

『中納言』概説 CEJC

2019年3月3日

ログイン画面 <https://chunagon.ninjal.ac.jp/auth/login>

コーパス選択画面 （どれか1つを選ぶ）

コーパス名	略称	状態	関連サイト	備考
現代日本語書き言葉均衡コーパス 通常版	BCCWJ-NT	利用可能	関連データ配布	従来より利用していた BCCWJ のデータです。コーパスの紹介ページへのリンク
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 版 (OT) とは
日本語歴史コーパス	CHJ	利用可能		
日本語話し言葉コーパス	CSJ	利用可能		
多言語母語の日本語学習者横断コーパス	I-JAS	利用可能	関連データ配布	発話の音声ファイルや書き起こしたトランスクリプトは、関連データ配布サイトからダウンロードできます。その他の資料については、I-JAS 関連資料を参照ください。
名大会話コーパス	名大会話コーパス	利用可能		
現日研・職場談話コーパス	職場コーパス	利用可能		
日本語日常会話コーパス モニター公開版	CEJC	利用可能		こちらのサイトの利用ガイドラインをご一読の上ご利用ください。

検索画面（初期画面）

1. 四つの検索方法

（一番上の赤で囲んだタブ）

- | | | |
|-------|---|------------------------------|
| 短単位検索 | } | 形態論情報を使った検索 ※今回主に解説するのは短単位検索 |
| 長単位検索 | | |
| 文字列検索 | } | 形態論情報を使わない検索 |
| 位置検索 | } | 語の出現位置を指定する検索 |

検索方法の使い分け

短単位検索

- 名詞＋「的」のような組み合わせ検索で「〇〇的」を一度に検索できる
「基本的」で検索してもヒットしない（短単位では「基本」＋「的」）

長単位検索

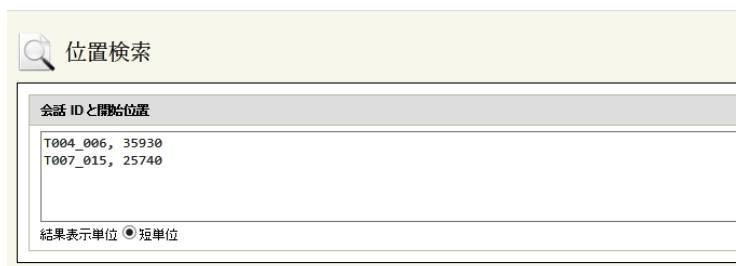
- 「基本的」でヒットする
ただし、「基本的人権」に含まれる「基本的」はヒットしない。

文字列検索

- 単位を気にせずに任意の文字列で検索できる（たとえば「と言った」など）
- ただし、自分が検索したいものと一致するとは限らない
例：「プリン」で検索すると「サンプリング」「チャップリン」などもヒットする。
- 語ではなく、**表記**を検索するので「おこなう」で「行う」は検索できない。
- 語彙素が分からないときに利用すると効果的。

