

日本語教育に役立つ コーパスの使い方

砂川 有里子

国立国語研究所客員教授

日本台湾交流協会2018年度第2回日本語教育研修会 1
実践大学高雄市教学中心 2018/11/17

BCCWJ

現代日本書き言葉均衡コーパス Balanced Corpus of Contemporary Written Japanese

書籍、雑誌、新聞、白書、ブログなど
1億430万語のデータ

4

コーパスとは何か

実際に使用された言語データを大量に集めて
電子化した言語研究のための言語資料

現実に使用された言語を
大量に分析する

内省では気づけない問題に
気づくことができる

2

国立国語研究所

コーパスデータベース

[https://www.ninjal.ac.jp/
database/](https://www.ninjal.ac.jp/database/)

3

いろいろなジャンルから 日本語母語話者の書き言葉を バランス良く集めたコーパス

現代日本語の書き言葉の多様性を把握し
その全体像を捉える

国立国語研究所
2011年に公開

5

BCCWJの用例検索システム

オンライン・無料

タイプ	概要	申込
少納言	文字列検索のみ 用例は500まで	不要
中納言	短単位／長単位検索 すべての用例をダウンロードできる	必要
NINJAL-LWP for BCCWJ	BCCWJの調査結果を示し、 語の特徴的な振る舞いを提示する	不要

6

少納言
KOTONOHA「現代日本書き言葉均衡コーパス」

検索結果
4085件の結果が見つかりました。そのうち 500 件を表示しています。

前文脈	外語	学年	後文脈	執筆者
とです。私たち今までの英語学習は、常にこの逆だったように思っています。まず外へ出でたことが実感となっていました。 ■ 20	国語	を読んで、その後に日本語に訳して理解してきたのが結果、英語をマスター	松木 修三(著)	
内省をつかむことが実感となっていました。 ■ 20	国語	、「手帳トヨタ」、「手帳トヨタ」は中国四川省の一		
である。「手帳トヨタ」は、中国語	中国語	学」「手帳」、「手帳トヨタ」、することを「誤り」と断じ、「国語」と「学」にわ	安田 敏郎(著)	
行く、というと彼女の夢は母さんだったはずがこの前開いたら、実感が変わったか外へ出でたことがあります。このようないい基本的には「	国語	大学に行く、って言ってました。こういって、人の意図に気付かれない人はいるのです		
国語を「共通語」にすら意識の内的仕方をみることができます。このようないい基本的には「	国語	」の誤認の及ばない地域という誤認が言説レベルで一致してたたかれていたといつも	安田 敏郎(著)	
中国語入力ソフトをインストールする必要はありません。WindowsのSIC中	国語	入力のソフトがいいていますから、セットアップ		
、中国語は表記文字による文字体系(すなはちアルファベット)であるために、中華中	韓国語	の表記文字でさえ、発音の再構成がむずかしい。音声(音の)の要素がよくわからないから	スティーヴン・ロジャー・チャーチャー(著)/ 鈴木 晶(訳)	
文書書士が応じます。英語、中国語、韓	国語	の適度あり、つまりしないと必ずしも男を見てない		

少納言

文字列検索しかできない。

前文脈	検索	後文脈
1. ジャンル: その言葉がとっても嬉しいです。(私はデブだと思うのですぐ...) それから	から	Kgになりました。彼は初めて1つうそをしてくれて8:
2. お腹が膨らむのが嫌なに見えるのです。しかし、入るような人はお腹持ちの人	から	ばかりなようなさじますから、自分では言わない
3. 大学に行く、って言ってました。こういって、人の意図に気付かれない人はいるのです	から	ちょっと動いていただるんだ」と言つたのですが、教えて下さい。私も姉がいる人
4. お腹の意図も言葉も動いています。体力がないと強引やすいですか?主人が「体力がない	から	ありません。見た目で分かれないのでなんですか?バツ
5. 伸びて行ったのですが、身体を動かさないといけないんだと断られました。ジャ	から	ンジンだ
6. は外で友達と一緒に歩いて「歩みに説きや！」となつたからです。P.C.はなく、携帯	から	歩みに説きや！」といつも歩いていたので、社
7. 女の子とあまり恋愛の話をしていませんです。それから男を見てない	から	からね。それから、『歩みに説きや！』と言つて見てください。貴方が男として見て
8. 中国語入力ソフトをインストールする必要はありません。WindowsのSIC中	から	い？それは良い。貴方が男として見て
、中国語は表記文字による文字体系(すなはちアルファベット)であるために、中華中	韓国語	ね。それから、『歩みに説きや！』といつも歩いていたので、社
文書書士が応じます。英語、中国語、韓	国語	からね。それから、『歩みに説きや！』といつも歩いていたので、社

