

在第二外語教學中導入生成式 AI (Large Language
Models Generative AI) 之考察
— 以 JFL 學習者為研究的對象 —

王睿琪

東京外國語大學大學院國際日本學研究院兼任講師

林俊成

東京外國語大學大學院國際日本學研究院教授

摘要

在第二語言的習得過程中，「說話」、「寫作」、「會話 (Conversational Interaction)」等語言輸出，在了解自己語言能力極限、驗證中間語言假說、促進句法統整、語言知識自動化等方面，擔任著極重要的角色。

本研究的焦點是第二外語的語言輸出。為此本研究開發了一套生成式 AI 語音會話系統「GPT-Talk」。這套系統是架構在大規模語言模型系統上的。學習者透過與 GPT-Talk 的交談，不但可以提升外語的說話能力，並同時增長資訊檢索及收集能力。

本研究的目的是要驗證這套生成式 AI 語音會話系統，是否具有促進語言輸出的效果。研究的方法是，在簡報課裡讓學習者製作 PPT 與發表簡報。對象是 JFL (Japanese as a Foreign Language) 學習者。研究的方法是「課堂觀察」、「評量規準」。透過這兩個分析來驗證這套生成式 AI 語音會話系統的有效性與教學上的效果。

關鍵詞：生成式 AI、第二外語習得、評量規準、語言輸出、
ChatGPT

受理日期：2024 年 03 月 09 日

通過日期：2024 年 05 月 24 日

DOI：10.29758/TWRYJYSB.202406_(42).0003

Efforts to Teach Second Language Classes Using Generative AI (Large Language Models Generative AI) : Targeting JFL Learners

Wang, Juei-Chi

Lecturer, Tokyo University of Foreign Studies Graduate School
of Studies

Lin, Chun-Chen

Professor, Tokyo University of Foreign Studies Graduate School
of Studies

Abstract

In the process of acquiring a target language, productive activities such as "speaking," "writing," and "conversational interaction" play important roles in recognizing gaps, testing the interlanguage hypothesis, facilitating syntactic processing, and automating language knowledge. This study focuses on the production in a second language and endeavors to develop a voice conversation system based on large language models, named "GPT-Talk." It is believed that through interactions with GPT-Talk, learners can simultaneously enhance their productive abilities and information retrieval and gathering skills in the target language.

The objective of this study is to verify whether GPT-Talk has an effect on promoting production. Therefore, presentation classes were conducted targeting Japanese as a Foreign Language (JFL) learners, and analyzed using "class observations," "rubric assessments."

Keywords: Generative AI, Second Language Acquisition,
Rubrics, Output, ChatGPT

第二言語における生成型 AI (Large Language Models Generative AI) を活用した授業の試み — JFL を対象に —

王睿琪

東京外国語大学大学院国際日本学研究院非常勤講師

林俊成

東京外国語大学大学院国際日本学研究院教授

要旨

目標言語の習得過程において、「話す」「書く」「会話 (Conversational Interaction)」などの産出活動は、ギャップの認識、中間言語仮説の検証、統語的処理の促進、言語知識の自動化など、重要な機能を果たしている。

本研究では、第二言語の産出に焦点を置き、大規模言語モデルに基づく生成型 AI を活用した音声会話システム「GPT-Talk」の開発に取り組んだ。この GPT-Talk との対話を通して、学習者は目標言語での産出能力と情報検索・収集能力を同時に向上させることが可能であると考えられる。

本研究の目的は、GPT-Talk が産出促進効果を有するかどうかを検証することである。そのため、JFL 学習者を対象に、プレゼンテーション授業を実施した。「授業観察」「ルーブリック評価」を用いて分析し考察した。

キーワード：生成型 AI、第二言語習得、ルーブリック、
アウトプット、ChatGPT

第二言語における生成型 AI (Large Language Models Generative AI) を活用した授業の試み

王睿琪

東京外国語大学大学院国際日本学研究院非常勤講師

林俊成

東京外国語大学大学院国際日本学研究院教授

1. はじめに

コロナ禍を契機として加速した DX (Digital Transformation) の流れの中で、生成型 AI (Generative AI) の発展、特に ChatGPT のようなツールの浸透が、言語教育分野にも顕著な影響を与えている。具体的には、「教材・教具の IoT (Internet of Things) 化」「ロボット教師」「xR (xReality) によるバーチャル英会話」「バーチャル留学体験」が実現し、教授方略や学習方略においてもすでに変革が始まっている。DX の適用により、言語のインプットおよびアウトプットの強化が可能となり、従来の教育方法に比べて、より多様な手段やツールを用いて言語教育の本質である言語能力の向上を図ることができるようになった (図 1)。

言語習得においては、インプットだけでは不十分であり、「話す」「書く」「会話 (Conversational Interaction)」といった産出活動が、ギャップの認識、中間言語仮説の検証、統語的処理の促進、言語知識の自動化などの重要な機能を果たしている (村野井 2011; Swain 1985, 1995)。これらは学びを定着させ、スキルアップするために欠かせない要素である。しかし、特に海外の目標言語への接触機会が限られた環境では、「話す」「会話」の機会が制限され、これらの能力の発展が阻害される傾向がある。

この現状を踏まえ、本研究は第二言語の産出に着目し、大規模言語モデルに基づく生成型 AI (LLM-GAI/Large Language Models - Generative Artificial Intelligence) を用いた対話型 AI 音声会話システム「GPT-Talk」を構築した。この GPT-Talk との「対話」を通して、学習者は目標言語の産出を行い、同時に情報を「検索」「収集」する能力を向上させることができると考えられる。本研究の目的は、GPT-Talk の導入が産出促進効果を有するかどうかを検証することである。

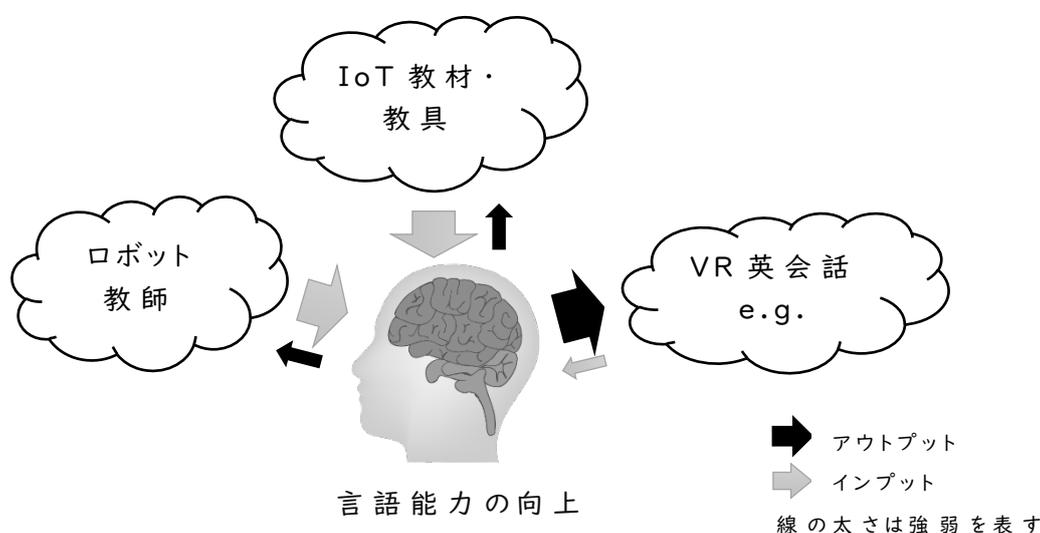


図 1 DX の活用と言語運用の相互作用

2. 先行研究

2.1 アウトプット仮説

Swain (1985) は、第二言語習得における Krashen の「インプット仮説」に疑問を投げかけ、言語学習には「話す」「書く」といったアウトプットも不可欠であると主張した。

