

危機言語の記録・保存・復興・研究・習得のための沖縄語の自然会話コーパスの構築 — 展望と課題¹ —

新垣友子・新里瑠美子

要 旨

琉球諸語・沖縄語は、2009年、ユネスコにより消滅危機言語に指定され、保存・復興への取り組みが始まってきたが、自然会話の記録は見落とされてきた。自然会話データには、elicitation methodでは現れない言いさし、言い換え、よどみ、沈黙、間なども記される。それらを反映した自然会話コーパスを構築することができれば、従来の研究方式とは異なる文法記述の可能性や言語習得のための有効活用が期待できる。

本稿では、沖縄語の自然会話・簡易コーパス構築を目指し、その計画概要と進捗、そして直面している課題と展望をまとめた。主な課題としては、危機言語特有の課題ともいえる、堪能話者の協力確保、そして音声データを文字起こしする協力者確保の困難さを挙げた。主な展望としては、終助詞や接続詞・感嘆詞的用法、敬語、テンス・アスペクト、evidentialityなど様々な研究分野において自然会話であるからこそ抽出できる事例の一部を挙げ、今後の言語研究発展に貢献できる可能性を示した。

キーワード：危機言語、コーパス、自然会話、量的分析、検索可能簡易コーパス

1. はじめに

琉球諸語に関しては、国内外の博論、著書に見られる文法研究も多岐にわたり、幾多の優れた辞書にも恵まれている。しかし、そのほとんどが研究者による話者へのインタビューをもとにした文法記述で、話者同志の自由な会話を長時間録音し、その会話記録の中で、対象となる文法形式が実際どのように使われているかを分析した研究は少ない。またそのような研究を推進する上で欠かせない自然会話の記録も、現時点では僅少、皆無に近い。本研究は、これまで研究対象となることがほぼなかった沖縄首里・那覇を中心とした中南部の言葉（以下沖縄語と称す）の自然会話に的を絞り、録音・記録し、その後、対象文法項目などの検索が可能となる簡易コーパスの作成を目指す取り組みの中間報告である。

沖縄語を始めとする琉球諸語は、2009年、国際連合教育科学文化機構（ユネスコ）により消滅危機言語に指定され、行政、学術機関、民間における保存・復興への取り組みも始まり、功を奏しつつあるものの、自然会話の記録が見落とされてきた。自然会話の記述は、言語が実際にはどのように話されているかの忠実な記録で、それには、言語話者が何気なく使用する言いさし、言い換え、よどみ、沈黙、間などはもちろん、複数話者の同時発話、発話の共同構築、話題転換なども記される。また、話者の性別、年齢、出身地、言語行

動の様相などもメタデータとして含まれる。更に、それをコーパスという形に構築すると、ただ単に自然会話の資料というだけでなく、ある言語形式を検索・抽出することも可能になり、従来の研究方式とは異なる文法記述ができるようになる。

欧米の言語学や日本語研究においても、このような自然会話の記録はその意義が認められ注目を集めてきている。いわゆるCorpus LinguisticsやConversational Analysis (CA) の分野の研究がそれで、前者では、文法事項の実際の使用例を統計学の方法を駆使して、量的分析・記述するもので、後者では会話がどのように進行するかを、映像を通し、参加者の表情、視線、仕草なども考慮し、会話の構築の観点から、それらの関連性を分析している。昨今の日本語教育においても、作例の会話ではなく、本物に近い会話を重要視する傾向にあり、自然会話の意義が高まってきている。

日本語においては、国立国語研究所の「中納言」のようなコーパス検索アプリケーション²があり、その中に完成度の高い検索可能なコーパスや、学術機関における種々の簡易コーパスも作成されているが、沖縄語の場合は、未だ自然会話のデータ自体が乏しく、検索可能なコーパスは皆無である。その打開策として、本プロジェクトはこれまで看過されてきた自然会話に価値を認め、積極的に言語資料として取り入れ、それをオープンソースの簡易コーパスとして一般に公開す

るという意図のもとに推進するものである。本稿においては、現在進行中の簡易コーパス構築に向けて取り組んできた試みを紹介する。

本稿は、2章で琉球諸語に関する現状を概観し、3章でコーパスの種類や有用性をまとめる。4章でコーパス構築の計画概要と進捗を提示し、5章で現在直面している課題を挙げる。そして、6章において、これまでのデータ収集・整備のなかでみえてきた展望をまとめる。

