

国立国語研究所学術情報リポジトリ

An experimental study on usage of Kanji (Chinese characters) in newspaper

メタデータ	言語: jpn 出版者: 公開日: 2017-03-31 キーワード (Ja): キーワード (En): 作成者: 野村, 雅昭, NOMURA, Masaaki メールアドレス: 所属:
URL	https://doi.org/10.15084/00000999

新聞使用漢字の試行的分析

野村雅昭

第三資料研究室では、昭和42年度から、「新聞語彙調査⁽¹⁾に伴う漢字および表記の研究」を行なっている。これは、語彙調査によって作成されたデータに、さらに、電子計算機による処理を施し、各種漢字表・表記一覧表などを作成して、その分析・記述を行なおうというものである。最初に、漢字調査から着手することにし、昭和42年度には、準備段階として、機械処理による手順の研究を行なった。そして、43年度には、実際の機械処理プログラムの作成および一紙一年分⁽²⁾（全体の約1/3の量）についての機械処理を実施した。この処理システムは、長単位語のデータによるもので、層別度数付きの各種配列順の漢字表、長単位語の用例表および出典表などを作成しようとするものである。その処理システムの一部については、以前に述べたことがある⁽³⁾。

ここで述べようとするのは、そのうち、朝日新聞朝刊（1～6月分）の結果についてである。（以下、これをA調査とよぶ。）先にも述べたが、これは長単位語のデータによるものなので、漢字のよみ（音訓の別）やどのような語に使用されたか（人名・地名などの別）について、くわしいことは、まだわかっていない。したがって、ここでは、結果にあらわれた、漢字の数量的データについて分析を加えることによって、新聞における漢字の量的構造を解明する手がかりをつかむとともに、その分析法についても、ある種の試みを行なうことを目的とする。それによって、新聞における漢字使用の一端をうかがうことができれば幸いである。

- (1) 新聞語彙調査については、「電子計算機による国語研究」（国研報告31）を参照のこと。
- (2) 標本全体は、昭和41年の、朝日・毎日・読売の朝夕刊の全紙面から、1/60の抽出比でサンプリングしたものである。
- (3) 「漢字調査の機械処理について」（国研LDP（月報別冊）1）

1. 新聞使用漢字の構造

戦前・戦後を通じて、用字調査あるいは活字調査といわれるものは、多くの数に上る。そして、現在も、各種の機関によって行なわれているものと思われる。それらの調査結果について、比較・分析を試みたものも、すでにくつかが発表されている(4)。上述のように、ここで問題とする調査結果は、部分的なものであり、既に行なわれた調査との完全な意味での比較はむずかしいと思われる。そこで、比較的最近行なわれたいくつかの調査(5)とくらべてみることによって、新聞漢字使用の特徴を考えてみることにしたい。対象とした調査を、以下、次のような略号でよぶことにする。

M調査…毎日新聞、S調査…総合雑誌、

G調査…現代雑誌九十種、N調査…NHK

表1は、各調査のことなり字数とのべ字数とを比較したものである。これで見ると、調査の対象によって、ことなり字数とのべ字数との間に一定の関係が生じることが予想できる。すなわち、調査対象がいろいろな内容を含んでいるものは、のべ字数が増加するにつれて、ことなり字数も増えつづけるが、内容が一定の傾向しか持たないものは、いくらのべ字数が多くなっても、ことなり字数は、それほど多くは、なりそうもない。新聞にかぎっていえば、これよりも調査規模を大きくしても、ある程度まで、ことなり字数は、ふえるであろうが、M調査やA調査の結果から、そう隔ったものにはならないであろう。その

表1 各種調査のことなり字数とのべ字数

	ことなり字数			のべ字数
	当用漢字	表外字	合計	
M	1,801	812	2,613	317,280
S	1,777	1,004	2,781	117,149
G	1,835	1,493	3,328	280,094
N	1,689	430	2,109	338,129
A	1,773	650	2,423	202,480

表2 各種調査の総使用度数中の表外字の割合

M	2.6%
S	3.5
G	4.5
A	2.3

推測の根拠となるもう一つの理由がある。それは、新聞では、当用漢字表の範囲を守ることに、きわめてきびしい態度をとっているからである。当用漢字は1850字であり、当用漢字補正案⁽⁶⁾によって加える字や人名用漢字をいれても、せいぜい2000字程度であって、ことなり字数が多いということは、表外字が多いということである。したがって、新聞では、ことなり字数が飛躍的に増えるということはまず考えられない。