インプット仮説は、第二言語習得において役に立つのは、目標言語の流暢さを高める点にある。しかし、流暢さと正確さは言語パフォーマンス上では異なる次元にあり、練習が流暢さを向上させる可能性があるものの、正確さの向上が保証されるわけではない (Swain 1995:125)。その一方で、アウト

プットの重要性について、学習者がインプットよりもアウトプットを通じて、言語をより深く処理するよう促されると指摘している (Swain 1995:126)。

Swain (1995: 125-6) は、第二言語習得におけるアウトプットの役割に関して、「気づき」「仮説検証」「統語処理」の3つの機能を促進するという仮説を提唱している。

① 気づき機能 (Noticing function) :

アウトプットの機会において、学習者は伝えたいことと伝えられることとの間にあるギャップ (言語的問題) を意識的に認識するよう促される。このギャップに気づくことで、学習者は新たな言語知識の獲得や既存の知識の補強の契機を得る。

② 仮説検証機能 (Hypothesis-testing function) :

アウトプットは学習者が自らの中間言語仮説を検証する機会を提供する。特に誤ったアウトプットはしばしば、学習者が中間言語仮説を形成し検証するという作業を伴う。誤ったアウトプットが相手からフィードバックを引き出すこともあり、それによって学習者は自らの中間言語法則を修正することにつながる。

③ メタ言語的機能 (Metalinguistic function) :

学習者は自身の目標言語使用を振り返ることで、アウトプットはメタ言語的機能を果たし、言語知識の制御および内面化が可能になる。

Swain 氏の研究により、第二言語習得におけるアウトプットの重要性が明らかにされている。学習者が積極的にアウトプットすることにより、自らの中間言語仮説を検証し、メタ言語的な思考を促すことは、言語習得をより深化させ、効果的な学習へと導く鍵となるだろう。

2.2 生成型 AI 研究

コロナ禍と生成型 AI「ChatGPT」の台頭により、日本語教

育における生成型 AI の活用に関する実践が多数報告されている。しかしながら、これらの実践研究が教育上、どこまでの役割を果たしているのかは明らかにされていない。本研究は、まず、Ruben R. Puentedura (2010) によって考案された SAMR モデルを概観し、続いて、SAMR モデルを用いて、日本語教育方法研究会 (2022 年以降) にて発表された生成型 AI の実践研究を分類・整理する試みを行う。そして、これからの教授方略や学習方略に与える影響について考察する。

2.2.1 SAMR モデル

Ruben R. Puentedura が 2010 年に提唱した SAMR モデルは、教育分野における新たなテクノロジーが、従来の教授方略や学習方略と比較し、学習活動にどんな影響を与えるかを示す尺度である。このモデルは、テクノロジーが教育プロセスにどのように組み込まれ、教育の質をどのように変化させるかを理解するために使われる。また、教育者がテクノロジーをどのように利用しているかを理解し、より効果的な教授方略へと進化させるためのガイドとしても使われる。

SAMR モデルは、Substitution (代替)、Augmentation (拡大)、Modification (変形)、Redefinition (再定義) の 4 つの段階で構成され、段階的により深い学習の変容を促すものである。SAMR モデルの各段階における内容と、生成型 AI を活用した際のその定義および具体例を、表 1 に示す。

表 1 SAMR モデルの各段階と具体例

段階	筆者訳	定義	具体例
S 代替	技術により従来ツールを直接に置き換えるが、機能的な変更はない	ChatGPT を用いて資料検索・要約・翻訳・添削・データ分析などの作業が行われる	ChatGPT を用いて動画の台本を生成する
A 拡大	技術により従来ツールを置き換え、さらに新たな機能が付与される	ChatGPT、Midjourney、Whisper など複数の生成ツールを組み合わせることで、学習活動を視覚のみならず聴覚にも展開する	画像生成ツールを使用して画像を生成・編集する
M 変形	技術により重大な学習活動の再設計を可能にする	Padlet、classroom、Google スライドなどを活用し、学習成果の共有を行うと共に、ピアリーディング・ディスカッションなどの活動を通じて、学びを深める	動画学級内に共有し、発表を通して相互評価する
R 再定義	技術により全く新しい、以前は想像できなかった学習活動を可能にする	学習成果は、Zoom などを利用して他校との交流を同時に行うことや、SNS や HP を通じて社会にリリースする	YouTube などのプラットフォームに発信する

注：表 1 の定義では、Puentedura(2010) が述べた「Tech」を「ChatGPT」に置き換えて解釈する。

2.2.2 生成型 AI の研究

2022 年以降、日本語教育方法研究会における生成型 AI に関する実践研として、潘(2022)、寺嶋ほか(2023)、夏(2023)、王(2024)、劉(2024)、孫(2024)、寺嶋(2024) の 7 件を抽出した。

潘(2022) は、スマートフォン向け AI チャットアプリ「Airfriend」を活用し、「導入」「練習」「発表」「省察」のフェーズから構成された会話授業を実施した。その結果、学習者の Airfriend とのやり取りを①「教師が指定した会話授業の課題」、②「Airfriend をバーチャルな友人として捉える」、③

「Airfriend を日本語練習のツールとみなしながら、ゲーム感覚での自然な会話の限界を試す」の3つのタイプに分類している。

寺嶋ほか（2023）と寺嶋（2024）は、ChatGPT を利用し、学習者が ChatGPT からの作文フィードバックをどのように受け入れるかを検討した。その結果として、学習者は自身の判断に基づいて ChatGPT のフィードバックを活用するかどうかを選択する傾向があることが示唆されている。

夏（2023）は、ChatGPT を活用して学習者の作文添削を行った実践研究を報告している。プロンプトの設定を通じて日本語レベルに応じた添削が可能であるものの、ChatGPT には専門的な知識が不足していることや、学習者の意図や文脈を正確に把握する能力に限界が存在すると指摘している。

王（2024）は、プレゼンテーション授業において対話型 AI を用いた音声会話システムを用い、その導入が目標言語使用とプレゼンテーションパフォーマンスに及ぼす影響を調査した。その結果、目標言語の発話量の増加と発表の流暢性に向上が見られたと報告している。

劉（2024）は、音楽生成 AI「Suno AI」を使用し、学習内容の復習ツールとしての有効性を検証した。その結果、音楽を介した学習は、被験者にとって魅力的であり、理解の強化、記憶効果の向上、学習意欲の増加が観察されたと述べている。

