

『中納言』を用いた 『日本語日常会話コーパス』の活用

柏野 和佳子 小磯 花絵



『中納言』について

『中納言』の概要(約5分) <https://youtu.be/AovMD2N4Qag>

「コーパス検索アプリケーション『中納言』音声配信」(約5分)

<https://youtu.be/XJTPLMj2yIs>

「まとめて検索『KOTONOHA』」(約5分半)

<https://youtu.be/CEdGQwIWa28>

『中納言』で公開中のコーパス



書き言葉	現代日本語書き言葉均衡コーパス 中納言版	BCCWJ
書き言葉	国語研日本語ウェブコーパス 中納言版	NWJC
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 モニター公開版	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
方 言	日本語諸方言コーパス モニター公開版	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

『日本語日常会話コーパス』モニター公開版

『日本語日常会話コーパス』のデータ量2倍に増加! (2021.2.17より)

短単位検索

検索フォームで検索 検索条件式で検索

前方共起条件の追加

キー -- 1 語 キーの条件を指定しない

書字形出現形 が

後方共起条件の追加

検索対象 設定を隠す

検索対象を選択

検索動作 設定を隠す

ダウンロードオプション 設定を隠す

列の表示 設定を隠す

検索対象の選択

公開時期

☐ 2018年度版 ☐ 2020年度版

左だけチェックをすれば
データ追加前と同じ
検索結果になります

ここをクリックすると
検索対象が選択可能に

コーパスの規模

	2018年度版	2020年度版
時間数	50時間	50時間
会話数	126会話	141会話
セッション数	116セッション	108セッション
ディスクサイズ	286.3ギガバイト	(HDDによる提供なし)

NEW!

* 一時的に会話の場に加わる人(店員など)を除く話者の数

全てのチェックを外す 全てにチェックを入れる OK



(約16分半)

<https://www.youtube.com/watch?v=7fSbEvbTmjc>

※2018年度版(収録50時間分)を使用

「中納言」の検索画面

日本語日常会話コーパス (モニター公開版) CEJC

↑コーパス選択画面 ログアウト

中納言 2.4.5 データバージョン 2018.12 マニュアル 語数について

現在のサーバ負荷状況:

現在 7 人ログイン中

コーパス検索アプリケーション

日本語日常会話コーパス (モニター公開版) CEJC

短単位検索長単位検索文字列検索位置検索

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

前方共起条件の追加

キー

--

1

語

キーの条件を指定しない

書字形出現形

 が

短単位の条件の追加

後方共起条件の追加

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

ダウンロードオプション 設定を表示する

検索

検索結果をダウンロード

条件クリア

キャンセル

オンライン検索システム「中納言」 4種類の検索法

日本語日常会話コーパス (モニター公開版) CEJC

短単位検索 長単位検索 文字列検索 位置検索

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

▼ 前方共起条件の追加

キー -- 1 語 ■ キーの条件を指定しない ⓘ

書字形出現形 ▼ が 短単位の条件の追加

▲ 後方共起条件の追加

- ① 短単位検索
- ② 長単位検索（日常会話コーパス・モニター版では未実装）
- ③ 文字列検索
- ④ 位置検索

オンライン検索システム「中納言」 4種類の検索法

日本語日常会話コーパス (モニター公開版) CEJC

短単位検索 長単位検索 文字列検索 位置検索

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

▼ 前方共起条件の追加

キー -- 1 語 ■ キーの条件を指定しない ⓘ

書字形出現形 ▼ が 短単位の条件の追加

▲ 後方共起条件の追加

- ① 短単位検索
- ② 長単位検索（日常会話コーパス・モニター版では未実装）
- ③ 文字列検索
- ④ 位置検索

文字列検索

①「文字列検索」のタブをクリックして選択

The screenshot shows the 'Text String Search' (文字列検索) interface. At the top, there are four tabs: '短単位検索' (Short Unit Search), '長単位検索' (Long Unit Search), '文字列検索' (Text String Search), and '位置検索' (Position Search). The 'Text String Search' tab is highlighted with a red dashed box and a red arrow pointing to it from the instruction box above. Below the tabs, there are three sub-tabs: '検索フォームで検索' (Search with Search Form), '検索条件式で検索' (Search with Search Conditions), and '履歴で検索' (Search with History). The '検索フォームで検索' sub-tab is active. In the search form, the '検索文字列' (Search String) input field contains the text 'はい' (Hai), which is highlighted with a red dashed box and a red arrow pointing to it from the instruction box below. To the right of the input field, there are options for '結果表示単位' (Result Display Unit) set to '短単位' (Short Unit) and 'キー表示形式' (Key Display Format) set to '語' (Word). Below the input field, there is a note about using wildcard characters like '%' and '#' for searching specific patterns, with examples like '国立%所' and '日本%'.

短単位検索 長単位検索 文字列検索 位置検索

文字列検索

検索フォームで検索 検索条件式で検索 履歴で検索

検索文字列 ① はい

結果表示単位 ☒ 短単位 キー表示形式 ① ☒ 語 ☐ 文字列全体

検索文字列では、^① 意味を持つメタ文字として以下に挙げる文字を使用できます。
% 任意の文字列 例：国立 % 所 → 「国立国語研究所」「国立社会保障・人口問題研究所」など
任意の文字列 例：日中友好 → 「日中友好」「日韓友好」など
→ 「日本人」「日本国」「日本語」
→ 「日本海」「日本史」「日本は」「日本の」など

② 検索したい文字列を入力
(例：「はい」)

検索動作 設定を隠す

文脈中の区切り記号 ① | ▾

文脈中の発話単位区切り記号 ① # ▾

前後文脈の語数 ① 20 ▾

ダウンロードオプション 設定を表示する

検索

検索結果をダウンロード 条件クリア キャンセル

③ 「検索」ボタンをクリック

文字列検索結果の例 (「はい」を分析した場合)

8,801 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。

検索対象語数: 1,247,791 記号・補助記号・空白を除いた検索対象語数: 1,247,080

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	公開時期	形式	話者ラベル	年齢	性別	出生地
T006_002	16720	10130	みんな 受ける から みんな A 日程 はい たま 市の 福祉 を受け て 春日部 市の 福祉 は	い	はい っすね # うん # ああ # あ # そっか そっか # あー # で 行政 も 僕 だけ な ん で # うん	イル	居る		イル	動詞-非自立可能	上一段-ア行	未然形一般	2018年度版	用談相談	IC03_青木	20-24歳	男性	埼玉県
T011_017	202430	113940	点数 を 言っ て くれた 人 に 確認 して る だけ です # うーん # で も 七十 点 台 は	い	はい って # あたし としお ん が 同じ だ もん # 意外 で しょ # あー # はい # あたし としお ん 同じ 点数	イル	居る		イル	動詞-非自立可能	上一段-ア行	未然形一般	2020年度版	雑談	IC03_高木	35-39歳	女性	長崎県
T005_019	75280	43320	あります の で その の 辺 を ちょっと 披露 は して えー もら いたい と 思 います の で #	はい	# あれ 入れる こと に した の かな # うん # はい # ア 新しい の ある の # この 前 の 医師	ハイ	はい		ハイ	感動詞-一般			2020年度版	会議会合	IC03_山川	35-39歳	女性	東京都
T004_003	98290	55630	閉 じ て # 鳥 賊 と 光 り です ね # で 座 っ て で また 幕 閉 じ て 帰 っ て み たい に こ #	はい	# マ グ ロ と 甘 エ ビ 三 つ です ね # 歩 け ない の よ # 支 え られ ない と # わ かり まし た # あー # す い	ハイ	はい		ハイ	感動詞-一般			2020年度版	雑談	N10A_酒井	70-74歳	男性	新潟県
T001_008	46380	27270	個人 で みんな 繋 が り 持 っ て や っ た ま う が なん か よ っ ま ど い い じ ゃ ない す か っ つ つ て #	はい	はい はい はい はい # そう です ね そう です ね # そう です ね そう です ね # ね # 今 度 井 上 は ん に	ハイ	はい		ハイ	感動詞-一般			2020年度版	雑談	IC03_広瀬	35-39歳	男性	鹿児島県

