

目録データの用語分析結果について

国立公文書館業務課

上記は「国立公文書館デジタルアーカイブ・システムにおける辞書サブシステム登載データ作成に係る調査・検討等業務作業報告書」(平成 16 年度委託調査)より抜粋したものである。章立ては関連する部分のみの構成とした。

調査対象目録データ件数は、昭和元年から 29 年までの下記の公文書類の目録データ約 10 万件である。

公文類聚：主に閣議案件

公文雑纂：公文類聚以外に内閣の授受した文書

御署名原本：法律勅令等の公布原本

枢密院会議筆記：枢密院会議の議事録関係

時代	公文類聚	公文雑纂	御署名原本	枢密院会議筆記	合計(件)
S1-20	19,560	40,539	12,972	1,046	74,117
S21-25	7,736	1,379	4,484	158	13,757
S26-29	7,069		2,957		10,026
	34,365	41,918	20,413	1,204	97,900

1. 目録データの事前整理

1.1. 目録データの重複の整理

作業開始後に判明したことだが、一定の年代をへだてて同一の目録が出現している。そこで、重複を一つにまとめてみたところ、97900件 78255件(80%)になった。つまり、20%の目録は除外することができた。なお、以下のように複数の資料群間で重複が認められた。(は、重複しているもの)

類 02993100 : 運輸省各局長及び鉄道局長の命免に関する件

類 02985100 : 運輸省官制の一部を改正する

類 02985100 : 運輸省官制の一部を改正する

類 03050100 : 運輸省官制の一部を改正する

御 29828100 : 運輸省官制の一部を改正すること等に関する勅令

類 02985100 : 運輸省官制の一部を改正すること等に関する勅令を定める

類 03105100 : 運輸省官制の一部を改正する政令

類 03105100 : 運輸省官制の一部を改正する政令

類 03105100 : 運輸省官制の一部を改正する政令

御 30811100 : 運輸省官制の一部を改正する政令

御 30855100 : 運輸省官制の一部を改正する政令

御 30865100 : 運輸省官制の一部を改正する政令

御 29888100 : 運輸省官制の一部を改正する勅令

御 30140100 : 運輸省官制の一部を改正する勅令

御 30297100 : 運輸省官制の一部を改正する勅令

類 03340100 : 運輸省官制の一部を改正する等の政令

御 31970100 : 運輸省官制の一部を改正する等の政令

類 03312100 : 運輸省官制の一部を改正する等の政令案及び気象官署官制の一部を改正する政令案撤回の件

類 03420100 : 運輸省官制の一部を改正する等の政令案等の撤回について

御 30041100 : 運輸省官制の一部を改正する等の勅令

類 02985100 : 運輸省官制の一部を改正する等の勅令を定める

類 03204100 : 運輸省官制の一部を改正する法律

御 31108100 : 運輸省官制の一部を改正する法律

類 03312100 : 運輸省官制の一部を改正する法律案

類 02911100 : 運輸省官制外三勅令中ヲ改正ス

御 29066100 : 運輸省官制外三勅令中改正ノ件

1.2. 改正や改編などに関わる目録の整理

法令や組織などの改正や改編では、制定・創設時の公文書において単語の抽出作業を行っているため、新しい情報はないと思われる。たとえば、以下のような表現を含む目録データは、同様の内容の再出現と推測される。

「衆議院議員選挙法施行令の一部を改正する勅令」のように「の一部を改正する」を含む目録データは 6860 件(7.0%)。

「輸出絹織物検査手数料令、輸出人造絹織物検査手数料令及輸出毛織物検査手数料令ヲ廃止ス」のように「ヲ廃止ス」を含む目録データは 2178 件(2.2%)。

1.3. タイトルの定型を利用した整理

「有島準次郎外二名ヲ鉄道局書記等二任用」のような職員の異動や待遇についての公文書では、辞書に載っていない一般人の姓名が未知語になったり、誤分割されたりすることが多かった。後述するように、抽出語のうち、実に半分以上が人名、特に一般人の姓名であることを考えると、こうした内容のものは、本作業の対象としなくてもよいであろう。試みに、いくつかの表現のパターンを含む目録データをカウントしてみた。