表2からもそうした事情をうかがうことができる。新聞における表外字の使用率は、M調査以降、さらに小さくなっているという推測をたてても、ほぼまちがってはいないだろう。(M調査には、広告が含まれていない。A調査の結果から、広告に使用されたものを除くと、表外字の使用率は、もうすこし小さくなると思われる。)しかしながら、別の理由によって、表外字の使用率が、これ以上、極端に減少することもないであろう。なぜならば、新聞における表外字の使用は、広告を除いて、ほとんどが、固有名詞にかぎられており、その方針はかなりよく徹底されていると考えられるからである。

上に述べた、ことなり字数とのべ字数の関係は、全体のある割合を占めるのに必要とする漢字数という考え方にも、あてはめることができる。図1は、表3・表4などをもととして使用度数の多い順にならべた漢字の度数の上位のも

図1 各調査の漢字数と使用率の関係

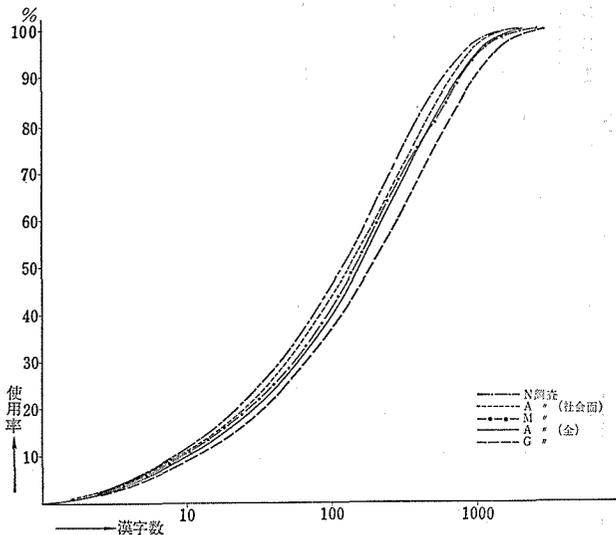


表3 A調査(朝月新聞朝刊1~6月分)使用度数分布表

	ことなり 度数	のべ度数	累 計		累 計 の べ	
			度 数	使用率(%)	度 数	使用率(%)
~1001	23	35,020	23	0.9	35,020	17.3
1000~ 501	52	34,711	75	3.1	69,731	34.4
500~ 301	94	36,313	169	7.0	106,044	52.4
300~ 201	97	24,000	266	11.0	130,044	64.2
200~ 101	231	32,493	497	20.5	162,537	80.3
100~ 91	42	3,989	539	22.2	166,526	82.2
90~ 81	46	3,894	585	24.1	170,420	84.2
80~ 71	52	3,918	637	26.3	174,338	86.1
70~ 61	60	3,920	697	28.8	178,258	88.0
60~ 51	74	3,843	771	31.8	182,371	90.1
50~ 41	82	3,747	853	35.2	186,118	91.9
40~ 31	120	4,288	973	40.2	190,406	94.0
30~ 21	164	4,084	1,137	47.0	194,490	96.1
21~ 11	304	4,521	1,441	59.5	199,011	98.3
10	30	300	1,471	60.7	199,311	98.4
9	46	414	1,517	62.6	199,725	98.6
8	52	416	1,569	64.8	200,141	98.8
7	48	329	1,617	66.7	200,470	99.0
6	56	343	1,673	69.0	200,813	99.2
5	65	325	1,738	71.4	201,138	99.3
4	97	388	1,835	75.7	201,526	99.5
3	99	297	1,934	79.8	201,823	99.7
2	168	336	2,102	86.8	202,159	99.8
1	321	321	2,423	100.0	202,480	100.0
合 計	2,423	202,480				

表4 全体のある割合に達する使用度数の順位

	M調査	S調査	G調査	N調査	A調査
全体の50%	153	160	176	?	156
80%	494	546	638	370	491
85%	635	689	777	464	607
90%	780	890	992	597	768
95%	1,070	1,234	1,358	813	1,044
98%	1,414	1,665	1,832	1,099	1,388
99%	1,650	2,003	2,157	1,309	1,614
100%	2,613	2,781	3,328	2,124	2,423

表 5

	ことなり 字数	のペ字数	総使用字数に対す るのペ字数の比率
当 用 漢 字	1,773	197,885	97.7
教育漢字 教育漢字に準 ずる字 非教育漢字 補正案で削られ る予定のもの	878	175,034	86.4
	112	5,544	2.7
	769	17,180	8.5
	14	127	0.1
表 外 字	650	4,595	2.3
補正案で加え られる予定の もの 人名用漢字 そ の 他	27	318	0.2
	76	1,250	0.6
	547	3,027	1.5
計	2,423	202,480	100.0