2. 琉球諸語とは

奄美、沖縄で話される言葉は、2009年にユネスコが、6つの琉球諸語を危機言語として認めて以来 (Moseley 2009)、研究者の間では、「琉球諸語」という呼び方が定着してきた。具体的には、奄美語、国頭語、沖縄語、八重山語、宮古語、与那国語の6言語が琉球諸語に含まれている。近年、「しまくとぅば」という用語が用いられることが多いが、本稿では「琉球諸語」という名称を用い、「沖縄語」という場合は、沖縄島全域で話されている言語ではなく、中南部の言葉の総称を指す。琉球諸語と一口に言っても、その多様性は多岐にわたり、その言語差も深く、本土全域の地域語差に匹敵すると言われている。琉球諸語が日本語と相互理解ができないのは勿論のこと、琉球諸語の中でさえ相互理解は困難である。中でも北琉球で話されている奄美語、国頭語、沖縄語と、南琉球で話されている言語群 (宮古語、八重山語、与那国語) は、発音、語彙・文法的にも隔たりが大きく、相互理解はほぼ不可能である。

2009年、ユネスコに危機言語に認定されて以来、沖縄県においては、公民の両レベルで、沖縄語をはじめとする各地域の言語継承への取り組みも展開されてきている。メディアにおいてはROKラジオ沖縄の方言ニュースもあり、教育機関においては、沖縄語習得のための教科書も少なからず出版されている (例 西岡敏・仲原穰著『沖縄語の入門—たのしいウチナーグチ』、花蘭悟著『初級沖縄語』など)。また、1970年代から続く、学術誌『琉球の方言³』に見られるようなフィールドワークに基づく沖縄の各地方言の記述研究も盛んであり、奄美諸島から先島諸島にまたがる琉球弧の言葉は、幾多の優れた辞書にも恵まれている (国立国語研究所 (編) (1963) 『沖縄語辞典』、半田 (編著) (1999)

『琉球語辞典』など)。

言語の危機度に関しては、ユネスコ専門部会の言語の存続性と危機状況に関する基準から判断すると、世代間継承が途絶えている状態であり (石原2010)、実際、しまくとぅば県民意識調査⁴でも、沖縄語を始め各地の言葉を使う割合は減り、2013年の調査結果から最も低い結果となっている。このような事態を鑑みると若手の話者育成をサポートする教材開発に貢献するためにも早急のコーパスの構築が望まれる。

3. コーパスとは

本章では、本研究の基礎情報として、コーパスの種類を簡潔に述べて、コーパスがどのように言語研究に用いられるか、その有用性に関して概観する。

3.1 コーパスの定義・種類

はじめに、辞書によるコーパスの定義をあげてみると、英語では “a collection of written texts, especially the entire works of a particular author or a body of writing on a particular subject”, “a collection of written or spoken material in machine-readable forms, assembled for the purpose of linguistic research.” (文字化されたテキストの集積。特に特定の著者による全巻や、特定のテーマに関する一連の作品など。言語研究のための機械処理可能な書き言葉、または話し言葉資料の集積 (『オックスフォード新英英辞典』第2版) とされおり、日本語でも「語彙検索など、言語研究のための資料。特にコンピューターを利用してデータベース化された大規模な言語資料」 (『デジタル大辞泉』第二版) とされている。このように、コーパスとは、電子化された検索可能な言語テキストの集積であり、言語研究や言語教育などにより活用されるようになってきている。石川 (2012: 13) では、「1) 書き言葉や話し言葉等の現実の言語を、2) 大規模に、3) 基準に沿って網羅的・代表的に集約し、4) コンピューター上で処理できるデータとして保存し、5) 言語研究に使用するもの」、コーパスの5つの観点として挙げている。本研究で目指すコーパスは、危機言語のデータ収集の困難さゆえ、2) で挙げた「大規模に」という基準を満たすことは厳しいが、その予備的土台構築を目指している。

3.2 コーパスの種類

前節で述べたように、コーパスは言語分析のためのデータベースであるが、データの収集方法や収集データの内容、趣旨等も様々であり、その規模や用途は多岐にわたる。この節では、代表的なコーパスのみを挙げる。

まず、世界で初めてのコンピューターによるコーパスは、Brown Corpusで1964年、アメリカで完成した書き言葉のコーパスである。イギリスでは、書き言葉を収集したLancaster-Oslo/Bergen Corpus、また、話し言葉を収集した、London-Lund Corpus of Spoken Englishがあり、対話や、電話での会話やインタビュー、討論等幅広い文脈での会話が収録されている。その他にもThe British National Corpusや、The Bank of English等があり、辞典の編集や第二外国語学習に役立っている。

日本では、2004年に国立国語研究所が『日本語話し言葉コーパス』(Corpus of Spontaneous Japanese) (CSJ)を完成させ、2011年には『現代日本語書き言葉均衡コーパス』(Balanced Corpus of Contemporary Written Japanese) (BCCWJ) が作られた。話し言葉のコーパスは、他にも『名大会話コーパス』(NUCC)や『現日研・職場談話コーパス』(CWPC)、宇佐美まゆみ監修の『BTSJによる日本語話し言葉コーパス』(トランスクリプト・音声)、出版物としては自然会話を忠実に文字化した現代日本語研究会編『女性のことば・職場編』『男性のことば・職場編』(ひつじ書房)などがある。