中納言

結果が見つかりました。そのうち 500 件を表示しています。

#	前文脈	キーワード	後文脈	語種	語種読み	語種素	語種固有名詞	固有名詞	活用形	活用型	サブコード名
1	(...)として「マレー(人)」の家庭にいる、つまりいる」という印象で、台湾で	等 が 挙げ られる	国語 ク 国語 台湾で	国語	クゴ	国語					出版・書籍
21	たゞ「皇室化イデオロギー」の「性別」の異なるものである。 そこでも、 には性別			国語	コクゴ	国語					出版・書籍
70	J. かねてから、『歩みに説きや！』といつも歩いていたので、社	全ての用例がダウンロードできる。		国語	コクゴ	国語					川口

9

NLB

中納言

前文脈	キー	後文脈
を飲みます。おなか(が)りか、と食つた。から(だ)が示す。食べ物の残り(が)	冷え	ているときは、腹巻きをしたり、かけていたから、それと、わずて示す。食べ物の残り(が)
...が冷える 115種類		
コロケーション	コーパス全体	
頻度	MI	LD
体が冷える	48	8.66 5.04
身体が冷える	15	9.06 5.43
足が冷える	12	8.28 4.65
(人名)が冷える	8	1.31 -2.29
お腹が冷える	8	9.72 6.02
下半身が冷える	8	12.40 8.24
からだが冷える	7	9.64 5.93
手足が冷える	6	11.09 7.16
マグマが冷える	5	12.73 8.17
空気が冷える	5	8.45 4.78
頭が冷える	4	5.87 2.25
足元が冷える	4	9.48 5.71

NLB



I-JAS

多言語母語の日本語学習者横断コーパス

International Corpus of Japanese as a Second Language

国立国語研究所で構築中の学習者コーパス

2020年3月に完成予定(現在一部公開中)

学習者の母語

◆ 多言語母語話者 (12言語)

表1 調査対象の12言語		
インドネシア語	タイ語	ベトナム語
韓国語	トルコ語	中国語
ハンガリー語	英語	フランス語
スペイン語	ドイツ語	ロシア語

2020年完成時的人数

2018年現在660名分公開

◆ 大規模(合計1,050人)

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

特徴

- すべての音声データとテキストデータがダウンロードできる。
- さまざまな比較が可能。
 - 母語話者と学習者
 - 教室環境と自然環境
 - 母語別
 - 海外学習者と国内学習者、など
- 中納言による検索が可能。

1
6類義語の研究
(冷える vs. 冷める)

砂川(2014)

多様なタスク

データの種類	タスク名
発話	ストーリーテリング (2種)
	対話(インタビュー)
	ロールプレイ (2種)
	絵の描写
作文	ストーリーライティング (2種)

1
5

冷える VS. 冷める

- 料理が冷める前に食べて下さい。
- ? 料理が冷える前に食べて下さい。
- 風邪を引くから体を冷やしちゃいけないよ。
- ? 風邪を引くから体を冷ましちゃいけないよ。

1
8

N LB

...が冷える 110種類			
コロケーション	頻度	MI	LD
体が冷える	44	8.63	4.98
身体が冷える	15	9.18	5.50
足が冷える	12	8.40	4.73
【人名】が冷える	8	1.43	2.22
お腹が冷える	8	9.81	6.07
からだが冷える	7	9.77	6.02
手足が冷える	6	11.21	7.24
下半身が冷える	6	12.09	7.90
空気が冷える	5	8.58	4.87
頭が冷える	4	6.00	2.34
足元が冷える	4	9.60	5.79
腰が冷える	4	7.95	4.25
脚が冷える	4	10.50	6.55
全身体が冷える	4	9.09	5.32
先端が冷える	4	5.60	1.95
マグマが冷える	4	12.46	7.90
エンジンが冷える	4	8.54	4.81
筋肉が冷える	3	6.05	2.39
脳が冷える	3	10.63	6.57
汗が冷える	3	7.93	4.21
肌が冷える	3	9.13	5.32
地域が冷える	3	7.41	3.72

...が冷める 77種類			
コロケーション	頻度	MI	LD
熱が冷める	31	11.86	7.59
ほとばしりが冷める	17	17.92	11.34
気持ちが冷める	17	9.03	4.79
【数字】が冷める	6	0.68	3.54
お湯が冷める	6	11.46	7.06
酔いが冷める	5	11.95	7.44
目が冷める	4	5.32	3.09
料理が冷める	3	7.38	3.13
愛情が冷める	3	9.80	5.46
豊かが冷める	3	8.13	3.87
心が冷まる	3	5.62	1.39
紅茶が冷める	2	10.41	5.93
潮が冷める	2	10.97	6.39
空気が冷める	2	7.84	3.57
熱狂が冷める	2	11.76	6.96
熱意が冷める	2	11.14	6.52
怒りが冷める	2	8.43	4.14
体が冷める	2	4.75	0.52
スパゲッティが冷める	2	11.05	6.45
ショックが冷める	2	8.75	4.44
桃が冷める	1	8.92	4.51