孫（2024）は、LLM-GAI を利用して日本語教材の試作を行い、ChatGPT Plus の Advanced Data Analysis 機能を用いて「祭り」「旅行」「映画」「食事」「読書」の5つのテーマに基づくモデル会話を生成したと報告している。孫（2024）が開発した AI 日本語教材は「本文」「語彙説明」「文型解説」「ドリル練習」「自由会話練習」で構成されており、特に「自由会話練習」においては、Create a GPT と ChatGPT、Whisper、TTS の API を組み合わせることで、リアルタイムでの自由会話が実

現できるシステムを開発したとしている。

孫（2024）の研究においては、具体的な実践的研究が行われていないため、SAMR モデルに基づく分析からは、その研究は除外された。生成型 AI を活用した学習活動がどの段階に影響を与えるのかについての分析結果は、表 2 に示されている。

表 2 SAMR モデルに基づいた分類した結果

段階	実践研究
Substitution 代替	寺嶋ほか（2023）、夏（2023）、 劉（2024）、寺嶋（2024）
Augmentation 拡大	該当なし
Modification 変形	潘（2022）、王（2024）
Redefinition 再定義	該当なし

表 2 に示されている通り、生成型 AI に関連する研究は「代替」レベルにとどまるものが多く、生成型 AI を活用しているものの一方的で新しい知識を取り入れる受動的な学習が主流である。一方で、「変形」レベルに達している研究は少ないものの、学習者が自身の既存の知識や経験と結びつけたり、自分の意見を他者に伝えたりするような能動的な学習を支援する研究が既に存在している。

生成型 AI を授業に取り入れることは、知識獲得に寄与すると言えるが、学習をより深めるためには、生成型 AI の活用に加えて、協調活動、表現活動、リフレクションを組み合わせる必要がある。すなわち、学習活動の内化と外化を繰り返すことにより、学習の深化が図られる。

これらの先行研究を基に、本研究では以下の研究課題を設定する。

3. 研究課題

本章では、本研究で開発された自由会話システム「GPT-Talk」について紹介する。続けて、研究課題についても述べる。

3.1 GPT-Talk とは

本研究では、OpenAI によって開発された音声認識技術である Whisper と音声合成 API の TTS を用いて、ChatGPT を中核とした自由会話システム「GPT-Talk」を構築した。

本システムでは、Whisper の音声認識機能によって学習者の発話をテキスト化し、このテキストを ChatGPT に送信する。



その後、ChatGPT によって生成された応答テキストが TTS の音声合成機能を介して音声として学習者に返されるというプロセスから構成されている。また、学習者が GPT-Talk との談話内容を保存すると、その内容は自動的に教師側に送信される。教師はこれを閲覧することで、学習者の学習状況を把握し、さらなる指導を行うことも可能である¹。図 2 は本システムの画面の例である。

図 2 GPT-Talk の画面

GPT-Talk の特徴は、ChatGPT に音声認識機能と音声合成機能を組み込むことにより、音声によるプロンプト入力を可能にし、ChatGPT によって生成されたテキストを音声またはテキスト形式で提示することができる

¹ 本研究で開発された自由会話システム「GPT-Talk」は、学習者が送信ボタンを押すことにより、会話内容をすべて文字で記録する機能を有している。この記録されたデータを分析することで、学習者の発話の質の変化を観察することが可能である。特に、質問スキルの欠如や会話のつまづきが明らかになる。これらの点に対して教師は適切なフィードバックを提供することができる。

ようになった。これにより、学習者の言語産出を促進し、自由な会話を実現することが可能となった。

3.2 研究課題

稲葉（2014）は、プレゼンテーション（以下、プレゼン）活動がアウトプットの機会を提供し、学習者同士のインタラクションを通すことによって発表内容や言語理解が深まると指摘している。この点に鑑み、本研究ではプレゼン授業を用いて以下の三点について考察し、GPT-Talk の教育分野への適用可能性を検討する。

1. GPT-Talk の導入が学習者の目標言語の発話を促すかどうか
2. GPT-Talk を導入することで、プレゼン資料の制作内容にどのような変化が生じるか
3. GPT-Talk を導入した場合に学習者の目標言語使用にどのような変化が生じるか

4. 調査概要

4.1 対象者と調査期間

本研究では、中国語を母語とする JFL 学習者（Japanese as a Foreign Language）を対象に、プレゼン授業を実施した²。調査の概要は表 3 の通りである。事前に GPT-Talk とプレゼン授業の効果を評価するための予備調査を実施した。

予備調査により、目標言語の使用頻度が顕著に増加したことが示されたため、本調査ではプレゼン授業の回数を 4 回から 2 回に減らした。さらに、予備調査では学習者にとって身近で興味深い題材を選定したが、本調査ではより積極的な発

² 本研究で実施されたプレゼンテーション授業は、正規の授業とは異なる実験的な授業であった。また、研究者と協力校の都合により、予備調査と本調査で異なる大学の JFL 学習者が対象となった。

話を促すために、学習者間での情報ギャップが大きい題材に変更した。

表 3 調査情報と学習者プロフィール

対象者	予備調査 (n=8)	本調査 (n=14)
調査機関	台湾の K 大学	台湾の T 大学
調査期間	2023 年 9 月 26 日～ 2023 年 10 月 12 日	2023 年 12 月 25 日～ 2024 年 1 月 3 日
日本語レベル	N1 取得者：2 名 N2 取得者：6 名	N1 取得者：10 名 N2 取得者：4 名
年齢	M:21 歳 1.6 ヶ月 (SD:7.64)	M:21 歳 2.1 ヶ月 (SD:24.48)
授業回数	4 回	2 回
プレゼン題目	① 台湾の食文化 ② 台湾の交通事情 ③ 台湾の年中行事 ④ 台湾の観光名所	① 日本の教育 ② 日本の社会問題

注：M は平均値 (Mean)、SD は標準偏差 (Standard Deviation)。

4.2 調査手順

調査は以下の手順に従い実施された。

- I. 実験の目的、手続きなどを説明し、参加者には同意書を提出してもらった (初回)。
- II. 評価基準となるルーブリックを配布し、その評価項目について説明した。
- III. プレゼン授業を実施した。
 - ① 主題についての導入 (5 分)
 - ② 副題の選定とパワーポイント (PowerPoint、以下 PPT) 資料の作成 (30 分)
 - ③ プレゼンの実施と質問応答 (10 分)

本調査の 1 回目のプレゼン授業では、資料検索に用いる検索エンジンの選択を学習者に委ねた。一方、2 回目のプレゼン

ン授業では、資料検索のツールは GPT-Talk に限定した³。

データの収集に関しては、Ⅲの②およびⅣでは IC レコーダーによる音声録音を行い、Ⅲの③ではビデオカメラで撮影した。

4.3 分析方法

本研究は、「発話量の測定」「ループリック評価」を用いて分析を行った。具体的には、目標言語の発話量については、調査手順のⅢの②「副題の選定と PPT の資料作成」で得られた音声データを基に、母語と目標言語の発話量を計測し分析した。PPT の作成およびプレゼンのパフォーマンスに関しては、ループリック評価を使用して分析を行った。なお、PPT の作成の評価では、「スライドデザイン」「スライドの枚数」「スライド内の文字数」「スライドの形式」「スライドの構成」「情報の正確性」の 6 つの項目を設け評価した。プレゼンのパフォーマンスの評価では、「発表の態度」「発表時間」「発表の進行」「語彙の豊かさ」「単語および文法の正確性」「発音の明瞭さ」「流暢性」の 7 項目を評価基準とした（表 4）。