○ 教育漢字に準ずる字とは、昭和43年7月11日に告示された新学習指導要領で小学校での学習を認められたものをいう。

○ 補正案で加えられるものの中には、人名用漢字と重複する尚・杉・斉・竜の4字が含まれている。

のから累積した分布曲線である。(S調査は、図が繁雑になるので、省いてある。実際は、A調査とG調査の間にある。)ほぼ、同一の曲線を描いているが、のペ字数の多少にかかわらず、ことなり字数が少ないものほど、少ない字数で全体をまかなえる率が高い傾向を示している。M調査とA調査がほとんど同一のカーブを描いているのは、新聞の使用度数分布に一定の傾向があることを物語っているものといえよう。N調査が、新聞に比べて、少ない字数でまかなわえる率が高いのは、新聞にくらべて、ニュースでとり上げられる内容が限られているからであろう。参考としてあげた、A調査の社会面の分布曲線のほうがN調査に近いことが、この推論を裏づけていると思われる。

表5は、A調査について、字種別に、総使用字数に対する使用比率を示したものである。総使用字数のうち、97.7%が当用漢字表内の漢字でまかなわれているわけである。これに補正案で加えられる予定になっているものをたすと、100字の内、約98字が当用漢字であるという結果になる。そのうち、教育漢字の占める割合は、新学習指導要領で、小学校で学習を認められたものを加えると、約89%である。こうした数字からも、現代の新聞では、使用率という点からみると、かなり少ない漢字数で全体がまかなわれているということがうかがわれる。

表6 よく使用される表外字

順位	A調査	M調査	G調査
1	藤	阪	頃
2	迄	藤	云
3	阪	岡	藤
4	崎	崎	袖
5	岡	菱	衿
6	伊	之	僕
7	菱	阿	岡
8	塚	鶴	崎
9	之	鳩	脇
10	杉	淀	誰
11	彦	奈	阪
12	幡	塚	之
13	竜	彦	伊
14	韓	杉	巾
15	乞	堀	廻
16	須	龍	裾
17	仙	熊	糰
18	旭	鹿	那
19	荻	弘	杉
20	鹿	幡	竜
21	熊	畿	鹿
22	奈	錦	頁
23	窪	旭	錦
24	鶴	韓	或
25	鎌	笠	戻
26	阿	辻	管
27	尚	垣	吾
28	堀	亀	貫
29	弘	塚	唄
30	頁	云	彦

表外字は、のべ字数の全体に対する割合は低い、ことなり字数では、約27%を占めている。表6は、表外字のうち、使用度数の多い順に、三十位までを示したものである。A調査とM調査をくらべると、かなりの一致がみられる。A調査にあって、M調査にみえない、迄・乞・頁などは、主として広告欄で用いられたものである。G調査で高い順位の頃・云・袖・衿・僕・誰などが新聞ではみられないのは、新聞では、人名・地名などの固有名詞以外の表外字に対してきびしい態度をとっていることを裏がきしている。さきに述べた、新聞における表外字の使用率が、これ以上大きく変化することはないだろうというのは、このような意味からである。

今、かりに、漢字調査の規模を、その母集団に含まれると仮定することなり字数の、一定の(かなり高い)割合を得ることができる量という意味で考えるならば、新聞は、その大きいほうのグループに属するだろう。(大小の境界をどこに置くかということは、今は問題としない。また、ここで比較の対象とした調査には、小さいほうのグループに属する

ものは含まれていない。)しかしながら、そのグループの中で、ことなり字数の大小という点からみると、新聞は少ないほうに属すると思われる。新聞が内容的には、かなり広範囲にわたるものを含みながら、少ない字数でまかなわれている理由は、くり返し述べるように、範囲としての当用漢字表がきびしく守られているからである。(それによって新聞の表記に、具体的な、どのような特色があらわれているかということをはっきりとすることは、重要な問題だ

が、今は、それにふれている余裕がない。)つまり、新聞の漢字の構造は、上述のような意味で、二面的な性格を持っているということができよう。今後、新聞の漢字ということを考えていく場合に、こうした位置づけをする必要があると思われる。

(4) 斎賀秀夫「漢字使用の実情」(「言語生活」121)

同 「現代における漢字・漢語の実態」(「文学・語学」41)

