漢字調査における統計的尺度の問題

漢字調査における統計的尺度の問題

田　中　章　夫

0. はじめに

従来，漢字の計量的調査における総計的尺度としては，もっぱら，その出現頻度（F）と，それに基づく漢字使用率（P）とが用いられてきた。

しかし，漢字の度数分布の特性として，多数の漢字が同一頻度で集まりやすいため，使用率は，漢字のウェイトを測る有効な尺度となりえない場合がすくなくない。そのうえ，使用率は，データの中の漢字のみを対象とする尺度であるために，それによって表記されている語彙群との関連性が失われ，大量語彙調査によって得られる有効な情報が，漢字調査の上に反映してこない欠点をもっている。

そこで，漢字調査の統計的尺度として，カバー率を提唱して，漢字調査の計量的処理の面に，語彙調査の調査結果の反映をはかろうというのが，この論の主旨である。

$$\text{漢字使用率（P）}‰=\frac{\text{ある漢字の度数（F）}}{\text{延べ漢字数（N）}}\times 1000$$

1. カバー率（C）の概念

漢字の中には，同じ程度の使用頻度であっても，用法の広いものと，狭いものとがある。いま，「雑誌九十種の漢字調査[1)]」の度数9・使用率0.032を見てみると，ここには，61個の漢字が並んでいるが，この中の「瘍」という漢字は，「腫瘍」一語にしか使われていないのに対して，「筒」は「円筒・筒型・筒抜け……」など8語に使われている。したがって，同一頻度・同一使用率の漢字

ではあるが，もし，「瘍」という活字がなくても，その影響は「腫瘍」一語にしか及ばないが，「筒」の場合は八語に及ぶということになる。

ここに，使用頻度あるいは，それに基づく使用率（すなわち，延べ字数に対する，各漢字の出現頻度の比率）といった，従来の尺度では，カバーしきれない問題が浮かびあがってくる。これは，簡単にいえば，従来の尺度には，個々の漢字が，どれだけの語を表記しうるかという観点が欠けていたということにほかならない。

漢字というものが，本来，語の表記のさいに現われるものである以上，一つ一つの漢字について，それが，どれだけの範囲の語を表記しうるかを調べ，それに基づいて，各漢字のウェイトを統計的に測定することは，可能なはずである。これを，仮に「カバー率」と名づけるならば，これは，「ある漢字で表記される語の量が，語彙の総量に対して，どれだけの比率になるか」を表わすものとなる。この概念を式の形で表わせば，つぎのようになる。

$$\text{カバー率 (C) ‰} = \frac{\text{ある漢字の用いられた語の数 (W)}}{\text{語彙集団の総体 (V)}} \times 1000$$

したがって，「カバー率」は，それぞれの漢字の影響が，語彙集団の，どれだけの範囲に及ぶか，影響の広さを測る尺度だということができる。

注 1)　国立国語研究所報告22「現代雑誌九十種の用語用字・第二分冊(漢字表)」1963

2.　延べ語カバー率 (Cn)

延べ語の総体に対して，ある漢字で表記される語の量が，どれだけの比率になるかを表わすものを「延べ語カバー率」と呼ぶならば，その計算式は，つぎのように表わされる。

$$\text{延べ語カバー率 (Cn) ‰} = \frac{\text{ある漢字の用いられた語の数 (Wn)}}{\text{延べ語総量 (Vn)}} \times 1000$$

いま，「雑誌九十種の漢字調査」における度数 150 の 6 字について，延べ語

表1　度数150の漢字の延べ語カバー率 (Cn)

全体使用率（P）=0.536‰

全体			一般			人名・地名		
漢字	語数	カバー率	漢字	語数	カバー率	漢字	語数	カバー率
座	150	0.514	申	150	0.546	阪	131	7.516
申	150	0.514	個	146	0.532	英	106	6.081
英	150	0.514	追	148	0.539	座	26	1.492
阪	150	0.514	座	124	0.452	追	2	0.115
追	150	0.514	英	44	0.160	個	0	—
個*	146	0.500	阪	19	0.069	申	0	—

＊ 個個4例あり

カバー率 (Cn) を求めると，表1のようになる。いずれも，この時の漢字調査の対象となった「雑誌九十種の語彙調査[2]」の中段[3] までの延べ語数 291910 に対する比率である。この表で「個」の「全体」のカバー率が他の字よりも，やや低いのは，「個々」が4例あったためである。「雑誌九十種の調査」では，おどり字（々）を，先行漢字に変換して数えているため，今の例のように，「個々」が4回出現した場合，漢字の頻度は8になるが，延べ語数は，4語ということになる。したがって，カバー率の計算においても，当然4語と計算されるので，「個」の度数150マイナス4の146語が，この漢字で表記された延べ語の数ということになる。この結果，こうした用法のなかった他の漢字よりも，カバー率が低く算出されるわけである。

この表1をみてみると，「座」の「全体」のカバー率は，0.514パーミルである。このことは，この漢字の用いられる語が，延べ語一万当り5語強の割合で現われるということを示している。また，この表から「個」「申」「追」などで表記される語は，人名・地名を除いた一般語において高いカバー率を示し，それに対して，「阪」「英」は人名・地名に用いられる語において千語当り7.5語強・6語強といった割合で，かなり頻繁に現われてくることを示している。

表2も，同様に，「雑誌九十種の漢字調査」の度数20の引字について延べ語カバー率を示したものである。この表からわかるように「岐」は，ほとんど人

表 2　度数 20 漢字の延べ語カバー率 (Cn)

全体使用率（P）=0.150%

全体			一般			人名・地名		
漢字	語数	カバー率	漢字	語数	カバー率	漢字	語数	カバー率
伎	20	0.069	伎	20	0.073	岐	19	1.090
促	20	0.069	促	20	0.073	誠	10	0.574
偵	20	0 069	偵	20	0.073	磨	9	0.516
凝	20	0.069	凝	20	0.073	孫	8	0.459
呆	20	0.069	呆	20	0.073	裕	7	0.402
岐	20	0.069	憤	20	0.073	麗	7	0.402
憤	20	0.069	是	20	0.073	貝	5	0.287
是	20	0.069	滞	20	0.073	菱	4	0.229
滞	20	0.069	熟	20	0.073	雀	4	0.229
熟	20	0.069	瓶	20	0.073	譲	3	0.172
瓶	20	0.069	腐	20	0.073	伎	0	—
磨	20	0.069	膝	20	0.073	促	0	—
腐	20	0.069	諾	20	0.073	偵	0	—
膝	20	0.069	誕	20	0.073	凝	0	—
菱	20	0.069	賊	20	0.073	呆	0	—
裕	20	0.069	賛	20	0.073	憤	0	—
誠	20	0.069	購	20	0.073	是	0	—
諾	20	0.069	遣	20	0.073	滞	0	—
誕	20	0.069	釦	20	0.073	熟	0	—
譲	20	0.069	鉢	20	0.073	瓶	0	—
貝	20	0.069	齢	20	0.073	腐	0	—
賊	20	0.069	譲	17	0.062	膝	0	—
賛	20	0.069	菱	16	0.058	諾	0	—
購	20	0.069	雀	16	0.058	誕	0	—
遣	20	0.069	貝	15	0.055	賊	0	—
釦	20	0.069	裕	13	0.047	賛	0	—
鉢	20	0.069	麗	13	0.047	購	0	—
雀	20	0.069	孫	11	0.040	遣	0	—
麗	20	0.069	磨	11	0.040	釦	0	—
齢	20	0.069	誠	10	0.036	鉢	0	—
孫*	19	0.065	岐	1	0.004	齢	0	—

*（子子）孫孫 1 例あり

名・地名の表記にのみ現われ，延べ千語当り約一語の表記に使われているが，一方「伎」は，人名・地名の表記には全く現われず，一般語のみに使われ，この字で表記された語は，延べ語十万語当り7.3語程度ということになる。

延べ語カバー率は，当然なことであるが，その増減は，頻度数の増減と，ほぼパラレルである。ということは「使用率(P)」の増減とも，ほぼ一致する。したがって，延べ語カバー率順の漢字表を作成すれば，順位においては，使用率順漢字表を作成すれば，大体一致する。

