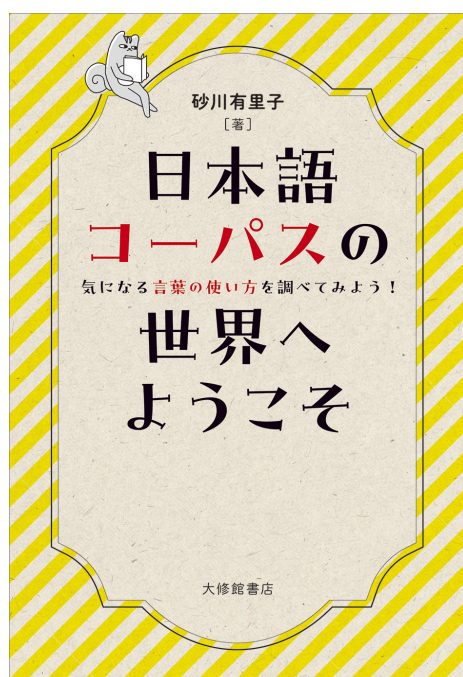


# 日本語コーパスの世界へようこそ —気になる言葉の使い方を調べてみよう—

砂川有里子（筑波大学名誉教授）  
加戸玲子（編集者・フリーランス）  
松岡滯（大修館書店企画推進部）

2024/11/30  
アクラスzoom寺子屋

1



砂川有里子 [著]

『日本語コーパスの世界へようこそ  
—気になる言葉の使い方を調べてみよう—』

大修館書店 2024年

2

## 国立国語研究所の コーパスの使い方を知りたい人は

- ▶ 中納言コーパスの動画をご覧ください。
  - ▶ <https://clrd.ninjal.ac.jp/tutorial.html>
    - ▶ 中納言の概要
  - ▶ 中納言を活用したコーパス日本語研究入門解説動画（中俣尚己）
    - ▶ part02 簡単に検索してみよう
    - ▶ part04 基礎的な検索
    - ▶ part06 コロケーションの検索
    - ▶ part14 中納言で検索できるいろいろなコーパス

(など多数)

3

## 今日のお話

### I. 砂川（著者）

- ▶ 出版に至った経緯
- ▶ コーパスとは何か
- ▶ この本の内容

### II. 加戸（編集者）

- ▶ わかりやすい文章を書くには
- ▶ 論文とエッセイはどう違う

### III. 松岡（出版社）

- ▶ 企画から出版まで

4

## 出版に至った経緯

ことば文化特設サイト

エッセイ初体験

一般向け書籍初体験

- I. 砂川（著者）  
初体験の面白さ
- II. 加戸（編集者）  
論文とエッセイの違い
- II. 松岡（出版社）  
企画から出版まで

5

## コーパスとは何か

- ▶ 話し言葉や書き言葉の貯蔵庫
  - ▶ 実際に使われた言葉
  - ▶ 大量に集める
  - ▶ 電子化されている

6

## コーパスの種類

- ▶ 母語話者コーパス 日本語母語話者の日本語の話し言葉や書き言葉のコーパス
- ▶ 学習者コーパス 日本語学習者者の日本語の話し言葉や書き言葉のコーパス
- ▶ 横断コーパス 同じ時期に大量に集めたコーパス
- ▶ 縦断コーパス 数年間にわたって同じ人々の言葉を集めたコーパス

7

## コーパスの種類

### 国立国語研究所のコーパス

コーパス名	略称
書き言葉 現代日本語書き言葉均衡コーパス	BCCWJ
話し言葉 日本語話し言葉コーパス	CSJ
話し言葉 日本語日常会話コーパス	CEJC
話し言葉 昭和話し言葉コーパス	SSC
話し言葉 名大会話コーパス	NUCC
話し言葉 現日研・職場談話コーパス	CWPC
通時 日本語歴史コーパス	CHJ
通時 昭和・平成書き言葉コーパス	SHC
方言 日本語諸方言コーパス	COJADS
日本語学習者 中国語・韓国語母語の日本語学習者縦断発話コーパス	C-JAS
日本語学習者 多言語母語の日本語学習者横断コーパス	I-JAS
日本語学習者 北京日本語学習者縦断コーパス	B-JAS

