

コーパスを使った日本語の歴史の研究

H1

言語変化研究領域 小木曾智信

ことばの研究方法

1. 自分の頭の中で考える(内省する)
2. 人に尋ねる(聞き取り, アンケート)
3. 実際の用例を集めて考える
→古い時代の言葉の研究には3.しかない



用例を集めるには

- ひたすら書物を読む
- 総索引を使う
- 電子テキストを検索
- コーパスを使う



コーパス

大量の言葉の使用例を集めて言語研究に必要な情報を付けたデータベース

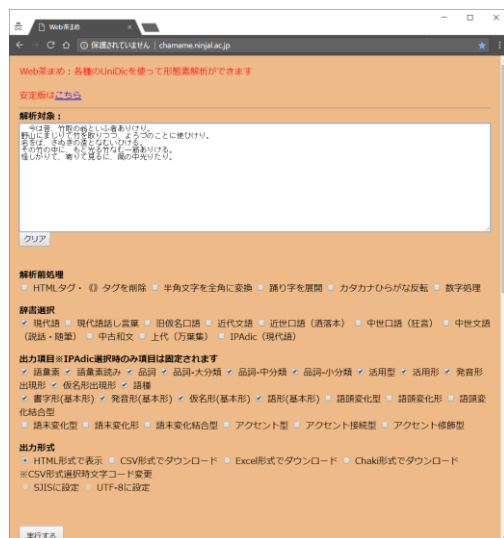
必要な情報: 単語の情報, 出典情報(いつ、どこで、誰が)、著者(話者)情報...

世界中の研究者が研究に使えるように古文のコーパスを作ってネット上で公開する



古文の自動解析

- 古文のコーパスに単語の情報を付与するため、形態素解析という技術によってコンピューターに自動で品詞分解を行わせる
- そのための電子辞書(古文用のUniDic)を自分たちで作成
- 実際に解析が行えるWebページ: Web茶まめ <http://chamame.ninjal.ac.jp/>



- 万葉集から近代文学まで対応

「日本語歴史コーパス」(CHJ)

2016-2022年で構築してきたものから公開コーパスに入れる重要な日本語研究資料

奈良時代	☒ 万葉集 ☐ 宣命	
平安時代	☒ 仮名文学	☐ 和歌
鎌倉時代	☒ 説話・随筆 ☒ 日記・紀行 ☐ 軍記	
室町時代	☒ 狂言 ☒ キリシタン資料	
江戸時代	☒ 洒落本 ☐ 人情本 ☐ 近松	
明治・大正	☒ 雑誌 ☒ 教科書 ☐ 文学作品 ☐ 新聞	

☒ はすでに公開したもの

「中納言」による「日本語歴史コーパス」の公開

- <https://chunagon.ninjal.ac.jp>
- 要登録、無料



「日本語歴史コーパス」を使った研究

- 各時代の表記・語彙・文法など、毎年40本以上の研究発表・論文
2016年:46本、2015年:49本、2014年:48本、2013年:41本、2012年:42本
- 関連書籍
 - 『コーパスと日本語史研究』2015.10(ひつじ書房)
 - 「日本語史研究と歴史コーパス」『日本語学』2014.11臨時増刊号(明治書院)