表3は，「雑誌九十種の漢字調査」のデータで，延べ語カバー率（全体）順上位30字について使用率と対照したものである。第一位の漢字「一」のカバー率は，延べ語全体29190において，15.539パーミル，すなわち，延べ語千語当り15.5語強が，この漢字で表記されているというわけである。「使用率(P)」の順位と，延べ語カバー率(全体)の順位とが，12位の「上」と13位の「本」のところで入れかわっているが，この理由は，「上」に「上々(上上)」「上池上」の表語が，各一例存在していたためである。すなわち，「上」の度数は1644，「本」の度数は1643なので，「上」の用いられた語の数は，1644マイナス2（上々・上池上）で1642となり，「本」の語数1643を下まわって，順位の逆転が起こるわけである。

表3の延べ語カバー率を検討すると，人名・地名におけるカバー率順位と，人名・地名以外の一般語におけるカバー率順位とが，大きく食い違ってくることが明らかになる。人名・地名では，「子・本・日……」などが，きわめて高いカバー率を示し，「子」は，5%を上まわり，「本」も5%近くのカバー率を示しているが，一般語においては，3パーミル前後のカバー率を示すにすぎない。

表3の数値にしたがって，「使用率（P）」と，延べ語カバー率（Cn）との関係を示すグラフを描くと，表4の通りになる。「使用率」と，延べ語カバー率（全体）とは，ほぼ平行に変化しているが，人名・地名の延べ語カバー率は，全く別のカーブを描いている。すなわち，人名・地名では，きわめて限られ

表 3　延べ語カバー率（全体）順上位30字と漢字使用率

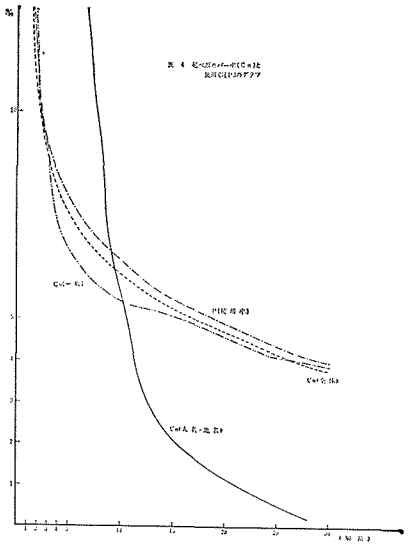
単位は‰

順位	使用率（P）	漢字	度数	全体			一般			人名・地名		
				字	語数	カバー率	字	語数	カバー率	字	語数	カバー率
1	16.346	一	4578	一	4536	15.539	一	4333	15.786	子	911	52.266
2	11.412	人	3196	人	3050	10.448	人	3022	11.010	本	850	48.766
3	9.052	二	2535	二	2534	8.641	二	2422	8.824	日	739	42.398
4	8.570	大	2400	大	2399	8.218	出	2067	7.531	大	448	25.703
5	7.906	日	2214	日	2186	7.489	大	1951	7.108	三	333	19.105
6	7.523	出	2107	出	2107	7.218	十	1831	6.605	中	315	18.072
7	7.209	三	2019	三	2017	6.910	年	1703	6.204	一	203	11.647
8	6.691	十	1874	十	1874	6.420	三	1684	6.135	上	166	9.524
9	6.559	子	1837	子	1836	6.290	中	1508	5.494	二	112	6.426
10	6.513	中	1825	中	1823	6.245	方	1501	5.469	五	96	5.508
11	6.163	年	1726	年	1706	5.844	上	1476	5.377	前	72	4.131
12	5.870	上	1644	本	1643	5.628	見	1464	5.334	十	61	3.500
13	5.867	本	1643	上*	1642	5.625	手	1448	5.275	見	45	2.582
14	5.538	方	1551	方	1513	5.183	分	1448	5.275	出	40	2.295
15	5.388	見	1509	見	1509	5.169	日	1447	5.272	間	39	2.238
16	5 235	手	1406	手	1466	5.022	生	1416	5.159	生	31	1.779
17	5.206	分	1458	分	1458	4.995	五	1348	4.911	人	28	1.606
18	5.178	生	1449	生	1447	4.957	行	1330	4.846	目	23	1.320
19	5.163	五	1446	五	1444	4.947	合	1299	4.733	行	22	1.262
20	4.913	前	1376	前	1370	4.693	前	1298	4.729	時	21	1.205
21	4.749	行	1352	行	1352	4.632	時	1243	4.529	合	19	1.090
22	4.706	合	1318	合	1318	4.515	目	1213	4.419	手	18	1.033
23	4.670	時	1308	時	1264	4.330	間	1177	4.288	方	12	0.688
24	4.413	目	1236	目	1236	4.234	思	1163	4.237	来	10	0.574
25	4.346	間	1217	間	1216	4.166	来	1133	4 128	四	10	0.574
26	4.163	思	1166	思	1163	3.984	女	1125	4.099	分	10	0.574
27	4.031	来	1143	来	1143	3.916	事	1113	4.055	年	3	0.172
28	4.031	女	1129	女	1128	3.864	四	1102	4.015	女	3	0.172
29	3.974	事	1113	事	1113	3.813	子	925	3.370	思	0	—
30	3.971	四	1112	四	1112	3.809	本	793	2.889	事	0	—

* 上々・上池上各1例あり

いくつかの漢字で表記されるものが，頻繁に現われることを示している。これは，いうまでもなく，表3の「子」「本」「日」などの類である。この影響が，人名・地名以外のものの延べ語カバー率に及んだため，延べ語カバー率(一般)のグラフは，特殊なカーブを描いている（ということは，「雑誌九十種の漢字調査では，使用率を全体についてしか計算していないが，もし人名・地名とそれ以外とに分けて計算していれば，表4のグラフに示されたような情報を得ることができたはずである)。

表4　延べ語カバー率（Cn）と使用率（P）のグラフ



今回は，使用率の上位30字について，延べ語カバー率を計算して，表4のグラフを作成したため，延べ語カバー率（一般）のグラフが，特殊なカーブを描いたが，延べ語カバー率（一般）の順位にもとづいてグラフを作成すれば，ナチュラルな対数曲線を描くであろうことは，言うまでもない。

注 2)　国立国語研究所報告21「現代雑誌九十種の用語用字・第一分冊（総記および語彙表)」1962

注 3)　「雑誌九十種の語彙調査」は，前段・中段・後段の三段階に分れるが，「漢字調査」は，その前段と中段を対象に実施している。上記「国立国語研究所報告21」の 314 ページ参照。

3.　通算延べ語カバー率（∑Cn）

つぎに，延べ語カバー率を，最上位の漢字から，ある順位までの漢字群全体について計算すれば，その範囲の漢字によって表記された語の集団が，延べ語総量に対して，どれだけの比率を占めているかを，表わすことになる。いま，試みに表3に挙げた使用率順の上位30位の漢字について，この計算をすると結果は表5の通りである。すなわち，使用率順位，第30位までの漢字群の字が使用される語の集団は，延べ語総量の 16.5% 弱を占めているという結果になる。この計算においては，例えば「一人」と書いた「ひとり」は，漢字「一」についてのカバー率計算のさいに算入ずみなので，「人」の計算のさいには，この語は算入しないわけだから，カバー率の単純な累算にはならない。

これを「通算延べ語カバー率（∑Cn）」と名づけると，その最終的な（すなわち，最下位の漢字までの）演算結果は，延べ語総量に対する，漢字表記語の総体を比率の形で表わしたものとなり，語彙全体の中における漢字表記語のウェイトを測る一つの尺度となる。

$$\text{通算延べ語カバー率}\ (\sum Cn)\ ‰ = \frac{\text{ある順位までの漢字で，表記された延べ語の累計*}\ (\sum Wn)}{\text{延べ語総量}\ (Vn)} \times 1000$$

* 同語重出は算入しない

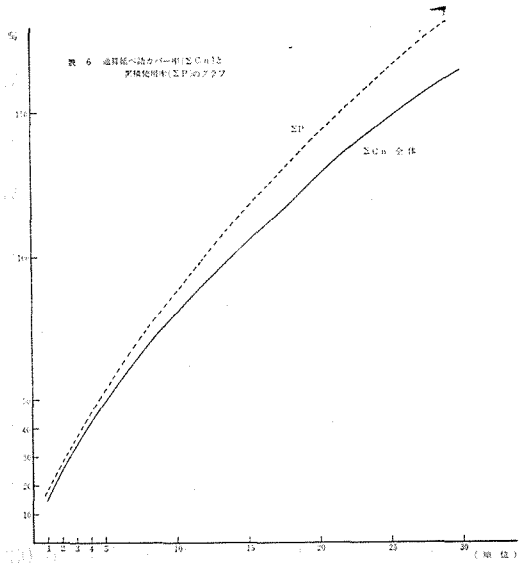
さきに，表5に示した「通算延べ語カバー率(全体)」の増加の具合を，使用率(P)の累加とともにグラフに示すと，表6のようになる。このグラフには，延べ語カバー率の高い漢字群の，延べ語表記における活躍ぶりが顕著に表われ，カバー率は，ほぼ等差的に累加している。このグラフでわかるように，通算延べ語カバー率は，累積使用率と，ほぼ並行して累加していくが，両者の性格は，まったく異なる。すなわち，累積使用率（∑P）は，使用率順漢字表の

