「何もしなければ」消滅してしまう琉球のことばを、記録、共有して、敬称するために

代表研究者	林　由華	国立国語研究所／日本学術振興会　特別研究員
共同研究者	山田　真寛	国立国語研究所　言語変異領域　准教授
共同研究者	山本　史	京都市立芸術大学 大学院美術研究科　大学院生
共同研究者	小川　晋史	熊本県立大学　文学部　准教授

概要

世界に 6000-8000 あるとされる言語の半数が、「何もしなければ」今世紀中に消滅すると言われている。UNESCO(2009)によれば、日本国内に少なくとも８つの消滅危機言語が存在し、その内の６つは、日本語と祖先を同じくする琉球諸語である。地域言語の消滅は、言語の多様性、言語が支える文化や価値観の多様性、また言語コミュニティのアイデンティティの喪失にもつながる可能性がある。琉球諸語の流暢な話者がおおむね 60 代以上であることを考慮すると、これらの言語の記録・継承は喫緊の課題である。本研究は、このような事態に鑑み、情報通信技術を利用することで、言語と文化の記録を蓄積し、同時に若い世代への言語継承サポートと、文法記述研究の促進を目指す総合的な消滅危機言語の実践研究である。

本助成では、上記の目的のもと、（１）琉球諸語の学習・世代間継承に特化した自然談話資料（字幕付き動画、文意訳付きスクリプト）の制作と蓄積、（２）これを誰もが容易に行うためのジェネレータ・オンラインインターフェースの開発、さらに（３）研究者や地域言語コミュニティにこれら普及させ、上述の自然談話資料の蓄積および（２）を利用したインターネット上で利用可能な地域言語学習コンテンツの制作を行うことで、消滅危機言語の記録・保存と、継承・復興のための研究を行った。具体的な成果としては、字幕・スクリプト付動画の談話資料を表示するジェネレータ、および動画その他のコンテンツを公開するためのウェブサイトを作成した（「しまのことば、しまからのことば」 http://shimashima.ryukyu/）。また、地域コミュニティと協働で談話資料収集やコンテンツ作成を行い、上記ジェネレータやウェブサイトを利用してコンテンツを公開した。そのほか、ウェブサイトでは、地域語の絵本などの既存の言語資料の電子コンテンツ化したものや、地域語による文章などを、琉球各地の作者ごとに順次掲載・公開している。

1 背景と目的

1-1 背景

琉球諸語は、日本語と祖先を同じくする唯一の姉妹語群である。7 世紀より前に日本語と分岐した後、琉球列島という日本とは地理的にも離れた地で、また琉球王朝という別国家のもと、日本語とは別の発展の道を歩んだ。集落ごとに異なる多様な方言群から成っており、大まかに奄美語、沖縄語、宮古語、八重山語、与那国語の５つに区分することができる（UNESCO は奄美語と沖縄語を合わせたものを３つに分け、全部で６つの区分を提示している）。これらは日本語とは互いに通じ合わないだけでなく、それぞれの言語間でも相互理解性がない。これらの言語は、日本語の古い姿、日本語と琉球語の共通祖先の姿の解明の重要なデータとなるなど、研究対象として貴重であるだけでなく、地域固有の文化や価値観を支えるものでもあり、地域コミュニティにとっては先祖から受け継いだ遺産である。しかし、琉球列島の島々が明治以降に日本に組み込まれて以来、日本語が「標準語」として教育や行政などの場から普及し、現在は地域の高齢者が使用するのみであり、次世代に継承されない言語群になっている。本研究の代表者及び共同研究者は、これまでの研究の中で、琉球列島での言語の調査記述研究に加え、研究者だけでなく地域コミュニティのための記録保存活動を行ってきた。

言語の記録には、参照文法・辞書・談話資料の３つが必須であると言われている（Tsunoda 2005）。琉球諸語に関しては、2000 年代から包括的な文法体系の記述が進められ、研究者・言語コミュニティメンバーによる辞書制作も進められている。自然談話資料は、実際の言語使用を知ることができる、音声・動画資料およびこれを文字化・分析したものが該当するが、各言語の大規模なコーパスを構築できるほど十分な蓄積はなく、言語学習・継承に利用可能なものは、ほとんどない。自然談話資料の作成には膨大な時間が必要ではあ