文字列検索結果の例

8,801 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。

検索対象語数: 1,247,791 記号・補助記号・空白を除いた検索対象語数: 1,247,080

- 検索件数が冒頭に表示
- その下に検索結果がランダムに表示
(同じ語でも検索するごとに出力結果の順番は異なる)
- 表示件数は500件まで

文字列検索結果の例

ピンクの列:コーパスの基本情報

▶ をクリック: 前後の音声を再生

「会話ID」をクリック: 前後の文脈の詳細な情報を表示

会話 ID	開始位置	連番
T007_007	25330	14550
T015_010	33070	18610
T009_021b	98530	57750
K004_013	86460	48380

T007_007	14420	IC01_塚田	する	スル	為る	動詞-非自立可能
T007_007	14430	IC01_塚田	って	ッテ	って	助詞-副助詞
T007_007	14440	IC01_塚田	ゆう	イウ	言う	動詞-一般
T007_007	14450	IC01_塚田	こと	コト	事	名詞-普通名詞-一般
T007_007	14460	IC01_塚田	だ	ダ	だ	助動詞
T007_007	14470	IC01_塚田	な	ナ	な	助詞-終助詞
T007_007	14480	IC05_栗山	はい	ハイ	はい	感動詞-一般
T007_007	14490	IC01_塚田	じゃ	デ	で	接続詞
T007_007	14500	IC01_塚田	これ	コレ	此れ	代名詞
T007_007	14510	IC01_塚田	これ	コレ	此れ	代名詞
T007_007	14520	IC01_塚田	しまっ	シマウ	仕舞う	動詞-非自立可能
T007_007	14530	IC01_塚田	とい	トク	とく	助動詞
T007_007	14540	IC01_塚田	て	テ	て	助詞-接続助詞
T007_007	14550	IC04_清水	はい	ハイ	はい	感動詞-一般
T007_007	14560	IC04_清水	ん	ンッ	んっ	感動詞-一般
T007_007	14570	IC04_清水	あっ	アッ	あっ	感動詞-一般
T007_007	14580	IC05_栗山	ん	ンッ	んっ	感動詞-一般
T007_007	14590	IC04_清水	これ	コレ	此れ	代名詞
T007_007	14600	IC04_清水	はい	ハイ	はい	感動詞-一般
T007_007	14610	IC05_栗山	うん	ウン	うん	感動詞-一般
T007_007	14620	IC05_栗山	イ			言いよどみ
T007_007	14630	IC05_栗山	切れ端	キレハシ	切れ端	名詞-普通名詞-一般
T007_007	14640	IC05_栗山	も	モ	も	助詞-係助詞
T007_007	14650	IC05_栗山	いい	ヨイ	良い	形容詞-非自立可能
T007_007	14660	IC05_栗山	よ	ヨ	よ	助詞-終助詞
T007_007	14670	IC04_清水	しまっ	シマウ	仕舞う	動詞-非自立可能
T007_007	14680	IC04_清水	とい	トク	とく	助動詞
T007_007	14690	IC04_清水	て	テ	て	助詞-接続助詞
T007_007	14700	IC04_清水	いい	ヨイ	良い	形容詞-非自立可能

文字列検索結果の例

紫の列

検索キーと前後の文脈
形態論情報(あとで詳述)

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
やっぱり机の上を綺麗にするってゆうことだな # はい # じゃこれこれしまっというて #	はい	# ん # あっ # ん # これ # はい # うん # イ切れ端 もいいよ # しまっというていい # うん # しまっ というて # きょう	ハイ	はい		ハイ	感動詞-一般		
はい # でもほらあのね皆さんに配れるものだった たらいいかなって # うんうん # うん #	はい	# お願いします # はい # はい # ぜひ # はい # は い # え # ****の # **** # **** # はい # はい # ホームページ とか	ハイ	はい		ハイ	感動詞-一般		
でえーとプレゼンします # あ # ごめんなさい # はい # はい # へー # はい # はい # 何月ぐらいだったか かな #	はい	# 七月だ # 七月 # うん # はい # 年ー二回 # う ん # でねえっとね七月にー	ハイ	はい		ハイ	感動詞-一般		
ですね # いや # 理系は理系関係ないと思います # 理系な # で関係ないと思いますけれど #	はい	# これは # じゃあ救急箱新刊 # これもねすご いあのこれ # あ # これはあれだね # あ # すご い	ハイ	はい		ハイ	感動詞-一般		

※ 検索結果の上にあるボタンで表示する列を選択できる

☐ 形態論情報

☒ 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 ☐ 書字形 ☐ 発音形出現形 ☐ 発音 ☐ 語種 ☐ 仮名 ☐ タグ付き書字形

文字列検索結果の例

緑の列（左側） 会話の基本情報
黄緑の列（右側） 話者の基本情報

会話概要	会話時間	話者数	話者間の関係性	形式	話者ID	話者ラベル	年齢	性別	出生地	居住地
地域活動の会の仲間8人とイベントの事前研修	32	9	友人知人・先生生徒・仕事関係	用談相談	T007_011	IC04_清水	45-49歳	女性	東京都	東京都
取材先の飲食店でオーナーに取材しながら	18	2	仕事関係	用談相談	T015	IC01_小川	50-54歳	男性	東京都	東京都
喫茶店でサークルの後輩に役員の業務の引継ぎ	38	2	友人知人	用談相談	T009_023	IC02_若菜	20-24歳	女性	北海道	埼玉県
職場の書店で先輩スタッフと会議	18	5	同僚	会議会合	K004_016	IC02_佐和	50-54歳	女性	東京都	神奈川県

※ 検索結果の上にあるボタンで表示する列を選択できる

☐ 会話情報

☐ セッション ID ☒ 会話概要 ☒ 会話時間 ☒ 話者数 ☒ 話者間の関係性 ☒ 形式 ☐ 場所 ☐ 活動

☐ 話者情報

☒ 話者 ID ☒ 話者ラベル ☒ 年齢 ☒ 性別 ☒ 出生地 ☒ 居住地 ☐ 職業 ☐ 協力者からみた関係性

検索結果のダウンロード

短単位検索 長単位検索 **文字列検索** 位置検索

文字列検索

検索フォームで検索 検索条件式で検索 履歴で検索

検索文字列 ① はい 結果表示単位 ☒ 短単位 キー表示形式 ① ☒ 語 ☐ 文字列全体

検索文字列では、特別な意味を持つメタ文字として以下に挙げる文字を使用できます。

%	任意の文字列	例：国立 % 所 → 「国立国語研究所」「国立社会保障・人口問題研究所」など
_	任意の1文字	例：日 _ 友好 → 「日米友好」「日中友好」「日韓友好」など
[]	文字クラス	例：日本 [人国語] → 「日本人」「日本国」「日本語」
[^]	文字クラスの否定	例：日本 [^ 人国語] → 「日本海」「日本史」「日本は」「日本の」など

検索動作 設定を隠す

文脈中の区切り記号 ① | ▾

文脈中の発話単位区切り記号 ① #

前後文脈の語数 ① 20 ▾

ダウンロードオプション 設定を表示する

検索 ② 検索結果をダウンロード 条件クリア キャンセル

② 「検索結果をダウンロード」ボタンをクリック
500件を超える検索結果も全てダウンロード可
非表示の情報（たとえば「発音」の情報など）もすべて含まれる

① 必要に応じて「ダウンロードオプション（設定を表示する）」をクリックして表示し
分析環境に合わせて適宜文字コード等を選択する

ダウンロードオプション 設定を隠す

システム ① Excel 向け ▾ 文字コード ① UTF-16LE ▾ 改行コード ① CRLF ▾ ☒ 出力ファイルが一つの場合は Zip 圧縮を行わない ①