³ 予備調査における 1 回目および 2 回目のプレゼン授業では GPT-Talk を導入していないが、3 回目および 4 回目の授業においては GPT-Talk の導入を行った。

表 4 ルーブリック評価

評価項目	4 (優)	3 (良)	2 (可)	1 (不可)
スライドデザイン ・文字の提示の適切さ ・図表や動画の提示の適切さ	文字は色、大きさ、字体などが工夫されていて美観に富みやすく、スライド間で統一されていて一体感があった。図表も見やすく、アニメーションや動きを使用した場合も非常に効果的であり、文字、図表とともにプレゼンテーション内容を引き立てていた。	文字の色、大きさ、字体などが工夫されており、大部分の図表が見やすかったが、一部見にくい箇所があったものの、全体的に文字、図表、写真などのバランスをうまく取っていた。アニメーションや動きを使用した場合も効果的であり、プレゼンテーション内容をわかりやすく伝えていた。	文字よりも写真が主体であったため、情報伝達が多分であった。また、スライド間の統一性は欠けた。図表に関しては若干の不具合が指摘されたが、それでも、プレゼンテーション内容との関連性も不明瞭だった。アニメーションや動きを使用した場面も効果的ではなかった。	すべての文字が読みやすかったが、写真が主体となっており、情報の伝達が効果的ではなかった。図表をわかりやすくする工夫が見られず、プレゼンテーションの内容との関連性もなかった。
スライドの枚数 (発表時間 5分)	発表時間を厳密に考慮し、全ての用意されたスライドを計画通り時間内に発表できた。また、スライドの重み発表時間に合わせて適切に調整された。	発表時間の管理は十分でなく、多めの重なるスライドを用意してしまったり、そのため、大部分のスライドを発表することができず、時間内に全てをカバーするために若干の調整が必要だった。	全体的にスライドは発表時間内に紹介できたが、用意したスライドの重なりが多かったため、発表が計画よりも早く終了してしまったり、時間内に全てをカバーするために若干の調整が必要であった。	ほとんどのスライドは、文字が多すぎた。あるいは文字が極端的に少ない(図やイラストのみの提示)という問題があった。
スライドの文字量	各スライドに適切な文字が配置され、聞き手が容易に理解できる。情報は整理され、視覚的に分かりやすい。	大部分のスライドは、適切な文字であるが、一部で文字が多すぎることがあるスライドが存在する。	ほとんどのスライドは、文字が多すぎることがあり、調整が必要であった。	ほとんどのスライドは、文字が多すぎた。あるいは文字が極端的に少ない(図やイラストのみの提示)という問題があった。
スライドの形式	スライドの表紙には、テーマと発表者の氏名、所属などが明示されている。また、目次、引用先 (参考文献、引用文庫、サイトなど)、著作権に関する情報も完全に記載される。	テーマと発表者の氏名、所属、目次、引用先、著作権に関する情報の大半は表示されているが、一部の情報が欠けている。	テーマと発表者の氏名、所属、目次、引用先、著作権に関する情報の大半が欠けている。	テーマと発表者の氏名、所属、目次、引用先、著作権に関する情報のほとんどが明示されていない。
スライドの構成	PPT は、「序論、本論、結論」や「SOS 法 (Summary, Details, Summary)」, 「PREP 法 (Point, Reason, Example, Point)」などの構成にわたって、論理的に組み立てられた。	PPT は、概ね「序論、本論、結論」や「SOS 法」、「PREP 法」などの構成にわたって、いくつもの部分でその流れが不明確または不足していた。発表を組み立てた。	PPT は、「序論、本論、結論」、「SOS 法」、「PREP 法」などの構成を用いてはなかったが、いくつかの重要なポイントに基づいて内容が組み立てられていた。	PPT は、「序論、本論、結論」、「SOS 法」、「PREP 法」などの構成をほとんど使わず、非論理的であった。
情報の正確さ ・引用先提示の有無 ・正確な情報の伝達	スライドで信頼性のある引用先を記載するなど、スライド上の情報が正確であり、聞き手も信頼性の高い、他者で自分の意見と区別するなど正確な情報を聞き手に伝えた。また、スライドや発言には聞き手が伝えない情報や誤解を生む情報が多かった。	一部のスライドで引用先を明らかにしていない情報があつたが、信頼性に乏しい引用先があつたが、概ね正確な情報を聞き手に伝えている。また、スライドや発言には必要な情報はほとんど伝えている。	スライド上で情報の引用先がなく、またスライドや発言内容に十分な正確な情報があつた。スライドや発言には聞き手に伝えない情報が入っていたが、いくつかが重要な情報や誤解を招く情報があつた。また、誤解を生むような発言もあつた。	スライドや発言内容が全くわからなかった。明らかに間違っている情報があつた。スライド上で情報の引用先が全くなく、かつ発言内容も正確性を欠いていた。
発表姿勢 ・体向き・視線等の姿勢や動作を適切に取るか	聞き手を常に意識して発表していた。聞き手の方を向いてはしゃび大きく、無声で大きな声で発表し、表情やジェスチャーなどの非言語的コミュニケーションも適切に用いて、聞き手の理解を助け、関与を促していた。	聞き手を意識する努力を示していたが、時折、手元の資料や PPT に目を落していた。表情やジェスチャーなどの非言語的コミュニケーションも適切に用いられ、聞き手の理解を促していた。	聞き手を意識しようとしたが、聞き手がよく見られなかった。全く聞き手の方を向かず、手元の資料や PPT を主に見ていた。また、声も小さく、全く聞き取れなかった。	聞き手に対して十分な配慮がなされなかった。全く聞き手の方を向かず、手元の資料や PPT を主に見ていた。また、声も小さく、全く聞き取れなかった。
発表時間 (発表時間 5分)	発表時間が指定時間内で終了した(4:超過-6分未満)。	発表時間が指定時間より少し短かった(3:超過-4分以下)。	発表時間が指定時間より少し長かった(6分以上-7分未満)。	発表時間が指定時間よりかなり長かった(7分以上)。
発表の進め方 ・挨拶・接続表現・敬語	発表前と終了時には礼儀正しく挨拶を行った。発表中の接続表現や敬語を適切に使った。	発表前と終了時には適切に挨拶を行った。しかし、発表中の接続表現や敬語の使用が不十分であった。	発表前と終了時には挨拶を行った。しかし、発表中の接続表現や敬語の使用が不十分であった。	発表前と終了時のおいざつを怠らなかつた。発表中に接続表現をほとんど使用せず、敬語を全く使わなかった。
言葉の豊富さ	豊富な言葉と多様な表現を巧みに用いて、自身の意図や発表の内容を明確に伝えた。また、敬語を多用し、丁寧な表現を用いた。	豊富な言葉と表現を使えなかったが、適切な言葉と表現を選んで、自身の意図や発表の内容を明確に伝えた。	限られた言葉と表現を使って、自身の意図や発表の内容を伝えることができた。しかし、時折、適切な言葉と表現を見つけれず、伝えたい内容をうまく表現できない場面があつた。また、時折、同じ言葉の繰り返し、回りくどい表現が見られた。	言葉や表現のバリエーションが乏しく、簡単な動詞や名詞などに頼り、言い直しが多く、発表の流れが断ち切られることがあつた。また、敬語を多用し、丁寧な表現を用いた。
単語と文法的正確さ	単語の読み方や文法的な誤りはなく、発表内容が明確に伝わり、完全に理解できた。	単語の読み方や文法的な誤りは少なく、発表内容の大部分は理解できた。	限られた単語と表現を使って、自身の意図や発表の内容を伝えることができた。しかし、時折、適切な言葉と表現を見つけれず、伝えたい内容をうまく表現できない場面があつた。また、時折、同じ言葉の繰り返し、回りくどい表現が見られた。	単語の読み方や文法的な誤りが頻りに見られ、発表内容の理解が困難な部分が多かった。
発音	非常に流暢に話せていた。発音やイントネーションの誤りはほとんどなく、母語の流暢さはほとんど感じられない。発表内容を完全に理解できた。	概ね流暢に話せていたが、若干の発音やイントネーションの誤り、母語の流暢さがあつた。また、発表内容はほとんど理解できた。	発音やイントネーションの誤りが頻りに見られ、発表内容の大部分が理解できなかった。	発音が不明瞭で、イントネーションの誤りが目立ち、母語の流暢さを欠き、大部分の内容を理解するのが困難であった。
流暢さ	スムーズに一定の速さで、止まらず話すことができて、自然な間やたまり、言い淀みはほとんどなく、内容の流れが保たれており、聞き手の理解を最大限にサポートした。	文法や言葉の間違いがあつた場面はあつたが、適切な内容の流れは保たれていた。言い淀みはほとんどなかった。	文法や言葉の間違いが多かった。また、内容の流れは保たれていなかった。	たまりや言い淀みが多く、発表の流れが断ち切られることがあつた。また、敬語を多用し、丁寧な表現を用いた。

