ゆうかくで働く女性とくらしている男は、はおりを着て出かけます。男は、後ろで火を起しているようぼうの顔を見るのが最後かもしれないと思います。男は、十人を一度にきる刀を持って無理心中をしようとして失敗しました。男は、うらみはこのまま一生続くのだと思います。「死ぬ」と言ったら、本当に命をかけて働くつもりです。男

娘は、七軒のうちのどこかの家の客として呼ばれています。娘は、提灯を持って、ちよこちよこ走って行ったり来たりしています。それは、仕事のための練習です。人々は、娘が一人前になったらどうなるのかと思っています。とにかく、立派な舞台に出るつもりでいるのは面白いのです。垢抜けた三十歳過ぎの女がいます。小ぎれいな縞模様の着物を着て、紺色の足袋をはき、雪駄をちゃらちゃら鳴らしています。女は、忙しそうです。女は、小包を横抱きにして

います。茶屋が棧橋に合図を送っています。茶屋は、料理をここから運びます。このあたりでは、茶屋のことを立って屋と呼びます。このあたりの習慣は、他の場所と違います。女は、帯をきちんと結んでいる人が少ないです。女は、袴がって幅の広い帯をしています。年上の女ならまだいいですが、十五、六歳の小娘がホオズギをふくらませても、こんな格好です。人によっては、見たくないと思う人もいるでしょう。この場所柄、仕方がありません。昨日、私は川岸のお店で「何紫」という源氏名を聞きました。今日、地回りの吉という男が、慣れていない焼き鳥屋を始めたようです。地回りの吉は、財産を使い果たしたら、また元の場所に戻って、女房のような格好をするつもりです。地回りの吉は、その時は、どんな素人よりも、女房らしく見えるでしょう。このあたりに住んでいる子供はみんな、こんなふうになっています。

秋の九月の仁和賀の頃の大きな道を見てほしいです。子供たちは、露八の物真似をして、栄喜の作ったものを着ています。子供たちの成長は、孟子の母も驚くほど早いです。「上手だ」と褒められて、今日も張り切っています。生意気な子供は、七つか八つくらいから威張っています。生意気な子供は、そのうち、肩で風を切って、鼻歌を歌うようになります。十五歳の少年は、大人びていて、学校の音楽に「ぎっちょんちょん」とリズムを取ります。運動会で木遣り音頭を踊るかもしれません。それでも、教育は難しいです。先生は苦勞しているでしょう。入谷の近くに育英舎という学校があります。育英舎は私立ですが、生徒は千人近くいます。狭い校舎に生徒がたくさんいます。先生は人望があります。このあたりでは、「学校」といえば育英舎のことです。学校に通っている子供は様々です。子供たちの父親は、火消しや鳶職人などです。ある子供は、父親が刎橋の番屋にいることを知っています。子供たちは、梯子登りの真似をして忍び返しを壊してしまいます。子供

は、遊びに行くように見えるのもおもしろいです。大籬の下には、娘がいます。娘は、しんぞと呼ばれています。娘は、七軒のうちのどこかの店の客に呼ばれています。娘は、ちよちんを持って、ちよこちよこ走って行ったり来たりしています。それは、仕事のための練習です。人々は、娘が一人前になったらどうなるのかと思っています。とにかく、りっぱなぶたいに出るつもりでいるのはおもしろいです。

あかぬけた三十才過ぎの女がいます。小ぎれいなしまもようの着物を着て、こん色のたびをはき、せったをちゃらちゃらならしています。女は、いそがしそうです。女は、小づつみを横だきにしています。茶屋がさん橋にあいずを送っています。茶屋は、料理をここから運びます。このあたりでは、茶屋のことをしたて屋とよびます。このあたりの習慣は、ほかの場所と違います。女は、帯をきちんと結んでいる人が少ないです。女は、いきがってはばの広い帯をしています。年上の女ならまだいいですが、十五、六才のお店で「何紫」という源氏名を聞きました。今日、地回りの吉という男が、慣れていない焼き鳥屋を始めたようです。この場所が、しかたがありません。昨日、私はかしのお店で「なにむらさき」というげんじなを聞きました。今日、地回りの吉という男と、慣れていないやきとり屋を始めたようです。なにむらさきは、お金を使い果たしたら、また元の場所に戻って、おかみさんのようなかっこうをするつもりです。その時は、おかしなしくよりも、おかみさんらしく見えるでしょう。このあたりに住んでいる子どもはみんな、まねをしています。

