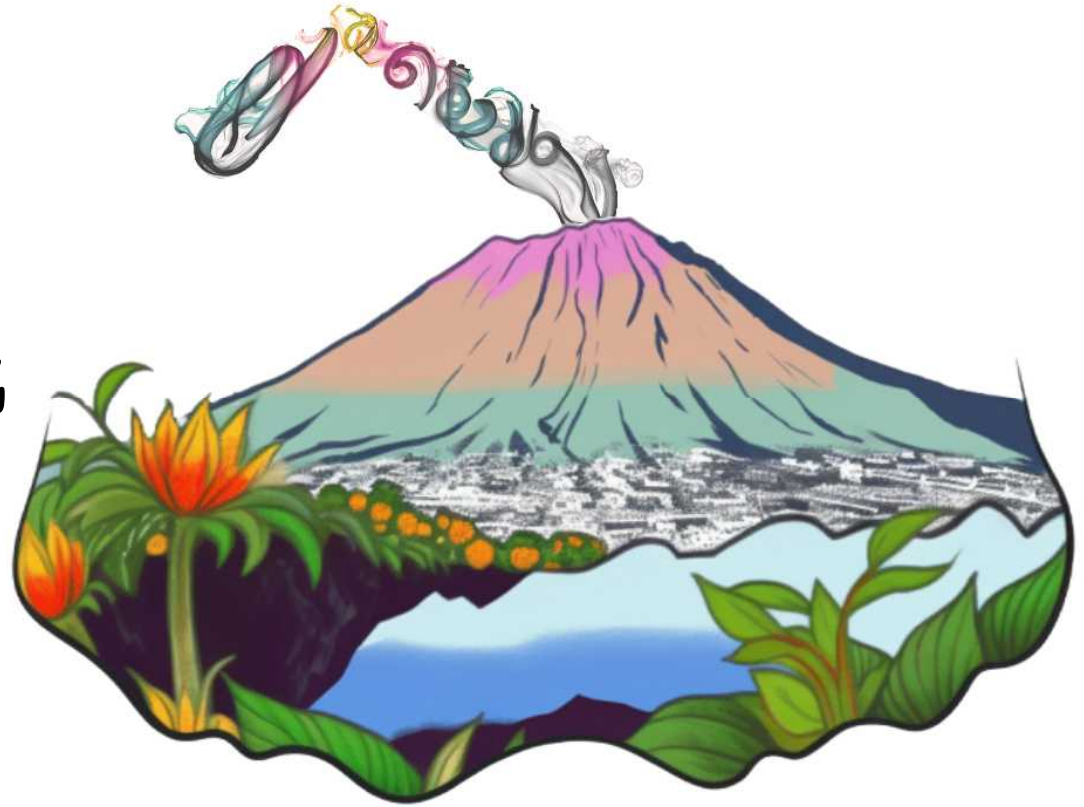


第16回産業日本語研究会・シンポジウム 2025/2/5
「生成AIと言葉、そして私たちの生活はどのように変わっていくか」

方言AIは方言の消滅 危機を救えるか

鹿児島大学
共通教育センター
坂井美日



Gokai Lab 

ことばの消滅危機

- 言語は世界に6000～7000。
- その半分（最悪約90%）が今世紀中に消滅と予測される（Krauss1992）。
- 2009年ユネスコは消滅危機言語を発表（Atlas of the World' s Languages in Danger）。
⇒約2500語が危機にあると報告。



日本の消滅危機言語（方言）

- 日本列島については、ユネスコのリストに8つ登録。

| Vitality（危機度） | Language（言語） |
|----------------------------------|---|
| Critically endangered （極めて深刻） | Ainu（北海道—アイヌ） |
| Severely endangered （重大な危機） | Yaeyama（八重山諸島）、Yonaguni（与那国島） |
| Definitely endangered （危険） | Hachijō（八丈島，青ヶ島），Amami（奄美大島，喜界島，徳之島）， Kunigami（与論島，沖永良部島，沖縄本島北部），Okinawan（沖 縄本島中部および南部，周辺諸島），Miyako（宮古島，周辺諸島） |

注：ユネスコは、「言語」と「方言」の区別に厳密ではなく、一緒にして扱っている。

ユネスコのリストは氷山の一角

ユネスコ未調査の言語・方言にも、消滅危機のもの多数。

(例)

