

# 漢字のORのための基礎データ

日本語の情報処理においては、なんらかの形で漢字の入力がある。本号の渡辺定久氏の論文“漢字入力装置の操作性”にもあるように現在一般に使われている入力装置にはつぎの4種類がある。

- (1) 多段シフト形
- (2) タブレット形
- (3) 邦文タイプライター形
- (4) カナタイプライター形

本資料編では(1)の多段シフト形入力装置の例として日本科学技術研修所で使用中のTORAY 8511のキーボード(図1)と、(4)のカナタイプライター形入力装置の例としてラインプット(渡辺定久氏の論文参照)の盤面(図2)を示す。後者の場合には48個のキーがあり2回のキーの使用すなわち

2ストロークで入力するので、文字の種類は2304である。

この方式では漢字についてのコードを記憶することになる。図3にその例を示す。

説明 東レ漢字プリンター(TORAY 8511)は入力用の紙テープを盤面(キーボード)を操作することによってつくる。扱う総字数は7056字で、このうち盤面から入る字は3240、盤面外からコードとして入れる盤外字は3816である。このうち盤面上の字の分類はつぎのようにになっている。

漢字: 2698, 英大文字: 26, 英小文字: 26, ギリシア大文字: 18, ギリシア小文字: 11, ひらがな: 79, カタカナ(重複あり): 121, 洋数字: 10, 和数字: 19, ローマ数字: 12, 記号類: 220, 合計: 3240。

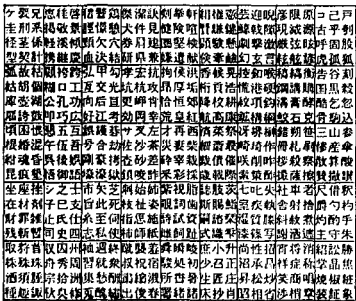


図1 紙テープ出力・キーボード(TORAY 8511)盤面文字種と文字配列

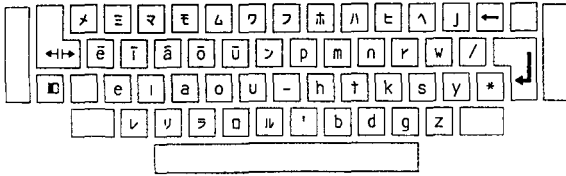


図 2 ラインプットのキーボード (出所: 川上晃, 川上義, タッチ打法による漢字入力, 情報処理, vol. 15 (1974), no. 11, 863~867.)

つぎの文は、ラインプ  
ットで記録したもので  
ある。

tuginoヲha'-rā'in>pū  
wūtōdeヲepsitamode  
'aru-'

世の中には、規格品ぎ  
らいという思想があつ  
て、それはおそらく個  
性尊重主義とつながつ  
ているのだらうが、と  
くに高級知識人によく  
みられる考え方である。  
しかし問題はどんな規  
格品をつかつても、そ  
れによつてつくりださ  
れる知的生産物が個性  
的創造的であればいい  
のである。

j-noānhiha'-ルヨEgi  
ra'ito'i'uz-dyga'awu  
te'-soreha'osorakumb  
j^dw^tūzr^tototunagawu  
te'irunodaro'uga'-to  
kunikn^bsz^f^~niyoku  
mirarerumg'e^ide'aru-'  
sikasi^kkt^hadon-na^  
ヨEwotukawutemo'-so  
reniyowutetukuridasa  
neruszyn^r^u^hgambj^  
yr^dyrd^areba'i'i  
node'aru-'

図 3 ラインプットによる入力例 (出所: 図 2 に同じ)

漢字の使用状況についての統計結果は世の中に  
いくつかある。まず、どのような漢字の集合を利用  
しているかについては、JIS C 6226 情報交換  
用漢字符号系の解説の 3 ページに解説表 2 があり、  
調査対象漢字表一覧として載せてある。ここには