秋の九月の仁和賀のころの大きな道を見てほしいです。子どもたちは、露八のものまねをして、栄喜の作ったものを着ています。子どもたちの成長は、孟子の母もどろくほど早いです。「じょうずだ」とほめられて、今日もはりきっています。なまいきな子どもは、七つか八つくらいからいばっています。なまいきな子どもは、そのうち、かたで風を切って、鼻歌を歌うようになります。十五才の少年は、大人びていて、学校の音楽に「ぎっちょんちょん」とリズムを取ります。運動会で木やりおんどをおどるかもしれません。それでも、教育はむずかしいです。先生は苦勞しているでしょう。入谷の近くに育英舎という学校があります。育英舎は私立ですが、生徒は千人近くいます。せまいこうしゃに生徒がたくさんいます。先生はじんぼう

<p>私たちは、文句を言われても言い訳をします。よくしゃべる子供もいるでしょう。ある子供は、父親が馬の世話をする人だからかわれて、恥ずかしそうにしています。</p> <p>お金を貸すところの秘密の子が、寮に住んでいます。秘密の子は、貴族の息子のようなふりをして、房のついた帽子をかぶって、洋服を着て、きれいな場所を歩いています。坊ちゃん坊ちゃんと呼ばれて、子供たちがついていきます。</p> <p>龍華寺の信如という少年がいます。信如は、たくさんの筋の入ったと称される黒髪をしています。信如は、十五歳で普通の身長です。信如の栗のいがのような色の髪も、やがて、黒く染めることになるでしょう。信如は、袖の色も変わるでしょう。人々は、信如は、本当に仏の道に入る決意をしたのだろうかと思ひます。信如は、親からもらった勉強好きですが、友達は、信如がおとなしいのを心配しています。友達は、信如に様々ないたづらをするので、ある時、猫の死体をロープにくくりつけて、信如に投げつけました。友達は、お役目だからお経をあげて供養するよう頼むと言いました。</p> <p>それは昔の話です。今は、信如は学校で一番の人です。信如をばかにする者はいません。信如は十五歳で、普通の身長です。信如は、栗のいが色の髪をしています。信如の髪は、意外と普通ではありません。信如は、藤本信如と名乗っていて、どこかしらお坊さんのようです。</p>	<p>があります。このあたりでは、「学校」といえば育英舎のことです。学校に通っている子どもは様々です。子どもたちの父親は、火消しやとび職人などです。ある子どもは、父親が刳橋のぼん屋にいることを知っています。子どもたちは、はしごのぼりのまねをしてしのび返しをこわしてしまいます。子どもたちは、文句を言われても言いわけをします。よくしゃべる子どももいるでしょう。ある子どもは、父親が馬のせわをする人だからかわれて、はずかしそうにしています。</p> <p>お金をかすところのひみつの子が、りょうに住んでいます。ひみつの子は、きぞくの息子のようなふりをして、ふさのついたぼうしをかぶって、洋服を着て、きれいな場所を歩いています。ぼっちゃんぼっちゃんと呼ばれて、子どもたちがついていきます。</p> <p>龍華寺の信如という少年がいます。信如は、たくさんのすじが入っているとされる黒かみをしています。信如は、十五歳で普通の身長です。信如のくりのとげのような色のかみも、やがて、黒くそめることになるでしょう。信如は、そでの色も変わるでしょう。人々は、信如は、本当にほとけの道に入る決意をしたのだろうかと思ひます。信如は、親ににて勉強が好きですが、友達は、信如がおとなしいのを心配しています。友達は、信如にさまざまないたづらをするので、ある時、ねこの死体をロープにくくりつけて、信如に投げつけました。友達は、おやくめだからおきょうをあげてくようするよう頼むと言いました。</p> <p>それは昔の話です。今は、信如は学校で一番の人です。信如をばかにする者はいません。信如は十五歳で、普通の身長です。信如は、くりのいが色のかみをしています。信如のかみは、意外と普通ではありません。信如は、藤本信如と名乗っていて、どこかしらおぼうさんのようです。</p>
--	---

4.4.2 量的評価 (リーダビリティ・スコアの測定)

生成された教材の量的評価は、日本語文章難易度判別システム (<http://jreadability.net/>) を用いて行った。リーダビリティ・スコアや語彙レベル構成を測定することで、教材の難易度が学習者に適しているかどうかを数値的に評価した。リーダビリティ・スコアは高いほど文章が易しい。

表 3 リーダビリティ・スコア (概要) の対照

	LLM	日本語教師
文章難易度	中級前半 ふつう	中級前半 ふつう
リーダビリティ・スコア	4.09	4.13
総文数	103	103
総形態素数 (延)	1799	1816

