

外国人介護職員のためのやさしい日本語を用いた オノマトペ変換辞書の自動生成手法の提案

小野愛佳¹ 中島陽子¹ 本間宏利¹ 秋葉友良² M. Ptaszynski³ 福田未来⁴ 榎井文人³
¹ 釧路工業高等専門学校 創造工学科 ² 豊橋技術科学大学 情報・知能工学系
³ 北見工業大学 地域未来デザイン工学科 ⁴ 泰日工業大学
{yoko, honma}@kushiro-ct.ac.jp

概要

介護現場において、介護サービスをしたり入居者の状況を職員間で共有する際の会話や介護記録の中でオノマトペは状況やニュアンスの伝達が容易なため頻繁に使用されている。近年、外国人介護士が増加している一方、日本語を修得した外国人介護士でもオノマトペのように文化的な背景を暗示する日本語を理解するには時間を要するため、外国人介護士が日本人介護士や入居者とのコミュニケーションに苦慮している。一方、日本人介護士は、オノマトペを使わない作業指示や状況説明を外国人介護士にわかりやすく行うことに苦慮している。そこで、本研究では、やさしい日本語や母国語を用いた説明文や例文など、オノマトペの理解に有効な5項目の情報を自動的に収集、生成し介護に特化したオノマトペ変換辞書を自動生成する手法を提案する。本辞書は、介護方法支援システムに導入することで外国人介護士とのコミュニケーションを円滑にし、介護職員の作業効率の向上が期待できる。

1 はじめに

近年、日本では高齢化の進行に伴い、介護施設の需要が急増している。厚生労働省によると、2025年には34万人の介護人材が不足すると報告されており、その対策として、外国人労働者の受け入れが進んでいる[1]。日本で働く外国人労働者を国別状況を見ると、ベトナム人が最も多く外国人労働者全体の26.2%を占めている。次いで、中国が23.0%、フィリピンが11.1%となっている[2]。しかし、日本語を習得して働く外国人介護士は、日本文化特有のオノマトペを理解することが難しく、自然習得だけでは体得しきれないという問題がある[3]。また、オノマトペは状況やニュアンスの伝達が容易なため、介護施設

において頻繁に使用されるが、日本人介護士がこれらのオノマトペを外国人介護士に適切に説明することに苦慮するという問題も生じている。

オノマトペ辞書に関連する研究として、浅賀ら[5]がオノマトペの用例に特化したオノマトペ用例辞典を構築しているが、登録されている辞書の項目はオノマトペの用例のみであり、使用されている日本語についても、外国人が理解するには難解なものがある。また、介護に関するオノマトペを扱っているウェブサイトや書籍の出版もみられるが、現場で即活用できる形式を兼ね備えていない。オノマトペを含む会話や文章を外国人介護士に理解しやすい説明を提供できる、かつ、現場で即時に理解できるオノマトペ辞書は現在のところ見当たらない。介護支援の関連研究では、結城ら[4]が介護施設の入居者の困りごとを入力するとその対応方法候補を出力する介護方法支援システムが開発されているが、外国人介護士に理解しやすい機能は考慮されていない。そこで、本研究では、介護施設が所有しているケア会議の会議音声データおよび議事録から、オノマトペとそれを含む文章を抽出し、やさしい日本語と4ヶ国語の説明文の生成、および、関連画像を検索により追加できるオノマトペ変換辞書の自動生成を提案する。

2 オノマトペ変換辞書の構成

オノマトペ変換辞書は、オノマトペ、やさしい日本語の説明文、英語の説明文、多言語の説明文、例文、画像の6項目で構成する(表1)。オノマトペを除く5項目は、日本語を母国語としない日本語学習者(泰日工業大学学生20名)にアンケート調査を行い決定した。アンケートはオノマトペ4種類を説明に、(1)「やさしい日本語のみ」、(2)「母国語のみ」、(3)「優しい日本語と画像」、(4)「画像のみ」をそれぞれ提

表1 オノマトペ変換辞書の項目と例

	項目	例
1	オノマトペ	うろろう
2	やさしい日本語の説明文	目的もなくあちこち動き回ること
3	英語の説明文	Moving around aimlessly without a specific purpose.
4	ベトナム語 中国語 フィリピン語	Mục đích không rõ ràng khi đi chuyển động khắp nơi 漫無目的地四處走動 Ang pag-iikot-ikot nang walang layunin
5	例文	家の前をうろろうしている。 ひとりでうろろうしている。
6	画像	