8

## コーパスの利用

「中納言」の利用申込みが必要

<https://clrd.ninjal.ac.jp>

すべて無料で使えます。有料版もあります

9

## 学習者コーパスで何が分かる？

多言語母語の日本語学習者横断コーパス (I-JAS)

協力者の内訳		人数
海外の学習者数		850人
国内の学習者数	教室環境	100人
	自然環境	50人
日本国内の日本語母語話者数		50人

10

## 学習者コースで何が分かる？

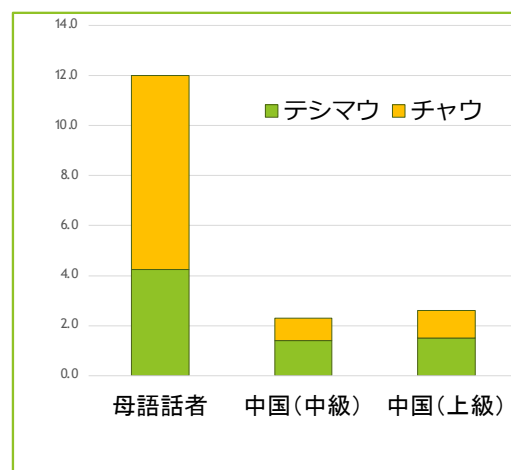
I-JASの「対話」を使って

中国語話者200人のテシマウとチャウの使い方

11

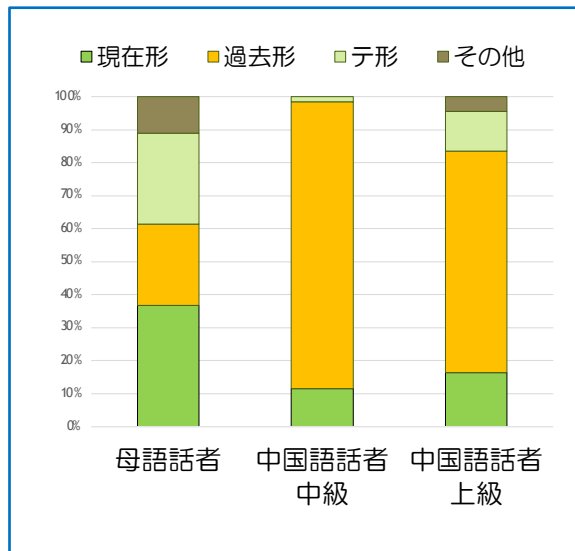
## 学習者コースで何が分かる？

テシマウとチャウの使用頻度（5千語あたり）



12

## 学習者コーパスで何が分かる？



13

## 学習者コーパスで何が分かる？

### この傾向は他の言語を母語とする学習者にも共通する

- ▶ 中国語話者のテシマウとチャウの使用頻度は母語話者に比べて非常に少ない。
- ▶ 母語話者は現在形を最も多く使うが、中国語話者は過去形を最も多く使う。
- ▶ 母語話者は現在形・過去形以外もかなり使うが、中語話者はあまり使わない。
- ▶ 以上の傾向は上級になってもそう大きくは変わらない。

14

## 学習者コースで何が分かる？

- 教室ではテシマウの過去形の練習ばかりする。
- 「完了の強調」や「後悔や遺憾の意」という  
「過去」の意味と強く結びつく
- 初級で導入されたら、それ以降は既習事項として指導しない。

15

## 学習者コースで何が分かる？

上級になっても使う頻度が低いのはなぜなのか。  
初級で教えられていない用法があるのかも。。。

中国語話者の使用が少なく、日本語母語話者の使用が多い「現在形」や「その他」の用例を観察する。

16



## 学習者コースで何が分かる？

1. 〈カロリーとかあんまり気にせずに、〉 あ、みんな気にしないで〈あー〉それは**食べちゃえ**みたいな感じで {笑}
2. まあ、いわまあ害虫みたいなもので、〈ん、ん、ん〉まあ**殺してしまえば**いいんじゃないかという話になるんですけど
3. みんな授業は**寝ちゃってます** {笑}
4. 私はすごい好きなので〈そうなんですねー〉時間を忘れて**没頭してしまうん**ですねー

