

オンライン沖縄語辞典

中村仁宣 (Hisanobu Nakamura)

March 23, 2024

自己紹介

初めていうがなびら！

はいさい、ぐすーよー、初めていうがなびら！わんねー...

- 名前のーじや中村なかむらひさのぶ仁宣いんでい言いちょーいびーん。
- 出身すだちや愛知県あいちけん名古屋なごや市しやいびーん。イギリスいぎりすんかい
ん暮くらちょーたくとうる事ことんあいびーん。
- 2021 年ねんに那覇なはうてい暮くらし始はじみびたん。
- 大学・大学院うとーてい専攻せんこうさる分野ぶんぎや数学・物理すうがく・ぶつりや
いびーたん。
- 前職ぜんしつや機械学習きかいがくしゆエンジニアえんじニアやいびーたん。
- 言語くとうば習ならいしえー好しちやいびーん。

はじめに

今日、話す内容は、元々沖縄に縁がなかった登壇者が、沖縄語を学習するにあたり体験した困難と、持ち合わせていたプログラミングの知識を使ってそのほんの一部を自分なりに解決したストーリーについてです。よって、この話は

1. ^{いち}一沖縄語学習者としての視点
2. ツールの作成者としての視点

の2つの視点から書かれています。このお話が、マイノリティ言語の新規学習者に対してどのように情報発信、または学習環境の整備等をすれば良いかのヒントになれば、と思っています。

今日話す内容

1. オンライン沖縄語辞典開発の経緯
2. オンライン沖縄語辞典の紹介
3. 得られた知見
4. 今後の展望

オンライン沖縄語辞典開発の経緯

オンライン沖縄語辞典開発の経緯

開発者が沖縄語学習を開始

- 2021 年 6 月に那覇に移住。
- 2022 年 6 月に沖縄語学習を始める。
- 教材・辞書を探し始めるが、選択肢は少ない。
- Web 上で使える辞典は、カジュアル・断片的なものが多い。

オンライン沖縄語辞典開発の経緯

『沖縄語辞典』と『うちなーぐち活用辞典』

- 沖縄語辞典、国立国語研究出版、1963 年発行、約 12000 語、¥15,000～
- うちなーぐち活用辞典、編・著者宮良信詳、国立国語研究所言語変異研究領域、2021 年発行

オンライン沖縄語辞典開発の経緯

価値

- 『沖縄語辞典』
 - 豊富な語彙数、例文、用言活用の解説等を含む。
 - 士族・平民発音の区別。古い言葉・意味の記載。
- 『うちなーぐち活用辞典』
 - 現代の沖縄語の豊富な例文を含む。

オンライン沖縄語辞典開発の経緯

NINJAL データにたどり着く

- 紙の『沖縄語辞典』『うちなーぐち活用辞典』を入手するが、重さのため持ち運べず。
- このような語彙数・例文などが豊富なオンライン辞典があればいいのになと思う。
- NINJAL で『沖縄語辞典』の **Excel データがダウンロード可能**だと知る。
- ないのなら、『沖縄語辞典』のデータをオンライン上で簡単に検索できるようにすれば良い！
- どうせ作るのなら、他の人にも使って貰いたい。

オンライン沖縄語辞典開発の経緯

開発の開始

- しかし、Excel のデータをそのまま使ったのでは、一般ユーザーにとっては使いにくい所もある...

沖縄語辞典 ホーム	
<hr/>	
アサ	
発音	/asa/, アクセント◎
品詞	【名】
意味	朝。早朝にはあまり使わない。asa(sakene)〈朝晩〉, asajasa (朝夕) などの複合語に使われる。普通はasumamiti。また,asumamitiの方がasaよりも早い時間をさす感じがある。
<hr/>	
アサ	
発音	/asa/, アクセント◎
品詞	【名】
意味	阿波。+他・9月。
<hr/>	

オンライン沖縄語辞典開発の経緯

『沖縄語辞典』を使う上でのハードル

音素記号を知らない利用者、初級学習者、などの一般ユーザーにとって、、、

1. 音素記号を知らないと**使う事ができない**（検索も例文を読むこともできない）
 - ・ 沖縄語は音素記号のみの表記。
2. 「解説篇」を読まないと**動詞活用がわからない**
 - ・ 動詞の活用は自分で考える必要がある。
3. **初見では解読困難な知識が必要。**
4. 語彙の説明文の空白が圧縮されていて**読みづらい**

オンライン沖縄語辞典開発の経緯

電子情報の利点を活かす

『沖縄語辞典』は、元々紙の辞書として作成され、紙面の節約の必要性があるため、重複を省き、変化部分のみを記す経済的な表記法にならざるをえないが、電子空間上ではその制限は緩くなり、わかり易さのための冗長性を(許容範囲内で)持たせられる。テキスト処理と Web 技術でそれが可能。