- 鹿児島県甑島方言 → severely endangered (重大な危機)
(木部・山田・下地 2011, pp.88-90)
- 宮崎県椎葉村方言 → severely endangered (重大な危機)
(坂井2018)

言語・方言を失うことの重大さ

言語はかけがえのない人類の文化遺産でありながら、今世界各地で十分な記録もないまま永遠に失われようとしている。文化遺産というと古文書や遺跡などの‘モノ’が注目されがちだ。だがモノを 生みだし、文化をかたちづくるのは言語であり、そこには社会の知恵や独自の世界観が集積している。言語の喪失は文化の活力の喪失にもつながり、モノも喪失よりはるかに重大ともいえる。

中山俊秀 (2002) p.104

従来の方策

- 文化庁の国語施策：「危機的な状況にある言語・方言サミット」の開催や、調査研究の委託事業。
- 国立国語研究所等による大規模な研究プロジェクト。
（例）「日本の消滅危機言語・方言の記録とドキュメンテーションの作成」等

従来の方策

- 文化庁の国語施策：「危機的な状況にある言語・方言サミット」の開催や、調査研究の委託事業。
- 国立国語研究所等による大規模な研究プロジェクト。
（例）「日本の消滅危機言語・方言の記録とドキュメンテーションの作成」等

でも、いまだに消滅危機を脱していない

なぜいまだ危機を脱しえない？

- 一時的なイベントでは継承に結びつかない。

なぜいまだ危機を脱しえない？

- 一時的なイベントでは継承に結びつかない。
- 継承…人々（特に子供世代）が当該方言を習得。

なぜいまだ危機を脱しえない？

- 一時的なイベントでは継承に結びつかない。
- 継承…人々（特に子供世代）が当該方言を習得。
- 自然習得？教室習得？ ←

なぜいまだ危機を脱しえない？

- 一時的なイベントでは継承に結びつかない。
- 継承…人々（特に子供世代）が当該方言を習得。
- 自然習得？ 教室習得？
- 日常で方言に触れる機会は激減（核家族化等）。

なぜいまだ危機を脱しえない？

- 一時的なイベントでは継承に結びつかない。
- 継承…人々（特に子供世代）が当該方言を習得。
- 自然習得？ 教室習得？
- 日常で方言に触れる機会は激減（核家族化等）。
- （教育に盛り込むにも）教えることができる人がいない。
- 教材がない。教育カリキュラムに入れ込む隙間が無い。

なぜいまだ危機を脱しえない？

• 一時的なイベントでは継承に結びつかない。

• 継承

方言を習得できる機会がない

- 自然習得 × 教室習得 ×
- 日常で方言に触れる機会は激減（核家族化等）。
- （教育に盛り込むにも）教えることができる人がいない。
- 教材がない。教育カリキュラムに入れ込む隙間が無い。

なぜいまだ危機を脱しえない？

- 一時的なイベントでは継承に結びつかない。
- 継承…人々（特に子供世代）が当該方言を習得。
- 自然習得×教室習得×
- 日常で方言に触れる機会は激減（核家族化等）。
- （教育に盛り込むにも）教えることができる人がいない。
- 教材がない。教育カリキュラムに入れ込む隙間が無い。
- **若年層の方言への無関心。**

補足：若年層の方言への無関心

「近年、少数言語・方言が消滅しています。どう思いますか。」

【仕方ない、あきらめる】

71.4%

(理由)

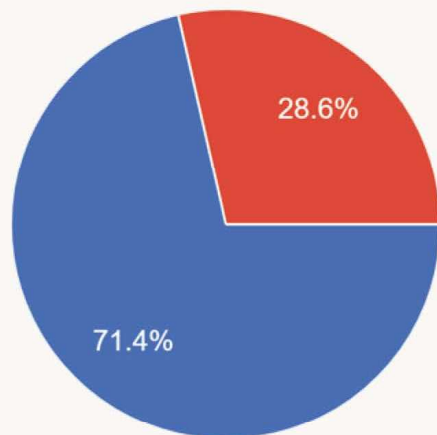
- 標準語を覚えればいい。
- 普段方言を使わないからあまり関係ない。
- 優先度が低いものにリソースを割けない。
- 翻訳がいらないから便利
- 話者数が少ないため消えても迷惑はかからない。

【事態を止める行動をする】

28.6%

(理由)

