

## 「女(ひと)」のような二重表記が検索できる歌詞コーパスの設計と構築

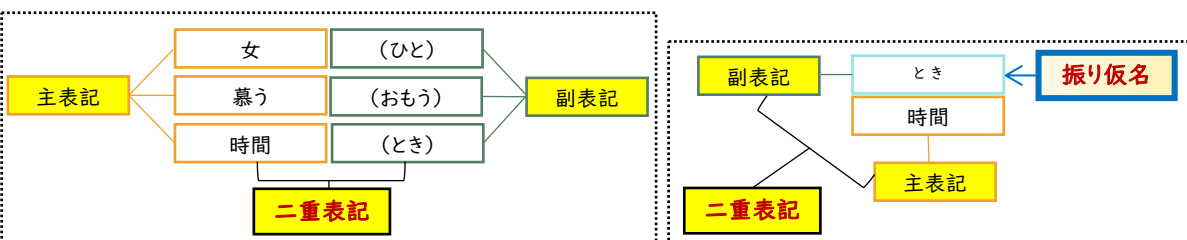
一橋大学大学院言語社会研究科第2部門博士課程

胡 佳芮(こ かせい)

## 「二重表記」の定義

日本流行歌の歌詞では:

- (1) 誰かいい女(ひと) 出来たのね(「うそ」、歌:中条きよし、詞:山口洋子、1974)
- (2) 慕(おも)っていたけど(「心凍らせて」、歌:高山巖、詞:荒木とよひさ、1994)
- (3) 時間(とき)が過ぎても(「風を待つ」、歌:STU48、詞:秋元康、2019)



泉(1993)によると、「二重表記」とは:

- 「時間(とき)」……発音形を( )の中に示すもの <主表記(副表記)>
- 「時間」……発音形を振り仮名で示すもの <副表記>

本研究では:二重表記→【主表記と副表記とが一致していない二重表記】

- a. 副表記に示す発音形を『大辞林』で検索し、該当する見出しの表記欄には主表記にあたるものが存在しない。
- b. 主表記に示す書字形を『大辞林』で検索し、該当する見出しの表記欄に主表記が表示されていても、その見出しと副表記とが一致していない。

## 先行研究における二重表記

二重表記について、

- 佐竹(1980): 振り仮名は日本語表記の特徴的な現象とし、表記の表現性を豊かなものとさせている存在
- 泉(1993): 二重表記は漢字と仮名の関係、表記と読みの関係、日本人の習慣・情緒性と言語意識の関係からみれば、二重表記が必要悪とされる、日本語の様々な特徴と問題点を集中的に体現している存在
- 今野(2013): 歌詞における振り仮名は、現代の振り仮名の代表的なものの一つ

## 目的・意義

1960年代から2010年代までの日本流行歌の歌詞における二重表記が検索できる歌詞コーパスの作成を目指す。

- 日本の流行歌が1960年代から現在にいたるまでの約50年間において、歌詞に出現している二重表記の使用実態が確認できる
- 小説や短歌、漫画等と比較でき、歌詞における二重表記の特徴がわかる
- 二重表記の使用から現代日本語の使用上の変化が見える

## 設計と構築手順

収録楽曲について、

- 参考資料: 1968年から開始した音楽ヒットチャート「オリコン年間売上ランキング」
- 選定基準: 「オリコン年間ランキング」にランクインした楽曲の中から、歌詞が日本語で書かれた楽曲(両A面CDシングルや複数のメイン曲入りの場合はすべてのメイン曲を対象とする)
- 歌詞出典: 歌詞検索サイト「歌ネット」
- 収録楽曲の発売年度: 1969~2019年の約50年間で5年ごとに区切り、1969、1974、1979、1984、1989、1994、1999、2004、2009、2014、2019の11の年

歌詞全文コーパス 構築手順:

- ① 歌詞検索ソフト「Lyrics Master」を利用し、「歌ネット」を情報源とし、収録対象である楽曲の歌詞テキストファイルをダウンロードする
- ② CDケースに置いてある歌詞カードから二重表記の箇所を確認し、手元のテキストファイルと照らし合わせる
- ③ 確認済みの歌詞テキストファイルを全文検索システム「ひまわり」(「ひまわり」には形態素解析エンジン「MeCab」および形態素解析用辞書「UniDic」を設定する)にインポートし、歌詞全文コーパスを作成する