表5　上位30字の通算延べ語カバー率（∑Cn）……全体

順位	漢字	累積使用率（∑P）	漢字	語数（全体）	新出語数	通算語数 Wk	通算カバー率 ∑Cu
1	一	16.346	一	4536	4532	—	15.539
2	人	27.758	人	3050	2833	7369	25.244
3	二	36.810	二	2534	2336	9705	33.247
4	大	45.380	大	2399	2375	12080	41.383
5	日	53.286	日	2186	2131	14211	48.638
6	出	60.809	出	2107	2095	16306	55.860
7	三	68.018	三	2017	1917	18223	62.427
8	十	74.709	十	1874	1290	19513	66.846
9	子	81.268	子	1836	1825	21338	78.098
10	中	87.781	中	1823	1808	23146	79.292
11	年	93.944	年	1706	1705	24851	85.132
12	上	99.814	本	1643	996	25847	88.544
13	本	105.681	上	1642	1636	27483	94.149
14	方	111.219	方	1513	1450	28933	99.116
15	見	116.607	見	1509	1480	30413	104.186
16	手	121.842	手	1466	1421	31834	109.045
17	分	127.048	分	1458	1376	33210	113.768
18	生	132.226	生	1447	1377	34587	118.485
19	五	137.389	五	1444	1355	35942	123.127
20	前	142.302	前	1370	1293	37235	127.556
21	行	147.051	行	1352	1326	38561	132.099
22	合	151.757	合	1318	1278	39839	136.477
23	時	156.427	時	1264	1228	41067	140.684
24	目	160.840	目	1236	1211	42278	144.832
25	間	156.186	間	1216	842	43120	147.717
26	思	169.349	思	1163	1080	44200	151.417
27	来	173.430	来	1143	878	45078	154.424
28	女	177.461	女	1128	1056	46134	158.042
29	事	181.435	事	1113	1043	47177	161.615
30	四	185.406	四	1112	895	48072	164.681

最上位から，ある順位までの使用率の単なる累算結果であり，累積度数（∑F）の伸び具合を示すに過ぎないが，通算延べ語カバー率（∑Cn）は，漢字の影響力の広がり具合を示す尺度である。

表6　通算延べ語カバー率（∑Cn）と累積使用率（∑P）のグラフ



$$\text{累積使用率}(\sum P)\ ‰ = \frac{\text{ある順位までの漢字の使用度数の合計}\ (\sum F)}{\text{延べ漢字数}\ (N)} \times 1000$$

4. 異り語カバー率（Ck）

このカバー率の考え方を，異り語に適用すれば，「異り語カバー率」は，「ある漢字で表記される語の，異り語総量に対する比率」を表わすことになる。

$$\text{異り語カバー率}(Ck)\ ‰ = \frac{\text{ある漢字の用いられた異り語数}\ (Wk)}{\text{異り語総量}\ (Vk)} \times 1000$$

ただし，この計算の場合，表記が一通りに統一されている語については問題ないが，カナ書きや，別の表記のある語のについては，表7に示すような計算法をとらなくてはならない。すなわち，カナ書きや，別の表記をもつ語の場合には，その語の表記全体の中における，当該漢字のウェイトを算定するわけである。別の言い方をすれば，異り語全体の中における，その漢字の活躍する範囲，広さを算出するということである。

表7　異り語カバー率の Wk の計算法

漢字	度数	異り語の表記　　（ ）内は用例数	Wk
冗	10	{冗数(10)}	Wk=1
妨	9	{妨害(8)}{害妨げる(1)}	Wk=2
械	61	{機械（器械）(61)}	Wk=1
脹	1	{膨脹(1)・膨張(1)}	$Wk=\frac{1}{1+1}$
鞄	10	{鞄(10)・かばん(1)・カバン(5)}	$Wk=\frac{10}{10+1+5}$
較	42	{比較(41)}{較べる(1)・比べる(20)・くらべる(23)}	$Wk=1+\frac{1}{44}$
轄	6	{所轄(1)}{管轄(4)・管括(1)}{総轄(1)・総括(1)}	$Wk=1+\frac{4}{5}+\frac{1}{2}$
蒐	9	{蒐集（蒐収）(8)} {蒐める(1)・集める(31)・あつめる(7)}	$Wk=1+\frac{1}{39}$

「雑誌九十種の漢字調査」において，度数9・使用率0.032として現われた漢字61字について，異り語カバー率を計算すると，表8のようになる。表8を見てみると，同じ頻度の漢字であっても，その異り語カバー率には，大きな違いが認められる。前に挙げた「瘍」の異り語カバー率は0.029‰に過ぎないが，「筒」のカバー率は，同一使用率の漢字でありながら，0.235‰を示している。異り語カバー率から見た場合，同じ度数9の漢字ではあるが，「筒・鼠・幽・……」などの漢字は，表記の対象となる語の範囲が，きわめて狭く，その用法は，前者に比べて，著しく狭いものと見ることができる。

また，表8の「人名・地名」の欄と，人名・地名以外の一般語の欄とを比べると，「柏」は人名・地名として使われる異り語一万語当り7.3語強をカバーし，「敦」は約5語の表記にみられる割合になるが，一般語の表記には全く影響を与えない。一方，「鼠」は，一般語一万当り2.6語弱をカバーし，「渦」は2.3語強の表記に現われる割合だが，人名・地名の表記には，全く影響を与えないというようなことになる。

以上のように，異り語カバー率は，同一頻度（同一使用率）で並ぶ漢字の間

表 8　度数 9（使用率 0.032）の 61 字の異り語カバー率（Ck）

全体				一般				人名・地名			
筒	0.235	⇩		鼠	0.258	⬇		柏	0.734	⇩	
鼠	0.196	臆	0.092	渦	0.223	曇	0.091	敦	0.490	惹	—
幽	0.176	聯	0.090	愁	0.206	蔽	0.089	柴	0.367	愁	—
柏	0.176	后	0.088	排	0.203	阻	0.087	筒	0.367	憧	—
閑	0.176	携	0.088	筒	0.194	兜	0.077	兜	0.245	挺	—
渦	0.169	柴	0.088	幽	0 194	宰	0.077	堺	0.245	排	—
殻	0.167	梢	0.088	閑	0.194	妨	0.077	梢	0.245	携	—
愁	0.157	炙	0.088	侮	0.194	訟	0.077	薫	0.245	椀	—
排	0.154	睨	0.088	疾	0.194	憧	0.076	鍛	0.245	渦	—
鍛	0.154	靖	0.088	穀	0.194	椎	0.072	靖	0.245	炙	—
侮	0.147	惹	0.086	燐	0.187	吠	0.069	韓	0.245	燐	—
呪	0.147	椎	0.122	殻	0.181	呪	0.069	升	0.122	殿	—
堺	0.147	殿	0.083	碑	0.155	挺	0.065	吠	0.122	疾	—
恭	0.147	吠	0.082	赦	0.155	只	0.060	曇	0.122	瘍	—
敦	0.147	顎	0.078	茸	0.149	盃	0.059	椎	0.122	睨	—
疾	0.147	稀	0.075	糞	0.147	姐	0.051	殻	0.122	碑	—
穀	0.147	蔽	0.067	臆	0.134	薫	0.045	糞	0.122	盃	—
燐	0.142	阻	0.066	剤	0.126	蒐	0.040	肘	0.122	稀	—
糞	0.141	升	0.059	鍛	0.126	升	0.039	閑	0.122	穀	—
肘	0.121	妨	0.059	井	0.123	恭	0.039	井	—	聯	—
兜	0.118	訟	0.059	肘	0.120	梢	0.039	侮	—	臆	—
宰	0.118	韓	0.059	聯	0.119	瘍	0.039	剤	—	茸	—
碑	0.118	憧	0.058	后	0.116	践	0.039	只	—	蒐	—
赦	0.118	挺	0.049	携	0.116	陛	0.039	呪	—	蔽	—
践	0.118	只	0.047	炙	0.116	靖	0.039	后	—	訟	—
茸	0.113	盃	0.045	睨	0.116	椀	0.029	妨	—	赦	—
曇	0.099	姐	0.039	惹	0.113	蛋	0.027	姐	—	践	—
剤	0.096	蒐	0.030	殿	0.110	堺	—	宰	—	蛋	—
井	0.093	瘍	0.029	顎	0.102	柏	—	幽	—	阻	—
薫	0.093	陛	0.029	稀	0.098	敦	—	恭	—	陛	—
		椀	0.022			柴	—		—	顎	—
⇩		蛋	0.020	⬇		韓	—	⇩		鼠	—

