

国立国語研究所学術情報リポジトリ

「中納言」を活用した分析

メタデータ	言語: Japanese 出版者: 研究社 公開日: 2026-07-24 キーワード: 作成者: 小磯, 花絵 メールアドレス: 所属: 国立国語研究所
URL	https://doi.org/10.15084/0002000725

This work is licensed under a Creative Commons
Attribution-NonCommercial-ShareAlike 4.0
International License.



第5章 | 「中納言」を活用した分析

本章では、コーパス検索アプリケーション「中納言」で『日本語日常会話コーパス』を活用する方法を紹介する。1節で中納言で検索可能な短単位情報・長単位情報について解説した上で、2節では中納言の基本的な機能とその利用法を、3節ではCEJCの活用事例を紹介する。

□1 短単位情報・長単位情報の基本

本節では、コーパス検索アプリケーション「中納言」の説明に入る前に、中納言で検索可能な短単位情報と長単位情報の基本について解説する。両単位の詳細については小椋ほか（2011a, b）を参照されたい。

1.1 単位の認定基準

『日本語日常会話コーパス』（CEJC）では、形態論情報として短単位と長単位という2種類の言語単位の情報を提供している。

短単位は、単位の斉一さを指向して設計されたものである。現代語において意味を持つ最小単位を規定した上で、最小単位が和語や漢語の場合は2つまで結合させることによって、また外来語や助詞、助動詞などの場合は結合させないことによって、短単位を認定する。例えば「国立国語研究所」の場合、最小単位は「国|立|国|語|研|究|所」と分割されるが、いずれも漢語であるため、短単位は「国立/国語/研究/所」と分割される。一方、「オレンジ|色|だ|けど」の場合、外来語「オレンジ」や助動詞「だ」、助詞「けど」は結合させないため、短単位は「オレンジ/色/だ/けど」となる。

一方、長単位は、文節を認定した上で、各文節を自立語部分と付属語部分に分割することにより認定される。「国立国語研究所」のような複合名詞、「勉強する」のようなサ変動詞、「について」のような複合辞、「ところが」や「要するに」のような接続詞などは、短単位では細かく分割されることがあるのに対し、長単位では全体で1単位と認定される。

1.2 階層的見出しの体系

CEJCの短単位情報は、階層的見出し構造に基づく電子化辞書UniDicに準拠して付与されている。UniDicでは、異形態や異表記、発音の変異などを整理した上で、語を、**語彙素**、**語形**、**書字形**のように階層化した形で登録することにより、書字や語形の違いに関わらず同じ語を同一の見出し（語彙素）でまとめて表現する（表1）。

表1 階層的見出しの例

語彙素（語彙素読み）	語形	書字形	発音形
本当（ホントウ）	ホントウ	本当	ホントー
		ほんとう	
		ホントウ	
	ホント	ほんと	ホント
		ホント	

1.3 短単位・長単位で提供する基本情報

短単位・長単位で提供する基本情報を表2に示す。

表2 短単位・長単位で提供する基本情報

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（第3章参照）を付与したもの
発音形	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	表3に示すセットの品詞の情報
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

CEJCでは、語形に対応する標準的な**発音形**に加え、言い誤りなどを含む実際の発音の情報も併せて提供している。第3章で解説したように、CEJCの転記

テキストには、例えば「けれども:」(非語彙的な母音の引き延ばし)や「改%革」(非語彙的な音の詰まり)、「(W ギーツ|技術)」(言い誤りや音のなまけ)などの**転記用タグ**が付与されている。これらの事例では、解析対象は「けれども」「改革」「技術」であり、対応する標準的な発音形は「ケレドモ」「カイカク」「ギジュツ」となる。一方、タグで表現された語末母音の延長や語中の詰まり、発音のなまけを含む「ケレドモー」「カイッカク」「ギーツ」などの実際の発音の情報も、発音形とは別に提供される。

1.4 品詞情報

短単位・長単位の**品詞情報**の一覧を表3に示す。

表3 短単位・長単位の品詞情報の一覧(*の品詞は短単位のみ)

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

短単位は語彙主義的な品詞体系であり、文脈により取りうる品詞を「名詞-普通名詞-形状詞可能」のように「～可能」として定義する。一方、長単位は用法主義的な体系であり、実際の文脈でどの用法かを判断するため、「～可能」という品詞は存在しない。例えば「安全」は短単位では「名詞-普通名詞-形状詞可能」だが、長単位の場合、「安全(です)」では「形状詞-一般」、「安全(が大

