

在 CEFR 的 Can-do 指標下應用生成 AI 來編寫日語教材之 研究：以 B1 等級為主的事前轉換手法之驗證

甘利實乃

東京外國語大學大學院綜合國際學研究科博士後期課程

摘要

本研究聚焦於 CEFR 框架下之 Can-do 陳述，探討其在日語教學領域與生成 AI 技術結合之可行性。研究中，先對抽象度偏高之 Can-do 陳述進行「事前轉換」，包含解析教學任務、指定難度層級、限制關鍵詞彙與文法結構，然後將其輸入 GPT-4o 模型以產出教學用文本。評估結果顯示，經事前轉換之 Can-do 陳述所對應之生成文本，整體可讀性與文法一致性皆更貼近 CEFR B1 層級，教師後續修訂量亦相對降低。然而，本研究侷限於機器指標（可讀性與形態素分析）之驗證，尚未納入學習者實際理解度與學習動機等因素。未來可透過課室實證與人為評量，進一步確認此手法對學習成效之實際貢獻。同時，考量不同推論模型（如 o1、o3）各自之特性，於多種教學情境下比較模型與轉換策略效益，亦為推動 AI 應用於日語教育的重要方向。本研究雖以 B1 層級為例，然此 Can-do 導向生成策略在 A2、B2 乃至更高層級之教材開發上亦具潛在擴張空間。

關鍵詞：CEFR、Can-do 陳述、生成 AI、教材自動生成、提示設計

受理日期：2025 年 03 月 10 日

通過日期：2025 年 06 月 06 日

DOI：10.29758/TWRYJYSB.202506_(44).0001

Generative AI-Based Design of Japanese Language Teaching Materials Using CEFR Can-Do Statements: An Empirical Evaluation at the B1 Level

Amari, Mino

Doctoral Program, Graduate School of Global Studies, Tokyo University
of Foreign Studies

Abstract

This study investigates how generative AI can be leveraged to convert CEFR-based Can-do statements into Japanese language learning materials. Through a “preprocessing” approach that refines these statements with explicit lexical and grammatical constraints, we employed GPT-4o to generate sample texts at the B1 level. Machine-based readability scores and morphological analyses confirm that structured prompts yield texts more closely aligned with the intended proficiency range, promoting coherence and vocabulary suitability. While this pilot study relied on automated metrics, future research must incorporate learner feedback and classroom implementation. The findings suggest that systematic prompt design can bridge abstract Can-do statements and precise language outputs, thus offering educators an efficient means of producing adaptable and level-appropriate teaching resources.

Keywords : CEFR, Can-do Statements, Generative AI, Automated Teaching Material Generation, Prompt Design

CEFR の Can-do ステートメントを活用した生成 AI による 日本語教材設計

—B1 レベルを例とする事前変換手法の効果検証—

甘利実乃

東京外国語大学大学院総合国際学研究科博士後期課程

要旨

CEFR の Can-do ステートメントを生成 AI に活用する方法として、本研究では抽象度の高い Can-do を事前に要素分解し、場面設定や語彙制限を加える手法（事前変換）を提案した。実験では日本語教育用 LLM として GPT-4o を使い、CEFR B1 レベルを例に抽出した Can-do を対照群（素の Can-do）と実験群（事前変換）で比較した。その結果、可読性指標（jReadability）や形態素解析により、事前変換群が文章難易度の安定や文末統一度の向上を示すことが統計的に示唆された。機械評価のみの検証であるため、最終的には学習者評価や教室実践との突合が必須であるが、教師の専門知識を LLM に適切に伝達する枠組みとして、Can-do の具体化による教材自動生成の可能性が示唆された。また、この手法は他のレベルや多様な場面に適用できる余地も大きく、今後はモデル別の出力特性と結びつけた精緻化研究が求められる。本研究はあくまで B1 レベルを例としたが、A2 や B2、さらには C レベルへの拡張や会話型タスクへの応用も検討すべき課題である。今後の研究発展も大いに期待される。本研究は AI 活用へ向けた一段階となり得る。

キーワード：CEFR、Can-do ステートメント、生成 AI、教材自動生成、プロンプト設計

CEFR の Can-do ステートメントを活用した生成 AI による 日本語教材設計

—B1 レベルを例とする事前変換手法の効果検証—

甘利実乃

東京外国語大学大学院総合国際学研究所博士後期課程

1. はじめに

本研究は、欧州言語共通参照枠（CEFR）の B1 レベルに示される Can-do ステートメントを用い、生成 AI による日本語教材作成の可能性と課題を検証するものである。近年、大規模言語モデル（Large Language Models: LLM）が飛躍的な性能向上を遂げ、様々な分野での応用が模索される中、教育領域においても教材自動生成や学習支援への活用が注目されている。日本語教育では、これまで教師が時間と専門知識を要して行ってきた教材の難易度調整や語彙制限を、LLM の自動生成によって効率化することが期待される。しかし、抽象度の高い Can-do ステートメントをそのまま入力すると、学習者のレベルと乖離した文章や不統一な文体を含むテキストが生成される可能性がある。

CEFR B1 レベルは、日常的な話題であればある程度自立してコミュニケーションを行える一方、文章や文体が複雑になると理解が追いつかない学習者も多い段階とされる（Council of Europe, 2001；独立行政法人国際交流基金, 2010）。したがって、学習者に無理のない語彙・文法範囲の教材を提供するためには、Can-do ステートメントから教師が必要要素を細分化し、使用語彙や文体、場面設定を具体的に提示することが効果的と考えられる。本研究では、この作業を「事前変換」と呼び、事前変換を行ったプロンプト（実験群）と抽象的 Can-do のみのプロンプト（対照群）で LLM が生成する文章を比較する。

具体的には、「旅行と交通」「自由時間と娯楽」「人との関係」とい

う三つの場面を対象に選定したうえで、対照群・実験群それぞれ 5 回ずつ生成 AI を呼び出し、計 30 本文を収集する。その際、文末が敬体か常体かを判定し、混在度を数値化する方法や接続詞の過剰連結を測定する方法等を形態素解析によって導入し、jReadability スコア（Hasebe & Lee, 2015）をあわせて参照することで、文章難易度を客観的に把握する。本研究は、CEFR B1 レベルで想定される表現範囲をおおまかな基準とし、教師の修正負担（文体の不統一や難度の偏りなど）がどの程度減らせるかを検討する。

本論文の構成は以下のとおりである。第 2 章では、CEFR B1 レベルの学習者像と Can-do ステートメント活用の意義を整理し、三つの場面を取り上げる背景を示す。第 3 章では、生成 AI の技術的基盤となる Transformer モデル（Vaswani et al., 2017）と、言語教育における可読性評価や形態素解析の先行研究を概観する。第 4 章にて事前変換手法と実験デザインを提示し、第 5 章で可読性指標や形態素解析指標をもとに得られた結果を示す。第 6 章では、事前変換が学習者レベルに適した教材生成をどの程度促すかを考察し、併せて日本語教育への応用の可能性と限界を議論する。最後の第 7 章で結論と今後の課題をまとめる。

以上を通じて、本研究は、抽象的な Can-do をどのように具体化・細分化すれば CEFR B1 相当の学習者に適切な文章生成を誘導できるかを明らかにする。教師の作業負担を削減しながら、質の高い教材作成を支援する仕組みとして、事前変換プロンプトの有用性と今後の課題を提示することが本研究の目的である。