...が冷える

- 「体」「身体」「足」「お腹」「手足」「下半身」など身体やその一部を表す名詞との共起が多い。
- 身体以外の名詞では「空気」「マグマ」など。
- 具体物を表す名詞と共にしやすく、抽象物を表す名詞との共起は少ない。

...が冷める

- 身体に関わる名詞との共起が少なく、その代わりに「料理が冷める」「紅茶が冷める」など食べ物に関する名詞が目に付く。
- 「熱」「ほとばしり」「気持ち」「愛情」「熱狂」「ショック」など抽象物を表す名詞との共起は「冷える」よりも多い。

具体物を表す名詞との共起

① ○「冷える」、○「冷める」

- 4の桃が冷めたら／冷えたら、くし形に切って皿に盛りつける。(ル・コルドンブルー・パリ校編『ル・コルドンブルー・パリ校のマルシェノート』)
- つまり体が冷めにくく／冷えにくくなると言ふことです。(Yahoo!知恵袋)

1
9

具体物を表す名詞との共起

② ○「冷える」、×「冷める」

- 水のなかを歩くとどうしてもからだが冷える／×冷めるため、真夏でもあたたかいものが食べたくなります。(越谷英雄『はじめてのリバー・トレッキング』)
- 空気がきーんと冷えた／×冷めた夜ふけに、ポンちゃんがふと目をさますと、となりのベッドで母さんが体を起こして、顔をしかめていました。(甲斐裕美『ゆたかな命のために』)

2
2

具体物を表す名詞との共起

✧「冷える」と「冷める」

物体の温度が下がる。

- 「冷える」：元々の温度からさらに低い温度に下がる。
- 「冷める」：元々の温度より高い温度から元々の温度に下がる。

2
5

抽象物を表す名詞との共起

① ○「冷える」、○「冷める」

- 夫婦の愛が冷めて／冷えてお金の為だけのつながりになってしまってとにかく両親さえ揃っていれば子供は幸せ？(Yahoo!知恵袋)
- 妹との間柄が急速に冷えて／冷めてきたレオは、家を出て、イタリアに去ってゆく。(金関寿夫『現代芸術のエポック・エロイク』)

2
0

具体物を表す名詞との共起

③ ×「冷える」、○「冷める」

- お料理が冷めない／×冷えないように、お皿の下に電熱器のようなものがありました。(Yahoo!ブログ)
- お風呂はお湯が冷めない／×冷えないうちに間隔をおかずに入るようにして、追い焚きをしないようにしましょう。(広報南アルプス)

2
1

抽象物を表す名詞との共起

② ○「冷える」、×「冷める」

- 景気が冷えれば／×冷めれば、税収は落ちます。
<http://oshiete.goo.ne.jp/qa/6787496.html>
- 現在のデフレは国際的な競争に晒されていること、また国内では少子化人口減などでマーケットが冷えて／×冷めているからでしょう。
http://kiyotani.at.wery.info/201001/article_5.html

2
7

抽象物を表す名詞との共起

③ ×「冷える」、○「冷める」

- 60年安保の熱狂が冷めて／×冷えてから社会に出たものの、これといった職は見つかなかつた。(大住昭『アジア日本人群像』)
- 帰国後のパーティは拳式の感激が冷めない／×冷えないうちに開くのがベターで、一般には帰国後、1週間以内に開くカップルが多いようです。(小林淳『海外オリジナルウェディング』)

2
8

抽象物を表す名詞との共起

「冷える」：相手に対する感情や関係が悪化する。

「冷める」：親愛の気持ちや関係が薄れる。高ぶった感情が静まる。

2
9

冷える

① 具体物を表す名詞と共に

「元々の温度からさらに下がる」

- 風呂に入って冷えた体を温める。
- マグマは高温だが、冷えると岩石になる。

② 抽象物を表す名詞と共に

「元々の状態から悪化する」

- 二人の関係が急速に冷えている。
- 増税すれば、ますます景気が冷えるだろう。

③ 慣用表現：肝が冷える、背筋が冷える

3
0

冷める

① 具体物を表す名詞と共に

「高い温度から元々の温度へ下がる」

- 料理が冷めないうちに食べましょう。
- お風呂のお湯が冷めてしまった。

② 抽象物を表す名詞と共に

「高いレベルになった状態が元々の状態に戻る」

- ひとときの熱狂が冷めて、町はふたたび静かになった。
- 歳を取っても情熱が冷めることなく、仕事を続けている。

③ 慣用表現：ほとぼりが冷める、酔いが冷める、混乱が冷めやらぬ

3
1

中国語母語話者の漢字の発音 (非漢字圏学習者との比較)

砂川・黒沢(2017、近刊)

