

LLM を活用した返答提案機能を備えた
多言語コミュニケーション支援システムの提案
A Multilingual Communication Support System
with Response Suggestion Functionality using Large Language Models

古澤 慧音[†] 金 智恩[†] 別所 正博[†]
Keito Furusawa Jee-eun Kim Masahiro Bessho

1. はじめに

グローバル化が進展する現代社会では、ビジネス、教育など多岐にわたる分野で言語の壁を超えたコミュニケーション能力が求められる。しかし、日本では十分な英語教育が行われておらず[1]、日本の英語能力指数は 113 カ国中 87 位である[2]。T. Konishi らによれば、国際的な人的交流は経済取引を促すとされており[3]、現在の状況は日本の国際競争力に悪影響を与え、経済発展の妨げになっている可能性がある。

多言語コミュニケーションを支援する翻訳ツールはすでに普及しているが、既存のツールは不自然な翻訳を出力することがあり[4]、会話の文脈や場面などのコンテキストに応じた翻訳が不十分である。さらには、言語だけでなく文化的ニュアンスの理解が重要であり、翻訳ツールは直訳以上の情報を提供する必要がある。

そこで本研究では、生成 AI のひとつである大規模言語モデル(LLM)に着目した。その中でも、Open AI が開発したトランスフォーマーベースのモデルである GPT[5]は事前学習の過程で文脈の理解、文法的構造の認識[6]、そして豊富なパターンの異言語間翻訳を実現する能力を獲得している[7]。そのため、従来のニューラル機械翻訳よりも広い文脈を認識し、対話全体の意図を理解することができる。これは従来の翻訳ツールでは不可能だった、文脈に応じた翻訳を提供するための重要な要素である。

本研究の目的は、英語が不得意な日本人が、英語話者と円滑にコミュニケーションを取れるようにすることである。この目的を達成するために、LLM によって適切な返答案を提案する機能を備えた、インタラクティブな翻訳支援システムを提案する。提案システムは、英語話者の発話をリアルタイムに日本語訳し、さらにそれに対する複数の返答案を英語と日本語の両言語で表示する。これにより、ユーザはサポートを受けながら英語話者と会話することができる。

本研究では、提案システムの有用性を評価するため、13 人の大学生を対象とした実験を行った。実験ではレストランで外国人を接客する場面を想定し、英語の音声に対して返答するロールプレイを行い、提案システムの有無で返答内容や応答速度にどのような差異がみられるかを記録した。加えて、実験後にシステムの使用体験に関するアンケート調査を行った。

実験結果から、提案システムはユーザの聞き返しを減らし、妥当な答えを増やすことが明らかとなり、自然な会話を支援する可能性を示した。また、アンケート調査を通じ

てシステムのユーザビリティと実用性に関して、良好な評価が得られた。

本論文では、2 節で関連研究を概観したのち、3 節でシステムの提案を行う。4 節で評価実験に関して説明し、5 節で考察を行った後、6 節で結論と今後の展望を述べる。

2. 関連研究

本研究では、既存の翻訳支援システムと、LLM を活用したコミュニケーション支援について、先行研究を調査した。

翻訳支援の分野では、数多くの研究がなされているが、近年では文脈を踏まえた翻訳に向けた取り組みも多くなされている。福島らは、医療従事者と外国人患者のコミュニケーション支援のための手法として、事前に作成された用例対訳と機械翻訳を組み合わせた方法を示している[8]。Sameen らは、バイリンガル多人数会話の翻訳において、ソースとターゲットの両方の会話履歴を活用する手法[9]により BLEU スコアの向上を示した。Agrawal らは、自己注意に基づく非再帰的アーキテクチャである Transformer を使用して、前後の文を含む文脈ウィンドウをソース側とターゲット側の文脈として組み込み、より効果的なニューラル機械翻訳 (NMT) を実現するアプローチを取った[10]。これらの研究では、自由形式対話での限界や、データセットの規模の問題、関連性のない情報の流入などの課題があったが、近年では LLM を用いることで、より自然な形で文脈に応じた翻訳を実現することが期待されている[11]。