2. CEFR B1 レベルと日本語教育における Can-do ステートメント

2.1 CEFR B1 レベルの概要

欧州言語共通参照枠（CEFR）は、学習者の言語運用能力を「実際に何ができるか」という行動面から示す枠組みとして広く知られている（Council of Europe, 2001）。A1～C2 まで 6 段階に区分されたレベルのうち、B1 は初級から中級へ移行する段階であり、日常的なや

り取りや身近な話題であればある程度自立してコミュニケーションを行えるとされる。一方、文体が複雑化したり、敬意表現の使い分けが求められたりすると、理解や表現に困難をきたしやすい中間レベルでもある。こうした B1 学習者の特徴は、日本語教育においても概ね同様に考えられ（独立行政法人国際交流基金, 2010）、具体的な文法・語彙の枠組みを踏まえつつ、ある程度実践的な場面でのタスクを想定しながら教材設計を行うことが重要視されている。

2.2 Can-do ステートメントの意義と抽象度の課題

CEFR が提示する Can-do ステートメントは、学習者がどのような場面・行為をこなせるかを明確に示し、学習目的を行動ベースで示す点に強みがある（Council of Europe, 2001）。日本語教育に導入されている JF 日本語教育スタンダード（独立行政法人国際交流基金, 2010）でも、同様の考え方を踏襲し、学習者が「旅行先で要点を理解できる」「趣味や娯楽についてある程度詳しく語れる」といった能力を指標化している。ただし、Can-do ステートメントは実社会のタスクを網羅的・横断的に記述しているため、学習現場でそのまま使用すると、語彙レベルや文体などの調整が不足し、教材としての難易度にばらつきが生じやすいという課題がある。

こうした抽象度の高い記述を有効に活用するには、教師が場面・語彙・文法などの要素をあらかじめ細分化する必要がある。たとえば塩澤真季・石司えり・島田徳子（2010）が提案するように、〔条件〕〔話題・場面〕〔対象〕〔行動〕などの観点で Can-do を分解すれば、学習者に合ったタスク設定を行いやすくなる。本研究では、教師が要素分解や語彙制限を行う工程を「事前変換」と位置づけ、大規模言語モデルへの入力（プロンプト）の具体化を試みる。

2.3 三つの場面（旅行と交通、自由時間と娯楽、人との関係）の選定理由

本研究では、B1 レベルにおける Can-do ステートメントの中から、以下の三場面を取り上げることとした。

① 旅行と交通

観光や移動手段のやり取りは、初級から中級にかけて頻度が高いトピックであり、道案内や体験談の理解など、多様な言語機能を含む。日本語の教科書でも取り上げられることが多く、B1 学習者にとって比較的身近かつ実用度の高い場面といえる。

② 自由時間と娯楽

自分の趣味や好きな作品について話す場面は、学習者同士で話題にしやすく、語彙制限や文法統制を行いながら、比較的長い発話を引き出せる可能性がある。映画や音楽、スポーツなど興味を持ちやすいテーマが多いため、Can-do ステートメントによって具体的なタスク（好きな映画の理由や魅力を説明するなど）を設定しやすい。

③ 人との関係

友人や知人とのやり取り、近況報告、自己紹介の拡張などは、初級段階から継続的に扱われる内容であり、B1 レベルでも尚一層の発話力が求められる部分である。普段の生活や人間関係に即した話題は指導上の汎用性が高く、日本語学習者にとっても負荷が適切だと考えられる。

いずれの場面も教室実践で扱いやすく、学習者の動機づけに資する話題である。さらに、Can-do を要素分解しやすい点もあり、大規模言語モデルへ事前変換プロンプトを与えた場合と、抽象的な Can-do だけを与えた場合の差異を明確に検証できると期待される。

3. 生成 AI と日本語可読性評価の先行研究

3.1 Transformer モデルと大規模言語モデル

Transformer アーキテクチャは、Vaswani et al. (2017) によって提案され、自己注意機構（Self-Attention）を用いて入力文脈を並列的かつ多層的に学習する点に特徴がある。従来の RNN 系モデルでは系列長が増すと学習が困難になる問題があったが、Transformer は文脈依存を効率よく捉えられるため、翻訳や要約、対話など多様なタ

スクで高い性能を示すようになった。これを大規模化した大規模言語モデル（Large Language Models: LLM）は、膨大なテキストコーパスを事前学習し、Few-Shot 学習や Zero-Shot 学習にも適応できる柔軟性を備える（Brown et al., 2020）。

教育分野においても、こうした LLM は教師の指示（プロンプト）に合わせて多彩な文章や会話例を生成できることが注目されている。しかし、教師が意図する難易度や文体を満たす文章が常に得られるわけではなく、抽象的な指示だけでは学習者水準を超えた文章や文体的混乱を生む恐れがある。本研究は、LLM に対して「事前変換プロンプト」を与えることで、学習者の到達度や場面情報を明示的に伝え、文章を安定化させる方法を検討する。

3.2 日本語可読性評価と機械的解析

3.2.1 jReadability

Hasebe & Lee (2015) が提案する jReadability は、日本語教科書と BCCWJ を基に、平均文長や語種（漢語率・和語率など）を変数とする回帰分析によってテキストを六段階（初級前半～上級後半）に振り分ける仕組みである。ただし、これらのレベルは CEFR とは直接対応しておらず、CEFR B1（中級相当）との結びつけは示唆されていない。

一方、本研究では、B1 レベルの文章を 3.5～4.0 付近と見なす仮定をおき、対照群・実験群で生成されたテキストを jReadability スコアで比較する。これはあくまで本研究の独自設定であり、Hasebe & Lee (2015) 自身が CEFR B1 を 3.5～4.0 帯と規定しているわけではない。最終的には、形態素解析による文末敬体率や接続詞連続率など複数の指標ともあわせ、文章難易度を総合的に評価する予定である。

3.2.2 形態素解析と文法的特徴の定量化

機械評価をより細かく行うには、形態素解析による文法的特徴の抽出が不可欠である。日本語は文末形態や接続表現が学習者の理解を左右しやすく、文末敬体／常体の混在度や接続詞の過剰連結度等を数値化することで、文体的一貫性や冗長表現を可視化できる。

本研究では、対照群と実験群それぞれの生成テキストに MeCab を適用し、敬体・常体の混在率と接続詞連続率等を算出する。これらの指標は、既存のリーダビリティ研究（Hasebe & Lee, 2015）で直接扱われているわけではなく、本研究が独自に導入した評価法である。学習者レベルに適した教材を提供するためには、文章が過度に複雑化していないかを複数の観点で検証する必要があると考える。

3.3 事前変換プロンプトの意義

大規模言語モデルは、抽象的指示でも文章生成を行える柔軟性を持つ反面、文体や難易度を精密に制御するには具体的な制約条件が必要となる。教師が Can-do ステートメントを「事前変換」し、場面情報や語彙範囲・敬体表現などを明示すれば、モデルが適切な文法・語彙を用いる確率が高まる。

一方で、過度に詳細な条件を付与すると文章が平板化する可能性も指摘される。そのバランスを探るために、本研究では抽象的な Can-do だけを入力した対照群と、教師が細分化した実験群とで LLM 出力を比較し、jReadability と形態素解析によって両群の文章特徴を定量的に示す。次章では、具体的に事前変換プロンプトをどのように作成し、実験デザインをどう構築したかを述べる。

