

# 特集 「方言AI」を開発しています

生成AIを用いて方言で対話できる「方言AI」を開発する

——この画期的な取り組みを進めている鹿児島大学准教授 坂井美日さんに、方言AIの始まりから将来展望まで、お話を伺いました。

聞き手は、危機言語研究を専門とする国語研特任助教 横山晶子さんです。



坂井美日

鹿児島大学 共通教育センター 准教授

方言AIとは？

方言を話す人と標準語を話す人がスムーズに対話できることを目指した、生成AI\*を用いた翻訳および対話システムです。

\* 生成AI：学習データをもとにテキストや画像などのコンテンツを新たに作り出す人工知能



横山晶子

国立国語研究所 研究系 特任助教



方言AIの開発を始めたきっかけは？

熊本地震です。

横山：坂井さんは熊本のご出身でしたね。でも、AIの専門家ではなかったと思うのですが。

坂井：2016年に熊本地震が発生したとき、県外から来た医療従事者と地元の人との意思疎通で方言が障壁になっていることが問題となりました。地震直後、方言研究者を中心に方言集をつくって意思疎通を支援しようという動きが起こり、私も当初

は、地元の役に立ちたいという思いで参加していました。しかし配布した方言集は、現場ではあまり使われなかったと聞いています。災害後の医療現場は、冊子を開いて方言の意味を一つ一つ調べることができるような状況ではなかったのです。

そのとき、ハンズフリーで使えて、そばに置いておくだけで方言を標準語に翻訳できるものがあればいいな、と考えました。しかし2016年当時、それを実現するのは到底無理な話でした。でも最近になって生成AIの技術が大幅に進化したので、今なら私でも実現できるのではないかと考えた。2023年ごろから方言AIの開発に取り組み始めました。



医療現場における方言の壁というのは、例えば？

「ビンタが痛い」と言われて頬を診てしまう、といった誤診やミスコミュニケーションです。



横山：頬を打たれて痛い、という意味ではないのですか？

坂井：鹿児島方言では「ビンタ」は「頭」を意味するので、患者さんは「頭が痛い」と訴えていたのです。また、若い看護師さんから聞いた話ですが、入院しているおばあさんが「私、はっちきたいのよね」と言ったそうです。その看護師さんは意味が分からず、ナースステーションに戻って先輩の看護師さんに伝えると、先輩は慌てておばあさんのもとへ向かいました。「はっちきたい」というのは鹿児島方言で「死にたい」という意味で、すぐにケアしなければいけないことばだったのです。

このように、方言を理解できないと、誤診や対応の遅れにつながる恐れがあります。

横山：以前、消滅危機言語の継承支援に活用できないかと、生成AIに興味を持ったことがありました。しかし、生成AIの学習にはビッグデータが必要で、精度は学習データ量に依存するため、方言の生成は難しいと感じました。その状況が変わってきているのでしょうか？

坂井：言語データの量が少ない低資源方言は生成AIに向かないというのは、変わりません。そこで、通常AIに学習させる①鹿

児島方言の会話と標準語の対訳テキスト、②会話に出現する方言固有語彙の標準語対訳辞書に加えて、③鹿児島方言の文法情報など言語知識の概説を学習させたところ、生成精度が上がっ

たのです。この記述言語学の3点セットを教えてあげることが、方言の生成精度の向上に有効であることが明らかになりました。現在の鹿児島方言の生成精度は85%です。

生成精度85%というのは、とても高いですね。方言AIを使うと、どのような鹿児島方言の文がつけられるのでしょうか？



「鹿児島方言で東京スカイツリーについて200字くらいで説明して」という指示で生成された文をお見せします。

坂井：精度85%というのは、生成結果を文節ごとに鹿児島方言の話者に評価してもらったものです。なお、学習させた方言会話にスカイツリーの話は含まれていません。今後、学習させ

る会話や語彙を増やすことで、さらに精度を向上させられると見込んでいます。音声生成にも取り組んでいます。アクセントやイントネーションの精度を十分に高めるには5,000文以上を学習させる必要があるというので、頑張っているところです。

横山：2023年に「鹿児島方言週間フェスティバル」に展示された「AIせごどん」は大きな注目を集め、マスコミでも紹介されました。これを開発したのも坂井さんの研究室ですね。

坂井：西郷隆盛さんが鹿児島方言を話し、しゃべりに合わせて顔や口も動きます。「AIせごどん」は決まった文しか話せませんでしたが、この技術と方言AIを組み合わせることで、西郷さんと鹿児島方言でおしゃべりができるようになるかもしれません。

## 生成結果

東京スカイツリーは、東京のスマダクにあるどーC。このBタワーは、ふととかで、世界ん中で一番高かタワーじゃった。スカイツリーんてっぺんから見ると、東京ん景色が一望できっと、よか眺めじゃらいね。特に夜ん景色は、光がきらきらしてB、きれいかどー。スカイツリーん中にはBおみせもずんばいあつて、おきゃっさーが買い物Bしたり、食事したりできっと。また、そこにはB展望台もあつてB、わざわざ高かところからB東京を見がでくつど。天気よか日にゃ、富士山も見えつど。スカイツリーへの行き方は、電車で行くのがB便利じゃつど。東京スカイツリー駅や押上駅からちかいからB、そっから歩いていけつど。

Bは不十分(方言文脈の中にあっても良いが、より伝統的な言い方がある)、Cは不自然(意味は分かるが、当該方言として違和感がある)と評価された文節。D評価の非文(文法的、意味的に不可)はない。



西郷でございます。ゆくさおさいじゃしたー！(西郷です。ようこそいらっしやいました)

「AIせごどん」YouTubeで動画をご覧ください。https://www.youtube.com/watch?v=7XnWfClg\_sU



鹿児島方言で音声会話ができる「AI犬 かるかん」。鹿児島大学 坂井研究室と monoDuki 合同会社の共同開発中。

方言AIについて、今後どのような展開をお考えですか？



医療や介護現場での方言の壁の解消に加え、方言学習にも役立てたいと考えています。

坂井：まずは、スマートフォンで方言AIを使えるようにする計画です。さらに、このシステムをロボットに搭載することも検討しています。高齢者は標準語で話しかけられると身構えてしまい、うまく会話できないこともあるそうです。介護現場に方言AIロボッ



トがあれば、方言でリラックスしておしゃべりできるようになるのではないかと考えています。見守りロボットからの声かけも、方言だと喜ばれると思います。

方言に触れる機会を増やしたいという思いも、方言AI開発のきっかけの一つでした。方言を学びたいけれども、母語話との会話は気後れするという人もいると思います。何より、方言の話者は減っています。スマートフォンが相手ならば気軽にいつでも方言に触れ、学ぶことができます。学校の教室にも方言AIロボットを置きたいですね。ロボットが子供たちに方言で話しかければ、方言に興味を持つようになり、方言学習が自然と進むのではないのでしょうか。

横山：消滅危機言語の継承にも活用できそうですね。方言AIのこれからに期待しています。

方言AIをどのように作るのですか？



方言の言語データをAIに学習させます。鹿児島方言は利用できる言語データが少ないため、工夫が必要でした。