

国立国語研究所学術情報リポジトリ

総合雑誌の語彙調査：中間報告

メタデータ	言語: jpn 出版者: 公開日: 2019-12-05 キーワード (Ja): キーワード (En): 作成者: メールアドレス: 所属:
URL	https://doi.org/10.15084/00002380

総合雑誌の語彙調査
中間報告

昭和30年4月

林

国立国語研究所

書きことば研究室

総合雑誌の語彙調査

中間報告

目次ならびに概要

- 0 総説 2
- 。この調査全体の範囲と規模
 - 。ここで中間報告をするのは延べ六万語の集計の結果
- 1 使用度数の多い199語 3
- 。標本度数40以上の語が199語
 - 。199語全体の使用率
 - 。語彙表
- 2 かたかなを用いた語 7
- 。一部または全部をかたかなで書いた例のある語が989語ある。
 - 。その種類および使用率
 - 。1/4ページごとに見られる割合
- 3 表外字を用いた語 13
- 。当用漢字表外の漢字を用いた例のある語が865語ある。
 - 。用いられた表外字の種類は616字。
- 4 語彙の総量の推定 24
- 。総合雑誌1年分の本文(延べ九百万語)は、約二万の異なった語で書かれていると推定される。
 - 。その推定法について。

0. 総説

0・0 この調査の範囲及び規模

雑誌の種類： 総合雑誌およびそれと近い内容を持つ雑誌13種
改造・解放・世界・世潮・中央公論・文芸春秋・
学園評論・国民・心・人生手帖・日本及び日本人・
ニューエイジ・平和

範囲： 昭和28年7月号～昭和29年6月号の本誌および付録
雑誌の期間中に出た増刊号

標本： 全体約二万三千ページから1120ページを無作為に抽
出し、更にそれぞれの $\frac{1}{2}$ ページ分を無作為に選んだ。

母集団の大きさ： 900万語

標本の大きさ： 24万語

なお、この調査は、「当用漢字の適用によって生ずる問題とその解決法の研究」の題で、文部省試験研究費の交付を受けて行った作業をふくむ。

0・1 ここで中間報告をするのは、
改造・解放・世界・世潮・中央公論から採集した延べ十二万語の標本につい
て、もう一度無作為に選んだ六万語の集計結果である。

この延べ六万語は、異なり一万語から成る。

1 使用度数の多い199語

1・0 延べ六万の標本について整理集計した結果、そこには約一万の異なつた語が用いられていることがわかつた。そのうち一語で40回以上くりかえし用いられているものが199語ある。

この199語は、標本全体の

異なり語数で 2% 延べ度数で 42%

を占める。

1・1 語彙表

。見出しのかたかなの書き方は『婦人雑誌の用語』の例による。ただし、カリルは一段活用、カリルは五段活用の動詞を示す。

。漢字は意味の注記として加える。必ずしも実際に用いられたものではない。

。見出しの左の数字は度数順位、右の数字は標本度数である。

順位	ことば	標本 度数				
			15	イチ	1	299
			16	コレ	此	293
1	シ・スル	1394	17	ナイ	無	278
2	イル	1184	18	ワタクシ	私	256
3	イイ・ウ	1033	19	サレル		253
4	コト	905	20	二	2	237
5	ナリ・ル	613				
6	ソの	557				
7	テキ	499	21	キ・クル	来	233
8	アリ・ル	479	22	シャ	者	218
9	モノ	464	23	ネン	年	218
10	コの	441	24	タメ	為	205
			25	アメリカ	米	197
11	ヨウ	410	26	ミル	見	196
12	ソレ	406	27	ヨリ・ル	由	195
13	ニッポン	334	28	デキル	出来	192
14	ノ	(準体) 306	29	一〇	十	180