4. 研究方法：事前変換手法と機械評価の設計

4.1 研究仮説と全体的アプローチ

本研究では、CEFR B1 レベルの Can-do ステートメントを教師が事前変換して具体化したプロンプト（実験群）を、抽象的 Can-do のみのプロンプト（対照群）と比較し、生成 AI（LLM）が出力する日本語教材の難易度や文体統一度にどのような差が生じるかを検証する。主な仮説として、以下の点を想定する。

① 文章難易度（可読性）の安定化

事前変換プロンプトを与えることで、B1 レベル向けの適切な語彙・文法範囲が維持され、可読性スコア（jReadability）が対照群より安定する。

② 文体統一度合いの向上

事前変換により、敬体・常体など文末表現が統一されやすくなり、また接続詞の過剰使用を抑制できると考えられる。

4.2 事前変換の手順

4.2.1 Can-do ステートメントの細分化

抽象度の高い B1 レベルの Can-do (例:「旅行体験談を理解できる」「趣味の魅力をある程度詳しく語れる」など) を、以下の観点で要素分解する。

① 語彙レベル

B1 学習者が理解しやすい単語を中心にリストアップし、難読漢字や専門用語を避ける。

② 文体制御

文末は敬体（です／ます）を基本とし、丁寧度合いを明示する。助詞や接続詞の使い方も簡潔なものを優先させる。

③ 場面情報

「旅行と交通」「自由時間と娯楽」「人との関係」など具体的シチュエーションを設定し、必要最低限の情報を盛り込む。

4.2.2 対照群との比較

① 対照群

抽象的な Can-do 文言をそのまま提示（例:「CEFR B1 レベルで以下の Can-do を元に 500～700 字で書いてください」程度）。

② 実験群

上記の要素分解に基づき、語彙・文体・長さ制限を明記したプロンプト（事前変換）を提示。

4.3 実験デザイン

4.3.1 三場面・二条件での生成

第 2 章で述べた三場面（旅行と交通、自由時間と娯楽、人との関係）を対象とし、対照群と実験群それぞれ 5 回ずつ大規模言語モデルを呼び出して文章を生成する。合計 30 本文を取得し、その難易度や文法的特徴を比較する。呼び出しごとの確率的揺らぎを含めて、

教師の意図を反映しやすい設定かどうかを観察する。

4.3.2 生成 AI のパラメータ

本実験では、OpenAI 社の GPT-4o（バージョン名: gpt-4o-2024-11-20）を用いる。これは 2025 年 2 月時点で利用可能であった大規模言語モデルであり、生成結果の安定性と柔軟性から本研究の実装条件に合致すると判断した。具体的な API 呼び出しパラメータとしては、以下の設定を採用する。

① temperature: 0.7

学習者向け教材である程度の多様性を保ちつつ、過度の乱数性を抑えるために設定している。

② max_tokens: 700

1 本文あたり約 500～700 字程度の日本語テキストを生成できるようにしつつ、長文になりすぎないように上限を設定する。

③ その他

top_p や n といったパラメータはデフォルト値のままとし、大幅な揺らぎを回避する。system メッセージでは「You are a helpful assistant.」という簡素な指定のみを行い、プロンプト（user メッセージ）との重複を避ける。

上記のようにパラメータを固定して再現性を確保しつつ、三場面×二条件（対照群・実験群）×生成 5 回の合計 30 本文を取得し、データセットとして分析を行う。呼び出しごとに微妙な確率的差異はあるが、複数回生成することで対照群・実験群それぞれの平均的傾向を比較可能にする狙いがある。

4.4 評価指標と分析手法

4.4.1 可読性スコア（jReadability）

本研究では、3.2.1 で述べたように jReadability (Hasebe & Lee, 2015) を用いて、対照群・実験群が生成した文章の難易度を相対的に数値化する。具体的には、B1 レベルの文章を 3.5～4.0 付近と見なしており、両群のスコア平均や分散がどのように分布するかを比較する。

ただし、jReadability の六段階は CEFR と直接対応していないため、

3.5～4.0 という範囲は本研究独自の目安である。最終的に、得られたスコアが B1 相当と推測される範囲内にどれほど分布するかを確認し、形態素解析（4.4.2 参照）による文法的統制度と合わせて検討する。

4.4.2 形態素解析に基づく文法的特徴

本研究では、形態素解析ツール MeCab を用いて、以下の 4 種類の文法的指標を抽出し、B1 学習者への適合性を評価する。なお、これらの指標は本研究が独自に設定したものであり、文末敬体や接続詞の連続使用といった要素を先行研究（Hasebe & Lee, 2015）が直接扱っているわけではない。

① 文末敬体混在率

各文末が敬体（～です／～ます）か常体（～だ／～である）かを判定し、両者が混在する度合いを算出する。事前変換プロンプトで敬体を統一するよう指示した場合、混在率の低下が期待される。

② 接続詞の過剰連結率

「そして」「しかし」「それで」などの接続詞が連続出現する割合、あるいは段落内に過度に多用される頻度を計測する。冗長表現の可視化によって、事前変換が文章の簡潔化に寄与したかを検証する狙いがある。

③ タイプ・トークン比（TTR）

語彙の多様度を示す指標である TTR を算出し、同じ単語の繰り返しや、難易度の高い語彙が極端に多く含まれていないかを把握する。B1 学習者の語彙範囲を逸脱していないかをチェックするうえで有用と考えられる。

④ 平均文長

1 文あたりの形態素数、あるいは文字数を計算し、長文化の度合いを測定する。文が過度に長いと読解負荷が上がる恐れがあり、B1 レベルの学習者には適度な文の長さが求められると推定される。

以上 4 種類の数値を対照群・実験群それぞれで集計し、統計的に比較する。たとえば文末混在率や接続詞過剰連結率が有意に低下し、TTR や平均文長が適度な範囲に収まる場合、事前変換によって B1 学習者に適した文体や難易度へ寄与したと評価できる可能性がある。一方で、指標同士のトレードオフ（文末は統一されたが、平均文長が極端に長文化する等）も想定されるため、複数観点から総合的に評価することが必要である。

5. 実験と結果

5.1 実験概要とサンプル生成データ

本研究では、第 4 章の実験デザインに基づき、JF575(旅行と交通)、JF555 (自由時間と娯楽)、 JF561 (人との関係) の 3 つの Can-do ステートメント（「JF575」などは JF 日本語教育スタンダードの Can-do の整番）に対し、対照群（素の Can-do に近い Can-do を元にしたプロンプト）と実験群（事前変換プロンプト）の 2 条件を設定し、大規模言語モデル（GPT-4o）を呼び出した。各 Can-do につき対照群・実験群それぞれ 5 回ずつ生成させ、合計 30 本文の日本語テキストを得た。

本節では、実際に用いたプロンプトと、生成された文章を 1 組ずつ例として示すことで（表 1）、両条件の差を把握しやすくする。なお、パラメータは第 4 章で述べたように `temperature=0.7`, `max_tokens=700`, `top_p`（デフォルト）などに統一している。生成 AI の性質上、全く同じ生成文を同じパラメータとプロンプトからであっても再現することは困難である。