- 言語が無くなるのはその土地の文化がなくなること。
- 言語はその人のアイデンティティ。
- 方言はその地域の個性。無くなるのはつまらない。



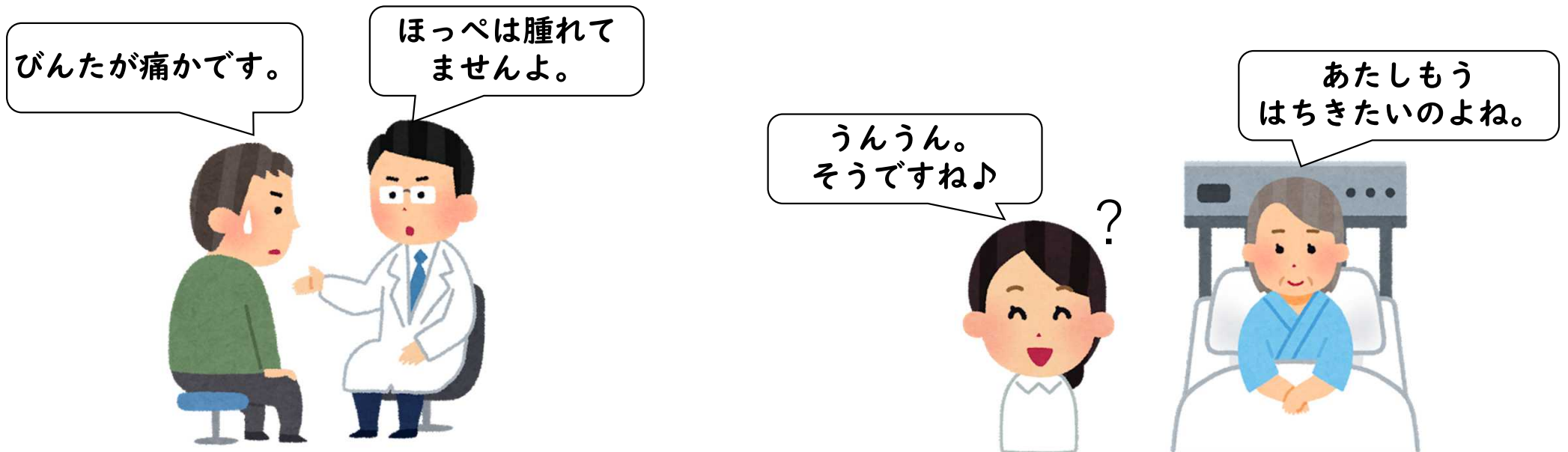
(2021年鹿児島の大学生35名への調査)

なぜいまだ危機を脱しえない？

- 一時的なイベントでは継承に結びつかない。
- 継承…人々（特に子供世代）が当該方言を習得。
- 自然習得？教室習得？
- 日常で方言に触れる機会は激減（核家族化等）。
- （教育に盛り込むにも）教えることができる人がいない。
- 教材がない。教育カリキュラムに入れ込む隙間が無い。
- 若年層の方言への無関心。
- **ことばの多様性の壁。**

補足：ことばの多様性の壁

- 多様性に起因するミスコミュニケーション。
- 特に医療現場で深刻（正確な問診や寄り添いの弊害）



補足：ことばの多様性の壁

- 多様性に起因するミスコミュニケーション。
- 特に医療現場で深刻（正確な問診や寄り添いの弊害）



継承を実現するには？

- 現代の家族構成や教育カリキュラムのなかで、
- 一時的ではなく日々持続的に、
- 日常生活か教室かで、
- 興味関心のもと当該方言を苦痛なく習得し、
- その習得と継承で生じる多様性が壁にならないようにしなければならない。

継承を実現するには？

- 現代の家族構成や教育カリキュラムのなかで、
- 一時的ではなく日々**持続的に**、
- 日常生活か教室かで、
- 興味関心のもと当該方言を苦痛なく習得し、
- その習得と継承で生じる多様性が壁にならないようにしなければならない。

継承を実現するには？

- 現代の家族構成や教育カリキュラムのなかで、
- 一時的ではなく日々持続的に、
- **日常生活**か**教室**かで、
- 興味関心のもと当該方言を苦痛なく習得し、
- その習得と継承で生じる多様性が壁にならないようにしなければならない。