本研究にもっとも関連があるのは、2019年に公開された『日本諸方言コーパス モニター公開版』(COJADS)で、『日本のふるさとことば集成』⁵全20巻と「各地方言収集緊急調査」⁶がデータ源となっている。COJADSには沖縄県国頭郡今帰仁村、平良市、与那国町の言葉も収録されており、貴重なデータであるが、本研究でターゲットにしている沖縄語は収集されていないため、本研究が簡易コーパス作成の一步となれば、補完的な位置づけになると期待できる。

3.3 コーパスの有用性

欧米の言語学会においては、すでに言語学の中にコーパス言語学の分野が確立されており、録画された自然会話を言語要素のみならず、参加者の表情、視線、

仕草、間の取り方など総合的に分析し、一見無秩序と思われる何気ない非言語行動にも意義を見出す会話分析の分野も発展を見せている。日本語教育の分野においては、教科書の会話をより自然に近いauthenticなものに近づけるよう推奨する傾向が強い。

日本国内においては、従来、elicitation methodを用いる研究も多かったが、現在は、前節でみたようにコーパスが整えられ、言語研究や言語学習に多用されてきている。特に言語研究に関しては、elicitation methodは、ターゲットとなる言語表現が抽出しやすいという利点はあるが、内省が働くことで、実際の運用との間にズレが生じる等の留意点もある。例えば、内省が働くと規範文法に則ったり、ある用法、ある語句を使うと言って実際は使わなかったり、使わないと言って使ったりということがある。また、自然会話ではないため、よどみやフィラーなどが出てこない。Web検索も同様で、記事の掲載削除が作者の意図で突然行われるため、流動的で、コーパスのような安定した結果を得ることは難しい⁷。

一方、自然会話コーパスは実際使用する表現がそのまま抽出できるので、用法や語句の使用法のばらつきなど、数で提示できる。一見何の規則性もなくランダムに使用されているように思える用法も精査されることにより、規則性が明確になったり、実際の言語の運用をより正確に記述することも可能となる。さらに、自然会話には、作例では出現しない口語特有の言語標識、言い直し、会話を円滑に進めるためのあいづち・終助詞などの使用、機能語・構文のゆれ、沖縄語から日本語へのいわゆるコードスイッチング、その他、予期しない表現や言葉の使用が見られる。一例を挙げるなら、「やはり、やっぱり、やっぱ」の変異をコーパスで精査したShinzato and Masuda (2009)があり、これらの三つの形が場の状況、対話相手との関係、話す内容により使い分けられていること、出現位置(文中か文末か)に偏りが見られること、年齢差により使用頻度が違うこと、更にフィラー的機能の有無などが統計的に実証されている。また、Shinzato (2014)では、文法化研究の領域で、「やはり、やっぱり、やっぱ」と「ばかり、ばっかり、ばっか」「あまり、あんまり」の平行関係を、コーパスを用いて分析している。

また、沖縄語の観点から、新垣(2006)では、貴族階級の男性と王家直系の女性との首里言葉の対談

(1955年:ラジオ放送)を文字化し、文法語彙の分析し、日本語訳化を施した。その対談の中で、王家直系の女性に対して、貴族の男性はnuNzu(目上の貴人に用いる二人称)を用いるが、女性の方は二人称代名詞を一切用いないという人称代名詞の使用に関する発見があった。また、応答の感動詞の使用についても、男性は首里における三段階の敬語体系の中でもっとも丁寧な返事?uu⁸「はい」を用いるが、女性は男性に対して、?Nという辞書にない形式を用いるという特徴が見られた。この研究は、社会的地位が非常に高い話者の対話という点で一般の会話とは異なる背景ではあったが、このような人称代名詞の使用や応答の感動詞の使用に関してelicitation methodでは現れにくい貴重な言語現象の明らかになった。

このように、自然会話においては、内省にもとづく文レベルの記述・分析では窺うことのできない、コミュニケーションの核に迫る豊かな言語事象を映し出してくれるため、実際の言語使用を記述・分析する意味で大変意義がある。ただし、elicitation method、コーパスを用いた研究のどちらかが正しいというわけではなく、言語研究において、双方が相補的な関係であるということはいままでのない。

4. 自然会話コーパス構築の計画概要と進捗状況

この章では、自然会話コーパス構築の計画とその進捗を概観する。2.2で挙げたような大規模コーパスと違い、大学や国立国語研究所のような組織ではなく、少人数で行っているプロジェクトであるため、作業のスピードは遅いが、各段階を丁寧に行っている。