菅野謙「漢字から見たラジオニュースの用語」(NHK文献月報1968.9)

(最後の論文には、各種の漢字調査について、要を得た解説があり、各種調査の結果の比較が試みられている。)

(5)○毎日新聞…昭和28年3月から5月までの3カ月間の新聞から抽出された7日分84頁の紙面の文字使用度数調査。(「本社使用活字使用度数調査表」毎日新聞大阪本社(昭28)所収)

○総合雑誌…昭和28年7月号から昭和29年6月号までの総合雑誌13種から抽出された標本の調査。(国研報告19(昭30)所収)

○現代雑誌九十種…昭和31年1月号から12月号までの五部門九十種の雑誌から抽出された標本の調査。(国研報告22(昭38)所収)

○NHKラジオニュース…昭和40年度のラジオニュースの放送原稿の漢字調査。

(注(4)論文所収)

(新聞の調査としては、昭和23年11月から昭和24年7月にわたる朝日新聞の調査と昭和21年12月から昭和23年11月にわたる毎日新聞の調査があるが、どちらも、使用文字調査ではなく、鋳造した活字の本数の調査なので、ここでは、取り上げないことにした。)

(6) 昭和29年3月に、国語審議会で、漢字部会から、当用漢字表を補正する際の資料として報告されたもの。新聞では、すでに実施している。

2. 層別による漢字使用の傾向

新聞語彙調査では、四つの角度からなる、層別区分を行なっている(7)。漢字調査としては、まず、層別による使用状況の概観を得るために、T種層別(話題による区分)を中心として、それにG種層別(文章の種類による区分)を組み合わせた十二の層別区分によって、処理を行なった。以下で、層別というのは次の十二の区分をさす。

1. 政治(国内)・労働

2. 外交

3. 経済（主として株式に関するもの）
4. 社会（社会面にのるもの）・地方（地方版）
5. 国際（主として外電によるもの）
6. 文化（学芸・教育・出版など）
7. スポーツ（囲碁・将棋を含む）
8. 婦人・家庭
9. 芸能・娯楽（ラジオ・テレビ番組。その解説）
10. 小説・漫画
11. 商業広告（商品や事業の広告）
12. 案内広告（いわゆる三行広告。通知・社告などを含む）

以上の区分をもとにして、ここでは、層別による漢字使用の比較を試みることにする。

表7は、層別区分ごとの、ことなり字数とのべ字数および総使用字数に対するのべ度数の割合を示したものである。どの区分において、漢字がよく使用されているかということは、ことなり字数やのべ字数の大小だけでは、一概に判定できない。紙面の面積に対する漢字の量や、かなや符号をも含めた文字（活字）の総数に対する漢字の割合なども、判定の要素として考える必要があるだろう。今、ここで問題にしようとするのは、語彙の量に対する漢字の量である。語彙調査における漢字ののべ度数は、その長単位語⁽⁸⁾ののべ度数とほぼひとしいということが経験的にいわれている。A調査の長単位語ののべ度数は、218,231である。これで、漢字のべ度数202,480を割ってみると、1語あたりに含まれる漢字数は、0.93字となり、さきの経験的法則とほぼ一致している。しかし、各区分ごとに、長単位語に対する漢字の割合を計算してみると、その比率は、表8に示すように、必ずしも一様ではない。区分11と12は、1語あたり、1字以上の漢字を含んでいることになる。これは、広告欄では、漢字の使用が比較的自由であることとあまり文章のスタイルをとらないことによるものであろう。特に、三行広告などでは、紙面の節約上、短い語形を漢字によって表わす傾向の強いためである。その他の区分を数値の高いものから順にながめてみると、いわゆる紙面のかたさといったものと似た順になる。区分7（スポ

表7 層別区分ごとのことなり字数とのべ字数

区分	話 題	ことなり 字数	のべ字数	総のべ字数に対す るのべ字数の割合
1	政治・労働	1,240	16,974	8.38%
2	外 交	622	2,776	1.37
3	経 済	1,172	19,727	9.74
4	社会・地方	1,582	26,671	13.17
5	国 際	1,171	14,030	6.93
6	文 化	1,358	11,560	5.71
7	ス ポ ー ツ	1,214	15,842	7.82
8	婦人・家庭	1,007	5,554	2.74
9	芸能・娯楽	1,414	13,458	6.65
10	小説・漫画	313	767	0.38
11	商業広告	1,779	26,768	13.22
12	案内広告	1,554	48,353	23.88
	全 体	2,423	202,480	100.00

