

令和 6 年 6 月 21 日現在

機関番号：62618

研究種目：挑戦的研究（萌芽）

研究期間：2021～2023

課題番号：21K18376

研究課題名（和文）フィールドデータのアーカイブに向けた問題点の整理と解決策

研究課題名（英文）Towards linguistic field research data archiving

研究代表者

中川 奈津子（Nakagawa, Natsuko）

大学共同利用機関法人人間文化研究機構国立国語研究所・研究系・准教授

研究者番号：50757870

交付決定額（研究期間全体）：（直接経費） 5,000,000円

研究成果の概要（和文）：少数の研究者間でのフィールドデータの共有を実現し、一部のデータは構造化して公開した (a)。フィールド調査のメタデータ構造をRDF化したモデル (b)、方言談話資料と辞書のTEI構造化を提案し、学会で発表した (c)。定期的な会合を行い、調査データのTEI構造化、データベース化、現地との協力体制、データの共有システム、メタデータなどについて議論を行った。また、特に日本本土の方言の言語コードが不足しているため、言語コードをどのように設定すれば良いかという議論を方言研究者、規格化の専門家などを交えて行った。

研究成果の学術的意義や社会的意義

データ共有のモデルケースを作ったことにより、研究データがオープンになるための仕組みに貢献できた。また、地域のデータがオープンになることで、地元の人々を中心とする市民にもデータを共有することができるようになった。

研究成果の概要（英文）：We realized the sharing of field data among a small number of researchers and made part of the data structured and publicly available (a). We proposed a model for RDF conversion of field survey metadata structures (b) and TEI structuring of dialect discourse materials and dictionaries, and presented it at a conference (c). Regular meetings were held to discuss the TEI structuring of survey data, database creation, cooperation frameworks with local communities, data sharing systems, and metadata. Additionally, since there is a shortage of language codes for dialects, particularly those of mainland Japan, we held discussions with dialect researchers and standardization experts on how to set these language codes.

研究分野：言語学

キーワード：言語学 人文情報学 アーカイビング 方言 琉球諸語

様式 C - 19、F - 19 - 1、Z - 19 (共通)

#### 1. 研究開始当初の背景

国内の言語学・方言学では一般的に、個々の言語学者が自分のフィールドデータを管理、研究で利用し、他の研究者がこの研究成果の妥当性を詳細に検討するためには、自分自身がそのフィールドに行って調査するか、個人的にデータをコピーさせてもらうしかない。このような慣習には、特に消滅危機言語・方言を研究する上で以下のような問題点がある。1. 研究者の主張の妥当性を確かめることが困難で、再現性も担保されない。2. 消滅危機言語・方言は、一部地域で話されている話者数の少ない言語でもあり、研究の再現性のために複数の研究者が同じ話者に調査協力を依頼すると話者の負担が増す。3. 個々の研究者が「データを囲い込んでいる」ことが問題だが、研究者の問題であるとも言い切れず、フィールドデータを発表しても業績にならないので発表する動機づけがないという構造的な問題をはらんでいる。4. 語彙リストや談話資料は業績となりうるが、発表媒体が主に紙であり、音声・動画の公開は研究者の技術に制約されやすく、また公開しても業績にならないので、公開する動機づけがない。5. 電子ジャーナルとして語彙リストや談話資料が公開されることもあるが、フォーマットは pdf が主流であり、定量的な分析に向かない。6. 調査に協力した話者を含むコミュニティへのデータの還元がなく、地元の言語や文化の継承に貢献できない。上述の問題を解決するためには、日本国内にデジタルアーカイブができる設備が必要である。

#### 2. 研究の目的

その第1歩として以下を遂行することを、本プロジェクトの目標とする。

- (a) 少数の研究者間でのフィールドデータの電子的共有 (1, 2 の解決策)
- (b) 共通メタデータの提案・改良 (1, 2 の解決策)
- (c) 電子ジャーナル以外でのデータの公開方法、引用方法の検討 (3, 4, 5, 6 の解決策) (d) 地域との協働 (6 の解決策)

