『現日研・職場談話コーパス』中納言版公開データの作成

著者	柏野 和佳子, 大村 舞, 西川 賢哉, 小磯 花絵
雑誌名	言語資源活用ワークショップ発表論文集
巻	3
ページ	495-510
発行年	2018
URL	http://doi.org/10.15084/00001684

『現日研・職場談話コーパス』中納言版公開データの作成

柏野 和佳子（国立国語研究所音声言語研究領域）*
大村 舞（国立国語研究所コーパス開発センター）
西川 賢哉（国立国語研究所コーパス開発センター）
小磯 花絵（国立国語研究所音声言語研究領域）

Supplemental Arrangement for Public Data Available in the Chunagon Versions of "Gen-Nichi-Ken Corpus of Workplace Conversation"

Wakako Kashino (National Institute for Japanese Language and Linguistics)
Mai Omura (National Institute for Japanese Language and Linguistics)
Ken'ya Nishikawa (National Institute for Japanese Language and Linguistics)
Hanae Koiso (National Institute for Japanese Language and Linguistics)

要旨

『現日研・職場談話コーパス』は，現代日本語研究会が作成した，首都圏の有職女性 19 名（20 代～50 代）と，首都圏の有職男性 21 名（20 代～50 代）の職場での自然談話を文字起こししたテキストを元に作成したコーパスである。その元となっている文字化テキストは,『合本　女性のことば・男性のことば（職場編）』（現代日本語研究会編，2011 年，ひつじ書房）の付録 CD－ROM に収録されている。国立国語研究所に提供されたその文字化テキストを MeCab＋UniDic で解析し，オンライン検索システム『中納言』にて『現日研・職場談話コーパス』として公開する。

本発表では,『現日研・職場談話コーパス』の概要と特徴を述べる。

１．はじめに

現在，国立国語研究所の機関拠点型基幹研究プロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」（プロジェクトリーダー：小磯花絵）にて，日常会話場面を対象とした大規模な『日本語日常会話コーパス』を構築中である(小磯ほか 2018)。その公開を前に，国語研に提供いただいた既存の会話データをオンライン検索システム『中納言』にて公開するということを進めている。2016 年には，『名大会話コーパス』(藤村ほか 2011)の中納言版(柏野ほか 2017)を公開した。それに続けて，このたび『現日研・職場談話コーパス』の中納言版を 2018 年 8 月より一般公開する運びとなった。

『現日研・職場談話コーパス』は，現代日本語研究会が作成した，首都圏の有職女性 19 名（20 代～50 代）と，首都圏の有職男性 21 名（20 代～50 代）の職場での自然談話を文字起こししたテキストを元に作成したコーパスである。その元となっている文字化テキストは，現代日本語研究会編(2011)の付録 CD－ROM に収録されている。「大規模日常会話コーパスに基づく話し言葉の多角的研究」のプロジェクトにおいて，その文字化テキストを対象に，形態素解析用辞書『UniDic』と形態素解析器『MeCab』を用いて形態論情報

* waka ＠ninjal.ac.jp

（短単位）を自動付与し，メタ情報として発話者の属性（性別・年齢層・職業・出身地など）と会話の情報（場面・場所など）を整理し付与している。

本稿では，『現日研・職場談話コーパス』の概要と特徴を述べる。

２．『現日研・職場談話コーパス』の概要

『現日研・職場談話コーパス』は，以下の 2 つの調査研究(a)と(b)より得た談話の文字化テキストを元に作成したものである。

(a)『女性のことば・職場編』

1993 年 9 月から 10 月にかけて，現代日本語研究会が首都圏の有職女性 19 名（20 代～50 代）を調査協力者として，それぞれの職場での自然談話を録音した。録音方法はレコーダーを首から下げたり，近くに置いたりしてもらい，行った。19 人の職場は皆異なる。職場に着いてからの朝の 1 時間，会議・打ち合わせの 1 時間，休憩の 1 時間を録音したうちから，おのおの 10 分前後のまとまった談話を選択し，文字起こしした。その文字起こしデータを収録した CD－ROM と，それにもとづく研究論文 10 本が『女性のことば・職場編』（現代日本語研究会編）としてひつじ書房から刊行された（現在品切れ，下記合本参照）。

(b)『男性のことば・職場編』

1999 年 10 月から 2000 年 12 月にかけて，現代日本語研究会が首都圏の有職男性 21 名（20 代～50 代）を調査協力者として，それぞれの職場での自然談話を録音した。録音方法は前回と同じである。21 人の職場は皆異なる。前回同様に，職場に着いてからの朝の 1 時間，会議・打ち合わせの 1 時間，休憩の 1 時間を録音したうちから，おのおの 10 分前後のまとまった談話を選択し，文字起こしした。その文字起こしデータを収録した CD－ROM と，それにもとづく研究論文 12 本が『男性のことば・職場編』（現代日本語研究会編）としてひつじ書房から刊行された（現在品切れ，下記合本参照）。

2011 年，上記 2 つの書籍内容・CD-ROM データを合わせ，『合本　女性のことば・男性のことば（職場編）』(現代日本語研究会編 2011)としてひつじ書房から刊行され，現在も販売されている[1]。また，引き続き，日常場面での会話の収集・調査が行われている(現代日本語研究会編 2016)。

現在，国語研にて構築を進めている『日本語日常会話コーパス』は，収録者自身に収録機器を預け，自然な談話を収録してもらっているものである。これと同様に，職場での会話を調査協力者自身に録音してもらい，自然な談話を収録するという方法を，すでに 1990 年代にいち早く行っていたということになる。非常に先駆的な試みであると言える。