「完了」でも「遺憾」でもない

教室では習わない用法。

17

## 学習者コースで何が分かる？

- ① 何らかの抵抗や障害を乗り越えて実現する
  1. カロリーとかあんまり気にせずに**食べちゃえ** vs. **食べろ**
  2. たまにはちょっと**外食しちゃおうかな** vs. **外食しろ**
  3. まあ**殺してしまえば**いいんじゃないかと vs. **殺せばいいじゃないかと**
- ② 自分のコントロールを離れて、思いがけず実現する
  1. みんな授業は**寝ちゃってます** vs. **寝てます**
  2. 時間を忘れて**没頭してしまうん**です vs. **没頭するん**

わだかまり

あえて気にしない

思いがけず

意図せず

18

## 学習者コーパスで何が分かる？

1. テシマウは初級段階で導入され、「完了の強調」や「遺憾の意」を表すものであると説明される。→過去形の多用に繋がる。
2. テシマウやチャウは既に習った形式として扱われ、初級以降で注意を向けられることがない。→非用に繋がる。
3. テシマウやチャウには「完了の強調」や「遺憾の意」を表す用法以外の用法もある。
4. これらの用法は教えられないまま過ぎてしまう。→上級になっても使えない。
5. これらの用法は母語話者にも気付かれていないことが多い

19

## 母語話者コーパスで何が分かる？

現代日本語書き言葉**均衡**コーパス (BCCWJ)

- ▶ 書籍・新聞・白書・ブログなどから集めた1億語の書き言葉コーパス

中納言への申込み申請なしで使えるソフトを紹介します

20

## 母語話者コーパスで何が分かる？

- ▶ 夜空で星がきらきら輝いている。
- ▶ 夜空で星がキラキラ輝いている。



21

## 母語話者コーパスで何が分かる？

	書籍	雑誌	Yahoo! ブログ	韻文
きらきら	328	14	49	12
キラキラ	233	77	175	5

22

## 母語話者コーパスで何が分かる？

▶ コロケーションや類義語の調査に便利なソフト

The screenshot shows the homepage of the NINJAL-LWP for BCCWJ (NLB) corpus. At the top, the title "NINJAL-LWP for BCCWJ (NLB)" is displayed in a blue dashed box. To the right is a link to the "国語研ホームページ". Below the title are two buttons: "検索を開始する" (Start Search) and "操作説明書を開く" (Open Manual). A navigation menu includes links for "NLBとは", "使い方", "機能表現検索", "オノマトペ検索", "使用コーパス", "ご利用にあたって", "更新履歴", and "お問い合わせ". A section titled "■ NINJAL-LWP for BCCWJ とは" explains that the corpus was constructed by the National Institute of Japanese Language (国語研) and is used for searching with the National Institute of Japanese Language and the Lingo Language Research Institute (Lago言語研究所).

23

## 母語話者コーパスで何が分かる？

▶ コロケーションや類義語の調査に便利なソフト

コンピューターを〇〇

NINJAL-LWP for BCCWJ の 見出し語検索

24

# 母語話者コーパスで何が分かる？

NINJAL-LWP for BCCWJ

→ 2語比較検索

コンピューター

絞り込み 元に戻す

すべて 名詞 動詞 形容詞 連体詞 副詞 オノマトペ

見出し	読み	ローマ字表記	頻度
コンピューター	コンピューター	konpyuta	5,168

25

# 母語話者コーパスで何が分かる？

コンピューター 頻度=5,168

グループ別 パターン頻度順 基本

コンピューターを... 153種類

コロケーション	コーパス全体		
	頻度	MI	LD
コンピューターを使う	87	8.30	5.74
コンピューターを利用する	53	9.50	6.89
コンピューターをい	44	11.17	8.42
コンピューターを用	21	8.37	5.76
コンピューターを活	20	9.95	7.21
コンピューターをす	17	21.63	9.93

» 助詞+動詞

パターン	頻度	比率
コンピューターが...	312	
コンピューターは...	91	
コンピューターも...	15	
コンピューターの...	15	
コンピューターを...	571	

26

## 母語話者コーパスで何が分かる？

- ▶ コロケーションや類義語の調査に便利なソフト

海外 VS. 国外

NINJAL-LWP for BCCWJ の 2語比較検索

27

## 母語話者コーパスで何が分かる？

NINJAL-LWP for BCCWJ

海外 国外

絞り込み

元に戻す

2語比較

名詞

動詞

形容詞・連体詞

副詞

オノマトペ

見出し

読み

ローマ字表記

頻度

?