昨今、LLM を活用して実際の対話をサポートする研究が盛んになされている。Yuyang Fang らは、LLM 活用した聴覚障がい者を支援するシステム[12]を開発した。Park らは LLM を活用したスピーチ練習アプリケーションを開発した[13]。このシステムはより多様な聴衆の人物像を生成し、スピーチのリアルタイムフィードバックを提供する。Fangkai Yang らは LLM を用いて人間同士の会話型質疑応答をシミュレートする手法[14]を示した。提案手法は、会話のダイナミクスを効果的にシミュレートし、文脈的に適切で一貫性がある対話を生成できることを示している。

これらの先行研究は、LLM を活用する手法がコミュニケーション支援において有効であることを示している。本研究では、LLM を活用し、より複雑なコンテキストに対応した、提案型の多言語コミュニケーション支援システムの実現を目指す。

3. 提案

3.1 システムの概要

[†] 東洋大学 大学院情報連携学研究所 Toyo University
Graduate School of Information Networking for Innovation and Design

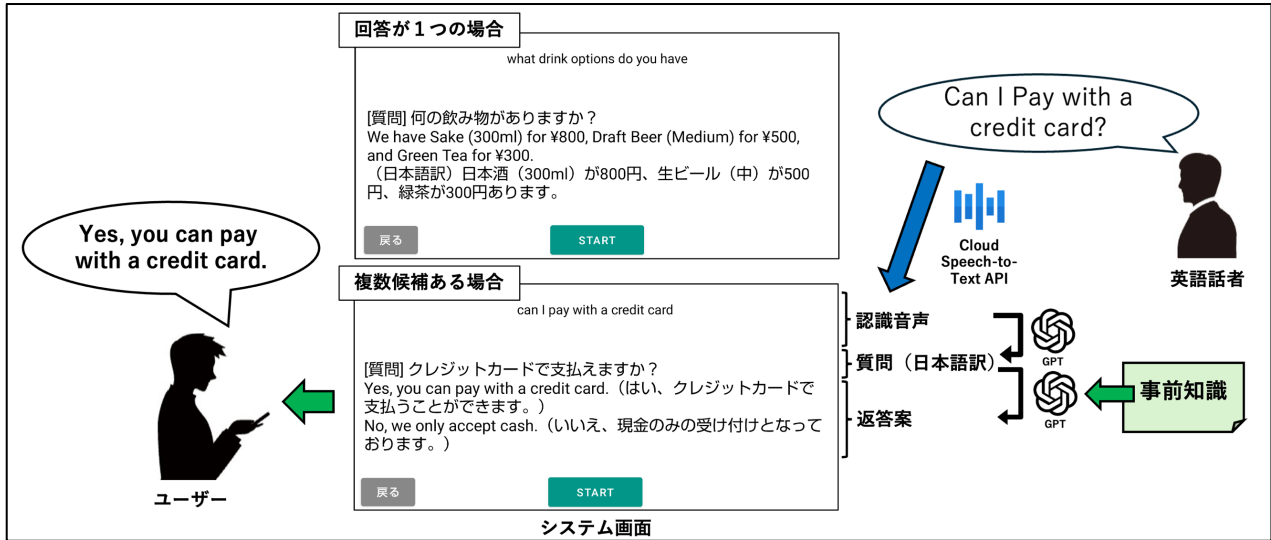


図 1 システムの動作例と使用例

GPT、以下の指示に従い対応することをお願いします：

1. 第一に、英語の質問のその意味を日本語に訳したものを提示すること(例:[質問] こんにちは(改行))。第二に、その英語の質問に対して簡潔な返答を提供すること。
2. 返答は英語コミュニケーションを支援するため、適切でなければならない。
3. 各返答には、その日本語訳も併記すること。英語と日本語の行は必ず分けて出力すること。
4. スマートフォンの画面表示に配慮し、短く情報過多にならないような返答を心がけること。
5. Yes/No で答えられる質問には、肯定的 (Yes) と否定的 (No) 両方の立場からの返答を提供すること。例：“Do you like chocolate?” に対しては “Yes, I love chocolate!” (はい、チョコレートが大好きです!) と “No, I`m not fond of chocolate.” (いいえ、チョコレートはあまり好きではありません。) のように答える。
6. 答えが一様に決まっている場合でも、肯定/否定の両サイドでの答え方を出力せよ。必ず日本語訳も出力すること。
7. 答えが複数考えられる場合には、それらをそれぞれ簡潔に出力すること。また必ずそれぞれについて日本語訳を出力すること。
8. 提出された英語文の話し手が大人か子供かを判断し、それに適したトーンを選択すること。
9. 回答できない、または関連性がない質問に対しては、標準的なフレーズを避け、より自然な人間の反応を示すこと。
10. 「AI」や「GPT」のような言葉を避け、人間の自然な会話のように表現を用いること。