5. 分析結果

5.1 予備調査の分析結果

予備調査の分析においては、1回目と2回目の授業を GPT-Talk 導入前（以降、導入前）のデータとして、3回目と4回目の授業を GPT-Talk 導入後（以降、導入後）のデータとして扱い、分析を行った。

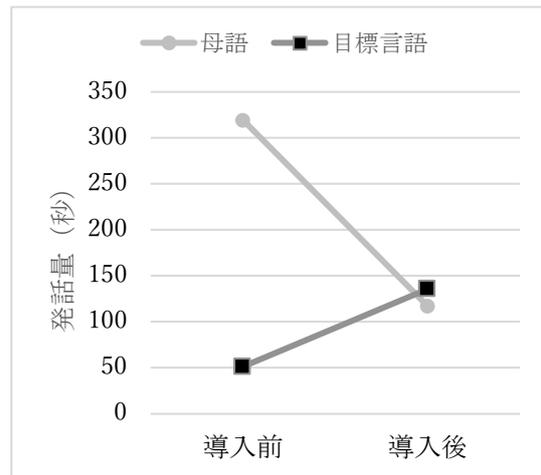


図 3 発話量の分析結果
(予備調査)

の平均は 135.7 ($SD=80.96$) 秒に増加した。導入後、母語の発話量は顕著に減少したのに対し、目標言語の発話量は明確に増加したことが示された (図 3)。

次に、PPT の完成度を分析した。その結果を表 5 と図 4 に示した。導入前後の平均値に統計的な差があるか否かを明らかにするために、対応のある t 検定を行った。その結果、「スライドデザイン ($t(31) = 2.49, p < .05$)」、「枚数 ($t(31) = 2.10, p < .05$)」の項目において有意な差が見られた。

さらに、PPT の資料内容を分析した結果、導入前の PPT では多くのスライドに写真が使用されていたが、これらの写真に対する適切な解説が欠如していた。そのため、視覚情報のみ依存した内容の理解は困難であった。加えて、写真の多用によりスライドの枚数が多くなっていた。一方、GPT-Talk 導入後は写真が提供されないため、生成された内容は文字情

報のみとなり、デザインがシンプル化され、写真を使用しないことからスライドの枚数が減少した。文字数の変化に関しては、導入前後で大きな差は見られなかったが、導入前には長文や文体の混在が多く見られたのに対し、導入後には箇条書きや短文の使用が増加したことが明らかになった。

表 5 PPT の作成の分析結果（予備調査）

	導入前	導入後
スライドのデザイン	$M=2.6$ ($SD=.5$)	$M=2.4$ ($SD=.5$)
スライドの枚数	$M=2.4$ ($SD=.9$)	$M=2.1$ ($SD=.9$)
スライドの文字数	$M=2.4$ ($SD=1.0$)	$M=1.9$ ($SD=.9$)
スライドの形式	$M=1.0$ ($SD=.0$)	$M=1.0$ ($SD=.0$)
発表構成	$M=2.1$ ($SD=.5$)	$M=2.0$ ($SD=.8$)
情報の正確さ	$M=2.5$ ($SD=.5$)	$M=2.5$ ($SD=.5$)

注：M は平均値（Mean）、SD は標準偏差（Standard Deviation）。

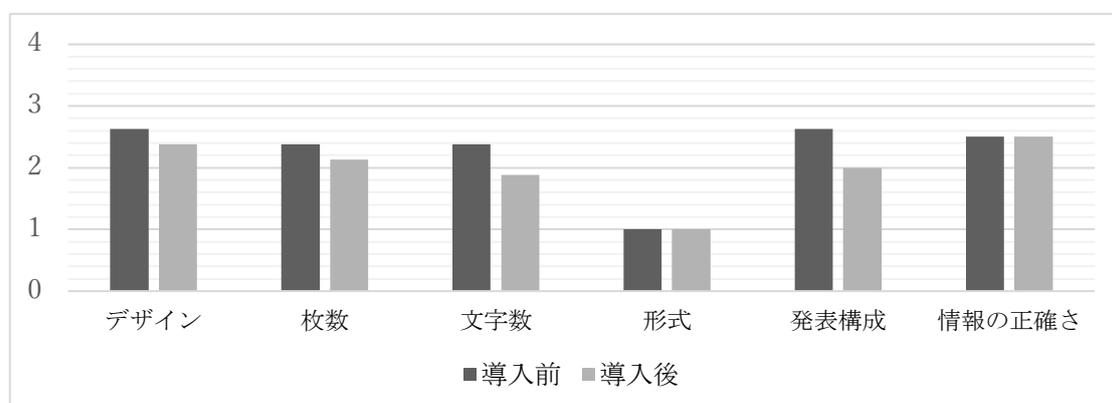


図 4 PPT の作成の分析結果（予備調査）

続いて、プレゼンのパフォーマンスに関する分析を行った。その結果を表 6 および図 5 に示した。導入前後の平均値に統計的な差があるか否かを明らかにするために、対応のある t 検定を行った。その結果、「発表時間 ($t(31) = 4.21, p < .01$)」、「単語および文法的正確さ ($t(31) = 3.49, p < .01$)」の項目に

