

国立国語研究所学術情報リポジトリ

方言コーパスを使ってみよう

メタデータ	言語: jpn 出版者: 公開日: 2022-08-19 キーワード (Ja): キーワード (En): 作成者: 木部, 暢子 メールアドレス: 所属:
URL	https://doi.org/10.15084/00003623

方言コーパスを使ってみよう

言語変異研究領域 木部暢子 nkibe@ninjal.ac.jp

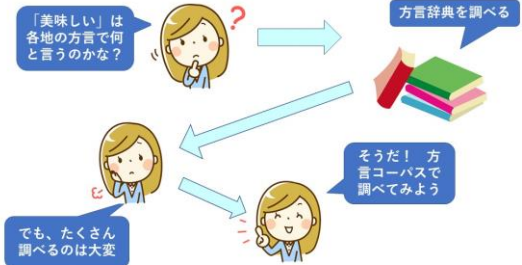
コーパスとは

- 自然な発話や文章をたくさん集め、それらを整理して、それぞれの単語にことばの情報(品詞・活用形など)をつけて、調べやすくしたものです。
- 「日本語諸方言コーパス」は日本各地の方言を集めたコーパスです。
- 略称 COJADS (Corpus of Japanese Dialects)



1

方言コーパスを使うと何が出来る？



2

でも、何でも調べられるわけではありません。

- COJADSは、文化庁が1977～1985年に行った「各地方言収集緊急調査」のデータをもとにしています。
- もとのデータは、全国 200 地点あまり、約4,000時間の方言談話の録音テープからなっています。
- そのうち、現在のCOJADSには、47都道府県62地点、約60時間のデータが搭載されています。(というより、しが搭載されていません。)
- 全国方言のデータ量としては、非常に少ないものです。
- したがって、COJADSに載っていない言葉がたくさんありますが、現在のところ、COJADSがいちばん大きな方言コーパスです。



3

- 共通語「苦しい」から方言形を検索した結果
- 各地の方言が横並びで検索できる。

データ名	サンプルID	原文	検索結果	場所	話者	性別	年齢
山形県 山形市	001	苦しい	苦しい	山形県 山形市	男性	70歳	無
山形県 山形市	002	苦しい	苦しい	山形県 山形市	女性	70歳	無
山形県 山形市	003	苦しい	苦しい	山形県 山形市	男性	70歳	無
山形県 山形市	004	苦しい	苦しい	山形県 山形市	女性	70歳	無
山形県 山形市	005	苦しい	苦しい	山形県 山形市	男性	70歳	無
山形県 山形市	006	苦しい	苦しい	山形県 山形市	女性	70歳	無
山形県 山形市	007	苦しい	苦しい	山形県 山形市	男性	70歳	無
山形県 山形市	008	苦しい	苦しい	山形県 山形市	女性	70歳	無
山形県 山形市	009	苦しい	苦しい	山形県 山形市	男性	70歳	無
山形県 山形市	010	苦しい	苦しい	山形県 山形市	女性	70歳	無

4

- 方言形「シンドイ」を検索した結果
- 各地の方言形の共通語訳が検索できる。

データ名	サンプルID	原文	検索結果	場所	話者	性別	年齢
山形県 山形市	001	シンドイ	シンドイ	山形県 山形市	男性	70歳	無
山形県 山形市	002	シンドイ	シンドイ	山形県 山形市	女性	70歳	無
山形県 山形市	003	シンドイ	シンドイ	山形県 山形市	男性	70歳	無
山形県 山形市	004	シンドイ	シンドイ	山形県 山形市	女性	70歳	無
山形県 山形市	005	シンドイ	シンドイ	山形県 山形市	男性	70歳	無
山形県 山形市	006	シンドイ	シンドイ	山形県 山形市	女性	70歳	無
山形県 山形市	007	シンドイ	シンドイ	山形県 山形市	男性	70歳	無
山形県 山形市	008	シンドイ	シンドイ	山形県 山形市	女性	70歳	無
山形県 山形市	009	シンドイ	シンドイ	山形県 山形市	男性	70歳	無
山形県 山形市	010	シンドイ	シンドイ	山形県 山形市	女性	70歳	無

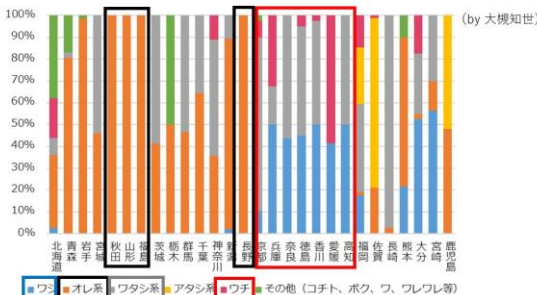
5

- タグ「一人称単数」(「私」)を検索した結果
- 共通の条件で検索しているため、地域間の比較やいろいろな比較ができる。

データ名	サンプルID	原文	検索結果	場所	話者	性別	年齢
山形県 山形市	001	私	私	山形県 山形市	男性	70歳	無
山形県 山形市	002	私	私	山形県 山形市	女性	70歳	無
山形県 山形市	003	私	私	山形県 山形市	男性	70歳	無
山形県 山形市	004	私	私	山形県 山形市	女性	70歳	無
山形県 山形市	005	私	私	山形県 山形市	男性	70歳	無
山形県 山形市	006	私	私	山形県 山形市	女性	70歳	無
山形県 山形市	007	私	私	山形県 山形市	男性	70歳	無
山形県 山形市	008	私	私	山形県 山形市	女性	70歳	無
山形県 山形市	009	私	私	山形県 山形市	男性	70歳	無
山形県 山形市	010	私	私	山形県 山形市	女性	70歳	無

6

「一人称代名詞」1語形使用タイプ(男女差なし) / 2語形以上使用(男女差あり)



7

まとめ

- 現在のCOJADSには、約60時間分の各地のデータが搭載されています。
- データ量としては、非常に少ないものですが、現在のところ、日本でいちばん大きな方言コーパスです。
- 検索方法には、共通語から方言形を検索する方法、方言を直接検索する方法、タグを使う方法があります。
- COJADSを使って各地の方言を調べてみると、いろいろな方言の違いを知ることができます。
- さあ、あなたも調べてみてください



8