「陸軍歩兵大佐木村政寿同上（枢密院書記官長二上兵治賞与ノ件外一件）ノ件」のように、「賞与ノ件」を含む目録データは 7159 件(7.3%)。

「上田速実高知県視学二再就職追認申請ノ件・昭和八年十月六日付ニテ載ス」のように、「再就職」を含む目録データは 612 件(0.6%)。

「内閣属池之上武雄外二名同上（内閣雇岩永トヨ賞与ノ件）ノ件」のように、「賞与ノ件」を含む目録データは 1300 件(1.3%)。

「東京高等学校教授山口察常同上（企画院委員八木芳之助外五名手当給与ノ件）ノ件」のように、「給与ノ件」を含む目録データは 1122 件(1.1%)。

「鉄道省事務官猪川清同上（鉄道局副参事平野浩吉奏任文官俸給制限外下賜ノ件）ノ件」のように、「下賜ノ件」を含む目録データは 4230 件(4.3%)。

1.4. 目録データの配列

前述したが、調査を開始すると同じような目録データが散在することがわかったため、試みに目録データのソート方法を変えて、精査作業のやりやすさを調査した。

a) 元の配列のまま、b) 件名でソート c) 件名の逆引き（件名を逆転させた文字列）でソート。

a) 資料群別・年代順（元ファイルの配列）

纂 01751100：鍼灸師法制定ノ請願ノ件

纂 01751100：血液循環治療法規制定ノ請願ノ件

纂 01751100：無楽療術法規制定ノ請願ノ件

纂 01751100：機船底曳網漁業取締ノ請願ノ件

纂 01751100：農村振興政策ニ関スル請願ノ件

纂 01751100：医師法第十三条改正二関スル請願ノ件
纂 01751100：仏教各宗派管長ヨリ貴族院議員互選二関スル請願ノ件外二件
纂 01751100：実業補習教育振興二関スル請願ノ件
纂 01751100：魚市場法制定二関スル請願ノ件
纂 01751100：小学校並中等学校ニ神祇科加設ノ請願ノ件
纂 01751100：家畜伝染病予防法中改正二関スル請願ノ件
纂 01751100：高砂港改修二関スル請願ノ件
纂 01751100：貧弱町村ノ合併二関スル請願ノ件
纂 01751100：水産防疫費国庫支弁ノ請願ノ件
纂 01751100：水産増殖法制定ノ請願ノ件
纂 01751100：部落問題ノ国策確立二関スル請願ノ件
纂 01751100：鉱業法改正ノ請願二件
纂 01751100：根室郡ニ国立種馬所設置ノ請願ノ件
纂 01751100：千曲川改修工事速成二関スル請願ノ件
纂 01751100：町村有給吏員退隠料法制定ノ請願ノ件
纂 01751100：牛ノ伝染性肋膜炎ニ関スル勅令継続二関スル請願ノ件

b) 目録の文字コード順

纂 02039100：ソヴィエト社会主義共和国聯邦アレクサンドロフスク駐在総領事緒方整
肅へ御委任状御下付ノ件
纂 02909100：「ソヴィエト」社会主義共和国聯邦「アレクサンドロフスク」駐在総領事
小柳雪生へ御委任状御下付ノ件
纂 02039100：ソヴィエト社会主義共和国聯邦ノヴォシビルスク駐在領事代理副領事小
柳雪生へ御委任状御下付ノ件
纂 02444100：「ソヴィエト」社会主義共和国聯邦浦潮斯德駐在総領事宮川船夫へ御委任
状御下付ノ件
纂 02039100：ソヴィエト社会主義共和国聯邦浦潮斯德駐在総領事渡辺理恵へ御委任状
御下付ノ件
纂 02093100：ソヴィエト社会主義共和国聯邦堪察加管区ペトロパウロフスク駐在領事
斎藤輝宇良へ御委任状御下付ノ件
纂 02543100：「ソヴィエト」社会主義共和国聯邦駐箚特命全権大使建川美次御信任状並
前任特命全権大使東郷茂徳御解任状御下付ノ件
纂 01987100：ソヴィエト社会主義共和国聯邦駐箚特命全権大使大田為吉御信任状並前
任特命全権大使広田弘毅御解任状御下付ノ件
纂 02038100：ソヴィエト社会主義共和国聯邦駐箚特命全権大使大田為吉御信任状並前
任特命全権大使広田弘毅御解任状捧呈済ノ件
纂 01744100：ソヴィエト聯邦アレクサンドロフスク駐在ノ総領事佐々木静吾へ御委任
状御下付ノ件
纂 01896100：ソヴィエト聯邦オデッサ駐在領事田中文一郎へ御委任状御下付ノ件