るが、参照文法や辞書の作成に比べると、学術的な訓練をそれほど必要としない（山田 2015)。よって、限られた数の地域言語研究者が行うだけでなく、地域言語コミュニティメンバーが自然談話資料を容易に記録・蓄積できるようになれば、消滅危機言語の記録が飛躍的に進むことが期待できる。さらにアノテーション・字幕付き動画は、言語学習にとって最適な学習教材の一つであり、この蓄積は消滅危機言語の復興にも貢献する。

琉球諸語は、日本語の仮名文字では表記できない音を多く持つが、本研究メンバーである小川、林、山田は、琉球諸語を正確、統一的かつ容易に表記するための、ひらがなにいくつかの補助記号を追加した表記法を考案し（小川 2015)、山本とともに民話絵本の制作（Yamada et al. 2016、下地ほか 2017）や、地域コミュニティへの普及活動を行っている。この表記法は言語学者以外でも容易に利用することができるため、地域言語コミュニティメンバーが地域言語の絵本や辞書の制作などに利用している（松村ほか 2016 など)。さらに、林、山田が参加しているデジタル博物館プロジェクト（田窪ほか 2014：kikigengo.jp）では、多数の琉球諸語研究者がアノテーションと字幕付き動画の制作を行っている。誰もが容易にこれを行うことができるようなジェネレータ・オンラインインターフェースを開発すれば、研究者と地域言語コミュニティが協働したり、地域言語コミュニティメンバーが独自に自分たちの言語資料を記録し、若い世代への言語継承に利用したりすることが可能になる。これらを実現するために、本研究を実践するに至った。

1-2 本研究の目的

上記のような背景のもと、本研究では、地域コミュニティのメンバーなど、専門的知識を持つ研究者以外でも、言語の記録や学習・継承において重要となる字幕・アノテーション付談話資料を中心としたコンテンツの作成やオンライン公開、また既存の紙媒体資料のデジタル化・オンライン化が容易に行える仕組みを作り、それを実践につなげることを目的としている。これにより、危機的状況にある言語資料が、研究者だけでなく地域コミュニティからも盛んに生まれ、蓄積し共有されていくことを目指す。

このような資料の蓄積と共有は、実は、話者コミュニティが非常に大きい日本語標準語などの大言語であれば自然に行われていることでもあるが、危機言語である少数言語（minority language）はしばしばその基盤を持っていない。その基盤を作り、少数言語であってもそのような営みを自然に行うことができるようにするというのも、代表者らのプロジェクト全体が目指すところである。大言語と異なり、少数言語は、そのコンテンツを共有する可能性のある人の絶対数は比較にならないほど少ない。ウェブでの公開は、その中でも共有できる人の市場を広げる一つの方法である。このため、これまで紙媒体などでローカルに共有されてきた資料をデジタル化・オンライン化し、ウェブを介して一か所でより多くの人が共有できるようにすることも、平行して行っていく。

2 プロジェクトの全体像と本研究との関連

ここでは、代表者らが行う関連研究を含めたプロジェクトの全体像を示しながら、特に本研究で行うことについて具体的に述べる。

代表者らはこれまで、琉球列島各地で言語調査を行いつつ、記録保存活動を続けてきた。その中で、多くのテキスト資料や音声・動画などを収集し備蓄しており、随時書籍などの紙媒体での公表、オンライン公開などを行っている。代表者らは、これらの活動を地域コミュニティと協働で、また地域コミュニティ主体でできるようにすることを目指している。次の図 1 では、その概要を図示している。図のうち点線の矢印は、これまでの資料公開の流れで、本研究ではそれを活かしつつも、赤と緑の矢印で示されるデジタルコンテンツとしてのオンライン公開を行う。1 節でも述べたように、記録保存や学習に際しては、字幕やその他アノテーションのある動画が効果的なコンテンツとなるが、このようなコンテンツの作成は誰もが簡単にできるわけではなく、本研究おいては特に、そのような字幕・アノテーション付談話資料を容易にするシステムを作ること、また既存の紙媒体の資料をデジタイズしオンライン化することを主たる目的としている。このために、動画や音声およびそれにつけられた書き起こしや訳などのファイルから簡単に字幕・アノテーション付談話資料を生成するオンラインジェネレータの開発を行う。本研究では、それらのコンテンツ作成や公開、既存資料のオンライン公開、およびそれに先立つ資料収集について、実際に地域コミュニティと協働で実践する。