37 表の名称、作成年月などが記されている。この  
漢字集合がどのように利用されているかを、使用  
度数として求めることが考えられる。

表 1 はこの種の資料をまとめたものである。

表 1 各種資料における漢字使用率の比較

資料略名	調査漢字 延べ字数	使用度数順位による使用率累計					調査漢字 異なり字 数
		200 位 まで	500 位 まで	1,000 位 まで	1,500 位 まで	2,000 位 まで	
A 大西	80万	50%	71%	79%	84%	87%	?
B <sub>1</sub> 印刷局	328万	61.8	84.5	95.5	98.3	99.2	3,948
B <sub>2</sub> 印刷局	4,591万	68.8	81.5	94.2	98.0	99.3	5,122
B <sub>3</sub> 凸版印刷	530万	51.8	75.6	91.3	96.9	98.7	4,520
C <sub>1</sub> 婦人雑誌	17万	53	75	90	96	99	3,048
C <sub>2</sub> 総合雑誌	12万	55.5	78.0	92.0	97.1	99.1	2,781
C <sub>3</sub> 雑誌九十種	28万	52.0	74.5	90.0	96.0	98.6	3,328
C <sub>4</sub> 現代新聞	99万	56.1	79.4	93.9	98.4	99.6	3,213
D <sub>1</sub> カナモジカイ	45万	54.0	77.3	91.7	96.9	98.7	3,542
D <sub>2</sub> 朝日	1,711万	47.5	66.0	83.8	94.0	99.4	2,309
D <sub>3</sub> 毎日	32万	56.9	80.3	94.4	98.5	99.7	2,643
D <sub>4</sub> 共同	46万	59.9	84.9	95.7	99.1	99.9	2,279
E ユニバック	272万	80.4	94.9	99.0	99.7	99.9	2,826

資料注

- A 大西雅雄・日本基本漢字(1941)——8分野の刊行物
- B<sub>1</sub> 印刷局・本邦常用漢字の研究(第4回)(1941)——第55, 65, 67帝国議会議院本会議速記録
- B<sub>2</sub> 印刷局・常用文字の調査——1962~66年鑄造掛取箱作業
- B<sub>3</sub> 凸版印刷・漢字出現頻度数調査(第2回)(1976)——単行本・雑誌・辞書・百科辞書等
- C<sub>1</sub> 国立国語研究所・婦人雑誌の用語(1953)——1950年1~12月, 主婦の友\*
- C<sub>2</sub> 国立国語研究所・総合雑誌の用字(1960)——1953年7月~54年6月, 13誌\*
- C<sub>3</sub> 国立国語研究所・雑誌九十種の用語用字II(1963)——1956年1~12月, 5部門90誌\*
- C<sub>4</sub> 国立国語研究所・現代新聞の漢字(1976)——1966年1~12月, 3紙朝夕刊\*

- D<sub>1</sub> カナモジカイ(岡崎常太郎・漢字制限の基本的研究・1938)——1935年1~12月, 5紙60日分
- D<sub>2</sub> 朝日新聞社(東京)・活字使用度数調査——1948年11月~49年7月
- D<sub>3</sub> 毎日新聞社(大阪)・活字使用度数調査表——1953年3~5月, 7日分
- D<sub>4</sub> 共同通信社・文字出度調査100万字集計結果——1971年7月15~21日
- E 日本ユニバック・漢字システムデザイン資料(4)——1971年, 66万人の姓名

\* C<sub>1</sub>~<sub>4</sub>は, すべて標本調査である。なお, C<sub>1</sub>の数値はC<sub>3</sub>での引用による。C<sub>1</sub>では『婦人生活』でも6万字の調査を行なったが, ここには加えてない。(出所: 林大, 漢字の問題, 岩波講座日本語, 第3巻101-134, 1977.)

日本語で記述された情報に含まれる漢字の出現度数の統計の例として, 表1に示されているB<sub>3</sub>の凸版印刷株式会社の昭和51年の調査から, 使用度数の高い100文字を選んで示すと表2のようになる。

(出所: 凸版印刷株式会社, 漢字出現頻度数調査(II)昭和51年2月。これは表1のなかのB<sub>3</sub>である。)調査対象漢字本数はつぎのとおり: (イ)単行本 37点3,041ページ, (ロ)雑誌22点1,259ページ, (ハ)用語辞典4点1,922ページ, (ニ)その他辞典5点2,921ページ, (ホ)百科辞典2点1,820ページ 漢字本数5,302,308本。

表2 多頻度漢字百選 (注: 当用漢字でないものは56番めの○)