- 纂 01938100 : ソヴィエト聯邦堪察加管区ペトロパウロフスク駐在領事島田滋へ御委任
状御下付ノ件
- 纂 02099100 : ソヴィエト聯邦赤十字及赤新月社聯合ヨリ風害見舞ノ件
- 纂 01833100 : 「ソウエト」聯邦反革命罪ニ関スル刑法
- 類 03527100 : ソ同盟引揚者に対する人権じゅうりんに関する件 梨木作次郎
- 纂 02267100 : 「ソ」聯邦行旅券査証ニ関スル件
- 類 04055100 : ソ連における農業及び工業の視察旅行に参加の件
- 纂 03119100 : 「ソ連及其の占領地区における日本人の状況」送付の件
- 類 03813100 : ソ連地域からの帰還に関する件
- 類 03641100 : 「ソ連地区」未帰還者数に関する件 林百郎
- 類 03188100 : ソ連領からの復員促進に関する請願外三件 (第九十一号)

c) 目録の逆引き・文字コード順

- 類 01646100 本邦及加奈陀間公使ヲ交換ス
- 類 01816100 日本印度間ノ通商ニ関スル条約ノ効力存続ニ関スル暫定取極ニ関スル公文ヲ交換ス
- 類 01648100 日本国及ニュージールランド間通商暫定取極ニ関スル公文ヲ交換ス
- 類 01712100 日本国ルーマニア国間通商暫定取極ニ関スル公文ヲ交換ス
- 類 01677100 日本国西班牙国間通商暫定取極ニ関スル公文ヲ交換ス
- 類 01712100 日本国埃及国間通商暫定取極ニ関スル公文ヲ交換ス
- 類 01677100 日本国及土耳其国間通商暫定取極ニ関スル公文ヲ交換ス
- 類 01612100 日本国勃爾牙利国間通商暫定取極ニ関スル公文ヲ交換ス
- 類 01677100 日本国ペルシア国間修好通商暫定取極ニ関スル公文ヲ交換ス
- 類 01677100 日本国キューバ国間通商暫定取極締結ニ関スル公文ヲ交換ス
- 類 01612100 日本国ポーランド国間通商航海条約ニダンチヒ自由市加入ニ関スル公文ヲ交換ス
- 類 01677100 漢口事件解決方ニ関スル公文ヲ交換ス

目録をなんらかの配列に並べ替えてみると、自ずと表現の類似した目録が集まってくる。法令名・組織名・地名が文頭にある場合には b) が、人名や年号が文頭にある場合には c) が、作業しやすかった。

2. 抽出語

2.1. 抽出語の分析

資料群別抽出結果

資料群	時代	目録 件数	文字 総数	語彙 総数	抽出語 総数	目録 あたり 文字数 (平均)	目録 あたり 語彙数 (平均)
公文 雑纂	S1-S20	40,539 件	1,253,616 字	523,477 語	39,456 語	30.9 字	12.9 語
	S21-25	1,379 件	32,043 字	13,500 語	3,160 語	23.2 字	9.8 語
公文 類聚	S1-S20	19,560 件	645,697 字	294,585 語	14,466 語	33.0 字	15.1 語
	S21-25	7,736 件	187,978 字	77,889 語	7,695 語	24.3 字	10.1 語
	S26-29	7,069 件	192,424 字	79,606 語	6,244 語	27.2 字	11.3 語
御署名 原本	S1-S20	12,972 件	483,724 字	182,272 語	7,112 語	37.3 字	14.1 語
	S21-25	4,484 件	132,657 字	52,131 語	4,378 語	29.6 字	11.6 語
	S26-29	2,957 件	72,353 字	29,812 語	2,422 語	24.5 字	10.1 語
枢密院 会議 筆記	S1-S20	1,046 件	31,934 字	13,147 語	1,387 語	30.5 字	12.6 語
	S21-25	158 件	5,265 字	2,170 語	425 語	33.3 字	13.7 語
全 体		97,900 件	3,037,691 字	1,268,589 語	5,8790 語	31.0 字	13.0 語