位置検索

- [会話 ID],[語の開始位置]の指定で検索
- 複数の同時検索が可能（ひと組ずつの改行が必要）



位置検索

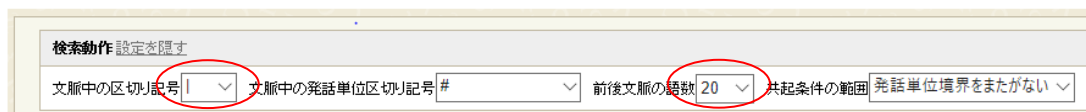
会話 ID と開始位置

T004_006, 35930
T007_015, 25740

結果表示単位 ☒ 短単位

2. 検索の前にやること（検索動作の指定）

検索動作の指定



検索動作 設定を隠す

文脈中の区切り記号 文脈中の発話単位区切り記号 # 前後文脈の語数 20 共起条件の範囲 発話単位境界をまたがない

発話単位境界をまたがない／またぐ：#で表示される部分をまたいで検索するかどうか。

3. 形態論情報を利用した検索

今回は短単位検索を例に説明するが、長単位でも方法は同じ。

3.1. 形態論情報利用の長所

活用語の一括検索

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

異語形の一括検索

②検索例：語彙素読み「ヤハリ」

→ 「やはり」「やっぱり」「やっぱ」「やっぱし」

異表記の一括検索

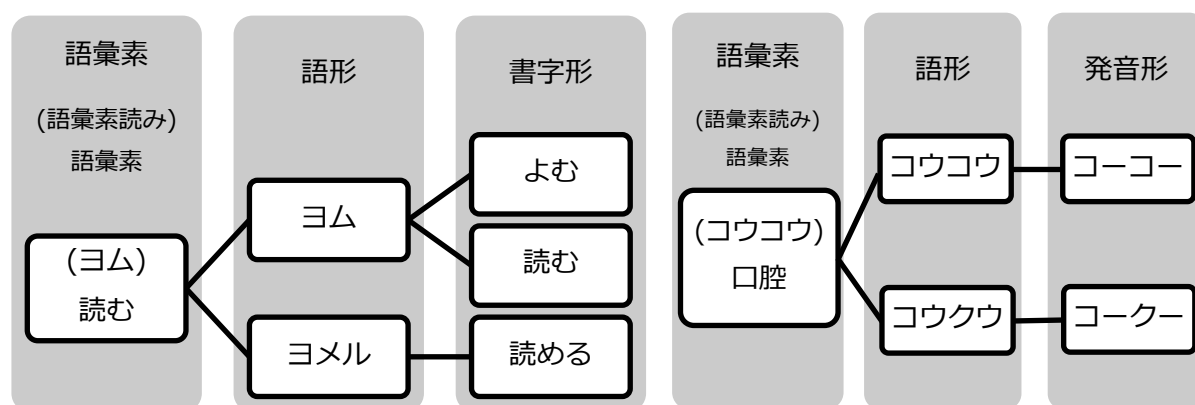
(BCCWJ の場合有効。CEJC や CSJ では異表記は原則ないため関係しない。名大会話・職場コーパスも異表記はあまりない。)

③検索例：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」

(CJEC はすべて「玉葱」，CSJ はすべて「玉ネギ」，名大会話はすべて「タマネギ」)

3.2. 形態論情報の階層構造



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 (CSJ・名大会話ではあまり関係しない)

発音形：語形の下にあり，語形を表音的に表す

☞ 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。

（「やはり」？「やっぱり」？）→（「やはり」で検索，表示される語彙素は「矢張り」）

3.3. 検索語の詳細な条件指定

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。

条件を指定する属性はデフォルトで「書字形出現形」になっている。
 検索したい属性（「語彙素」「品詞」など）をプルダウンメニューから選び、右の空欄にその中身を入力する。

 短単位条件の追加 ボタンで1つの単位について詳細な条件指定を追加できる。

④検索例：語彙素「読む」 + 活用形（の大分類）「連体形」

活用形など選択肢が決まっている属性はドロップダウンメニューから選択できる。

⑤ 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」

→用例数 0 件（但し、検索エラーにはならない）

☞ 短単位条件の追加で追加した条件を削除するには、検索ボタンの右にある「条件クリア」をクリック。

3.4. 複数単位の組み合わせ（共起・連接）

 ボタンで前方に出現する単位を指定。

 ボタンで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  

- 前方と後方とで合わせて最大 10 個まで共起条件を追加できる。

⑤検索例：「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 -- 1 語 ☐ キーの条件を指定しない

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

検索条件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 

品詞 の 大分類 が 形容詞 

AND 活用形 の 大分類 が 連体形  

キー -- 1 語 ☐ キーの条件を指定しない

語彙素 が 言葉  



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。

⑥検索例：修飾表現「○○な言葉」→

「品詞：形状詞」「書字形出現形：な」＋「語彙素：言葉」

前方共起条件の追加

① キー 1 語 キーの条件を指定しない

品詞 の 大分類 が 形状詞 短単位条件の追加

② 後方共起1 キーから 1 語 キーと結合して表示

書字形出現形 が な 短単位条件の追加

③ 後方共起2 キーから 2 語 キーと結合して表示

語彙素 が 言葉 短単位条件の追加

後方共起条件の追加

3.5. ワイルドカード

語彙素などの検索指定では次のワイルドカード（記号部分は半角）も利用できる。

- % 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（一部利用に制限がある）
- [abc] 文字クラス… 括弧内の文字いずれか一文字。
- [^abc] 文字クラスの補集合… 括弧内の文字以外のいずれか一文字。

⑦検索例：心% 「心」ではじまる「心がけ」「心地」etc.

