『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・大村舞・西川賢哉・小磯花絵　（国立国語研究所）

背景と目的

- 『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』として公開した。『現日研・職場談話コーパス』中納言版の概要を報告する。

データ量

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

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

データ情報：性別，年代

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

- 男性、女性、ほぼ同じくらい。
- 20代～40代が多い。

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

品詞の分布の比較

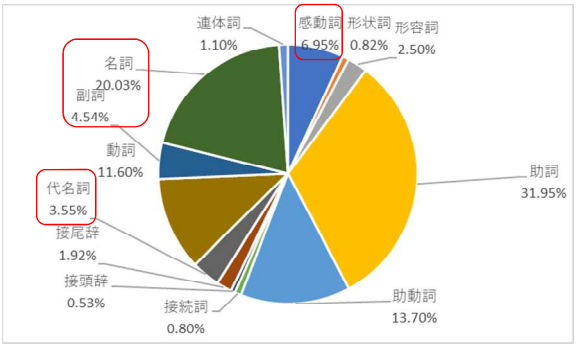

図4 『現日研・職場談話コーパス』の品詞の分布

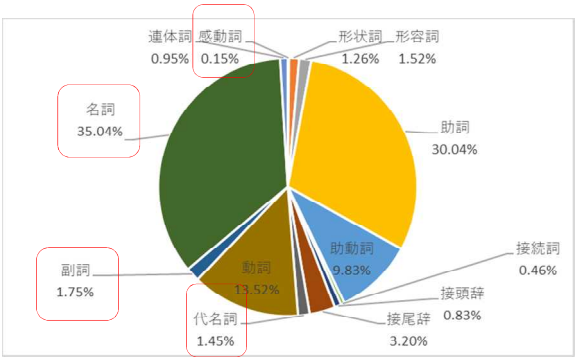

図5 『BCCWJ』の品詞の分布

中納言にて8月20日公開！

- 国立国語研究所で開発されたコーパスを検索することができるWebアプリケーション。
- 短単位と文字列（と長単位）によって，コーパスに付与された形態論情報を組み合わせた高度な検索を行うことができる。

https://chunagon.ninjal.ac.jp/

（利用には申請が必要，無料）

上位語の比較

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 語彙素読み	名大会話コーパス 語彙素	名大会話コーパス 品詞	職場会話コーパス 語彙素読み	職場会話コーパス 語彙素	職場会話コーパス 品詞	BCCWJ 語彙素読み	BCCWJ 語彙素	BCCWJ 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

中納言の検索例

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。
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＃＃＃。

まとめ

- 『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。
- 『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。

© 2018 NINJAL