30	三	3	176	59	五	5	107
				60	セイジ	政治	106
31	ユキ・ク	行	172	61	イマ	今	105
32	ダイ	第	170	62	二〇	二十	105
33	モンダイ	問題	169		エル	得	103
34	オモイ・ウ	思	161	63	コクミン	国民	103
35	マデ	迄	158		センソウ	戦争	103
36	シュギ	主義	151	67	トコロ	処	103
37	ナニ	何	151		六	6	103
38	ヨイ	好	149	68	クミアイ	組合	102
39	シカシ	(接)	145	69	ドウ	如何	102
40	トキ	時	143	70	ナカ	中	100
41	タイシ・スル	対	137	71	ロウドウ	労働	95
42	ソウ	然	136	72	コウ	斯	94
43	オ	御	134	73	シホン	資本	94
44	マタ	又	134	74	ハ	8	91
45	ケイザイ	経済	133	75	ヒトツ	1	90
46	モチ・ツ	持	133	76	ヤリ・ル	遣・試	89
47	オキ・ク	於	131	77	トリ・ル	取	87
	タチ	達	131	78	五〇	五十	85
49	トウ	党	131	79	シャカイ	社会	85
50	ダケ	丈	124	80	ジブン	自分	84
51	カレ	彼	121	81	カ	化	83
	ツキ・ク	就いて	121	82	一〇〇〇	千	83
53	ラ	等	121	83	クニ	国・地	81
54	ガツ	月	119	84	ニチ	日	81
55	セイフ	政府	118	85	オオキい <small>な</small>	大	80
56	ヒト	人	114	86	ウンドウ	運動	78
57	カンガエル	考	112		タイ	度	78
58	ナド	等	111				

88	ホウ	方	78				
89	コク	国	76				
90	セカ界	世界	76				
91	キョウサン	共産	75				
92	サン	様	75				
93	グン	軍	74				
94	デル	出	74				
95	七	ワ	73				
96	ジユウ	自由	72				
	ジン	人	72				
98	バアイ	場合	72				
99	カイケツ	解決	70				
100	九〇〇	九百	70				
101	カンケイ	関係	69				
	ソレン	ソ連	69				
105	テン	点	69				
	ホド	程	69				
	ワレワレ	我々	69				
106	ヘイワ	平和	68				
107	ワケ	訳	67				
108	カタ	方	66				
109	万	万	66				
110	ジケン	事件	65				
111	ダ	(接 ^{だが} _{から})	64				
112	マエ	前	63				
113	シリル	知	62				
114	ソシテ	(接)	60				
115	メ	目・眼	60				
116	カイ	会	58				
	セイ	性				58	
119	セイサク	政策				58	
	リョク	力				58	
120	オナジ	同				57	
	セイカツ	生活				57	
121	オヨビ	及				56	
	カ	家				56	
122	グンジ	軍事				56	
	コクサイ	国際				56	
129	ダイ	大				56	
	バカリ	許				56	
	ハナシ	話				56	
	ヒツヨウ	必要				56	
130	シマイ	了				55	
	ガクセイ	学生				54	
131	キョウテイ	協定				54	
132	ウエ	上				53	
	ウチ	内				53	
133	オリル	居				53	
138	ケイカク	計画				53	
	ココ	此処				53	
	フ	不				53	
	エン	円				52	
139	シレル	知				52	
	ヒトビト	人々				52	
142	四〇	四十				52	
	イミ	意味				51	
143	タ	他				51	

146	ダシス	出	51	億	億	43
	一〇〇	百	51	サイ	再	43
	アイダ	間	50	173 サラ	更	43
	ガワ	側	50	181 テ	手	43
	九	9	50	ノウミン	農民	43
	ケンキュウ	研究	50	ヒトリ	一人	43
147				ボク	僕	43
	ゴ	後	50	ワが	我	43
154	コトバ	言葉	50	アル(い)は	或	42
	キンゲン	人間	50	オオイ	多	42
	マカリル	分	50	182 ツクリル	作	42
	イジヨウ	以上	49	187 ドイツ	独逸	42
155	サセル	させる	49	ミンゾク	民族	42
157	ソコ	其処	49	ヨオロッパ	欧	42
	アル	或	48	ウケル	受	41
158	ニン	人	48	オキク	置	41
160	ヨ(シ)	4	48	カノジョ	彼女	41
			47			
161	シ	氏	47	188		
162	ジョウ	上	47	193 コロ	頃	41
	エンジョ	援助	46	三〇	三十	41
	ゲンザイ	現在	46	タダ	唯	41
163	四	4	46	クレル	與	40
168	セイサン	生産	46	コンニチ	今日	40
	ホウ	法	46	シハイ	支配	40
	モウ	(副)	46	194 ジンミン	人民	40
	ジジツ	事実	45	199 ツギ	次	40
169	ボウエキ	貿易	45	マダ	未	40
171			44			
	ミエル	見	44			
172	アイウ	会	44			
	インド	印度	43			