NO	漢字	度数%	累積度数	NO	漢字	度数%	累積度数	NO	漢字	度数%	累積度数	NO	漢字	度数%	累積度数
1	一	1.215	1.214	26	発	0.426	15.438	51	事	0.304	24.146	76	数	0.233	30.714
2	年	1.066	2.281	27	二	0.420	15.858	52	気	0.303	24.449	77	手	0.233	30.947
3	人	0.876	3.157	28	化	0.412	16.271	53	力	0.300	24.749	78	通	0.224	31.171
4	大	0.794	3.951	29	間	0.410	16.680	54	同	0.290	25.039	79	九	0.224	31.394
5	的	0.738	4.688	30	体	0.399	17.080	55	高	0.286	25.325	80	戦	0.223	31.618
6	国	0.722	5.410	31	主	0.393	17.473	56	○	0.285	25.610	81	最	0.221	31.839
7	中	0.703	6.113	32	性	0.389	17.861	57	長	0.280	25.889	82	以	0.220	32.059
8	学	0.616	6.729	33	時	0.380	18.242	58	入	0.273	26.163	83	形	0.220	32.279
9	生	0.584	7.313	34	動	0.377	18.618	59	教	0.273	26.435	84	実	0.219	32.498
10	分	0.564	7.877	35	三	0.357	18.976	60	民	0.272	26.707	85	経	0.219	32.717
11	地	0.551	8.428	36	代	0.354	19.329	61	理	0.269	26.977	86	平	0.218	32.936
12	日	0.545	8.973	37	後	0.337	19.666	62	下	0.265	27.241	87	類	0.218	33.154
13	合	0.511	9.484	38	定	0.336	20.003	63	前	0.265	27.506	88	心	0.217	33.371
14	行	0.499	9.983	39	家	0.334	20.336	64	水	0.258	27.764	89	目	0.216	33.587
15	出	0.486	10.470	40	自	0.333	20.670	65	政	0.256	28.020	90	機	0.216	33.803
16	物	0.472	10.941	41	作	0.333	21.003	66	現	0.256	28.276	91	産	0.213	34.016
17	法	0.466	11.408	42	場	0.323	21.326	67	内	0.255	28.531	92	第	0.213	34.229
18	上	0.460	11.868	43	立	0.320	21.646	68	五	0.251	28.781	93	結	0.213	34.449
19	者	0.455	12.322	44	世	0.319	21.965	69	原	0.248	29.029	94	全	0.213	34.655
20	部	0.453	12.775	45	成	0.317	22.282	70	見	0.246	29.275	95	山	0.210	34.865
21	本	0.452	13.228	46	業	0.317	22.599	71	文	0.243	29.518	96	面	0.209	35.074
22	子	0.452	13.679	47	多	0.317	22.916	72	関	0.242	29.760	97	期	0.209	35.283
23	方	0.446	14.125	48	度	0.309	23.225	73	要	0.242	30.002	98	名	0.207	35.490
24	用	0.446	14.571	49	対	0.309	23.534	74	小	0.241	30.243	99	社	0.206	35.696
25	会	0.441	15.012	50	月	0.309	23.842	75	四	0.233	30.481	100	表	0.206	35.902