☒ 検索条件式ごとに出力ファイルを分割する ①

☐ インラインタグを使用 ①

タグの区切り記号 ▾ ☐ 語彙素読み ☐ 語彙素 ☐ 語形 ☐ 品詞 大分類 ▾ ☐ 活用型 大分類 ▾ ☐ 活用形 大分類 ▾ ☐ 書字形 ☐ 発音形出現形 ☐ 語種

文字列検索の問題点

文字列として一致するものを検索しているだけ。

意図に反する検索結果が含まれている可能性あり。

K001_003a ▶	うん # でだからこの並縫い又結構細かく早くできるんですよ # はい はいはい # はい # 作業 #	はい	# 細か # あ # うん # やるのができるわけね # なるほどね # ふーん # それ	はい	感動詞- 一般
T015_008a ▶	さあ新しくこうリーダーになった人とかさそうゆう人が来るんだ けど #	はい	はい	はい	感動詞- 一般
T015_008a ▶	あと新しくこうリーダーになった人とかさそうゆう人が来るんだ けど # はい	はい	# はい うん # はい	はい	感動詞- 一般
T010_007 ▶	てもいいですか # うん # いいよ # じゃあ # あの # あ # 日程は自分で 出す分には	い	い # はい うん # はい	良い	形容詞- 非自立 可能
T016_007 ▶	はい # 先に払うの # 先払いだね # 五百円 # はい # ですって # 五百円 です # ン # はい # おいしょ # はい	はい	# ありがとうございました # 影になりそうないところがいいよね # こ っちかな # そっちはもう	はい	感動詞- 一般

感動詞「はい」以外の結果も
(出す分には**はい**んですけど)



適切な分析をするには形態論情報(短単位・長単位)を用いる!

オンライン検索システム「中納言」 4種類の検索法

- ① 短単位検索
- ② 長単位検索（日常会話コーパス・モニター版では未実装）
- ③ 文字列検索
- ④ 位置検索

日本語日常会話コーパス (モニター公開版) CEJC

短単位検索 長単位検索 文字列検索 位置検索

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

前方共起条件の追加

キー -- 1 語 ☐ キーの条件を指定しない

書字形出現形 が ☐ 短単位の条件の追加

後方共起条件の追加

短単位検索

①「短単位検索」タブをクリックして選択

The screenshot shows the 'Short Unit Search' (短単位検索) interface. At the top, there are four tabs: '短単位検索' (Short Unit Search), '長単位検索' (Long Unit Search), '文字列検索' (Text Search), and '位置検索' (Position Search). The '短単位検索' tab is selected and highlighted with a red dashed box. Below the tabs, there are three buttons: '検索フォームで検索' (Search with Search Form), '検索条件式で検索' (Search with Search Condition), and '履歴で検索' (Search with History). The '検索フォームで検索' button is selected. Below these buttons, there is a search form. The form has a dropdown menu for 'キー' (Key) with a value of '--', a dropdown for '1', and a dropdown for '語' (Word). There is a checkbox for 'キーの条件を指定しない' (Do not specify key conditions). Below this, there is a text input field for '書字形出現形' (Character shape appearance) with a value of 'はい' (Yes). This input field is highlighted with a red dashed box. Below the input field, there is a button for '短単位の条件の追加' (Add short unit conditions). At the bottom of the form, there are several options: '文脈中の区切り記号' (Delimiter in context), '文脈中の発話単位区切り記号' (Delimiter in context for utterance unit), '前後文脈の語数' (Number of words in context), and '共起条件の範囲' (Range of co-occurrence conditions). The '検索' (Search) button is highlighted with a red dashed box. Below the '検索' button, there are three buttons: '検索結果をダウンロード' (Download search results), '条件クリア' (Clear conditions), and 'キャンセル' (Cancel). The '検索' button is highlighted with a red dashed box.

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

前方共起条件の追加

キー -- 1 語 ☐ キーの条件を指定しない

書字形出現形 が はい

短単位の条件の追加

後方共起条件の追加

②検索したい短単位の
情報を入力(例:「はい」)

検索

検索結果をダウンロード

条件クリア

キャンセル

③「検索」ボタンをクリック

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

文字列検索時 8,801 件 → 8,115 件の検索結果が見つかりました。

短単位として提供される基本情報

情報	説明
語彙素	国語辞書の見出しに相当
語彙素読み	語彙素の読み
語形	同じ語彙素に所属するものとして語形の変異を区別したもの
書字形	同じ語形に所属するものとして表記の変異を区別したもの
タグ付き書字形	書字形に転記テキストのタグを付与したもの
発音形出現形	語形に対応する発音の形
発音	言い誤りなどを含む実際の発音の形
品詞	別表参照
活用型	動詞・形容詞・助動詞・一部の接尾辞の活用型
活用形	動詞・形容詞・助動詞・一部の接尾辞の活用形
語種	和語・漢語・外来語・混種語の別

短単位の品詞

形容動詞（静かだ、堂々たる）
は存在しない
語幹⇨形状詞
語尾⇨助動詞「だ」「たり」

2字漢語+する（例：勉強する）
⇒ 名詞+動詞「する」

1字漢語+する（例：称する）
⇒ 動詞（活用型：サ変変格）

名詞-普通名詞-一般	形状詞-一般	接頭辞
名詞-普通名詞-サ変可能	形状詞-タリ	接尾辞-名詞的-一般
名詞-普通名詞-サ変形状詞可能	形状詞-助動詞語幹	接尾辞-名詞的-サ変可能
名詞-普通名詞-形状詞可能	副詞	接尾辞-名詞的-副詞可能
名詞-普通名詞-副詞可能	接続詞	接尾辞-名詞的-助数詞
名詞-普通名詞-助数詞可能	動詞-一般	接尾辞-動詞的
名詞-数詞	動詞-非自立可能	接尾辞-形容詞的
名詞-固有名詞-人名-名	形容詞-一般	接尾辞-形状詞的
名詞-固有名詞-人名-姓	形容詞-非自立可能	記号-一般
名詞-固有名詞-人名-一般	助動詞	記号-文字
名詞-固有名詞-地名-国	助詞-格助詞	感動詞-一般
名詞-固有名詞-地名-一般	助詞-準体助詞	感動詞-フィラー
名詞-固有名詞-一般	助詞-係助詞	言いよどみ
名詞-助動詞語幹	助詞-副助詞	歌
代名詞	助詞-接続助詞	伏せ字
連体詞	助詞-終助詞	形態論情報付与対象外

話し言葉に特有

語彙素で検索

例) 語彙素が「凄い」であるものを検索(書字形:すごい、すごく、すんごい、すげえ、etc)

The screenshot shows the 'Short Unit Search' (短単位検索) interface. At the top, there are tabs for 'Short Unit Search' (短単位検索), 'Long Unit Search' (長単位検索), 'Character String Search' (文字列検索), and 'Position Search' (位置検索). Below these, there are buttons for 'Search with Search Form' (検索フォームで検索), 'Search with Search Conditions' (検索条件式で検索), and 'Search with History' (履歴で検索). The main search area has a dropdown menu for 'Key' (キー) with 'Lexeme' (語彙素) selected, followed by the word 'Amazing' (凄い) in the search input field. A red box highlights the 'Lexeme' dropdown and the search input field. A red arrow points to the 'Lexeme' dropdown with the label '①検索対象「語彙素」選択'. Another red arrow points to the search input field with the label '②検索ワードを入力'. Below the search area, there are buttons for 'Search' (検索), 'Download Search Results' (検索結果をダウンロード), 'Clear Conditions' (条件クリア), and 'Cancel' (キャンセル). At the bottom, there are settings for 'Search Action' (検索動作) and 'Hide Settings' (設定を隠す). The settings include 'Delimiter in Context' (文脈中の区切り記号) set to '|', 'Delimiter in Utterance Context' (文脈中の発話単位区切り記号) set to '#', 'Number of Words Before and After Context' (前後文脈の語数) set to 10, and 'Scope of Co-occurrence Conditions' (共起条件の範囲) set to 'Do not cross utterance unit boundaries' (発話単位境界をまたがない).