3
2

同形語と非同形語

△ 同形語：

- 中国語にも同じ形の漢字単語があるもの
- 学校、経済、政治、地図、能力、歴史
- 緊張、専門、得意、無理、試験、先生、人間、勉強

△ 非同形語：

- 中国語にはその形の漢字単語がないもの
- 映画、会社、授業、自分、卒業

電影、公司、課業、自己、畢業

3
3

中国人の発話データ 発音の誤用は同形語のほうが多い

正用と誤用の頻度(頻度2以上の漢語名詞)

	正用頻度	誤用頻度	合計
同形語	1,943 (91.4%)	183 (8.6%)	2,126 (100%)
非同形語	965 (94.4%)	57 (5.6%)	1,022 (100%)

砂川・黒沢(2017)

(χ²=8.57; df=1; p<.005)
3
4

フランス語母語話者との比較

—正用と誤用の頻度と比率—

母語	フランス語		中国語		
	頻度/比率	頻度	比率	頻度	比率
正用	2,735	93.3%		3,115	92.0%
誤用	195	6.7%		271	8.0%
計	2,930	100%		3,386	100%

砂川・黒沢(近刊)

(χ²=3.98; df=1; p<.05)
3
5

考察

- 中国人学習者は、自然発話を行う際に、母語である中国語の漢字音が介在してしまう(負の転移)。
- フランス人学習者に比べて中国人学習者のほうが漢字単語の使用頻度が高い。
- 使用に対する誤用の割合は中国人学習者のほうが大きい。
- 漢字圏の学習者の漢字単語の習得は、発音に関しては、非漢字圏の学習者に比べて不利である。

3
6

テシマウの使用状況 (母語話者との比較)

砂川(2018)

- 母語話者**
- ・ まだ分僕は、ま企業に、**言っちゃう**んだろうな。
 - ・ あたかい所にばっかり**いたくなってしまう**ので。
 - ・ ぶどうとか、の皮も**入れちゃつたり**して。
 - ・ 年寄りになつたら**動けなくなっちゃう**から。

- 学習者**
- ・ ちょっとえーがを**しちやつた**のでー。
 - ・ 重い病気なのではい**亡くなってしま**いました。
 - ・ 今はちょっと急げ者**になつてしま**いました。
 - ・ 知つていて**忘れてしま**いました。

第2部 ワークショップ

1. BCCWJの検索法

- NLB (NINJAL-LWP for BCCWJ)
- 中納言

2. I-JASの検索法

- 中納言

テシマウ(チャウ)の使用頻度

I-JASの発話データ

表2 テシマウの使用頻度と使用比率

母語話者の使用頻度を100とした場合、
学習者の使用頻度平均は22にすぎない。

母語話者との比較	100	19	17	26	24	13	27	25	22
----------	-----	----	----	----	----	----	----	----	----

過去形と非過去形

母語話者		学習者							合計
		スペイン	フランス	ベトナム	ハンガリー	韓国	自然環境	教室環境	
過去形	64 (29%)	36 (86%)	31 (84%)	54 (95%)	38 (73%)	15 (52%)	30 (51%)	38 (70%)	242 (73%)
非過去形	155 (71%)	6 (14%)	6 (16%)	3 (5%)	14 (27%)	14 (48%)	29 (49%)	16 (30%)	88 (27%)

母語話者は非過去形が7割、
学習者は過去形が7割。

頻度差の原因

中級学習者は、「完全に済ませることを強調する」、または「遺憾・後悔を表す」の用法だけ。

母語話者はそれ以外の用法も使う。

まあ害虫みたいなもので、まあ殺してしまえばいいんじやないか。
先輩におごってもらつちやつた！

遺憾以外の気持ちを表すテシマウ

1. きれいな景色を見ると、ついカメラを向けてしまいます。
2. 店内が静かなので、何時間でもいてしまいそう。
3. なんと大企業に受かっちゃいました。
4. 宝くじ買って大金あてちゃおうと！
5. いつそのこと殺してしまおうか。

NLBの用途

コロケーション

- 「被害を」の後にどんな動詞が使われるか。

類義語

- 「冷める」と「冷える」の使い分けは。

表記

- 「タバコ」「たばこ」「煙草」はどの表記が多いか。

NLB

<http://nlb.ninjal.ac.jp>

NINJAL-LUP for BCCWJ (NLB)

使い方 操作説明書を聞く

ここでは特に重要な点だけを説明します。
詳しくは、「使い方」や「操作説明書」を
読んでください。

ここでは、「言葉」について調べています。まずは画面上の入力ボックスに「言葉」または「ことば」（カタカナも可）を入力して、[読み込み]ボタンをクリックします。下のリストに「言葉」という見出しが現れますのでそこをクリックします。

NINJAL-LUP for BCCWJ (NLB)