もう一つの漢字使用度数の調査として、表1に示されているC<sub>4</sub>の現代新聞の漢字使用度数分析から一部を引用すると表3のようになる。

これは新聞記事を政治、社会、経済などの6個の層にわけ、層別の五十選を示す。運動のところを見るといかにも野球の記事が多そうである。

表3 現代新聞漢字調査——多頻度漢字層別五十選

層内順位	漢字	層内使用率	全体順位	層内順位	漢字	層内使用率	全体順位	層内順位	漢字	層内使用率	全体順位	層内順位	漢字	層内使用率	全体順位	層内順位	漢字	層内使用率	全体順位				
第I層(政治)				第II層(社会)				第III層(経済)				第IV層(文化・家庭)				第V層(運動・芸能)				第VI層(広告)			
1	日	18.151	1	1	十	18.273	3	1	一	19.176	2	1	一	15.148	2	1	日	18.768	1	1	日	16.484	1
2	国	17.143	9	2	一	17.317	2	2	十	15.720	3	2	日	13.529	1	2	大	16.263	5	2	円	15.940	19
3	会	14.706	8	3	日	16.192	1	3	日	14.888	1	3	人	12.695	6	3	一	15.905	2	3	東	11.961	13
4	十	14.518	3	4	二	16.176	4	4	二	13.697	4	4	大	10.523	5	4	二	12.586	4	4	一	10.109	2
5	一	13.938	2	5	三	11.851	7	5	三	12.177	7	5	学	9.962	23	5	三	10.861	7	5	大	9.315	5
6	大	9.501	5	6	人	11.511	6	6	大	10.556	5	6	十	9.738	3	6	十	10.316	3	6	本	8.954	12
7	二	9.485	4	7	五	8.444	14	7	年	9.667	10	7	二	9.025	4	7	人	10.295	6	7	京	7.768	27
8	政	8.649	50	8	年	8.398	10	8	業	9.624	47	8	年	8.848	10	8	木	9.633	12	8	新	7.664	31
9	議	8.441	53	9	年	8.058	5	9	株	8.433	240	9	木	7.855	12	9	中	8.156	11	9	中	6.845	11
10	中	7.769	11	10	同	7.589	20	10	上	8.218	18	10	生	7.558	24	10	手	7.514	43	10	月	6.295	21
11	人	6.777	6	11	会	7.320	8	11	五	8.161	14	11	三	7.109	7	11	田	6.928	33	11	人	6.069	6
12	年	6.777	10	12	四	7.299	16	12	四	7.946	16	12	中	6.700	11	12	出	6.914	17	12	田	5.623	33
13	三	6.673	7	13	時	6.758	15	13	出	7.501	17	13	会	6.404	8	13	時	6.893	15	13	会	5.592	8
14	長	6.669	22	14	東	6.519	13	14	月	7.272	21	14	時	6.155	15	14	勝	6.383	198	14	年	5.562	10
15	党	6.481	109	15	中	6.431	11	15	国	6.942	9	15	国	5.835	9	15	打	6.251	185	15	上	5.476	18
16	同	6.133	20	16	事	6.129	26	16	円	6.798	19	16	子	5.466	41	16	東	6.203	13	16	三	5.391	7
17	米	6.033	91	17	上	5.839	18	17	高	6.770	49	17	出	5.370	17	17	回	6.141	62	17	代	5.342	58
18	対	5.873	56	18	学	5.743	23	18	会	6.684	8	18	五	5.186	14	18	分	5.996	28	18	社	5.238	39
19	相	5.725	69	19	六	5.483	35	19	社	6.354	39	19	上	5.121	18	19	四	5.969	16	19	学	5.055	23
20	事	5.713	26	20	出	5.357	17	20	場	6.081	37	20	者	4.849	29	20	年	5.962	10	20	子	5.012	41
21	発	5.629	38	21	区	5.328	72	21	金	5.909	59	21	東	4.793	13	21	子	5.913	41	21	生	4.963	24
22	行	5.537	25	22	本	5.298	12	22	六	5.852	35	22	分	4.785	28	22	会	5.686	8	22	時	4.951	15
23	員	5.489	51	23	生	5.126	24	23	壳	5.837	130	23	気	4.761	68	23	五	5.603	14	23	電	4.682	126
24	本	5.405	12	24	行	5.097	25	24	本	5.837	12	24	的	4.705	44	24	戦	5.368	86	24	区	4.260	72
25	的	5.329	44	25	地	4.984	36	25	百	5.522	77	25	方	4.665	32	25	山	5.175	65	25	壳	4.168	130
26	民	5.117	90	26	長	4.929	22	26	期	5.393	122	26	間	4.641	30	26	上	5.141	18	26	高	4.162	49
27	上	4.897	18	27	八	4.820	46	27	合	5.221	40	27	家	4.592	83	27	後	5.113	42	27	店	4.150	208
28	問	4.865	76	28	京	4.803	27	28	資	5.178	220	28	教	4.520	100	28	場	5.030	37	28	品	4.009	137
29	四	4.861	16	29	前	4.774	34	29	九	5.149	55	29	自	4.304	16	29	間	4.733	30	29	分	3.930	28
30	方	4.637	32	30	月	4.623	21	30	産	5.120	137	30	自	4.192	54	30	生	4.451	24	30	金	3.906	59
31	出	4.629	17	31	国	4.593	9	31	八	5.120	46	31	長	3.911	22	31	前	4.361	34	31	地	3.833	36
32	五	4.597	14	32	者	4.585	29	32	小	4.991	64	32	行	3.871	25	32	第	4.271	80	32	発	3.771	38
33	者	4.357	29	33	員	4.468	51	33	同	4.848	20	33	新	3.751	31	33	者	4.181	29	33	用	3.722	102
34	連	4.229	92	34	間	4.346	30	34	買	4.805	326	34	思	3.727	143	34	見	4.147	48	34	場	3.637	37
35	定	4.153	60	35	市	4.321	75	35	千	4.776	78	35	見	3.703	48	35	演	4.037	219	35	下	3.576	73
36	府	4.145	156	36	田	4.321	33	36	中	4.776	11	36	作	3.703	93	36	八	3.981	46	36	名	3.569	117
37	時	4.065	15	37	部	4.275	45	37	七	4.747	57	37	月	3.695	21	37	国	3.954	9	37	内	3.490	52
38	月	4.045	21	38	七	4.233	57	38	万	4.733	98	38	地	3.647	36	38	六	3.933	35	38	二	3.490	4
39	合	4.025	40	39	後	4.048	42	39	気	4.704	68	39	事	3.599	26	39	京	3.919	27	39	国	3.410	9
40	決	4.009	96	40	九	4.036	55	40	前	4.604	34	40	手	3.583	43	40	行	3.774	25	40	特	3.404	155
41	開	3.909	79	41	分	4.019	28	41	発	4.590	38	41	後	3.543	42	41	野	3.774	87	41	都	3.270	89
42	社	3.817	39	42	方	3.893	32	42	方	4.518	32	42	前	3.543	34	42	合	3.733	40	42	全	3.258	63
43	題	3.813	103	43	午	3.851	100	43	化	4.389	97	43	京	3.502	27	43	球	3.726	379	43	町	3.093	125
44	軍	3.761	190	44	町	3.830	125	44	度	4.331	71	44	部	3.486	45	44	選	3.643	118	44	映	3.087	232
45	共	3.745	158	45	場	3.759	37	45	東	4.331	13	45	私	3.470	221	45	日	3.560	66	45	作	3.087	93
46	委	3.681	170	46	百	3.721	77	46	長	4.317	22	46	場	3.406	37	46	月	3.547	21	46	式	3.087	182
47	間	3.681	30	47	都	3.696	89	47	部	4.274	45	47	高	3.366	49	47	新	3.547	31	47	入	3.056	74
48	内	3.669	52	48	業	3.624	47	48	行	4.145	25	48	代	3.270	58	48	長	3.443	22	48	行	3.044	25
49	見	3.645	48	49	手	3.582	43	49	的	4.088	44	49	文	3.262	141	49	高	3.409	49	49	方	2.983	32
50	首	3.593	206	50	社	3.415	39	50	価	3.959	203	50	合	3.246	40	50	作	3.409	93	50	出	2.964	17