⑧検索例：わたし[はがも] 「わたしは」または「わたしが」または「わたしも」



174 件の検索結果が見つかりました。
検索対象話数: 611,309 記号・補助記号・空白を除いた検索対象話数: 610,959

会話 ID	開始位置	連番	前文脈	キー	後文脈	語彙素読み	語彙素	語彙素細分類	語形	品詞	活用型	活用形	会話概要	会話時間	話者数	話者間の関係性	形式
T016_004b	70600	39550	だけ # お酒と # うん # うーん # ふーん # ちょっとだけ酒油入れたいんだね # なん # か # あ # わたし	は	入れない # おいしい # これおいしいね # これ # はい # この汁め # もうめっちゃ # 早よ	ハ	は		ハ	助詞-係助詞			自宅で妻・息子2人・義母と夕食を食べながら	22	5	家族	雑談
C001_001	120760	66500	あのシャビーの感じの # シャビーのね # うーん # うん # これ早く欲しいらしい # え # わたし	も	欲しい # あー # 重い # あの えーと ね 重い # あの わたし 包んだ の で ちと 辛い なんです # えー	モ	も		モ	助詞-係助詞			飲食店で友人4人と誕生日会	31	5	友人知人	雑談
T009_010	38140	20370	乗らなくするの # うん # うん # おい # うん # うん # うん # あ # やっぱそうなんだ # いつもわたし	も	蒸し焼きやっ て た ね # あの て け 置 は するんだ けど # うん # うん # うん # お 水 は	モ	も		モ	助詞-係助詞			自宅で両親・祖母と夕食を食べながら	17	4	家族	雑談
T003_004	51190	29250	の 朝から 電話 や けて きて OG の 予定 を 先に 聞 か ない に と を ます わ た し	は	怒られた わ け # あー # 勝手に に 二十 七 と 決 め や が っ て に の 野郎 っ て ゆ う 話 し に	ハ	は		ハ	助詞-係助詞			子供の幼稚園時代のママ友2人とバザーの打合せ	28	3	友人知人	用談相談
T007_017	67800	39130	は どう し た の # 持 っ て 帰 っ た の リン グ は だ っ て 結局 も ら わ な っ た に だ す # わ た し	が	同 本 も 持 っ て る の で そ の ー 本 を 今 け っ て 置 っ て ま す # も ら わ な っ た の # 彼	ガ	が		ガ	助詞-格助詞			公共施設で地域活動の会の仲間3人と役員会	29	4	友人知人	会議会合
K004_017	46900	26430	ね # こう けん は # けん ちゃん ト マ ト いく っ つ 食 べ た # ー # ぜ っ ケ ビ ン 吉 野 な だ は # よ し # わ た し	が	四 つ # 悠 次 郎 は 三 つ な # よ う い # う # い や # さ ら に ま た 余 っ て し ま っ た # あ と も う ー 個	ガ	が		ガ	助詞-格助詞			自宅で息子3人と夕食を食べながら	18	4	家族	雑談
T006_001	12290	7270	涼しくなっ た り と か # うん うん # うん # うん # あー # それ は 楽 だ か も し ん だ い # はい # あ の わ た し	も	ね あ の 夕 方 ぐ ら い あ の 大 体 ね 四 時 と け り 五 時 ぐ ら い に 可 能 と それ まで ほと ん ど	モ	も		モ	助詞-係助詞			大学の教員室でサークル顧問の先生と雑談	28	2	先生生徒	雑談