継承を実現するには？

- 現代の家族構成や教育カリキュラムのなかで、
- 一時的ではなく日々持続的に、
- 日常生活か教室かで、
- **興味関心**のもと当該方言を**苦痛なく習得**し、
- その習得と継承で生じる多様性が壁にならないようにしなければならない。

継承を実現するには？

- 現代の家族構成や教育カリキュラムのなかで、
- 一時的ではなく日々持続的に、
- 日常生活か教室かで、
- 興味関心のもと当該方言を苦痛なく習得し、
- その習得と継承で生じる多様性が壁にならないようにしなければならない。

継承を実現するには？

- 現代の家族構成や教育カリキュラムのなかで、
- 一時的ではなく日々持続的に、
- 日常生活か教室かで、
- 興味関心のもと当該方言を苦痛なく習得し、
- その習得と継承で生じる多様性が壁にならないようにしなければならない。

もう…一体どうやって…

AIの急成長による一筋の光

AIの急成長による一筋の光

- 対話型AIは、語学との相性がよい。
- 自分専用の語学教師をデバイスに。
例：Duolingo Max等 (Duolingo2023)

AIの急成長による一筋の光

- 対話型AIは、語学との相性がよい。
- 自分専用の語学教師をデバイスに。
例：Duolingo Max等 (Duolingo2023)
- 方言を生成し対話できるシステムを開発すれば、継承支援になるのではないか。



方言AIシステム研究（2024～）

危機が特に深刻な九州・琉球の方言AIシステム開発開始。

- 基盤研究B「九州・沖縄方言の継承支援に資する音声対話型生成系AIの開発」
- 鹿児島大学先端研究プロジェクト「九州・沖縄の方言の消滅危機対策に資する生成AIコンテンツの開発」

坂井美日（代表／鹿大）
重野裕美（奄美・喜界担当／広大）
加藤幹治（徳之島担当／Nii）
横山晶子（沖永良部担当／国語研）
宮川創（与論担当／筑波大）
久保菌愛（鹿児島古典担当／岡大）
高城隆一（鹿児島本土担当／九大）
山田高明（熊本担当／有明高専）
當山奈那（沖縄担当／琉球大学）
中川奈津子（八重山担当／九大）
高道慎之介（音声部門担当／慶大）



方言AI構想

背景：生成AIの急速な発達と普及

- 「AIの開発はかねて進められてきたが、オープンAIが22年11月にChatGPTを公開したのを機に、コンテンツ（文書やソースコード、画像、音声など）を自動でつくる生成AIがにわかに大衆化」（日経2023年12月5日）
- 今後の言語生活には、AIが当然のように入り込む。

生成AI（テキスト生成）の仕組み

例（GPT-4o）

M You

方言の消滅危機について、あなたはどのように思いますか？

 ChatGPT

方言の消滅危機は、とても残念なことだと思います。方言はその地域の文化や歴史を反映していますし、独自の表現や温かみがあります。方言が失われると、その地域のアイデンティティや伝統も一部失われてしまう感じがしますね。あなたはどのように思いますか？

生成AI（テキスト生成）の仕組み

例（GPT-4o）

M You

方言の消滅危機について、あなたはどのように思いますか？

 ChatGPT

方言の消滅危機は、とても残念なことだと思います。方言はその地域の文化や歴史を反映していますし、独自の表現や温かみがあります。方言が失われると、その地域のアイデンティティや伝統も一部失われてしまう感じがしますね。あなたはどのように思いますか？

- 大規模言語モデル（LLM）がベース。
- プロンプト（指示文）を入力。
- AIが次の単語の出現確率を予測し（トランスフォーマーアーキテクチャ）文章を生成。

大きな問題—低資源言語の生成は非常に難しい—

- 生成AIのベースは大規模言語モデル=膨大なデータが必要。
(例) GPT-3：約1750億個のパラメータで45TBを事前処理した570GB以上のテキストを学習 (Brown 2020)
- 低資源の方言で、生成AIモデルを1からは作れない。
- 既存の生成AIモデルを使用するにも、精度は学習データ量に依存するため低資源の方言はうまく出力できない。

既存LLMで低資源言語を生成しようとする…

プロンプト「鹿児島方言で話してください。東京スカイツリーについて説明して下さい。200文字程度でお願いします。」

(7)Zero-shot (鹿児島方言を学習していないGPT4)