短単位検索

短単位検索 長単位検索 文字列検索 位置検索

検索フォームで検索 検索条件式で検索 履歴で検索

前方共起条件の追加

キー -- 1 語 ☐ キーの条件を指定しない

① 語彙素 が 凄い 短単位の条件の追加

①検索対象「語彙素」選択

②検索ワードを入力

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 10

共起条件の範囲 発話単位境界をまたがない

検索 検索結果をダウンロード 条件クリア キャンセル

3,691 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。

検索対象語数: 1,247,791 記号・補助記号・空白を除いた検索対象語数: 1,247,080

会話 ID	前文脈	キ	後文脈	語彙素	語形	発音
C002_015	ね みたい な # あー # うーん # あんまり 住ん で ない ね # うん # うん # そう です よ ね # ベビー カー も 電車	すごい	増え ます た よ ね # ね # ベビー カー 率 って ゆう か # 増え た # ね # また あっ きい の よ ね	凄	スゴイ	スゴイ
T005_011a	うん # うん # それ を 取る の も もう ト あ あ の ま か の 人 に 取 ら れ て る から もう	すごい	稼働 率 の 低い 和室 で やる と # うん # うん # うーん # ゆう # で 三 十 人 ぐ ら い 来 る から もう ぎ ぎ ち	凄	スゴイ	スゴイ
T015_014	な # で も 俺 なん か も あ の マコッ ちゃん と か さ ハル ちゃん と 長い か ら あ の 二 人 が	すごい	の ま ね 人 の 悪口 を 絶対 ゆ わ ない ね # うん うん うん # あー # 偉 い よ な # あ い つ ら	凄	スゴイ	スゴイ
C001_001	って # ね # 子供 は ね あ の 一 輪 の こ う ゆ う 二 個 付 い て る や つ の と か さ	すごい	やっ て ん じゃ ん # ね # うーん # 上手 だ よ ね # うーん # ね # もう 普 通 で やっ て る も ん ね # 今 # うーん	凄	スゴイ	スゴイ
K004_008	館 行 っ た り と か 大 人 も 続 け る 趣 味 な の ね # へー # ふーん # うん # なん か ね # へー # みたい #	すごい	ね # いい ね # そう な の # そう な の ね # うん # そう # ま お 兄 ちゃ ん バ イト も 始 め た	凄	スゴイ	スゴイ
T011_017	か 聞 い た ら う ん だ い じ ょ ぶ た だ 喉 が 痛 く て っ て ゆう か ら # だ っ て コ ス モ ス だ か マン ボウ が	すごい	似 て ん の は # うん # 一 気 に # うーん # ね # そ こ か ら です よ っ て ゆ っ た の # う そ 誰 が	凄	スゴイ	スゴイ
T001_011	いい ね # その 年代 で か っ こ いい ね # うーん # いい 味 出 し て る よ ね # うーん # うーん # うーん # なん か	すごい	爽 やか で き ら き ら だ っ た けど 今 こ う 洪 く て か っ こ いい み たい な # 太 平 記 の 時 は か わ い か っ	凄	スゴイ	スゴイ
T016_004b	よ ね # うーん # 甘い # 貝 # あ の さ お 母 さん 貝 貝 さ 海 で 拾 っ て き た の #	すご	甘 い # お い しい # いや # オ お い しい # お い しい で しょ # うーん # 拾 っ て き た # シ 慎 吾 さん に れ 食 べ て # こ の は	凄	スゴイ	スゴク
T002_013	って ゆう の が あ っ て # うん # うん # うん うん # うん うん うん # はい # うん # うん # うーん # うーん # で その	すごい	力 を 持 ち すぎ た が 故 に こ で 沸 々 と あ い つ ら は ま ん と じゃ な い の に	凄	スゴイ	スゴイ
C001_012	で さ ぴ ー ん で さ 思 う と は 天 才 少 年 と 一 緒 じゃ ん # うん # それ に な っ ちゃ う の が	すごい	怖 く て さ # うん # もし ま ん と に 思 い 付 い て や っ て ん だ っ た ら さ い づ も の く だ り で	凄	スゴイ	スーゴイ
T004_013	くる も ん だ か ら も の す ご い 個 人 差 は あ る じゃ ない です か # うん # 違 う # うん うん うん うん # あ る #	すごい	あ る と 思 い ま す よ # うーん # そう そう そう そう そう そう # ね # うん # うーん # うーん # あ っ # で も そ も そ も	凄	スゴイ	スッゴイ



後方共起条件を利用して検索

例) 語彙素「凄い」に**形容詞が後続**する事例を検索

短単位検索

長単位検索

文字列検索

位置検索

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

▼ 前方共起条件の追加

キー

--

1

語

☐ キーの条件を指定しない

語彙素

が

凄い

☐ 短単位の条件の追加

▲ 後方共起条件の追加

①「後方共起条件の追加」をクリック

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

ダウンロードオプション 設定を表示する

検索

検索結果をダウンロード

条件クリア

キャンセル

後方共起条件を利用して検索

例) 語彙素「凄い」に**形容詞が後続**する事例を検索



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

前方共起条件の追加

キー 語 1 キーの条件を指定しない

語彙素 が 凄い 短単位の条件の追加

後方共起1 キーから 語 1 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

書字形出現形 が 短単位の条件の追加

後方共起条件の追加

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

ダウンロードオプション 設定を表示する

検索 検索結果をダウンロード 条件クリア キャンセル

後方共起条件を利用して検索

例) 語彙素「凄い」に**形容詞**が**後続**する事例を検索

The screenshot shows the '短単位検索' (Short Unit Search) interface. At the top, there are tabs for '短単位検索', '長単位検索', '文字列検索', and '位置検索'. Below these are buttons for '検索フォームで検索', '検索条件式で検索', and '履歴で検索'. The main search area has a section for '前方共起条件の追加' (Add Forward Co-occurrence Condition) and a section for '後方共起' (Backward Co-occurrence). The '後方共起' section is highlighted with a green background. It contains a dropdown menu for 'キー' (Key) set to '語彙素' (Lexeme), a text input field containing '凄い', and a button '短単位の条件の追加'. Below this, there is a section for '後方共起1' (Backward Co-occurrence 1) with a dropdown menu for 'キーから' (Key from) set to '1', a dropdown menu for '語' (Word) set to '語', and a checkbox 'キーと結合して表示' (Combine key and display). To the right of this section are buttons 'この条件をキーに' (Use this condition as key) and 'この共起条件を削除' (Delete this co-occurrence condition). Below the '後方共起1' section, there is a dropdown menu for '品詞' (Part of speech) set to '品詞', a dropdown menu for '大分類' (Major category) set to '大分類', and a dropdown menu for '形容詞' (Adjective) set to '形容詞'. A red box highlights the '品詞' and '大分類' dropdowns, and a red arrow points to the '形容詞' dropdown. A blue box contains the text '② 後方共起の検索対象として「品詞」を選択' (Select 'Part of speech' as the search target for backward co-occurrence). Another blue box contains the text '③ 品詞(大分類)として「形容詞」を選択' (Select 'Adjective' as the part of speech (major category)). At the bottom of the interface, there are buttons for '検索' (Search), '検索結果をダウンロード' (Download search results), '条件クリア' (Clear conditions), and 'キャンセル' (Cancel). There are also input fields for '文脈中の発話単位区切り記号' (Contextual utterance unit separator), '前後文脈の語数' (Number of words in context), and '共起条件の範囲' (Range of co-occurrence conditions).

