

Web 資料 2

語種構成比率の変遷の調査方法

第 7 章の**解説 1**の「1. 語種から見た日本語の歴史」(pp.142~143)に示した、図 7.2「『日本語歴史コーパス』による自立語の語種構成比率の推移(異なり語数)」は、『日本語歴史コーパス』の開発者用のツールである「大納言」を用いた、平安時代、室町時代、明治・大正時代を対象とした、やや大がかりな語彙調査の結果(田中 2017)に基づいている。同じく、「1.3 漢語の減少とその背景」(pp.144~146)に示した、図 7.3「『明治・大正編』に含まれる雑誌による語種構成比率の推移(異なり語数)」は、明治・大正時代の雑誌を対象とした語彙調査の結果(田中 2015)によっている。

このような「大納言」を使った大がかりな語彙調査ではなく、各時代から一部の資料を選んで、「中納言」を用いて個人が調査を行うことでも、語種構成比率の変遷はとらえることができる。その例として、奈良時代、平安時代、室町時代、江戸時代、明治時代の口語体資料から、次の【調査対象】に対して、語種構成比率の調査を行う手順を【調査方法】に示す。【調査方法】の示した Excel の操作については、本冊の付録に掲載した「8.4 語種構成比率の調査」(p.171~172)も参照してほしい。

【調査対象】

奈良時代：万葉集、平安時代：枕草子、室町時代：虎明本狂言集-脇・大名類、江戸時代：洒落本-江戸作品、明治時代：明治初期口語-4 作品(安愚楽鍋、開化問答、交易問答、文明開化)

※鎌倉時代を対象にしなかったのは、この時代に口語体の文献が存在しないため、文体を揃えて時代間比較を行うことができないためである(→第 1 章「3. 『日本語歴史コーパス』とは」p.2~7)。また、各時代で選択した文献は、総語数が、「中納言」のダウンロード上限である 10 万件に達しない範囲で選んだ。

【調査方法】

- ① 「中納言」で、「検索対象の選択」から、『万葉集』を選ぶ。
- ② 「短単位検索」で、「語彙素」に「%」(半角)を入力し、「検索結果をダウンロード」のボタンを押す。

- ③ ②でダウンロードしたデータを Excel で開き、「挿入」→「ピボットテーブル」を選択して「ピボットテーブルのフィールド」のところで、「行」と「値」のボックスに「語種」をドラッグする。この操作でできたテーブルが、『万葉集』の延べ語数による語種の集計である。
- ④ ②でダウンロードしたデータを Excel で開き、「挿入」→「ピボットテーブル」を選択して「ピボットテーブルのフィールド」のところで、「行」のボックスに「語彙素 ID」「語彙素読み」「語彙素」「語種」を、「値」のボックスに「語種」をドラッグする。
- ⑤ 「デザイン」→「レポートのレイアウト」から、「表形式で表示」と「アイテムのラベルをすべて繰り返す」を選択する。
- ⑥ ⑤によって作成されたテーブル全体をコピーし、その値のみを、新しいワークシートに貼り付ける。これが、編集可能な「語彙頻度表」である。
- ⑦ ⑥で作成した「語彙頻度表」をもとに、「挿入」→「ピボットテーブル」を選択して「ピボットテーブルのフィールド」のところで、「列」「値」のボックスに「語種」をドラッグする。この操作でできたテーブルが、『万葉集』の異なり語数による語種の集計である。
- ⑧ 『枕草子』、『虎明本狂言集』-脇・大名類、「洒落本」-江戸作品、「明治初期口語」-4 作品についても、①～⑦の操作を行う。
- ⑨ 五つの文献の語種集計の結果を、一つのワークシートにまとめ、帯グラフを作成して、時代的推移を確認する。

調査結果

調査の結果得られた集計表とグラフは、延べ語数は、表 1、表 2、図 1、異なり語数は、表 3、表 4、図 2 の通りである。延べ語数においても、異なり語数においても、室町時代の虎明狂言と、江戸時代の洒落本との間には、ほとんど差がないが、それ以外は、時代が進むに従って、和語の比率が減少し、漢語の比率が増加する傾向のあることが確認できる。また、外来語の比率は、いずれの資料でもきわめて低い、低い数値のなかでも、延べ語数においても異なり語数においても、室町時代と江戸時代の間には差があり、異なり語数においては、江戸時代と明治時代の間には差があり、その時期に増加することがわかる。

表 1 語種別の語彙統計（延べ語数の実数）

語種	万葉集	枕草子	虎明狂言_脇・大名	洒落本_江戸	明治初期口語_4 作品
和語	95743	62650	57783	67431	53627
漢語	29	2590	5508	6583	8974
外来語	2	17	8	152	146
混種語	6	307	2014	1560	1124
合計	95780	65564	65313	75726	63871