東京スカイツリーは、東京の墨田区にある電波塔で、高さ634メートルで日本一の高さを誇るばってん。

観光スポットとしても有名で、展望台からの眺めは絶景じゃけんね。

下の部分にはショッピングモールもあって、飲食店やお土産店がたくさんありますよ。夜にはライトアップされて、とても綺麗なんじゃけど、その景色を見るのもおすすめですよ。

既存LLMで低資源言語を生成しようとする…

プロンプト「鹿児島方言で話してください。東京スカイツリーについて説明して下さい。200文字程度でお願いします。」

(7)Zero-shot (鹿児島方言を学習していないGPT4)

東京スカイツリーは、東京の墨田区にある電波塔で、高さ634メートルで日本一の高さを誇るばってん。

観光スポットとしても有名で、展望台からの眺めは絶景じゃけんね。

下の部分にはショッピングモールもあって、飲食店やお土産店がたくさんありますよ。

夜にはライトアップされて、とても綺麗なじゃけど、その景色を見るのもおすすめですよ。

- 鹿児島方言固有の要素は一つも入っていない。
- 標準日本語の分量が多い（標準日本語としての文レベルも低い）。
- 方言要素は中国四国（じゃけん、じゃけど）、九州肥筑（ばってん、ばい）。
- 「高さを誇るばってん（高さを誇るけれども）」は文脈にもあわない。

(補足) GPT-4oでもハルシネーション祭り

「鹿児島方言を話してみて。」 (Zero-shot, 2024.10.27)

× (関西方言) × (「ん」) × (「くれっせー」)
おおきに、鹿児島の方言に興味を持ってくれて嬉しかあ！

× (「じゃ」) × (完全に嘘) × (「っちゅー」)
鹿児島弁では、「こんにちは」を「こんちわっ」と言ったり、

× (奄美語が混ざっている?) × (「っちゅー」) × (「もす」)
「ありがとう」を「ありがっさま」と言うたりします。

方言AI構想の実現のためにクリアすべきこと

- 低資源言語たる方言の生成を実現すること。
- 標準語との距離があり難易度が高い方言も高精度に生成すること。
- 方言AIを実装し、継承支援や多様性の壁克服に実用化すること。

これまでの進捗

既存LLMを活用して日琉諸語・諸方言を高精度で生成する手法を発見。

- 1.標準日本語（日琉諸語で最もLLMに学習された言語）を活用（対訳）するのが有効（坂井2024a）。
- 2.当該方言の標準語対訳テキスト、標準語対訳辞書、言語知識の概説の3点を入力（あるいは学習）すると、方言の生成精度が上がる（坂井2024a, b）。
- 3.この手法は、本土方言でも、琉球諸語でも有効である（坂井2024c）

自然言語処理 (回答内容の生成)



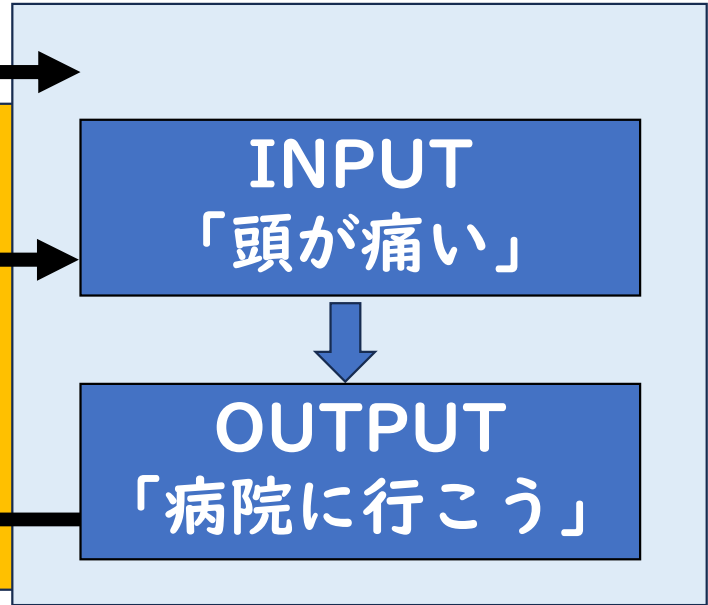
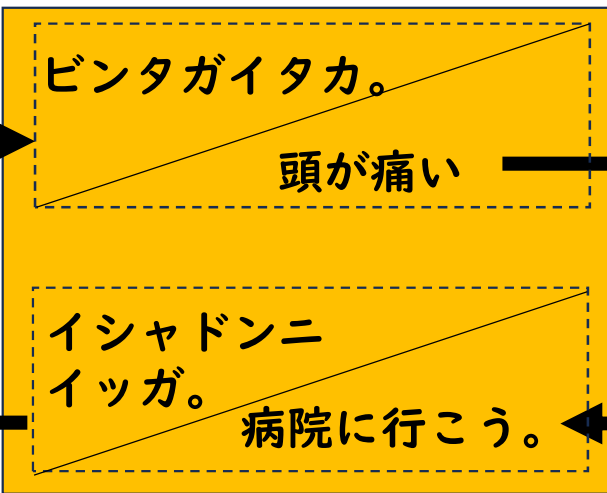
頭が痛いです。