そこで得られた談話データは，当時よりたいへん画期的なものであると評価されているデータである。これら談話データを分析し，現代日本語研究会編(2011)では，男性語や女性語と呼ばれる，性別の違いにあわせて使われる言葉や，使われる傾向のある言葉は，現状ではそのような区別がなくなりつつあるということが，明らかにされている。また，いわゆる書き言葉とは異なる話し言葉の実態が様々にとらえられている。

[1] 『現日研・職場談話コーパス』の中納言版は，前後最大 300 文字ずつまでが表示可能である。文脈を確認したい場合などは，この CD－ROM に収録されている元データを参照されたい。

２．１　『現日研・職場談話コーパス』のデータ仕様

２．１．１　ファイル名

元データは，『女性のことば・職場編』，『男性のことば・職場編』，それぞれ一つのテキストファイルで提供されている。本コーパスでは，新たに次のファイル命名規則に基づき分割し，ファイル名を付与している。

例：　F　01 A　01 1
　　(1)　(2) (3) (4) (5)

表１　ファイル名の表すもの

(1)	女性／男性	F,M	F:『女性のことば・職場編』出典データという意味
			M:『男性のことば・職場編』出典データという意味
(2)	協力者コード	01,02,…	元データと同じ調査協力者の識別コード
(3)	場面１	A,K,Q	元データの「朝」「会議」「休憩」の別を示す
(4)	場面２	01,02,…	連番: 新規に付与　[場面１か２が変わる毎に付与]
(5)	場所	1,2,…	連番: 新規に付与　[場所が変わる毎に付与]

２．１．２　メタ情報

元データに付与されている項目から，下記の項目を収録している[2]。

◆会話情報◆
＜場面１＞…「朝」「会議」「休憩」の別。
＜場面２＞…「場面１」の細分類。「挨拶」「院生の指導」「客との対応」「雑談」「仕事」など，69 種あり。
＜調査日＞…調査した年月。ただし，『女性のことば・職場編』出典データのみ。『男性のことば・職場編』出典データはすべて「＊」となっている。1999 年 10 月から 2000 年 12 月の間である。
＜場所＞…会話の場所。ただし，『女性のことば・職場編』出典データのみ。「室内」「廊下」「うなぎ屋」「路上」「店先」「店内」「会社内」「不明」。『男性のことば・職場編』出典データはすべて「＊」となっている。
＜会話参加者数＞…1 人から最大 12 人まで。この数値は元データから算出して新たに会話単位に付与したものである。

◆話者情報◆
＜発話者コード＞…発話者の識別コード。元データでは，発話者コードが，『女性のことば・職場編』では「01Ａ」，『男性のことば・職場編』では「０１Ａ」，のように，数字部分に全角と半角とが用いられている。本コーパスでは，発話者コードはすべて半角にした。また，「M01A」や「F01A」のように，先頭に F（『女性のことば・職場編』出典データという意味）あるいは M（『男性のことば・

[2] 元データには，直前文の話者との関係，相手の情報，相手との関係，職場の規模など，『中納言』には収録してない情報も付与されている。それらは，現代日本語研究会編(2011)を参照されたい。

職場編』出典データという意味）を付与して，元データを区別する。「M」や「F」は，話者の性別を表示するものではなく，元データがどちらの出典のものであるかを区別するものであることに注意が必要。なお，元データの発話者コードに含まれている不明を表す全角のクエスチョンは全角のままになっている（「＊」の使用はない）。また，元データにて，数字とアルファベットの組み合わせではなく，「F03 男」「F04 多」「お客①」「他者(女)」などとなっているものは，そのまま用いている。

＜性別＞…発話者の性別。「男」「女」「？」「＊」が入力されている。『女性のことば・職場編』出典データは不明が「？」であり，『男性のことば・職場編』出典データは不明が「＊」である。両出典データともに，個人が特定できていないものは空白である。

＜年齢層＞…発話者の調査当時の年齢層。『女性のことば・職場編』出典データでは，わかる場合は具体的な年齢が入力されている。9 歳から 60 代まである。『男性のことば・職場編』出典データは 10 年刻みになっている。10 代から 70 代まである。両出典データともに，個人が特定できていないものは空白である。

＜職業＞…発話者の職業。53 種ある。「アルバイタ」と「アルバイター」のゆれなども含んでいる。『女性のことば・職場編』出典データは不明が「？」になっている。両出典データともに，未調査は「＊」であり，個人が特定できていないものは空白である。半角は全角にした。

＜職種＞…発話者の職種。83 種ある。職業と重なるものもある。「？」，「＊」，空白は上に同じ。半角は全角にした。

＜役職＞…発話者の役職。53 種ある。ここにも「アルバイト」と「アルバイター」がある。フェイスシートに役職がないことが明示してある場合は「（なし）」と入力。ただし，「（なし：一般職）」は別にある。また，「無」も別にある。「？」，「＊」，空白は上に同じ。

＜出身＞…発話者の出身都道府県。『女性のことば・職場編』出典データは未調査のため，すべて「＊」となっている。『男性のことば・職場編』出典データも不明は「＊」である。

＜最長居住地＞…発話者の 4 歳～15 歳の最長居住都道府県（≒言語形成地）。『女性のことば・職場編』出典データは未調査のため，すべて「＊」となっている。『男性のことば・職場編』出典データも未調査は「＊」である。

２．１．３　会話データ

下記のとおり，元データと異なる点があることに注意を要する。

- 半角は全角にした。（例：That→Ｔｈａｔ）
- 「［名字］さん」における［名字］のように伏せ字された要素は，全体を一つの単位とし，「伏せ字」という品詞を付与している。
- ＜笑い＞＜間 7 秒＞＜咳ばらい＞＜独り言＞など，元データに付与されている言語情報以外の要素については，除外している。
- 元データにある，上昇「↑」，下降「↓」，発話途中で次の話者の始まった時点の「★」，重なった部分の始まり「→」と終わり「←」は，いずれも除外している。

