

日琉諸語を対象とした 方言談話データベース試作版の概要

高城隆一・下地理則
(九州大学)

謝辞

本研究は文部科学省研究拠点形成費等補助金（デジタルと掛けるダブルメジャー大学院教育構築事業）の助成を受けています。



九州大学大学院人文情報連係学府のHP

<https://dh.kyushu-u.ac.jp/>



本発表の内容

1. 日琉諸語を対象とした方言談話データベースの試作版 (<https://danwa.odjl.net/>) について開発の趣旨を説明する
2. 実際のページを使用してデモンストレーションを行う
3. 目下の課題について述べる

※本発表で提示する内容は、後述するコンセプトに基づいた開発途中のものであり、決して確定したものではありません。コメントやアドバイスを来年度以降の開発に活かすことで、最終版の形が大きく変わる可能性があります。

趣旨～ドキュメンテーションの基本に立ち返る～

大規模コーパス:

大量の談話データが用いられており、直接的な研究利用を目的とした設計になっている。研究者向け。

例: 日本語諸方言コーパス (COJADS)

本データベース:

研究利用というよりも成果公開。博物館的。研究者、コミュニティ、一般向け。

趣旨～ドキュメンテーションの基本に立ち返る～

➤博物館的な設計

✓ガチ研究への利用をあえてゴールにしない

- それはCOJADSなどのシマ

✓「この言語はこういうものです」というサンプルの提示

- 当該言語の辞書や文法（概説）とのリンク
- 話者とコラボした談話資料を、1編から気軽にアップできる
- 後続のガチ研究を触発するきっかけにはなるかもしれない

趣旨～ドキュメンテーションの基本に立ち返る～

- 研究者の成果公開のプラットフォーム
 - ✓記述の副産物としての談話データを、そのままにしている人は多い
 - ✓公開までのハードルを低くしたい
 - 短い挨拶、歌、ことわざ（とその解説）、etc.でも五月雨式に公開して良い
 - ✓自分の対象言語の談話資料が一定の場所に上がっていると、いろいろな機会に便利
 - 授業で使う、紹介する、書籍で引用する、etc.

機能

基本機能：談話テキスト+音声データ

- 1つの談話当たり数分～数十分のものを想定

機能

基本機能：談話テキスト+音声データ

談話テキストの詳細（試作版段階）

- テキスト
- 形態素
- グロス
- 標準語訳
- 注

※追加予定

- かな表記
- 英語訳

機能

※追加予定項目

- プロンプト素材の共有（画像、映像、話題、など）
- 地域の解説
- 言語の解説（簡易文法記述）
- 談話の解説（民話の内容、祭りの背景知識等）
- 語彙の解説（ODJLとの連携）
- その他（ジャンル検索など）

談話データ提供のタイミング

- 音声ファイル (wavファイル)
- 書き起こしに使ったELANのファイル
- リンク (個人のサイト等で公開済みの場合)

※公開のための同意書・公開プロセスについては、九大・カルリノ共同研究のODJLの経験も活かし、現在調整中。

※談話データの提供は、最終版の完成後、NINJAL共同研究会などで告知してから開始したい。

【問い合わせ先】takiryu1@kyudai.jp (高城隆一)

目下の（開発上の）課題

✓「日琉諸語オンライン辞書（ODJL）」（カルリノ・サルバトーレ氏の開発したもの）と、連携できるところはしたい

- 例えば、談話中の語彙と辞書をリンクさせる

✓文法（概説）をどこに置き、
どうリンクすれば見やすく、
使いやすいか工夫する必要あり



参考

- 「日琉諸語オンライン辞書 (ODJL)」 <http://odjl.net> (2024年3月更新版)
- 「日本語諸方言コーパス (COJADS)」
<https://www2.ninjal.ac.jp/cojads/index.html>