このうち、ジェネレータについて、計画当初予定していた機能のすべてをもたせることが技術的に難しい

ことが研究途中でわかった。このため、当初は主として別財源で改めて作成する予定であった、コンテンツ公開のためのポータルサイトを、本研究内で作成した。

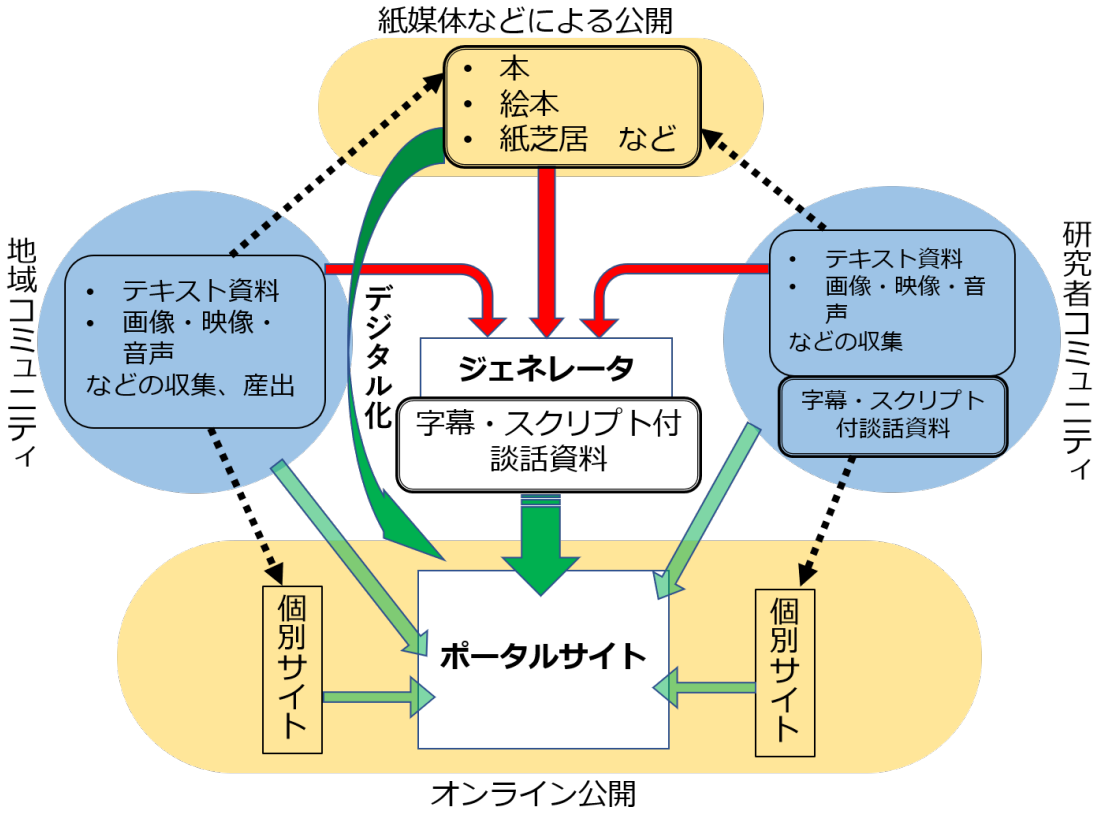


図 1：プロジェクト全体像

ジェネレータは、当初、(i) 字幕付き動画、(ii-a) 原語の書き起こし、(ii-b) 文意訳、(ii-c) 逐語訳、(ii-d) 言語学的分析アノテーション、それらに加え、ウェブブラウザ上で、(i)の再生表示されている字幕部分が、(ii-a)上でハイライトされるしくみ　を備えた談話資料を生成することを計画していた(図 2 参照)。このうち、(ii-cd) については現時点では含めることが困難だったため、本研究内では、上記の (i) (iia~d) のうち、(i)と(iiab) が示されるスクリプト、およびそれがハイライトされる仕組みを備えた談話資料ジェネレータを作成している（詳細は 3.1 節)。