<input checked="" type="checkbox"/>	海外	カイガイ	kaigai	8,424
<input checked="" type="checkbox"/>	国外	コクガイ	kokugai	591

28

## 母語話者コーパスで何が分かる？

海外 頻度=8,424      国外 頻度=591

グループ別    パターン頻度順    基本

助詞+動詞

海外		国外	
パターン	頻度	パターン	頻度
海外に...	987	国外に...	162
海外で...	559	国外で...	51
海外から...	223	国外から...	15
海外へ...	212	国外へ...	43
海外を...	47	国外を...	4
海外は...	18		
海外と...	17	国外と...	1
海外も...	13	国外も...	2
海外が...	12		
海外の...	6		
海外より...	5		
海外まで...	3	国外まで...	1

海外に... 190    国外に... 57

コロケーション	海外			国外			LD差
	頻度	MI	LD	コロケーション	頻度	MI	
海外に進出する	28	11.72	8.91				8.91
海外に依存する	30	11.17	8.66				8.66
海外に派遣する	20	10.22	7.85				7.85
海外に在留する	5	13.32	7.30				7.30
海外に留学する	7	10.49	7.23				7.23
海外に在任する	5	12.38	7.23				7.23
海外に赴任する	6	10.78	7.17				7.17
海外にシフトする	5	11.13	7.05				7.05
海外に旅行する	6	10.25	7.00				7.00
海外に住む	43	8.77	6.89				6.89
海外に出勤する	4	10.49	6.66				6.66
海外に滞在する	5	9.35	6.48				6.48
海外に送金する	3	10.70	6.37				6.37
海外に委託する	4	9.41	6.32				6.32
海外に行く	137	7.97	6.19				6.19
海外に離飛する	2	14.58	6.04				6.04

頻度 すべて 2以上 5以上 10以上  
LD差 すべて ±2以上 ±3以上 ±5以上

ページ / 3

29

## 母語話者コーパスで何が分かる？

▶ ちょっと気になる言葉を調べるのにも使える

〇〇活

NINJAL-LWP for TWC の 見出し語検索

30

# 母語話者コーパスで何が分かる？

NINJAL-LWP for TWC

活  絞り込み 元に戻す

すべて 名詞 動詞 連体詞 形容詞 副詞

見出し	読み	ローマ字表記	頻度
活	カツ	katsu	13,336

→ 2語比較検索

31

# 母語話者コーパスで何が分かる？

活 頻度=13,336

グループ別 パターン頻度順 基本

グループ	頻度	比率
活+名詞・接尾	2,555	
活+の+名詞	473	
活的(な)+名詞	4	
活って+名詞	7	
名詞+活	7,107	
名詞+の+活	502	
名詞+的(な)+活	15	
並立	264	

接頭辞・動詞辞

パターン	頻度	比率
接頭辞+活	1,216	

名詞+活 328種類

コロケーション	頻度	MI	LD
婚活	5,417	16.23	13.23
朝活	201	8.24	5.86
憲活	172	10.35	7.64
特活	79	10.28	7.23
失活	71	12.41	7.98
学活	70	5.81	3.47
産活	64	7.99	5.47
磁気活	62	10.31	7.09
結婚活	58	6.72	4.32
【数字】活	34	-1.58	-3.87
エコ活	32	7.65	5.00
就職活	30	6.23	3.79
夜活	25	5.14	2.76
婚活	25	8.78	5.66

婚活 5,417件

婚活？  
(ジムをたまり場にして話し相手さがし)

婚活中です。  
(好きです。の意味。(婚活中) 婚活 総)

婚活も同じだ。  
(本当に大切なものを見つけるために |

婚活ポイント！  
(飲み会で嫌われる男の行動)

婚活中である。  
(大阪吃音教室 ことば文学賞・最優秀)

恋愛・婚活講座

32



## 本の目次

1. 「きらきら」と「キラキラ」
2. 無意識に使ってしまう「～てしまう」
3. 「〇〇音痴」と「〇〇外れ」
4. 「〇〇活」と「〇〇感」
5. 逃亡先は「海外か」「国外か」

## 本の目次

6. オネエ言葉と女言葉
7. 「男」と「女」の描き方
8. 職場で人を何と呼ぶ？
9. 「うそだい！」は古めかしい？
10. 「全然」と「全く」