表8

区分	漢字／長単位語
1	0.95
2	0.94
3	0.83
4	0.84
5	0.84
6	0.71
7	0.78
8	0.56
9	0.84
10	0.42
11	1.07
12	1.36
全	0.93

一ツ)や区分9(芸能)の数値がかなり高いのは、人名が多く現われることと表の占める割合の大きいことによると思われる。

つぎに、ことなり字数をみると、大部分が1,000~1,500字であるが、のべ字数は、区分によって、かなりの違いがみられる。しかも、のべ字数の多いものが必ずしもことなり字数が多いとは限らない。のべ字数に比して、ことなり字数が少いとみられるのは、区分3(経済)や区分12(案内広告)である。経済欄は、株式の相場表がその大部分を占めている。そのため、社名など、使用される漢字は、ほとんど一定しているので、のべ字数が多くなっても、ことなり字数は、それほど変化がみられない。案内広告欄も、求人・土地・建物などが主な内容で、同一の語(漢字)がくりかえし用いられるという点で、経済欄と同様のことが考えられよう。これに対して、同じ広告でも、一般の商業広告は、内容も多種にわたり、広告主によっては、かなり自由な用字が行なわれるため、案内広告と同一の傾向をとってはいない。

のべ字数に対してことなり字数が多いのは、区分6(文化)と区分9(芸

能)である。文化欄は、学問・芸術・教育・科学・図書など、内容が広範囲で、いろいろな分野の語彙があらわれるため、ことなり字数が多くなるものと思われる。また、芸能欄は、主となるテレビ・ラジオの番組欄の人名や題名に表外字が用いられるため、バラエティに富んだ内容を持っているものようである。

図2は、図1と同じように、各区分ごとの使用度数分布を曲線で表わしたものである。繁雑さを避けるため、省いたものもある。表9と合わせみると、おおよその傾向をよみとることができよう。(区分2と区分10は、他とくらべて、のべ度数が小さいので、ここでは、直接比較の対象とはしない。)

- ①使用度数の多い漢字800字では、すべての区分が使用率90%に達しており、ほぼ1,000字~1,500字で100%になるが、100字~300字あたりでは、区分によって、使用率にかなりの違いがみられる。
- ②スポーツ欄、経済欄、案内広告などは、100字で、50%~60%に達するが、商業広告や文化欄は、40%前後である。
- ③案内広告と婦人欄をのぞいた、他の区分は、使用度数の多い漢字の上位から40%~50%で、使用率90%に達している。別の言い方をすれば、ことなり字数の約半分で、全体の90%をまかなっていることになる。また、約55