において有意な差が見られた。

これらの結果から、導入後はスライドの枚数が減少し、発表時間が3分の2に以下となり、早く終わった学習者が数名いた。導入前は写真のみを用いたスライドが多かったため、発表では学習者がスピーキング能力を即興的に発揮するしかなかった。しかし、導入後は各スライドに短文や箇条書きの内容が記載されているため、学習者はスライドに書かれた内容を参考にしてプレゼンを行うことが可能となった。この結果、単語の適切な使用および文法的正確さが向上したことが確認された。

表 6 プレゼンのパフォーマンスの分析結果（予備調査）

	導入前	導入後
発表姿勢	$M=2.5$ ($SD=.8$)	$M=2.5$ ($SD=.7$)
発表時間	$M=2.3$ ($SD=1.4$)	$M=1.5$ ($SD=.9$)
発表の進め方	$M=2.3$ ($SD=.7$)	$M=2.4$ ($SD=.8$)
言葉の豊富さ	$M=2.5$ ($SD=.7$)	$M=2.6$ ($SD=.7$)
単語と文法的 正確さ	$M=2.5$ ($SD=.6$)	$M=2.8$ ($SD=.4$)
発音	$M=2.9$ ($SD=.6$)	$M=2.9$ ($SD=.5$)
流暢さ	$M=1.8$ ($SD=.7$)	$M=1.8$ ($SD=.6$)

注：Mは平均値（Mean）、SDは標準偏差（Standard Deviation）。

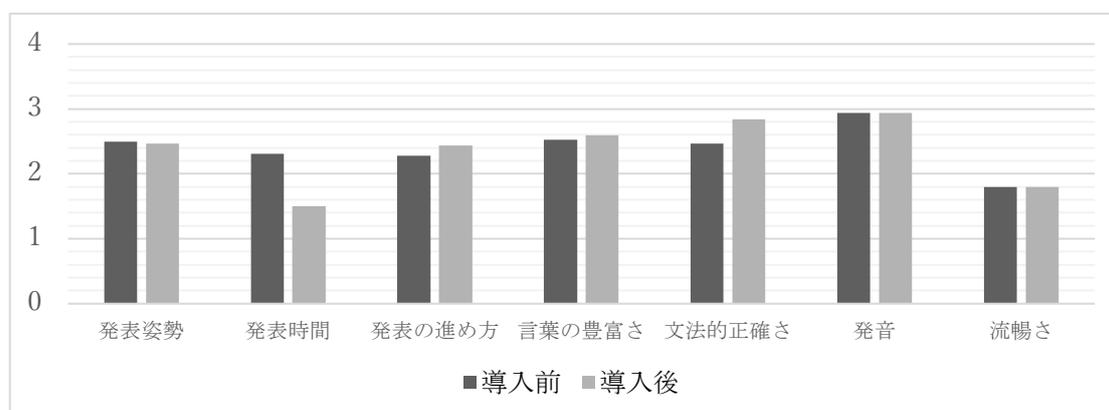


図 5 プレゼンのパフォーマンスの分析結果（予備調査）

予備調査の結果、GPT-Talkの導入が目標言語の発話量の増加に寄与し、その結果、母語の発話量を上回るという結果に至ったことが明らかになった。また、PPTのデザインはシンプル化され、スライドの枚数が減少した一方、箇条書きや短文の使用は増加した。発表時間は短縮されたものの、文法的な正確さや整合性は向上した。これらの結果から、自由会話システム GPT-Talk の導入が学習者の言語使用およびプレゼン能力に肯定的な影響を与えたことが示唆される。

しかし、副題の選定とPPTの作成に際して、学習者間の発話が活発に行われなかったことと、総発話量が多くなかったことが明らかになった。学習者には情報の検索や資料作成において、主に個別に作業を進める傾向が多く観察された。この点は課題として認識し、本調査においては改善を図る必要がある。

5.2 本調査の分析結果

予備調査の結果および残された課題に基づき、本調査では以下の4点について変更を加えた。

1. プレゼン授業の回数を4回から2回に減らした。
2. 予備調査が一斉授業の形式で実施されたのに対し、本調査では1対2(教師1名、学習者2名)の形式で行った。
3. プレゼンテーションの題目を、学習者にとって身近な話題から、背景知識が少ない話題へと変更した。
4. PPTの制作に際しては、写真の使用を禁止した。

本調査の発話量を分析した結果、導入前は母語の発話量の平均は685.1($SD=301.2$)秒であり、目標言語の発話量の平均は114.9($SD=80.2$)秒であった。しかし、導入後、母語の発話量の平均は327.3($SD=163.2$)秒に減少し、一方で目標言語の発話量の平均は400.7($SD=144.5$)秒に増加した。

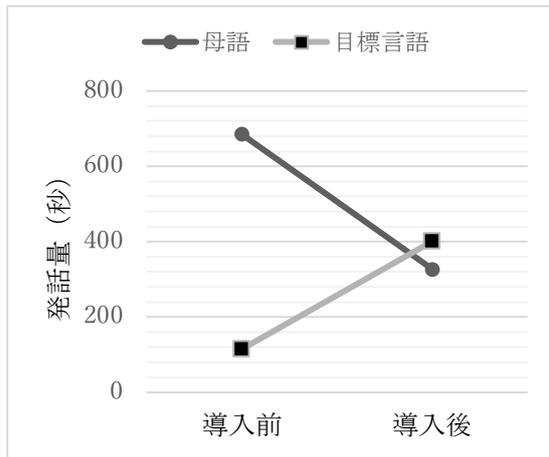


図 6 発話量の分析結果
(本調査)

そして、導入前後の総発話量の平均値に統計的な差があるか否かを明らかにするために、対応のある t 検定を行った。その結果、導入前後の総発話量に関しては、($t(6)=0.596, n.s.$) 有意な差は見られなかった。これらの結果から総発話量は大きな変動が見られなかったものの、GPT-

Talk の導入が目標言語の発話量を顕著に増加させ、母語の発話量を減らす結果となった (図 6)。

次に、PPT の作成に関する分析を行った。その結果は表 7 および図 7 に示した。導入前後の平均値に統計的な差があるか否かを明らかにするために、対応のある t 検定を行った。その結果、「枚数 ($t(6) = 2.50, p < .05$)」、「文字数 ($t(6) = 2.83, p < .05$)」、「構成 ($t(6) = 3.87, p < .01$)」の項目において有意な差が見られた。

そして、PPT の資料内容を分析した結果、導入前のスライドではキーワードのような単語羅列が主であったが、導入後には箇条書きや体言止めを用いた記述が増えた。この変化から、導入後はスライドの枚数と 1 枚あたりの文字数が増加したことが明らかになった。さらに、構成面では、「問題提起」「本題」「解決策」といった明確な構造を持つ PPT が増えたり、各スライドの目的がより明確になったりしたことが示された。

表 7 PPT の完成度の分析結果（本調査）

	導入前	導入後
スライドデザイン	$M=2.7$ ($SD=.5$)	$M=2.9$ ($SD=.3$)
スライドの枚数	$M=2.1$ ($SD=.6$)	$M=2.9$ ($SD=.6$)
スライドの文字数	$M=2.4$ ($SD=.5$)	$M=3.0$ ($SD=.0$)
スライドの形式	$M=2.6$ ($SD=.5$)	$M=2.1$ ($SD=.3$)
発表構成	$M=2.0$ ($SD=.5$)	$M=2.7$ ($SD=.5$)
情報の正確さ	$M=2.7$ ($SD=.7$)	$M=2.1$ ($SD=.3$)