上記の指示に基づいて、効果的なコミュニケーションが行えるよう準備してください。
 再度繰り返すが、複数の答え方があるものには、どちらの立場でも答えること。
 <以下、シナリオ用のロールプロンプトが続く>

図 2 指示プロンプト

本研究では、LLM を活用し、文脈に沿った返答提案機能をもつ、新しい多言語コミュニケーション支援システムを提案する。提案システムの重要な特徴は、LLM に前提知識を提供することで、可能性のある返答案を、複数提示する点にある。これにより、多言語でのコミュニケーションにおいて、日本人が英語での会話をスムーズに行うことを目指す。システムの利用イメージを図 1 に示す。

システムは片手で簡単に操作でき、さらに屋外でも通信できる媒体として、スマートフォンを採用した。提案システムの動作は次の 3 段階となる。第 1 に音声認識された発話内容がテキストデータに変換される。音声認識は START ボタンの長押しにより動作し、英語音声のみが抽出される仕組みを利用する。第 2 に指示と事前知識が書かれているプロンプトと発話内容を組み合わせ GPT に認識させる。第 3 に GPT から返答された内容を表示する。返答には質問の日本語訳と、それに対する返答案が含まれる。返答案は、日本語と英語の両言語で表示される。このよう

にすることで、相手の発話内容に対する適切な返答案を日本人が理解できる形で提供することを実現する。

3.2 プロンプト

翻訳と返答案のインタラクティブな生成を実現するために、本システムでは図 2 に示すようなプロンプトを採用した。相手の発話の日本語訳と、会話相手に自然な印象を与えるように文脈に合わせた返答案を、日本人が理解しながら発言できる形で、生成することを指針とし、理想の形になるように試行錯誤を繰り返し、不自然な出力がないように調整した。プロンプトの構成は、基本的な指示プロンプトと、シナリオに合わせてチューニングするロールプロンプトに分かれている。指示プロンプトでは、各指示を簡潔かつ明瞭にするために箇条書きで情報を整理し、プロセスの明確化を図った。さらに、プロンプトの処理を高速化し日本語での正確な返答を実現するために、各指示文を日本語の短い文で記述した。また、複雑な指示においては、具体例を示すことで内容を正確に把握させている。最後に、

「AI」や「GPT」のような直接的に LLM の存在を示唆する言葉の使用を避け、より人間らしい返答の生成を心がけるようにした。ロールプロンプトには、利用する場面に合わせた指示文を挿入することで、その場面で求められる翻訳サポートを提供することが可能である。

3.3 実装

提案システムは Android アプリケーションとして実装した。音声認識には、Google Cloud Speech to Text API[15]を使った。本研究では英語の会話における音声認識に特化するため、言語の設定を en-us(アメリカ英語)にした。これにより、録音された音声から英語の発音のみが抽出される。