示し、回答して頂いた。その結果、(1)が7.69%、(2)と(3)が34.61%、そして(4)が23.07%となった。(2)と(3)が学生がわかりやすい説明であることが明らかになった。母国語の説明は理解に役立つが、日本の介護施設で仕事をするうえで日本語で学ぶことはコミュニケーションに役立つことを考慮すると、(2)のように日本語と画像を組み合わせることが有効である。また、例文を示した実験結果では、(4)例文を示さない、(5)例文を示す場合の正誤率は(4)は65%、(5)は35%であった。以上のことから、複数の説明形式を合わせることで理解の支援になると考え、5項目に設定した。各項目の詳細を以下に述べる。

辞書項目で生成するやさしい日本語の説明文のやさしい日本語は、自治体で使われているやさしい日本語¹⁾を参考にする。やさしい日本語とは、難しい言葉や表現、漢字などを言い換えるなど、相手に配慮したわかりやすい日本語のことである。やさしい日本語は、災害時に外国人に素早く的確に情報を伝えることを目的に考案され、現在では自治体や外国人支援団体に生活情報や観光情報などを伝える手段としても使われている。「雨がザーザー降る。」のオノマトペ：「ザーザー」を説明する場合、「非常に激しく」よりも「たくさん」のような簡潔な表現とする。

また、オノマトペを説明する際に、(a)短い簡潔な説明文と(b)長い詳細な説明文のどちらが理解しやすいかについてもアンケート調査を行った。オノマトペとその説明をやさしい日本語の(a)と(b)を示し、理解しやすい説明文を選択させた。その結果、(a)が68.75%、(b)が31.25%であったことから、

1) やさしい日本語：
<https://www.moj.go.jp/isa/content/930005563.pdf>

やさしい日本語での説明は短い簡潔な説明文を採用する。

多言語の説明文は、日本で働く外国人労働者の国籍が多い、ベトナム人、中国人、フィリピン人と評価協力していただくタイ人を対象とし、ベトナム語、中国語、フィリピン語およびタイ語による説明文とする。前述の言語を母国語としない外国人介護士のために、また、翻訳精度の信頼度が高いとされる英語の説明文を生成する。例文は、浅賀ら[5]のオノマトペ用例辞典があるように、オノマトペの例文から使用されているオノマトペの意味を推測するのに役立つ。説明文や例文に加えて、画像を提示することで、オノマトペの利用場面をイメージできたり、表意の理解の支援が期待できる。

3 提案手法

オノマトペ変換辞書の項目を生成するために、オノマトペリストを用いて、介護施設データからオノマトペとそれを含む文章の抽出を行う。抽出したオノマトペと文章から、オノマトペ変換辞書の項目の生成を行い、生成したそれぞれの項目をオノマトペ変換辞書に登録する。オノマトペ変換辞書構築の手順を以下に示す。

Step1: 介護施設音声データのテキスト化 介護施設データには音声データとテキストデータがある。まず、音声データをテキストデータに変換するために、音声データの書き起こしを行い、句読点の自動挿入を行う。書き起こしには、OpenAIが提供している文字起こしAPI Wisper midiumモデル²⁾を用いる。出力結果は、句読点が少なく文章が長い傾向がみられるため、BERT日本語版事前学習言語モデルを用い句読点を挿入する。出力された際に付与された句読点を一度削除し、全ての文を対象に100文字ごとにマスキングを行い、文字間に入る確率の高い句読点を文脈から予測し、確率が高い位置に句読点を挿入する。テキスト化された音声データとテキストデータは不要な記号や文字列を削除しクリーニングを行う。

Step2: オノマトペリストの作成 オノマトペリストは、ひらがな表記、カタカナ表記、ひらがな表記+「と」、カタカナ表記+「と」の項目で構成する。まず、オノマトペを抽出するための手