のウェイトの違いを知る尺度になりうると考えられる。

つぎに，さきに，表 3 に示した延べ語カバー率（全体）順上位 30 字の漢字に

ついて，異り語カバー率 (Ck) を計算すると，表9のようになる。

この表9の，まず，「全体」の欄を見てみると，「子・大・出・一……」といった順位になり，使用率や延べ語カバー率とは，かなり相様を異にしている。

表9　延べ語カバー率上位30字の異り語カバー率

P	Cn	順	異り語カバー率 (Ck) ‰					
	全体	位	全体		一般		人名・地名	
一	一	1	子	14.393	出	9.250	子	46.879
人	人	2	大	11.131	大	9.231	三	18.115
二	二	3	出	7.733	手	7.760	大	17.136
大	大	4	一	7.673	人	7.714	一	13.953
日	日	5	上	7.577	上	7.690	中	8.323
出	出	6	人	6.508	合	6.292	本	7.834
三	三	7	手	6.101	一	5.686	二	7.589
十	十	8	中	6.090	日	5.511	上	7.222
子	子	9	日	5.628	中	5.383	日	5.998
中	中	10	三	5.606	見	5.025	五	5.630
年	年	11	本	5.355	目	4.728	十	4.039
上	本	12	合	5.133	前	4.722	四	3.427
本	上	13	見	4.523	本	4.570	出	2.938
方	方	14	前	4.293	生	4.494	見	2.938
見	見	15	生	4.062	分	4.332	前	2.938
手	手	16	目	3.797	子	4.118	人	2.693
分	分	17	行	3.443	行	3.913	生	2.693
生	生	18	分	3.438	年	3.182	間	2.507
五	五	19	二	2.891	女	2.970	行	1.958
前	前	20	年	2.505	事	2.861	時	1.591
行	行	21	間	2.381	間	2.321	合	1.469
合	合	22	女	2.315	来	2.314	来	1.102
時	時	23	事	2.174	方	2.288	方	0.857
目	目	24	五	2.093	時	1.783	手	0.857
間	間	25	来	2.023	思	1.729	目	0.857
思	思	26	方	1.944	三	1.650	分	0.612
来	来	27	四	1.768	二	1.405	年	0.367
女	女	28	時	1.737	四	1.243	女	0.245
事	事	29	十	1.629	五	0.974	思	—
四	四	30	思	1.313	十	0.867	事	—

「子」が，第一位を占めたのは，人名・地名の影響であろうが，異り語100語当り1.5語弱の語に使われる勘定になっている。一方，使用率においては，著しく大きな数値を示し，常に上位を占めやすい「一，二，三，十……」などの漢数字が，全般に順位の低落を示しているのも，異り語カバー率の特徴であろう。これは，数値情報の表記において，漢数字は，常に算用数字と競合関係にあるためである。たとえば，「ツイタチ」という語の表記は，「雑誌九十種の調査」では，「一日」が16例，「1日」が6例なので，表7で述べたWkへの算入値は，22（16＋6）分の6となり，1に達しない。同様に「イチ」「フタリ」「サンルイ」「トオカ」などの単語の表記における漢数字の現われ方を示してみると，それぞれ表10の通りである。この表からわかる通り，これら，数値情報を表わす語の表記においては，漢数字は，常に算用数字の強い圧迫を受けるため，必然的にWkへの算入値が伸びてこない。その結果，異り語カバー率においては漢数字のウェイトが低くおさえられ，「大・出・上・人・手…」などが，上位に進出してくるわけである。

一般に，漢数字は，きわめて高い頻度で繰り返し現われやすい数値関係の語の表記に使用されるため，使用率や延べ語カバー率は，当然，高くなる。しかし，その出現領域は，主として数値情報に限られる上，そこで上述のように算用数字との競合が起こるので，異り語の表記における漢数字の活躍領域は，使

表10 数値情報の表記

（ ）内の数字は用例数（延べ数）

漢字	語例	表記	Wkへの算入値
一	イチ	一(2003)・いち(6)・1(1225)・I(2)	2003／3236
二	フタリ	二人(197)・2人(2)・ふたり(11)	197／210
三	サンルイ	三塁(17)・3塁(3)	17／20
十	トオカ	十日(25)・10日(5)	25／30
五	ゴ	五(731)・伍(1)・ゴ(1)・5(690)・V(3)	731／1426
四	ヨッカ	四日(27)・4日(9)	27／36

用率が示すほど広いものではないという結果になる。

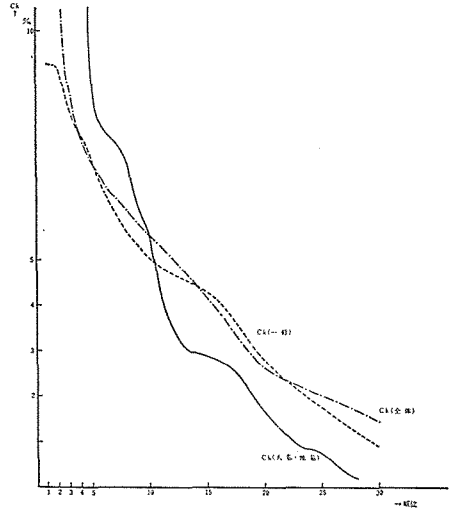
異り語カバー率（全体）において，漢数字「三」が，辛うじて十位をたもちえたのは，人名・地名の表記において活躍しているためであり，「四」の順位が，使用率順位よりも若干上昇したのも，同じ理由による。

人名・地名の影響を排除した一般語の異り語カバー率においては，一般的な用法の広い「一」を除いて，漢数字は軒並み，順位の著しい低落を示し，算用数字との競合の激しさを，端的に物語っている。これは，簡単にいえば，漢数字は，たとえ無くなっても，算用数字によって，かなりの程度カバーされうるということにほかならない。

つぎに，人名・地名欄では，「子」が，圧倒的に引き離し，4.7パーセントにのぼる異り語カバー率を示している点が，印象的である。

以上述べた異り語カバー率の変化の様子をグラフに表わしてみると，表11のようになる。全般的にみて，異り語カバー率は，表4のグラフに示した使用率（P）や延べ語カバー率（Cn）の変化の様相とは，かなり違い，表4のグラフに

表 11　異り語カバー率（Ck）のグラフ



比べると，直線的な変化を示している。その中で，もっとも特異な変化を示しているのは，人名・地名のグラフである。すなわち，人名・地名の表記においては，高い順位を占める一部の限られた漢字が，著しく高いカバー率を示すことを表わしている。これは，結局，人名・地名の表記で広く活躍する漢字は，たいへん偏っているということにほかならない。

人名・地名における，こうした傾向は，異り語カバー率（全体）のグラフにも強く反映し，使用率や延べ語カバー率の場合とは全く異なる，角度の急な直線的な変化となって，その影響が現われている。それに対して，人名・地名の影響を受けない一般語のグラフは，上限も低く，他の二本のグラフに比べて，ゆるやかな変化になっている。

総体的にいって，表11のグラフは，異り語カバー率というものが，きわめて人名・地名の影響を受けやすいことを示していると見ることができよう。

5. 通算異り語カバー率（$\sum$ Ck）

異り語カバー率についても，延べ語カバー率の場合と同様に，最上位の漢字から，ある順位までの漢字群，全体について異り語カバー率を計算すれば，「通算異り語カバー率」といったものが算出しうる。これは，いうまでもなく，ある範囲の漢字群によって表記された語の集団が，異り語総量に対して，どれだけの比率を占めるかを表わすことになる。

いま，試みに表9に掲げた使用率順上位30字の漢字群について，この計算をしてみると，結果は，表12のようになる。この表でわかるように，使用率上位30字の漢字群の漢字が用いられる語は，異り語全体の127.824パーミル，すなわち13%弱を占めるという結果になる。もちろん，この計算においては，「大人」と書いた「おとな」は，第2位「大」までの $\sum$ Wk に算入されるので，第6位「人」の計算のときには，算入しないこととなる。したがって，通算異り語カバー率の値は，異り語カバー率の単純な累計ではない。

$$\text{通算異り語カバー率}\ (\sum Ck)\ ‰ = \frac{\text{ある順位までの漢字で,表記された語の累計*}\ (\sum Wk)}{\text{異り語総量}\ (Vk)} \times 1000$$