表 1 各 Can-do に対するプロンプトおよび生成例

Can-do 整番 ／言語活動 ／カテゴリー／第一 ピック／ Can-do 本文	群	プロンプト	生成例 括弧内はリーダビリティ・スコア／判定 レベル／文字数／一文の平均語数
JF575／受容 ／講演やプ レゼンター	対照群	CEFR B1 レベルで以下の Can-do を元に 500 ～700 字で書いてください。 「【Can-do 本文】」	私の友人が最近旅行に行った話を聞きました。彼はとても楽しそうで、たくさんの興味深い話をしてくれました。友人の発

ションを聞く／旅行と交通／発音がはっきりしていれば、友人の旅行の体験談を聞いて、出来事や感想、気に入った場所や食べ物など、話の要点を理解することができる。			音ははっきりしていたので、彼の話をよく理解することができました。 友人は、まず旅行先で訪れた美しい場所について話してくれました。一番気に入ったのは、海が広がる静かなビーチだったそうです。砂浜は白く、海の水は青く澄んでいて、とてもリラックスできる場所だったと言っていました。特に夕日がとてもきれいで、写真をたくさん撮ったそうです。その話を聞いて、私もその場所に行ってみようと思いました。 次に、友人は旅行中に食べた料理について話してくれました。彼が一番好きだったのは、地元の新鮮な魚を使った料理でした。魚はとてもおいしく、スパイスが程よく効いていたそうです。また、デザートも特別だったと言っていました。甘すぎず、フルーツの味が生きていて、何度も食べたいような味だったそうです。 旅行中、友人は地元の人々ともたくさん話をしたそうです。みんな親切で、旅行者を歓迎してくれたと言っていました。地元の人々からおすすめの場所や食べ物を教えてもらい、それがとても役に立ったそうです。 友人の旅行の体験談を聞いて、私も旅行に行きたい気持ちが強くなりました。彼の話はとても楽しく、まるで自分がその場にいるような気分になりました。 (3.99／中級前半／567／20.17)
同上	実験群	以下の Can-do ステートメントを踏まえて、CEFR B1 レベルの学習者向けに 500～700 字の文章を作成してください。 「【Can-do 本文】」 指示 ① 語彙レベル 基本語彙を中心に：旅行に関する動詞（行く・歩く・見る など）、形容詞（楽しい・美しい・おいしい など）、名詞（町・景色・料理 など）を用いてください。 難読漢字・専門用語は避ける：「観光地」は「観光の場所」、「郷土料理」は「地元の料理」のように言い換えます。 新出語は《10 語以内》に抑え、初出時に（ ）の簡単な言い換えや説明を添えます。 ② 文体制御 終始敬体：「～です／ます」で統一し、やわらかい丁寧さを保ちます。 一文 20～25 字前後を目安に区切り、読点「、」の連続使用を避けます。 接続詞はシンプル：「まず」「次に」「それから」「だから」「最後に」を主に使い、複雑な逆接は使いません。 主語や代名詞の重複を減らす：同じ主語が続く場合は省略し、文を滑らかにします。 ③ 場面情報 シチュエーション：「旅行と交通」—友人が帰国後、カフェで旅行体験を語り、聞き手が内容を理解する場面。 盛り込む要素（最低限） 旅行先の国・都市名。 旅行中の出来事や感想を 2 点程度（例：きれいな海、親切な人）。 気に入った場所・食べ物を 1 つずつ具体的に。 話の要点をまとめる簡単なリアクション（例：「それは楽しそうですね」）。 入れない要素：詳細な旅行日程、費用、歴史的データなど専門情報。	先日、友人がフランス旅行の話をしてくれました。友人はパリに行ったそうで、とても楽しい時間を過ごしたと言っていました。発音がはっきりしていたので、話の内容をよく理解することができました。 まず、友人が一番感動したのはエッフェル塔だったそうです。夜に訪れたとき、ライトアップされたエッフェル塔がとても美しかったと言っていました。塔の上に登って、パリの夜景を見たそうです。街全体が光に包まれていて、まるで夢の中にいるようだったと話していました。写真も見せてもらいましたが、本当にすばらしい景色でした。 次に、友人はパリの街をたくさん歩いたそうです。セーヌ川の近くを散歩しながら、かわいいお店やカフェをいくつも見たと言っていました。その中の一つのカフェで、クロワッサンを食べたそうです。このクロワッサンは外がサクサクしていて、中はふんわりしていて、とてもおいしかったと言っていました。フランスのパンはどこで食べてもおいしいと感じたそうです。 それから、友人は地元の人たちが親切だったとも話していました。道に迷ったとき、近くにいた人が笑顔で助けてくれたそうです。また、お店の店員さんも気さくに話しかけてくれて、楽しい会話ができたと言っていました。 最後に、友人が一番気に入った場所はルーブル美術館です。有名な「モナ・リザ」を見たとき、とても感動したと言っていました。展示されている絵や彫刻はどれも美しく、たくさんの時間を過ごしたそうです。 友人の話聞いて、私もいつかパリに行きたいと思いました。エッフェル塔やルーブル美術館を見て、フランスのパンを食べてみたいです。「それは楽しそうですね」と心から思いました。

		上記①～③を守り、500～700 字で書いてください。文字数には句読点も含めます。	(4.3／中級前半／705／20.41)
JF555／産出／経験や物語を語る／自由時間と娯楽／自分の好きな映画などについて、好きな理由や魅力、印象に残ったシーンなどを、ある程度詳しく友人に語ることができる。	対照群	CEFR B1 レベルで以下の Can-do を元に 500～700 字で書いてください。 「【Can-do 本文】」	私が好きな映画のひとつは、「君の名は。」です。この映画は日本のアニメ映画で、新海誠監督が作った作品です。私はこの映画が大好きで、何度も見ました。ここでは、なぜこの映画が好きなのか、そしてその魅力や印象に残ったシーンについてお話しします。 まず、この映画のストーリーがとてもユニークで感動的です。田舎に住む女の子・三葉と、東京に住む男の子・瀧が、夢の中でお互いの体が入れ替わってしまうという不思議な物語です。最初はコメディのような雰囲気ですが、ストーリーが進むにつれて、二人がある大きな運命に巻き込まれることがわかります。この展開がとても面白くて、最後まで目が離せませんでした。 次に、この映画の魅力のひとつは、その美しい映像です。新海誠監督の作品は、いつも色彩がとてもきれいで、背景の描写が細かいのが特徴です。「君の名は。」も例外ではなく、東京の街の風景や田舎の自然がとてもリアルに描かれています。特に夜空や星のシーンは、本当に美しく印象的でした。 また、音楽もこの映画の大きな魅力です。RADWIMPS というバンドが作った音楽が、映画のストーリーにとってもよく合っています。特に「前前世」という曲は明るくて元気な気持ちにさせてくれるので、今でもよく聴いています。 印象に残ったシーンはたくさんありますが、一番好きなのは、二人が彗星が落下する夜にやっと会うシーンです。その瞬間はとても感動的で、映画を見ている私も涙が出そうになりました。二人が時間や場所を超えてお互いを探し続ける姿が、とても心に響きました。 この映画は、ストーリー、映像、音楽のすべてが素晴らしく、何度見ても新しい発見があります。友人にこの映画をおすすめしたいですし、また一緒に観たいと思います。
同上	実験群	以下の Can-do ステートメントを踏まえて、CEFR B1 レベルの学習者向けに 500～700 字の文章を作成してください。「君の名は。」を見て、特に主人公たちの心の変化が感じられる印象的な場面について、敬体で書き、難易度を抑えた語彙を用いること。 以下の Can-do ステートメントを踏まえて、CEFR B1 レベルの学習者向けに 500～700 字の文章を作成してください。 「【Can-do 本文】」 指示 ① 語彙レベル 基本語彙を中心に：日常生活でよく使う動詞（見る・感じる・思う など）、形容詞（おもしろい・美しい・すばらしい など）、名詞（物語・主人公・音楽 など）を選んでください。 難読漢字や専門用語は避ける：例えば「俳優」は「はいゆう」、「映像美」は「えいぞうび」と読みが分かりにくい語は使わず、「俳優さん」「画面の美しさ」のように置き換えます。 語彙数は絞る：新しい単語は《10 語以内》に収め、文中で自然に意味が分かるよう軽	私が最近見た映画で特に好きだったのは「天気の子」というアニメ映画です。この映画はファンタジーと恋愛の物語で、見ている間ずっと引き込まれました。今日は、この映画の魅力をお話したいと思います。 まず、この映画のストーリーがとてもおもしろかったです。高校生の帆高（ほだか）という少年が、東京にやって来て、不思議な力を持つ少女、陽菜（ひな）と出会う話です。陽菜には天気を晴れにする力があり、それを使って人々を幸せにします。しかし、その力には大きな代償があり、二人はその運命に立ち向かっていきます。この物語は、ただの恋愛映画ではなく、自然や人間関係についても考えさせられる深い内容でした。 それから、映画の画面の美しさにも感動しました。特に雨のシーンでは、水の表現がとてもリアルで、まるで本物の雨の中にいるように感じました。空や街の風景も細かく描かれていて、映画館の大きな画面で見ると、本当に美しいと思いました。 印象に残ったシーンの一つは、帆高が陽