② 後方共起の検索対象として「品詞」を選択

③ 品詞(大分類)として「形容詞」を選択

後方共起条件を利用して検索

例) 語彙素「凄い」に形容詞の**連体形**が後続する事例を検索

後方共起条件を利用して検索

例) 語彙素「凄い」に形容詞の**連体形**が後続する事例を検索

The screenshot displays the '短単位検索' (Short Unit Search) interface. At the top, there are tabs for '短単位検索', '長単位検索', '文字列検索', and '位置検索'. Below these, there are buttons for '検索フォームで検索', '検索条件式で検索', and '履歴で検索'. The main search area is divided into sections for adding conditions. The '前方共起条件の追加' (Add Forward Co-occurrence Condition) section is currently active, showing a search for '凄い' (Amazing). Below this, the '後方共起1' (Backward Co-occurrence 1) section is highlighted in green. It shows a search for '品詞: 形容詞' (Part of speech: Adjective) and '活用形: 連体形' (Form: Attributive). A red arrow points to the 'AND' button, indicating that the search is for 'Amazing' followed by 'Adjective in Attributive form'. The '検索動作' (Search Action) section at the bottom allows for setting search parameters like '文脈中の区切り記号' (Delimiter in context), '文脈中の発話単位区切り記号' (Delimiter in utterance unit in context), '前後文脈の語数' (Number of words in context), and '共起条件の範囲' (Range of co-occurrence conditions). A 'ダウンロードオプション' (Download Options) section is also visible at the bottom left.

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

前方共起条件の追加

キー -- 1 語 キーの条件を指定しない

語彙素 が 凄い 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

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

後方共起条件の追加

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

ダウンロードオプション 設定を表示する

検索 検索結果をダウンロード 条件クリア キャンセル

後方共起条件を利用して検索

例) 語彙素「凄い」に形容詞の**連体形**が後続する事例を検索

The screenshot shows a search interface with the following elements:

- Search Type:** 短単位検索 (Short Unit Search) is selected.
- Search Method:** 検索フォームで検索 (Search with Search Form) is selected.
- Front Condition:** 前方共起条件の追加 (Add Front Co-occurrence Condition). Key: 語彙素 (Lexeme) が 凄い (Amazing).
- Back Condition:** 後方共起1 (Back Condition 1). Key: 品詞 (Part of Speech) の 大分類 (Major Classification) が 形容詞 (Adjective).
 - AND:** 活用形 (Inflection Form) の 大分類 (Major Classification) が 連体形 (Attributive Form).
- Annotations:** Red boxes and arrows highlight the selection of '活用形' (Inflection Form) and '連体形' (Attributive Form) in the back condition.
- Search Action:** 検索 (Search) button.
- Download Options:** 検索結果をダウンロード (Download Search Results), 条件クリア (Clear Conditions), キャンセル (Cancel).
- Search Settings:** 検索動作 設定を隠す (Hide Search Settings).
 - 文脈中の区切り記号 (Delimiter in Context): |
 - 文脈中の発話単位区切り記号 (Delimiter in Utterance Unit): #
 - 前後文脈の語数 (Number of Words in Context): 20
 - 共起条件の範囲 (Range of Co-occurrence Conditions): 発話単位境界をまたがない (Does not cross utterance unit boundaries)
- Download Options:** ダウンロードオプション 設定を表示する (Show Download Options Setting).

後方共起条件を利用して検索

例) 語彙素「凄い」に形容詞の**連体形**が後続する事例を検索

▼ 前方共起条件の追加

i キー

— —

▼

1

▼

語

▼

☐ キーの条件を指定しない

i

語彙素

▼

が

凄い

☐

☒ 短単位の条件の追加

i 後方共起1

キーから

▼

1

▼

語

▼

☐ キーと結合して表示

i

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

▲ 後方共起条件の追加

後方共起条件を利用して検索

例) 語彙素「凄い」+形容詞の連体形に、**更に普通名詞が後続**する事例を検索

▼ 前方共起条件の追加

キー 1 語 ☐ キーの条件を指定しない

語彙素

後方共起1 キーから 語 ☐ キーと結合して表示

品詞 が 形容詞

AND 活用形 が 連体形

▲ 後方共起条件の追加

①「後方共起条件の追加」をクリック

後方共起条件を利用して検索

例) 語彙素「凄い」+形容詞の連体形に、更に普通名詞が後続する事例を検索

The screenshot shows a web-based search interface for Japanese linguistic data. At the top, there are four tabs: "短単位検索" (Short Unit Search), "長単位検索" (Long Unit Search), "文字列検索" (String Search), and "位置検索" (Position Search). The "短単位検索" tab is active.

Below the tabs, there is a search bar with a magnifying glass icon and the text "短単位検索". To the right of the search bar are three buttons: "検索フォームで検索" (Search with Search Form), "検索条件式で検索" (Search with Search Condition Formula), and "履歴で検索" (Search with History).

The main search area is divided into sections. The first section is "前方共起条件の追加" (Add Forward Co-occurrence Condition). It contains a form with a "キー" (Key) dropdown set to "--", a "1" dropdown, and a "語" (Word) dropdown. There is a checkbox "キーの条件を指定しない" (Do not specify key conditions) and an information icon. Below this, there is a text input field with "語彙素" (Lexeme) and "が 凄い" (is amazing), and a button "短単位の条件の追加" (Add short unit condition).

The second section is "後方共起1" (Add Backward Co-occurrence Condition 1). It contains a form with a "キーから" (From Key) dropdown, a "1" dropdown, and a "語" (Word) dropdown. There is a checkbox "キーと結合して表示" (Display combined with key) and an information icon. To the right, there are buttons "この条件をキーに" (Use this condition as key) and "この共起条件を削除" (Delete this co-occurrence condition). Below this, there is a text input field with "品詞" (Part of speech) and "の 大分類" (of major classification) and "が 形容詞" (is adjective), and a button "短単位の条件の追加" (Add short unit condition).

The third section is "後方共起2" (Add Backward Co-occurrence Condition 2). It contains a form with a "キーから" (From Key) dropdown, a "2" dropdown, and a "語" (Word) dropdown. There is a checkbox "キーと結合して表示" (Display combined with key) and an information icon. To the right, there are buttons "この条件をキーに" (Use this condition as key) and "この共起条件を削除" (Delete this co-occurrence condition). Below this, there is a text input field with "品詞" (Part of speech) and "の 中分類" (of sub-classification) and "が 名詞-普通名詞" (is noun-general noun), and a button "短単位の条件の追加" (Add short unit condition). A red arrow points to this section.

At the bottom, there is a section "検索動作 設定を隠す" (Search Action Hide Settings). It contains several dropdown menus: "文脈中の区切り記号" (Delimiter in context) set to "|", "文脈中の発話単位区切り記号" (Delimiter in utterance unit in context) set to "#", "前後文脈の語数" (Number of words in context) set to "20", and "共起条件の範囲" (Range of co-occurrence conditions) set to "発話単位境界をまたがない" (Does not cross utterance unit boundary). There is a button "ダウンロードオプション 設定を表示する" (Download Option Show Settings).

At the bottom right, there are four buttons: "検索" (Search), "検索結果をダウンロード" (Download search results), "条件クリア" (Clear conditions), and "キャンセル" (Cancel).