表1 計画概要と進捗状況

	計画概要	進捗状況
①	(1)参加者へ配布する本プロジェクトの企画書ならびに同意書 (2)文字化に向けて、使用する表記、記号などに関する基本原則。(1)(2)の整備、確認。	終了。適宜、調整・修正を行っている。
②	参加者を募る。参加者を40人募る予定である。また、録音・録画時間は60時間(2時間×30組)	参加者、募集中。
③	録音データをAudacity上で、wavフォーマットに編集し、コーパスに使用可能な部分を抽出し、部分ファイルとしてまとめ、雑音の消音処理なども行う。	録音終了後、その都度作業中。

④	録音と、可能であれば、録画も進める。	現在は録音のみ実施。
⑤	サンプル会話を文字化する	終了。
⑥	文字化された自然会話データに文法・語彙分析を加える。	文字化の作業中。文法・語彙分析は文字化作業終了後に着手予定。
⑦	⑥に日本語訳を付加する。	部分的に実施。
⑧	メタデータを付加する。	未作業。
⑨	⑧をコーパスの専門家やコーパスの使用経験者に見ていただき、ご教示をいただく。	相談中。
⑩	⑨で得た知見をもとに、文字化の原則、メタデータの表示に必要な修正を行い、本格的な会話のコーパス化(④～⑧)へ移行する。	未作業。
⑪	可能であれば、⑩にタグを付し、検索機能を高めるよう工夫をほどこす。	未作業。
⑫	コーパス化された自然会話データと録画面面のリンク付をする。	未作業。

表1が示すように、①～⑤は終了、または継続となっており、録音作業、データ抽出、文字化の作業を継続している。文字化に必要な①の表記に関しては、コーパスの一般公開を目指していることもあり、研究者以外の利用者が使いやすいように国際音声記号(IPA)ではなく、通常のキーボードで入力容易なように整備した。具体的には、glottal stopはIPAの[ʔ]を使うべきであるが、取って[ʔ]で代用させるというように簡略化した。その他、表記の一貫性を保持するため、声門閉鎖音の有無によるミニマルペアの代表例のまとめや難解語彙表記の頻出語彙一覧を整備した。文字化の基本原則に関しては、宇佐美まゆみ(2007)の基本的な文字化の原則(Basic Transcription System for Japanese: BTSJ)の2019年改訂版を参考にした。BTSJには20ページにもわたる原則の詳細が記されているが、本プロジェクトには必要最低限の情報のみを付与とし、簡略化している。

②の参加者数は現段階ではまだ半数ほどで、規模の縮小を検討している。③のデータ編集には、GaragebandやAudacityなどのソフトを使用し、抽出作業を行っている。抽出方法に関しては、筆者らが同時に録音データを聞き、沖縄語が頻出している箇所、文化的、言語データの興味深い内容、表現の箇所を選定している。④の録音と録画に関しては、現在録音

のみで録画は行っていないが、今後録画も検討する予定である。現在、参加者が少ない状態であるが、会話の参加者の人数（2人のグループと2人以上のグループなど）、年齢差（同世代か世代差ありか）、性別、親疎関係なども考慮し、なるべく多様な組み合わせになるよう検討したい。⑤のサンプル会話の文字化は、①の文字化の原則を整えるために、初期段階で行った。

⑥の文字化された自然会話データに文法・語彙分析を加えるという作業に関しては、文字化の作業中で、文法・語彙分析は文字化が終了し次第、着手する予定である。文字化の作業をする人が非常に限られている状態で、文字化作業が難航している。この課題に関しては、5章で述べる。

本プロジェクトの達成目標としては、以下を目指している。①初期レベルの検索機能を持つ簡易自然会話コーパスの作成。②コーパス使用ガイドの作成。③一般公開すべく、幾つか候補の大学、機関のホームページへのリンクづけ、以上3点である。

本研究の中長期的な展望は、下記を検討中である。

①本プロジェクトのコーパスのデータ量を拡大（自然会話資料を春、夏、秋、冬の行事を含むもの）する。②英語での単語訳、文翻訳を入れた英語ベースの別コーパスも作成する。あるいは、このような要素も日本語ベースのコーパスにとり入れる。③形態素分析により頻度数が出せるコーパスを開発する。これにより統計学的に適切な分析が可能となり、より適確な数値を得ることが可能になる。④実際にコーパスを使い、リサーチを行う。コーパスを使い、生の例文を用いた沖縄語習得のための教材を作成する。

以上、これまでの取り組みと進捗状況を概観してきた。次章では、本プロジェクトにおける課題を挙げる。

5. 課題

危機言語の自然会話データの収集は、ただでさえ困難が伴うが、新型コロナウイルス感染拡大により、2020年-2021年の間、録音は限られてしまわざるを得なかった。この章では、その他の点も含めて以下の5点の課題を挙げる。