短単位検索を中心に、中納言の基本的な操作の流れを示す。

- ①**検索方法の種類を選択**：文字列による検索(2.2節)、短単位情報あるいは長単位情報を活用した検索(2.3節)、コーパス中の位置を一意に指定する検索(2.4節)の中から検索方法を選択する。図1は短単位検索の画面だが、長単位検索もほぼ同じである。
- ②**検索条件の指定方法を選択**：いずれの検索方法でも、**検索フォーム**で検索する方法(2.3.1節)と、**検索履歴**を活用して**検索条件式**で検索する方法(2.3.2節)がある。図1は検索フォームの画面である。
- ③**キーの条件を指定**：短単位・長単位検索において検索フォームで検索する場合、品詞や語彙素など、検索条件として指定したいキーを選び、具体的に条件を指定する。前後に共起する語の条件を指定したり、「発話単位の冒頭の語」といったように**発話単位**(第3章1.3節)を基準に位置の条件を設定したりすることもできる。
- ④**検索対象の選択**：**会話形式**(雑談、用談・相談、会議・会合、授業・レッスン)と**コア**か**非コア**か(コア：長単位を人手修正している範囲)の点から検索対象を絞り込むことができる。
- ⑤**検索の実行**：必要に応じて、検索結果に表示される前後の文脈の語数や区切り記号などを指定し、検索を実行する。検索件数など結果の概要とともに500件を上限に検索結果が表示される。
- ⑥**結果のダウンロード**：10万件を上限に結果をダウンロードできる。必要に応じて文字コードや改行コードなどを指定する。次の検索をするには「条件クリア」をクリックして指定した条件をリセットする。

2.2 文字列検索

■**文字列検索の基本** 図2に文字列検索の画面を示す。この図では「について」という文字列を検索している。

文字列検索

検索したい文字列を入力

検索フォームで検索

表示単位の切り替え

履歴で検索

検索文字列 ① について

結果表示単位 ① ○短単位 ●長単位

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

ワイルドカードの情報

図2 文字列検索の画面

結果の表示単位を短単位と長単位から選ぶことができる。表示単位を切り換えることで、短単位は「に/ついで」と3つに分割されることが分かる。文字列による検索であるため、「服装についてるボタン」のように複合辞以外のものも結果に含まれる。短単位や長単位がどのように区切られるか分からない場合は、まず文字列検索をし、単位や品詞、語彙素表記などの情報を把握した上で、短単位検索や長単位検索で詳細に条件を指定して検索するとよい。

■**ワイルドカード** 中納言では、表4に示すワイルドカードを用いた検索をすることができる。検索文字列の指定の下にも利用できるワイルドカードと使用例が示されている(図2)。これらのワイルドカードは、文字列検索だけでなく、短単位検索・長単位検索でも利用できる。

表4 中納言で利用できるワイルドカード

記号	説明	使用例	検索例
%	0文字以上の任意の文字列	心%	心、心配、心配り
_	任意の1文字(アンダーバー)	心_	心配、心地
		心__	心配り、心掛け
[]	カッコ内の文字いずれか1文字	[心表感]情	心情、表情、感情
[^]	カッコ内の文字以外の任意の1文字	[^心表感]情	事情、苦情、風情

■**検索結果** 検索結果の例を図3に示す。表示単位は長単位である。

193 件の検索結果が見つかりました。
検索対象語数：2,190,917 記号・補助記号・空白を除いた検索対象語数：2,189,952 検索対象会話数：577

会話 ID	開始位置	前文脈	キー	後文脈	語彙素	品詞	活用形	収録年	形式	年齢	性別
T001_003	184230	んだってゆう(話し)をしていて(その)	について	も(だ)そう(いう)意味(では)紛争(を)する	について	助詞-格助詞		2016年	雑談	35-39歳	男性
W001_001a	6900	6900 (ダイソンの掃除機)を(くれ)	について	# 廣松利が # 廣松利が # 一緒(に)行っ	について	助詞-格助詞		2019年	会議・会合	50-54歳	女性
T008_009		ダイソンの掃除機(を)くれ	について	# 廣松利が # 廣松利が # 一緒(に)行っ	について	助詞-格助詞		2016年	雑談	30-34歳	女性

図3 文字列検索の結果(表示単位：長単位)

