NINJAL OPEN HOUSE 2021

語誌情報ポータルの活用

山崎　誠（国立国語研究所研究系言語変化研究領域）

1.コーパス化になじまない資料

- コーパスの利用が真価を発揮するのは，文脈をもつ言語資料である。
- 文脈を持たない言語資料は，コーパス化にはなじまない。
- 例えば，古辞書の多くは見出しのみで構成されるため，文脈らしい文脈がなく，コーパス化にはなじまない。
- しかし，複数の古辞書を串刺しで検索できれば，語の歴史的な変遷の情報を得ることができる。
- このような情報は，言わば当該の語にとってメタ情報であり，コーパスから得られた用法を補完するものとなる。

参考文献：山崎誠(2020)語誌情報ポータルについて，「日本語学」39(2)，124-129

2.どのような言語資料があるか

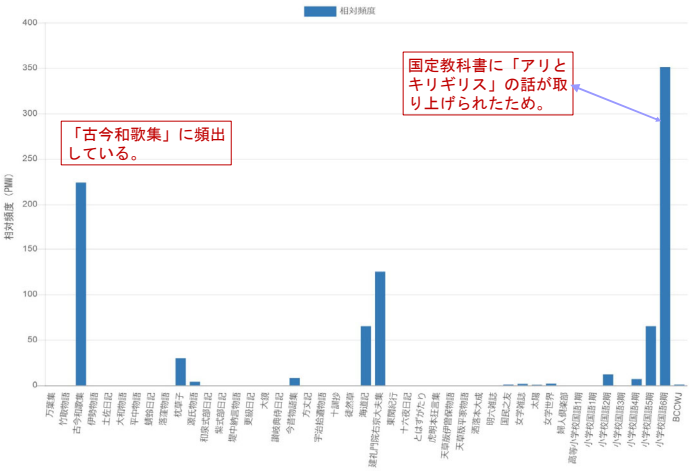
- 古辞書：語の時系列的な情報が得られる。デジタル化され，インターネット上で画像で確認できるものもある。
- 言語地図：言葉の地理的分布が視覚的に得られる。「国立国語研究所言語地図データベース」（作成：大西拓一郎ほか）の利用。
- 言語記事：言葉について書かれた新聞記事，雑誌の記事の書誌情報。『読売新聞』（1913年，1917年）『文藝春秋』（1923～45年）が対象（作成：新野直哉）。
- コーパスからの統計情報：『日本語歴史コーパス』および『現代日本語書き言葉均衡コーパス』における相対頻度。

3. 検索例

- 語彙素か語彙素読みかを自動判別：かな（ひらがな，カタカナ）は，語彙素読み，漢字は語彙素とみなして検索する。
- 以下は，「きりぎりす」で検索した例。
- 検索結果はダウンロードできる。

URL https://goshidb.ninjal.ac.jp/goshidb/

作品別相対頻度

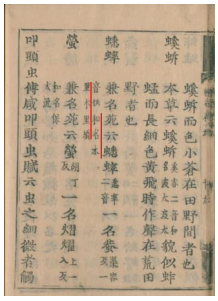


時代別相対頻度

古辞書情報

古辞書情報

辞書名	辞書見出し	所在
和名類聚抄	蟋蟀	巻19・虫豸部第31・虫豸類第240・19丁裏5行目



言語記事情報

要旨が分かる（文藝春秋のみ）

言語記事情報

資料名	発行年月（日）	掲載位置	欄名・連載名等	執筆者	題名	内容	コード	キーワード	備考
文藝春秋	1942年9月	146-152		大町文衛	蟋蟀を尋ねて	「古文書と蟋蟀」古来、キリギリス、コオロギ、イトドの三つの名が様々な虫について使われてきたこと、マツムシとスズムシの指すものが二度入れ替わったことを挙げ、昔の人の実証精神の欠如が原因とする。（150-151）	動植物名 語源・語誌	キリギリス	

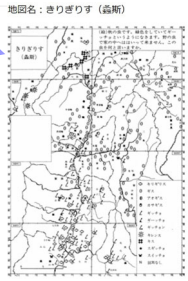
言語地図情報

言語地図情報

地図ID	地図名	よみがな	質問文	分野	品詞	分類	地図集ID	地図集書名
9848	きりぎりす	キリギリス	夏から秋にかけて，野原の草むらなどにいる虫で，長い後ろ足でよくとびはねます。触角が体の長さより長い。 共通語形：キリギリス 予想語形：ギッチョン	語彙	名詞	虫	187	『神奈川県言語地図』
9854	きりぎりす（蟲斯）	キリギリス（キリギリス）	（絵）秋の虫です。緑色をしていてギーッチョというようになきます。野の虫で家の中へははいって来ません。この虫を何と言いますか。	語彙	名詞	虫	201	『上伊那の方言』

地図名：きりぎりす

地図名：きりぎりす（蟲斯）



4.今後の課題

- 検索の統一キーとしている，UniDicにない語の扱い（現在はUniDic登録の語のみが検索対象となっている）。
- 検索に適さない語の見出しの付け方（言語地図の地図名「頭の後下部」に「盆の窪」を検索語として付与するなど）。

本発表は，国立国語研究所の機関拠点型基幹研究プロジェクト「通時コーパスの構築と日本語史研究の新展開」及びJSPS科研費JP17K18505の成果です。