目録あたり文字数 = 文字総数 ÷ 目録件数

目録あたり語彙数 = 語彙総数 ÷ 目録件数

抽出語総数は、語彙総数のうち重複を取り除いた語彙数

2.2. 抽出語の分類整理

2.2.1. 目的

抽出語の頻度表を作成したところ、次表のように上位を一般語や助詞・助動詞などが占めており、目的の語種を探ることが困難だった。そのため、一般的な語や助詞・助動詞を排除することと類似の意味を持つ語彙をグループ化することを目的として、暫定的な分野を設定して、分類作業を試行した。この作業では、正確さを期すことはしなかった。

抽出語の出現頻度表(上位 50 語)

順位	抽出語	出現回数	順位	抽出語	出現回数
1	ノ	76981	26	シ	5801
2	件	65441	27	俸給制限	5774
3	二	44237	28	政令	5626
4	ヲ	30934	29	定ム	4764
5	改正	30768	30	為	3968
6	中	25049	31	外一名	3645
7	の	24225	32	鉄道局	3635
8	勅令	20626	33	下賜	3516
9	に	16341	34	事務官	3207
10	同上	15692	35	以下	3185
11	ス	15636	36	所管	3102
12	御署名原本	15470	37	施行令	3048
13	関スル	13626	38	支給	3031
14	官制	11352	39	昭和十六年	2976
15	を	10093	40	奏任文官	2938
16	法律	9994	41	大正十五年	2866
17	する	9085	42	判任官	2707
18	等	8549	43	都市計画	2694
19	賞与	8373	44	並	2679
20	一部	7673	45	副参事	2646
21	関する	7564	46	外二件	2633
22	任用	7505	47	条	2572
23	外	7120	48	朝鮮総督府	2561
24	及	7083	49	及び	2497
25	外一件	6670	50	技師	2373