検索件数など検索結果の概要が示され、その下に500件を上限に検索結果が表示される。検索の度に結果の並びは変わる。また500件以上ヒットした場合、表示される500件の内容も検索の度に異なる。

図3の結果では前後文脈の語数を10語に設定しているが、CEJCでは最大で300語まで表示できる。ただし、前後文脈に話者情報は表示されないため、会話内容を把握することが難しいこともある。検索結果の一番左の列の会話IDをクリックすると、話者の情報や形態論情報を含む前後文脈の結果が表示される。また会話IDの下のマークをクリックすると、当該箇所とその前後の音声を聴くことができる。

2.3 短単位検索・長単位検索

本節では短単位検索を例に説明するが、長単位検索も方法は同じである。

2.3.1 検索フォームで検索

■キーを指定して検索 図4に示すように、検索キーの種類をプルダウンから選び①、検索ワードや品詞等の検索条件を指定する②。

キーの種類は、書字形出現形、語彙素、語形のように文字列を入力して指定するもの、品詞、活用形、語種のように候補をプルダウンで選択するもの、WHERE句のように検索条件をWHERE句¹の書式で入力するものがある。

図4 短単位・検索フォームの画面

辞書の見出しに相当する語彙素を指定することで、語形や表記、活用形によらずまとめて検索することができる。語彙素は「愈(イヨイヨ)」など見慣れない表記のこともあるため、その場合は語彙素読み(片仮名)や文字列検索など

1| 第8章で紹介するリレーショナルデータベースを操作するためのSQL言語のうち、抽出条件を指定するWHERE句の説明も参照のこと。

で表記を調べるとよい。語形(片仮名)をキーに用いると表記や活用形を一括して検索することができる。また書字形出現形では会話音声を文字化したテキストの表記が検索対象となる。

品詞は大分類(例:名詞)、中分類(例:名詞-普通名詞)、小分類(例:名詞-普通名詞-サ変可能)から、活用形は大分類(例:連用形)と小分類(例:連用形-撥音便)から選ぶ。

■**複数条件の設定** 「難い(ガタイ/ニクイ)」や「会う/合う(アウ)」、「様(サマ):名詞/接尾辞」のように、語彙素か語彙素読みの一方のみ、あるいは両方を指定しても、検索対象が一意に定まらないこともある。その場合には複数の条件を設定する。「短単位の条件の追加」(図4③)をクリックすると、図5のようにANDのあとで新たな条件を指定することができる。

図5 複数条件の設定

■**発話単位中の位置の設定** 第3章で説明した発話単位を基準に、発話単位の冒頭あるいは末尾からの位置を指定できる(図4④)。例えば図6(左)は、代名詞が発話単位冒頭(頭から1語)に出現するもの、図6(右)は「助詞-接続助詞」が発話単位末から3語以内に出現するものが検索対象となる。

図6 発話単位中の位置の設定

■**前方共起条件・後方共起条件の設定** キーの前後に共起する語の条件を指定できる。前方あるいは後方の**共起条件**を指定する場合は図4の⑤あるいは⑥をクリックする。図7では、「動詞」(キー)の1語前が「副詞」で、「動詞」のあと3語以内に「助詞-終助詞」が現れるものが検索対象となる。前方共起・後方共起を合わせて10個まで指定できる。