総形態素数 (異)	485	501
総文字数	2671	2813
一文の平均語数	17.47	17.63

表 4 語彙レベル構成の対照

	LLM	日本語教師
初級前半	230	241
初級後半	160	169
中級前半	136	143
中級後半	78	75
上級前半	32	29
上級後半	9	7

日本語教師に対しては、一から全て人間がやり直したのでは AI との協働の意義が失われるため、Gemini 1.5 Pro が生成した文章を基に改善するよう依頼した。その前提での結果の分析と考察であることを勘案しなければならない。

まず、リーダビリティ・スコアに関しては、Gemini 1.5 Pro による自動変換が 4.09、日本語教師による手動調整後のスコアが 4.13 という非常に近い値を示した。このことから、Gemini 1.5 Pro による自動変換は、学習者にとって読みやすい文章を生成する上で十分に有効であることが示唆される。総文数や一文の平均語数といった指標もほぼ同一であり、全体として文章の構造や複雑さに大きな差異は見られなかった。総形態素数 (延) や総形態素数 (異) の値も類似しており、語彙の使用頻度や多様性についても両者が一致していることがわかる。これらの結果は、AI が生成する文章が、人間の手による修正を経て、教育的に適切な形で提供できる可能性を示している。

総文数や総形態素数、一文の平均語数といった指標が類似していることは、Gemini 1.5 Pro が生成する文章が、形式的には人間の手による調整後の文章とほぼ同等の品質を持つことを示している。これにより、AI が生成する初期文章の質が高く、効率的な教材生成プロセスが実現できることが確認された。

ただし、AI が生成した文章を基にすることで効率化を図りながらも、手動修正には限界があることも認識される。特に「たけくらべ」のような高難易度の近代文学作品を完璧に多読教材化するには、AI の力だけでなく、専門家の意見をさらに取り入れて修正していく必要があると考えられる。

4.4.3 質的評価 (日本語教育専門家のフィードバック)

表 1 の IU 分割時に、主語を文脈から推定するように指示を出していたが、概ね正確であったものの、いくつかの誤りが見られた。今回は実験記録として残しているが、実用段階では日本語専門教員が気づいた時点で修正を加えていくべきである。そうしないと、IU での誤りが文章再構築時まで引き継がれてしまう問題があることが判明した。人間側も必ずしも一人で全ての誤りに気づけるわけではないため、ダブルチェックを行ったり、プレビュー公開や Wikipedia 方式での公開が望ましいと考えられる。

また、実際の運用ではふりがなが使われることを明記しておく。ふりがなは、学習者が難しい漢字や語彙を理解する助けとなり、読みやすさを向上させるために有効である。ふりがなの使用は、特に初級および中級レベルの学習者にとって重要である。

なお、本研究の実践研究では、あえて抽出した全ての IU を使用したが、実際の運用に際しては、LLM の要約機能を使用して作品を凝縮した多読教材の作成も必要であろう。その場合も、どの部分を取捨選択したかを可視化する上で、IU という謂わばブロックにしておくことは有効であると考えられる。さらに、厳密な多読教材作成という観点からは、もっと厳格な語彙・文法・文型の管理と、教材の段階性を考慮すべきである。しかし、青空文庫のような文学作品を多く収録する題材に対しては、JLPT のある程度のレベル別の作り分けが可能であるとしても、JLPT 自体が現状、レベルによる語彙・文法・文型の指定などを行っていないことを鑑みても、AI と人間の協働による、若干緩めのレギュレーションによる新しい多読教材の形を模索していくことにも意味があると考えられる。

4.4.4 総合評価

量的評価と質的評価の結果を総合すると、以下のような結論が得られる。

量的評価の結果:

リーダビリティ・スコアおよび関連指標は、Gemini 1.5 Pro による自動変換と日本語教師による手動調整後の文章が非常に近い値を示しており、AI が生成する文章の質が高いことを示している。総文数や一文の平均語数などの指標もほぼ同一であり、文章の構造や複雑さに大きな差異は見られなかった。これにより、AI が生成する初期文章の質が高く、効率的な教材生成プロセスが実現できることが確認された。

質的評価の結果:

日本語教育専門家のフィードバックにより、生成された教材が JLPT N4 レベルの学習者に適していることが確認された。内容の妥当性、文化的背景の理解、学習者の理解促進の観点からも、AI による自動変換が有効であることが示された。しかし、特に複雑な文や文化的背景を含む部分については、専門家の手動調整が必要であることも確認された。

総合評価:

