

『職場編』作成側から見る『中納言版』

2018『現日研・職場談話コーパス』公開記念シンポジウム：
パネル討論：今後の研究の可能性について

1

小林美恵子

現代日本語研究会・早稲田大学

2018・9・3

『職場編』データについて

- 収録した録音データを文字化し、データベース化したもの
- 1発話（文）を1行とし、各行に場面・談話内容・話者情報などを加えてデータベース化した（遠藤・発表20180903）
- TEXTファイルCD-ROMによって提供し、利用者がExcelなどで読み込めるようになっている
→ 現代日本語研究会（2011）『合本女性のことば・男性のことば（職場編）』ひつじ書房
- 検索・絞り込みについては、各列の文字列によって行うが、絞り込んだキー（を含む行）をマークすることにより、そのキーの含まれた文字列の前後を含む流れや、その場面全体の位置づけなどを談話の流れとして把握することも可能である

『中納言版』にできること①

- ▶ 『職場編』元データに形態素解析を施してある
- ▶ 短単位検索によりキーとなる語句の品詞や後方共起する語の指定などの組み合わせにより、『職場編』の単純な文字列検索ではできなかった検索が可能になった
- ▶ 尾崎（1997（2011） 本田（発表20180903）にて詳述）

女性専用文末形式とされる「わ（よ）（ね）」-対立形式「∅（よ）（ね）」を抽出することの困難さから、「形容詞（い）+わ（よ）（ね）」「形容詞型の助動詞（ない・たい・らしい）+わ（よ）（ね）」によって「わ（よ）（ね）」使用の全体像を類推した



『中納言版』は「動詞+わ（よ）（ね）」も「形容詞+わ（よ）（ね）」も容易に検索できる→総体的な使用頻度・出現頻度などの実数把握が容易にできる

『中納言版』にできること②

- キーの前後文脈の語数を指示することにより、そのキーがどのような文脈で使われているか 見当をつけられる

キーが抽出されても、それを含む1文（行）でしか見ることができず、談話の流れでどのような位置にあるかは、キーをマークしたうえで、いったん元の談話の流れに戻して把握するしかなかった

『職場編』

『中納言版』には「できない」こと

- ▶ 見当はあくまでも見当である
 - ・ 抽出されたキーの発話者は話者情報によって確認できる
 - ・ 前後の発話については話者情報は示されず確認しにくい→「詳細な文脈情報」へ
 - ・ 初期検索画面、「詳細な文脈情報」ともに<笑い><間><咳払い>、アクセントの上昇下降などを示す「↑」「↓」は示されない、会話の重なるのマークも示されない
 - ・ 発話中に含まれる他の談話参加者のあいづちも、発話から独立させた形であいづちのみが示され、あいづちの位置は特定できない（「詳細な文脈情報」では発話者のみ特定できる）
- ▶ 各発話の会話情報（場面・場所・参加人数など）話者情報（発話者コード・性別・年齢）などからの絞り込み・検索はできない

『中納言版』の検索はサンプルである

- キーとして抽出された文字列や、言語形式をサンプルとして、出現頻度を数量的に計量する
- 検索によって明示される項目の出現順はランダムである
 - ・ 検索ヒットが500件を超える場合は、ランダムに選ばれた500件のみを表示する
 - ・ 500件以下の場合も、検索するたびに表示される項目の順番は入れ替わる
- 並べ替えや二次的な検索を行う場合は、検索されたものをダウンロードし、Excelで読み込んだものによって行う。
 - ・ ・ ・ これはあくまでもそのキーに関する検索結果項目のデータであり、ここから『職場編』全体のデータに戻ることはできない
- 『中納言版』からは元データの全体を見渡すことはできない

『中納言版』と『職場編』

- ▶ ある言語形式の出現頻度、使用量、前後の語との接続関係の頻度などを検索し数量的に算出するというタイプの研究

『中納言版』（+他の大規模コーパスなど）が力を発揮！

- ▶ 談話の流れの中である言語形式や言語形式以外の要素などの出現状況を明らかにし例示するというタイプの研究 = ある一定以上の分量の談話の流れの中で言語形式などを見ていく場合