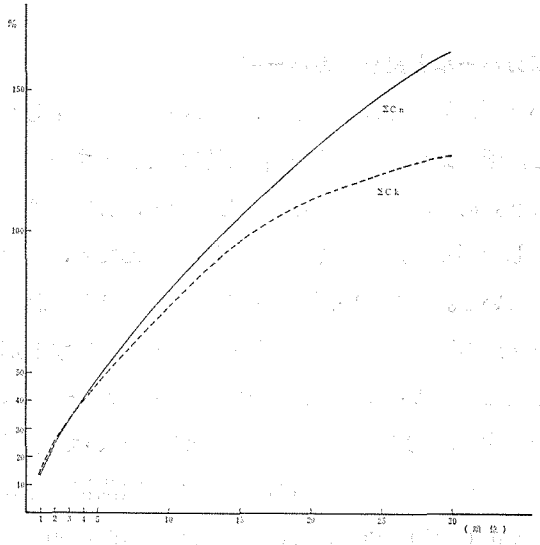
* 同語重出は算入しない

表 12　使用率上位 30 字の $\sum Ck$（全体）と $\sum Cn$（全体）

単位は（‰）

通算異り語カバー率（全　体）		順位	通算延べ語カバー率（全　体）	
子	14.393	1	一	15.539
大	25.494	2	人	25.244
出	33.198	3	二	33.247
一	37.900	4	大	41.383
上	45.420	5	日	48.638
人	51.576	6	出	55.860
手	57.478	7	三	62.427
中	63.450	8	十	66.846
日	68.666	9	子	78.098
三	73.981	10	中	79.292
本	78.953	11	年	85.132
合	83.910	12	本	88.544
見	88.151	13	上	94.149
前	92.248	14	方	99.116
生	96.107	15	見	104.186
目	99.617	16	手	109.045
行	102.943	17	分	113.768
分	106.126	18	生	118.485
二	108.522	19	五	123.127
年	110.851	20	前	127.556
間	113.050	21	行	132.099
女	115.160	22	合	136.477
事	117.206	23	時	140.684
五	119.095	24	目	144.832
来	120.680	25	間	147.717
方	122.486	26	思	151.417
四	123.763	27	来	154.424
時	125.326	28	女	158.042
十	126.600	29	事	161.615
思	127.824	30	四	164.681

表 13　通算異り語カバー率（∑Ck）と通算延べ語カバー率（∑Cn）のグラフ



上に述べたようにして，もし通算異り語カバー率を，異り語カバー率（Ck）の最上位の漢字から，順次，最下位の漢字にいたるまで計算したならば，その最終的な計算結果は，異り語総量に対する，漢字表記語の総体を比率の形で表わしたものとなり，語彙集団における漢字表記語の重みを測る一つの尺度となる。

つぎに，通算異り語カバー率の増加のしかたを，通算延べ語カバー率の累加の状況とともにグラフに示すと，表13のようになる。この30字の漢字群について見るかぎりでは，延べ語カバー率は，漢字が加わるにつれて，比例的に増加し，ほぼ直線的なグラフになるが，一方，通算異り語カバー率のグラフの方は，20位以下のところから伸びが鈍くなり，頭打ちになっている。これは，この辺から下の漢字が用いられる語のバラエティーが，上位に比べて，ぐんと少なくなっていることを示している。したがって，もし，「雑誌九十種の漢字調査」のデータに現われた，すべての漢字について異り語カバー率を算出し，異り語カバー率順の漢字表を作成した場合には，表12の20位以下のあたりの漢字は，

かなり，別のものに入れかわってしまうのではないかと推定される。

6. 延べ語カバー率と異り語カバー率

簡単にいえば，異り語カバー率は，個々の漢字が表記しうる語のバラエティーを表示し，延べ語カバー率は各漢字が影響を及ぼす語彙の広さを示すということになろう。前者は用法の広さを示し，後者は影響の強さを示すと言ってもよかろう。したがって，同じくカバー率とは呼んでも，この二つは，かなり性格を異にしている。その違いが，もっともはっきり表われるのは，すでに前節までのデータから明らかなように，順位である。重複するが，今まで扱って来た30字の漢字群について，使用率（P）・延べ語カバー率（Cn）・異り語（Ck）の順位表を示すと，表14のようになり，これに基づいて順位相関を算出すると，計算結果は，つぎの通りである（スピアマンの順位相関計算法による）。

P と Cn（全体） $r=0.9996$，（一般） $r=0.9929$

Cn と Ck（全体） $r=0.6449$，（一般） $r=0.1472$

P と Ck（全体） $r=0.6476$，（一般） $r=0.2940$

すなわち，人名・地名も含めた全体の順位の場合，延べ語カバー率順位は使用率順位との間に，きわめて高い相関を示すが，異り語カバー率順位との間の相関は，それほど高くない。これを，人名・地名の影響を排除した一般語の方でみると，異り語カバー率順位の，延べ語カバー率順位との間の相関，あるいは，使用率順位との間の相関は，きわめて低くなる。

ここに，人名・地名の強い影響が見られるとともに，異り語カバー率に基づく漢字の段階づけの特異な性格がうかがわれる。

そこで，異り語カバー率と延べ語カバー率の間の関係をとらえるために，表3・表9に示した両者のカバー率を，グラフの上にプロットしてみると，表15・表16のような分布となり，これらの間のスピアマンの相関係数は，つぎの通りである。

（全体） $r=0.4734$

表 14　使用率 (P)・延べ語カバー率 (Cn)・異り語カバー率 (Ck) の順位

P				Cn				Ck			
全	体	一	般	全	体	一	般	全	体	一	般
1	一	1	一	1	一	1	一	1	子	1	出
2	人	2	人	2	人	2	人	2	大	2	大
3	二	3	二	3	二	3	二	3	出	3	手
4	大	4	出	4	大	4	出	4	一	4	人
5	日	5	大	5	日	5	大	5	上	5	上
6	出	6	十	6	出	6	十	6	人	6	合
7	三	7	年	7	三	7	年	7	手	7	一
8	十	8	三	8	十	8	三	8	中	8	日
9	子	9	方	9	子	9	中	9	日	9	中
10	中	10	中	10	中	10	方	10	三	10	見
11	年	11	上	11	年	11	上	11	本	11	目
12	上	12	日	12	本	12	見	12	合	12	前
13	本	13	見	13	上	13	手	13	見	13	本
14	方	14	手	14	方	14	分	14	前	14	生
15	見	15	分	15	見	15	日	15	生	15	分
16	手	16	生	16	手	16	生	16	目	16	子
17	分	17	五	17	分	17	五	17	行	17	行
18	生	18	前	18	生	18	行	18	分	18	年
19	五	19	行	19	五	19	合	19	二	19	女
20	前	20	合	20	前	20	前	20	年	20	事
21	行	21	時	21	行	21	時	21	間	21	間
22	合	22	目	22	合	22	目	22	女	22	来
23	時	23	間	23	時	23	間	23	事	23	方
24	目	24	思	24	目	24	思	24	五	24	時
25	間	25	四	25	間	25	来	25	来	25	思
26	思	26	来	26	思	26	女	26	方	26	三
27	来	27	女	27	来	27	事	27	四	27	二
28	女	28	事	28	女	28	四	28	時	28	四
29	事	29	子	29	事	29	子	29	十	29	五
30	四	30	本	30	四	30	本	30	思	30	十

(一般)　$r=0.3027$

一般に，使用率が高く，延べ語カバー率の高い漢字については，それらの漢字によって表記される語の範囲も広く，バラエティーも豊かであり，当然，異

表 15　カバー率（全体）の分布

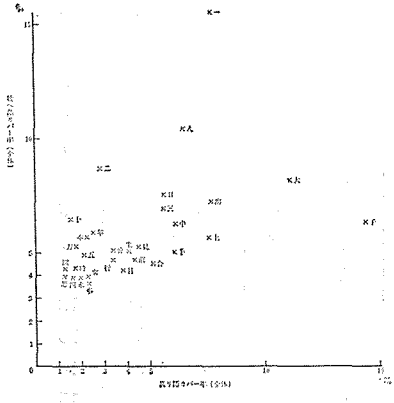
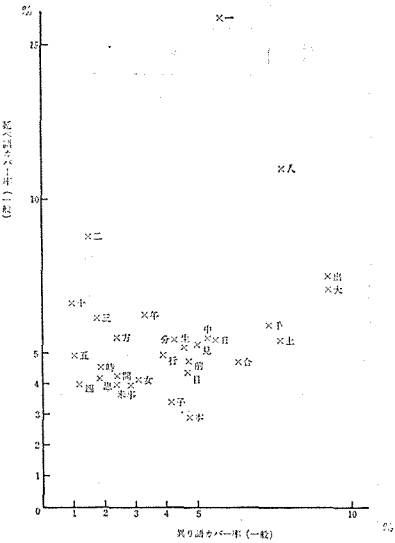