#### 3. 研究の方法

少数の研究者間でのフィールドデータの共有を実現し、一部のデータは構造化して公開した (a)。フィールド調査のメタデータ構造を RDF 化したモデル (b)、方言談話資料と辞書の TEI 構造化を提案し、学会で発表した (c)。定期的な会合を行い、調査データの TEI 構造化、データベース化、現地との協力体制、データの共有システム、メタデータなどについて議論を行った。また、特に日本本土の方言の言語コードが不足しているため、言語コードをどのように設定すれば良いかという議論を方言研究者、規格化の専門家などを交えて行った。

#### 4. 研究成果

研究成果として以下の発表を行った。

- [1] Natsuko Nakagawa, So Miyagawa, A multi-media dictionary of endangered languages with TEI Lex-0: A case study of Hatoma, Yaeyama Ryukyuan, *Joint MEC TEI conference 2023*, 2023/9/7, 査読有.)
- [2] 中川奈津子「少数言語の資源」国際シンポジウム デジタル・ヒューマニティーズと研究基盤, 人文学のためのデジタル研究基盤及び DH の人材育成に関する事例報告, 2023/11/18, 査読無.
- [3] Natsuko Nakagawa, Jun Ogawa, Ikki Ohmukai, and Akihiro Kameda, Metadata of linguistics field research in RDF, *International Conference on Language Documentation and Conservation 2023*, 2023/3/, 査読有.

- [4] Natsuko Nakagawa, Masataka Ogawa, Nominal marking and word order in varieties of the Japonic family, *The 14th International Conference of the Association for Linguistic Typology*, Online & the University of Texas at Austin, 2022/12/, 査読有.)
- [5] Natsuko Nakagawa, Toshinobu Ogiso, The network of lexical resources in Japanese, *Tools of the Trade Conference*, Harvard University, 2023/3/, 査読無.
- [6] Natsuko Nakagawa, Digital dictionaries in Ryukyuan, *Tools of the Trade Conference*, Harvard University, 2023/3/, 査読無.
- [7] 中川奈津子「琉球諸語の言語調査データのアーカイビング」第7回デジタルアーカイブ学会研究大会サテライトセッション琉球文化のテキストアーカイビング, 2022/11/23, 査読無.
- [8] 中川奈津子、林由華「日本語における対比の八は焦点ではなく主題である」日本言語学会第164回大会, 2022/6/18, 査読有.
- [9] Natsuko Nakagawa, Yuka Hayashi Contrastive topic =gyaa in Ikema-Nishihara Miyakoan of Southern Ryukyus, *Japanese/Korean Linguistics*, 2021/10/9, 査読有.
- [10] Natsuko Nakagawa, An experimental method for eliciting (zero) case-marking in spoken Japonic dialects, *Japanese/Korean Linguistics pre-workshop*, 2021/10/8, 査読無.
- [11] 中川奈津子、加治工真市、宮良信詳「消滅危機言語の辞書データベースの構築と公開：「鳩間方言 音声語彙データベース」、「うちなぐち 活用辞典テキストデータベース」の事例報告」(人文学とコンピュータ研究会, 2021/5/21, 査読無.)
- [12] Kato Kanji, So Miyagawa, Natsuko Nakagawa, Language Atlas of Japanese and Ryukyuan (LAJaR), In Hahn et al. (Eds.), *Proceedings of the 6th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, 2024.  
<https://aclanthology.org/2024.sigtyp-1.7>
- [13] 中川 奈津子, 岡田 一祐, 永崎 研宣, 北崎 勇帆, 王 一凡, 曹 芳慧, 藤原 静香, 塚越 柚季, 小川 潤, 片倉 峻平, 左藤 仁宏, 王 雯璐, 石田 友梨, 宮川 創, 佐久間 祐惟, 塩井 祥子, 井上 慶淳, 村瀬 友洋, 関 慎太郎, 嵩井 里恵子, 渡邊 眞儀, 中町 信孝, 幾浦 裕之共著「日本語方言談話資料のTEIによる構造化の試み」, 『じんもんこん 2023 論文集』, 2023. <http://id.nii.ac.jp/1001/00231237/>
- [14] 中川奈津子単著「日琉諸語における主題標識の種類と格標識」『日本語文法学会第24回大会発表予稿集』, 2023.
- [15] 中川奈津子単著「琉球諸語の辞書の編纂方法」『日本語学 42(2)』明治書院, 2023.
- [16] 中川奈津子、宮川創、小川潤「方言研究資料のためのデータベース構築」『方言の研究 9』ひつじ書房, 2023.
- [17] 加治工真市、中川奈津子、加藤幹治「鳩間方言 音声例文データベース」, 国立国語研究所, 2022. <http://doi.org/10.15084/00003659>

