

『中納言』の使い方



凡例

- 検索例① 検索例
- ☞ 参考情報
- ② 注意点

1. 3つの検索方法

(一番上のタブ)

- | | |
|-------|----------------|
| 短単位検索 | } 形態論情報を使った検索 |
| 長単位検索 | |
| 文字列検索 | } 形態論情報を使わない検索 |

3つの検索方法の使い分け

短単位検索

- 国名+「語」のような組み合わせ検索で「〇〇語」を一度に検索できる
- 「日本語」で検索してもヒットしない（短単位では「日本」+「語」）

長単位検索

- 「日本語」でヒットする
- ただし、「日本語研究」「日本語教育」を含めるには「日本語%」（「日本語」で始まるもの）とする必要がある

文字列検索

- 単位を気にせずに文字列で検索できる（たとえば「日本語の研究」など）
- あくまでも検索対象は表記なので、「日本」では「にっぽん」にヒットしない

2. 形態論情報を利用した検索

2.1. 形態論情報利用の長所

活用語の一括検索

検索例① 語彙素「読む」（終止形）
→ 「読ま」「読み」「読む」「読め」「読もう」（各活用形）

異表記の一括検索

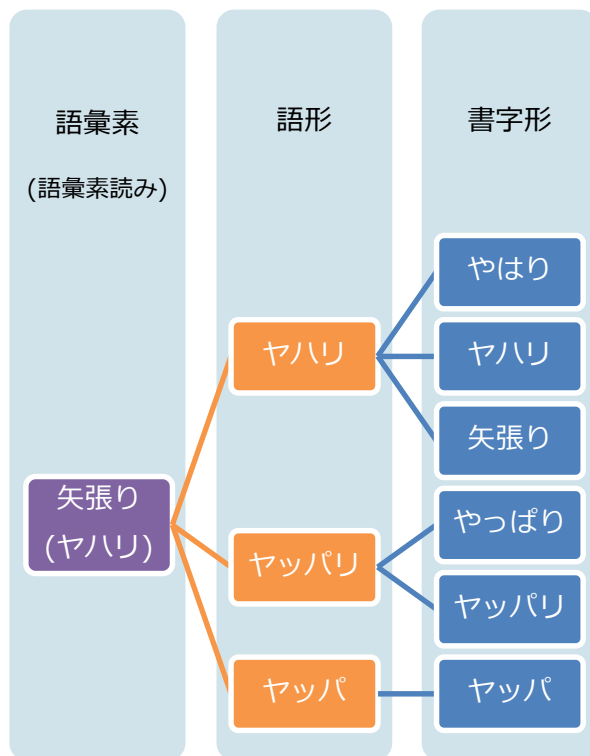
検索例② 語彙素読み「タマネギ」
→ 「玉葱」「玉ねぎ」「たまねぎ」「タマネギ」「玉ネギ」

異語形の一括検索

検索例③ 語彙素読み「コミュニケーション」
→ 「コミュニケーション」「コミニケーション」「コミニケーション」

2.2. 形態論情報の階層構造

BCCWJ の形態論情報の見出し語は次のような階層構造を持っている
(UniDic の見出し階層)



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 (見出し語の代表表記) が分からないときは「文字列検索」で検索して確認してみる (たまねぎ? タマネギ? 玉葱? → 「たまねぎ」で文字列検索、表示される語彙素「玉葱」を見て確認)

2.3. 検索語の条件指定

形態論情報を使った検索では、次の画面で検索条件を設定する

キー (--- ▼ 10 ▼ 語 ▼) キーを未指定

--選択-- ▼ が 短単位の条件の追加

「--選択--」とある部分で条件指定する属性 (「語彙素」「出現書字形」など) を選び、右の空欄でその中身を指定する

キー (--- 10 語) キーを未指定

語彙素 が 読む

ボタンで一つの単位について詳細な条件指定を追加できる
 検索例④ 語彙素「読む」 + 活用形 (の大分類) 「連体形」

キー (--- 10 語) キーを未指定

語彙素 が 読む

AND 活用形 の 大分類 が 連体形

(活用形など選択肢が決まっているものはドロップダウンメニューから選択する)

④ ここで追加される条件は AND 指定 (この画面上では OR 指定はできない)

❖ 誤った検索例: 語彙素「読む」 + 語彙素「書く」 →エラー

2.4. 複数単位の組み合わせ (共起・連接)