検索結果を見る 操作説明書を聞く

ここでは、「言葉」について調べています。LBでは、検索語句を入力する代わりに、必ず調べたい語（LBではこれを「見出しどと」と呼びます）を選びます。見出しどととして選べるのは、名詞、動詞、形容詞、連体詞、副詞の5種類の内部語です。

コロケーションの検索

ここでは、「言葉」について調べています。まず画面上の入力ボックスに「言葉」または「ことば」（カタカナも可）を入力して、[読み込み]ボタンをクリックします。下のリストに「言葉」という見出しが現れますのでそこをクリックします。

NINJAL-LUP for BCCWJ

とめる 続込み 元に戻す

すべて 名詞 動詞 形容詞 連体詞 副詞 オノマトペ ローマ字表記 頻度 +

語	例文	読み	意味
いる・非自立	イル	iru	968,064
こと	コト	koto	701,470
の	ノ	no	686,025
する	スル	suru	601,911
なる	ナル	naru	484,861
ある	アリ	aru	471,783
の	ソノ	sono	347,337
よう	ヨウ	you	333,829
買う	イフ	iu	281,579
この	コノ	kono	272,773
もの	モノ	mono	234,829
ん	ン	n	205,301
思う	オモウ	omou	199,915
できる	デキル	dokiru	194,228
くる・非自立	クル	kuru	171,836
ない	ナイ	na	163,065
人	ヒト	hito	145,992

NINJAL-LUP for BCCWJ

とめる 続込み 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 +

語	見出し	読み	意味	頻度
止める	トメル	tomeru	8,137	
とめる	トメル	tomeru	1,489	
泊める	トメル	tomeru	277	
富める	トメル	tomeru	61	

止める 頻度=8,137

グループ別 パターン頻度順 基本 ?

名詞+助詞 ▲

パターン	頻度	比率
...が止める	451	
...は止める	954	
...も止める	193	
...の止める	15	
...を止める	4,228	
...に止める	1,823	
...へ止める	7	
...で止める	884	
...と止める	41	
...から止める	52	
...まで止める	31	
...よし止める	3	

名詞+複合助詞 ▲

パターン	頻度	比率
...として止める	2	

コロケーション コーパス全体 頻度 + MI LD

「グループ別」とは

1. 名詞+助詞+止める

- …が止める
- …を止める

2. 止める+名詞

- 止める+名詞
- 止め+名詞

3. 止める+助動詞

- 止め+れる
- 止め+ない

文法パターンを提示

止める 頻度=8,137

グループ別 パターン頻度順 基本 ?

名詞+助詞 ▲

パターン	頻度	比率
...が止める	451	
...は止める	954	
...も止める	193	
...の止める	15	
...を止める	4,228	
...に止める	1,823	
...へ止める	7	
...で止める	884	
...と止める	41	
...から止める	52	
...まで止める	31	
...よし止める	3	

名詞+複合助詞 ▲

パターン	頻度	比率
...として止める	2	

コロケーション コーパス全体 頻度 + MI LD

止める 頻度=8,137

グループ別 パターン頻度順 基本 ?

名詞+助詞 ▲

パターン	頻度	比率
...が止める	451	
...は止める	954	
...も止める	193	
...の止める	15	
...を止める	4,228	
...に止める	1,823	
...へ止める	7	
...で止める	884	
...と止める	41	
...から止める	52	
...まで止める	31	
...よし止める	3	

名詞+複合助詞 ▲

パターン	頻度	比率
...として止める	2	

コロケーション コーパス全体 頻度 + MI LD

足を止める 全1059種類

語	頻度	比率
足を止める	431	9.49
火を止める	272	9.19
車を止める	199	7.41
動きを止める	163	8.83
手を止める	163	6.40
のを止める	151	2.93
ことを止める	102	1.97
鳥を止める	82	8.57
それを止める	70	3.57
目を止める	70	4.80
【人名】を止める	49	0.02
鳥の飛行を止める	37	13.84
歩みを止める	36	10.41
名乗りを止める	30	10.50
言葉を止める	30	4.37
【一般】を止める	28	0.05
流れを止める	28	6.77
名を止める	27	4.71
呼吸を止める	26	7.61
愛を止める	26	5.07
圓形を止める	25	10.18
エンジンを止める	24	7.08
形を止める	23	4.61
ワガを止める	21	6.36
タクシーを止める	20	7.54