▼ 前方共起条件の追加

① 前方共起1 キーから 1 語 ☐ キーと結合して表示 ①

品詞 の 大分類 が 副詞

① キー -- 1 語以内 ☐ キーの条件を指定しない ①

品詞 の 大分類 が 動詞

① 後方共起1 キーから 3 語以内 ☐ キーと結合して表示 ①

品詞 の 中分類 が 助詞-終助詞

▲ 後方共起条件の追加

図7 前方共起条件・後方共起条件の設定

2.3.2 履歴で検索・検索条件式で検索

図1②の検索条件の指定方法で「履歴で検索」を選択すると、図8に示ような検索履歴が表示される。

				検索フォームで検索	検索条件式で検索	履歴で検索
検索履歴 ID	検索日時	コーパス略称	データバージョン	検索条件式		
☐ 15738714	2025-02-04 22:18:15	CEJC	2023.03	キー: 品詞 LIKE "動詞%" AND 前方共起: 品詞 LIKE "副詞%" ON 1 WORDS FROM キー AND 後方共起: 品詞 LIKE "助詞-終助詞%" WITHIN 3 WORDS FROM キー WITH OPTIONS tglKugiri="1" AND tglBunKugiri="#" AND limitToSelfSentence="1" AND tglWords="20" AND unit="1" AND encoding="UTF-16LE" AND endOfLine="CRLF"; 編集して検索		
☐ 15738657	2025-02-04 22:03:32	CEJC	2023.03	キー: 品詞 LIKE "名詞%" WITH OPTIONS tglKugiri="1" AND tglBunKugiri="#" AND limitToSelfSentence="1" AND tglWords="20" AND unit="1" AND encoding="UTF-16LE" AND endOfLine="CRLF"; 編集して検索		

図8 検索履歴の画面

「編集して検索」をクリックすると、図9のように「検索条件式で検索」の画面に推移し、選択した検索条件式が表示される。

🔍 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 "名詞%"
 WITH OPTIONS tglKugiri="1" AND tglBunKugiri="#" AND limitToSelfSentence="1" AND tglWords="20" AND unit="1" AND encoding="UTF-16LE" AND endOfLine="CRLF";

図9 検索条件式の確認・編集画面

この図ではキーの条件が「品詞 LIKE "名詞%"」となっている。LIKEは条件をパターンで指定し、表4に示したワイルドカードが利用できる。この条件式では「名詞」で始まる品詞が検索対象となる。同じ条件のまま検索することでもできるが、「名詞」を「動詞」に編集するなどして検索することもできる。

2.4 位置検索

位置検索は、コーパス中の位置を一意に指定する情報を用いて検索する方法である。CEJCでは会話IDと開始位置を半角カンマでつないだ値が位置情報となる。例えば図3の一番上の結果を指定する場合は「T001_003,184230」となる。改行して複数指定することもできる。

□3 中納言の活用事例

3.1 事例1

「早いです」や「あります」が「早いっす」「あるっす」となるように、助動詞「です/ます」が「っす/す」と縮約して発話されることがある。こうした縮約形の「っす/す」(以下「っす」)を男性と女性のどちらが用いやすいか、年齢層による差が見られるかを調べてみよう。

■ステップ1 下準備として助動詞「です/ます」にどのような語形・書字形・書字形出現形があるかを調べる。短単位検索で品詞が「助動詞」でありかつ語彙素が「です」または「ます」であるという条件で検索し、結果をダウンロードした上で語形・書字形・書字形出現形をまとめると、表5ようになる。

表5 助動詞「です/ます」の語形・書字形・書字形出現形

語彙素	語形	書字形	書字形出現形 (一例)
です	デス	です	でし (連用形-一般)、です (終止形-一般/連体形-一般) でん (終止形-撥音便)、でしょう/でしょ (意志推量形)
	ツス	っす	っす/す (終止形-一般) っしょう/しょう/っしょ/しょ (意志推量形)
	デシュ	でしゅ	でしゅ (終止形-一般)
	デチュ	でちゅ	でちゅ (終止形-一般)
	デツ	でつ	でち (連用形-一般)
ます	マス	ます	ませ (未然形-一般)、まし (連用形-一般/命令形) ます (終止形-一般/連体形-一般)、まっ (終止形-促音便) ましょう/ましょ (意志推量形)
	マシュ	ましゅ	ましゅ (終止形-一般)

表5を見ると、「っす」の活用形が「終止形一般」と「意志推量形」に限られるため、このあとの手順では、条件を揃えるためにこの2つの活用形に限定する。また「終止形一般」のうち「でしゅ/でちゅ/ましゅ」は計12件と限られるため対象外とし、「です/ます/っす」に限定する。

■**ステップ2** 検索フォームで語彙素「です」、品詞「助動詞」、活用形「終止形一般」を指定して検索する。検索履歴から次の検索式が確認できる。

キー：(語彙素="です" AND 品詞 LIKE "助動詞%" AND
活用形="終止形一般")

この検索式を編集して上述の条件を一括検索してみよう。語彙素が「です」あるいは(OR) 語彙素が「ます」であり、かつ (AND) 品詞が「助動詞」で開始し、かつ活用形が「終止形一般」あるいは活用形が「意志推量形」である、という条件である。OR条件よりAND条件の方が優先されるため、OR条件をカッコで囲む必要がある。

キー：((語彙素="です" OR 語彙素="ます") AND 品詞 LIKE "助動詞%"
AND (活用形="終止形一般" OR 活用形="意志推量形"))