ボタンでキーの前方に出現する単位を指定

ボタンでキーの後方に出現する単位を指定

- 「キーから or 文頭から」「N 語 or N 語以内」のように、共起位置を指定可能

後方共起1 (キーから 1 語) キーと結合して表示

- 前方後方合わせて最大 10 個まで共起条件を追加できる

検索例⑤ 「言葉」を連体修飾する形容詞

キー = 品詞「形容詞」+ 活用形 (の大分類) 「連体形」

後方共起 1 (キーから 1 語) = 語彙素「言葉」

キー (--- 10 語) キーを未指定

品詞 の 大分類 が 形容詞

AND 活用形 の 大分類 が 連体形

後方共起1 (キーから 1 語) キーと結合して表示

語彙素 が 言葉

☞ 集計にしたいもの (この場合は形容詞) をキーの位置に持ってくるのがコツ

条件指定を入力したあとでも ボタンで、キーの位置を移動できる。

前方共起1 (キーから 1 語) キーと結合して表示 この条件をキーに この共起条件を削除

品詞 の 大分類 が 形容詞

AND 活用形 の 大分類 が 連体形 短単位の条件の追加

キー (--- 10 語) キーを未指定

語彙素 が 言葉 短単位の条件の追加

- ⑨ 係り受け情報はアノテーションされていないので、離れた場所にあるものを修飾している例は取得できないし、直後に来ても必ず修飾しているとは限らない
- ⑩ 短単位と長単位を組み合わせた検索はできない

2.5. ワイルドカード

語彙素などの検索指定では通常 of 文字の代わりに、次のワイルドカードが利用できる。

- % 任意の文字列 0文字以上の文字列、何でも良い
- _ 任意の1文字 なにか1文字、何でも良い
- [abc] 文字クラス 括弧内の文字 (abc) いずれか一文字
- [a-c] a~c の文字いずれか一文字
- [^abc] 文字クラスの補集合 abc 以外のいずれか一文字

検索例⑥ 飛び% 「飛び」ではじまる「飛び込む」「飛び入り」etc.

検索例⑦ [光風] 景 「光景」または「風景」

☞ _% で1文字以上の任意の文字列を表せる

2.6. 特殊な検索指定

キーを指定しない検索

前方・後方共起が指定されている場合のキーに限り、「キーを未指定」をチェックすることにより、条件指定なしで検索することができる。

キー (--- 10 語) キーを未指定

---選択--- が 短単位の条件の追加

KWIC 中のキーの範囲変更

前方・後方共起が指定されている場合、キーに隣接する共起条件の「キーと結合して表示」をチェックすることで、KWIC のキーに隣接する前後条件を含めることができる。

キー (--- 10 語) キーを未指定

品詞 の 中分類 が 名詞-固有名詞 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 この条件をキーに この共起条件を削除

語彙素 が 的 短単位の条件の追加

↓

PB15_00078	の貴重な資料となるだろう。そのなかでも目立つのは、言語学モデルだ	デリダ的	脱構築の結合(一時)のピーター・アイゼンマンに代表されるゆから、コンピュータ	デリダ	デリダ	Derrida	名詞-固有名詞-人名-一般		出版・書籍	磯崎 新(著)	建築と時間	磯崎新, 土居義岳 著	岩波書店	2001
PB32_00246	的なものが見直される時代が東の間に来したが、敗戦を迎えるとき、再び	日本的	な伝統をいなくなり捨て、経済大国の道にむかひたすら突っ走って来た。そしてリアルが	ニッポン	日本		名詞-固有名詞-地名-国		出版・書籍	延原 啓(著) 澤田 和子(著)	いま、街から伝えたい	澤田和子述	街から舎	2003
LBh3_00023	あった。それはアメリカ経済の強さを表すがけではなく、アメリカ文化、	アメリカ的	生活のシンボルでもあった。その後すぐにアルフレッド・スローン(車)いる(GM)が	アメリカ	アメリカ	America	名詞-固有名詞-地名-国		図書館・書籍	佐伯 啓思(著)	「欲望」と資本主義	佐伯啓思著	講談社	1993

2.7. 検索にあたっての注意点

- ⑧ 固有名詞の「語彙素」はカタカナ（固有名詞-一般、国名は除く）
- ⑨ 「語彙素」の指定だけでは一意に決まらない場合がある（例：市場→ いちば・しじょう）
- ⑩ 語彙素レベルで例外なく一意にするには 語彙素・語彙素読み・品詞・語彙素細分類 の4つを指定する必要がある
- ⑪ 可能動詞は語形レベルで定義されている（「読める」の語彙素は「読む」）