- 元データにある，疑問下降の「？」と，聞き取り不明の「＃」はそのまま残している。
- 「相づちなどの挿入要素」は，包含する発話から独立させ，本来の発話場所とは異なる位置（原則，直後）に記述している。

上記「相づちなどの挿入要素」の処理について補足する。例えば，下記のような相づち（網掛け部分）を含む例の場合，『中納言』の検索結果では，図 1 のように表示される。

例：F01A021 の[元データ]

F01B　やっぱりさ、（うん　Inf(女)）どちらかってゆうとさ、こうやってぱっと見たときさ、（うん　Inf(女)）目ってこっちを見ていない↑

F01A　うん。

前文脈	キー	後文脈
#やっぱりさ、どちらかってゆうとさ、こうやってぱっと見たときさ、目ってこっちを見てい	ない	#うん#うん#うん。

図 1　F01A021 の相づちの挿入要素のある部分の『中納言』の検索結果

図 1 で前文脈と後文脈にある半角の「#」は発話単位区切り記号である。検索結果の画面では，後文脈では「うん」という発話が 3 度繰り返しているようにしか見えない。しかしながら，『中納言』にある「詳細な文脈情報表示」[3]という機能を使うと，「発話者コード」の欄を見ることにより，最初の二つの「うん」は「Inf(女)」の発話であることがわかるため，相づちの挿入要素らしいとあたりをつけることはできるようにはなっている。

詳細な文脈情報

会話 ID	連番	書字形出現形	語彙素読み	語彙素	語彙素細分類	品詞	活用型	活用形	発音形出現形	語種	原文文字列	発話者コード	性別	年齢層	職業	職種	役職	出身	最長居住地
F01A021	2690	目	メ	目		名詞-普通名詞一般			メ	和	目	F01B	女	31	会社員	社長秘書・一般事務	？	＊	＊
F01A021	2700	って	ッテ	って		助詞-副助詞			ッテ	和	って	F01B	女	31	会社員	社長秘書・一般事務	？	＊	＊
F01A021	2710	こっち	コチラ	此方		代名詞			コッチ	和	こっち	F01B	女	31	会社員	社長秘書・一般事務	？	＊	＊
F01A021	2720	を	ヲ	を		助詞-格助詞			オ	和	を	F01B	女	31	会社員	社長秘書・一般事務	？	＊	＊
F01A021	2730	見	ミル	見る		動詞-非自立可能	上一段-マ行	連用形-一般	ミ	和	見	F01B	女	31	会社員	社長秘書・一般事務	？	＊	＊
F01A021	2740	て	テ	て		助詞-接続助詞			テ	和	て	F01B	女	31	会社員	社長秘書・一般事務	？	＊	＊
F01A021	2750	い	イル	居る		動詞-非自立可能	上一段-ア行	未然形-一般	イ	和	い	F01B	女	31	会社員	社長秘書・一般事務	？	＊	＊
F01A021	2760	ない	ナイ	ない		助動詞	助動詞-ナイ	終止形-一般	ナイ	和	ない	F01B	女	31	会社員	社長秘書・一般事務	？	＊	＊
F01A021	2770	うん	ウン	うん		感動詞一般			ウン	和	うん	Inf(女)							
F01A021	2780	うん	ウン	うん		感動詞一般			ウン	和	うん	Inf(女)							
F01A021	2790	うん	ウン	うん		感動詞一般			ウン	和	うん	F01A	女	28	会社員	イベント企画開発	無	＊	＊
F01A021	2800	。		。		補助記号-句点				記号	。	F01A	女	28	会社員	イベント企画開発	無	＊	＊

図 2　F01A021 の相づちの挿入要素のある部分の『中納言』の「詳細な文脈情報表示」

[3] 検索結果の「会話 ID」をクリックして表示する。

２．２.『現日研・職場談話コーパス』のデータ量

はじめに，文字化テキストの例(F01A011 の冒頭)を図 3 に示す。

会話 ID：F01A011
調査日：1993 年 10 月　場面１：朝　場面２：電話　場所　室内
会話参加者数：1
発話者コード：F01A　性別：女　年齢層：28　職業：会社員
職種：イベント企画開発　役職：無　出身：＊　最長居住地：＊

はい、お電話代わりました。
はい、お世話になっております。
はい。
はい。
はい、受け取っております。
はい。
ええ。
ええ。
ええ。
はい。
はい。
わかりました。
じゃ、これーはお受けします。

図 3　F01A011 の冒頭の文字化テキスト

上記のような文字化テキストの全体のデータ量は表 2 のとおりである[4]。

表 2 『現日研・職場談話コーパス』の全体

ファイル数	1,324
会話数	22,372
語数(全て)	248,677
語数(記号等除外・全て)	186,906

以下，次のメタ情報のうち主なものについてのデータ量を示す。

会話情報：　場面１，場面２，調査日，場所，会話参加者数
話者情報：　性別，年齢層，職業，職種，出身，最長居住地

[4] 本コーパスには，『女性のことば・職場編』の＜通番＞1,571～1,718 までの 148 行分の会話は含まない。

２．２．１　会話情報①：場面１，場面２

表 3 に場面１の，表 4 に，場面１と２別の語数(記号等除外・全て)の内訳を示す。以降，語数は「記号等除外・全て」のものを示す。

表 3　場面１のファイル数と語数(記号等除外・全て)

