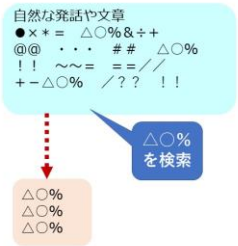
方言コーパスを使ってみよう

言語変異研究領域　木部暢子　nkibe@ninjal.ac.jp

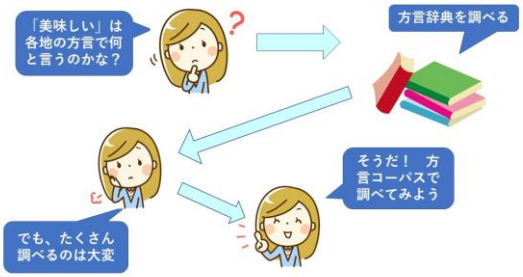
コーパスとは

- 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できる？



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、しか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。

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。

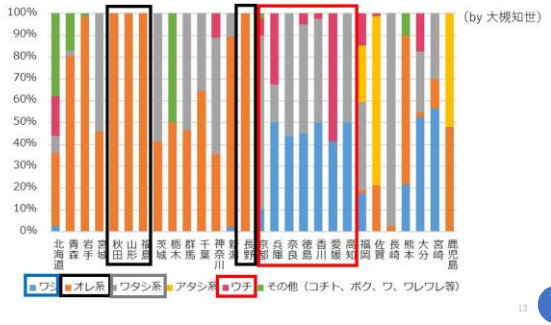
5

- 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ので、地域間の比較やいろいろな比較ができる。

（男女差なし）

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