1. 話者の確保（話者間の相性、人間関係、日時設定）
録音に入る前に人間関係の構築を入念におこなう必要がある。

2. 趣旨の理解（自然対話ではなくモノローグになりがち）
3. 環境の整備（場所、録音機、コロナ対策）
4. 話者間のレベルのバランス（秀でた堪能話者がいると談話になりにくいいため、同等が理想的。）
5. 文字化作業への協力者確保

まず、1番目に話者の確保であるが、しまくとぅば普及センター、沖縄語普及協議会、宜野湾市うちなあくち会、北中城村文化協会しまくとぅば部等の方々、知人等を通して話者を紹介してもらい、協力者を募った。調査者と協力者の関係を築くことが重要であり、2番目に挙げた調査の趣旨を理解してもらうことが不可欠であるため、研究倫理の配慮も併せて初回は丁寧に説明を行った。自然に会話してもらうことが大切であるため、話がしやすいように雑誌や写真等資料を準備してもらうこともあった。

3番目の環境の整備は、コロナ禍になる前からの共通の課題としては、録音機器に慣れてもらうことであるが、その段階がクリアできると、ジェスチャーや雑誌をめくる音、筆記用具の音などが入りやすいため、消音配慮が必要である。また、協力者の話に興味を示しつつも、調査者は会話に入らないという点も注意点として挙げられる。調査者に向けて、日本語で説明が入ったり、部外者がいる違和感などで、自然な会話にならないことがあるためである。また、コロナ禍以降の録音としては、感染予防対策として、パーティションを使用したのが、会話のキャッチボールが上手くいかず、不自然な展開となったり、話者がパーティションの前を離れて席を移動したりすることがあった。慣れることで改善される可能性もあるが、今後、試行錯誤しながら対応策が必要である。

4番目に挙げた点は、話者の組み合わせである。秀でた堪能話者がいると、その話者のみがターンをもち、他の参加者はあいづちに徹するというスタイルになりがちで談話にはなりにくい傾向がある。様々な組み合わせが理想的ではあるが、熟達度が同等の話者が好ましい。

最後に5点目の文字化作業についてであるが、現在、主に文字化作業を担当しているのは本人が堪能話者であるというパターンと、20代の若い世代が祖父母と共に録音データを聞き、祖父母に手伝ってもらいながら

文字起こしをするというパターンである。しかし、いずれのパターンも継続的に協力者力者を確保することが難航しており、その背景には、作業自体も大変煩雑であり、事務的手続きも煩雑であるという点が挙げられる。

若者が協力者として文字化に関わる場合は、次章で述べるように、本人の沖縄語習得に役に立つという利点はあるが、祖父母などの堪能話者の協力がなければ困難であるため、そのような条件を満たし、かつ時間的にも協力体制が整う人材の確保が困難である。文字化練習用の教材を作って事前に配布しているが、その教材をこなせる人材も少ない状態である。

6. 展望

前章で課題を挙げたが、この簡易コーパス構築の試みから得られる暫定的な成果もあり、展望もある。この章では、その展望について3点述べたい。まず、1点目にコーパスだからこそ観察できる興味深い言語形式が挙げられる。特にyaa, doo, dee, saa, Qsa等終助詞が多用されていることが分かる。終助詞研究に関しては、その分野の遅れがすでに指摘され、崎原（2018）に見られる文字資料による分析もあるが、自然会話における記述分析はまだない。コーパスを用いた分析はその点でも意義があると思われる。更に、自然会話においては、elicitation methodではまず出てこないであろう、yakutu, anshigaなどが文頭に現れる用法も目立つ。作文と違い実際の会話では、句や節を繋いで一文がかなり長い場合が多く、yakutyoo, anshigayoo等の接続表現で、ターン交替が起こることが多い等が挙げられる。日本言語学の中の文法化の分野では、文中の「から」「けど」がコピュラと合体し、「だから」「だけど」として接続詞となる脱範疇化現象が取り上げられるが（Higashiizumi 2006, Shinzato 2018）、沖縄語のyakutu, anshigaもそれに類似した文法化現象として注目される。沖縄語の感嘆詞の用法は、接続詞用法から更に進んだ文法化の段階として興味深い。

加えて、うちなーやまとうぐち⁹の「だからよ」との比較なども発展可能性のある研究テーマである。葦原（2016）は、「だからよ」を単なるあいづちではなく、「接続表現」の一つとして捉え、文脈展開機能を示している。日本語の「だから」が有している文脈展