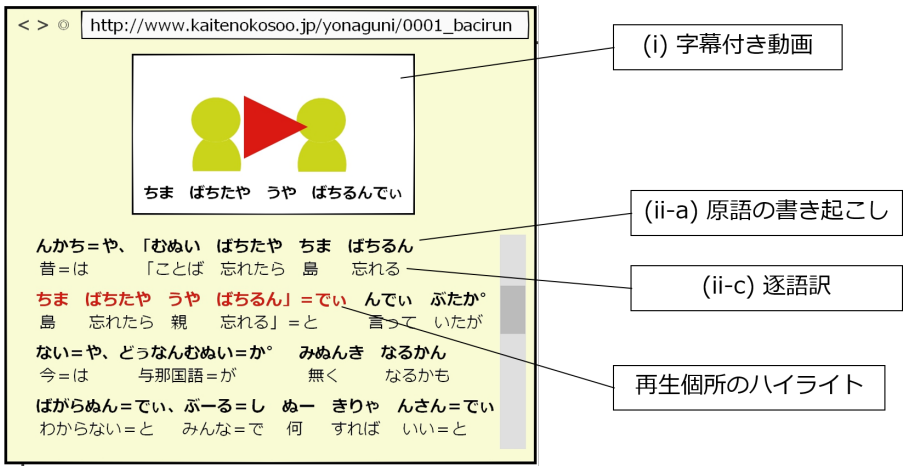


図 2：ジェネレータによって生成される資料のモデル（予定）

3 成果

以下では、1 節や 2 節で示した計画をもとに行った実践研究の成果について、まずは成果物として作成されたジェネレータ（3-1 節）とポータルサイト（3-2 節）について説明したあと、それらを利用した地域コミュニティとの協働活動や資料収集などについて（3-3 節）述べる。

3-1 ジェネレータの作成

危機言語の記録保存および公開に関してオンライン上で閲覧できるアノテーション付き動画・音声はコンテンツとして大変有用なもので、実際にさまざまなプロジェクトで用いられている。しかし、動画とスクリプトが別々に表示される、音声が早すぎてスクリプトの利用が不便、特殊な制作環境を必要とするなど、多くの人にとって作成また学習に利用するのには問題があった。本研究では、これらの問題に鑑み、学習に利用しやすく、また簡単な手順でアップロードできるジェネレータを作成した。これはウェブブラウザを利用して用いるもので、現在は本研究で作成したウェブサイト（「しまのことば、しまからのことば」http://shimashima.ryukyu/）および代表者のウェブサイト（「みゃーくの方言―宮古語諸方言の記録」http://miyakogo.ryukyu/）で稼働している。また、「しまのことば、しまからのことば」（http://shimashima.ryukyu/）にてダウンロードできるようにする予定があり、他サイトにも自由に組み込み可能である。以下では、ジェネレータで生成される資料のデザイン、生成方法について説明したうえで、作成のうえで生じた問題点について述べる。

（1）生成される談話資料のデザイン

生成される資料は、以下の図 3 のように表示される。

図 3：生成される談話資料のデザイン

字幕付きの動画と同一画面にスクリプトを表示、スクリプト内では書き起こしおよび訳が見られ、再生個所にハイライトがつく。これによって、スクリプトで前後を確認しながら再生個所の書き起こしや訳を見ることができ、再生されている間のみ表示される字幕で追いつくことができない非ネイティブスピーカー、初学者にも利用できる資料となる。（図 3 では参考のためにアルファベットでの書き起こしも表示している。）通常の字幕では学習者は内容を追うことができず、文字を読むために一時停止すると音声も止まるという弱点を解決し、カラオケのようなハイライトを用いた表示方法で、学習者の意見を取り入れてデザインした。入力ファイルを区別することによって、字幕付き動画を表示する YouTube 上の言語切り替え機能を使って書き起こしと訳を別々に表示することも可能である。

（2）生成方法

用意する素材は、動画（もくしは音声）ファイルと srt 形式の字幕・スクリプト用ファイルである。これら（動画、srt ファイル）を YouTube（https://www.youtube.com/）にアップロードし、YouTube 上での動画 URL を本ジェネレータを組み込んだページで入力するだけで、自動で（1）で示した資料が表示されるようになっている（図 4）。srt ファイル自体の用意については、Elan（https://tla.mpi.nl/tools/tla-tools/elan/）などの無償で直感的に使用できるアノテーションソフトで動画や音声に書き起こしや訳をつければ、それを srt ファイルでエクスポートすることができる。YouTube 上で字幕を編集することも可能で、後者を選んだ場合は特に srt ファイルの存在を意識する必要もない。このようにして、Youtube で字幕付きの動画を作ることさえできれば、本ジェネレータを用いて図 3 のような談