カタカナの語は
定めて
置いて。

2. かたかなを用いた語

2.0 まえがき

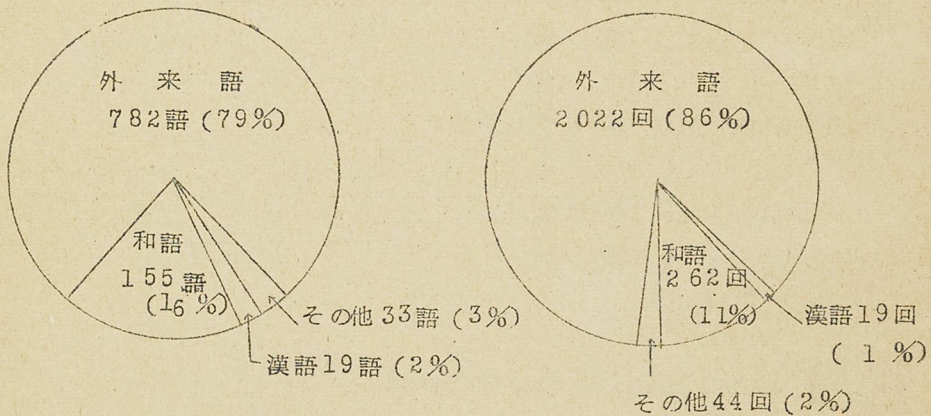
一部または全部をかたかなで書いた語は、異なり一万語の中に約一千語(989語-99%)ある。

その一千語が繰返された回数(使用度数)は2347回で、これは総使用度数六万に対して約39%にあたる。この一千語がどのような種類の語によって占められるかを見るために、まず、次の四つに大別してその分布の状態を調べた。

1. 外来語の表記に用いられたもの
2. 和語の表記に用いられたもの
3. 漢語の表記に用いられたもの
4. 記号、音表記、ルビ等に用いられたもの

(異なり989語の分布)

(使用度数2347回の分布)



以上の分布をみれば、異なり語、使用度数の双方ともかたかな書きは外来語に用いられたものが圧倒的に多いことがわかる。

2.1 外来語の表記に用いられたかたかな

種類	異なり語数	使用度数	例 語
地名	164	889	アジア, イスラエル, コロンビヤ, ソヴェト, ビキニ etc.
人名	222	387	アチソン, オッペンハイマー, シェクスピア, ネルソン etc.
その他	396	746	アジテーター, アパート, インフレ, エネルギー, コスト etc.
	782	2022	

(8)

(地名) ソヴェト の略語形たる<ソ>を含んでいる。(使用度数3回)
 なお、<ソ連, 日ソ, 米ソ>(異なり3語, 使用度数75回)や<対ソ, 駐ソ, 入ソ, 反ソ>(異なり4語, 使用度数5回)も便宜上この中で数えてある。
 ソヴェトの略語としての<ソ>がさかんに使われるのは、<蘇>が制限字であることに原因している。

(その他) ここには、地名、人名以外の外来語、外国語をおさめたが、この中には、単独で、もしくは他の語と複合して、固有名詞として用いられたもの48語(延べ69語)を含んでいる。

(例えば、タイム(雑誌名) 朝日ビール(商品名))

(なお、<ガス弾, 急カーブ>などの混種語を18語(延べ25回)含んでいる。

ここで注意したいのは、上の782語のうち14語については、かたかなを用いない表記が同時にとられていることである。それぞれの語ごとに、その表記のゆれ方を見ると、以下のようにになっている。