足を止めた。
火は足を止めた。
くは足を止めた。
開者は足を止めた。
竹は足を止めた。
おれは足を止めた。
ゆうきりん著「勢人姫」、1994, 913

近づくにつれ、闇が退き、色が目覚めていく。脛者は足を立て燃え、彼の行く手をさえぎっている。

ゆうきりん著「勢人姫」、1994, 913

閉じる

「被害を」のあとにどんな動詞が来るか

被害を... 132種類		
コロケーション	コーパス全体	
	頻度	MI
被害を受ける	598	10.01
被害をうける	120	11.04
被害をもたらす	118	10.37
被害をかる	111	8.50
被害をこうむる	73	13.32
被害を及ぼす	66	10.64
被害を防ぐ	54	11.32
被害を止める	44	10.05
被害を減らす	42	12.89
被害を軽減する	36	9.49
被害を軽くする	31	11.06
被害を出す	24	5.52
被害を取る	23	8.77
被害をじる	21	7.12
被害をうける	19	9.08
	7.30	

5

5



直接の係り受けがない場合があるので

注意が必要

「被害+に」

コロケーション	コーパス全体		
	頻度	MI	LD
被害に遭う	201	14.87	11.76
被害にある	96	5.35	2.71
被害にあう	63	12.79	9.77
被害にいる	32	10.81	7.97
被害につく	19	6.59	3.93
被害に係る	14	8.60	5.85

5

7

「被害にある？」

山・書籍: 0.82	図・書籍: 0.73	雑誌: 0.4
白書: 0.42	ベスト: 0.24	国会: 0.2
■ 何人も 被害にあってるよ。 (神保史郎著『タクシードライバーのとっておきの話』, 1991, 049)		
■ 一番多く 被害にあったのが私。 (加藤正夫著『攻めの構図、読みの力』, 2004, 795)		
何かの 被害にあったこともありません。 (Yahoo!知恵袋, 2005, Yahoo!オーバーク)		
嬉しい反面、何か 被害にあった気分です (阪本亮一著『成果が即上がる営業のコツ』, 2005, 673)		
子どもから高齢者まで 被害にあります。		

あう

5

8

NINJAL-LWPの注意点

- 形態素の誤解析がけっこうある。
- 特にひらがな書きに注意！
- 同じ漢字で読みが違うもの(「とめる」と「やめる」など)も注意！
- 直接の係り受け関係にない場合がある。(「被害をいる」など)

目的に合った検索が行われているか、
必ず用例をみて確認すること。

「準備」と「用意」

NINJAL-LWP for BCCWJ

ここに調べたい語を入力 (読みも可) 続込み 元に戻す

すべて 名詞 動詞 形容詞・連体詞 副詞 オノマトペ

いる・非自立 こと する なる ある ソノ ジュンビ

こと の る る る よウ ヨウイ

する する る る る よウイ

なる ナル る る る よウイ

ある アル る る る よウイ

その ソノ る る る よウイ

よう ヨウ る る る よウイ

言う イウ る る る よウイ

2語比較検索

類義語の検索

NINJAL-LWP for BCCWJ

準備 用意 続込み 元に戻す 2語比較

すべて 名詞 動詞 形容詞・連体詞 副詞 オノマトペ

いる・非自立 こと する なる ある ソノ ジュンビ

こと の る る る よウイ

する する る る る よウイ

なる ナル る る る よウイ

ある アル る る る よウイ

その ソノ る る る よウイ

よう ヨウ る る る よウイ

言う イウ る る る よウイ

NINJAL-LWP for BCCWJ

準備 用意 続込み 元に戻す 2語比較

すべて 名詞 動詞 形容詞・連体詞 副詞 オノマトペ

見る 読み ローマ字表記 頻度

いる・非自立 イル iru 968,064

こと コト koto 701,470

の ノ no 686,025

する スル suru 601,911

なる ナル naru 484,861

ある アル aru 479,785

その ソノ sono 347,357

よう ヨウ you 333,829

言う イウ iu 285,579

6

1

NINJAL-LWP for BCCWJ

準備 用意 続込み 元に戻す 2語比較

すべて 名詞 動詞 形容詞・連体詞 副詞 オノマトペ

見る 読み ローマ字表記 頻度

準備 ジュンビ junbi 7,270

用意 ヨウイ youi 1,553

6

2

6

3

準備 頻度=7,270		用意 頻度=1,553	
グループ別	パターン頻度順	基本	
▶ 助詞+動詞			
準備 用意			
パターン	頻度	パターン	頻度
○ 準備を…	2271	用意を…	423
○ 準備が…	758	用意が…	286
○ 準備に…	380	用意に…	33
○ 準備は…	304	用意は…	93
○ 準備も…	163	用意も…	49
○ 準備で…	59	用意で…	10
○ 準備と…	26	用意と…	3
○ 準備の…	22	用意の…	9
○ 準備から…	15		