■**ステップ3** ステップ2の結果をダウンロードし、語形「デシュ/デチュ/マシュ」を除外する。その上で、性別と年齢層毎に「デス/マス」と「ッス」の頻度を求めて「ッス」と発話される率(以下、縮約率)を算出する。結果を表6に示す。なお人数の少ない20歳未満は対象外とし、60歳以上はまとめた。結果から、若干の増減はあるものの年齢が若い人ほど縮約率は高く、またどの世代も女性より男性の方が多く用いていることが分かる。

■**ステップ4** 第1章4節で述べたように、CEJCには調査協力者など収録の機会が多く総語数の多い話者もいれば、語数の少ない話者もいる。ステップ3のようにまとめて集計すると、語数の多い話者の傾向が結果に強く影響する可能性がある。そこで、話者毎に縮約率を求めた上で性別・年齢層毎の縮約率を算出する方法を試してみよう。ダウンロードした結果に含まれる話者IDを活用し、話者ID毎に「デス/マス」と「ッス」の頻度を求め、縮約率を算出する。その上で性別・年齢層毎に縮約率の平均値を算出すると表7のようになる。

表6 性別・年齢層毎に見た助動詞「です/ます」の縮約率

年齢層	男性			女性		
	「デス/マス」	「ッス」	縮約率	「デス/マス」	「ッス」	縮約率
20代	2804	1240	30.7%	2376	190	7.4%
30代	4014	820	17.0%	3028	152	4.8%
40代	2520	437	14.8%	3919	83	2.1%
50代	1950	391	16.7%	4593	54	1.2%
60歳～	3708	249	6.3%	2888	57	1.9%

表7 性別・年齢層毎に見た助動詞「です/ます」の縮約率の平均（話者別集計）

年齢層	男性		女性	
	話者別集計	全体集計（参考）	話者別集計	全体集計（参考）
20代	18.1%	30.7%	5.2%	7.4%
30代	17.9%	17.0%	4.4%	4.8%
40代	13.6%	14.8%	1.7%	2.1%
50代	12.5%	16.7%	2.5%	1.2%
60歳～	7.8%	6.3%	1.1%	1.9%

参考までにステップ3で求めた全体集計の縮約率も記載する。概ね両者は一致するが、20代男性は話者別集計の縮約率は18.1%、全体集計の場合は30.7%と、かなり異なる。これは縮約率の高い話者の語数が多かったためである。

話者別集計では性別・年齢層毎に複数話者の縮約率が得られるため、データの分布を把握できる。話者別集計の結果を図10に示す。ここでは標準誤差によるエラーバーを付している。

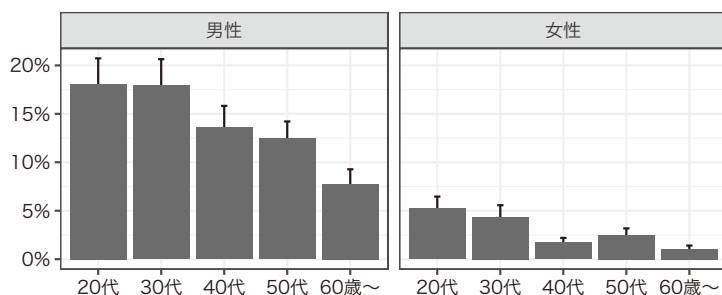


図10 性別・年齢層毎の助動詞「です/ます」の縮約率（話者別集計）

3.2 事例2

次に、事例1で対象とした助動詞「です/ます」の縮約形「ッス」の前に、どのような品詞が出現しやすいかを調べてみよう。

■**ステップ1** 手順を図11に示す。まず語彙素「です」(助動詞・終止形一般)の語形「デス」を検索し、結果に問題がないかを確認する(①)。次に「前方共起条件の追加」をクリックすると(②)、図11中段の画面に推移する。更に「この条件をキーに」をクリックすると(③)、図11下段のように語彙素「です」の条件が「後方共起1」に移る。最後に「キーの条件を指定しない」をクリックした後、検索を実行する(④)。

Figure 11 illustrates the four steps for searching for the pre-occurrence of the contraction 'ッス' (ssu) from the verb 'です' (desu):