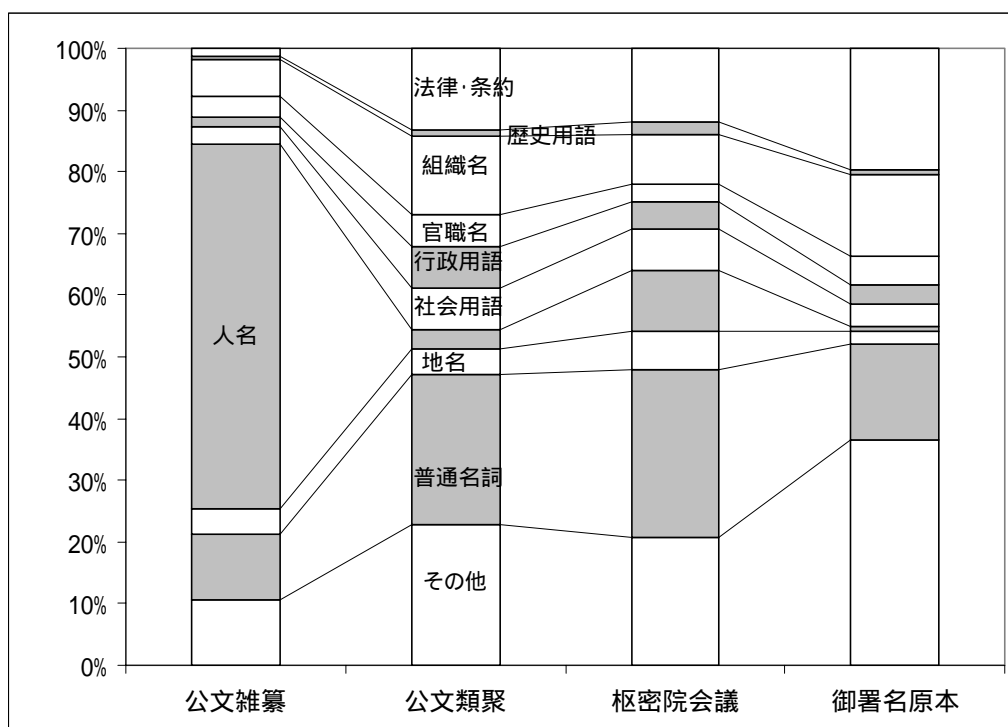
2.2.2. 抽出語のジャンル分け（試行）の結果

抽出語を分類した結果を以下に示す。

歴史用語は、事件・出来事などを含む。
 その他は、動詞・助詞・助動詞・数詞など。
 各資料群において、普通名詞とその他を除いた頻度上位2つにマーカーをつけた。

抽出語の分野別・資料群別の出現頻度表

#	分野	公文雑纂		公文類聚		枢密院会議筆		御署名原本		全体	
		語数	%	語数	%	語数	%	語数	%	語数	%
1	法律・条約	523	1%	2,943	13%	917	12%	1,490	20%	3,658	6%
2	組織名	2,646	6%	2,842	13%	605	8%	991	13%	4,643	8%
3	官職名	1,348	3%	1,122	5%	211	3%	351	5%	1,827	3%
4	行政用語	727	2%	1,528	7%	354	5%	238	3%	1,935	3%
5	社会用語	1,206	3%	1,468	7%	511	7%	269	4%	2,290	4%
6	歴史用語	198	0%	200	1%	157	2%	60	1%	302	1%
7	人名	25,131	59%	703	3%	752	10%	49	1%	26,000	44%
8	地名	1,735	4%	899	4%	465	6%	168	2%	2,356	4%
9	普通名詞	4,565	11%	5,415	24%	2,070	27%	1,173	16%	6,251	11%
1	その他	4,540	11%	5,044	23%	1,592	21%	2,750	36%	9,528	16%
	計	42,619	100%	22,16	100%	7,634	100%	7,539	100%	58,790	100%



2.2.3.

資料群別の傾向

- ・ 公文雑纂では、人名、普通名詞、その他 で、81%を占めており、人名の多くが一般人である。また、法律関係の語は少ない。公文雑纂には、人事に関する目録が多いことがわかる。
- ・ 公文雑纂以外の資料群では、法律・条約と組織に関する語が多い。
- ・ 歴史用語は、全体的に少ない。
- ・ 平均して 25%は、普通名詞や動詞・助詞・助動詞などである。

2.3. 頻度表

以下のような各種頻度表を作成した(頻度表全体は CD-R に収録)。
 なお、これ以降の表では、普通名詞・その他を除いた。

2.3.1. 資料群全体での頻度順 (上位 50 位)

順位	抽出語	頻度(回)	順位	抽出語	頻度(回)
1	勅令	20626	26	専売局	1025
2	御署名原本	15470	27	賞賜	986
3	俸給制限	5774	28	特命全権公使	959
4	鉄道局	3635	29	駐劄	953
5	下賜	3516	30	執行年度割	941
6	事務官	3207	31	内閣属	920
7	奏任文官	2938	32	東京	878
8	判任官	2707	33	勅諭	866
9	都市計画	2694	34	帝国議会	866
10	副参事	2646	35	答弁書	835
11	朝鮮総督府	2561	36	大蔵省	832
12	技師	2373	37	日本国	831
13	逓信局	2259	38	施行規則	811
14	台湾総督府	2250	39	建議	780
15	臨時職員	1983	40	逓信省	753
16	設置制	1921	41	事務囑託	744
17	教授	1778	42	地方事務官	714
18	鉄道局技手	1731	43	助教授	712
19	施行期日	1686	44	内務省	705
20	召集免除	1449	45	街路事業	655
21	通信事務官	1216	46	第二予備金	653
22	特別会計	1166	47	褒章下賜	649
23	法律案	1110	48	関東局	642
24	高等官官等俸給	1103	49	鉄道省	631
25	御下付	1055	50	文部省	626