ログダイス差

準備を... 118		用意を... 41	
準備	用意	準備	用意
コロケーション 頻度 MI LD	コロケーション 頻度 MI LD	コロケーション 頻度 MI LD	コロケーション 頻度 MI LD
準備を進める 203 9.66 8.86		準備を進める 8.86	
準備をする 18 9.06 7.24		準備をする 7.24	
準備を開始する 16 7.56 6.37		準備を開始する 6.37	
準備を済ませる 7 7.34 5.72		準備を済ませる 5.72	
準備をしまう 4 6.74 5.01		準備をしまう 5.01	
準備を重ねる 6 6.11 4.93		準備を重ねる 4.93	
準備を行う 71 5.44 4.85		準備を行う 4.85	
準備をまいる 6 5.74 4.68		準備をまいる 4.68	
準備をとく 2 6.6 4.33		準備をとく 4.33	
準備を急ぐ 3 5.22 4		準備を急ぐ 4	
準備をいく 17 5.35 4.64	用意をいく 1 3.69	準備をいく 0.68	用意をいく 3.96
準備を欠く 2 5.62 3.96		準備を欠く 3.96	

準備を... 118		用意を... 41	
準備	用意	準備	用意
コロケーション 頻度 MI LD	コロケーション 頻度 MI LD	コロケーション 頻度 MI LD	コロケーション 頻度 MI LD
用意を中する 1 14.52 6.24		用意を中する 6.24	
用意を督する 1 14.65 6.24		用意を督する 6.24	
用意をさせる 1 10.16 5.66		用意をさせる 5.66	
用意を書いてける 1 9.77 5.51		用意を書いてける 5.51	
用意を書かせる 1 7.71 4.31		用意を書かせる 4.31	
用意を希望する 1 6.93 3.69		用意を希望する 3.69	
用意を知らせる 1 6.89 3.66		用意を知らせる 3.66	
用意をする 1 6.04 2.91		用意をする 2.91	
用意を置む 1 5.82 2.72		用意を置む 2.72	
準備を忘れる 1 1.84 1.06		準備を忘れる 1.06	
用意を示す 3 5.09 2.09		用意を示す 2.09	
準備をなす 1 2.29 1.44		準備をなす 1.44	
準備を指示する 1 5.04 3.13		準備を指示する 3.13	
準備をしつかりす 1 4.96 3.1		準備をしつかりす 3.1	
用意をしつかりす 1 7.39 4.06		用意をしつかりす 4.06	

準備を見る 2 0.68 0.08
準備を命じる 7 6.95 5.52 用意を命じる 4
準備を考える 3 0.77 0.18
準備をかける 1 0.91 0.24
準備をとる 1 0.94 0.26
準備を下さる 3 1.12 0.52

BCCWJ

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

現代日本語書き言葉均衡コーパス 通常版

中納言 コーパス選択
コーパス検索アプリケーション

この画面になりたいコーパス名をクリックしてください。

コーパス名	場所	状態	選択サイト	番号
現代日本語書き言葉均衡コーパス 通常版	BCCWJ-NT	利用可能	関連アーカイブ	日本語学習者用
現代日本語書き言葉均衡コーパス 非 numTrans 版	BCCWJ-OT	利用可能	非 NumTrans (OT)	日本語学習者用
日本国歴史コーパス	GHJ	利用できません	利用申請	日本語学習者用
日本語翻訳者用コーパス	CSJ	利用可能	関連アーカイブ	日本語学習者用
多言語母語の日本語学習者傾斜コーパス	I-JAS	利用可能	関連アーカイブ	日本語学習者用
名大会話コーパス	名大会話コーパス	利用可能		日本語学習者用

パスワード変更 メールアドレス変更 ユーザ情報の変更 ニュース登録 収集履歴インポート

現代日本語書き言葉均衡コーパス (通常版) BCCWJ-NT

短単位検索 長単位検索 文字列検索

外国人|旅行|者|数 外国人旅行者|数

検索対象 指定を選択 検索対象を選択 検索対象をクリア 全て

検索動作 設定を戻す 文字列中の文区切り記号 # 文字列中の文区切り記号 # 前後文例の括弧数 20 検索実行

ダウンロードオプション 設定を表示する

キーの前
キー
キーの後

を
読む
せる

前方共起 キー 後方共起

前方共起1 キーから 1 語 キーと結合して表示

書字形出現形 が を 短単位の条件の追加

キー キーから 1 語 キーと結合して表示

語彙素 が 読む 短単位の条件の追加

後方共起1 キーから 1 語 キーと結合して表示

語彙素 が せる 短単位の条件の追加

「語彙素」とは何か

「つくる」「作る」「造る」のような異表記や

「作ら」「作り」「作れ」のような異なる活用形を一括して検索する。

<注意事項>

- 助詞や助動詞など漢字が当てられないもの以外は漢字を使って入力する。例 矢張り、成る、ある etc.
- 漢字が分からない場合は「Web茶まめ」を使って調べてください。
<http://unidic.ninjal.ac.jp/chamame/>
- 受身や使役は「れる」と「られる」、あるいは「せる」と「させる」のどちらも検索する必要があります。