開機能は3種に分類されているが、うちなーやまとうぐちの「だからよ」及び派生表現には、9種の文脈展開機能が見られると報告されており（葦原 2016）、日本語の「だから」と共通点がありながらも、独自の機能も持っていることが分かる。実際のコーパス構築のために収集したデータの中にも、「だからよ」の汎用性を認める話者の発言もあり、今後、日本語の「だから」と沖縄語のyakutyoo等、またうちなーやまとうぐちの「だからよ」との比較研究の可能性が期待できる。

その他にも、saiやtaiなどの敬語表現の使用方法に関しても、自然会話のデータは有効である。3.3で挙げた新垣（2006）の対話分析では、いずれの形式も一切表れなかったが、これまで収集したデータの中でもsai, taiがよく用いられるグループと、そうでないグループに分かれる傾向がみられる。

また、沖縄語には日本語にはないテンス・アスペクトの体系があり、どのようにして情報を獲得し、そしてそれらを内在化していくか認識容態に関する研究がなされてきたが（Shinzato 1991, 2003）、それらのカテゴリーに関連してevidentiality（証拠性）の体系的な研究も報告されている（Arakaki 2013）。evidentialityとは、文で表される情報の源を示す範疇で、その定義や分類に関しては研究者によって異なるが¹⁰、これまで収集したコーパスのデータにもevidentialityを示す例があり、例えば、伝聞証拠を示すNdiが自然会話でよく用いられている。このNdiの使用に関して、伝聞情報であるにもかかわらず伝聞証拠Ndiの付与が免除され、断定形式が用いられることがある事例が報告されているが¹¹、沖縄語における伝聞証拠の言語化は日本語よりも厳格であると指摘されているが（新垣 2018）、今後、そのような日本語との対照研究もより発展する可能性がある。他にも、直接証拠や間接証拠のevidentialsの用法に関する自然会話データの蓄積があれば、実際の文脈でどのように用いられているか検証することができる。

このように終助詞や接続詞・感嘆詞的用法、敬語、テンス・アスペクト、evidentialityなど様々な研究分野において自然会話から実際の用例を抽出できることは、語彙、文法等の言語研究にとって有益なデータとなることが期待できる。

2点目に、上記1点目に述べたように、実際の会話を文字で観察した際、テキストやモノローグ、辞書の

用例とかなり違うため、新話者がコーパスを利用すれば自然な会話が学べると期待できる。また、危機言語の新話者育成にも貢献できる。Shinzato and Iwasaki (forthcoming) はそのような新しい視点から試みた教科書作りである。

最後に、3点目として文字化作業を行う若者の沖縄語に対する理解や学ぶ動機に繋がるという点が挙げられる。実際に文字化に関わった20代の協力者にヒアリングしてみると、ほとんど沖縄語を理解できなかった初期の状態から徐々に慣れ、聴き取り能力や単語力が向上し、あいづちのタイミング等、会話のリズムがつかめるようになってきたということが判明した。また、言語面だけでなく地元の歴史や文化を学ぶ意欲が湧いたということ報告も受けている。今後、文字化の作業に関わることで、地元の言語や歴史文化を学ぶきっかけに繋がる可能性がある。

7. おわりに

危機言語のコーパス作成は、話者を探して録音することが難しく、さらにそれを文字化する協力者の確保の難しさ等、多くの課題に直面することが多い。その困難な状況に加えて、2020年度は新型コロナウイルスの影響により調査の中断を余儀なくされるなか、限られた状況で出来る事を重ねてきた。課題もあるが、前節で述べたように、この簡易コーパス構築の試みから得られる暫定的な成果もあり、今後の言語研究に貢献できる可能性も示すことができた。

注釈

¹ 謝辞：本研究はJSPS科学研費補助金（科研費）19K13171の助成を受けたものです。

² 形態素情報を用い、検索結果がExcel形式でダウンロード可能で、Excelのピボットテーブル機能を用いて頻度表を作成したり、フィルター機能を使って特定の表現を抽出することが出来る。詳細は、中俣（2021）参照。

³ 法政大学沖縄文化研究所発行。

⁴ 2020年（令和2年）『しまくとうば県民意識調査報告書』「しまくとうば」に対する県民の意識は全体として、高い好感度を示しており、全体で84.8%が「親しみを持っている」と答えている。しかし、

「①しまくとうばを主に使う（3.6%）」、「②しまくとうばと共通語を同じくらい使う（17.7%）」、「③挨拶程度使う（21.9%）」と使用する人の割合の合計は、43.2%と半数に満たない。2019（令和元年）の56.7%から大きく減少している。一方、「④あまり使わない（37%）」、「⑤全く使わない（18%）」と答えた人の割合は、合計55%と半数以上であった。