場面１	ファイル数	語数
朝	550	52,773
会議	351	68,613
休憩	423	65,520

表 4　場面２と場面１の語数(記号等除外・全て)

場面２	場面１朝	場面１会議	場面１休憩	語数合計
コンピュータの操作方法の相談と説明		1,495		1,495
シャンプー中の応答	260			260
スタッフルームでの雑談	198			198
パソコン操作の指導と相談			1,049	1,049
ブロー中の応答	1,033			1,033
ミーティング・報告	2,339			2,339
レジでの応答	14			14
挨拶	173	24	78	275
挨拶(電話)	211	25	4	240
院生の指導			2,402	2,402
応対	419	214		633
応対（説明）	556			556
会議		8,099		8,099
客との応対	404	23		427
客との対応	48			48
休憩時雑談	9,986	3	10,492	20,481
教師生徒の会話	6			6
業務電話	44			44
検討会		410		410
研究室会議		3,260		3,260
講義	2,187			2,187
雑談	8,144	3,462	23,750	35,356
雑談（パソコン）			451	451
雑談（パソコンの記憶媒体）			505	505
雑談（レストランの食事）			374	374
雑談（交通規制）			271	271
雑談（自転車）			200	200
雑談（朝食）			68	68
雑談（徹夜）			100	100
雑談（転居）			39	39
雑談(電話)	174			174
仕事	200			200
仕事（応対）		993		993
仕事（相談）	321			321
仕事（打合せ）		172		172

場面２	場面１朝	場面１会議	場面１休憩	語数合計
仕事（打合せ？）	234			234
仕事の話	576			576
仕事上の確認			147	147
仕事中の雑談	1,559			1,559
始業前雑談	353			353
指導		62		62
取引先との電話折衝		866		866
出張報告		2,244		2,244
商品管理業務	1,276			1,276
小会議		10,009		10,009
接客と応答			54	54
相談	3,950	359	1,430	5,739
相談(電話)	238			238
打合せ	10,645	23,643	5,124	39,412
打合せ（商談）		1,459		1,459
打合せ（説明）		2,102		2,102
打合せ(電話)	2,668	12	108	2,788
大会議		3,770		3,770
昼食時雑談			17,967	17,967
昼食時雑談・電話			3	3
朝礼	1,791			1,791
電話	692	225	241	1,158
電話（打合せ）	50			50
電話・雑談	21		43	64
電話・打合せ	1,865		620	2,485
電話依頼	91			91
電話引き継ぎ		50		50
電話取り次ぎ		7		7
電話取り次ぎ(電話)		11		11
独り言	32			32
反省会		1,618		1,618
報告		3,946		3,946
《その他》	15	9		24
《不明》		41		41
語数合計	52,773	68,613	65,520	186,906

表 3 に示したとおり，朝，会議，休憩はおおよそ同じくらいのデータ量である。表 4 に示した場面２の分類は多岐にわたっている。そこで，小磯ほか(2016)で用いている会話の形式 4 タイプ＋そのほかという 5 分類に分類しなおした結果の内訳を次の表 5 に示す。

表 5　場面２の 5 分類と場面１の語数(記号等除外・全て)

会話のタイプ	場面１ 朝	場面１ 会議	場面１ 休憩	語数合計
雑談	20,848	3,514	54,302	78,664
用談・相談	17,015	12,547	3,692	33,254
会議・会合	12,670	52,440	5,124	70,234
授業・レッスン・講演	2,193	62	2,402	4,657
そのほか	47	50		97
語数合計	52,773	68,613	65,520	186,906

表 5 でみると，データ量が多いのは，場面１休憩の雑談と，場面１会議の会議・会合であることがわかる。場面１の合計では，雑談と会議・会合がおおよそ同じくらいのデータ量である。フォーマルな場面とそうでない場面の会話がそれぞれ十分な量とれていることがうかがえる。なお，場面２の再分類の内訳は，表 6 のとおりである。

表 6　場面２の 5 分類

場面２	5分類
コンピュータの操作方法の相談と説明	用談・相談
シャンプー中の応答	用談・相談
スタッフルームでの雑談	雑談
パソコン操作の指導と相談	用談・相談
ブロー中の応答	用談・相談
ミーティンク・報告	用談・相談
レジでの応答	用談・相談
挨拶	雑談
挨拶(電話)	雑談
院生の指導	授業・レッスン・講演
応対	用談・相談
応対（説明）	用談・相談
会議	会議・会合
客との応対	用談・相談
客との対応	用談・相談
休憩時雑談	雑談
教師生徒の会話	授業・レッスン・講演
業務電話	用談・相談
検討会	会議・会合
研究室会議	会議・会合
講義	授業・レッスン・講演
雑談	雑談
雑談（パソコン）	雑談
雑談（パソコンの記憶媒体）	雑談
雑談（レストランの食事）	雑談
雑談（交通規制）	雑談
雑談（自転車）	雑談
雑談（朝食）	雑談
雑談（徹夜）	雑談
雑談（転居）	雑談
雑談(電話)	雑談
仕事	用談・相談
仕事（応対）	用談・相談
仕事（相談）	用談・相談
仕事（打合せ）	会議・会合