2.8. その他の検索例

- 検索例⑧ 可能動詞「書ける」
- 検索例⑨ 「大きな」+名詞（※コアのみ対象）
- 検索例⑩ 「綺麗な」+名詞（※コアのみ対象）

☞ どう指定したらいいかわからないときは「文字列検索」で該当する例を検索して確認してみる（用例のサンプル ID をクリックすると前後の単位にどのような形態論情報が付与されているか確認できる）

サンプル ID	前文脈	キー	後文脈	語彙素読み
LBm9_00083	濃いもみあ けと、信頼を 呼びそうば	大きな	顔の男だっ た。ほれいに ブラシ	オキ

大きな 検索 検索結果をダウンロード

使用できるワイルドカード以下の通りです
% : 0回以上の任意の
_ : 任意の1文字(例
[] : 文字クラス(例日
[] : 否定文字クラス(例
【検索動作】設定を隠す
文脈中の区切り記号
【列の表示】設定を隠す
 形態論情報
 サンプルID 連
 語種 出現発音
 コーパス情報
 サブコーパス名
 山崎情報

LBm9_00083	9180	呼び	ヨブ	呼ぶ	動詞一般	五段- バ行	連用形- 一般		和	ヨビ	流通・書 籍		1	呼び	1443.0
LBm9_00083	9190	そう	ソウ	そう	形動詞-助動詞 語幹			様態	和	ソー	流通・書 籍		1	そう	1445.0
LBm9_00083	9200	な	ダ	だ	助動詞	助動 詞-ダ	連体形- 一般		和	ナ	流通・書 籍		1	な	1447.0
LBm9_00083	9210	大きな	オオキ ナ	大きな	連体詞				和	オオキ ナ	流通・書 籍		1	大きな	1448.0
LBm9_00083	9220	顔	カオ	顔	名詞-普通名詞- 一般				和	カオ	流通・書 籍		1	顔	1451.0
LBm9_00083	9230	の	ノ	の	助詞-格助詞				和	ノ	流通・書 籍		1	の	1452.0
LBm9_00083	9240	男	オトコ	男	名詞-普通名詞- 一般				和	オトコ	流通・書 籍		1	男	1453.0

3. 検索条件式

検索画面で指定した検索条件は、「検索条件式」として履歴に記録される（「履歴で検索」で再検索可能）

検索例⑤の検索条件式：

```
キー: (品詞 LIKE "形容詞%" AND 活用形 LIKE "連体形%") AND 後方共起: 語彙素 = "言葉" ON 1 WORDS FROM キー WITH OPTIONS unit="1" AND tglWords="20" AND limitToSelfSentence="0" AND endOfLine="CRLF" AND tglKugiri="|" AND encoding="UTF-8" AND tglFixVariable="2"
```

- 「検索条件式」を使うことで、中納言ユーザーなら誰でも、同じ検索を行うことができる

短単位検索

検索フォームで検索 **検索条件式で検索** 履歴で検索

検索条件式

```
キー: (品詞 LIKE "形容詞%" AND 活用形 LIKE "連体形%") AND 後方共起: 語彙素 = "言葉" ON 1 WORDS FROM キー WITH OPTIONS unit="1" AND tglWords="10" AND tglKugiri="|" AND tglFixVariable="2"
```

- 研究の再現性のために論文などで使用した検索条件式を明記するとよい

☞ 検索条件式を修正することで OR 指定も可能

検索例⑤の修正版（「言葉」または「言語」）：

```
キー: (品詞 LIKE "形容詞%" AND 活用形 LIKE "連体形%") AND 後方共起: ( 語彙素 = "言葉" OR 語彙素 = "言語" ) ON 1 WORDS FROM キー-WITH OPTIONS unit="1" AND tglWords="20" AND limitToSelfSentence="0" AND endOfLine="CRLF" AND tglKugiri="|" AND encoding="UTF-8" AND tglFixVariable="2"
```

☞ 複数の検索条件式を並べて検索すると、一括して検索・ダウンロードすることができる（この際、検索結果は zip 圧縮される）

4. 検索結果

4.1. 検索結果の表示

検索結果は、見つかった件数と、500 件までの該当例が画面上に表示される。

2041 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。

☐ テーブルの幅を固定 短

サンプル ID	前文脈	キー	後文脈	語彙素読み	語彙素	語彙素細分類	品詞	活用型	活用形	レジスタ	執筆者	書名/出典	編著者等	出版者	出版年
---------	-----	----	-----	-------	-----	--------	----	-----	-----	------	-----	-------	------	-----	-----

500 件以上ある場合には表示は 500 件まで。並び順は無秩序。

4.2. 検索結果の表示項目

前後文脈 (KWIC)