注：M は平均値（Mean）、SD は標準偏差（Standard Deviation）。

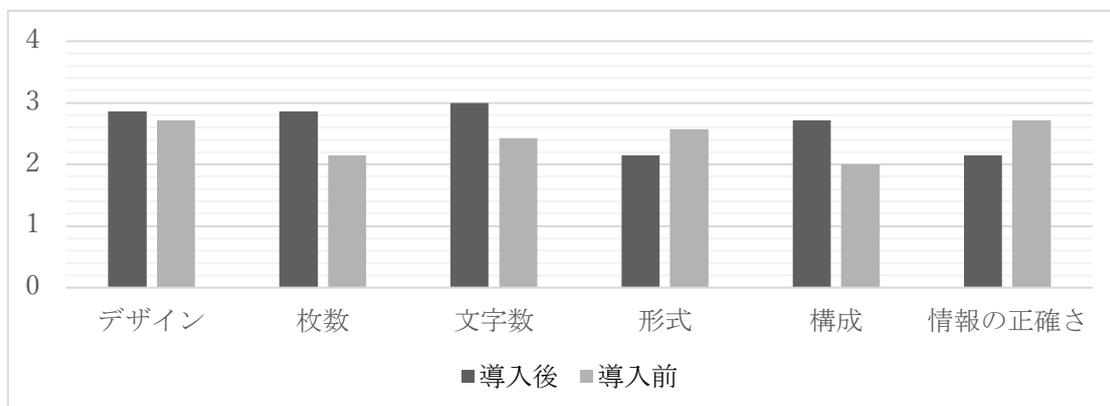


図 7 PPT の作成の分析結果（本調査）

続いて、プレゼンのパフォーマンスに関する分析を行った。その結果を表 8 と図 8 に示した。導入前後の平均値に統計的な差があるか否かを明らかにするために、対応のある t 検定を行った。その結果、「流暢さ」の項目においてのみ有意な差が見られた ($t(6)=2.83, p < .05$)。

この結果から、GPT-Talk の導入によって PPT の制作時間が短縮され、発表の練習時間が確保できるようになった。また、言い淀みやフィラーの減少が発表の流暢さの向上に寄与した。ただし、プレゼン能力（発表姿勢、進め方）の養成と、言語能力（語彙の豊富さ、単語と文法的正確さ、発音）の向上には、教育的介入は必要であり、長期的かつ継続的な指導が求

められる。

表 8 プレゼンのパフォーマンス分析結果（本調査）

	導入前	導入後
発表姿勢	$M=3.3$ ($SD=.7$)	$M=3.3$ ($SD=.7$)
発表時間	$M=2.6$ ($SD=1.0$)	$M=3.1$ ($SD=1.0$)
発表の進め方	$M=3.4$ ($SD=.7$)	$M=3.6$ ($SD=.7$)
言葉の豊富さ	$M=3.7$ ($SD=.5$)	$M=3.9$ ($SD=.3$)
単語と文法的 正確さ	$M=2.6$ ($SD=.5$)	$M=2.7$ ($SD=.5$)
発音	$M=2.7$ ($SD=.7$)	$M=2.7$ ($SD=.7$)
流暢さ	$M=2.0$ ($SD=.8$)	$M=2.6$ ($SD=.7$)

注：Mは平均値（Mean）、SDは標準偏差（Standard Deviation）。

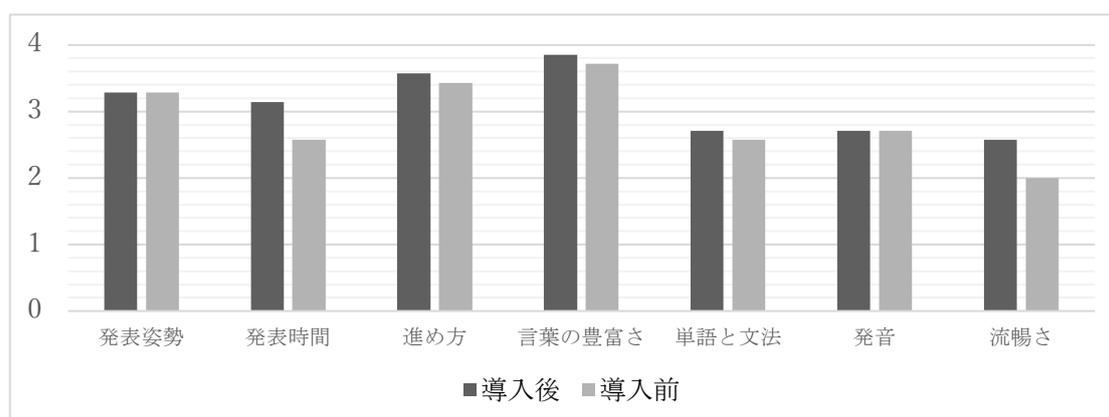


図 8 プレゼンのパフォーマンスの分析結果（本調査）

5.3 まとめ

本研究では、GPT-Talkが教育上に与える可能性を「目標言語の発話量」「PPTの作成」「目標言語使用」の3つの側面から分析した。「目標言語の発話量」においては、予備調査と本調査の結果が一致しており、GPT-Talkの使用により学習者の目標言語における発話量が増加することが確認された。一方、「PPTの作成」「目標言語使用」に関しては、予備調査と本調

査の一部の結果で異なる結果が観察された。これは、学習者の習熟度、PPT作成スキル、プレゼン能力に依存すると考えられる。対象とした学習者は、PPT資料作成に関する知識が不足していることが明らかになった。また、授業においてプレゼンを一度も経験していない学習者が多かったことも判明した。

6. 結論と考察

6.1 結論

本研究では、第二言語の産出に焦点を当て、大規模言語モデルに基づく生成型 AI を利用した対話型 AI 音声会話システム「GPT-Talk」を開発し、プレゼン授業におけるその使用に関して検討した。学習者の「書く」「話す」に関する産出を考察した結果、以下の3点が明らかになった。

- I. 「目標言語の発話量」に関する課題において、GPT-Talk の導入は目標言語での発話量を促進する効果があった。
- II. 「PPT制作」に関する課題では、スライドデザインが単純化される一方で、スライドの枚数と文字数、およびスライドの構成が学習者の習熟度や PPT 作成スキルによって異なる結果が見られた。しかし、全体的に学習者は PPT 資料作成に関する知識が不足していることが示された。内容の提示方法に関しては、長文や単語の羅列から、短文や箇条書きへの改善が見られた。
- III. 「目標言語使用」に関する課題では、中級学習者においては、語彙の適切な使用と文法的正確さが向上した。上級学習者においては、言い淀みやフィラーの減少と言語の流暢性の向上が見られた。

6.2 考察

本研究では、対話型 AI の教育への適用可能性をプレゼン授業における実践的な研究を通じて検討した。GPT-Talk は音

声を介してプロンプトの作成を行うため、学習者は目標言語を使用せざるを得ない状況に置かれ、これが目標言語の使用を促す要因となった。しかし、実際の発話は一問一答に留まり、話題の展開には限界があるという課題が残った。