- Step 1:** Initial search for 'です' (Verb, Termination Form General) with the form 'デス' (Desu). The interface shows search criteria: 語彙素 (Lexeme) 'です', 語形 (Form) 'デス', 品詞 (Part of speech) '助動詞' (Auxiliary verb), and 活用形 (Conjugation form) '終止形一般' (Termination form general).
- Step 2:** Clicking '前方共起条件の追加' (Add Anterior Co-occurrence Condition) to move to the '前方共起1' (Anterior Co-occurrence 1) screen. The interface shows search criteria: 前方共起1 (Anterior Co-occurrence 1) 'キーから' (From key), 1, 語 (Verb), and 書字形出現形 (Character form appearance form) 'デス'.
- Step 3:** Clicking 'この条件をキーに' (Use this condition as key) to move the condition to the '後方共起1' (Posterior Co-occurrence 1) screen. The interface shows search criteria: 後方共起1 (Posterior Co-occurrence 1) 'キーから' (From key), 1, 語 (Verb), and 書字形出現形 (Character form appearance form) 'デス'.
- Step 4:** Clicking 'キーの条件を指定しない' (Do not specify key condition) and then performing the search. The interface shows search criteria: 後方共起1 (Posterior Co-occurrence 1) 'キーから' (From key), 1, 語 (Verb), and 書字形出現形 (Character form appearance form) 'デス'.

図11 語彙素「です」(助動詞・終止形一般)の語形「デス」の前接語をキーに検索する手順

■**ステップ2** 検索履歴を次のように編集し、語彙素「です/ます」、語形「デス/マス」の「終止形一般」と「意志推量形」をまとめて検索する。この検索式では、

「語彙素 LIKE "[でま]す"」のようにワイルドカードを使用し、語彙素「です/ます」や語形「デス/マス」をまとめて検索する。

後方共起: (語彙素 LIKE "[でま]す" AND 語形 LIKE "[デマ]ス"
AND 品詞 LIKE "助動詞%" AND
(活用形="終止形-一般" OR 活用形="意志推量形"))
ON 1 WORDS FROM キー

■**ステップ3** ステップ1、ステップ2を参考に、語形「ッス」の「終止形-一般」「意志推量形」を検索する。その上で、結果をダウンロードし、キーの品詞毎に「デス/マス」と「ッス」の頻度を求めて縮約率を算出すると表8のようになる（「デス/マス」「ッス」の総頻度が1000以上に限定）。助詞以外は大分類の品詞を採用した。結果から、前接する語の品詞によって縮約率が大きく異なり、前接語が形容詞の場合に最も高く、動詞の場合はほとんど縮約しないことなどが確認できる。

表8 「デス/マス」と「ッス」に前接する語の品詞別頻度と縮約率

品詞	「デス/マス」	「ッス」	縮約率
形容詞	3142	1205	27.7%
助動詞	2957	506	14.6%
助詞-準体助詞	7843	1154	12.8%
副詞	3360	398	10.6%
形状詞	1112	131	10.5%
名詞	4811	188	3.8%
動詞	7739	48	0.6%

3.3 事例3

再度、事例1の表6を見てみよう。「ッス」の出現頻度が性別・年齢層毎に示されている。これを**粗頻度**という。表から20代の男性の「ッス」の粗頻度が1240と最も高いことが分かるが、この値だけで20代の男性が「ッス」を最も多く用いると結論付けることはできない。性別・年齢層毎に総語数に片寄りがあり、20代男性の総語数がほかの年齢層よりも格段に多いために必然的に「ッス」も多く出現しただけ、という可能性が残る。このように総語数が異なるデータを

比較する場合、粗頻度ではなく総語数を考慮した相対頻度や調整頻度を用いることがある（石川ほか 2010; 石川 2012）。コーパス分析でよく用いられるのは、**100万語あたりの調整頻度（Per Million Words, PMW）**である。この節ではPMWを用いた分析を取り上げる。

■**ステップ1** 中納言の検索画面（図1）の右上にある「語数について」をクリックすると、CEJCの**語彙表・語数表**を提供するページが開く。ここでは「CEJC短単位語数表_品詞構成表_語種構成表」（zipファイル）のうち「8_CEJC_語数表_形式と場所と年齢と性別と話者間の関係性.xlsx」を用いる。このファイルには、性別と年齢層毎の総語数の情報などがまとめられている。