図2 各層別区分の漢字数と使用率の関係

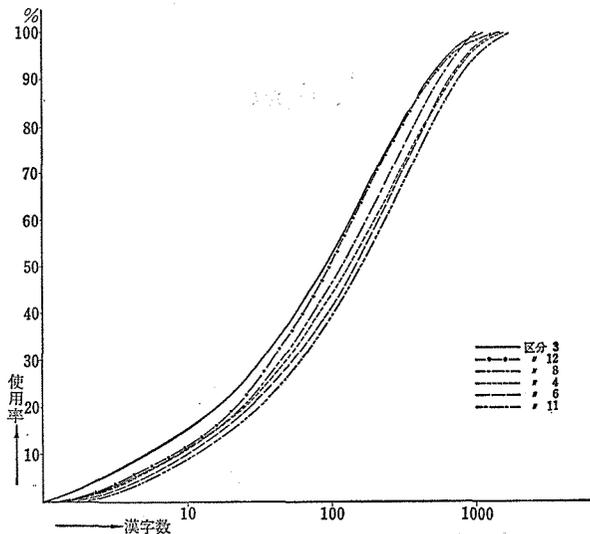


表 9 属別区分による使用度数分布

	1	2	3	4	5	6	7	8	9	10	11	12
最初の												
10字	11.9%	14.9%	15.1%	11.1%	13.2%	10.1%	16.8%	11.0%	10.8%	20.6%	9.0%	11.3%
20字	18.7	22.9	22.6	17.4	20.2	16.5	25.6	17.6	17.6	29.9	15.0	18.8
30字	24.3	29.0	28.5	22.4	25.7	21.3	31.9	23.2	23.2	37.0	19.7	25.1
40字	28.3	33.9	33.3	26.6	30.7	25.0	36.9	28.1	27.7	48.3	23.5	30.6
50字	32.9	38.2	37.5	30.4	35.0	28.5	41.2	32.0	31.4	43.5	26.7	35.4
60字	36.5	42.2	41.1	33.7	38.6	31.5	44.9	35.0	34.7	53.1	29.6	39.6
70字	39.6	45.7	44.3	36.6	41.9	34.3	48.2	38.7	37.6	57.0	29.6	43.3
80字	42.7	48.9	47.2	39.3	45.0	36.8	51.0	41.5	40.3	60.9	34.9	46.6
90字	45.4	51.8	49.9	41.7	47.7	39.2	53.5	44.0	42.8	64.0	37.3	49.6
100字	48.0	54.4	52.4	43.9	50.3	41.4	55.8	46.4	45.2	66.6	39.5	52.3
200字	66.1	73.2	70.0	59.8	68.5	58.6	71.0	63.4	62.0	85.3	55.6	69.8
500字	88.9	95.6	90.6	82.8	89.5	83.0	90.1	87.2	84.2		80.3	90.1
1000字	98.6		99.1	96.3	98.8	96.9	99.9	99.9	96.7		94.5	98.2
1500字				99.7							99.0	99.9

全体の	335字	256字	302字	445字	320字	445字	300字	374字	416字	159字	405字	302字
80%	414	307	376	551	400	542	381	458	519	197	607	380
85%	529	377	480	693	514	674	498	568	661	236	766	496
90%	711	482	660	910	696	890	691	729	885	274	1,034	695
95%	768	511	709	978	746	947	743	784	952	282	1,138	761
96%	836	538	777	1,064	817	1,011	822	840	1,020	289	1,207	849
97%	920	565	864	1,170	890	1,126	901	895	1,114	297	1,337	966
98%	971	594	973	1,312	930	1,242	911	951	1,279	305	1,511	966
99%	1,071	622	1,172	1,582	1,030	1,242	1,055	951	1,279	305	1,511	1,155
100%	1,240			1,582	1,171	1,358	1,214	1,007	1,414	313	1,779	1,554

%～65%で、使用率95%に達する。

④案内広告は、496字（ことなり字数の32.0%）で、全体の90%に達するが、100%になるのは、1,554字である。すなわち、比較的少ない字数で、全体をおおうことができるが、字種はかなり多い。

⑤婦人欄は、全体の90%に達するのに、568字（56.4%）を必要とするが、100%になるのは、1,007字である。このことは、婦人欄では、ある程度の使用率に達するためには、比較的多くの字種を必要とするが、全体としては、それほど多くを必要としないことを意味している。

以上、層別による漢字使用の特徴を大づかみに指摘したが、これによって、同じ新聞でも、話題別によって、かなりの相違があることがわかったと思う。なお、雑誌などの層別による特徴と比較してみると、また、興味ある結果が出るものと思われる。

(7) 林四郎「新聞語彙調査の概略と語彙分析法試案」（前掲国研報告31）

(8) 長単位語とは、文節を基準とし、1文節内の自立語の部分に1語とするものをいう。（注(7)論文を参照のこと。）

3. 層別区分による使用漢字の特徴

層によって、どのような漢字がよく使われるかということは、その層にどのような語彙が多く属するかということと不可分の問題である。その両者の関係を明らかにすることも、一つの大きな問題であるが、今、それを論ずる余裕はない。これから行なおうとするのは、語（または、語を構成する要素）を表記する文字としての漢字が、各層別区分にどのように特徴的に現われるかを、一応、語彙とは切りはなして、数量的な処理によって見ようとする試みである。結果として出てきたものについて、後に、語彙との関連を問題とすることになるが、それは、上に述べたことと矛盾はしない。なぜならば、ある漢字がどのような語に用いられたかということと、漢字と語彙との間に存する一般的な関係とは、別の次元に属することだからである。

一般に、基本語彙というようなものを統計的に用いようとする場合、一つの尺度として、深さ（使用度数）とか広さ（出現範囲）といったものが用いられ

ている。もし、新聞の基本漢字というものを求めようとするならば、やはり、そうした尺度を用いてみる必要があるだろう。しかし、すでに見てきたように、同じ新聞に用いられる漢字であっても、話題によって、記事の種類によって、あるいは、紙面上の位置によって、その用いられ方は一様でなく、それぞれ異った性格を持っているはずである。各区分の全般的な傾向は、すでに述べた。ここで、目的とするのは、各層別区分において、特徴的だと思われる漢字を具体的に抽出することである。語彙については、そうした試みが、すでにいくつか行なわれている⁽⁹⁾。それらを参考にしつつ、漢字について、そうした方法を試みることにする。