表 16　カバー率（一般）の分布



り語カバー率も，相当に高くなると考えられる。したがって，この種の漢字については，延べ語カバー率と異り語カバー率との間に，かなり高い相関が予想される。それにもかかわらず，延べ語カバー率のきわめて高い使用率順位上位30字の漢字についてすら，この程度の相関しか認められないということは，全般的にみれば，延べ語カバー率と異り語カバー率は，互いに，かなり独立した尺度ではないかと想像される。

また，表15・表16のグラフを見ると，延べ語カバー率・異り語カバー率がともに高い「一・人」などのグループと，延べ語カバー率だけが著しく高い「二・十」などのグループ，異り語カバー率だけが著しく高い「大・出」あるいは表15の「子」などのグループ，そして両方とも低い「四・思」などのグループが認められ，これら二つの尺度に基づく，漢字のグループ分け・段階づけの可能性が示唆されている。

7.　カバー率による漢字基本度の推定

以上述べてきたように，各漢字のカバー率を計算すれば，延べ語カバー率によって，語彙全体の表記に及ぼす，その漢字の影響力を推測することができ，異り語カバー率に基づいて，その漢字で表記しうる語のバラエティーの豊かさを測定しうる。

いま，「雑誌九十種の漢字調査」に現われた3328字の漢字について，一字当りの平均カバー率を計算すると，延べ語カバー率・異り語カバー率のいずれも，0.3パーミル前後となり，これは延べ語については，一字当り平均87.7語の表記をカバーすることを示し，異り語については10.2語平均の表記を担当する計算になる。そこで，この平均カバー率0.3パーミルの線で，データの3328字の漢字を分けてみると，延べ語カバー率・異り語カバー率とも，この平均を超える漢字は，2割程度と推定される。

本来，漢字の分布は，度数1など頻度の低い漢字がきわめて多く，頻度の高い漢字が，きわめて少ない，いわゆるL字型分布となることが予想される。ちなみに，雑誌九十種のデータの場合，度数1は473字もあり，これは3328字の約14.2%に当る。これらの漢字は，無論カバー率も，きわめて低い値を示すわけであるが，平均カバー率の算出には，比較的，数の少ない高頻度・高カバー率の漢字が，強い影響力をもちやすい。その結果，平均カバー率は，上記のように，かなり高く出てしまう。

ここでは，さきの平均カバー率を，一応，認めたうえで，それ以下については，0.1パーミルの間隔で分割し，表17に示すように，延べ語カバー率・異り語カバー率の各々について，三段階の段階づけを試みた。

まず，表17のAのグループは，いうまでもなく，語彙全体の表記に対する影響力も強く，用法も広いもので，当然，基本漢字とされるものの多くは，これに属する。延べ語カバー率を柱の高さに，異り語カバー率を柱の太さになぞらえれば，このAグループは，太くて高い柱ということになる。

それに対して，Iグループは，低くて細い柱に当たるもので，語彙全体の表記に対する影響も弱く，用法もきわめて限られている漢字群である。いわゆる

表17　カバー率（全体）による漢字の段階づけの例　　A～I各欄の（ ）内の数値は度数

Ck＼Cn	延べ語カバー率 (Cn) ‰ 大 (Cn ≧ 0.3)	中 (0.3 > Cn ≧ 0.2)	小 (0.2 > Cn)
異り語カバー率 (Ck) ‰ 大 (Ck ≧ 0.3)	A. 一(4578) Cn=15.539, Ck= 7.763 人(3196) Cn=10.448, Ck= 6.508 大(2400) Cn= 8.681, Ck=11.131 日(2214) Cn= 7.489, Ck= 5.628 出(2107) Cn= 7.218, Ck= 7.733 子(1874) Cn= 6.290, Ck=14.393 私(1080) Cn= 3.700, Ck= 0.488 様(383) Cn= 1.302, Ck= 0.373 際(214) Cn= 0.733, Ck= 0.358 雪(90) Cn= 0.308, Ck= 0.926	D. 筆 (87) Cn = 0.298, Ck = 0.824 附 (86) Cn = 0.295, Ck = 0.673 塩 (78) Cn = 0.267, Ck = 0.765 骨 (73) Cn = 0.250, Ck = 1.000 徳 (69) Cn = 0.236, Ck = 0.941 述 (69) Cn = 0.236, Ck = 0.361 速 (67) Cn = 0.229, Ck = 0.616 攻 (63) Cn = 0.223, Ck = 0.529 閣 (62) Cn = 0.212, Ck = 0.300 輪 (59) Cn = 0.202, Ck = 0.784	G. 暴 (58) Cn = 0.199, Ck = 0.560 弁 (55) Cn = 0.188, Ck = 0.676 敬 (54) Cn = 0.185, Ck = 0.598 略 (49) Cn = 0.168, Ck = 0.648 港 (46) Cn = 0.158, Ck = 0.471 盤 (37) Cn = 0.127, Ck = 0.559 燈 (33) Cn = 0.113, Ck = 0.319 熟 (20) Cn = 0.067, Ck = 0.353 垣 (15) Cn = 0.051, Ck = 0.338 棟 (13) Cn = 0.045, Ck = 0.300
中 (0.3 > Ck ≧ 0.2)	B. 的(1014) Cn= 3.474, Ck= 0.206 彼(799) Cn= 2.737, Ck= 0.219 頃(412) Cn= 1.411, Ck= 0.258 件(196) Cn= 0.671, Ck= 0.294 非(195) Cn= 0.668, Ck= 0.249 枚(177) Cn= 0.606, Ck= 0.206 昨(151) Cn= 0.517, Ck= 0.299 個(150) Cn= 0.500, Ck= 0.264 兄(94) Cn= 0.322, Ck= 0.259 歳(88) Cn= 0.301, Ck= 0.217	E. 週 (87) Cn = 0.298, Ck = 0.299 委 (81) Cn = 0.277, Ck = 0.236 迎 (79) Cn = 0.271, Ck = 0.204 競 (74) Cn = 0.254, Ck = 0.284 易 (71) Cn = 0.243, Ck = 0.240 償 (70) Cn = 0.239, Ck = 0.265 療 (65) Cn = 0.223, Ck = 0.205 頂 (63) Cn = 0.215, Ck = 0.292 系 (60) Cn = 0.206, Ck = 0.206 級 (59) Cn = 0.202, Ck = 0.206	H. 致 (55) Cn = 0.188, Ck = 0.265 承 (53) Cn = 0.182, Ck = 0.265 救 (49) Cn = 0.168, Ck = 0.294 宣 (41) Cn = 0.140, Ck = 0.235 坐 (37) Cn = 0.127, Ck = 0.200 液 (31) Cn = 0.106, Ck = 0.235 謀 (28) Cn = 0.096, Ck = 0.221 泳 (19) Cn = 0.065, Ck = 0.200 傑 (15) Cn = 0.051, Ck = 0.207 筒 (9) Cn = 0.031, Ck = 0.235
小 (0.2 > Ck)	C. 第(564) Cn= 1.932, Ck= 0.086 僕(214) Cn= 0.733, Ck= 0.138 係(213) Cn= 0.729, Ck= 0.182 億(170) Cn= 0.582, Ck= 0.053 誰(153) Cn= 0.524, Ck= 0.172 央(139) Cn= 0.476, Ck= 0.059 及(107) Cn= 0.367, Ck= 0.147 般(106) Cn= 0.363, Ck= 0.176 昔(98) Cn= 0.336, Ck= 0.173 杯(92) Cn= 0.315, Ck= 0.069	F. 糧 (85) Cn = 0.291, Ck = 0.010 撮 (79) Cn = 0.271, Ck = 0.119 皆 (78) Cn = 0.267, Ck = 0.090 批 (77) Cn = 0.263, Ck = 0.062 排 (71) Cn = 0.243, Ck = 0.176 頁 (67) Cn = 0.230, Ck = 0.187 或 (64) Cn = 0.219, Ck = 0.014 械 (61) Cn = 0.209, Ck = 0.029 功 (60) Cn = 0.206, Ck = 0.118 希 (59) Cn = 0.202, Ck = 0.032	I. 筈 (58) Cn = 0.199, Ck = 0.074 較 (42) Cn = 0.144, Ck = 0.030 灯 (38) Cn = 0.130, Ck = 0.190 拶 (27) Cn = 0.192, Ck = 0.025 箇 (23) Cn = 0.079, Ck = 0.026 釦 (20) Cn = 0.069, Ck = 0.129 鞄 (10) Cn = 0.034, Ck = 0.018 雑 (6) Cn = 0.021, Ck = 0.068 遡 (3) Cn = 0.010, Ck = 0.059 脹 (1) Cn = 0.003, Ck = 0.015