■**ステップ2** 性別・年齢層毎の総語数と、表6の「ッス」の粗頻度から、性別・年齢層毎のPMWを算出する。例えば20代男性の場合、総語数は220979、「ッス」の頻度は1240であることから、PMWは $1240 \times 1000000 / 220979 = 5611.4$ となる。結果を図12に示す。事例1と同じく、若い人ほど「ッス」の使用は多く、またどの世代も男性の方が多く用いる傾向が確認できる。

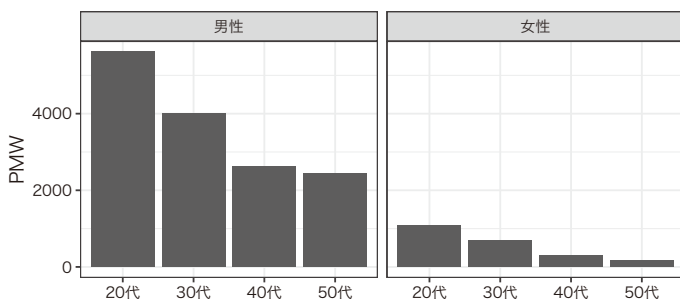


図12 性別・年齢層毎の「ッス」のPMW

■**ステップ3** 次に、事例1のステップ4で見たように、話者を考慮した分析を行ってみよう。「CEJC短単位語数表_品詞構成表_語種構成表」に含まれる「7_cejc_wc.xlsx」を用いる。会話ID・話者ID毎の総語数が、話者の情報（年齢層、性別など）とともにまとめられている。この表を活用して、話者毎の語数を計算すると、表9のようになる（「話者ID」から「総語数」まで）。この情報と、事例1のステップ4で求めた話者毎の「ッス」の頻度から、話者毎にPMWを求める（表9「PMW」）。

表9 話者毎の総語数・「ッス」頻度・PMWの一例

話者 ID	年齢層	性別	総語数	「ッス」頻度	PMW
C001	40-44歳	女性	21158	2	94.5
C001_001	45-49歳	女性	5167	1	193.5
C001_002	40-44歳	女性	5281	3	568.1
C001_003	35-39歳	女性	870	0	0
C001_004	40-44歳	女性	1889	0	0
C001_010	NA	男性	16	0	0

これをもとに、性別・年齢層毎に「ッス」のPMWの平均値と平均誤差を算出すると、図13のようになる。若い人ほど「ッス」のPMWは高く、またどの世代も女性より男性の方が高い傾向が確認できる。

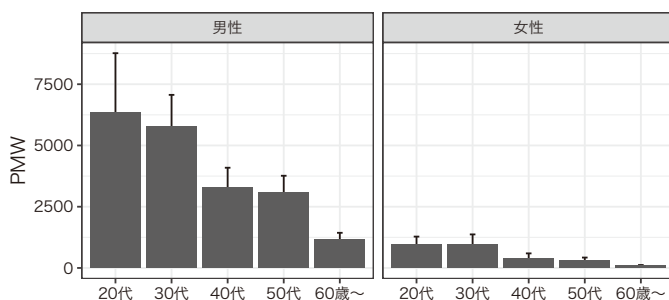


図13 性別・年齢層毎の「ッス」のPMW(話者別集計)

3.4 事例4

「食べて は (いけない)」など動詞・連用形に接続助詞「て」と係助詞「は」が後続する場合、「食べちゃ」のように「ちゃ」と縮約することがある。こうした縮約形がどのような関係性の相手に使用されやすいかを調べてみよう。

■**ステップ1** 下準備として動詞・連用形に後続する接続助詞「て」の語形を調べると、「テ」や「デ」(例：見て、飲んで)のほかに、「ては/では」の縮約形「チャ」や「ジャ」(例：見ちゃ、飲んじゃ)も含まれていることが分かる。

■**ステップ2** これを踏まえ、図14のように縮約しないケース(左)と縮約するケース(右)を検索する。語形について、ワイルドカードを用いて“[テデ]”や“[チジ]ャ”のようにまとめて指定している。このように検索フォームでもワイルドカードを用いることができる。

キー
--
1
語

品詞
の
大分類
が
動詞

AND
活用形
の
大分類
が
連用形

後方共起1
キーから
1
語
☐ キーと結合

語彙素
が
て

AND
品詞
の
中分類
が
助詞-接続助詞

AND
語形
が
[テ]