2.3.2. 資料群別頻度順（上位 50 位）

順位	公文雜纂	公文類聚	枢密院會議	御署名原本
1	俸給制限	勅令	勅令	勅令
2	下賜	特別會計	御署名原本	御署名原本
3	事務官	施行期日	帝国議會	臨時職員
4	都市計画	法律案	法律案	設置制
5	奏任文官	大蔵省	特別任用	施行期日
6	技師	日本国	文官	朝鮮總督府
7	逓信局	施行規則	任用分限	高等官官等俸給令
8	台灣總督府	歳入歳出	初叙陞叙	台灣總督府
9	通信事務官	文部省	衆議院議員選挙法	施行規則
10	朝鮮總督府	農林省	商工省	樞太
11	賞賜	逓信省	領事官	台灣
12	駐劄	厚生省	外務省	關東州
13	建議	外務省	裁判所構成法	樞太庁
14	事務囑託	台灣	世伝御料	特別會計
15	褒章下賜	朝鮮	逓信省	服制
16	御信任状	衆議院	外交官	地方官
17	答弁書	附則	事務官	貴族院
18	陸軍中將	歳出	通則	朝鮮
19	鐵道省	一般會計	俸給	補闕選挙
20	男爵	特赦	大蔵省	歳入歳出
21	理事官	服制	改正案	帝国議會
22	日本赤十字社	軍事費	設置制	議員
23	衆議院議員	議定書	厚生省	総予算
24	書記官	臨時措置	国民学校	關東局
25	書記官長	官吏	會計検査院法	予算外国庫
26	退職特別賜金	議員	會計法	歳入歳出予算
27	枢密院	歳入歳出予算	官吏	官吏
28	帝国議會	俸給	両院	特別任用令
29	裁判所	予備費	陸軍省	専売局
30	逓信技師	内閣總理大臣	高等試験令	特別會計
31	特別社員	裁判所	文部省	農林省
32	有功章	官庁	理事官	大蔵省
33	内閣總理大臣	災害復旧	樞太	北海道庁
34	再就職追認	衆議院議員	衆議院議員	商工省
35	逓信省	同年度	議院法	諸学校
36	衆議院	法令	陪審法	日本国
37	都市計画事業	東京	大使館	裁判所
38	法律案	文官	海軍省	厚生省
39	俸給	一般會計予算	軍需省	一般會計
40	鐵道官	北海道	農商省	商工部内
41	外六件	国民	貴族院令	奏任文官
42	貴族院	諸学校	内閣員	稅務署
43	北海道庁	移管	國務大臣	職員定員令
44	会長	運輸省	国民学校令	庁府県
45	情報局	通則	書記生	手数料
46	褒章	特別會計法	皇室儀制令	特別任用
47	飾版	旅費	戰時行政職權特例	文部省
48	褒状下賜	輸入稅	軌範	農林部内
49	紺綬	事業費	高等学校令	警視庁
50	情状位	勅語	中等学校令	逓信官署

2.3.3. 年代別頻度順（上位 50 位）

順位	戦前	戦後	順位	戦前	戦後
1	勅令	勅令	26	執行年度割	小川友三
2	御署名原本	御署名原本	27	内閣属	意見書
3	俸給制限	法律案	28	特命全権公使	厚生省
4	鉄道局	施行期日	29	駐箚	特別会計法
5	下賜	日本国	30	勅諭	内閣総理大臣
6	事務官	特別会計	31	東京	附則
7	奏任文官	参議院議員	32	特別会計	農林省
8	判任官	衆議院	33	帝国議会	災害復旧
9	都市計画	設置制	34	事務囑託	帝国議会
10	副参事	施行規則	35	建議	大水害
11	朝鮮総督府	答弁書	36	地方事務官	暫定措置
12	技師	退職手当	37	助教授	予備費
13	逓信局	ポツダム宣言	38	逓信省	米国対日援助
14	台湾総督府	使用方	39	街路事業	運輸省
15	臨時職員	国家公務員	40	大蔵省	「ポツダム」宣言
16	教授	アメリカ合衆国	41	第二予備金	歳入歳出
17	鉄道局技手	一般会計	42	関東局	議員提案
18	設置制	特赦	43	褒章下賜	同年度
19	召集免除	一般会計予備	44	内務省	行政協定
20	施行期日	臨時措置	45	鉄道省	政府職員
21	通信事務官	臨時職員	46	理事官	所得税法
22	高等官官等俸給	参議院	47	樺太庁	一般会計予算
23	御下付	大蔵省	48	教諭	日本国憲法
24	専売局	地方自治法	49	台湾	繰入金
25	賞賜	法令	50	視学	俸給