後方共起条件を利用して検索

例) 語彙素「凄い」+形容詞の連体形に、更に普通名詞が後続する事例を検索

短単位検索 長単位検索 文字列検索 位置検索

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

前方共起条件の追加

キー --- 1 語 ☐ キーの条件を指定しない

語彙素 が 凄い

後方共起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

2 後方共起2 キーから 2 語 ☐ キーと結合して表示

品詞 の 中分類 が 名詞-普通名詞

後方共起条件の追加

88 件の検索結果が見つかりました。

会話 ID	前文脈	キー	後文脈	語彙素	語形
T003_017	は ない # 石山 何 ちゃん # 石山 志乃 ちゃん # 志乃 ちゃん #	すごく	(かわいい) (子) # メール で あ の 次 の 方 を 探し	凄い	スゴイ
T007_015	こう マップ に 表わ してく わけ だよ # あー # あれ	すごく	(いい) (こと) で # ガイド マップ を 作る ん す ね	凄い	スゴイ
T004_006	思っ ちゃう よ ね # うん # うーん # 設備 的 に は	すんごく	(いい) (ところ) だ っ た の # あー # そう # うーん # 新し く でき た て	凄い	スゴイ

後方共起条件を利用して検索

例) 語彙素「凄い」+形容詞の連体形に、**更に普通名詞が後続**する事例を検索

The screenshot shows a search interface with the following conditions:

- 前方共起条件の追加** (Add anterior co-occurrence conditions)
 - キー** (Key): 語彙素 (Lexeme) が 凄い (Amazing)
- 後方共起1** (Posterior co-occurrence 1)
 - キーから** (From key): 品詞 (Part of speech) の 大分類 (Major classification) が 形容詞 (Adjective)
 - 活用形** (Inflection form) の 大分類 (Major classification) が 連体形 (Attributive form)
- 後方共起2** (Posterior co-occurrence 2)
 - キーから** (From key): 品詞 (Part of speech) の **中分類** (Subclassification) が 名詞-普通名詞 (Noun - Common noun)

名詞-普通名詞-一般

名詞-普通名詞-サ変可能

名詞-普通名詞-サ変形状詞可能

名詞-普通名詞-形状詞可能

名詞-普通名詞-副詞可能

名詞-普通名詞-助数詞可能

大分類

中分類

小分類

大分類

名詞-普通名詞-一般	形状詞-一般	接頭辞
名詞-普通名詞-サ変可能	形状詞-タリ	接尾辞-名詞的-一般
名詞-普通名詞-サ変形状詞可能	形状詞-助動詞語幹	接尾辞-名詞的-サ変可能
名詞-普通名詞-形状詞可能	副詞	接尾辞-名詞的-副詞可能
名詞-普通名詞-副詞可能	接統詞	接尾辞-名詞的-助数詞
名詞-普通名詞-助数詞可能	動詞-一般	接尾辞-動詞的
名詞-数詞	動詞-非自立可能	接尾辞-形容詞的
名詞-固有名詞-人名-名	形容詞-一般	接尾辞-形状詞的
名詞-固有名詞-人名-姓	形容詞-非自立可能	記号-一般
名詞-固有名詞-人名-一般	助動詞	記号-文字
名詞-固有名詞-地名-国	助詞-格助詞	感動詞-一般
名詞-固有名詞-地名-一般	助詞-準体助詞	感動詞-フィラー
名詞-固有名詞-一般	助詞-係助詞	言いよどみ
名詞-助動詞語幹	助詞-副助詞	歌
代名詞	助詞-接統助詞	伏せ字
連体詞	助詞-終助詞	形態論情報付与対象外

中分類

名詞-普通名詞-一般	形状詞-一般	接頭辞
名詞-普通名詞-サ変可能	形状詞-タリ	接尾辞-名詞的-一般
名詞-普通名詞-サ変形状詞可能	形状詞-助動詞語幹	接尾辞-名詞的-サ変可能
名詞-普通名詞-形状詞可能	副詞	接尾辞-名詞的-副詞可能
名詞-普通名詞-副詞可能	接続詞	接尾辞-名詞的-助数詞
名詞-普通名詞-助数詞可能	動詞-一般	接尾辞-動詞的
名詞-数詞	動詞-非自立可能	接尾辞-形容詞的
名詞-固有名詞-人名-名	形容詞-一般	接尾辞-形状詞的
名詞-固有名詞-人名-姓	形容詞-非自立可能	記号-一般
名詞-固有名詞-人名-一般	助動詞	記号-文字
名詞-固有名詞-地名-国	助詞-格助詞	感動詞-一般
名詞-固有名詞-地名-一般	助詞-準体助詞	感動詞-フィラー
名詞-固有名詞-一般	助詞-係助詞	言いよどみ
名詞-助動詞語幹	助詞-副助詞	歌
代名詞	助詞-接続助詞	伏せ字
連体詞	助詞-終助詞	形態論情報付与対象外

小分類

名詞-普通名詞-一般

名詞-普通名詞-サ変可能

名詞-普通名詞-サ変形状詞可能

名詞-普通名詞-形状詞可能

名詞-普通名詞-副詞可能

名詞-普通名詞-助数詞可能

名詞-数詞

名詞-固有名詞-人名-名

名詞-固有名詞-人名-姓

名詞-固有名詞-人名-一般

名詞-固有名詞-地名-国

名詞-固有名詞-地名-一般

名詞-固有名詞-一般

名詞-助動詞語幹

代名詞

連体詞

形状詞-一般

形状詞-タリ

形状詞-助動詞語幹

副詞

接続詞

動詞-一般

動詞-非自立可能

形容詞-一般

形容詞-非自立可能

助動詞

助詞-格助詞

助詞-準体助詞

助詞-係助詞

助詞-副助詞

助詞-接続助詞

助詞-終助詞

接頭辞

接尾辞-名詞的-一般

接尾辞-名詞的-サ変可能

接尾辞-名詞的-副詞可能

接尾辞-名詞的-助数詞

接尾辞-動詞的

接尾辞-形容詞的

接尾辞-形状詞的

記号-一般

記号-文字

感動詞-一般

感動詞-フィラー

言いよどみ

歌

伏せ字

形態論情報付与対象外

「頻度」について

「度数・頻度」「粗頻度」

- 度数・頻度：コーパスから取り出された生の件数の値。
- 粗頻度 raw frequency：コーパスから観測された頻度情報。rawには、「生の、未加工の」といった意味があり、コーパスから取り出された状態のまま、調整を加えていないことをふまえ、このように呼ぶ。
- 粗頻度の特性：元となったコーパスサイズの情報が含まれていない。「10回」という粗頻度が1万語コーパスから得られた10回なのか、100万語コーパスから得られた10回なのか示されていない。



サイズの異なるコーパスから取り出された粗頻度は互いに比較できない。

「調整頻度」

- 調整頻度 adjusted frequency : 異なるコーパスから得られた頻度の相互比較が可能になるよう、粗頻度を一定の基準によって調整した値のこと。
- 調整頻度の特性:
 - たとえば、500万語コーパスから得られた10回と、2,000万語コーパスから得られた30回を、そのまま比較することに意味がない。
 - 共通の物差しを用意して、個々の頻度をそれにあわせて調整する必要がある。
 - 最も単純な方法は、度数÷総度数 で求めること。
 - 0.000002回($10 \div 500$ 万)と、0.0000015回($30 \div 2,000$ 万)