話資料を 1 ステップで簡単に生成できる。

図 4：ジェネレータ使用時の入力欄（miyakogo.ryukyu 上）

（３）作成時の問題点など

2 節でも触れたように、資料のデザインについては、当初予定していたものとは異なるものとなった。当初は、スクリプトは（i）字幕付き動画、（ii-a）原語の書き起こし、（ii-b）文意訳、（ii-c）逐語訳、（ii-d）言語学的分析アノテーション を含むものを考えており、特に図 2 に示したような書き起こしに逐語訳を語のまとまりごとに表示する予定だった。しかし、このような複数層にわたる要素でアラインメントをとって表示し、かつ再生個所に合わせてハイライトするということを、準備中のシステムの中で実装することができなかった。再生個所に合わせたハイライトは必ず入れたい要素だったためこれを優先させ、アラインメントを必要とする(ii-cd)の逐語訳、言語学的分析アノテーションなどを組み込むことは、今回の研究内での実現を見送ることとした。

また、動画再生に必要なストリーミング配信サービスとして YouTube を使用していることにより、YouTube の仕様や規定の変更に伴った不都合が生じる可能性もある。実際に、計画進行中にそれまで自由記述ができた字幕の「言語名」が、選択制になり、自由に言語名を付けることができなくなった。このため、字幕入力・選択時の言語名は、YouTube 内の既存のものから選ばざるを得なくなっている。しかし、このような仕様の変更は、どのようなサービスを使っていても避けきれないものである。また、YouTube を使っている理由として、現段階で最も広く用いられており、使いやすい無料のサービスであることがあげられる。「誰でも使えて見られる」ことや「持続可能性」を考えると、そのような形にせざるを得ない。現時点での最善策として、YouTube を選んでいる。

3-2 ポータルサイトの作成

2 節で述べたように、ジェネレータが当初の予定通りに作りこむことができないことが判明したため、その分の予算などを、全体計画（2 節参照）として次にやることを計画していたポータルサイト作成に充てることとした。作成したウェブサイトは、琉球の諸地域のことばに関連するコンテンツ（地域語で作成されたコンテンツ、あるいは地域語への興味を喚起するようなコンテンツ）を掲載するものとして、サイト名を「しまのことば、しまからのことば」とし、http://shimashima.ryukyu/ で公開している。

ポータルサイトは、図 1 で示したように、代表者らのプロジェクトで作成するすべてのデジタルコンテンツの公開の場となる。これまでも個別のプロジェクトが個別のテーマで作ったサイトは複数あったが、「島（言語）を横断した」「様々なタイプの言語資料に対応した」「ちいさな個別のコンテンツでもすぐに公開しやすい」「個別のコンテンツにアクセスしやすい」ような、本研究で作成する島々のコンテンツを一か所で公開するのに適したものはなかった。また、これらの性質は、既存のサイトから個別コンテンツを引用して表示するポータルサイトとしても有効に機能する。本研究では、これらの性質を満たすウェブサイトとして、作成依頼先の coban.lab 上田寛人氏との相談のうえ、図 5 に示すような構成でのウェブサイトを作成した。

個別コンテンツページは、作者、島、カテゴリ（島の人の書き物、動画、絵本など）についてそれぞれタグ付けされている。個別コンテンツには、トップページから（新着順数件）、もしくは作者ごと、島ごと、カテゴリごとのそれぞれのリストから入れる。個別コンテンツはもちろん、作者も自由に増やすことができる。この枠組みのもと、どこの言語（島）でも、どの資料でも、各作者につき 1 コンテンツの小さいサイズから自然に公開していくことができる。

