

## 『中納言』の概要

著者	大村 舞, 石本 祐一, 岡 照晃
URL	<a href="http://doi.org/10.15084/00003206">http://doi.org/10.15084/00003206</a>

# 『中納言』の概要

大村舞・石本祐一・岡照晃 (国立国語研究所 コーパス開発センター)



## コーパス検索システム『中納言』

中納言はコーパスを単語や品詞などの形態論情報で検索可能な検索システム (要ユーザ登録)

### 9つのコーパスについて検索可能

### 検索結果の例

検索語の品詞情報、検索語が現れる資料の情報を表示

【個別検索】ご利用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	種別	個別検索	包括検索	備考
現代日本語書き言葉均衡コーパス 中納言版	BCCWJ	✓	✓	従来より利用されているBCCWJのデータです(コーパスの紹介ページ)。こちらのページからBCCWJ/アニメーションデータをダウンロードできます。笑天版はこちら
国際研日本語ウェブコーパス 中納言版	NIJWC	✓	✓	笑天版はこちら
日本語話し言葉コーパス	CSI	✓	✓	コーパスの紹介ページ
日本語日常会話コーパス モニター公開版	CEJC	✓	✓	コーパスの紹介ページ
昭和話し言葉コーパス モニター公開版	SSC	✓	✓	コーパスの紹介ページ
戦後話し言葉コーパス	NUCC	✓	✓	コーパスの紹介ページ
現代研・職場談話コーパス	CWPC	✓	✓	コーパスの紹介ページ
日本歴史コーパス	CHU	✓	✓	コーパスの紹介ページ
日本語方言コーパス モニター公開版	COJADS	✓	✓	コーパスの紹介ページ
多言語母語の日本語学習者購読コーパス	I-JAS	✓	✓	I-JASに関する詳細な情報は「I-JAS 運用ガイド」をご参照ください。ファイルテキスト・音声・ビデオは「データ配布」からダウンロードできます。

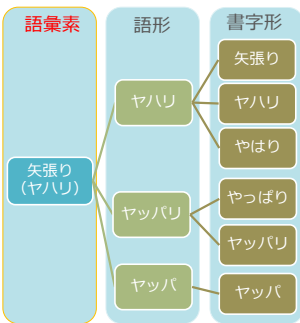
38,400件の検索結果が見つかりました。そのうち500件を表示しています。 表示している(NDC)の出力:国立国語研究所登録データ(https://iss.ndl.go.jp/information/npi/2019年4月3日現在)に基づき、『現代日本語書き言葉均衡コーパス』(以下登録)のNDCを再確認した。

検索対象語群: 124,100,954  
 記号・機能記号: 空白的な検索対象語群: 104,911,450

サンプルID	開始位置	検索語	品詞	後文	品詞	活用形	活用形	活用形	活用形	活用形	活用形	活用形	活用形	活用形	活用形	活用形	活用形	活用形	活用形
LBK2_00050	13140	7740	助詞	「さ」	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞
LBH7_00016	7800	4720	助詞	「と」	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞
PNAL_00012	1750	1110	助詞	「で」	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞	助詞

## 形態論情報

語彙素という単語の見出し語に付与されている情報



形態論情報の見出し語は図のような階層構造を持っている

語彙素「矢張り」の例

語彙素：辞書の見出しのレベル  
語形：異語形を区別するレベル  
書字形：異表記を区別するレベル

語彙素を使うことで表記の違いをまとめることができる

## 中納言の4つの検索方法

中納言では4種類の検索方法によって検索可能 (一部コーパスでは長単位検索はない)

短単位検索・長単位検索 形態論情報を使った詳細検索  
文字列検索 形態論情報を使わず出現形で検索  
位置検索 出現位置から(再)検索

細かく情報を指定して検索可

## 文字列検索

単語単位を気にせず完全一致する検索が可能

表記の違い (読みや書き方の違い) は考慮しないため その場合は短単位・長単位検索を利用する必要がある

## 位置検索

「ID」と「開始位置」でコーパス中の用例が一意的に決まるため、2つを保持していれば再現検索可能

## 短単位検索・長単位検索

形態論情報に基づいた検索ができる

短単位検索 短単位が一致する検索が可能

語彙素「日本」がヒットする → 「日本」「ニッポン」「にっぽん」「にほん」いずれも共通語彙素なのでヒット

長単位検索 長単位が一致する検索が可能

「日本語」「日本語研究」「日本語教育」などが長単位