翻訳を生成する言語モデルは、可能な限りリアルタイムにレスポンスをユーザに提供する必要があることから、Updated GPT-3.5 Turbo[16]を採用した。temperature パラメータは 0.5 に設定した。

4. 評価実験

4.1 実験の目的・概要

本研究では、LLM を用いたリアルタイム翻訳支援システムの有用性を検証するために実験を行った。この実験では、外国人に対する英語での接客を想定したロールプレイを通じて、提案システムが日本語話者の英語での返答速度、返答の適切さ、聞き返し回数にどの程度影響を与えるかを検証した。

実験参加者は学生 13 人で、実験は比較的騒音の少ない環境下で行った。実験で用いるシナリオは日本食レストランでの接客を想定して作成した。このシナリオに沿って、参加者に英語での接客を行ってもらい、提案システム使用の有無それぞれでデータを収集した。

4.2 実験の手順

はじめに、システムをシナリオに最適化するためのロールプロンプトとして、レストランで接客する日本人スタッフをサポートする旨を含んだ指示と、レストランのメニューを商品名、価格をそれぞれ列挙する形で設定した。同様の内容を参加者にもあらかじめ提供し、極力自然な接客を行うように指示した。

次に、ロールプレイにおいて、Google Cloud Text-to-Speech[17]によって生成した音声をノート PC で再生し、質問に答える形式で接客を行わせた。参加者が応答する質問の数はアプリの有無でそれぞれ 5 つずつとした。質問の内容は表 1 に示す通りである。Yes/No で回答可能な質問を 4 つ、複数の回答が考えられる質問を 4 つ、マニュアルに含まれない質問を 2 つ用意した。そのほかに、聞き返しの音声を 5 つ、表 1 に含まれるもの以外で音声確認、アプリの動作チェック用の音声を 4 つ用意した。合成音声の生成設定は、アメリカ英語の音声で、Voice name は en-US-Studio-O(女性)と en-US-Studio-Q(男性)の二種類を均等に割り当てた。また、ロールプレイに入る前には、音量の調節をし、参加者が問題なく聞くことができることを確認した。質問の再生は著者が行い、回答者が不明確な返答をした場合は聞き返しの音声を再生した。

最後に、実験データの収集は、ロールプレイ中の録音、システムの動作中のログを基に行った。録音データから、返答までの時間、内容の正確性、聞き返し回数を計測し、

表 1 質問一覧

1	Is the Assorted Sushi (Deluxe) made with fresh fish?
2	Can I have the Chicken Teriyaki as a vegetarian option?
3	Do you serve alcohol to underage customers?
4	Is the Matcha Ice Cream homemade?
5	What options do I have for Ramen flavors?
6	Can I choose different types of Yakitori, and if so, which ones?
7	For the Sashimi, what kinds of fish are included?
8	What drink options do you have?
9	Can I pay with a credit card?
10	Is there a parking area nearby?
	Could you explain that one more time, please
	Excuse me, could you speak a little slower, please?
	I'm sorry, could you please repeat that?
	I'm sorry, what was that last part again?
	Pardon me, I didn't catch that. Could you say it again?

システムが音声入力から返答表示に要した時間を動作ログから記録した。また、実験後にはシステムの使用感についてアンケートを行った。実験では、AQUOS sense4[18]を用いてシステムを動作させた。

実験の評価は、返答速度、返答の適切さ、聞き返し回数に対して行った。加えて、実験参加者に対するアンケートを通じて、システムのユーザビリティ、音声認識の適切さ、翻訳サポートの適切さ、接客のしやすさ、そして英語への自信の変化を評価した。これらの評価を通じて、翻訳支援システムが実際のコミュニケーションに与える影響を評価した。