表 2 語種別の語彙統計（延べ語数の比率）

語種	万葉集	枕草子	虎明狂言_脇・大名	洒落本_江戸	明治初期口語_4 作品
和語	100.0%	95.6%	88.5%	89.0%	84.0%
漢語	0.0%	4.0%	8.4%	8.7%	14.1%
外来語	0.0%	0.0%	0.0%	0.2%	0.2%
混種語	0.0%	0.5%	3.1%	2.1%	1.8%
合計	100.0%	100.0%	100.0%	100.0%	100.0%

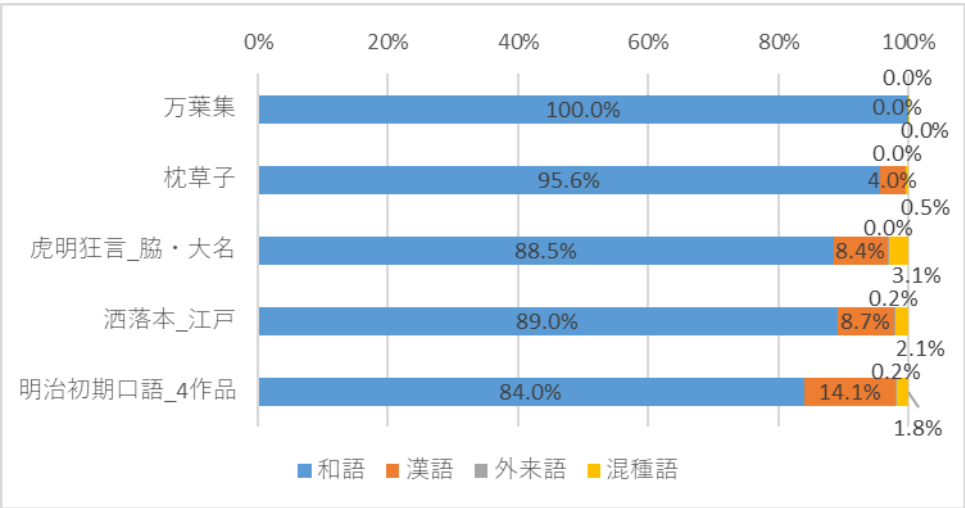


図 1 語種構成比率の推移（延べ語数）

表 3 語種別の語彙統計（異なり語数の実数）

語種	万葉集	枕草子	虎明狂言_脇・大名	洒落本_江戸	明治初期口語_4 作品
和語	4353	3155	2728	4652	3852
漢語	24	530	984	1561	2497
外来語	2	9	4	30	55
混種語	5	78	104	365	293
合計	4384	3772	3820	6608	6697

表 4 語種別の語彙統計（異なり語数の比率）

語種	万葉集	枕草子	虎明狂言_脇・大名	洒落本_江戸	明治初期口語_4 作品
和語	99.3%	83.6%	71.4%	70.4%	57.5%
漢語	0.5%	14.1%	25.8%	23.6%	37.3%
外来語	0.0%	0.2%	0.1%	0.5%	0.8%
混種語	0.1%	2.1%	2.7%	5.5%	4.4%
合計	100.0%	100.0%	100.0%	100.0%	100.0%

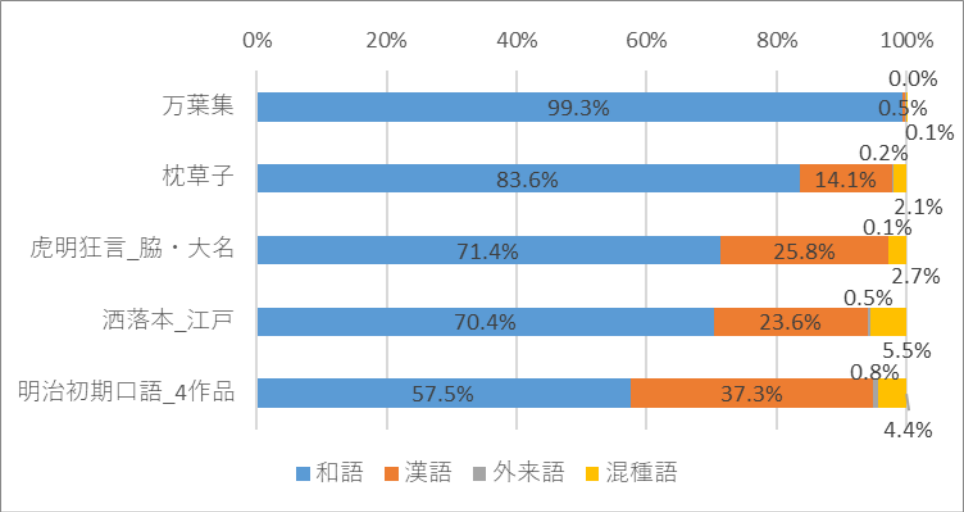


図 2 語種構成比率の推移（異なり語数）

参考文献

田中牧郎（2015）「明治後期から大正期に基本語化する語彙」（斎藤倫明・石井正彦編『日本語語彙へのアプローチ』（おうふう）

田中牧郎（2017）「日本語史」（計量国語学会編『データで学ぶ日本語学入門』朝倉書店）