3.6. 特殊な検索指定

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キーに隣接する前後条件を含めることができる。

- ⑨検索例： キー = 語彙素「的」
 前方 = 品詞「固有名詞」 （キーと結合して表示）

前方共起条件の追加

前方共起1 ☒ キーと結合して表示

の が

キー ☐ キーの条件を指定しない

が

後方共起条件の追加

4 件の検索結果が見つかりました。

検索対象語数: 611,309 記号・補助記号・空白を除いた検索対象語数: 610,959

会話 ID	開始位置	連番	前文脈	キー	後文脈	語彙素読み	語彙素	語彙素細分類	語形	品詞	活用型	活用形	会話概要	会話時間	話者数	話者間の関係性	形式
T016_006a	24670	14410	みたいな人々がちょっとやめちよってゆづのを見て喜ぶみたいなさ #	平野的	なとこがあるよね # # # 駄目 # 駄目でしょ # 中央局に入れるかもしれない # 佳大に	テキ	的		テキ	接尾辞 -形状詞 的			自宅で妻と食事・晩劇をしながら	33	2	家族	雑談
T002_020	83930	47560	ばんばん発生しててみたいな # ふーん # ああ # そうなんですね # うーん # # # # # # 今の	中国的	な # あー # そうそうそうそう # イメージかな # そんな感じ # あんな感じです # これだいいじぶなの	テキ	的		テキ	接尾辞 -形状詞 的			公園で友人2人とお茶を点てながら	51	3	友人知人	雑談
T006_009	35980	21080	あれは下明じゃないよ # あれはねオオ東京だから # うん # ネオ東京 # ネオ東京 # あの	アキラ 的	な感じだから # あー # 第三 # 第三新東京市 # うん # うん # あれあれでしょ # あそこそこだ	テキ	的		テキ	接尾辞 -形状詞 的			飲食店で中学時代の同級生と食事しながら	14	2	友人知人	雑談
T015_006	26250	14600	バスクリンで子供のころしか知らない # うん # ある # うち入れないからわかんないけど # ババ	バブ的	な # オレンジ色の粉でしょ # 今なんかいりんなのありますよ # そうだよ # 名湯なんたら	テキ	的		テキ	接尾辞 -形状詞 的			経営する会社の従業員4人と飲食店で送別会	24	5	同僚	雑談

キーを未指定にすることで広く「〇〇的」を検索することも可能。

- ⑩検索例： キー = 未指定
 後方共起 1 = 語彙素「的」 （キーと結合して表示）

前方共起条件の追加

キー ☒ キーの条件を指定しない

が

後方共起1 ☒ キーと結合して表示

が

後方共起条件の追加



下記に☒を入れると、「発音形出現形」「発音形」が確認できる。なお、これらは検索の前後どちらでもオンオフの切り替えが可能である。また、ダウンロードする際は、☒の有無は関係なく、すべてデータに含まれる。

列の表示  設定を隠す

<input checked="" type="checkbox"/> コーパス情報	<input type="checkbox"/> 形態論情報
<input checked="" type="checkbox"/> 会話 ID <input checked="" type="checkbox"/> 開始位置 <input checked="" type="checkbox"/> 連番	<input checked="" type="checkbox"/> 前文脈 <input checked="" type="checkbox"/> キー <input checked="" type="checkbox"/> 後文脈 <input checked="" type="checkbox"/> 語彙素読み <input checked="" type="checkbox"/> 語彙素 <input checked="" type="checkbox"/> 語彙素細分類 <input checked="" type="checkbox"/> 語形 <input checked="" type="checkbox"/> 品詞 <input checked="" type="checkbox"/> 活用型 <input checked="" type="checkbox"/> 活用形 <input type="checkbox"/> 書字形 <input checked="" type="checkbox"/> 発音形出現形 <input checked="" type="checkbox"/> 発音 <input type="checkbox"/> 語種 <input type="checkbox"/> 仮名 <input type="checkbox"/> タグ付き書字形
<input type="checkbox"/> 会話情報	<input type="checkbox"/> 話者情報
<input type="checkbox"/> セッション ID <input type="checkbox"/> 会話概要 <input type="checkbox"/> 会話時間 <input type="checkbox"/> 話者数 <input type="checkbox"/> 話者間の関係性 <input type="checkbox"/> 形式 <input type="checkbox"/> 場所 <input type="checkbox"/> 活動	<input type="checkbox"/> 話者 ID <input type="checkbox"/> 話者ラベル <input type="checkbox"/> 年齢 <input type="checkbox"/> 性別 <input type="checkbox"/> 出生地 <input type="checkbox"/> 居住地 <input type="checkbox"/> 職業 <input type="checkbox"/> 協力者からみた関係性
	<input type="checkbox"/> その他
	<input type="checkbox"/> permalink