場面２	5分類
仕事（打合せ？）	会議・会合
仕事の話	用談・相談
仕事上の確認	用談・相談
仕事中の雑談	雑談
始業前雑談	雑談
指導	授業・レッスン・講演
取引先との電話折衝	用談・相談
出張報告	用談・相談
商品管理業務	用談・相談
小会議	会議・会合
接客と応答	用談・相談
相談	用談・相談
相談(電話)	用談・相談
打合せ	会議・会合
打合せ（商談）	会議・会合
打合せ（説明）	用談・相談
打合せ(電話)	用談・相談
大会議	会議・会合
昼食時雑談	雑談
昼食時雑談・電話	雑談
朝礼	会議・会合
電話	用談・相談
電話（打合せ）	雑談
電話・雑談	用談・相談
電話・打合せ	用談・相談
電話依頼	用談・相談
電話引き継ぎ	用談・相談
電話取り次ぎ	用談・相談
電話取り次ぎ(電話)	用談・相談
独り言	そのほか
反省会	会議・会合
報告	用談・相談
《その他》	そのほか
《不明》	そのほか

表 7　電話か否かの語数(記号等除外・全て)

電話か否か	語数
電話	7,217
電話以外	179,689
語数合計	186,906

本コーパスは，電話の会話も多く収録されている。電話全体の語数は，表 7 に示したとおりである。

２．２．２　会話情報②：場所，会話参加者数

次に，表 8 に場所の，表 9 に会話参加者数のファイル数と語数の内訳を示す。職場を中心に収録されたものであるため，室内での会話がほとんどである。会話参加者数は，3 人のものがもっとも多い。

会話参加者数 1 は，先の表 7 に示した電話の会話のほか，下記の表 10 に示したような，「独り言」(a)や，その場に発話相手はいるが，相手が一言も発しない場合の会話に付与されているもの(b)や(c)である。

表 8　場所のファイル数と語数

場所	ファイル数	語数
室内	598	79,723
廊下	35	2,305
うなぎ屋	8	1,468
路上	7	589
店先	6	664
店内	4	104
会社内	3	393
《不明》	2	6
*	661	101,654
合計	1,324	186,906

表 9　会話参加者数のファイル数と語数

会話参加者数	ファイル数	語数
1	136	11,353
2	270	33,446
3	298	49,160
4	216	28,258
5	138	22,690
6	102	16,750
7	35	2,870
8	46	8,938
9	60	10,843
10	9	1,262
12	14	1,336
合計	1,324	186,906

表 10　会話参加者数 1 の電話以外の会話例

項目	元データの通番	ファイル名	会話データ（元データ）
(a)	1982	F05A011	ここらにおいとけばいいんだな。＜独り言＞
(b)	1983	F05A021	今日［名前］ちゃん来たらさあ、これ、これをさ、こっちで、こっちで、うってもらって、ってゆうか、あいだにいっぱいはいるんでねえ、あたしうち始めちゃってもいいんだけど、＃＃＃が来る、あの、あれ、やらなくちゃいけないからとりあえず。＜間＞
(b)	1984	F05A021	これもやんなくちゃいけない、今日、リーダーも。＜間＞
(b)	1985	F05A021	どれをさきにやるかなあ。＜間＞
(c)	2002	F05A041	ともかくあれをやっちまわないと、うんん。＜咳ばらい＞＜間２８秒＞

表 11　12 人の会話の内訳(M05A071)

発話者コード	会話数
M05B	13
M05E	6
M05F	9
M05L	12
M05M	6
M05P	4
M05Q	7
M05R	4
M05S	3
M05T	2
M05U	6
M05W	9
M05 δ	4
M05 ε	17

表 12　10 人の会話の内訳(M18Q011)

発話者コード	会話数
M18A	60
M18B	15
M18C	8
M18D	1
M18E	16
M18G	2
M18I	2
M18J	32
M18K	1
M18L	1

会話者数最多 12 人の会話は，朝の朝礼時の会話である。また，次に多い 10 人の会話は，休憩の雑談時の会話である。発話者別の会話数の内訳は表 11 と表 12 に示すとおりである。12 人の会話に参加しているのは，M05B から M05W の 12 人である。M05δ と M05ε は，その場にいる人であるが特定できなかったため，別のコードが付与されているものである。10 人の会話のうち，M18L は，元データで「@＜笑い　複数＞」となっている会話であり，『中納言』では除外されている。よって，実質は実は 9 人である。

２．２．３　話者情報：性別，年齢層

最後に，表 13 に性別の，表 14 に年齢層別のファイル数と語数の内訳を示す。男性と女性はおおよそ同じくらいのデータ量である。年齢層は，職場を中心とした収録であるため，20 代から 50 代のデータが多く，最も多いのは 30 代である。

表 13　性別のファイル数と語数

性別	ファイル数	語数
男	596	96,657
女	450	86,419
?	6	343
*	53	692
(空白)	219	2,795
合計	1,324	186,906

表 14　年齢層別のファイル数と語数

年齢層	ファイル数	語数
～9	4	57
10代	11	319
20代	256	48,407
30代	292	51,907
40代	265	48,694
50代	152	27,020
60代	32	4,447
70代	6	463
?	84	2,749
*	3	48
(空白)	219	2,795
合計	1,324	186,906

３．『現日研・職場談話コーパス』の特徴

各種語彙表を用いて,『現日研・職場談話コーパス』の話し言葉としての特徴を概観する。

３．１　上位語

同じく会話が収録されているが，すべて雑談である『名大会話コーパス』と，書き言葉の代表として『現代日本語書き言葉均衡コーパス』（以下,『BCCWJ』）の語彙表より，上位語を比較する。比較の結果を表 15 に示す。