ある特定の層別区分における漢字（または語彙）の特徴度を数量的に測定する場合、その基準となるのは、およそ次のようなものであろう。

- ①使用度数および順位
- ②実度数（使用度数）と理論度数の差
- ③ " " 比
- ④その層別区分への集中度

ここでいう理論度数というのは、配分係数（総のべ度数に対する各層別区分ののべ度数の比率。表8に%で示してあるのがそれにあたる。）によって、それぞれの漢字の使用度数を比例配分して算出したものである。たとえば、使用度数1位の「日」の実度数と理論度は、表10のようになる。また、集中度というのは、基本語彙でいう、ちらばり度（出現の範囲の広さ）の反対の概念である。

もし、①に重点をおいて考えるならば、表11のように、各区分ごとに使用度数の多い順に漢字をならべた表を作ってみればよい。しかしながら、それぞれをたてにながめると、たしかに、その区分で特徴的に用いられているらしい漢字を認めることができるが、横にながめてみると、使用度数の多い字ほど、ど

表 10

区 分	1	2	3	4	5	6	7	8	9	10	11	12	計
実 度 数	258	66	565	370	275	132	275	59	154	6	372	674	3,206
理 論 度 数	269	44	312	422	222	183	251	88	213	12	424	766	3,206

表 11 各層別区分でよく使われる漢字 (数字は使用回数)

順位	区分	1	2	3	4	5	6	7	8	9	10	11	12
1	1	328	68	565	440	275	192	382	91	210	43	372	676
2	2	258	66	369	395	272	132	368	88	185	24	278	674
3	3	236	55	367	370	205	131	347	71	154	17	272	647
4	4	225	40	331	305	199	119	275	66	145	15	263	580
5	5	222	34	286	271	170	103	252	59	144	12	242	550
6	6	161	31	270	270	158	119	228	59	138	11	240	550
7	7	158	31	235	238	155	99	219	53	133	10	194	528
8	8	147	30	188	235	140	92	217	43	121	10	193	433
9	9	142	30	183	230	138	89	186	41	110	8	181	431
10	10	138	28	179	209	130	89	181	40	107	8	178	405
11	11	132	27	175	197	123	83	177	38	107	8	176	389
12	12	129	24	164	184	118	82	161	38	104	8	176	386
13	13	119	24	160	175	112	80	152	38	100	8	175	376
14	14	117	24	159	173	110	78	151	38	94	7	162	371
15	15	116	24	152	168	95	78	133	38	89	7	161	369
16	16	112	21	143	166	92	74	132	37	88	7	159	353
17	17	111	21	140	165	89	67	129	35	87	7	157	349
18	18	110	20	135	153	87	66	128	34	86	7	146	348
19	19	108	19	134	151	85	66	120	34	85	6	143	346
20	20	105	18	131	150	83	65	114	34	84	6	143	337

の区分でも上位を占める傾向があり、どの層別区分において、より特徴的かということ判断するのは困難である。②または③に重点を置くのは、①にくらべて、より求めるものに近づくことができそうな方法である。先にあげた、層別特徴語彙の試みも、この二つのどちらかに重点を置いて処理している。これから行なおうとする方法も、③に重点を置いた方法である。理想として考えられるのは、上にあげた四つの基準（まだほかにも加えるべきものがあるかもしれない）をなんらかの方法で総合して、統計学的にも意味のある処理を行なうことである。しかしながら、今は、それに足るだけの準備ができていない。そこで、将来、よりよい方法を作り上げるための捨石として、以下に述べるような一つの試みを行なってみることにした。

①対象は、各層別区分ごとに現われたすべての漢字とする。ただし、区分2と区分10は、全体に対する比率が小さいので、対象から除く。

②各漢字ごとに、理論度数を算出する。算出された理論度数が1に満たないものは、対象から除く。

③実度数から理論度数を引き、その差を理論度数で割ったものを、その漢字の理論度数からの逸脱度とする。ただし、理論度数が3以下で、実度数が、10以下のものは、対象から除く。

逸脱度 = $\frac{\text{実度数} - \text{理論度数}}{\text{理論度数}}$ (例) 実度数 = 565 理論度数 = 312 逸脱度 = $\frac{565 - 312}{312} = 0.81$