2) <https://github.com/openai/whisper>

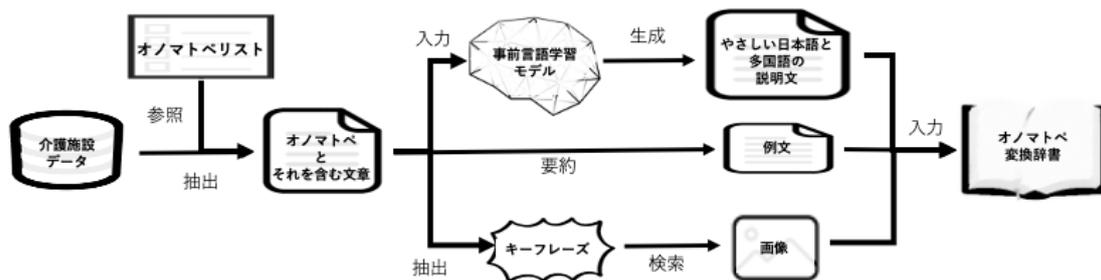


図1 介護データを用いたオノマトペ変換辞書自動生成の概要図

がかりとして、介護・看護のオノマトペ辞典 [6] を参考にオノマトペリストを作成する。次に、介護施設データ中に作成したオノマトペリストに含まないオノマトペがある場合、オノマトペリストに追加する。「はっとした」のような小さな「っ」で終わるオノマトペの場合、「はっ」を検索キーとした場合、「はっきり」のようなオノマトペを抽出してしまう。このような誤抽出を防止するために、小さな「っ」で終わるオノマトペには、語尾に「と」を追加し「はっと」のような形式で登録する。オノマトペリストは、テキストデータからオノマトペを含む文章を抽出する際の検索キーとして利用する。

Step3: オノマトペを含む文章の抽出 オノマトペリストのオノマトペを検索キーとして、テキストデータから検索キーを含む文章を抽出する。文章を形態素解析し、検索キーと一致した単語があった場合、その文章の抽出を行う。抽出した文章は、やさしい日本語の説明文、英語および多言語の説明文を生成する際の入力となる。

Step4: やさしい日本語と多言語の説明文の生成 まず、やさしい日本語の生成を行う。

オノマトペとそれを含む文章と質問文「この文章に含まれる [] というオノマトペはどういう意味ですか？やさしい日本語で教えてください」を入力とし、大規模言語モデル GPT3.5³⁾ の文章生成タスクを用いてやさしい日本語の説明文を3文生成する。出力された文章がやさしい日本語であるかの判定は、Web アプリケーションやさしにチェック [7] に入力し行う。やさしい日本語と判定された説明文を辞書に登録する。いずれもやさしい日本語と判定されなかった場合は、質問文を変更し再度生成する。次に、やさしい日本語の説明文を GPT3.5 の翻

訳タスクを利用し英語および多言語への翻訳を行い、辞書に登録する。

Step5: 例文の生成 例文の生成は、Step3 で抽出したオノマトペを含む介護施設の事例を原文とし、オノマトペを欠落させることなく抽出型要約手法を用いて文章要約を行う。

Step6: 画像検索 対象となるオノマトペの特徴が現れている画像（著作権フリーのイラスト）を Web 検索をし収集する。検索は、Step3 で抽出した文章中の単語の特徴を表す指標のひとつである TF-IDF を用いて特徴的な単語を特定し、(オノマトペ AND 特徴的な単語) を検索キーとして検索を行う。画像検索は複数の結果が出力され、僅かなニュアンスの違いがある場合があるため、出力結果から画像を選択する作業は人手で行う。

4 実験結果

(a) 介護施設音声データのテキスト化、(b) オノマトペリストの作成、および、(c) オノマトペを含む文章の抽出について実装と検証を行なった。本実験では、介護施設で行われるケア会議を介護職員が作成した8ヶ月分の議事録データ（テキストデータ）と会議を録音した6ヶ月分の音声データを用いた。各実験ごとの結果について述べる。

4.1 介護施設音声データのテキスト化

介護施設音声データのテキスト化では、6ヶ月分ケア会議の録音データを用い実験を行った。文字起こし API Wisper midium モデルの結果と録音データを聞き手で書き起しをした結果の比較例を表 2 に示す。人手による書き起しを正解として比較すると、単語については、Wisper の書き起し結果の下線部分の1箇所を除き、正確に書き起こされていた。しかし、句読点に注目すると、適切な位置に句読点

3) OpenAI: <https://platform.openai.com/docs/models/gpt-3-5>

が欠落していることがわかる。次に、Wisper の出力を入力とし、句読点の自動挿入を行なった結果の例を表 3 に示す。挿入前と挿入後の文を比較すると、人間が挿入したような自然な句読点の挿入ができている。しかし、「～はありますでしょうか。」のような文章では、「～はありますでしょう。か」また、「～と考えておりますので、いちご狩り」のような文章では、「～と考えております。のでいちご狩り」と適切な位置に句点が挿入されない場合がある。

表 2 Amivoice と Wisper の書き起こし文の例

書き起こし方法	書き起こし結果
人手による書き起こし	6月の行事は、誕生日者該当なし。避難訓練と食事作り、いちご狩りを予定しております。このいちご狩りの時に2台のハイエースを使っていこうかなと考えておりますので1回行けたらいいなと思っています。
Wisper (midium)	6月の行事は誕生日会は外装だし、避難訓練と食事作りいちご狩りを予定しておりますこのいちご狩りの時に、2台のハイエースを使って行こうかなと考えておりますので1回で行けたらいいなと思っています。