歌詞における二重表記の一覧表 作成手順:

- ① 歌詞全文コーパスで、“(”と“)”をキーワードに検索し、丸括弧がついている文字列を検出する
- ② 検索結果の中から、本研究における二重表記の定義にあたるものを抽出し、リストアップする
- ③ 分析用の付加情報として、二重表記の形態論的情報・音韻論的情報・意味論的情報、該当楽曲の関連情報等を入力する
- ④ 形態素解析エンジン「MeCab」および形態素解析用辞書「UniDic」を用い、収集した二重表記について形態素解析をし、品詞、語種、文字種の情報をつけ加える
- ⑤ 主表記と副表記に分け、それぞれのモーラ数を数える
- ⑥ 漫画・短歌における二重表記の意味内容や分類について考察する先行研究を参考にし、本研究の意味タグを設定し、二重表記に意味タグを付け、主表記と副表記との関係から分類する
- ⑦ 該当楽曲の関連情報(歌手、作詞者、作曲者、発売元、ジャンル、テーマ等)を収集する
- ⑧ 以上の手順に従い、歌詞における二重表記の一覧表を作成し、問題やミスの有無を協力者にチェックしてもらい、校正する

## 現段階の規模

【表1】歌詞全文コーパスの内訳

収録年度	オリコンCDシングル年間売上ランキング集計曲数	収録曲数(A)	延べ語数	異なり語数
1969	50	42	6,640	1,162
1974	100	98	18,369	2,329
1979	100	87	20,190	2,925
1984	100	100	22,675	3,266
1989	100	101	27,205	3,465
1994	100	109	30,241	3,689
1999	100	108	34,569	3,984
2004	100	114	33,567	4,028
2009	100	123	43,007	4,318
2014	100	121	42,733	4,669
2019	100	128	52,401	5,523
合計	-	1,131	331,597	39,358

【表2】歌詞における二重表記の一覧表の内訳と集計結果

収録年度	歌詞に二重表記のある曲数(N)	出現率(N/A%)	二重表記の延べ個数(n1)	一曲*あたりの延べ個数の平均(n1/N)
1969	4	9%	4	1.00
1974	12	12%	21	1.75
1979	9	10%	10	1.11
1984	17	17%	33	1.94
1989	12	12%	22	1.83
1994	27	25%	55	2.04
1999	20	19%	54	2.70
2004	24	21%	57	2.38
2009	28	23%	62	2.21
2014	27	22%	58	2.15
2019	27	21%	52	1.93
合計	207	18%	428	2.07

【参考資料】

泉文明(1993)「二重表記の現在:短歌・俳句の表記の調査」『日本語学』12(3), pp.95-104, 明治書院。  
 今野真二(2013)「振り仮名:二つの言語の架け橋(特集:ことばの名脇役たち—書きことば)」『日本語学』32(5), pp.144-156, 明治書院。  
 佐竹秀雄(1980)「日本語の表記の表現性(文学表現総論<特集>)—表現者の姿」『国文学解釈と鑑賞』45(8), pp.59-65, 至文堂。  
 オリコンミュージックストア配信サービス「オリコン年間CDシングルランキング」  
<https://music.oricon.co.jp/php/special/Special.php?pcd=yrankindex>  
 歌詞検索サービス「歌ネット」<https://www.uta-net.com/>  
 歌詞検索ソフト「Lyrics Master 2」ver. 2.5.4 <http://www.kenichimaehashi.com/lyricsmaster/>  
 形態素解析エンジン「MeCab」(0.996) <https://taku910.github.io/mecab/>  
 形態素解析用電子辞書「UniDic」ver.3.1.0 <https://cod.ninjal.ac.jp/unicd/>  
 クラブハウス(1998)『オリコンno.1 hits 500:オリコンチャート1位ヒットソング集上(1968~1985)』三省堂(2010)『スーパー大辞林3.0』  
 全文検索システム「ひまわり」ver.1.7.1 <https://csd.ninjal.ac.jp/lrc/index.php?himawari>