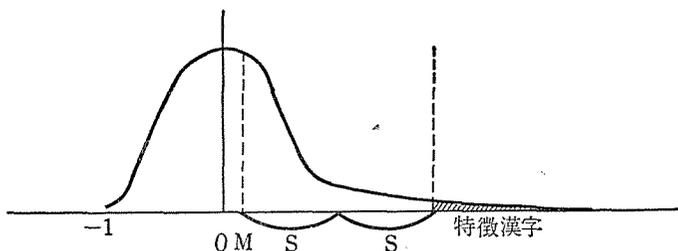
④層別区分ごとに、逸脱度の大きい順に漢字をならべて、分布表を作り、平均値Mと標準偏差Sをもとめる。

⑤層別区分ごとに、平均値から標準偏差の2倍以上隔っている漢字を特徴漢字とする。

特徴漢字 $> M + 2S$

以上のような手つづきを経て、抽出された特徴漢字は、表12に示すおとりであるが、それについて分析を加える前に、この方法による問題点をいくつかあげてみることにする。

①この方法の特色は、特定の層にかたよることなく、それぞれの層別区分内における特徴漢字をひろえることと、使用度数の低い漢字でも、ある層別区分における使用率が他の区分に比べて著しく高い場合には、特徴漢字に



なりうることにある。ただし、使用度数が低いものは概して理論度数も小さく、逸脱度が大きくなる傾向を持つので、有意な範囲についての検定を行なう必要がある。

②実度数が0でも（その層別区分では1字も出現しなかった字でも）、理論度数が1以上の値を持つ場合がある。ここでは、そういう字は対象としなかったが、考え方によっては、消極的な特徴を持った漢字に含めてもよいと思われる。しかしながら、上述の方法では、たとえ対象にするとしても、理論度数の大小にかかわらず、それらの漢字の逸脱度は、いずれもマイナス1となり、処理上に問題がある。

③一般に、実度数が理論度数よりも小さい場合の逸脱度はマイナスになる。そして、その範囲は、 $-1 < x < 0$ である。ところで、層別特徴漢字の範囲を平均値から標準偏差の2倍以上隔っているものと定義した場合、 $M + 2S$ より大きいものとともに、 $M - 2S$ より小さいものも、消極的な意味での特徴漢字になりうるはずである。しかしながら、実際に処理してみると、 $M - 2S$ は、みなマイナス1よりも小さくなってしまい、該当するのは、あらわれない。したがって、この方法は積極的な意味での特徴漢字をとりあげるには適しているが、消極的な意味での特徴漢字をひろう方法としては、あまりよいものとは言えない。

以上のような条件を頭におきながら、表12をながめてみると、次のようなことが考えられる。

- ①特徴漢字の多少は、その区分の漢字数とは直接関係がなさそうである。むしろ、他の区分と異った内容を持つものほど、特徴漢字が多い。
- ②二つ以上の区分で特徴漢字となったものは、区分1と区分5の府と官、区

分3と区分11の綿と酒ぐらいである。ある区分における特徴度が高ければ高いほど他の区分では、特徴漢字となりにくい、この方法の性格によるものと思われる。

③話題による層別のため、その層別区分の話題と関連性の強い特徴語との結びつきが強い。

①社会面では、調査対象となった時期に、羽田空港沖の航空機墜落事故があったことを反映して、空・故・航・羽などの字がみられる。

②経済欄は、会社名および株式用語に関するもの（無・買・減・伸・据）が多い。

③芸能欄は、番組表独特の用字がみられる。特に人名に用いられる字（郎・夫・雄・彦・之）が多いのがめだつ。

④文化欄には、天気予報が含まれているので、気象関係の語に用いられる漢字が多い。

⑤案内広告欄は、土地建物に関するもの（迄・坪・歩・駅・完・邸）と求人に関するもの（万・給・歴・面・遇・卒）が多い。固有名詞以外の表外字、迄・乞なども、この欄独特の用字である。

④層別特徴漢字は、特徴語と強い結びつきを持つが、その逆に、特徴語に用いられる漢字が特徴字になるとは限らない。例えば、大鵬・巨人などは、一方が特徴字だが、一方は特徴度は、きわめて弱い。

以上、新聞の漢字使用について、一応整理のついたデータをもとに、その特色をさぐってきたが、はじめに述べたように、不十分な資料から推測し得たものは、おそらく九牛の一毛にもおよばないのではないと思われる。さらに、この調査が量的にも質的にも充実した時に、いっそう深い分析が行なわれることを期することにした。

- (9) 林四郎 前掲(7)論文
木村繁「層別特徴語の判別」(本書25頁)