(入力が標準語なら直行)

OR

ビンタガイタカ。

イシャドンニ
イツガ!



方言AI処理

既存LLM

- 標準語対訳辞書
- 標準語対訳テキスト
- 当該言語・方言の概説

現段階の状況

- 鹿児島島本土方言で精度8割程度。
- 宮古島城辺方言で精度7割程度。
- 今年度から基盤B（JP24K00074「九州・沖縄方言の継承支援に資する音声対話型生成系AIの開発」）で奄美・沖縄等作業中。
- 同科研で方言音声合成も開発中。
- テキスト生成と音声合成をあわせてコンテンツ制作も進行中。

鹿児島方言のチャットボット試作 (GPT-4o/miibo/2024.11月)

びんたて、どこのことか分かる？

【レビュー】

びんたは、頭のことじゃっど。びんたがいたか時は、注意が必要じゃね。



⚠ AIによる自動生成

さすがじゃ。びんたがいたかときは、重か病気かもしれん。気をつけんといかん

【レビュー】

ほんのこっじゃな。びんたがいたか時は、無理せんで、ちゃんと医者どんに診てもろた方がよかど。大事にせんといかんね。



⚠ AIによる自動生成

宮古語城辺方言のチャットボットの試作（GPT-4o／2024.7月）

了解しました！城辺方言での会話を始めましょう。

ウヴァー、アサムヌヤ ノーユガ フォーターリヤ？

バヤー バソーユバ フォーター！

実装試作「かるかんちゃん」

(鹿児島方言の音声合成は制作段階につき、音声は「ずんだもん」で代替中)



鹿児島方言⇔標準語の双方向対訳システム

鹿児島方言⇔標準語翻訳

方言や標準語を入力してください

翻訳

- 後ろで動いているのは、チューニングしたGPT4o

鹿児島方言⇔標準語の双方向対訳システム

宮古語城辺方言⇔標準語:

翻訳したい文章を入力してください

翻訳

翻訳結果:

後ろで動いているのは、チューニングしたGPT4o

方言AIを社会実装できれば

1. 方言学習の補助役



2. 「方言の壁」の解消

(低資源言語・方言の生成による可能性)