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 0件／うち国際共著 0件／うちオープンアクセス 0件）

1. 著者名 中川奈津子、宮川創、小川潤	4. 巻 9
2. 論文標題 方言研究資料のためのデータベース構築	5. 発行年 2023年
3. 雑誌名 方言の研究	6. 最初と最後の頁 173-193
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 中川奈津子、加治工真市、宮良信詳	4. 巻 2021-CH-126
2. 論文標題 消滅危機言語の辞書データベースの構築と公開：「鳩間方言 音声語彙データベース」	5. 発行年 2021年
3. 雑誌名 研究報告人文科学とコンピュータ	6. 最初と最後の頁 1-6
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件（うち招待講演 1件／うち国際学会 3件）

1. 発表者名 Natsuko Nakagawa, Jun Ogawa, Ikki Ohmukai, and Akihiro Kameda
2. 発表標題 Metadata of linguistics field research in RDF
3. 学会等名 International Conference on Language Documentation and Conservation 2023（国際学会）
4. 発表年 2023年

1. 発表者名 Natsuko Nakagawa, Toshinobu Ogiso
2. 発表標題 The network of lexical resources in Japanese
3. 学会等名 Tools of the Trade Conference, Harvard University（招待講演）（国際学会）
4. 発表年 2023年

1. 発表者名 Natsuko Nakagawa
2. 発表標題 Digital dictionaries in Ryukyans
3. 学会等名 Tools of the Trade Conference, Harvard University (国際学会)
4. 発表年 2023年

1. 発表者名 中川奈津子
2. 発表標題 琉球諸語の言語調査データのアーカイピング
3. 学会等名 第7回デジタルアーカイブ学会研究大会サテライトセッション琉球文化のテキストアーカイピング
4. 発表年 2022年

1. 発表者名 中川奈津子、加治工真市、宮良信詳
2. 発表標題 消滅危機言語の辞書データベースの構築と公開：「鳩間方言 音声語彙データベース」
3. 学会等名 「人文学とコンピュータ」研究会
4. 発表年 2021年

〔図書〕 計1件

1. 著者名 仲間 博之 , 田窪 行則 , 岩崎 勝一 , 五十嵐 陽介 , 中川 奈津子	4. 発行年 2022年
2. 出版社 国立国語研究所言語変異研究領域	5. 総ページ数 428
3. 書名 南琉球宮古語池間方言辞典	

〔産業財産権〕

〔その他〕

-

## 6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	田窪 行則  (Takubo Yukinori)  (10154957)	大学共同利用機関法人人間文化研究機構国立国語研究所・大学共同利用機関等の部局等・所長    (62618)	
研究分担者	籠宮 隆之  (Kagomiya Takayuki)  (10528269)	大学共同利用機関法人人間文化研究機構国立国語研究所・言語資源開発センター・プロジェクト非常勤研究員    (62618)	
研究分担者	山田 真寛  (Yamada Masahiro)  (10734626)	大学共同利用機関法人人間文化研究機構国立国語研究所・研究系・准教授    (62618)	
研究分担者	小西 いずみ  (Konishi Izumi)  (60315736)	東京大学・大学院人文社会系研究科(文学部)・准教授    (12601)	
研究分担者	大向 一輝  (Ohmukai Ikki)  (30413925)	東京大学・大学院人文社会系研究科(文学部)・准教授    (12601)	
研究分担者	宮川 創  (Miyagawa So)  (40887345)	大学共同利用機関法人人間文化研究機構国立国語研究所・研究系・助教    (62618)	
研究分担者	小川 潤  (Ogawa Jun)  (50965390)	大学共同利用機関法人情報・システム研究機構(機構本部施設等)・データサイエンス共同利用基盤施設・特任研究員    (82657)	

## 7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------