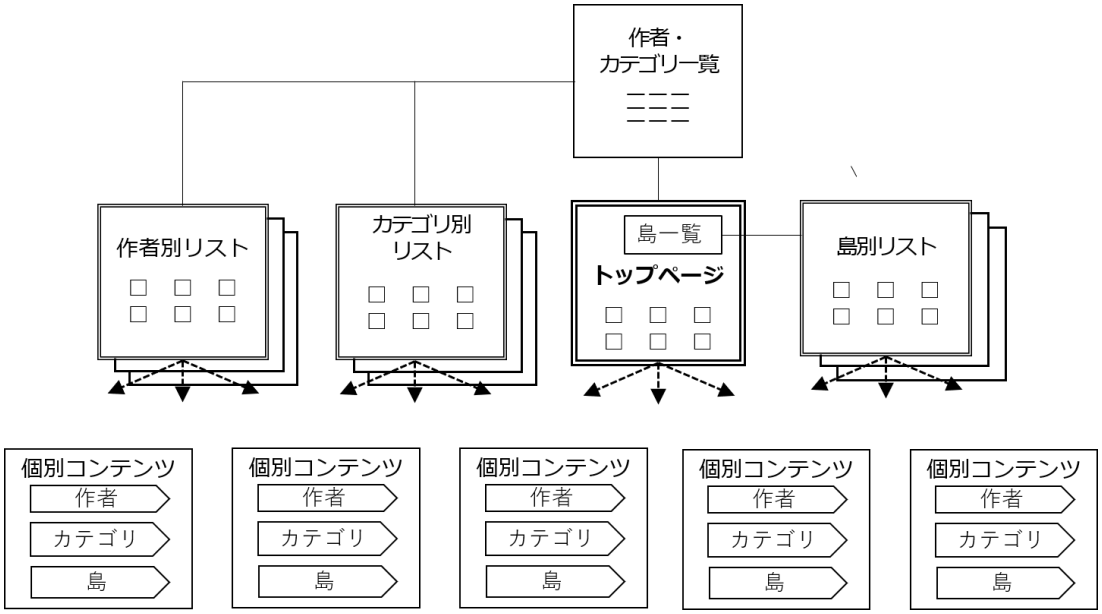


図 5：ポータルサイトの基本構成

作者は、このサイトで新規にコンテンツを公開していく場合と、他サイトからの引用コンテンツを表示する場合とがある。既存のコンテンツも活かしながら、それらがより関心のある人の目に触れやすいようにすることができる。現在のところ、新規作者６、別サイトからの引用４の合計１０作者の作品掲載が確定しており、これ以降も作者を増やしていく予定がある（詳細は 3-3 節（2）参照)。

3-3 ジェネレータやポータルサイトを利用した活動

（１）地域コミュニティとの協働でのコンテンツ作成と公開

本研究期間内においては、主として宮古島の高校生（宮古総合実業高校　担当教員：平良舟江）との共同研究として、資料収集やジェネレータやポータルサイトを利用した一連の記録保存活動を行った。具体的には、i）自然談話の収録 ii）地域語学習のための紙芝居および絵本の作成 iii）高校生自身の作業によるそれらの電子化・公開　を行った。

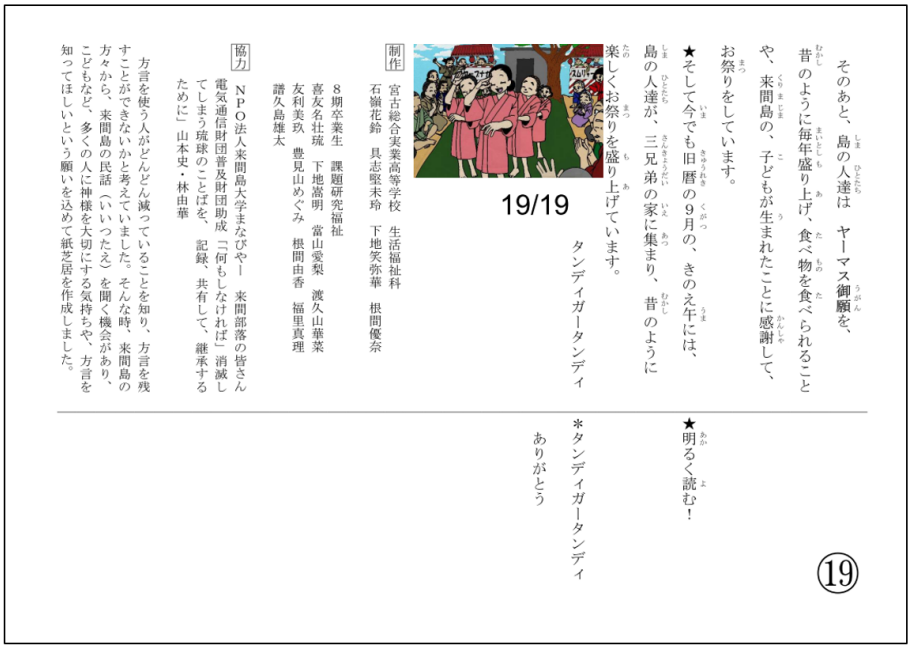
i）については、高校生たちの地域研究に即した地元の高齢者の方言による談話を収録し、書き起こしや訳などのアノテーションを付与した。図３で例示した談話資料コンテンツは、それをもとに作成したものである。ii）では、地域の子供たちが地域語を学習できるようにするめのツールとして絵本や紙芝居を選び、地元の伝承を題材として、共通語の中に効果的に方言をいれたものを作成している(図 6)。高校生自身が地域の高齢者に取材し、伝承話としての正確性を確認しながら小さな子供にも通じるものを作成し、実際に幼稚園などで複数回紙芝居の読み聞かせを行った。これらの絵本作成には、絵や構成、パフォーマンスに関して主としてデザインが専門である山本が、読み原稿については代表者（林）が、作成補助およびアドバイスにあたった。また、紙芝居は 5 部作成し、実際の読み聞かせに使用するほか、作品展への出展や図書館への納品を行った。