前文脈 ・ キー ・ 後文脈

形態論情報

語彙素読み ・ 語彙素 ・ 語彙素細分類

語形 ・ 品詞 ・ 活用型 ・ 活用形 ・ 書字形 ・ 発音形出現形 ・ 語種 ・ 原文文字列

コーパス情報

サンプル ID ・ 連番 ・ サブコーパス名 ・ コア ・ 固定長 ・ 可変長

出典情報

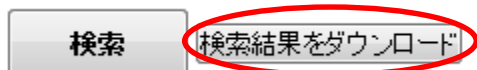
執筆者 ・ 生年代 ・ 性別 ・ ジャンル ・ 書名/出典 ・ 副題/分類 ・ 巻号 ・ 編著者等 ・ 出版者 ・ 出版年

画面上の「【列の表示】」でチェックボックスをオン・オフすることで、リアルタイムで表示を切り替えられる

ダウンロード時には以上の全ての項目に「反転前文脈」を加えたデータがダウンロードされる（【列の表示】とは無関係）

4.3. 検索結果のダウンロード

中納言自身には集計機能はないので、検索結果をダウンロードして利用する



- 検索画面では 500 例までしか表示されないが、ダウンロード時には最大 10 万件まで一度にダウンロードできる
- 検索画面の【ダウンロードオプション】で、文字コード等を指定できる

【ダウンロードオプション】設定を隠す

システム 文字コード 改行コード 出力ファイルが一つの場合はZIP圧縮を行わない

- 使っているパソコンに合わせて自動選択されるが、自分で変更することも可能
- そのまま Excel に読み込ませる場合はシステム「Excel(Windows)」が便利

4.4. 検索結果に関する注意

- ⑨ 共起検索で同じレコードが複数表示されることがある
 - たとえば、前方共起条件がキーから 3 語以内にある名詞である場合、同一のキーから 3 語以内に 2 つの名詞があれば、同じ ID の行が 2 度表示される
- ⑨ 並び順は無秩序
 - ランダム性は保証されない。ランダムサンプリングする際は全件取得後に別途並び替えを
- ⑨ 検索結果が 10 万件以上ある場合、10 万件で切れるので注意

5. Excel による集計（ピボットテーブル）

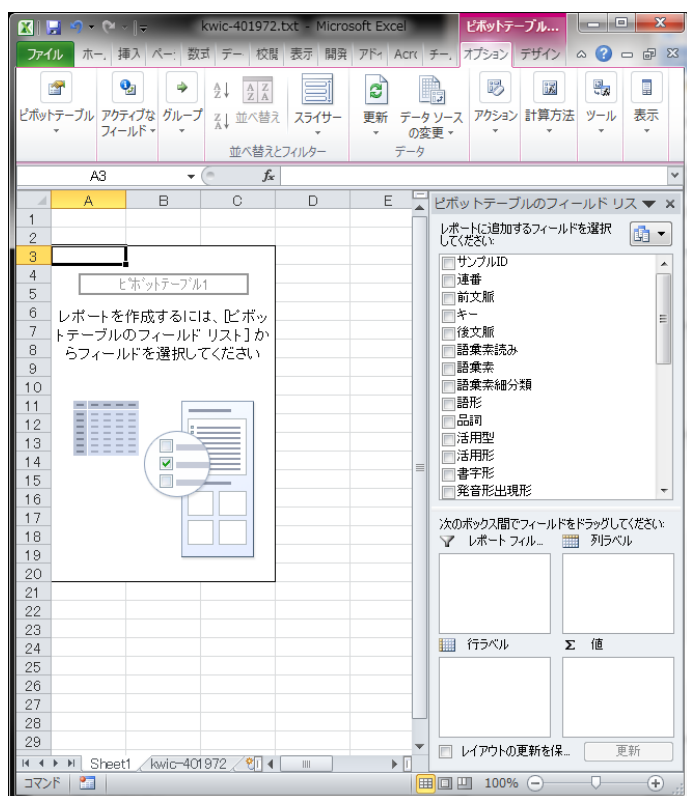
5.1. Excel での集計の流れ

1. 検索結果をダウンロード・解凍
 - ⑨ 複数検索時など、検索結果ファイルが zip 圧縮されている場合はダウンロード完了後、自分で解凍してテキストファイルを取り出しておく
2. 検索結果のインポート
 - ⑨ 「ファイルを開く」ダイアログでファイルの種類を「テキストファイル」または「全てのファイル」にして選択する
 - ⑨ ダウンロードオプションでシステム「Excel(Windows)」にしておけばドラッグアンドドロップで開ける
3. ピボットテーブルの挿入
4. ピボットテーブルの作成・集計
5. ピボットグラフの作成