表 15　『名大会話コーパス』『現日研・職場談話コーパス』『BCCWJ』の上位語

順位	名大会話コーパス			職場会話コーパス			BCCWJ		
	語彙素読み	語彙素	品詞	語彙素読み	語彙素	品詞	語彙素読み	語彙素	品詞
1	ダ	だ	助動詞	ダ	だ	助動詞	ノ	の	助詞-格助詞
2	ウン	うん	感動詞-一般	ノ	の	助詞-準体助詞	ニ	に	助詞-格助詞
3	タ	た	助動詞	テ	て	助詞-接続助詞	テ	て	助詞-接続助詞
4	テ	て	助詞-接続助詞	ネ	ね	助詞-終助詞	ハ	は	助詞-係助詞
5	ネ	ね	助詞-終助詞	**デス**	**です**	**助動詞**	ダ	だ	助動詞
6	ノ	の	助詞-準体助詞	ノ	の	助詞-格助詞	ヲ	を	助詞-格助詞
7	カ	か	助詞-副助詞	タ	た	助動詞	タ	た	助動詞
8	ト	と	助詞-格助詞	ハ	は	助詞-係助詞	スル	為る	動詞-非自立可能
9	ノ	の	助詞-格助詞	ニ	に	助詞-格助詞	ガ	が	助詞-格助詞
10	モ	も	助詞-係助詞	ト	と	助詞-格助詞	ト	と	助詞-格助詞
11	デ	で	助詞-格助詞	ガ	が	助詞-格助詞	デ	で	助詞-格助詞
12	ガ	が	助詞-格助詞	デ	で	助詞-格助詞	モ	も	助詞-係助詞
13	ニ	に	助詞-格助詞	モ	も	助詞-係助詞	イル	居る	動詞-非自立可能
14	ハ	は	助詞-係助詞	イウ	言う	動詞-一般	マス	ます	助動詞
15	ソウ	そう	副詞	ヨ	よ	助詞-終助詞	ノ	の	助詞-準体助詞
16	イウ	言う	動詞-一般	スル	為る	動詞-非自立可能	アル	有る	動詞-非自立可能
17	ッテ	って	助詞-副助詞	ソウ	そう	副詞	**デス**	**です**	**助動詞**
18	**ナニ**	**何**	**代名詞**	テル	てる	助動詞	イウ	言う	動詞-一般
19	テル	てる	助動詞	ウン	うん	感動詞-一般	コト	事	名詞-普通名詞-一般
20	ヨ	よ	助詞-終助詞	カ	か	助詞-副助詞	ナイ	ない	助動詞

表 15 において赤字が,『名大会話コーパス』と『現日研・職場談話コーパス』にあり,『BCCWJ』にはない。これらは話し言葉としての特徴を表す語と言えるだろう。また,『名大会話コーパス』の 18 位「何」は,『現日研・職場談話コーパス』では 28 位であり,『BCCWJ』では 62 位であることから，これも話し言葉としての特徴を表し得ると考えられる。一方,『BCCWJ』で 17 位である「です」は,『名大会話コーパス』にないが,『現日研・職場談話コーパス』では 5 位と非常に上位に位置する点が目立つ。雑談以外のフォーマルな会話を多く含む『現日研・職場談話コーパス』の特徴を表すと言えそうである。

３．２　LLR（対数尤度比）

『BCCWJ』の書籍の出版サブコーパス（PB）と，特定サブコーパスの一つ yohoo!知恵袋（OC),『名大会話コーパス』，主に学会講演が収録されている『日本語話し言葉コーパス』（CSJ）をベースに，『現日研・職場談話コーパス』の LLR（対数尤度比）により，

特徴語を求めた。表 16 に上位 10 語を示す。感動詞が多く入っていることがわかる。

また，『名大会話コーパス』をベースにした際に 3 位に入った「ゼロ」が目を引く。語彙素読み「ゼロ」の用例は全部で 406 件あり，ほとんどが数字の「０」の用例であった。数字の話が多いというのも，『現日研・職場談話コーパス』の特徴の一つかと推測される。また，書字形「ゼロ」の用例は全部で 18 件あった。そのうち 8 例を表 17 に示す。これらも職場の会話らしい用例と言えそうなものである。

表 16　『現日研・職場談話コーパス』の LLR 上位語

ベース	1	2	3	4	5	6	7	8	9	10
BCCWJ（PB）	ね	うん	あの	はい	てる	って	です	よ	で	そう
BCCWJ（OC）	うん	あの	はい	ね	そう	あっ	で	ええ	えー	ああ
名大会話コーパス	です	ます	ゼロ	はい	えー	あの	えーと	此れ	御早う	が
CSJ	よ	うん	あっ	ああ	はい	まあ	えー	ね	さ	の

表 17　『現日研・職場談話コーパス』の LLR 上位語「ゼロ」の用例

会話 ID	前文脈	キー	後文脈
F03A021	#うん#ここを	ゼロ	にして。#うん、あ、ここに、
F05K011	#お金はかからないわね。#	ゼロ	です。#うん
F06K011	あの、キャリアーとしての経験は	ゼロ	ですからー、
F14Q021	#それで、あのー、実施は	ゼロ	だった人がなん人かいたんですね。
F14Q021	#それーはー。#あの前期まったく	ゼロ	だったんで。#２、３人いるんですけど、
F14Q021	#でなんか［名字］さんもなんかまったく	ゼロ	だとなんかちょっといいわけー、ってゆうか
F14Q021	あの出勤簿とかもまったく空欄になりますよね、ぜ、	ゼロ	とかじゃなくてね。#わかりました。＃＃＃。
M12Q101	#ひさびさに	ゼロ	から組んだけど、＃＃＃＃。

３．３　品詞の分布

続いて，同じく『BCCWJ』の語彙表を用いて，『現日研・職場談話コーパス』と品詞の分布を比較する。それぞれの品詞の分布を図 4 と図 5 に示す。

図 4 と図 5 を比べると，赤の四角形で強調している通り，『現日研・職場談話コーパス』は名詞が少なく，感動詞，副詞，代名詞が多い。これは，『名大会話コーパス』と『BCCWJ』の比較(柏野ほか 2017)と同じ傾向であり，話し言葉としての特徴が表れていると考えらえる。

