

# MG2 日本の自然共生に関する伝統的知識を学習した 生成型対話システムの開発

Development of a Fine-Tuned Generative Dialogue System  
Using a Traditional Knowledge on Harmony with Nature in Japan

指導教員 町村尚准教授・地球循環共生工学領域 28H22057 花田拓実 (Takumi HANADA)

**Abstract:** Indigenous peoples and local communities have lived harmonized with nature and held traditional knowledge called Indigenous and Local Knowledge (ILK). In Japan, however, the urbanization of lifestyle and the depopulation make it difficult to inherit ILK. This research aims to develop an artificial intelligence for contributing to the inheritance of traditional knowledge and the conservation of biodiversity. The articles of “Kikigaki Koshien”, which is an archive of Japanese master’s ILK, were used as the ILK data frame. The Sentence BERT model and a Generative Pretrained Transformer (GPT) model pre-trained in Japanese were fine-tuned by the ILK dataset to create a generative dialogue system. The quality of generated answers by the dialogue system were evaluated from the view of naturality, preciseness, and utility. The developed dialogue system could successfully generate better answers compared to a general language model in terms of the inheritance of traditional master’s knowledge.

**Keywords:** indigenous and local knowledge, natural language processing, generative pre-trained transformer

## 1. 背景・目的

世界各地の先住民や地域コミュニティは自然との密接な関係の下で暮らしを営み、伝統的知識の担い手となっている。その知識は、持続可能な自然資源管理や、公正で持続的なライフスタイルへの転換に大きく貢献しうると考えられている<sup>1)</sup>。しかし日本では、都市的な生活様式への変化や人口減少を背景として伝統的知識の継承が困難になりつつあり、知識継承への支援は喫緊の課題である。本研究では、日本の自然共生についての伝統的知識のデータフレームから機械学習を行い、伝統的知識の継承と生物多様性の保全に寄与するための伝統的知識の生成型対話システムを開発することを目的とした。

## 2. 分析方法

### 2. 1 伝統的知識の記事コーパス学習用データフレームの構築

伝統的知識の生成型対話システムの学習用データには、高校生が自然とかかわる名人にインタビューをして伝統的知識を記録した「聞き書き甲子園」<sup>2)</sup>を用いた。2023年4月時点の「森を作る」(N=457)、多様な自然資源の利用方法を扱う「自然の恵み」(N=457)、伝統工芸をテーマとする「暮らしの手技」(N=707)、知識の継承活動を主なテーマとする「伝承と再生」(N=71)、および「その他」(N=58)の5つのカテゴリから1,750本の記事を用いて伝統的知識の記事コーパスデータフレームを作成した。

### 2. 2 伝統的知識の対話型生成システムの学習と評価実験

伝統的知識の生成型対話システムの回答の精度を高める Retrieval-Augmented Generation を実装するために、対話内容に類似した内容を記事コーパスデータフレームから取得するために日本語事前学習済みの Sentence-BERT<sup>3)</sup> を追加学習して情報検索モデルを構築した。追加学習の目的関数には Triplet Objective Function を選択し、記事コーパスデータフレームから作成した 111,285 の Triplet から追加学習を行った。伝統的知識生成モデルには日本語 Wikipedia や CC-100 等で事前学習し、対話学習・強化学習によって対話の精度の向上とアラインメント済みの Japanese-GPT-neox-3.6b-instruction-ppo<sup>4)</sup> を記事コーパスデータフレームから LoRA により追加学習した。最後に、「自然さ」「正確性」「名人の知恵の学習に有効か」の3観点から汎用生成モデルと伝統的知識生成モデルが生成した回答の比較により質的評価を実施した。