4.3 実験の結果

質問に対して返答を開始するまでに要した時間を「返答速度」、質問に対して参加者が聞き返した回数を「聞き返し回数」、質問に対して文脈上破綻していない適切な返答を「有効返答」とし、それぞれ表 2 にまとめた。

まず返答速度について、アプリ使用時の平均返答速度は 9.24 秒、使用しない時は 7.21 秒と計測された。ウェルチの T 検定の結果、P 値が 0.056 であり有意水準 0.05 を超えているため、統計的に有意差は認められないが、アプリ使用時に返答が遅くなる傾向を示唆している。

次に聞き返し回数について、アプリを使用した場合と使用しない場合で、聞き返しの頻度に顕著な違いが確認された。アプリ使用時の合計聞き返し回数の平均は約 0.385 回であり、使用しない場合は約 2.692 回である。T 検定の結果、P 値は 4.71×10^{-6} となり、アプリ使用により聞き返しが有意に減少することが確認された。

最後に有効返答数の平均値は、アプリを使用した場合は約 4.62 個であり、使用しない場合は約 3.08 個となった。T 検定の結果、P 値は約 1.941×10^{-5} であり、アプリを使用した場合の有効返答数は、使用しない場合に比べて有意に向上することが確認された。

表 2 実験結果

No.	返答速度(平均)		聞き返し回数		有効返答数	
	A	B	A	B	A	B
1	7.508	9.956	2	0	2	5
2	8.991	6.04	5	0	4	4
3	5.049	11.076	4	0	2	5
4	10.086	7.218	2	0	3	4
5	19.293	14.14	5	1	3	4
6	5.048	8.697	0	0	3	5
7	7.955	12.504	2	1	5	5
8	2.276	5.935	1	0	3	5
9	6.245	13.366	3	0	1	4
10	5.248	4.813	3	2	4	5
11	7.662	6.497	7	0	3	5
12	3.983	7.618	1	0	4	4
13	4.32	12.251	0	1	3	5
平均	7.205	9.239	2.692	0.385	3.08	4.62

A: システム未使用時 B: システム使用時

また、実験後に行ったアンケートにおいては、システムのユーザインタフェース (UI) の直感的な使いやすさについては、肯定的な評価が 100%であった。音声認識の精度が十分であったと回答した参加者は 92.3%であった。また、システムのサポートが適切であったと感じた参加者も 100%で、システムを使用することで英語に対する自信が向上したと感じた参加者は 69.2%であった。

5. 考察

本研究の提案する多言語コミュニケーション支援システムの評価実験では、いくつかの肯定的な結果が得られた。特に、システム使用によって聞き返し回数が大幅に減少し、システムの翻訳機能の有効性が明確に示された。音声の誤認識が発生しても翻訳精度は 97.7%と高く、GPT が返答の段階で軌道修正を行うことが観察された。この点は、システムの翻訳機能が音声認識ミスを補正し、有効な返答を生成できる能力を持つことを示している。また、返答の妥当性に関しては、システム使用時の有効返答数が未使用時に比べて有意に多く、スムーズで自然な返答が増えたことが確認された。アンケート調査の結果、システムの UI の直感的な使いやすさや音声認識の高い精度、システム使用による英語への自信の向上が示された。一方で、返答速度に関しては有意な改善が見られなかった。システムのレスポンス速度も重要な影響を及ぼしており、GPT へのメッセージ送信から翻訳表示までの平均時間が約 2.91 秒であったことが返答時間の増加の原因と推察される。また、質問を理解できなかった参加者は適切ではない回答を素早く返答する傾向にあったが、これは時間が短くなった一因と考えられる。反対に、質問を理解した参加者は返答に時間をかけることが多く、システムに表示された英単語の理解に戸惑うケースも見られた。これらの要素が、システム使用時に返答時間を増加させた要因として考えられる。