⁵ 国立国語研究所（編）2001～2008年。国書刊行会発行。

⁶ 文化庁が日本全国47都道府県で1977～1985年に行った調査。老若男女の自然談話のうち、その地域の伝統的言葉がもっとも現れている部分がデータベース化されている。

⁷ Google等のWeb検索は、検索の対象が秒単位で増大しているため結果が安定せず、一時的に流行っているごく稀な例も拾うことがあり、検索結果を処理しにくい。一方、コーパスで得られる結果は安定しており、非常に稀にしか使われない例は大量には出てこない。また、Excelで検索結果を容易に処理できるという利点がある（中俣 2021：8）。

⁸ 首里言葉では、貴族、士族、平民という3種の階級があり、話者が属する階級と、性別、年齢の差異に従って厳重に敬語が使い分けられていた（『沖縄語辞典』1963:19-20）。この対談では、女性の方が位や年齢が高かったため、この感動詞が用いられたと考えられるが、女性の方は、男性に対して、目下で使用するとされている?oohooも使用していない。日本語のあいづちも、「はい」はあらたまった時に使用され、「うん」はくだけた場面で使用されるとされるが、実際の談話を分析してみると「うん」の使用は「はい」の2倍以上で、フォーマル場面にも多用される事例が報告されている（中島 2011）。この女性が使用した?Nの使用に関しても、さらに調査が必要である。

⁹ 琉球の言葉を母語とする人々が日本語を第二言語として習得しようとした時に形成された接触言語。ロング（2010）はうちな一やまとうぐちを従来研究されてきたSingapore型クレオイド、Africaans型の2種類とも異なる新しいタイプの接触言語としている。

¹⁰ 例えば、Arakaki（2013）は、証拠性を独立した文法範疇とみなす狭義の定義（Aikhenvald:2004, de Haan:1999）を採用し、沖縄語の証拠性に直接証拠（direct）、間接証拠（inference, assumed, reportative）などが存在する

と指摘している。宮良(2019:129)では、「継続動作が終了したところに居合わせた場合(1)と、動作の起動が成立したところに居合わせた場合(2)」の二つの目撃形を示す形式を挙げている。

- ¹¹ 新垣(2018)は、Cuzco Quechua語の同様の現象に関して、assimilationの概念と「情報のなわ張り理論」が関係するという仮説(Faller 2002)を参考に、その2つの観点からNdiの付与の免除の要因を分析している。その結果、沖縄語における伝聞証拠の省略は日本語よりも条件が厳しく、(1)情報提供者本人やその人と関係がある情報を受け取った場合、(2)一定の情報処理課程を経てassimilationされた場合の(1)と(2)の両条件が満たされていないといけなことが指摘されている。

参考文献

- 葦原恭子(2016)「沖縄県の地域共通語『だからよ』の談話における機能」『Southern review: Studies in foreign language & literature』(31) 75-90. 沖縄外国文学会.
- 新垣友子(2006)「『旧正と大晦日の思い出』における敬語表現の研究」伊豆山敦子編『放送記録テープによる琉球・首里方言：服部四郎博士遺品』東京外国語大学アジア・アフリカ言語文化研究所 65-89.
- 新垣友子(2018)「琉球諸語・沖縄語における証拠性：伝聞証拠-Ndiの非言語化現象」『沖縄キリスト教大学院大学論集』(15), 1-11. 沖縄キリスト教大学院大学.
- 石川慎一郎(2012)『ベーシックコーパス言語学』ひつじ書房.
- 石原昌英(2010)「琉球語の存続と危機度 — 逆行的言語シフトは可能か」パトリック・ハインリッヒ、松尾慎(編著)『東アジアにおける言語復興』111-149. 三元社.
- 宇佐美まゆみ(2007)「改訂版：基本的な文字化の原則 (Basic Transcription System for Japanese: BTSJ) 2007年3月31日改訂版」『談話研究と日本語教育の有機的統合のための基礎的研究とマルチメディア教材の試作』平成15-18年度 科学研究費補助金 基盤研究B(2)(研究代表者 宇佐美まゆみ) 研究成果報告書.
- 沖縄県文化振興課(2020)『沖縄県しまくとぅば意識調査報告書』
- <https://www.pref.okinawa.jp/site/bunka-sports/bunka/r2simakutolubakenminisikityousa.html>
(アクセス日2021年11月6日)
- 『オックスフォード新英英辞典(第2版)』(2005) オックスフォード大学出版局(電子辞書)
- 現代日本語研究会(1997)『女性のことば・職場編』ひつじ書房.
- 現代日本語研究会(2002)『男性のことば・職場編』ひつじ書房.
- 国立国語研究所(1963)『沖縄語辞典』財務省印刷局.
- 崎原正志(2018)『琉球語沖縄首里方言のモダリティ：叙述・実行・質問のモダリティを中心に』琉球大学大学院人分社会科学研究科博士論文.
- 『デジタル大辞泉(第二版)』(2012) 小学館
- 中島悦子(2011)『自然談話の文法—疑問表現・応答詞・あいづち・フィラー・無助詞—』おうふう.
- 中俣尚己(2021)『「中納言」を活用したコーパス日本語研究入門』ひつじ書房.
- 花蘭悟(2020)『初級日本語』研究社.
- 半田一郎(1999)『琉球語辞典』大学書林.
- 西岡敏・仲原穰(2000)『沖縄語の入門—たのしいウチナーグチ』白水社.
- 宮良信詳(2019)『うちなーぐち しくみと解説』沖縄時事出版.
- ロング、ダニエル(2010)「言語接触論から見たウチナーヤマトウグチの分類」『人文学報 日本語教育学』No.428. 1-30. 首都大学東京都市教養学部人文・社会系.
- Aikhenvald, Alexandra. 2004. *Evidentiality*. Oxford: Oxford University Press.
- Arakaki Tomoko. (2013) *Evidentials in Ryukyuan: The Shuri Variety of Luchuan—A Typological and Theoretical Study of Grammatical Evidentiality*. (Brill's Studies in Language, Cognition and Culture,4) Leiden・Boston: Brill.
- De Haan, Ferdinand. 1999. *Evidentiality and Epistemic modality: Setting boundaries,*” *Southwest Journal of Linguistics* 18:83-102.
- Faller, Martina. 2002. *Semantics and Pragmatics of Evidentials in Cuzco Quechua*. Ph.D. thesis, Stanford University.
- Higashiizumi, Yuko. (2006) *From a Subordinate*