注：教育漢字ではないが、小学校での学習を認められた当用漢字として第VI層44番めの「映」がある。

出所：国立国語研究所，現代新聞の漢字，国立国語研究所報告56，1976，秀英出版発行（●162 東京都新宿区納戸町40）。

層別の各延べ調査字数は7万字から25万字の間にあるが、それらの延べ字数が、どの程度の異なり字数から成るかが表4に示してある。これによ

ると経済関係の異なり字数が1,557に対し、広告の関係が2,559字である。

表4 現代新聞漢字調査——層別の使用度数

	異 な り										延 べ
	当 用 漢 字					表 外 漢 字					
	教育漢字	準教育漢字	⊖補正漢字	その他	小 計	⊕補正漢字	人名用漢字	その他	小 計	計	
I 政治	873	113	7	713	1,707 (85.8)	21	52	211	284 (14.2)	1,990 (100.0)	249,962
II 社会	877	114	10	772	1,773 (79.4)	25	73	360	458 (20.6)	2,231 (100.0)	238,383
III 経済	836	99	3	496	1,434 (92.1)	15	28	80	123 (7.9)	1,557 (100.0)	69,722
IV 文化家庭	880	115	15	754	1,764 (74.9)	25	68	498	591 (25.1)	2,355 (100.0)	124,769
V 運動芸能	878	114	13	739	1,744 (71.0)	25	75	611	711 (29.0)	2,455 (100.0)	144,962
VI 広告	877	113	15	731	1,736 (67.8)	24	78	721	823 (32.2)	2,559 (100.0)	163,613

(出所：国立国語研究所，表3に同じ)