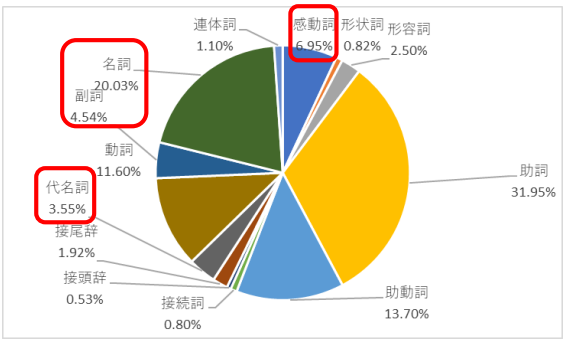

図 4　『現日研・職場談話コーパス』の品詞の分布

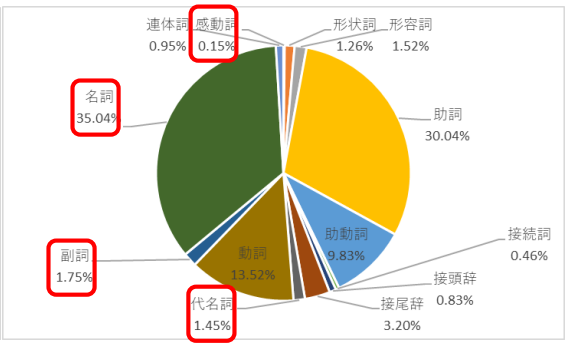

図 5　『BCCWJ』の品詞の分布

３．４　そのほかの特徴語

柏野ほか(2017)では，『BCCWJ』では用例が得にくいが，『名大会話コーパス』で頻出することが期待されるそのほかの特徴語として，表 19 に示す俗語的な用法のある a)〜h)の 8 語を取り上げた。今回，『現日研・職場談話コーパス』について調べた結果と比較して，表 19 にまとめて示す[5]。なお，改めて両コーパスの主な違いを表 18 に示す。

表 18　『現日研・職場談話コーパス』と『名大会話コーパス』の相違点

	『現日研・職場談話コーパス』	『名大会話コーパス』
語数	186,906	1,131,971
収録期間	1993-2000	2001-2003
会話の形式	雑談/用談・相談/会議・会合/ 授業・レッスン・講演/ほか	雑談のみ

表 19　『現日研・職場談話コーパス』と『名大会話コーパス』の用例検索結果

項目	語	『職場』 用例数	『職場』 ＰＭＷ	『名大』 用例数	『名大』 ＰＭＷ	検索方法
a)	微妙	1	5.4	156	109.8	語彙素「微妙」
b)	やば	14	74.9	168	118.2	文字列「やば」
c)	まじ	3	16.1	197	138.6	語彙素「まじ」
d)	無理	33	176.6	273	192.1	語彙素「無理」
e)	てか、	3	16.1	60	42.2	文字列「てか、」
f)	すごい＋形容詞	10	53.5	344	242.0	書字形出現形「すごい」＋形容詞
g)	うける	2	10.7	37	26.0	職場：文字列「うける」 名大：語彙素「受ける」の終止形
h)	みたいな	30	160.5	473	332.8	文字列「みたいな[、。？]」

表 18 に示したとおり，延べ語数は，『現日研・職場談話コーパス』は『名大会話コーパス』の約 7 分の 1 である。そこで，表 19 での用例数の比較は PMW を用いる。a)〜h)の 8 語は話し言葉のなかでも俗語的な言い方であるため，雑談以外の会話が収録されている『現日研・職場談話コーパス』では，全体的に『名大会話コーパス』よりも少ない値になっている。また，少しだけ『現日研・職場談話コーパス』の収録時期が『名大会話コーパス』よりも前になるため，今ある俗語的な用法は，まだ『現日研・職場談話コーパス』にはそうなかったのかもしれない。

たとえば，次の表 20 に実際の用例を示すが，a)「微妙」は，俗語的ではない用法の用例が 1 例あるのみである。俗語的には応答で「それ，びみょー」などと言うのを聞くが，平仮名表記，カタカナ表記でも『現日研・職場談話コーパス』に該当例はなかった。g)「受ける」も同様である。やはり俗語的に応答で「それ，うけるー」などと言うのを聞くが，その該当例はなかった。なお，両コーパスともに，d)「無理」の用例も，現在ほどの俗語的な，応答詞的に用いるような例はみられない。

[5] 表 19 の検索結果数は，当該語の「話し言葉」ならではの用法例の正確な件数ではない。検索もれ，あるいは，別語，別用法の例が少々混じっている。

ただ，それ以外では数は少ないながらも，『名大会話コーパス』と同様に，現在耳にするような俗語的な言い方の用例が得られている。表 20 に示す。なお，会話 ID が「data」から始まるものは『名大会話コーパス』の用例である。