		い説明を添えてください。 ② 文体制御 終始敬体：「～です／ます」で統一し、丁寧さを保ちます。 一文は 20～25 字程度を目安に切り、読点「、」を多用しすぎないようにします。 接続詞はシンプルに：「まず」「それから」「だから」「しかし」「最後に」を優先し、複雑な逆接（ところが・にもかかわらず など）は避けます。 代名詞の反復を減らす：同じ主語が続くときは省略して滑らかにします。 ③ 場面情報 シチュエーション：「自由時間と娯楽」—映画館で友人と鑑賞後、カフェで感想を語り合う場面。 必要最低限の要素 映画のタイトルとジャンル（例：SF アクション）。 好きな理由（ストーリー、音楽、俳優の演技などを1～2点）。 印象に残った具体的シーンを一つ描写。 友人へのおすすめひと言。 不要な情報は入れない：制作会社の詳細、興行収入、評論家の意見など専門的データは省きます。 上記①～③を守り、500～700 字で書いてください。文字数カウントには句読点も含めます。	菜を助けるために高いビルに登る場面です。空が晴れていく中、帆高が必死に陽菜に手を伸ばす姿に感動しました。このシーンでは、二人の強い絆や、困難に立ち向かう勇気がよく伝わってきました。また、音楽もその場面をさらに感動的にしていたと思います。 「天気の子」は、ストーリー、映像、音楽のすべてがすばらしい映画です。見終わった後、心が温かくなり、前向きな気持ちになりました。もしこの映画をまだ見ていないなら、ぜひおすすめします。あなたもきっと感動すると思います！ (4.16／中級前半／651／21.63)
JF561／産出／経験や物語を語る／人との関係／久しぶりに会った友人に、自分の近況などについて、まとものある話を語ることができる。	対照群	CEFR B1 レベルで以下の Can-do を元に 500～700 字で書いてください。 「【Can-do 本文】」	久しぶり！元気にしてた？会えて本当にうれしいよ。最近のことを少し話すね。私は去年、新しい仕事を始めたんだ。今までとはちょっと違う分野だけど、毎日新しいことを学べて楽しいよ。最初は慣れるのが大変だったけど、同僚が親切で助けてくれるから、だんだん自信がついてきたよ。 それから、趣味も増えたんだ。最近、料理にハマっていて、いろいろな国の料理を作るのが楽しいの。特にイタリア料理が好きで、パスタとかピザを自分で作るようになったよ。時間があったら、一緒に作ってみたいね。 あと、健康のために運動も始めたよ。週に 2 回、ジムに行ってるんだ。最初は筋肉痛がひどかったけど、今では体が少しずつ強くなっているのを感じるよ。ジムの友達もできて、運動がもっと楽しくなったよ。 そういえば、この前旅行にも行ったんだ。久しぶりに京都へ行って、お寺や神社をたくさん回ったよ。紅葉がすごくきれいで、写真を撮るの楽しかったんだ。今度見せるね！ 私の近況はこんな感じだけど、あなたは？最近何か面白いことがあった？またゆっくり話そうね！ (4.14／中級前半／455／13.14)
同上	実験群	以下の Can-do ステートメントを踏まえて、CEFR B1 レベルの学習者向けに 500～700 字の文章を作成してください。 「久しぶりに会った友人に、自分の近況などについて、まとものある話を語ることができる。」 指示 ① 語彙レベル 日常語を中心に：仕事・学校・家族・趣味・健康など、よく使う動詞（働く・学ぶ・楽しむ など）、形容詞（忙しい・元気な・新しい など）、名詞（会社・授業・趣味 など）を選びます。	久しぶりですね！会えて本当にうれしいです。最後に会ったのはいつでしたっけ？こんなふうにゆっくり話すのは久しぶりですね。今日は私の最近のようすをお話します。 まず、最近新しい仕事を始めました。前の会社で 3 年間働いていましたが、もっとやりがいのある仕事をしたいと思い、転職しました。今の会社ではチームで新しいプロジェクトに取り組んでいます。少し忙しいですが、毎日学ぶことがたくさんあって楽しいです。それに、新しい同僚（いっしょに働く人）とも仲良くなれて、とてもい

	難読漢字や専門用語を避ける：「近況報告」は「最近のようす」、「昇進」は「ポジションが上がる」のように言い換えます。 新語は 10 語以内：はじめて出す語は（ ）で簡単な説明を添えてください。 ② 文体制御 終始敬体：「～です／ます」で統一し、親しみと丁寧さを保ちます。 一文は 20～25 字前後を目安に区切り、読点「、」や複文の多用を避けます。 接続詞はシンプルに：「まず」「次に」「それから」「だから」「最後に」を優先し、複雑な逆接や重語は控えます。 主語の省略を活用：同じ主語が続くときはできるだけ省き、話を滑らかにします。 ③ 場面情報 シチュエーション：「人との関係」— 久しぶりに再会した友人とカフェで近況を語り合う場面。 最低限盛り込む要素 あいさつと再会の喜び（例：「久しぶりですね。会えてうれしいです。」）。 自分の近況を 3 点ほど（例：新しい仕事、最近始めた趣味、健康面）。 それぞれに短い感想（楽しい・大変・挑戦など）。 近い未来の計画や目標を 1 つ。 相手への一言質問や興味を示すリアクション（例：「あなたはどうですか。」）。 不要な要素：専門的な仕事の詳細、統計データ、硬い表現や敬語の多用。 上記①～③を守り、500～700 字で書いてください。文字数には句読点も含めます。	い環境だと思います。 次に、最近新しい趣味を始めました。ヨガです。最初は体が硬くて大変でしたが、少しずつ柔らかくなってきました。ヨガをすると気分がリフレッシュできて、ストレスが減る気がします。それに、健康にもいいので、これからも続けたいと思っています。 それから、最近健康に気をつけるようになりました。特に食べ物に注意しています。以前は外食ばかりしていましたが、今は家で野菜や魚を使った料理を作っています。まだ料理は得意ではありませんが、少しずつ上手になってきました。健康的な生活をする、体も軽くなった気がします。 最後に、近い未来の計画ですが、来月旅行に行く予定です。国内の温泉地（おんせんち）に行こうと思っています。温泉に入ってゆっくり休んで、リフレッシュしたいです。あなたも旅行が好きでしたよね？最近どこか行きましたか？ こうしてゆっくり話せるのはとてもいいですね。また近いうちに会って、もっとたくさん話しましょう！ (4.31／中級前半／647／14.59)
--	--	---