iii）については、ウェブサイト作成・システム構築の依頼先である coban.lab の上田寛人氏を迎えたワークショップを行い、高校生らと実際にこれらのコンテンツを電子化・アップロードする作業を行った。これらのコンテンツは現在代表者のウェブページ（「みゃーくの方言 - 宮古語諸方言の記録」http://miyakogo.ryukyu/）で一部が閲覧できるほか、3-2 節で述べたウェブサイトにも掲載準備中である。また、高校生らの地域研究の一環として、これらの活動をまとめたものを國學院大學主催の「第 14 回地域の伝承文化に学ぶコンテスト」に応募し、地域文化研究部門において優秀賞を受賞した（下地笑弥華、石嶺花鈴、具志堅未玲、根間優奈「広めよう！みゃーくふっと先人の郷土文化」　発表要約 http://www.koukouseishinbun.jp/common/dld/pdf/5bbc7cbe9e2675acef82b4b87d5dd758.pdf)。宮古総合実業

高校との共同研究は、今後も継続予定である。

(1) 表紙

(2) 読み原稿2ページ目



(3) 最終ページ

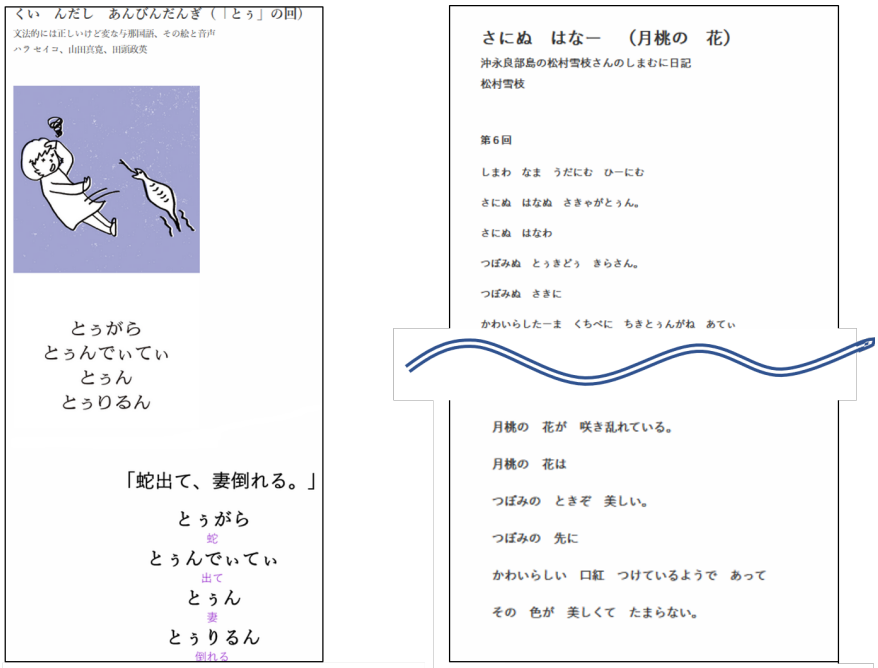
図 6：高校生による文化と地域語学習のための紙芝居

（2）各地の地域語のコンテンツの収集と公開

3-2 節で示したウェブサイトでは、現在１０作者のコンテンツの掲載が決まっている（既存サイトからの引用ベース：4 件、新規作者：6 件）。3-2 節で述べたように、本サイトでは、既存のサイトを活かしたポータルサイトの役割を果たしながら、そのようなサイトを持たない作者でもコンテンツ掲載ができる形式になっており、またコンテンツの種類としても、動画、文章、絵本など、様々なものに対応している（図 7 で例示）。

代表者らの手元の既存の資料に関しても、既に作成した絵本など（與那嶺ほか近刊　など）をデジタイズして掲載したり、これまで別サイトで公開していた字幕・スクリプト付動画を引用掲載したりしている。既に掲載が決まった作者の中には、このために地域語での日記をつけはじめてくれたというケースもある。