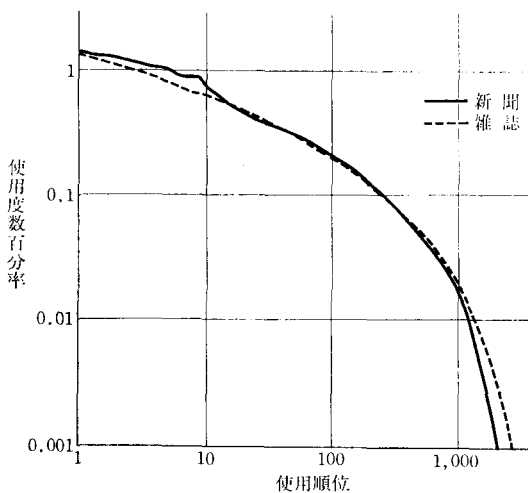


図4 現代新聞漢字調査，使用順位と使用度数百分率の関係

(出所：国立国語研究所，表3に同じ)

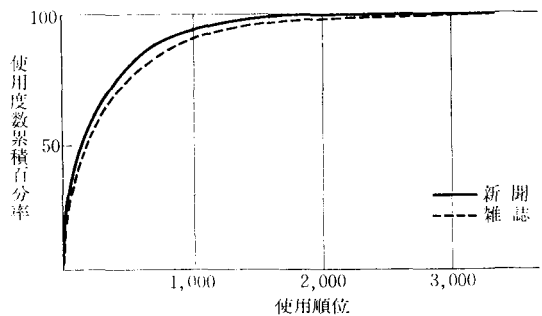


図5 現代新聞漢字調査，使用順位と使用度数累積百分率の関係

(出所：国立国語研究所，表3に同じ)

漢字を使用度数に関して区切ったとき、各使用度数区間に対応する異なり漢字がどの程度あるかに関心をもつことになろう。これを示すのが表5である。

また、図5は使用順位に沿って漢字を並べたとき、使用度数の累積曲線がどのようになるかを示す。

表 5 現代新聞漢字調査——標本全体の度数分布と漢字の種類

度数区間	当用漢字				表外漢字				合計		累 計		
	教育漢字	準教育漢字	⊖補正漢字	その他	⊕補正漢字	人名用漢字	その他	異なり	延 べ	異なり(%)	延 べ(%)		
～10001	5							5	64,307	5	0.2	64,307	6.5
10000～ 5001	13							13	87,427	18	0.6	151,734	15.3
5000～ 4001	9							9	40,275	27	0.8	192,009	19.4
4000～ 3001	26							26	91,410	53	1.6	283,419	28.6
3000～ 2001	50							50	121,018	103	3.2	404,437	40.8
2000～ 1001	144	1		2		1		148	208,294	251	7.8	612,731	61.8
1000～ 501	219	13		14			3	249	174,659	500	15.6	787,390	79.4
500～ 401	69	5		8				82	36,524	582	18.1	823,914	83.1
400～ 301	83	10		30		2	1	126	43,196	708	22.0	867,110	87.5
300～ 201	96	24		43		1	1	165	41,528	873	27.2	908,638	91.7
200～ 101	103	33		147	4	8	7	302	43,830	1,175	36.6	952,468	96.1
100～ 91	11	5		32	2	1	3	54	5,105	1,229	38.2	957,573	96.6
90～ 81	8	1	1	27	1		3	41	3,510	1,270	39.5	961,083	96.9
80～ 71	9	4		37	1	4	2	57	4,264	1,327	41.3	965,347	97.4
70～ 61	7	2		51		4	1	65	4,265	1,392	43.3	969,612	97.8
60～ 51	9	6		51	1	5	8	80	4,457	1,472	45.8	974,069	98.3
50～ 41	6	2	1	44	5	5	11	74	3,347	1,546	48.1	977,416	98.6
40～ 31	6	2		74	2	3	17	104	3,745	1,650	51.3	981,161	99.0
30～ 21	2	2	2	69	1	8	43	127	3,223	1,777	55.3	984,384	99.3
20～ 11	2	4	3	104	5	13	74	205	3,109	1,982	61.7	987,493	99.6
10			1	7	1	3	19	31	310	2,013	62.6	987,803	99.6
9				12		6	28	46	414	2,059	64.1	988,217	99.7
8				12	1	5	30	48	384	2,107	65.6	988,601	99.7
7			1	11	2	3	26	43	301	2,150	66.9	988,902	99.8
6	1	1		4	2	2	45	55	330	2,205	68.6	989,232	99.8
5			2	8		2	60	72	360	2,277	70.8	989,592	99.8
4			4	6	1	1	81	93	372	2,370	73.7	989,964	99.9
3	1		1	15		4	130	151	453	2,521	78.4	990,417	99.9
2	1		4	8		2	251	266	532	2,787	86.7	990,949	99.9
1	1		4	8		3	410	426	426	3,213	100.0	991,375	100.0
合 計	881	115	24	824	29	86	1,254	3,213	991,375				