AI による自動変換と日本語教育専門家による手動調整の協働により、効率的かつ教育的価値の高い多読教材の生成が可能であることが示された。Gemini 1.5 Pro による自動変換は、JLPT N4 レベルの学習者に適した文章を生成する上で十分に有効であり、専門家の手動調整を加えることで、さらに高品質な教材が完成する。このことから、AI と人間の協働が最適な多読教材の生成に不可欠であることが明らかになった。特に難易度の高い作品を教材化する場合には、AI の自動変換だけでは不十分であり、専門家の意見を取り入れることが重要である。また、実際の運用ではふりがなの使用や、LLM の要約機能を活用して作品を凝縮することも必要であろう。さらに、青空文庫のような文学作品を対象とする場合、JLPT レベルに基づいた厳密なレギュレーションを適用するのではなく、若干緩めのレギュレーションによる新しい多読教材の形を模索することにも意義があると考えられる。

5. おわりに

本研究は、生成 AI 型大規模言語モデルとアイデアユニットを活用した日本語教育のための多読教材自動生成システムの開発を目的とし、特に樋口一葉の「たけくらべ」を

題材に実践的な研究を行った。Gemini 1.5 Pro による自動変換と日本語教育専門家による手動調整の協働により、効率的かつ教育的価値の高い多読教材の生成が可能であることが示された。リーダビリティ・スコアや質的評価の結果からも、AI と人間の協働が最適な多読教材の生成に不可欠であることが明らかになった。

しかし、本研究にはいくつかの限界が存在する。まず、IU の分割における主語の推定や文脈の理解において AI が誤りを犯す場合があり、それが最終的な教材の質に影響を及ぼすことが確認された。また、ふりがなの使用や教材の段階性に関する管理についても、今後の改良が求められる。特に、難易度の高い文学作品を教材化する場合には、専門家の意見をさらに取り入れる必要がある。

今後の研究では、AI の精度向上に加え、さらに高度なフィードバックシステムやダブルチェック機能を導入することが重要である。また、JLPT レベルに基づいた厳格なレギュレーションを適用する一方で、柔軟な規制による新しい多読教材の形を模索することも意義があると考えられる。これにより、日本語学習者がより効果的に学習できる教材の提供が可能となり、教育現場における実践的な応用が期待される。

謝辞

私をこの興味の尽きない世界へといざなってくくださった全ての方々に深く御礼申し上げます。

参考文献

- [1] Devlin J., Chang M. W., Lee K., Toutanova K., "BERT: Pre-training of deep bidirectional transformers for language understanding", Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.1 (Long and Short Papers), pp.4171-4186 (2019).
- [2] Zhu J., Chen J., Chou K., Gao S., Zhang Y., "A survey of AI-assisted language learning", arXiv preprint arXiv:2106.12534 (2021).
- [3] Waring R., McLean S., "Exploration of the core and variable dimensions of extensive reading research and pedagogy", Reading in a Foreign Language, Vol.27, No.1, pp.160-167 (2015).
- [4] 樋口一葉, "たけくらべ" 青空文庫, https://www.aozora.gr.jp/cards/000064/files/389_15297.html (アクセス日: 2024/6/12), 本文献の底本: 「樋口一葉全集」第 2 巻、筑摩書房、1993(平成 5)年 5 月 25 日初版発行。入力: 高橋誠、校正: 大久保ゆう、2004 年 12 月 7 日公開、2011 年 12 月 19 日修正。
- [5] Day R. R., Bamford J., "Extensive reading in the second language classroom", Cambridge University Press (1998).
- [6] Yamashita J., "Extensive reading and development of different aspects of L2 proficiency", System, Vol.36, No.4, pp.661-672 (2008).
- [7] Genest P. E., Lapalme G., "Framework for abstractive summarization using text-to-text generation", Proceedings of the Workshop on Monolingual Text-To-Text Generation, pp.64-73 (2011).
- [8] Chun D. M., Plass J. L., "Effects of Multimedia Annotations on Vocabulary Acquisition", Language Learning & Technology, Vol.1, No.1, pp.60-81 (1996).
- [9] Paivio A., "Mental Representations: A Dual Coding Approach", Oxford University Press (1986).
- [10] Mayer R. E., "Multimedia Learning (2nd ed.)", Cambridge University Press (2009).
- [11] 甘利実乃, "自律的多読学習のための AI 支援アイデアユニット近似評価法—理解度の自己モニタリングと効果的フィードバックの提供—", 2024 年日本文学国際シンポジウム並びに西行学会台湾特別大会予稿集, pp.68-72 (2024).