オンライン沖縄語辞典開発の経緯

『沖縄語辞典』をより多くの人へ

そこで、以下の点を情報技術により克服することで、より多くの人にとって使いやすいオンライン沖縄語辞典を作る事を目標とした。

1. 音素記号をカナと IPA で表示する。
2. 語彙説明の中の沖縄語もカナ表示する。
3. 動詞の活用表を自動で生成する

オンライン沖縄語辞典

オンライン沖縄語辞典

概要

- 収録辞書：『沖縄語辞典』（沖日、日沖）、『うちなーぐち活用辞典』
- URL：https://okinawago.app/
- 使用料：なし

沖縄語辞典 ホーム 見出しとかな表記 この辞典について

沖縄辞典・日沖辞典・うちなーぐち活用辞典

このページでは、国立国語研究所の『沖縄語辞典』と『うちなーぐち活用辞典』内の単語を、かなの見出し語で検索できます。
かな表記については、こちらのページを参照して下さい。かな表記について

検索語の入力（かな表記）

検索する辞書の選択

☒ 沖・日

☒ うちなーぐち活用辞典

☒ 日・沖

検索語と見出し語の一致条件の選択

☒ 検索語を含む

☐ 前方一致

☐ 後方一致

検索の仕方と結果

動詞「アーキユン」での文章表示比較

ʔaaki=juN① (自 =raN, =ti) ㊦ 割れる。
 裂けて離れる。また、ひびが入る。裂け目
 ができる。ʔicanu kariti ~. 板が枯れて
 割れる。 muraziNmi sjaʃiga, ʔikuçi-
 nuNkai ʔaakitan. 村の協議をしたが、
 意見がいくつにも割れた。㊦言行が食い違
 う。つじつまが合わない。矛盾する。
 taiga ʔjuru kutubanu ʔaakitooN. ふた
 りの言うことばは矛盾している。ʔariga
 ʔjuru kutoo ʔatutu sacitu caa ~. 彼
 の言うことは前とあとといつも食い違
 う。

アーキユン

発音

音素	/ʔaaki=juN, ʔaaki=juN /
カナ	アーキユン
IPA	[ʔa:kiɸun]
アクセント	①

品詞

〔自〕

意味

1. 割れる。裂けて離れる。また、ひびが入る。裂け目ができる。

イチヤヌ カリティアーキユン。
 (ʔiɸanu karitiʔa:kiɸun.)

板が枯れて割れる。

ムラジンミ サシガ、イクチヌンカイ アーキタン。(主: ムラジンミ サシガ、イクチヌンカイ アーキタン。)
 (muradSimmi saʃiga, ʔikutiɸunɸai ʔa:kitas. (主: muradSimmi saʃiga, ʔikutsinunɸai ʔa:kitas.))

村の協議をしたが、意見がいくつにも割れた。

2. 言行が食い違う。つじつまが合わない。矛盾する。

タイガアエル クトゥバヌ アーキトーン。
 (taiga ʔjuru kutubanu ʔa:kitooN.)

ふたりの言うことばは矛盾している。

アリガアエル クトアットク サチトゥ チャー アーキヌン。
 (ʔariga ʔjuru kutoo ʔatutu ʃa: ʔa:kiɸun.)

彼の言うことは前とあとといつも食い違う。

活用

語幹	活用	音素	カナ	IPA
基本語幹 /ʔaaki-/	否定形	ʔaakiraN	アーキラン	[ʔa:kiɸan]
連用語幹 /ʔaaki-/	連用形	ʔaakii	アーキー	[ʔa:ki:]
	過去形	ʔaakitaN	アーキタン	[ʔa:kitas]
音便語幹 /ʔaakit-/	てい形	ʔaakiti	アーキティ	[ʔa:kiti]
	継続形	ʔaakitooN	アーキトーン	[ʔa:kitooN]

カナ表記について

開発者が今まで見かけた中で、取り入れやすい表記法をできるだけ多くのバリエーションで検索できるように努力した。「どれか一つでしか検索できない」ではなく、「どの表記法でも検索できる」を目指したかった。特に、グロツタルストップの有無をどう表記するかについては、現在様々な表記法が使われているので、よく使われているであろう書籍のものを取り入れた。

カナ表記で参考にした書籍、文献

- ・ 白水社『沖縄語の入門』
- ・ 研究社『初級沖縄語』
- ・ 研究社『沖縄語辞典』
- ・ 沖縄県の『沖縄県における「しまくとうば」の表記について』
- ・ しまくとうば普及協議会の教材

カナ表示の例

スン,シュン

発音

音素	/sjuN /
カナ	スン ,(土)シュン
IPA	〔sun〕 ,(土) 〔ʃun〕
アクセント	①

品詞

【他・不規則】

意味

1.する。

'ウン,'ウン,ウウン,ウウン,ヲウン

発音

音素	/'uN /
カナ	'ウン,'ウン,ウウン,ウウン,ヲウン
IPA	〔uN〕
アクセント	①

品詞

【自・不規則】

意味

1.いる。おる。（人間・動物などが）存在する。ンカシ スイナカイ

語彙説明文と沖縄語例文の分離

sjuN④ (他・不規則) ⑤する。'jamatuguci
～. 日本語を話す。bincoja siibusikoo
neeNkutu saN. 勉強はしたくないからし
ない。siijuusjuN. できる。なしうる。⑤
(強意の補助動詞として) 'jumidu sjuru.
読むのだ。?aee sani. ありはしないか。

意味

1. する。

ヤマトグチスン. (土: ヤマトグチシュン.)
['jamatugutʃisun . (土: 'jamatugutʃifun .)]