表 20　『名大会話コーパス』と『現日研・職場談話コーパス』の用例

会話 ID	前文脈	キー	後文脈
data077	この子は、Ｅ短の子だよ。あっ、そうなんだー、	微妙	、微妙。うんそういうのばっかり。
M06Q031	まー、驚くことが多いって話よー。	微妙	なニュアンスで教えてくれてー。
data072	なかなか時間がないんだよね。ねーあたしもだよ。	やばー	い。あっ、ＴＯＥＩＣさ、こないだあったけど、
M12Q031	#あの、これやうめーや、ちょっと、	やばい	んじゃない。#このランチメニューー。
data011	５級だったしね、一番最初受けたの。	まじ	？６年のときに５級。
M21Q011	３時までさー、ずっーとしゃべってて。#	まじ	でー。#まじ、もーあたし、その前の日とかに、
data046	うーん。無理。だから、なんで。とにかく	無理	。それは、そういうことしたら、
M21K011	#夜は	無理	っす、平日の夜は無理っす。
data056	うんどこで見たの。	てか、	あの、日本に来たときの。
M12Q101	#うん。#いや、	てか、	自動なんだよー、もー。
data103	んー、かっこいい。＊＊＊。	すごい	(かわいい)、この絵。
F11Q011	#すごい、なんだっけそれ。#	すごい	(おいしい)やつ。
data065	あ、君は日本文学専攻か、ふーん、とか言って。	受ける	ー。うーん話しかけやすい雰囲気なんじゃん。
F15K011	#だって、みんな、	うける	ものねー、あれ、すごく。#うけますねー。
data085	今日、さむ。な、何となく寒そう、	みたいな？	うん。雪が降ったときとか、
F15Q011	#でしょ#でビール、がーん、みたいな	みたいな。	#もっとほかのお母さまの意見も聞いた方が＃＃＃。

遠藤(2011)は，俗語，若者ことば，流行語として次の語の用例が『男性のことば・職場編』にあると指摘している。

やばい，おかまちっく，ぽい（の新しい用法）[6]，まじ・まじで

6 「ぽい」は本来名詞や動詞の連用形につくが，形容詞や動詞の終止的につく新しい用法例を指摘している。「かっこいいっぽい」「コピーしてるっぽい」の例があげられている。

このうち，「やばい」については用例があるということのみ指摘しているが，表 20 に示したとおり，ランチに対して「うめー」「やばい」と発話しており，すでにポジティブの「やばい」の用例が収録されている点は注目に値する。

さらに，「今どきの日本語」については，遠藤編(2018)にくわしく報告されている。

４．おわりに

『現日研・職場談話コーパス』の概要と特徴を述べた。書き言葉コーパスである『BCCWJ』と比べ，雑談を収録した『名大会話コーパス』同様に，「うん」，終助詞の「ね」「か」「よ」や，副詞の「そう」が頻出し，また，感動詞，副詞，代名詞が多く，話し言葉の特徴語の用例が多く得られるコーパスであることを示した。さらに，『名大会話コーパス』が雑談のみであるのに対し，『現日研・職場談話コーパス』は用談・相談，会議・会合，授業・レッスン・講演などの会話を含むため，「です」が頻出する点が『名大会話コーパス』とは異なる特徴であることを示した。そして，LLR を調べることにより，「ゼロ」のような仕事上の会話ならでは現れやすい語があることを示した。また，俗語的なものは，『名大会話コーパス』ほどではないが，ある程度収録されていることも示した。

『中納言』で公開するに際し，貴重な『現日研・職場談話コーパス』のデータが，今後さらにさまざまな研究に活用されることが期待される。

謝　辞

本研究は国立国語研究所の機関拠点型基幹研究プロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」（プロジェクトリーダー：小磯花絵）の研究成果を報告したものです。また，オリジナルのデータは，現代日本語研究会による研究成果です(現代日本語研究会編 2011)。遠藤織枝先生，髙﨑みどり先生，髙橋美奈子先生をはじめとする現代日本語研究会のみなさまと，出版元のひつじ書房に感謝申し上げます。そして，『中納言』版の構築と公開に際しては，形態素解析結果の人手修正をはじめ，多くのみなさまにご協力いただきました。みなさまに感謝いたします。

文　献

遠藤織枝(2011)「第２章　男性のことばの文末」現代日本語研究会編『合本　女性のことば・男性のことば（職場編）』pp.33-45, ひつじ書房.

遠藤織枝編(2018)『今どきの日本語−変わることば・変わらないことば』ひつじ書房.

柏野和佳子・西川賢哉・小磯花絵(2017)「『名大会話コーパス』中納言版・ひまわり版公開データの作成」『言語資源活用ワークショップ 2016 発表論文集』pp.324-335.

現代日本語研究会編(2011)『合本　女性のことば・男性のことば（職場編）』ひつじ書房.

現代日本語研究会編(2016)『談話資料 日常生活のことば』ひつじ書房.

小磯花絵・土屋智行・渡部涼子・横森大輔・相澤正夫・伝康晴(2016)「均衡会話コーパス設計のための一日の会話行動に関する基礎調査」『国立国語研究所論集』10, pp.85-106.

小磯花絵・天谷晴香・居關友里子・臼田泰如・柏野和佳子・川端良子・田中弥生・伝康晴(2018)「『日本語日常会話コーパス』構築状況と予備的分析」『言語処理学会第 24 回年次大会発表論文集』pp.889-892.

藤村逸子・大曽美恵子・大島ディヴィッド義和(2011)「会話コーパスの構築によるコミュニケーション研究」藤村逸子・滝沢直宏編『言語研究の技法：データの収集と分析』pp. 43-72, ひつじ書房.

関連 URL

国立国語研究所共同研究プロジェクト「大規模日常会話コーパスに基づく話し言葉の多角的研究」 http://pj.ninjal.ac.jp/conversation/

『現日研・職場談話コーパス』 http://pj.ninjal.ac.jp/conversation/shokuba.html

『名大会話コーパス』 https://nknet.ninjal.ac.jp/nuc/templates/nuc.html
http://pj.ninjal.ac.jp/conversation/nuc.html

コーパス検索アプリケーション『中納言』 https://chunagon.ninjal.ac.jp/

『現代日本語書き言葉均衡コーパス』語彙表
http://pj.ninjal.ac.jp/corpus_center/bccwj/freq-list.html