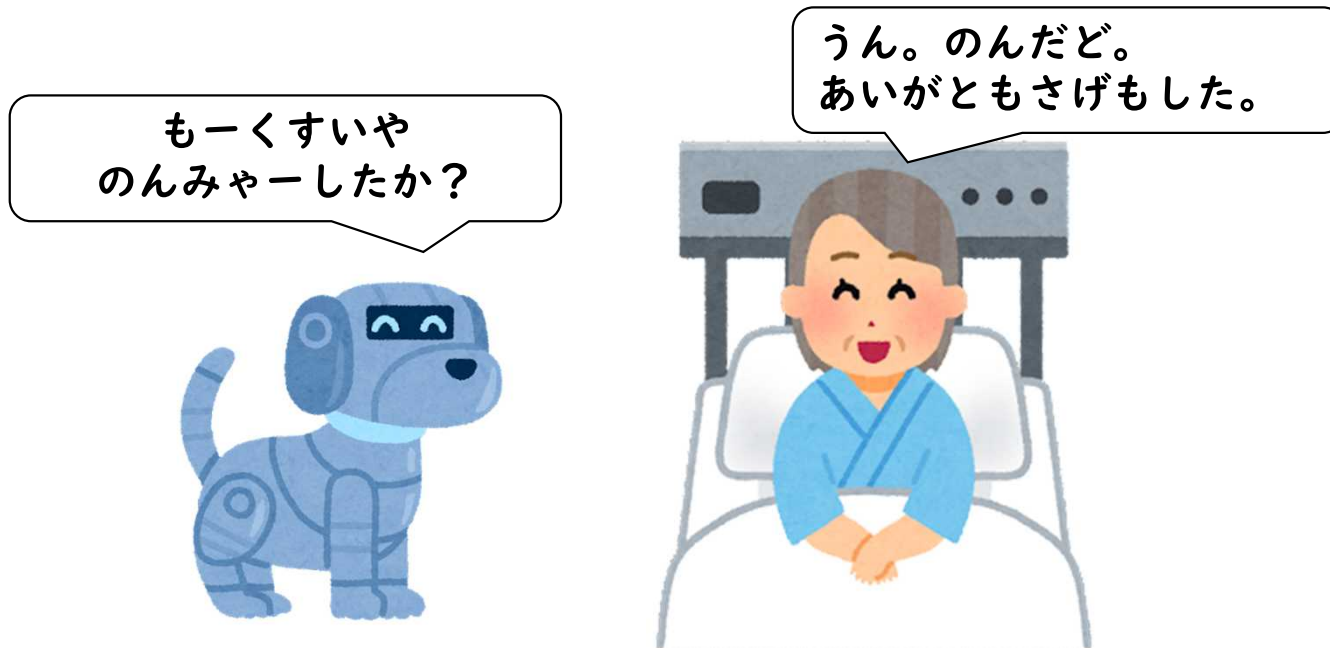
- 医療従事者へのコミュニケーション支援ツール



3. 「方言の壁」の解消

(低資源言語・方言の生成による可能性)

- 患者のQOL向上ツール



いま、0か100かの瀬戸際

開発にあたり留意しないといけないこと

- 精度は、ほぼ完璧でなければならない（間違った継承に繋がる）。
- 方言システムの精度を評価できるのは、当該方言の母語話者だけ。

間違いや違和感はネイティブにしか分からないため、話者と協働で開発する必要がある。

（実践例）鹿児島市方言AIの実証実験

母語話者がテスト利用中。生成にミスを見つけたら、正しい方言形を開発者（坂井）に随時フィードバック。

時間がない…

- 方言母語話者は、現在減少の一途。
- このままでは、約50年後に不在となる。

今、開発のラストチャンス

0か100かの瀬戸際

- 高精度の方言AIを開発し、社会実装することができれば、消滅危機を乗り越え多様性を維持する一助となりうる。



- 逆に、方言AIを開発できなければ、今後、当該方言は
‘AIが対応できないことば’として、AI時代に一層弱い立場となり、消滅危機に拍車がかかると考えられる。

仲間、求む。

参考文献

- 木部暢子 (2011) 「鹿児島県甑島方言」木部暢子・三井はるみ・下地賀代子・盛思超・北原次郎太・山田真寛『文化庁委託事業 危機的な状況にある言語・方言の実態に関する調査研究事業 報告書』, pp. 69-76: 国立国語研究所
- 木部暢子・三井はるみ・下地賀代子・盛思超・北原次郎太・山田真寛(2011)『文化庁委託事業 危機的な状況にある言語・方言の実態に関する調査研究事業 報告書』国立国語研究所.
- 坂井美日 (2018) 「宮崎県椎葉村方言のいまむかし—静かに消えてゆく、私達のことば—」『日本語学』37(7) 56 - 67
- — (2024a) 「生成AIを用いた鹿児島方言生成—日琉諸語の低資源言語・方言の生成に向けた試み—」言語処理学会 第30回年次大会 発表論文集
- — (2024b) 「日琉諸語・諸方言の生成に向けた手法の提案—鹿児島方言を例に—」第118回日本方言研究会発表原稿集
- — (2024c) 「生成AIを用いた日琉諸語・諸方言の生成対話システム開発において」2024年度沖縄言語研究センター総会・公開研究会・ワークショップ
- 中山俊秀 (2002) 「日本の危機言語研究」『日経サイエンス』2002年11月号日経サイエンス社、p.104
- 文化庁HP 「危機的な状況にある言語・方言サミット」
https://www.bunka.go.jp/seisaku/kokugo_nihongo/kokugo_shisaku/kikigengo/summit/index.html [2024.0831最終閲覧]
- Krauss, Michael E. (1992). The world's languages in crisis. *Language*, 68, 4-10