C001_001	123300	67920	さ # えらッピンぐねえのかよ と思って # うーん # あ # え # これ何 # じゃあ美沙ちゃん セレクト # あ #	わたし	わたしらッピンぐ # いいや # 嬉しい # やーん # じゃぎれ いに # 美沙ちゃん # 汚い けど # ぎれいに # かawaii いじゃん # なんかもったい ン	ワタン	私	ワタン	代名 詞	ワタン	ワタ シー	飲食店で友人 4人と誕生日 会	31	5	友人知人	雑談	C			
C001_001	62740	35800	ね # もう # へー # うーん # でしょう # なんかえ可奈子 ちゃんカーボンロードとかず るの # うーん # え # しない #	わたし	食事制限なんにもしない # あ # すこいね # うーん # えー # え # なんかもスト イックじゃないもん # なんだ	ワタン	私	ワタン	代名 詞	ワタン	ワタ シー	飲食店で友人 4人と誕生日 会	31	5	友人知人	雑談	C			
K003_012a	920	480	ぶりやないですか # うーん # やばい # ね # 何年ぶり # 結 構来てないよね # あ # でも	わたし	半年ぶりくらいかも # そうゆ うことやっちゃうんだ # うん # ゆめちゃうまい # 熱い # 熱 い # ボーション # あー # うまい	ワタン	私	ワタン	代名 詞	ワタン	ワタ シー	飲食店で高校 時代の友人と 食事をしてなが ら	12	2	友人知人	雑談	K			
C001_001	131670	72200	片せるじゃん # すこい # こ れでそう # ね # ちょっとね置 き場所今どうして今思っ た # やと # うーん #	わたし	もそれ見た時にねどこに置 くのかしらと思ったの # で もなんか	ワタン	私	ワタン	代名 詞	ワタン	ワタン	飲食店で友人 4人と誕生日 会	31	5	友人知人	雑談	C			
T013_008	18510	11330	たちはねあのソリヤ兄ちゃん ちにうちに待ってたて # う ん # うん # うん # それで	わたし	とねお父さんを浅間さんへ 残して # あー # ほしいでわフ タ人で浅間さん	ワタン	私	ワタン	代名 詞	ワタン	ワッタ ン	妻の実家で義 母・義弟の妻 と朝食後の雑 談	8	3	家族・親 戚	雑談	T			
T004_005b	37190	21470	あそこイタリアン # うん # ブドウあるある # うん # ブド ウ # あそこ結構するでしょ # そうでもない # ワイン # う ん #	わたし	お昼しか行かないからな # あそこ安いよ # あそこ # あ すこのシュウこうのナカシ ンのあすこだ	ワタン	私	ワタン	代名 詞	ワタン	ワッタ ン	町会役員の自 宅にて役員3 人と会計監査 後の慰労会	24	4	友人知人	雑談	T			
T011_007	57190	33150	ないと思うけど # ううん # 部 長さえならなければいいかな とか思ってた # うーん # うー ん # うん #	わたし	も部長は文化だからできた な # ね # やっぱ一回やっ た委員だしさ # そんな	ワタン	私	ワタン	代名 詞	ワタン	ワッヒ	飲食店で長男 の学校のママ 友2人と雑談	38	3	友人知人	雑談	T			
T013_008	30060	18430	で # あー # すこいじゃん # 大昔まで # へー # ほしいだ けどねそんなうちらであんな ミ身分が違うから	わたし	も行かぬなしさ # うん # うーん # あんたそれから今 度工場で会うのがうんとこう 間	ワタン	私	ワタン	代名 詞	ワタン	ワータ ン	妻の実家で義 母・義弟の妻 と朝食後の雑 談	8	3	家族・親 戚	雑談	T			
T009_020	89870	52320	二十分ぐらに三鷹駅を出 れば全然いいよね # いや # いいよ # うん # 前後ろで背 負う #	わたし	がこう抱えれば # いや # う ん # でも重いですよね # う ん # うん # だいいじぶ # あ の # 全然だいいじぶ # ね #	ワタン	私	ワタン	代名 詞	ワタン	ワータ ン	上京した父方 の祖父母と両 親・母方の祖 母と自宅で団 楽	22	6	家族・親 戚	雑談	T			
K004_017	70280	39270	が結局きょうやなかったな # 何それ # それはくそだな # なんつった # 今 # 惣次郎く ん #	わたし	の作戦はくそと言ったな # なんその作戦 # だから要 はスマホとかを	ワタン	私	ワタン	代名 詞	ワタン	ワータ ン	自宅で息子3 人と昼食を食 べながら	18	4	家族	雑談	K			
会話 ID	開始位 置	連番	前文脈	十一	後文脈	語彙素 読み	語彙素 素	語彙素 細分類	語形	品詞	活用 型	活用 形	発音形 出現形	発音	会話概要	会話 時間	話者 数	話者間の 関係性	形式	話者