これらの結果から、総合的にはシステムの使用により英語でのコミュニケーションが円滑化されたと言えるが、返答時間の改善が期待される。具体的には、GPT の応答速度の向上や、ユーザが英単語に戸惑うことなく利用できるような UI の改良が考えられる。また、日本語由来の借用語

についての音声の誤認識を減らすための対策も必要である。これらの改善を通じて、より使いやすく効果的な多言語コミュニケーション支援システムを目指したい。

6. 結論

本研究では、GPT-3.5 Turbo を活用した返答提案機能を備えた多言語コミュニケーション支援システムを提案した。実験結果から、提案システムは英語が不得意な日本語話者が、英語でのコミュニケーションを円滑化できることが明らかになった。聞き返し回数、返答の妥当性の面で顕著な改善が見られ、提案システムは言語間の障壁を低減し、より円滑なコミュニケーションを実現する見通しが得られた。

参考文献

- [1] 文部科学省, “令和 4 年度「英語教育実施状況調査」の結果について”, 初等中等教育局教育課程課外国語教育推進室 (2022).
- [2] Education First, “EF English Proficiency Index” (2023).
- [3] T. Konishi, “国際的な人的交流と経済取引に関する分析—国別別高度専門職者と経済取引に着目して—”, 社会情報学, vol. 8, no. 3 (2020).
- [4] 日経, “翻訳機で外国人と話してみた「やばい」は伝わる?“, Nikkei (2021).
- [5] OpenAI, “GPT-3 Turbo”, OpenAI (2021).
- [6] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, “Improving language understanding by generative pre-training” (2018).
- [7] R. Sawai, I. Paik, A. Kuwana, “Sentence Augmentation for Language Translation Using GPT-2,” Electronics, vol. 10, no. 3082 (2021).
- [8] T. Fukushima, “医療従事者と外国人患者の間の対話支援を目的とした多言語対話シート作成システム LuPaCa”, 研究報告デジタルコンテンツクリエイション (DCC), vol. 2020-DCC-24, no. 39 (2020).
- [9] S. Maruf, A. F. T. Martins, G. Haffari, “Contextual Neural Model for Translating Bilingual Multi-Speaker Conversations”, arXiv:1809.00344v1 [cs.CL] (2018).
- [10] R. Agrawal, M. Turchi, M. Negri, “Contextual Handling in Neural Machine Translation: Look Behind, Ahead and on Both Sides,” Proceedings of the 21st Annual Conference of the European Association for Machine Translation, pp. 11-20, Alacant, Spain (2018).
- [11] C. Lyu, Z. Du, J. Xu, Y. Duan, M. Wu, T. Lynn, A. F. Aji, D. F. Wong, S. Liu, L. Wang, “A Paradigm Shift: The Future of Machine Translation Lies with Large Language Models,” LREC-COLING, vol. 2024, pp. 1339-1352 (2024).
- [12] Y. Fang, Y. Xu, Z. Teng, Z. Jiang, W. Xiang, “SocializeChat: a GPT-based AAC Tool for Social Communication Through Eye Gazing,” in Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing (UbiComp/ISWC '23 Adjunct), New York, NY, USA (2023).
- [13] J. Park, D. Choi, “AudiLens: Configurable LLM-Generated Audiences for Public Speech Practice,” in Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23 Adjunct), New York, NY, USA (2023).
- [14] F. Yang, H. Sun, “Let the LLMs Talk: Simulating Human-to-Human Conversational QA via Zero-Shot LLM-to-LLM Interactions”, arXiv preprint arXiv:2312.02913 (2023).
- [15] Google Cloud, “Google Cloud Speech-to-Text”, Google Cloud Documentation (2021).
- [16] OpenAI, “GPT-3.5 Model Documentation,” OpenAI Platform (2023).
- [17] Google, “Cloud Text-to-Speech”, Google Cloud (2023).
- [18] Sharp Corporation, “AQUOS sense4 - Basic Specifications”, Sharp Japan (2023).