各作者の地域分布は、図 8 のようになっている。図に示したように、琉球列島だけでなく、同様に危機言語である八丈語が話される八丈島も含めることとなった。



（1）絵本の例　　(2) 文章の例

図 7:オンラインコンテンツの例

図 8：ポータルサイトのコンテンツ作成作者の地域（2019 年 3 月時点）

このように、多様な言語、様々なコンテンツを見せることは、表現の可能性・選択肢をより多くの話者・コミュニティメンバーに提示することにもつながる。1-2 節でも述べたように、本研究は、少数言語でも大言語と変わりなくオンラインでの発信の場を持てることも目標としている。発表の場と、その実例を提示することは、その活動を促進するのにも重要な役割を果たす。できるだけ多くの人が地域語の発信者、受信者となるよう、今後も、逐次作者およびコンテンツを増やす予定である。

4 まとめ

3 節でまとめた成果によって、1-2 節に示した本研究の目的である「地域コミュニティによる記録保存活動を容易にする基盤をつくり、それを実践する」ということは達せられたと考える。本研究での成果を含む活動は、国際学会 5 件、国内学会 4 件などによって、学会発表および紹介を行った。

3-1 節でも示したように、ジェネレータについては、特にスクリプトの構成について、予定通りの仕様にできなかったが、今後方法を検討し、より学習に使用しやすいコンテンツの作成を目指す。

今後も、本研究で行った活動を地道に続けていくことにより、琉球地域の言語の記録保存、および学習を促進し、ひいては復興・継承の力に繋げていきたい。

謝辞: この活動を可能にしていただいた電気通信普及財団よりの助成、ご助力に、心よりお礼申し上げる。

【参考文献】

松村雪枝・山田真寛・横山(徳永)晶子・元木環・浅川友里江 (2016)『みちゃぬ ふい』てぃんがまシリーズ 1 言語復興の港

小川晋史編 (2015)『琉球のことばの書き方 ―琉球諸語統一的表記法』くろしお出版

下地賀代子・山本史 (近刊)『カンナマルクールクの神』言語復興の港

Tsunoda, Tasaku. (2015). Language Endangerment and Language Revitalization. Berlin: Mouton De Gruyter.

Yamada, Masahiro, Fumi Yamamoto, Akiko Yokoyama-Tokunaga, Kayoko Shimoji, Yurie Asakawa, and Yukie Matsumura. (2017). Picture Project for All.The 5th International conference on Language Documentation and Conversation.University of Hawai'i at Manoa.

山田真寛 (2015)「島民による琉球与那国語の自然談話資料蓄積プロジェクト」日本方言研究会第 100 回大会

與那覇悦子・山本史 (近刊)『ディラブディ』言語復興の港

〈発 表 資 料〉

題　　名	掲載誌・学会名等	発表年月
Creating Contents for Language Learning: Collaboration with Language Communities	NINJAL-NMJH-UHM Workshop "Underdescribed Languages and Histories: Linguist's and Historian's Challenges", University of Hawai'I at Mānoa	2017 年 5 月
地域言語コミュニティと協働する消滅危機言語研究、専門家と協働する言語と文化の継承活動ーじぶんたちでバイリンガル社会をつくる。	平成 29 年度危機的な状況にある言語・方言サミット（北海道大会）	2017 年 12 月
琉球諸語による表現活動：島ことば絵本・プロダクトの制作とコンテンツの活用	第 10 回琉球継承言語研究会, 琉球大学	2018 年 3 月
Port Language Revitalization Project	Approaches to Endangered Languages in Japan and Northeast Asia: Description, Documentation, and Revitalization	2018 年 8 月
Documenting the rich dialectal variation of Miyakoan	Approaches to Endangered Languages in Japan and Northeast Asia: Description,	2018 年 8 月

	Documentation, and Revitalization	
消滅危機言語再活性化のための研究プラットフォーム：言語復興の港	日本音声学会，沖縄国際大学	2018 年 9 月
危機言語・方言の保全と復興のために：データ整備と公開の価値	NINJAL シンポジウム「データに基づく日本語研究	2018 年 12 月
Endangered Languages in Japan: Documentation and Revitalization	Digital Revolution for Under-resourced Languages in Asia，奈良先端科学技術大学院大学	2019 年 2 月
Preparing Infrastructure for Writing and Typing with non-Latin Symbols: A Case of Ryukyuan Languages	The 6th International Conference on Language Documentation and Conservation，University of Hawai‘i at Mānoa	2019 年 3 月