5.2 可読性指標（jReadability）の結果

5.2.1 スコア分布と分析対象

本研究では、第 4 章の実験デザインに基づき、JF575（旅行と交通）、JF555（自由時間と娯楽）、JF561（人との関係）の 3 つの Can-do ステートメントを対象に、対照群（抽象的 Can-do のみ）と実験群（事前変換プロンプト）でそれぞれ 5 回ずつ LLM（GPT-4o）を呼び出し、計 30 本文の日本語テキストを得た。対照群プロンプトは、ほぼ素の Can-do ステートメントを CEFR B1 レベル向けに書く程度の抽象指示にとどめ、実験群プロンプトは語彙・文体などの制限を明示的に施している。

各生成文章に対し、jReadability を用いて可読性スコアを算出し、CEFR B1 を 3.5～4.0 帯とみなす本研究の独自基準に当てはめて分析を行った。結果として、対照群の平均スコアは約 3.44（標準偏差 0.20）、実験群は約 3.59（標準偏差 0.18）であり、ともに 3.5～4.0 帯に近い

数値が得られた。ただし、両群の平均にはわずかに有意差($p<.05$)が認められ、事前変換群がやや B1 相当に近い文章を生成する傾向が確認できた。

5.2.2 GPT-4o の能力と差の縮小

当初の想定では、抽象的 Can-do のみを与えた対照群の文章は、文体・難易度などに大きなばらつきが生じると懸念していた。しかし、本実験の結果、GPT-4o が抽象的指示でもかなり良質な文章を生成することが判明し、対照群と実験群の絶対的な差は従来見込みより小さくなった。具体的には、対照群でも jReadability が概して 3.4 前後に集約し、B1 レベル下限を下回るほど難度が高くなる例は少なかった。

もっとも、事前変換群の方がスコアがわずかに高い（可読性がやや向上する）だけでなく、分散（標準偏差）も低いことが示されており、LLM の生成範囲を教師が明確に示すことで、より安定的に B1 水準へ近づける利点は依然として存在するといえる。

5.2.3 データの総合評価

jReadability 上の差分は大きくないものの、統計的検定では対照群 = 3.44、実験群 = 3.59 という平均スコアの違いに対して ($t(28)=2.11$, $p<.05$) の有意差が見られた。これは事前変換が B1 相当へ誘導するという仮説を支持するが、GPT-4o の性能向上により、対照群でも一定以上の品質を保っている現状を示唆している。

こうした状況は、教師が LLM にプロンプトを与える際、抽象度の高い Can-do でもそこそこの成果が得られることを意味する。ただし、後述する形態素解析の結果や文体面の統一度を総合すると、教師の意図（敬体か常体か、どの程度簡潔な文を求めるか）を正確に実現したい場合には、事前変換による制御が引き続き有用であると判断できる。

5.3 形態素解析による文法的特徴の結果

5.3.1 各生成文テキストの概要と分析方法

前節（5.2）では jReadability スコアを比較したが、本節では形態

素解析 (MeCab) を用いて、以下の四つの指標を抽出し、対照群 (抽象的 Can-do) と実験群 (事前変換プロンプト) の差を分析する。

① 文末敬体混在率

文末が敬体 (「です／ます」) か常体 (「だ／である」) かを判定し、同一の文章内で両者が混在する割合を%で示す。

② 接続詞過剰連結率

「そして」「それで」「しかし」「だから」など接続詞が連続使用される頻度を測定し、過剰な連結を数値化する (%)。

③ タイプ・トークン比 (TTR)

語彙多様度の指標。同じ単語の繰り返しが多いと TTR が下がり、学習者にとっては逆に負担が減る面もあるが、全体の語彙展開は単調になりがちである。

④ 平均文長 (形態素ベース)

1 文あたりの形態素数を算出し、文が長くなりすぎていないかを測定する。CEFR B1 では過度に長い文が読解を阻害する可能性がある。

5.3.2 文末敬体混在率：どの文章でも 0% か

① GPT-4o の自律的文体一貫性

今回の解析結果によれば、対照群・実験群いずれの本文についても、文末が敬体と常体の両方を混在させる事例は一切見られず、混在率 0% であった。従来、抽象度の高い Can-do のみを与えた場合に文体が乱れる懸念があったが、GPT-4o は自律的に常体か敬体のいずれかを選択して全文を貫く傾向が強く、対照群であっても文末が統一されていた。

② 敬体か常体かの選択

JF561 (人との関係) に対する対照群の出力例 (表 1 参照) では、一貫して常体を選んでいるが、文体ブレはなく、学習者向け文章として必ずしも読みにくくはない。ただし、教師が敬体を強く希望する場合、GPT-4o が常体を選ぶ可能性がある点を考慮すると、事前変換プロンプトに「～です／～ま

すの文末にする」と具体的に記述しておく方が、意図通りの文体を確実に得られる。

5.3.3 接続詞の過剰連結率

① 絶対値としての低さ

「そして」「それで」「しかし」「だから」などの接続詞が連続使用されていないかを測定したところ、対照群は平均 2.9%、実験群は 1.7%程度という推定値を示した。いずれの値も大きくはなく、GPT-4o が抽象的 Can-do でも冗長な接続詞を大量に挿入する事例は少数であった。

② 教師の意図との合致

とはいえ、実験群の方がさらに抑制されている(1.7%前後)という微差が認められ、接続詞を必要最小限にする指示が反映されやすいことが分かった。教師が「文章を簡潔に」と望む場合、事前変換プロンプトで「接続詞を多用しない」と指定することで、GPT-4o がさらに簡潔かつ読みやすい文章を生成する傾向が強まると推定される。

5.3.4 タイプ・トークン比 (TTR) と平均文長

① TTR

語彙多様度の指標である TTR を算出したところ、対照群・実験群ともに 0.46~0.48 前後に収まり、大きな群間差は認められなかった($p=.29$, n.s.). GPT-4o の能力により、抽象的 Can-do でも極端に難解な語彙を用いず、適度なバリエーションをもった文章を生成するため、事前変換プロンプトが語彙多様度を大幅に変えることは少ないようである。

② 平均文長

一方、平均文長(形態素数ベース)を見ると、対照群が約 16.5 形態素/文、実験群が 15.2 形態素/文という推定値となり、 $p<.05$ の有意差が確認された。たとえば JF555(自由時間と娯楽)では、実験群が「文を短めにする」という指示を忠実に守り、1 文を 20~25 字程度にまとめていたことが確認