5.2. ピボットテーブル

ユーザーの指示により動的にクロス集計表を作る機能

Excel2007以降では「挿入」タブ左端のボタンをクリック
(バージョンによって大きく違う)



ピボットテーブルの作り方

1. 作りたい集計表の形をイメージ

	たまねぎ	玉ねぎ	玉葱	...
書籍	10	2	8	
新聞	10	0	0	
ブログ	5	4	2	
:				

表記バリエーション
=キーの文字列

出現数
=キーの個数

テキストの種類
=レジスター

2. イメージに合わせて行ラベル/列ラベルをドラッグアンドドロップで指定 (ドロップすべき場所はイメージした表と位置関係が同じ)

次のボックス間でフィールドをドラッグしてください

レポートフィル... 列ラベル

キー

行ラベル Σ 値

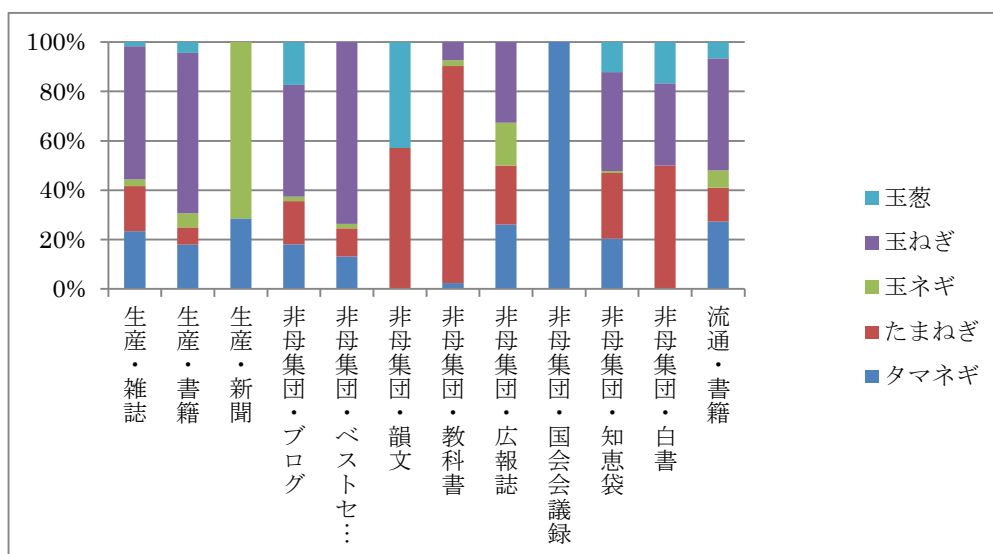
レジスター データの個数 ...

レイアウトの更新を保... 更新

3. これだけで自動的に集計表ができる

	A	B	C	D	E	F	G
1							
2							
3	データの個数 / キー	列ラベル					
4	行ラベル	たまねぎ	たまねぎ	玉ねぎ	玉ねぎ	玉葱	総計
5	生産・雑誌	42	33	5	97	3	180
6	生産・書籍	93	36	30	336	23	518
7	生産・新聞	2		5			7
8	非母集団・ブログ	90	87	9	225	86	497
9	非母集団・ベストセラー	7	6	1	39		53
10	非母集団・韻文		4			3	7
11	非母集団・教科書	1	36	1	3		41
12	非母集団・広報誌	12	11	8	15		46
13	非母集団・国会会議録	23					23
14	非母集団・知恵袋	47	61	2	92	28	230
15	非母集団・白書		6		4	2	12
16	流通・書籍	129	65	33	215	31	473
17	総計	446	345	94	1026	176	2087
18							

4. 「ピボットグラフ」ボタンでグラフ化



(割合を見るため 100%積み上げ棒グラフで表示)

5.3. 二つ以上の検索結果をまとめる

- 2回以上の検索結果を扱う場合、一つの表にまとめておくと集計が容易になる
 - それぞれの検索結果をコピーし、一つの表の下に続けて貼っていく
貼り付ける際、余分な列名の行（ヘッダー）は削除する
 - ピボットテーブルを作り語彙素を使って集計

表の結合イメージ



- 集計に必要な列を自分で追加しておけばその列も集計に使える
既存の列（語彙素など）ではどの検索結果か区別できない場合は、2つを区別するための列をあらかじめ追加しておく。

列の追加イメージ →