	表記	度数	表記	度数	表記	度数	表記	度数		
地	イギリス	31	英吉利	2	人名	キリスト	1	基督	1	
	インド	36	印 度	7		エム・एस・エー	1	M S A	34	
	ギリシャ	3	希 臘	1		そ	ドル	28	弗	2
	シナ	1	支 那	4		の	トン	4	噸, 廳	8
	ドイツ	40	独 逸	2		他	パーセント	3	%	28
	パリ	4	巴 里	5		メートル	3	米, m	4	
	フランス	32	仏 蘭 西	1		小 計	41		76	
	ベルシヤ	1	波 斯	1						
	小 計	148		23						

2.2 和語の表記に用いられたかたかな

種類	異なり語数	使用度数	例 語
地名	5	5	イタミ, 市ガ谷, 千駄ヶ谷, 丸ノ内 etc.
人名	12	31	カネ, ギン, シメノ, ヒデオ, テイ子 etc.
その他	138	226	(例は以下に示す)
	155	262	

(地名) 日本の地名を、全部かたかな書きにしたのは、<イタミ飛行場>

だけで、これは非常に特殊の場合であり、多くは、〈市ガ谷、丸ノ内〉のように助詞の部分をかたかなにしたものである。

(人名) ほとんどが女の名前(小説中の人物が多い)である。男の名前は〈ヒデオ〉(混血児の名前)〈タケル〉〈ヤマトタケル〉(共に日本武尊のこと)の3例だが、いずれも特殊の例と言えよう。

(その他) ここにはいろんな性質の語が含まれるので、これをまず、漢字では書けぬ語、すなわち漢字をあてる習慣のないと認められる語と、そうでないもの、すなわち漢字をあてようと思えばあてられるものなどに大別した。(その区別は「明解国語辞典 改定版」の見出しに漢字が示されているかいないかで見分けた。)

漢字では書けぬ語	異なり語数 (41)	使用回数 (54)	
1) 擬声・擬態語	24	34	アラ、ガーッと、ガツガツ、グズグズ、ニヤニヤ etc.
2) 活用語	7	7	ズレ、マゴツク、コジツケ;ゴタつく、トギツい、ワメキちらす etc.
3) 無活用語	9	12	インチキ、ケチ、チャテ、ニコヨン;ピカー、ガリ版 etc.
4) 接辞	1	1	オッタまげる

漢字でも書ける語	(97)	(172)	
1) 動植物名	24	40	イモ、イワシ、オットセイ、ネコ、ヘビ、ムギ、etc.
2) 活用語	15	18	コワイ、ツカマル、クツログ;ニラミ合せる、ネコ伏せる、ボケる etc.
3) 無活用語	53	82	アカ、アラ、ゼニ、デタラメ、バケモノ;ベラ棒 etc.
4) 接辞	5	32	カ年、カ月、オ、サマ、サン
合計	138	226	

(漢字では書けぬ語)

1) 擬声語・擬態語——これのかたかな書きが24語ある中で、表記のゆれて
いるものは、次の4語である。

表記	度数	表記	度数
ウロウロ	1	うろうろ	1
ゾツと	1	ぞつと	1
ドンドン	1	どんどん	3
ハッキリ	3	はつきり	27
	6		32

(10)

<ドンドン>は物事の進行する状態を示す擬態語の方。

なお、²⁾³⁾には表記のゆれた例は見られない。

(漢字でも書ける語。)

1) 動植物名で表記のゆれているものは次の二語である。

表 記	度数	表 記	度数
サル	5	猿	7
タケ	2	竹	2
	7		9

2) 活用語で、表記のゆれのあったものは次の6語である。

表 記	度数	表 記	度数
コワイ	2	恐い, こはい	2
ツカマエル	1	つかまえる	1
ヌキ	1	抜き	2
キツイ	1	きつい	1
ニラミつける	1	睨みつける	1
ねえ(無)	1	無い, ない等	277
	7		284