された。B1 学習者にとって、長文の多用は読解のハードルを上げる要因となるため、事前変換により文長がやや短縮する効果は依然として教育的意義をもつと考えられる。

5.3.5 総合的評価とテンプレート化の意義

以上の形態素解析結果から、GPT-4o が抽象的 Can-do（対照群）であっても、文末敬体と常体が混在するような文章をほとんど生み出さず、また接続詞連続使用率が 2～3%程度と高くはない点が確認できた。これは、従来の懸念ほど対照群が乱れやすいわけではなく、モデルの高い自己補正力が働いていると推測される。

ただし、事前変換プロンプトでは、(1)文末敬体を確実に指定できる、(2)接続詞をさらに抑制して簡潔な文体を誘導できる、(3)平均文長を短めにしやすい、という三つの追加効果が見られ、統計的にも小さいながら有意差を生み出している。さらに、本研究が注目してきたテンプレート化は、この事前変換プロンプトの設計を AI の支援下で進めるという新たな可能性を開く。すなわち、教師が Can-do ステートメントを要素分解しながら生成 AI からの提案を受け、語彙レベルや文体制御の細かい設定を洗練していくことで、より汎用的かつ高精度な教材作成プロセスを構築できると考えられる。

これらの点を踏まえて、本研究が提示した事前変換の教育的意義と限界を次節（5.4）において総合的に考察し、CEFR B1 レベル以外の適用範囲や今後の課題についても論じることとする。

5.4 総合考察：GPT-4o の高性能と事前変換の位置づけ

本章では、対照群（抽象的 Can-do）と実験群（事前変換プロンプト）の二条件下で生成された計 30 本文を対象に、(1) jReadability スコアによる可読性、(2) 文末敬体混在率や接続詞過剰連結率など形態素解析に基づく指標を比較した。

① GPT-4o 自体の高い能力

対照群でも文体が混在せず、接続詞連続使用率や可読性スコアが B1 レベル近辺に収まる例が多く、乱れは想定より軽微であった。

② 事前変換による追加効果

統計的にはわずかながら有意差を生み、敬体を確実に使わせたり、文長や接続詞多用を抑止したり、語彙の難易度をコントロールしたりできる。

③ テンプレート化

事前変換プロンプトを定型化し、AI と協働しながらプロンプトを仕上げるという運用が、教師の意図を最大限に反映できる道となる。既に GPT-4o は抽象的 Can-do でも高品質を発揮するが、細やかな制約を与えることで教材の完成度を一層高めることが可能である。

次章（第 6 章）では、以上の結果を踏まえ、事前変換手法の教育的意義と限界を考察し、日本語教育の現場や他レベル・会話タスクへの展開について議論を深める。

6. 考察

6.1 事前変換プロンプトの有効性と GPT-4o の高性能

第 5 章で示したように、抽象的 Can-do ステートメントを入力するだけでも、GPT-4o が比較的高い精度で B1 レベル相当の日本語教材を自動生成する可能性が確認された。対照群（素の Can-do）でも、文体が中途半端に混在することはほとんどなく、過度に難解な語彙や長文が多用されるわけでもない。このような傾向は、近年の大規模言語モデルの著しい性能向上を物語るものである。

一方、事前変換プロンプトを用いた実験群では、さらに敬体・常体の選択を教師が確実に指定できたり、接続詞の連続使用や文長などを微調整しやすいなどの利点が認められた。実験においては対照群よりわずかに可読性スコアが上昇し、文体が安定しやすくなることが統計的にも示唆されている。これらの結果から、GPT-4o の自律的な能力と、教師による事前変換の付加的効果が併存する形となり、「十分に高性能なモデルに任せつつも、教師の意図を精密に反映させたい場合は事前変換が依然有効」という結論を導くことができる。

6.2 具体的制約とテンプレート化：事前変換のさらなる展望

6.2.1 教師と生成 AI の協働によるプロンプト設計

第 5 章では、対照群でも一定以上の質が得られる状況を確認したが、教師の意図（例えば「敬体を使う」「難読漢字を避ける」など）を高い精度で実現するためには、事前変換プロンプトが依然として有用である。これまでは教師が自力で細分化や語彙制限を考え、プロンプト文面を決定するという想定であったが、近年の生成 AI はプロンプト自体の記述を AI に相談しながら作成するという新たなアプローチを可能にしている。

具体的には、教師が CEFR B1 レベルを想定したタスクに必要な語彙や文体要件を大まかに提示し、AI からテンプレート草案を提案してもらうという方法が考えられる。AI が提案するテンプレートを教師が修正し、それを最終的なプロンプトとして再度 AI に渡すことで、教師の専門知識と AI の文書生成能力が相互補完的に働くという実践形態が期待できる。

6.2.2 テンプレート化の教育的意義

テンプレート化された事前変換プロンプトを使えば、B1 相当以外の学習レベル（A2 や B2 など）や、口頭技能や会話型タスクといった異なる種類の Can-do にも汎用的に対応しやすくなる。教師の作業時間が限られている現場では、これらテンプレートを使い回すことで、多様な場面やレベルの教材を短時間で生成し、それを必要に応じて微修正する運用が実現する。したがって、単に抽象的 Can-do を入力するだけではなく、事前変換テンプレートを充実させることが、より高品質な教材作成プロセスを安定的に構築する上で重要な鍵となる。

6.3 CEFR B1 を超えた拡張と会話型タスクへの応用

本研究は B1 レベルを事例として取り上げたが、日本語教育現場には多様なレベルや技能が存在する。たとえば、A2 レベルの学習者には文法項目や語彙の範囲がさらに狭く、B2 以上の学習者には複雑な文型や議論展開が求められる。抽象的 Can-do ステートメントをレ

ベル別に事前変換することで、GPT-4o がそのレベルに応じた文体・語彙を選択しやすくなる可能性がある。加えて、会話型タスク（ロールプレイやインタラクティブ対話）への応用も視野に入れるべきであろう。

事前変換プロンプトで、会話シナリオの設定や発話ターン構成を詳細に指定すれば、学習者にとって自然な発話例を生成できる見込みがある。従来、会話教材の作成には教師の多大な時間がかかるとされてきたが、LLM を活用すれば、複数の話者や多段階の会話展開を想定したテンプレートを作成し、B1 以外のレベルにも適用することが可能となるであろう。

6.4 機械評価と人的評価の相互補完

本研究では、可読性スコア（jReadability）や形態素解析指標（敬体混在率、接続詞過剰連結率、TTR、平均文長）といった機械評価を軸に実験結果を示した。しかし、教材として真に有効かどうかは、最終的には学習者がその教材を使って達成感や理解度を得られるかによって判断される。すなわち、本研究で扱った客観的指標はあくまで教師が教材の難易度や文体を迅速に比較・検討するためのサポートツールであり、「教師・学習者の主観評価」や「教室実践での実際の効果測定」と併せて総合的に評価する必要がある。

加えて、対照群でも高品質な文章を生み出す GPT-4o の現状を見ると、「ある程度の品質が保たれれば OK」というケースでは機械評価だけで素早く判断し、教師の修正コストを減らすことができる。一方、文体面や語彙選択を厳密に統制したいケースでは、事前変換を施して丁寧な教材を生成し、学習者からのフィードバックを経て微調整する、というサイクルが有効となる。