⑨ 語彙素だけでは一意にならない（辛い→ つらい・からい）。

④ 可能動詞は（短単位では）語形レベルで定義されている（「読める」の語彙素は「読む」だが、語形は「ヨメル」）。

▼ 前方共起条件の追加

☒ キー

☐ キーの条件を指定しない

▲ 後方共起条件の追加

☞ どのように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）ただし、「語形」は表示されないので注意。

4. 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。検索履歴 ID にチェックを入れ、「編集して検索」で再検索ができる。

⑤の検索履歴

検索履歴 ID	検索日時	検索条件式
<input checked="" type="checkbox"/> 1920255	2017-08-23 13:58:13	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 tglKugiri=" " AND tglBunKugiri="#" AND limitToSelfSentence="1" AND tglWords="20" AND unit="1" AND encoding="UTF-16LE" AND endOfLine="CRLF";



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 tglKugiri="|" AND tglBunKugiri="#" AND limitToSelfSentence="1" AND tglWords="20" AND unit="1" AND encoding="UTF-16LE" AND endOfLine="CRLF";

この状態で、下の「検索」ボタンをクリックすると再検索ができる。

☞ 最初の2行だけでも検索可能。

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

最小限「キー: 語彙素 LIKE "言葉"」だけでも OK。

- 「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。
- 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。

☞ 例えば、単に「言葉」を「人」に修正するなど、同じ条件で別の語を検索したい場合にも便利に使える。

➤ 検索例⑤の修正版（「言葉」を「人」へ）：

```
キー: 品詞 LIKE "形容詞%"  
AND 後方共起: 語彙素="人" ON 1 WORDS FROM キー  
WITH OPTIONS tglKugiri="|" AND tglBunKugiri="#" AND limitToSelfSentence="1" AND  
tglWords="20" AND unit="1" AND encoding="UTF-16LE" AND endOfLine="CRLF"
```

☞ OR 指定も可能。

➤ 検索例⑤の修正版（「言葉」または「人」へ）：

```
キー: 品詞 LIKE "形容詞%"  
AND 後方共起: (語彙素="言葉" OR 語彙素="人") ON 1 WORDS FROM キー  
WITH OPTIONS tglKugiri="|" AND tglBunKugiri="#" AND limitToSelfSentence="1" AND  
tglWords="20" AND unit="1" AND encoding="UTF-16LE" AND endOfLine="CRLF"
```

⑨ OR, AND は半角大文字で、前後に半角スペースを入れる、括弧()も半角。

5. 検索結果

5.1. 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。
500 件以上ある場合には表示は 500 件まで。並び順は無秩序。

5.2. 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。

5.3. 検索結果に関する注意

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
- 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、並べ替えが必要。

検索結果が 10 万件以上ある場合、10 万件で切れる。

- ダウンロードしたファイルには特に警告がないので注意。