『中納言版』 + 『職場編』元データで！

『中納言版』は『職場編』によって補完されることにより、より力を発揮し、
『職場編』は『中納言版』の出現により、言語データとして分析の可能性が増し、
研究されることによりデータの寿命が延びることになる？

『中納言版』に望むこと

- ▶ 『中納言版』から『職場編』に「戻る」ために
 - ・ 「開始番号」と「連番」と「行番号」（柏野ほか2018 では「通番」）
 - ・ 『中納言版』の各行が『職場編』のどの行に一致するのかがわかるようなナンバーがほしい
- ▶ 形態素解析の問題点
 - ・ ミス（あるいは理由不明？の何らかの基準が働いている？）の検出がしにくい
 - ・ 発見された場合、どのように修正していけるのか
 - ・ 文字列検索「やろ」30件 「動詞 意志推量形/命令形」「未知語」「助動詞」判定に混乱がある？
短単位検索「書字出現形」「えー」1012件 + 「形容詞」47件（良い）
うち3件の「えーやん」を除く44件は「感動詞」とみるべき？

解析ミス？の一部

●は対照する正例 前後文脈については意味の取れる区切りまで

| 連番 | 前文脈 | キー | 後文脈 | 語彙素読み | 語彙素 | 品詞 | 活用法 | 活用形 |
|----|--------|--|-----|-----------------------------------|-----|----|-----------|----------------|
| ① | 20080 | もうえー、もうえー | やろ | 。さー、準備しよう。 | ヤル | 遣る | 動詞-非自立可能 | 五段-ラ行 命令形 |
| ② | 18730 | いや、こってり | やろ | 、ラーメンは。 | ヤル | 遣る | 動詞-非自立可能 | 五段-ラ行 命令形 |
| ③ | 8410 | 取引してなかったらー、向こうもーたぶんゆうてきーへん | やろう | しー。#うん#たとえば出してってゆわれてもー、もーないですよ、って | ヤル | 遣る | 動詞-非自立可能 | 五段-ラ行 意志推量形 |
| ④ | ●22380 | ゆうたことやらんやつは最低やでー。 [ニックネーム] 恐いのー。あたりまえ | やろー | 。[ニックネーム]、[ニックネーム]、きのうはだれと行ったのー。 | ヤ | や | 助動詞 | 助動詞-ヤ 意志推量形 |
| ⑤ | 230 | なんか編集部のほうでさあ、あーの、女性だけで、なんか | やろー | かってゆう話になってんだけど来るま、もし時間があつたら。 | | | 未知語 | |
| ⑥ | 2080 | 渋谷でー、忘年会。あ、そうだ忘年会。忘年会は別に | やろー | 。忘年会だからいいよ、ついたちの日でもー、やっちゃおうよ。 | ヤ | や | 助動詞 | 助動詞-ヤ 意志推量形 |
| ⑦ | ●370 | すべ、すべてをいっしょになんか | やろう | とするからー。んー。 | ヤル | 遣る | 動詞-非自立可能 | 五段-ラ行 意志推量形 |
| ⑧ | 4140 | そろそろ、お話ししなきゃいけないんでー、 | えー | 。えー、ちょっとご連絡ください###。 | ヨイ | 良い | 形容詞-非自立可能 | 形容詞 終止形-一般 |

(参考資料)

尾崎喜光 (1997 (2011))

「第2章 女性専用形式の今」 現代日本語研究会編『女性のことば・職場編』
pp. 33-58 ひつじ書房 (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)』ひつじ書房)

柏野和佳子・大村舞・西川賢哉・小磯花絵 (2018)

『現日研・職場談話コーパス』中納言データの作成 『言語資源活用ワークショップ
2018発表論文集』