6.5 まとめ

本章では、GPT-4o の高性能が対照群でもそこそこの質の文章を生み出す一方、事前変換プロンプトがなお細部制御と品質安定化に寄与する様相を明らかにした。特に文体（敬体・常体）や文長、接続詞の多寡といった細かな領域では、教師がプロンプトをテンプレー

ト化し、レベル別・場面別に応用する余地が大きい。さらに、機械評価のみならず学習者の主観評価や教室実践との連動を深めることで、本研究の成果をより実際の教育現場に役立つ形へ育てることが期待される。

次章（第7章）では、本研究全体を総括し、事前変換手法が示す可能性と限界、そして今後の研究課題を提示する。

7. 結論と今後の課題

7.1 本研究の総括

本研究は、CEFR B1 レベルに相当する Can-do ステートメントを活用し、大規模言語モデル（GPT-4o）の自動生成による日本語教材設計を検証したものである。とりわけ、抽象度の高い Can-do をそのまま入力する“対照群”と、場面設定・語彙制限などを明示的に行った“事前変換プロンプト（実験群）”を比較し、それぞれの生成文章を可読性指標（jReadability）や形態素解析に基づく文末敬体混在率、接続詞過剰連結率、タイプ・トークン比（TTR）、平均文長などの観点から評価した。

結果として、GPT-4o が抽象的 Can-do のみでも B1 レベルに近い文章を生成する例が多く、以前に懸念されていたような文体混乱や難易度のばらつきは限定的であった。一方で、事前変換プロンプト群では統計的にもわずかながら可読性スコアが向上し、文体の安定度や文長制限など、教師の細かい要望を反映しやすい利点が見いだされた。この差は、高性能なモデルに任せるだけでも一定以上の質が担保される現状において、教師がいかに「微調整しやすいか」という視点からなお意義をもつと判断される。

7.2 研究の意義と限界

7.2.1 研究の意義

① 事前変換による教材自動生成の枠組み

（ア） CEFR B1 レベルに焦点を当てつつ、教師が Can-do ステートメントを要素分解し、文体や語彙を制御する手法

が有効である点を数値的に示した。

- (イ) 高性能 LLM を前提とする現状でも、事前変換による「教師の意図の精密な伝達」が教材生成の安定性を高める一つの方策である。

② 具体例の提示と指標ベースの評価

- (ア) 対照群・実験群で得られた生成文章の一部を実際に提示し、jReadability や形態素解析指標などの観点から両者の差を明示した。本研究独自の基準（B1 を 3.5～4.0 帯）での結果比較は、今後の類似研究にも応用可能と考えられる。

③ テンプレート化と AI 支援によるプロンプト設計

- (ア) 高性能 LLM とのやりとりを活用してプロンプト自体をテンプレート化し、Can-do ステートメントの抽象度を補う案を議論した。これは多忙な教師の労力軽減と、教材設計精度の向上に寄与すると期待される。

7.2.1 本研究の限界

① 機械評価と学習者評価の乖離

- (ア) 本研究は機械評価（jReadability や形態素解析）を中心としたため、学習者が実際に「読みやすい」と感じるか、「興味をもって取り組める」かなどの主観的・教育的評価までは十分に検証できていない。
- (イ) 教師と学習者の実際の使用感や学習効果の観点から、本研究の指標をどこまで有効とみなせるかは、さらなる教室実践やアンケート調査が必要である。

② B1 レベル以外への適用範囲

- (ア) 本研究は B1 レベルを対象としたが、同様の事前変換手法が A2 や B2、あるいは C レベルへどのように拡張できるかは未確定である。レベルごとの文体や語彙統制の仕組み、タスク設計上の注意点などは今後検討の余地がある。

③ モデル依存性

- (ア) GPT-4o を使用した結果を示したが、他のモデル（OpenAI o1, o3, GPT-4.5, あるいは他社 LLM）で同様の効果が得られるかは未検証である。モデル固有の出力特性や更新に伴う振る舞いの変化を考慮すると、事前変換の適用方法も微調整を要する。

7.3 今後の課題と展望

7.3.1 学習者評価との統合

事前変換プロンプトによって得られた文章が、どの程度学習者の理解やモチベーション向上につながるかを、教室実践の場で評価することが求められる。たとえば、A/B テストの形で対照群・実験群の文章を学習者に読ませ、難易度感や興味度合いをアンケート調査することで、機械評価との関連を検証できると考えられる。

7.3.2 他レベル・スキルへの展開

B1 レベルに特化して検討した結果は、A2 や B2 などの異なるレベルにも応用可能であるが、それぞれのレベルで必要となる語彙やタスクの要求度は異なる。事前変換テンプレートをレベルごとに整備し、同様の手法で対話生成や会話教材に展開することが展望される。なかでも会話型タスクでは、ロールプレイに適したやりとりを自動生成する際に、事前変換によるターン構成・敬語選択などの効果を検証できよう。

7.3.3 複数モデル比較とバージョンアップへの対応

本研究は GPT-4o を対象に行ったが、LLM は日進月歩でモデルアップデートが続いており、他のモデルや後続バージョンとの比較検討を行えば、事前変換プロンプトの汎用性や最適化方法がさらに明確になるだろう。教師の専門知識をいかにモデルに反映するかを探るうえで、モデルの性質ごとに異なる事前変換テンプレートが有効となる可能性が高い。

7.4 結論

本研究は、CEFR B1 レベルの Can-do ステートメントを活用し、大規模言語モデル（GPT-4o）が自動生成する日本語教材の難易度・文体特性を検証したものである。抽象的 Can-do だけを入力しても、近年のモデル性能により比較的高品質な文章が生成される例が多いことが示唆されたが、より教師の意図を精密に反映させるには、事前に語彙・文体・場面情報を具体化する「事前変換プロンプト」手法が依然として効果をもつことが確認された。わずかな差ではあっても、文章の安定度や可読性スコアで有意な改善が見られる点は、教師の教材作成コストを低減するうえで重要な示唆である。

さらに、テンプレート化や AI との協働プロンプト設計により、多様なレベル・技能・場面に対応した教材自動生成が可能となりうる。最終的には機械評価を超えた学習者評価や教室実践を通じて効果を実証していく必要があるが、本研究が示した事前変換の枠組みは、教師と生成 AI の相乗効果を生かし、日本語教育の現場で今後より実践的に活用される土台となるだろう。

参考文献

- 独立行政法人国際交流基金（2010）「JF 日本語教育スタンダード」、国際交流基金
- 塩澤真季・石司えり・島田徳子（2010）「言語能力の熟達度を表す Can-do 記述の分析—JF Can-do 作成のためのガイドライン策定に向けて—」『国際交流基金日本語教育紀要』6、pp.23-39.
- Brown, T. et al. (2020). Language Models are Few-Shot Learners. *NeurIPS 2020*.
- Council of Europe (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press.
- Hasebe, Yoichiro and Lee, Jae-Ho (2015) ‘Introducing a Readability Evaluation System for Japanese Language Education’ *Proceedings*

*of the 6th International Conference on Computer Assisted Systems
for Teaching & Learning Japanese (CASTEL/J), pp.19-22.*

Vaswani, A. et al. (2017). Attention Is All You Need.” *Advances in Neural
Information Processing Systems*, 30, pp.5998-6008.

＜付記＞

本研究では、日本語文章難易度判別システム (<http://jreadability.net/>)
を使用した。