後方共起2
キーから
2
語
☐ キーと結合

語彙素
が
は

AND
品詞
の
中分類
が
助詞-係助詞

キー
--
1
語

品詞
の
大分類
が
動詞

AND
活用形
の
大分類
が
連用形

後方共起1
キーから
1
語
☐ キーと結合

語彙素
が
て

AND
品詞
の
中分類
が
助詞-接続助詞

AND
語形
が
[チ]ヤ

図14 動詞 + 「て/で」「は」と動詞 + 「ちゃ/じゃ」の検索フォーム画面

■ステップ3 中納言の検索結果には、第1章4節で紹介した「話者間の関係性」（会話中の全ての話者間の関係性の組合せ）と、「協力者から見た関係性」の情報が含まれる。ここでは前者の関係性を用いる。「家族」「友人知人」「同僚」「先生生徒」や「家族・友人知人」などの関係性（の組合せ）が含まれる。個人差を考慮するため、これらの関係性と話者ID毎に縮約率を算出した上で、平均と標準誤差を求めた。結果の一部を図15に示す。家族や友人知人同士の会話では縮約率は高く、同僚や仕事関係者、先生生徒同士では相対的に低い傾向が確認できる。

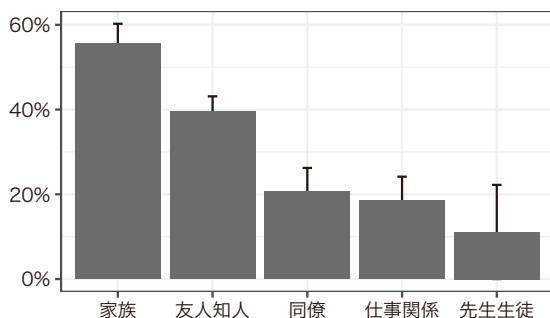


図15 話者間の関係性毎の縮約率

■ステップ4 ステップ3で取り上げた話者間の関係性では、例えば先生生徒間の会話の場合、先生から生徒への発話なのか、その逆なのかは分からない。そこで第1章4節で紹介した「話し手から見た聞き手との関係性」を用いて分析

してみよう。この関係性の情報は、CEJCのウェブサイトからダウンロードする必要がある²⁾。中納言の結果と合わせて分析する場合、中納言のセッションID、話者IDと、関係性データのsessionID、informantIDをキーにして結合すればよい。結果を図16に示す。結果から、学生から先生に対する発話よりも、先生から学生に対する発話の方が縮約率が高いことが分かる。

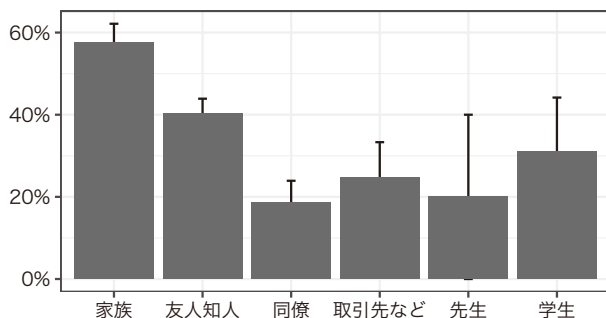


図16 話し手から見た聞き手との関係性毎の縮約率

□4 おわりに

本章では、「中納言」の基本機能とその利用法を概説し、CEJCを活用した具体的な分析方法を紹介した。形態論情報や話者の年齢層・性別などのメタデータを活用することで、多角的に言語現象を把握できることを示した。中納言では、BCCWJなどCEJC以外のコーパスも同様に検索可能であり、異なるレジスターの言葉と日常会話との比較も可能である。こうした比較は、日本語の使用実態をより深く理解する手がかりとなるだろう。

参考文献

- 石川慎一郎・前田忠彦・山崎誠(2010)『言語研究のための統計入門』くろしお出版。
 石川慎一郎(2012)『ベーシックコーパス言語学』ひつじ書房。
 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕(2011a)『現代日本語書き言葉均衡コーパス』形態論情報規程集第4版(上)』
<http://doi.org/10.15084/00002855>
 小椋秀樹・小磯花絵・富士池優美・宮内佐夜香・小西光・原裕(2011b)『現代日本語書き言葉均衡コーパス』形態論情報規程集第4版(下)』
<http://doi.org/10.15084/00002856>

2) https://www2.ninjal.ac.jp/conversation/cejc/data_relation.html

関連URL (2025年7月31日確認)

コーパス検索アプリケーション「中納言」

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

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

(小磯花絵)