(出所：国立国語研究所，表3に同じ)







字形の複雑さを示す目安として画数がある。表6には『新字源』に収めてある漢字の画数分布を示す。つぎに、表7には現代新聞漢字調査によって求められた使用率の高い字数の画数の平均を順位ごとに求めたものを示す。

表6 総画数の分布

(出所：林大、表1に同じ)  
注：漢字は『新字源』所収の9,767字。

画数	字数
1	2
2	26
3	60
4	133
5	192
6	256
7	441
8	621
9	703
10	784
11	825
12	883
13	798
14	703
15	704
16	604
17	494
18	377
19	322
20	224
21	192
22	131
23	104
24	75
25	48
26	23
27	22
28	9
29	6
30	2
31	1
32	1
33	1
計	9,767

表7 使用率と画数

使用率順位	字数	平均画数
1~50	50	6.4
51~100	50	7.3
101~150	50	8.4
151~200	50	8.2
501~550	52	10.0
1,001~1,048	50	11.4
1,496~1,539	51	11.1
1,958	56	11.9

(出所：林大、表1に同じ)

注：使用率順位は現代新聞漢字調査による。

JIS規格では6,349字の漢字がコードを与えられたが、これらの漢字の音訓がどうなっているかを示す一つの手がかりとして、表8にJIS漢字の音訓分布表を示す。これによると、音訓の多いのは、し、か、こ、き、はの順である。

表8 JIS漢字の音訓分布(%)

	全漢字	第1水準	第2水準		全漢字	第1水準	第2水準
あ	3.54	3.40	3.67	ね	0.41	0.43	0.38
い	2.92	2.90	2.95	の	0.74	0.91	0.58
う	1.98	2.09	1.89	は	4.03	4.15	3.91
え	1.57	1.40	1.73	ひ	3.21	3.02	3.39
お	3.18	3.24	3.13	ふ	2.31	2.23	2.39
か	8.00	7.73	8.26	へ	1.11	0.83	1.38
き	4.62	4.59	4.66	ほ	2.91	2.88	2.93
く	2.12	2.18	2.06	ま	1.59	1.80	1.40
け	2.67	2.34	2.99	み	1.29	1.47	1.12
こ	4.95	4.68	5.22	む	0.71	0.70	0.72
さ	3.29	3.24	3.34	め	0.58	0.68	0.47
し	9.41	9.87	8.96	も	0.95	1.10	0.80
す	1.90	1.85	1.94	や	1.09	1.24	0.94
せ	2.99	2.84	3.13	ゆ	1.02	1.15	0.89
そ	2.59	2.68	2.50	よ	1.75	1.71	1.78
た	3.97	4.17	3.77	ら	0.60	0.36	0.84
ち	2.06	2.14	1.99	り	1.54	1.40	1.68
つ	2.14	2.39	1.90	る	0.13	0.11	0.16
て	1.70	1.60	1.80	れ	0.78	0.61	0.94
と	3.55	3.72	3.39	ろ	0.82	0.65	0.98
な	1.30	1.35	1.26	わ	1.05	1.13	0.96
に	0.66	0.70	0.61				
ぬ	0.28	0.32	0.24				
				100.01	99.98	100.03	

出所：日本科学技術研修所漢字処理システムグループ資料(1978-04-05)

説明：

1. 表の内容：各漢字に音および訓を最大1個ずつ与え、五十音ごとに集計を行ない百分比で示してある。
2. 漢字の種類：JIS C6226に収めてある6,349字に関する五十音ごとの百分比を『全漢字』の列に、第1水準2,965字、第2水準3,384字に関する百分比をそれぞれ『第1水準』『第2水準』の列に示してある。

やじま・けいじ

日本科学技術研修所