※あまりに値が小さく、扱いがたい。

→ 単純な相対頻度に変えて、調整頻度を用いる。

「PMW」

- 調整頻度の基準値にはさまざまな値を使用できるが、値をある程度大きくして読み取りやすくするために、コーパス研究では、通例、**「100万語あたり調整頻度」**が使用される。
- PMW per million words: 100万語あたり調整頻度は、英語の頭文字をとり、PMWと表記する。
 - 0.000002回 $(10 \div 500万) \times 100万 = 2$ 回
 - 0.0000015回 $(30 \div 2,000万) \times 100万 = 1.5$ 回
- ただし、常にPMWが適当なわけではない。100万語未満のコーパスには不適とする立場がある。
 - 調整しようとするすべてのコーパスサイズの下限値を超えない範囲で最も大きな10の乗数(1万、10万、100万、1,000万…)を選ぶのが原則。

調整頻度を出すために必要な「語数表」①

「5_cejc_wc.xlsx」より集計し、
「6_CEJC_語数表_形式と場面と年齢と性別と
話者間の関係性.xlsx」を作成

年齢	男性	女性	計
10代以下	43,863	5,775	49,638
20代	118,009	100,957	218,966
30代	111,782	160,995	272,777
40代	91,318	159,261	250,579
50代	64,171	144,314	208,485
60代	55,265	71,048	126,313
70代以上	75,782	38,180	113,962
不明	2,033	4,327	6,360
合計	562,223	684,857	1,247,080

形式名	語数
雑談	931,137
用談相談	235,061
会議会合	80,882
合計	1,247,080

場面名	語数
自宅	231,556
職場	114,814
学校	66,278
施設	567,570
室内	170,632
屋外	61,126
交通機関	35,104
合計	1,247,080

調整頻度を出すために必要な「語数表」②

話者間の関係性	語数	
サービス場面関係	13486	
家族	295232	
家族・サービス場面関係	9851	
家族・親戚	79534	
A 家族・友人知人	6430	66,632
仕事関係	86508	
B 仕事関係・同僚	2437	5,721
仕事関係・同僚・友人知人	2452	
C 仕事関係・友人知人	24108	28,254
先生生徒	14501	
先生生徒・友人知人	27872	
先生生徒・友人知人・同僚	9891	
同僚	87499	
B 同僚・仕事関係	3284	×
友人知人	489609	
友人知人・サービス場面関係・初対面	4909	
A 友人知人・家族	60202	×
友人知人・家族・初対面	6383	
C 友人知人・仕事関係	4146	×
友人知人・仕事関係・初対面	9723	
友人知人・先生生徒・仕事関係	9023	
合計	1,247,080	

話者間の関係性	語数
家族	295232
家族・親戚	79534
友人知人	489609
同僚	7499
仕事関係	86508
サービス場面関係	13486
先生生徒	14501
合計	1066369

- ✓ 付与されたそのまま集計したもの
→ 分類にゆれがあるためそのまま使わない。
- ✓ ABCはアノテーション作業のゆれ。使用する場合は合計し、合計した数値を使う。
- ✓ なお、中黒で並列されているものは、いずれかの関係性であり特定する情報は付与されていない。
- ✓ 特定できるものだけを抜粋したものは右の表。
- ✓ 話者間の差異を見るのにはこの右の分類のみ使用することを推奨。

集計例

文末の「じゃん」を検索

短単位検

①「発話単位末から」を選択

検索フォームで検索 検索条件式で検索 履歴で検索

前方共起条件の追加

② 検索対象「語彙素」選択

③ 検索ワードを入力

④ここをクリックしてダウンロード

ここは好みに適宜変更する。

このオプションは基本的にはデフォルトでよい。

検索対象 設定を隠す

検索対象を選択 検索対象をクリア

検索動作 設定を隠す

文脈中の区切り記号 なし

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

ダウンロードオプション 設定を表示する

検索

検索結果をダウンロード

条件クリア

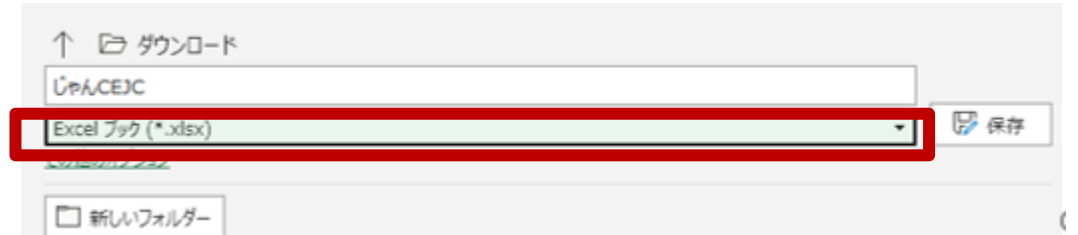
キャンセル

2,349 件の検索結果が見つかりました。

検索結果をダウンロードとしてファイル保存

- 特に指定しなければ「ダウンロード」フォルダに、「kwic-6848463.csv」のようなファイル名で保存される。
- ダウンロードしたファイルをクリックして開き（Excelが起動することを想定）、「ファイル」→「名前を付けて保存」という操作を行い、形式「Excelブック」を指定し、たとえば、「じゃんCEJC.xlsx」のような名前で保存する。

※必ず、形式「Excelブック」を指定すること。



参考（Googleスプレッドシートを使用する場合）：

- ① Googleドライブの「マイドライブ」の適当なフォルダに、「kwic-6848463.csv」をドラッグして、アップロードする。
- ② アップロードしたファイル名をダブルクリックし、上部に表示される「Googleスプレッドシートで開く」をクリックし、ファイルを開く。
- ③ 「ファイル」タブの「Googleスプレッドシートとして保存」をクリックして保存し、名前を「じゃんCEJC」のように変更するとよい。

ピボットテーブルの作成

1. 「挿入」→「ピボットテーブル」(一番左にあり)→「OK」。
 2. 「キー」を「値」のところへドラッグ&ドロップする。
 3. 「年齢」にチェックを入れる。→「行」に挿入される。
 4. 「性別」を「列」のところへドラッグ&ドロップする。
 5. 左のテーブル上の「列ラベル」を「降順」に。
- ※図の左のように集計される。

年齢	男性	女性	総計
0-4歳	4	2	6
10-14歳	65		65
15-19歳	24	14	38
20-24歳	279	224	503
25-29歳	83	5	88
30-34歳	7	98	105
35-39歳	235	229	464
40-44歳	145	170	315
45-49歳	74	71	145
50-54歳	78	146	224
55-59歳	16	116	132
5-9歳	14		14
60-64歳	18	53	71
65-69歳	55	25	80
70-74歳	63	23	86
75-79歳	4	2	6
80-84歳	1		1
85-89歳		3	3
90-94歳		1	1
NA	2		2
(空白)			
総計	1167	1182	2349

ピボットテーブルの作成

参考（Googleスプレッドシートを使用する場合）：

- ① 「データ」→「ピボットテーブル」→「作成」。
- ② 「行」の「追加」をクリックして「年齢」を選択。
- ③ 「列」の「追加」をクリックして「性別」を選択。
並べ替えを「降順」に。
- ④ 「値」の「追加」をクリックして「キー」を選択。