を+動詞の使役形



接続助詞「から」の検索
短単位の条件の追加

短単位検索

長単位検索

短単位検索

前方共起条件の追加



区切り記号の設定

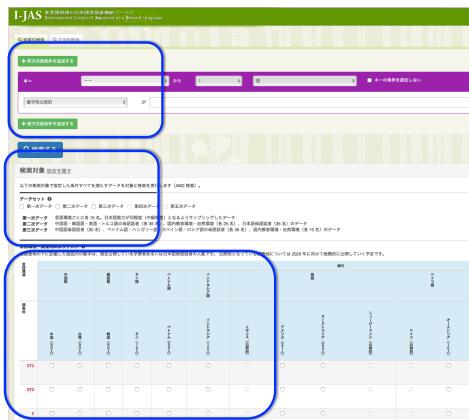
文脈の語数の設定



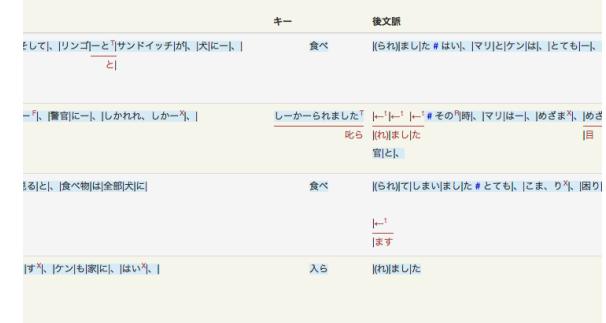
I-JAS

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

多言語母語の日本語学習者横断コーパス



受身形の検索



検索したデータをダウンロードしてみましょう。

The screenshot shows a table of search results with columns for ID, Name, and other details. At the bottom, there are download buttons for 'CSV' and 'Excel' formats.

ダウンロードしたデータ 解析向けに加工した文字列

H	I	J	K
前文脈(解析向けに加工した文字列)「H」に加工し、 <K>に、とびまし、あのー、飛びました#えーそして、リンゴとサンドイッチが、犬にー、	後文脈(解析向けに加工した文字列) 食べられました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、	食べられました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、	食べられました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、

<K>に、とびまし、あのー、飛びました#えーそして、リンゴとサンドイッチが、犬にー、

音声データと文字データ

The screenshot shows a search interface for selecting a corpus. It includes a search bar at the top and a list of corpora below, each with a name, type, and status. A red box highlights the 'I-JAS' entry in the list.

The screenshot shows the I-JAS course overview page. It features a user profile, a dashboard, and a course summary section. A red box highlights the 'I-JAS' link in the course summary.

発話通りの文字列

S	T	U
前文脈(発話通りの文字列)「S」に加工し、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、	後文脈(発話通りの文字列) 食べられました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、	食べられました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、
<K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、	食べられました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、	食べられました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、
<K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、	食べられました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、	食べられました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、 <K>に、とびまし、あのー、飛びました#[高い]マリとケンは、とてもー、あのーF]そして、リンゴ[ーT]サンドイッチが、犬にー、

<K>に、[とびまし=X]、[あのー=F]、飛びました#[えー=F]そして、リンゴ[ーと=T]

サンドイッチが、犬にー、

URL

◆国立国語研究所コーパスデータベース

<https://www.ninjal.ac.jp/database/>

◆BCCWJ

http://pj.ninjal.ac.jp/corpus_center/bccwj/

◆I-JAS

<http://lsaj.ninjal.ac.jp>

◆少納言

<http://www.kotonoha.gr.jp/shonagon/>

◆中納言

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

◆NINJAL-LWP for BCCWJ

<http://nlb.ninjal.ac.jp>

参考文献

- 李在鎬・石川慎一郎・砂川有里子(2018)『新日本語教育のためのコーパス調査入門』くろしお出版
- 砂川有里子・黒沢晶子(2017)「中国語を母語とする中級日本語学習者の発話に見られる日本語漢語名詞の使用状況—中国語の音素の影響を中心に—」『日中言語研究と日本語教育』10, 64-79.
- 砂川有里子(2014)「コーパスを活用した日本語教師のための類似表現調査法」『日本語/日本語教育研究』5, 7-27.
- 砂川有里子(2018)「中級以降で指導が必要なテシマウの用法について—学習者と母語話者の使用状況調査に基づく考察—」藤田保幸・山崎誠編『形式言語研究の現在』和泉書院 479-499.
- 砂川有里子(2018)「ストーリーテリングにおける順接表現の族語展開機能」庵功雄・石黒圭・丸山岳志編『時間の流れと文章の組立て—林言語学の再構築』ひじ書房 75-106.
- 砂川有里子・黒沢晶子(近刊)「フランス語と中国語を母語とする日本語学習者の漢語名詞の習得状況—自然会話にみられる発音の誤用分析」「フランス人日本語学習者の誤用から考える」ひじ書房