

草野翼\*, 小田まり子\*

(久留米工業大学工学部\*)

## 1. はじめに

法務省出入国在留管理庁によると、留学生が日本の大学に入学する際の日本語能力の目安として日本語能力検定 (JLPT) においてN2以上の認定を受けていることが望ましいとされている<sup>[1]</sup>。N2レベルでは日常的な場面に加えて、ある程度のビジネス会話も可能なレベルである。しかしながら、実際に日本で生活する際には「方言」が大きな壁となる。例えば、九州地方において「なおす」という語は「片付ける」という意味で用いられることがある。しかし、標準語において「なおす」は「修理する」という意味で使われるのが一般的であるため、この方言表現を理解していないと、意図が正しく伝わらない場面が生じる。このように、方言は留学生にとって日常的なコミュニケーションにおける障壁となり得る。

## 2. 目的

以上のような背景から本研究では、筑後方言に特化した言語モデルを作成し、標準語と筑後方言との相互翻訳を可能とすることを目的とする。これにより、留学生が地域社会において直面する表現によるコミュニケーション上の課題を解決するシステムを開発し、検証を行う。

## 3. 研究手法

本研究では、筑後方言を学習した言語モデルを構築し、その有効性を検証するために以下の手順で研究を進める。

- ① データ収集  
筑後方言の収集を行う。地元住民へのインタビューでデータを収集し、標準語との対応表を作成する。
- ② データセット整備  
収集した方言を標準語と対応付けて整形し、学習用のデータセットを構築する。
- ③ モデル構築  
構築したデータセットを用いて既存の自然言語モデルを基盤としてファインチューニングを行い、筑後方言と標準語の相互翻訳を行うモデルを作成する。
- ④ システム実装  
構築したモデルをチャットボットに組み込む。留学生が入力した筑後方言を標準語に変換、またはその逆を行う対話システムを構築する。
- ⑤ 評価実験  
実際の留学生や地元住民を対象に評価実験を行い、アンケートを用いてシステムの有効性を評価する。

## 4. 研究結果

現在、研究手法の3.モデル構築まで行った。

- ① データ収集  
久留米工業大学職員および久留米郷土研究会の協力を得て、地元住民へのインタビューを実施した。質問内容は「子供の頃の思い出」や「筑後地域の生活」など、自然に方言が出やすいテーマを設定した。これにより筑後方言が含まれる多様な音声データを収集することができた。
- ② データセット整備  
収録した音声データは音声認識ツールWhisperを用いて文字起こしを行った。その際、筑後方言特有の語彙や発音が正しく認識されていない箇所が多く見られたため、手動で修正を加えた。さらに文字起こし

のデータの中から筑後方言特有の語彙や語尾を抽出し、標準語との対応を整理したうえで、JSONL形式の学習用データセットとしてまとめた。

### ③ モデル構築

本研究では自然言語モデルとしてT5 (Text-to-Text Transfer Transformer) を採用した。このモデルに対してLoRA (Low-Rank Adaptation) という手法を用いたファインチューニングを実施した。T5を選択した理由は、入力テキストを出力テキストに変換する汎用モデルであり、翻訳や要約など幅広いタスクに対応可能であるためである。ほかの代表的な言語モデルと比較すると、BERTは入力文の理解に優れるが生成が苦手であり、GPTは自然な生成に強いが翻訳などの忠実な文変換には不向きである。これに対してT5は両者の特性を合わせ持ち、入力文を正確に別形式の文として出力できるため、本研究の目的に適している (表1)。また、データセットのデータ数が少ないことから、LoRAによる軽量ファインチューニングを併用することで効率的な学習が可能となる。

表1. 代表的な3つのモデルの比較

モデル	形式	得意な処理
T5	Encoder-Decoder	翻訳, 要約, 言い換え
BERT	Encoder	分類, 固有表現抽出
GPT	Decoder	対話, 文章生成

### ④ モデル構築の結果と課題

本研究で構築したモデルは、特に動詞の変換において一定の成果を示した。動詞が比較的安定して変換できた理由として、生成AIを活用することで過去形、疑問形など複数のパターンをデータセットに用意できたことが考えられる。一方で名詞など他の品詞については変換が不安定であり、十分な精度が得られていない。今後は学習データの拡充とモデルの改良を進め、より幅広い品詞に対応できるように翻訳精度の改善を進める。

## 5. まとめ

本研究では、筑後方言の音声データ収集および対応表の作成を通じて学習用データセットを整備し、T5を用いた翻訳モデルの開発に取り組んだ。現時点では、特に動詞の方言表現を標準語へ変換において一定の成果を得ることができた。今後はモデルの改良を進め、チャットボットへの取り組みと、留学生や地元住民による評価を実施する予定である。また、チャットボット化の際にはRAG (Retrieval-Augmented Generation) を活用し、久留米のお祭りや伝統工芸品など地域に密着した知識を与えることで、より地域に根ざした支援型AIの構築を目指す。

## 参考文献

- [1] 出入国在留管理庁, 高等教育機関へ入学するための日本語能力について,  
[https://www.moj.go.jp/isa/applications/resources/n\\_yuukokukanri07\\_00022.html](https://www.moj.go.jp/isa/applications/resources/n_yuukokukanri07_00022.html)