表 3 BERT による句読点の挿入例

挿入前後	挿入結果
挿入前	毎日楽しく元気に過ごし互いに遠慮のない関わりをできるようにということなので年間通してこのように行なって勤務していきましょう
挿入後	毎日楽しく元気に過ごし、互いに遠慮のない関わりをできるように、ということなので、年間通して、このように行なって、勤務していきましょう。

表 4 オノマトペリストの例

おのまとペ	オノマトペ	おのまとペ + 「と」	オノマトペ + 「と」
ぐったり	グツタリ		
ぐったり	グツタリ		
はっ	ハッ	はっと	ハッと
ぱっ	パッ	ぱっと	パッと
はっきり	ハッキリ		
はらはら	ハラハラ		

4.2 オノマトペリストの作成

生成したオノマトペリストの例を表 4 に記載する。オノマトペリストの項目は、ひらがな表記、カタカナ表記、ひらがな表記+「と」、および、ひらがな表記+「ト」からなる4つの要素で構成する。介護や看護の現場で多用されているオノマトペに特化

したオノマトペ辞典 [6] から 140 語と 14 ヶ月分のケア会議データから 140 語以外のオノマトペ 32 語を加え、合計 172 語のオノマトペリストを生成した。

4.3 オノマトペと文章の抽出

作成したオノマトペリストのオノマトペを検索キーとして、14 ヶ月分のケア会議データから検索キーを含む文章の抽出を行なった。形態素解析解析器は MeCab⁴⁾、辞書は unidic を使用した。オノマトペは 52 種類、それらを含む文章は 243 文抽出できた。オノマトペ「ぐったり」で抽出した文章に A: 「Nさんは、ぐったり疲れてベッドの上にいることが多かった。」また、B: 「高熱でぐったりしている。」のような文章があった。文章 A の「ぐったり」と文章 B の「ぐったり」のように複数の意味を持つオノマトペがあることも明確になった。

5 考察

介護施設音声データのテキスト化において、音声データをテキスト化し句読点を自動挿入すると、適切な位置に句点が挿入されない問題がある。オノマトペの例文を生成する際に影響が出ると考える。句読点挿入後の文章全体を形態素解析をし、品詞と記号の連結可能性を考慮するなど再確認し修正する必要がある。オノマトペリストの単語は 172 語の登録ができた。小さい「っ」で終わるオノマトペの語尾に「と」を追加することで、小さい「っ」で終わるオノマトペを検索キーとしても、それを含む文章が抽出できるようになった。テキストデータからオノマトペを含む文章を抽出はできたが、一つのオノマトペが複数のニュアンスを持つ場合、ユーザが混乱しないように、さらに、ニュアンスの分類タグを付与するなど、工夫が必要と考える。

6 おわりに

やさしい日本語を用いたオノマトペ変換辞書の自動生成手法を提案し、介護施設音声データのテキスト化、介護施設に特化したオノマトペリストの作成、および、テキストデータからオノマトペとそれを含む文章の抽出を行った。実装はできたが、十分な評価結果とは言えない。今後は、各 step の評価とやさしい日本語と多言語の説明文の生成、例文の生成、関連画像の追加の実装を行い、介護に特化したオノマトペ変換辞書の完成を目指す。

4) MeCab: <https://taku910.github.io/mecab/>

謝辞

本研究はJSPS 科研費 20K11093 の助成を受けたものです。

参考文献

- [1] 厚生労働省, 介護人材の処遇改善, “<https://www.mhlw.go.jp/content/12300000/001144293.pdf>”, 2023.
- [2] 厚生労働省, 外国人雇用状況, “<https://www.mhlw.go.jp/content/11655000/000887553.pdf>”, 2021.
- [3] 神村初美, 介護のオノマトペは自然習得が可能なのか—EPA 候補者へのヒアリングから探る—, 日本語教育方法研究会誌 Vol.23 No.2, 2017.
- [4] 結城 政宗, 介護事例データベースを用いた介護支援提示モデルの構築, 自然言語処理学会, 2023.
- [5] 浅賀 千里, Web コーパスを用いたオノマトペ用例辞典の開発, 電子情報通信学会, 2007.
- [6] 吉永 尚, 廣部久美子, 介護・看護のオノマトペ—「もぐもぐ」・「ぜいぜい」は中国語・英語・ベトナム語・インドネシア語・ネパール語・タイ語でどう言う?—, 和泉書院, 2022.
- [7] 岩田 一成, 森 篤 嗣, 松下 達 彦, 中 島 明 則 (2015) 「やさ に ち チ ャ ッ カ ー」 URL : <http://www4414uj.sakura.ne.jp/Yasanichi1/nsindan/>