PPT制作に関しては、GPT-TalkがChatGPTを中核としているために、実践時では写真・図・グラフの提供が不可能であり⁴、結果としてPPTに画像を含めることができなかった。これにより、スライドデザインは文字中心となり、結果的にはシンプルなものになった。ChatGPTの機能が今後拡張されれば、これらの課題を克服できる可能性があると考えられる。

予備調査と本調査では、主な対象者の日本語習熟度が異なるため⁵、「PPTの枚数」「文字数」「構成」において異なる結果が観察された。

「PPTの枚数」については、中級学習者は写真を多用する傾向があったが、ChatGPTの導入により、写真の使用が不可能となったため、スライドの枚数が減少した。また、導入前には中級学習者は検索した情報をそのままPPTに転載する傾向があったが、写真使用の制限により、中級学習者が作成したPPTでは多量の写真が使用されることはなくなった。しかし、GPT-Talkの効率的な情報整理により、PPTの枚数は増加したことが分かった。

「文字数」に関しては、中級学習者は導入前や導入後を問わず、検索した内容をそのままスライドに転載していた。ただし、内容を長文から短文や箇条書きへと変更したため、文字数には大きな変化は見られなかった。一方、上級学習者は導入前には情報を単語の羅列で提示する傾向にあったが、導

⁴ 現在「ChatGPT4」 Assistants API では、プロンプトから画像生成ができる。

⁵ 予備調査での主な対象者は中級学習者（JLPT N2 レベル）であり、本調査では上級学習者（JLPT N1 レベル）を主に対象とした。

入後は短文や箇条書きを用いるようになり、結果として文字数が増加した。

「構成」においては、上級学習者は GPT-Talk を活用して情報を素早く簡潔にまとめ、PPT 資料作成時間を短縮し、PPT の構成を論理的なものに改善した。中級学習者は GPT-Talk によって生成された情報の読み取りに時間を要し、情報の消化にエネルギーを消耗してしまい、PPT の構造改善には至らなかった。

「目標言語使用」においては、流暢性は PPT 資料作成の効率化により確保された発表練習時間の増加に伴って向上し、言い淀みやフィラーの減少が見られた。これにより、発表の流暢性が改善され、学習効果が即座に示された。しかし、発表時の視線や姿勢といった非言語的コミュニケーション能力、さらには語彙の豊富さ、文法的正確さ、発音といった言語能力の向上に関しては、教育的介入が必要であると判断された。

「PPT 作成」と「目標言語使用」の分析からは、全体的に学習者は PPT 作成とプレゼンに関する知識が不足していることが明らかになった。導入前後でスライドの枚数が増減することから、学習者は「スライドの枚数を発表時間に応じて決定する」という概念に関しては不明瞭であることが示された。また、導入前後で 1 枚当たりの文字数が増加したことは、スライド作成における情報量の適切な管理に関する知識の欠如を示している。プレゼン能力に関しても、発表の際の視線や体の向きなど、非言語的コミュニケーションの指導を受けていない様子が観察された。

SAMR モデルに基づけば、本研究は「変形」レベルに達しているといえる。本研究は生成型 AI 技術を基にし、プレゼンテーション授業および発表・質問応答の授業活動を通じて、学習を深めるものである。既存の生成型 AI を使用した研究と比較して、本研究は学習者間の協働活動および表現活動を

積極的に支援する授業設計であると評価できる。

6.3 教育への示唆

本研究では、ChatGPTを活用したプレゼン授業を実施した。その結果、学習者の「書く」能力と「話す」能力の向上に対する効果が明らかになった。また、GPT-Talkがプレゼン授業に適用可能であり、生成型AIが教育上の応用において大きな可能性を持つことが示された。しかし、生成型AIの使用においては、情報源の透明性に関する問題が存在するため、その使用にあたっては慎重な姿勢が必要である。さらに、生成型AIの情報提示方式において、従来の検索エンジン(Google、Microsoft Bing等)との間に顕著な差異が観察される。生成型AIは、一つの質問に対して、膨大なデータベースから選択された数個の答えを簡潔かつ箇条書き形式で提示することで、情報の選別に要する時間を劇的に短縮している。これに対して、従来の検索エンジンは、一つの質問に対して関連性のある情報を網羅的に表示し、ユーザー自身がその中から必要な情報を選び出して整理する必要がある。この違いは、情報処理の効率性における生成型AIの優位性を示している。本研究では、学習者の思考力や情報の統合・分析能力の育成についての考察は対象外であったが、生成型AIの導入による情報検索の利便性向上が明らかにされた。一方で、膨大な情報からの取捨選択、統合、分析を省略する学習行動が引き起こされることによる潜在的なリスクも指摘されるべきである。

生成型AIの急速な発展と、各産業におけるその積極的な導入の現状を踏まえると、言語教育分野においても、ただ危惧するだけではなく、これからどのように生成型AIと向き合っていくかを再考する必要があると考えられる。

参考文献

日本語

- 稲葉みどり (2011)「アウトプット重視の日本語授業の構想創り—自己紹介のプレゼンテーション作成と発表—」『教養と教育』(11)、愛知、愛知教育大学、pp.1-8
- 王睿琪(2024)「第二言語における LLM-GAI の産出促進効果」『日本語教育方法研究会誌』30(2)、東京、日本語教育方法研究会、pp.30-31
- 夏逸慧 (2023)「日本語作文教育における ChatGPT 添削の活用」『日本語教育方法研究会誌』30(1)、東京、日本語教育方法研究会、pp.112-113
- 孫彤 (2024)「LLM-GAI を用いた初中級日本語教材の開発」『日本語教育方法研究会誌』30(2)、東京、日本語教育方法研究会、pp.128-129
- 寺嶋弘道 (2024)「ChatGPT によるレポートへのフィードバックとその活用に関する一考察」『日本語教育方法研究会誌』30(2)、東京、日本語教育方法研究会、pp.138-139
- 寺嶋弘道・稲田栄一・坂井芳江・隈井正三 (2023)「ChatGPT による作文のフィードバックとその活用の試み」『日本語教育方法研究会誌』30(1)、東京、日本語教育方法研究会、pp.68-69
- 潘寧 (2022)「AI チャット bot サービス「エアフレンド」を活用した授業実践—台湾の大学における初級日本語会話授業から—」『日本語教育方法研究会誌』29(1)、東京、日本語教育方法研究会、pp.102-103
- 村野井仁 (2011)「アウトプットと第二言語」『東北学院大学論集 英語英文学』(95)、宮城、東北学院大学、pp.51-64
- 劉玲玲 (2024)「AI 歌教材を利用した言語学習の効果の調査」『日本語教育方法研究会誌』30(2)、東京、日本語教育方法研究会、pp.120-121

英語

PuenteDura, R.R. (2010) A Brief Introduction to TPCCK and SAMR

<http://www.hippasus.com/rrpweblog/archives/2011/12/08/BriefIntroTPCKSAMR.pdf> (2024年3月7日閲覧)

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Cambridge, MA: Newbury House.

Swain, M. (1995). *Three functions of output in second language learning*. In G. Cook, & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson* (pp. 125-144). Oxford: Oxford University Press.