日本語を話す。

ビンチョーヤ シーブシコーネンクトゥ サン.
[bintʃo:ja ʃi:buʃiko:ne:nkutu san.]

勉強はしたくないからしない。

シーユースン. (土: シーユーシュン.)
[ʃi:ju:sun. (土: ʃi:ju:fun.)]

できる。なしうる。

2. (強意の補助動詞として)

ユミドゥ スル. (土: ユミドゥ シュル.)
['jumidu suru. (土: 'jumidu furu.)]

読むのだ。

アエー サニ.
[?ae: sani.]

ありはしないか。

語彙説明文と沖縄語例文の分離

- 「ほとんど」の例文を分離できているが、一部できていない箇所もあるかもしれない。
- 正規表現による処理を行っているため、正確な処理は対象データに対する網羅的な知識が必要。

オンライン沖縄語辞典

動詞活用表の例

ʔaaki=juN① (自 =ran, =ti)

活用

語幹	活用	音素	カナ	IPA
基本語幹 /ʔaakir-/	否定形	ʔaakiraN	アーキラン	[ʔa:kiɾaɴ]
連用語幹 /ʔaakij-/	連用形	ʔaakii	アーキー	[ʔa:ki:]
音便語幹 /ʔaakit-/	過去形	ʔaakitaN	アーキタン	[ʔa:kiɾaɴ]
	てい形	ʔaakiti	アーキティ	[ʔa:kiɾi]
	継続形	ʔaakitooN	アーキトーン	[ʔa:kiɾo:ɴ]

sjun① (他・不規則)

活用

語幹	活用	音素	カナ	IPA
基本語幹 /s Qs-/	否定形	saN	サン	[saɴ]
連用語幹 /s sj-/	連用形	sii	シー	[ʃi:]
音便語幹 /sj-/	過去形	sjaN	サン(土:シャン)	[saɴ] (土: [ʃaɴ])
	てい形	Qsi	ッシ	[ʃʃi]
	継続形	sjoon	ソーン(土:ショーン)	[so:ɴ] (土: [ʃo:ɴ])

動詞活用表の生成について

- 元の『沖縄語辞典』では品詞の部分に否定形とティ形語幹末尾のみの記載だった。
- 不規則動詞の活用辞書を手で作成。規則動詞は自動生成。

うちなーぐち活用辞典に関する課題

PDF の構造化されていないデータを文字のフォント・色・位置・大きさなどの情報から手動で構造化を行った。そのため、まだ正しく構造化できていない箇所がある。

まとめ1

ユーザビリティ向上のための施策

1. シンプルで、一貫性のあるページデザイン
2. カナ表記の解説ページの作成
3. カナで検索可能に。
 - ・ 可能な限り多様な表記法に対応
4. 土族発音も併記
5. 語彙説明文内の沖縄語の例文の分離。
 - ・ 語彙説明文内の沖縄語の例文もカナで表示可能に。
 - ・ 改行を挿入する事で、視認性も向上。
6. 動詞活用表の生成。

到達しやすさへの施策

1. シンプルな URL

- URL を覚えてもらえる。チラシに乗せる時にコンパクト。

2. twitter アカウントとチラシ作り

- SNS での発信と対面での配布、店舗に配置

3. QR-code 掲載

- スマートフォンなどで、隣の人にすぐに URL をシェアなどできる

得られた知見

アクセシビリティ (accessibility) の重要性

アクセシビリティ := アクセスのしやすさとは...

- 見つけやすさ (searchability)
 - 見つけやすい場所にある。
- 入手しやすさ (reachability)
 - 普通の人がある手段・コストで手に入れられる事。
- 使いやすさ (usability)
 - 使いこなすのに特別な訓練を必要としない事。
- **利用する側からの視点が重要。**

XX にとって利用しやすいデータとは

- 一般ユーザーにとって
 - インターネットなら、検索にヒットしやすい、雑多な劣化した情報に埋もれていない。
 - 紙の書籍なら、街の本屋で購入可能である、図書館にある、高くない等
 - 利用料が相応、または無料。
 - スマホ・PC 上で、カナで検索できる等
- 開発者にとって
 - データがダウンロード可能で、ある程度構造化 (Excel, CSV, JSON, etc) されている。
 - ソースコードがレポジトリで公開されている。

今後の展望

アプリの改良・充実

- 辞書機能の充実 (活用表の充実、お気に入り語彙、履歴、モバイルアプリ化等)
- 学習サポート機能 (動詞活用練習、語彙暗記テスト等)
- 他辞書の取り込み (さらなる沖縄語、琉球諸語の辞書の追加)
- Wiki 作成
- 共同開発者・資金調達

う ち くい
終わいまでい聞ち呉みそーち
にふえー
いっぺー御拝でーびる。