制限漢字の多くが，このグループに属するのは当然といえよう。

一方，Gグループは，広く多種類の語の表記に用いられるが，それらの語の頻度が，いずれも低いグループである。頻度よりも，用法の広さの注目される漢字群であり，その意味で，低くて太い柱に例えられよう。逆に，Cグループのものは，きわめて限られた種類の語の表記に用いられる漢字群であるが，これによって表記される語の頻度が，いずれも高いわけである。たとえば，頻繁に出てくる特定の代名詞・接続詞・副詞・序数詞などの表記に専ら用いられるものなどが，多くこのグループに属する。その意味で，高くて細い柱といえよう。

また，Eは，用いられる語の種類も，それらの語の頻度も平均的なものであり，高からず，太からず細からずといった柱になぞらえうるものである。

漢字にウェイトをつける場合，単に頻度や，それに基づく使用率といったもののみに頼っていては，たとえば，表17のGグループや，Cグループの漢字の特殊な性格が把握できない。たとえば，外国人や児童生徒に学習させる漢字の提出順序などを決める場合，どの位の語彙の表記に影響があるか，あるいは，用法が広いかせまいかなどの情報を得ることは有効であろう。

こうした観点からの漢字の性格が，使用率よりは，多角的にとらえうる点でカバー率に基づく，漢字のウェイトづけは，一つの有効な方法ではないかと考える。

8. カバー率の可能性

カバー率の一つの拡張として「音訓カバー率 (Cy)」といったものも考えることができる。これは，データの中に現われた漢字の読み（音訓）が，それぞれ，どれだけの語に用いられているかを，語彙の総体に対する比率で示したものとなる。この概念を式の形で表わすと，つぎのようなことになる。

$$\text{音訓カバー率 (Cy) ‰} = \frac{\text{その音訓の用いられた語の数 (Wy)}}{\text{語彙の総体 (V)}} \times 1000$$

雑誌九十種のデータを用いて，漢字「座」と「坐」の音訓カバー率を求めると，それぞれ，表18のようになる。すなわち，漢字「座」の持っている音「ザ」は，延べ語全体（人名・地名を除く）の0.430パーミル（一万語当り約四語）に用いられ，漢字「坐」の音「ザ」は0.007パーミルの延べ語の表記に使われているということを示している。この計算式は

$$\underset{(Cny)}{\text{延べ語音訓カバー率}}\ \text{‰} = \frac{\text{その音訓の用いられた延べ語数 (Wn}y\text{)}}{\text{延べ語総量 (Vn)}} \times 1000$$

である。

一方，異り語について計算すると，漢字「ザ」の音は，異り語全体（人名・地名を除く）の0.699パーミルすなわち（一万語当り約7語）に用いられ，漢字「坐」の音「ザ」は0.077パーミルの異なり語の表記に使われているということになる。この計算式は

$$\underset{(Cky)}{\text{異り語音訓カバー率}} = \frac{\text{その音訓の用いられた異り語数 (Wk}y\text{)}}{\text{異り語総量 (Vk)}} \times 1000$$

である。

また，「座」の訓「すわる」の，延べ語についての音訓カバー率は0.095パーミル，「坐」の訓「すわる」のそれは0.128，異り語についての音訓カバー率

表 18　「座／坐」の音訓カバー率

座 (124)　Cn=0.546・Ck=0.700			坐 (37)　Cn=0.135・Ck=0.261		
音　訓	Cny ‰	Cky ‰	音　訓	Cny ‰	Cky ‰
ザ	0.430	0.699	ザ	0.007	0.077
すわる	0.095	0.001	すわる	0.128	0.184

（ ）内の数値は度数

表 19　類義的漢字の音訓カバー率

付 (237)　Cn=0.863・Ck=2.021			附 (86)　Cn=0.313・Ck=0.868		
音　訓	Cn y ‰	Ck y ‰	音　訓	Cn y ‰	Ck y ‰
フ	0.153	0.414	フ	0.124	0.257
つ　く	0.153	0.458	つ　く	0.033	0.125
つ け る	0.557	1.150	つ け る	0.156	0.486
個 (146)　Cn=0.532・Ck=0.347			**箇 (23)　Cn=0.084・Ck=0.034**		
音　訓	Cn y ‰	Ck y ‰	音　訓	Cn y ‰	Ck y ‰
カ	0.026	0.023	カ	0.033	0.013
コ	0.506	0.324	コ	0.051	0.021
嘆 (17)　Cn=0.062・Ck=0.295			**歎 (10)　Cn=0.036・Ck=0.126**		
音　訓	Cn y ‰	Ck y ‰	音　訓	Cn y ‰	Ck y ‰
タ　ン	0.036	0.213	タ　ン	0.018	0.097
な げ く	0.026	0.082	な げ く	0.018	0.029
炎 (17)　Cn=0.062・Ck=0.187			**焔 (10)　Cn=0.036・Ck=0.133**		
音　訓	Cn y ‰	Ck y ‰	音　訓	Cn y ‰	Ck y ‰
エ　ン	0.044	0.155	エ　ン	0.018	0.116
ほ の お	0.011	0.013	ほ の お	0.018	0.017
かげろお	0.007	0.019	—	—	—
峰 (2)　Cn=0.007・Ck=0.213			**峯 (2)　Cn=0.007・Ck=0.045**		
音　訓	Cn y ‰	Ck y ‰	音　訓	Cn y ‰	Ck y ‰
ホ　ウ	0.004	0.116	ホ　ウ	0.004	0.039
み　ね	0.004	0.097	み　ね	0.004	0.006

( ) 内の数値は度数

も「座(すわる)」の 0.001 パーミルに対して，「坐(すわる)」は 0.184 と，いずれも，「坐」の方が高率を示している。

したがって，表18の結果によると，音「ザ」の音訓カバー率は，延べ語についても，異り語についても，「座」の方が，「坐」よりも高い。それに対して訓「すわる」の音訓カバー率は，延べ語・異り語とも「坐」の方が高いという

ことになる。

なお，言うまでもないが，「座」なら「座」の，延べ語音訓カバー率 (Cn y) の合計は，当然，その文字の延べ語カバー率 (Cn) に一致し，異り語音訓カバー率 (Ck y) の和は，当然，異り語カバー率 (Ck) に一致する。

いまの「座／坐」のセットと同じように，「付／附」「個／箇」「嘆／歎」「炎／焔」「峰／峯」について調べてみると，表19のようになり，「付／附」「嘆／歎」のセットでは，いずれも「付」「嘆」のもつ音訓が，延べ語においても，異り語においても高いカバー率を示している。しかし「個／箇」のセットでは，音「カ」の延べ語についての音訓カバー率のみ，「箇」の方が「個」よりも高い値を示している。このことは，音「カ」の場合に限って，「箇」の方が「個」よりも，やや広く延べ語をカバーすることを示しているが，この音の用いられる語のバラエティーを示す，異り語音訓カバー率 (Ck y) は，「個」の方が「箇」を上まわっている。

また「炎／焔」においては，音「エン」の音訓カバー率は，延べ語の場合も異り語の場合も「炎」の方が「焔」の数値を上まわっているが，訓「ほのお」については，逆転している。ということは，音「エン」の場合には「炎」の方が広くさまざまな語の表記に用いられ，訓「ほのお」の場合には「焔」の方が，やや広く使われる傾向があることを示している。

最後の「峰／峯」については，余りにデータ数が少なく，この結果からものを言うのは危険だが，異り語音訓カバー率が一致して高いところをみると，あるいは，「ホウ」の場合も「みね」の場合も，「峰」の方が多種類の語の表記に現われるという傾向があるのかもしれない。