Google Sheets

じゃんCEJC

☆ 国

ファイル編集表示挿入表示形式データツールアドオンヘルプ最終編集: 数秒前

100% ¥ % .0 .00 123 Arial 10

A2 fx 年齢

	A	B	C	D	E	F
1	キーの COUNT 性別					
2	年齢	男性	女性	総計		
3				0	0	
4	0-4歳	4	2		6	
5	10-14歳	65			65	
6	15-19歳	24	14		38	
7	20-24歳	279	224		503	
8	25-29歳	83	5		88	
9	30-34歳	7	98		105	
10	35-39歳	235	229		464	
11	40-44歳	145	170		315	
12	45-49歳	74	71		145	
13	5-9歳	14			14	
14	50-54歳	78	146		224	
15	55-59歳	16	116		132	
16	60-64歳	18	53		71	
17	65-69歳	55	25		80	
18	70-74歳	63	23		86	
19	75-79歳	4	2		6	
20	80-84歳	1			1	
21	85-89歳		3		3	
22	90-94歳		1		1	
23	NA	2			2	
24	総計	1167	1182	0	2349	

kwic-6848463.csv!1:2352

候補

行 追加

年齢 閉

並べ替え 昇順 年齢 総計を表示

列 追加

性別 閉

並べ替え 降順 性別 総計を表示

値 追加

キー 閉

集計 COUNTA 表示方法 デフォルト

フィルタ 追加

年齢・性別の集計→調整頻度→グラフ

「5-9歳」の位置に注意。
たとえば、10代～70代
で集計してみる。

※赤字はセル名の例
※調整頻度、たとえば、
左上のセルにこのよう
に数式を入れれば、残
りのセルはそのコピーで
よい。

M4

語数表

年齢	男性	女性
10代以下	43,863	5,775
20代	118,009	100,957
30代	111,782	160,995
40代	91,318	159,261
50代	64,171	144,314
60代	55,265	71,048
70代以上	75,782	38,180

Q4

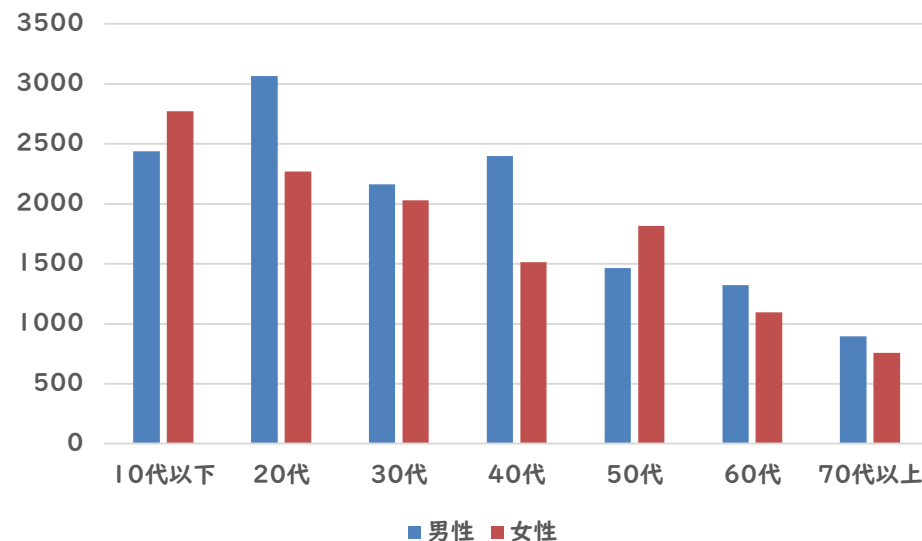
粗頻度

年齢	男性	女性
10代以下	107	16
20代	362	229
30代	242	327
40代	219	241
50代	94	262
60代	73	78
70代以上	68	29

=Q4/M4*1000000

調整頻度

年齢	男性	女性
10代以下	2439.413629	2770.5628
20代	3067.562644	2268.2924
30代	2164.928164	2031.119
40代	2398.212839	1513.2393
50代	1464.836141	1815.4857
60代	1320.908351	1097.8493
70代以上	897.310707	759.55998



「じゃん」の年齢別・性別の調整頻度 (PMW)

形式の集計→調整頻度→グラフ

語数表

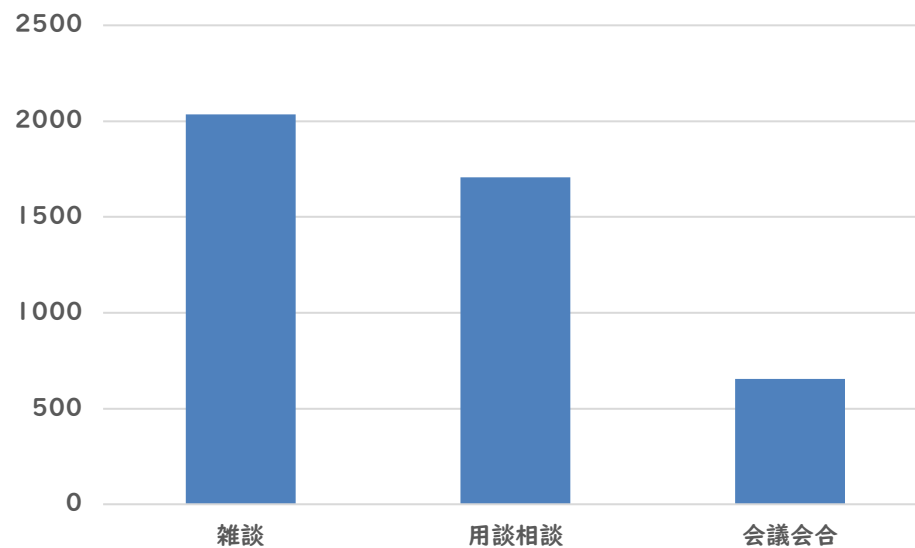
形式名	語数
雑談	931137
用談相談	235061
会議会合	80882

粗頻度

形式名	語数
雑談	1895
用談相談	401
会議会合	53

調整頻度

形式名	語数
雑談	2035.146
用談相談	1705.94
会議会合	655.2756



「じゃん」の形式別の調整頻度 (PMW)

話者間の関係の集計→調整頻度→グラフ

語数表

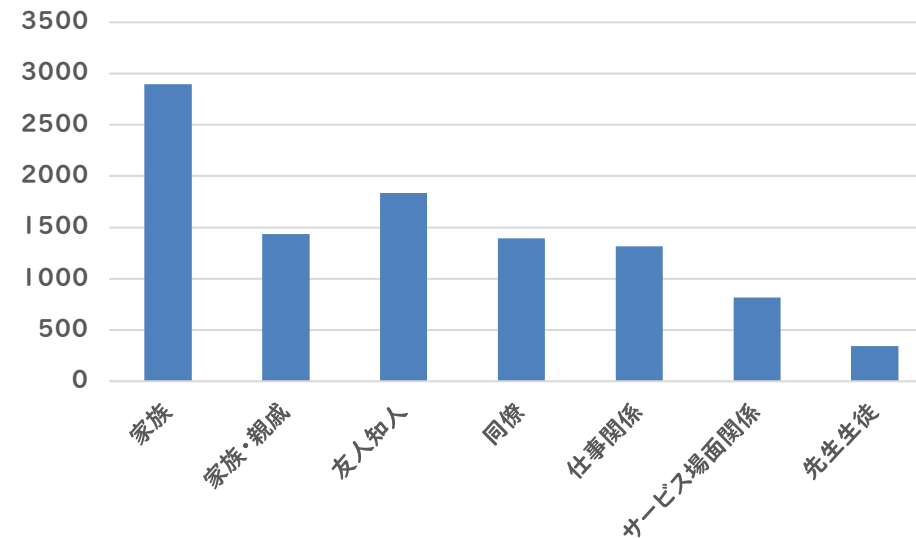
話者間の関係性	語数
家族	295232
家族・親戚	79534
友人知人	489609
同僚	87499
仕事関係	86508
サービス場面関係	13486
先生生徒	14501

粗頻度

話者間の関係性	語数
家族	855
家族・親戚	114
友人知人	898
同僚	122
仕事関係	114
サービス場面関係	11
先生生徒	5

調整頻度

話者間の関係性	語数
家族	2896.028
家族・親戚	1433.349
友人知人	1834.117
同僚	1394.302
仕事関係	1317.797
サービス場面関係	815.6607
先生生徒	344.8038



「じゃん」の話者間の関係別の調整頻度 (PMW)