表 1. 生成モデルとの対話の例

<p><b>質問：</b>初心者でもできる里山管理ボランティアって何がありますか？</p> <p><b>汎用生成モデルによる回答：</b>地元の森林組合や林業会社などに参加することができます。また、地元のボランティア団体に参加することもできます。地元のボランティアグループに参加する場合は、地元の森林管理局や県などの機関に連絡を取る必要があります。</p> <p><b>伝統的知識生成モデルによる回答：</b>まず、里山ボランティアっていうと、草刈りとか枝打ちとかそういう作業だよ。それは、誰でもできる。あとは、間伐とか。</p>
<p><b>質問：</b>山と海のつながりは何があるのですか？</p> <p><b>汎用生成モデルによる回答：</b>海藻やプランクトンなどの海洋生物が豊富になるため、海藻の森が作られています。また、川からの栄養素が海に流れ込み、海産物の成長にも役立っています。</p> <p><b>伝統的知識生成モデルによる回答：</b>海の魚が獲れなくなるっていうのはね、海の生態系が変わってきているからなんです。海の生態系の変化の原因は、地球温暖化やろね。</p>

### 3. 結果と考察

表 1 に、生成モデルとの対話の例を示す。質問とそれに対する伝統的知識の記事コーパスデータフレームによる事前学習をしない汎用生成モデルの回答と、事前学習を行なった伝統的知識生成モデルによる回答例を示す。概ね汎用生成モデルに比較して今回開発した伝統的知識モデルの方が名人らしい口調で口語的に変化しており、追加学習による効果が確認できた。ただし、生成モデルそのものが根本的に抱える、現実には存在しないことを発言するという幻覚問題が観測されており、質問によっては回答の正確性にばらつきが見られた。

次に図 1 はアンケートにより得た 31 問の質問に対して汎用生成モデルによる回答と伝統的知識生成モデルによる回答について、16 名の被験者が自然さ、正確性、名人の知恵の学習の 3 観点から一対比較によって評価を行った結果である。自然性と正確性については汎用生成モデルと伝統的知識生成モデルによる回答には統計的に有意な差がなかった一方で、名人の知恵の学習に有効かについては伝統的知識モデルが統計的に有意に高く、名人の持つ伝統的知識の継承への貢献の可能性が示唆された。

### 4. 今後の課題

近年 ChatGPT などの大規模言語モデル (Large Language Model) が公開されているが、これら LLM による回答との比較や API を通じた連携を検討する必要がある。また伝統的知識を保持している方々や伝統的知識の研究者など当事者との協働により、更に妥当性と有用性の高いシステムの開発が必要である。

### 参考文献

- 1) Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Indigenous and local knowledge in IPBES, <https://ipbes.net/indigenous-local-knowledge>, 2024.01.31 参照.
- 2) 聞き書き甲子園実行委員会, 聞き書き甲子園, <https://www.kikigaki.net/>, 2024.01.31 参照.
- 3) 鈴木正敏, Pretrained Japanese BERT Models / 日本語 BERT 訓練済みモデル, <https://github.com/cl-tohoku/bert-japanese>, 2024.01.30 参照.
- 4) Hugging Face, japanese-gpt-neox-3.6b-instruction-ppo, <https://huggingface.co/rinna/japanese-gpt-neox-3.6b-instruction-ppo>, 2024.01.30 参照.

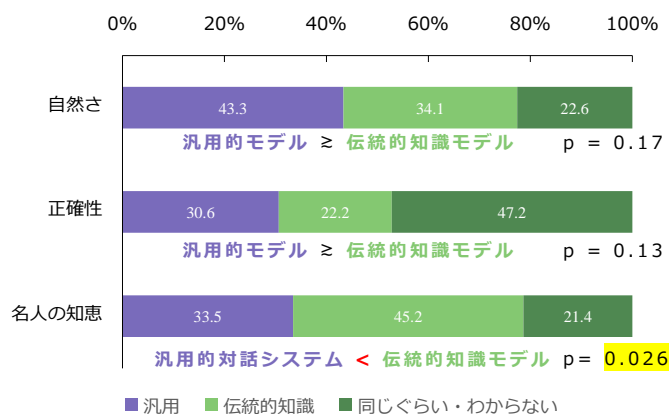


図 1. 汎用モデルと伝統的知識モデルに対する自然さ、正確性、名人の知恵の学習の評価の結果 (N=16)