つぎに，よく問題になる「燈／灯」のセットをとりあげると，表20のような結果になる。すなわち，このデータに共通に現われた音訓「トウ」と「ひ」についていえば，音「トウ」では延べ語においても，異り語においても，「燈」の音訓カバー率の方が高く，訓「ひ」では「灯」の方が高い。したがって「トウ」の場合には，「燈」の方が広く多種類の語の表記に用いられ，「ひ」の場

表 20　「燈／灯」の音訓カバー率

燈 (32)　Cn=0.117・Ck=0.381			灯 (38)　Cn=0.138・Ck=0.250		
音　訓	Cny ‰	Cky ‰	音　訓	Cny ‰	Cky ‰
トウ	0.106	0.339	トウ	0.051	0.068
ドン	0.007	0.039	ドン	—	—
チン			チン	0.011	0.029
ひ	0.004	0.003	ひ	0.044	0.074
ほ	—	—	ほ	0.007	0.039
あかり	—	—	あかり	0.004	0.006
ともしび	—	—	ともしび	0.018	0.024
ともる	—	—	ともる	0.004	0.010

（ ）内の数値は度数

合には「灯」の方が語の表記に及ぼす影響力が強いということになる。

このように，ある一定の読み（音訓）を共通に持ち，対立・競合の関係にある漢字について，音訓カバー率を計算してみると，表21のようになる（人名・地名を除く）。ここにとりあげた音訓の大部分については，延べ語における音訓カバー率も，異り語におけるそれも，ある漢字に偏りやすいということができる。たとえば「ケイ（形／型）」では「形」に偏り，「め（目／眼）」では「目」に偏っている。また「はかる（計／図／測／量／謀）」では「計→図→測→量・謀」の順になっている。したがって，このデータでみる限り，一定の音訓を共通にもつ漢字の間の競合関係は，延べ語における活躍の広さにおいても，異り語におけるバラエティーの豊かさにおいても，ある特定な漢字が他を圧倒する傾向が強いとみることができる。しかし，「キョウ（凶／兇）」の場合は，延べ語の表記における影響力は「凶」の方が強いといえるが，用いられる語のバラエティーは，差がないという結果になっている。また，最後の「あたたかい（暖／温／煖）」の場合は，延べ語の音訓カバー率は「暖→温→煖」の順であるが，異り語についてのそれは「温→暖→煖」の順になっている。すなわち，字訓としての「あたたかい（—まる・—める）」は，漢字「暖」の訓として，最も広く延べ語をカバーするが，この字訓の用いられる語のバラエティーの多様さ

表 21　共通音訓における漢字の競合関係

音訓	漢字	Cn y ‰	Ck y ‰
ケイ	型	0.302	0.920
	形	0.233	0.435
ヘン	編	0.284	0.450
	篇	0.062	0.183
ヨク	欲	0.164	0.389
	慾	0.036	0.147
カン	罐	0.051	0.205
	鑵	0.015	0.077
キョウ	凶	0.022	0.077
	兇	0.011	0.077
ゲン	弦	0.018	0.077
	絃	0.004	0.039
め	目	3.723	3.675
	眼	0.805	0.311
かわ	川	0.226	0.521
	河	0.044	0.047
まち	町	0.357	0.291
	街	0.128	0.135
はじまる (—める・—めて)	初	0.470	0.147
	始	0.244	0.064
	創	—	—
はかる	計	0.098	0.132
	図	0.044	0.009
	測	0.015	0.003
	量	0.004	0.001
	謀	0.004	0.001
はやい (—まる・—める)	早	0.590	0.765
	速	0.036	0.003
あたたかい (—まる・—める)	暖	0.066	0.028
	温	0.040	0.073
	煖	0.004	0.010

は「温」に及ばないということを示している。そして，「煖」の字訓「あたたかい」は，延べ語における活躍の広さも，これによって表記される語の種類も，「暖・温」の場合に比べて，きわめて限られているという結果になる。

最後に，雑誌九十種のデータについて，延べ語音訓カバー率（Cny）順上位20位までの音訓を示すと，表22の通りである（人名・地名を除く）。延べ語についての音訓カバー率では，やはり「イチ」「ニ」「ジュウ」「ネン」「ブン」など数値情報に関するものが高いカバー率を示し，「おもう」「みる」「て」などの和訓が，それに続いているが，異り語音訓カバー率の方では，数値関係のものは「イチ」を除いては大幅に後退し，「て」「キ」「チュウ」「ジン」「ホ

表 22　音訓カバー率上位20位の音訓

順位	延べ語			異り語		
	音訓	漢字	Cny ‰	音訓	漢字	Cky ‰
1	イチ	一	11.232	テ	手	6.142
2	ニ	二	7.410	イチ	一	4.606
3	ジュウ	十	6.438	キ	気	4.364
4	サン	三	5.713	チュウ	中	4.243
5	ネン	年	5.596	ジン	人	4.201
6	ゴ	五	4.813	ホン	本	4.197
7	ブン	分	3.906	め	目	3.675
8	おもう	思	3.855	みる	見	3.647
9	ジン	人	3.756	だす	出	3.463
10	みる	見	3.753	コク	国	3.446
11	て	手	3.749	ダイ	大	3.401
12	め	目	3.723	ジョウ	上	3.355
13	テキ	的	3.676	ブン	分	3.241
14	キ	気	3.629	こ	小	3.136
15	カイ	会	3.541	ドウ	同	3.134
16	ジ	自	3.454	ヨウ	用	3.130
17	セイ	生	3.399	おお	大	3.114
18	チュウ	中	3.264	チョウ	長	3.070
19	シャ	者	3.239	ナイ	内	3.000
20	ひと	人	3.228	ム	無	2.969

ン」「め」「みる」など，基本語の表記に用いられる音訓が上位に進出している。

以上述べてきた，音訓カバー率に基づく分析結果は，音訓カバー率による音訓のウェイト測定の可能性を物語るものといえよう。従来，音訓についての統計的尺度は，ほとんど用意されていなかったといってよい。しかし，漢字基本度の段階づけと並行して，その音訓のウェイトを測定することは，きわめて重要な課題である。なぜなら，基本漢字と見なされる漢字のもつ音訓のすべてが，表記の上で重要な音訓であるとはいえないからである。

また，音訓整理をめざせば，一つ一つの音訓について，それが，語の表記に，どれだけの影響をもっているかを把握しておく必要がある。

こうした問題に，今回試みた音訓カバー率による音訓のウェイト測定は，一つの手がかりを与えるものではないかと考えられる。

9. おわりに

従来，しばしば，語彙調査などの結果に基づいて，基本語彙とか基礎語彙とかいったものが選定されることがあったが，一方，表記調査・漢字調査の結果出てくる基本漢字の類と，これら基本語彙を表記するのに必要な漢字との間には，同一データによる調査の結果の場合においてすら，少なからぬギャップが生ずることが多かった。これは，いうまでもなく，語のウェイトを測る尺度と，漢字の重みを測定する尺度との間に，統計上の関連性が，全くなかったことに起因している。

ここに述べてきたカバー率の基本的な概念は，ある漢字がなくなった場合，その影響は，どれだけの範囲の語に及ぶか，その影響の広さを測る尺度であり，このことは逆にいえば，ある範囲の語彙を表記する場合の，漢字の必要度を段階づける尺度でもある。したがって，カバー率のような尺度を用いれば，同一のデータから計量的に選定されたものである以上，基本語彙の表記に必要な漢字と，基本漢字との間のギャップは，理論上は起りえないはずである。こ

こにおいて，大量語彙調査と漢字調査とが，統計的尺度の面で結びつくことになる。また，最後に触れた，音訓カバー率は，各漢字のもっている読み(音訓)が，語彙の総体のどれだけに用いられているかを算定し，音訓のウェイトを測る目的で試みたものであるが，表記のゆれの測定や，標準表記の設定などのさいにも利用しうる。

しかし，いずれにしても，カバー率は，その基本的な性格からいって，語彙調査の側における，同語・異語の認定や語の表記の扱い，あるいは語の読みの決め方などが，大きく影響してくる。もちろん，語彙調査の言語単位（語の長さ）の採り方なども，決定的な影響をもたらす。したがって，カバー率の採用に当っては，語彙調査そのもののデザインが，カバー率にゆがみをもたらさないように企画されなくてはならない。

一方，漢字調査・表記調査の側においても，漢字の同字・異字の認定や表記のゆれの扱いなどに慎重な配慮が必要である。なぜなら，それは，カバー率の数値全体に影響を及ぼすからである。

なお，この研究は，昭和51年度文部省科学研究費補助金（一般研究A）による研究の一部である。