- Clause to an Independent Clause: A History of English Because-Clause and Japanese Kara-Clause*. Tokyo: Hituzi Syobo Publishing.
- Moseley, C.(ed.) (2009) *Atlas of the World's Languages in Danger*. Paris: UNESCO.
<http://www.unesco.org/culture/languages-atlas/> (アクセス日 2021年11月6日).
- Shinzato, Rumiko and Kyoko Masuda. 2009. Morphophonological variability and form-function regularity: A usage-based approach to the Japanese modal adverb *yahari/yappari/yappa*. *Language Sciences* 31: 813-836.
- Shinzato, Rumiko and Shoichi Iwasaki. (forthcoming) *Basic Okinawan: From Conversation to Grammar*. Honolulu: University of Hawaii Press.
- Shinzato, Rumiko. 1991. Epistemic properties of temporal auxiliaries: A case study from Okinawan, Japanese and Old Japanese. *Linguistics* 29.1: 53-77.
- Shinzato, Rumiko. 2003. Experiencing self vs. observing self: The semantics of stative extensions in Japanese. *Language Sciences* 25.2: 211-238.
- Shinzato, Rumiko. 2014. From degree/manner adverbs to pragmatic particles in Japanese: A corpus-based approach to the diachronic parallel developments of *amari*, *bakari*, and *yahari*. In: Andreas H. Jucker and Irma Taavitsainen (eds.), *Diachronic Corpus Pragmatics: Intersections and Interaction*, 77-105. Amsterdam: John Benjamins Publishing Company.
- Shinzato, Rumiko. 2018. Layered structure, positional shifts, and grammaticalization. In: Yoko Hasegawa (ed.), *The Cambridge Handbook of Japanese Linguistics*, 40-64. Cambridge: Cambridge University Press.

Building a Spoken Natural Conversation Corpus of the Okinawan Language for Documentation, Revitalization, Investigation, and Acquisition of Endangered Languages — Challenges and Prospects —

Tomoko Arakaki · Rumiko Shinzato

Abstract

Since the *UNESCO Atlas of the World's Languages in Danger* included six Ryukyuan languages in 2009, various attempts at the revitalization of Okinawan language have been taken. However, the necessary effort in recording and documenting natural conversation has been largely overlooked by scholars. The data that comprise natural conversations appear as filler, hedging, and silence during the conversation which are difficult to extract from the elicitation method. Building a corpus that indicates such data should contribute to the more practical descriptions of grammar and language acquisition that usual methods have, so far, been unable to provide.

The purposes of this paper are to offer a brief overview of the ongoing corpus building project of Okinawan language including a discussion of the unexpected challenges we are facing as well as the potentials the data, hitherto, appear to suggest. A major challenge has been the lack of human resources in terms of both fluent speakers and transcribers. Regarding the project's potential to motivate future research was the discovery of various expressions (detectable only through observation of natural conversation) such as the use of final particles, conjunctions, exclamatory and honorific expressions, as well as possible key findings in the areas of tense, aspect evidentiality, etc.

Keywords: endangered languages, corpus, natural conversation, quantitative analysis, preliminary searchable corpus