3) 無活用語で表記のゆれているものは次の25語である。

表 記	度数	表 記	度数	表 記	度数	表 記	度数
アゴ	1	顎	1	セリフ	1	せりふ	1
イヤ	3	嫌, いや	13	タナ	1	棚	5
オバ	3	伯母, おば	2	ダメ	4	駄目, だめ	6
オラ	2	俺, おら etc.	8	チリ	1	塵	1
カナメ	1	かなめ	1	テンデ	1	てんで	1
カバン	1	鞆	2	バカ	2	馬鹿, 莫迦, ばか	7
キツカケ	1	きつかけ	2	ビックリ	1	吃驚, びっくり	2
キレイ	2	綺麗, きれい	3	ヒモ	2	紐	1
クビ	2	首	3	ボク	1	僕, ぼく	42
コトバ	5	言葉, ことば	45	ホンの	1	ほんの	6

表 記	度数	表 記	度数	表 記	度数	表 記	度数
マネ	1	まね	2	ワク	4	粹	4
ミコト	1	尊	1	まア(副詞)	1	まあ	4
ヤミ	4	闇, 暗	3		47		166

4) 接辞 <カ年>の一例を除き, 他はすべて表記がゆれている。

表 記	度数	表 記	度数
オ	9	御, お	125
カ月	11	箇月	1
サマ	2	様, さま	14
サン	1	さん	74
	23		214

2・3 漢語の表記に用いられたかたかな

種 類	異なり 語 数	使用 度 数	例 語
地名	1	1	ニッポン
その他	19	19	イジュク, カン(勘), コツケイ, ドンラン; 悪タイ, カツ然
	20	20	

ここで表記のゆれているものは次の通り

表 記	度数	表 記	度数
ニッポン	1	日本	333
アイマイ	1	あいまい	2
カイライ	1	傀儡	1
コタツ	1	炬燵	2
ミン	1	味噌	2
ムリ	1	無理	10
リクツ	1	理窟	2
	7		352

2・4 記号・音表記・ルビ等に用いられたかたかな

種類	異なり 語数	使用 度数	例 語
記号	3	9	イ, ロ, ハ,
音表記	19	23	赤ン坊, かッど, 引ッころがす, ただ今ア etc.
ルビ	11	12	木嶋 ^{キイ} さん, 木舞 ^{コマイ} , 手巾 ^{ハンカチ} , 横揺 ^{ローリング} れ,
	33	44	

(音表記)というのは、撥音、促音、長音等を示すために用いたもの、また、特別に語の読み方を示すために用いたものなどを含む。

(ルビ)も、語の読み方を示すものであるが、この中には、「^ウ共同」「^ア堅揺」^{アカウント}のような、一種の二重表記と目されるものが6語含まれている。^{ピッチング}

2・5 1/4ページごとにかたかな語の現われる割合

雑誌の1/4ページの面積を単位の広さとして、そこに含まれるすべての語の延べ語数と、かたかなを用いた語の延べ語数との割合を調べた結果は次の通りである。(表中16 4とあるのは、100語に16語の割合でかたかな語を含む場所(1/4ページの広さ)が四箇所あったことを示す。)

0	123	7	23	14	5	24	1
1	71	8	22	15	1		
2	76	9	18	16	4	51	1
3	55	10	15	17	2		5・58
4	56	11	10				
5	32	12	11	20	1		
6	22	13	9				

- この1/4ページと呼んだのは、この語彙調査の採集カード一枚にリプリントされるだけの大きさ(つまり雑誌の一ページの1/4の大きさ)である。調査した1/4ページの総数は558箇所である。
- かたかなで表記された語のきわめて多い場所は、外国の政治・経済事情の解説・論文で外国の地名人名がしばしば出るもの、または外国の地名などを表の中で列挙したものが、ほとんどを占めている。
- かたかなを一度も用いなかった1/4ページは数は多いが、目立った特徴は少ない。座談会・小説などにやや多く存在する傾向が認められる程度の事しか、まだ分っていない。

3 表外字を用いた漢

3・0 延べ六万語，異なり一万語のうち，当用漢字表外の漢字を用いた例のある語は，

異なりで 865語 (全体六万語の 8.7%)

延べで 5153語 (全体六万語の 8.6%)

そのうち，表外字を用いた例は，

延べで 1569語 (5153語の30.5%)

全体六万語の 2.6%

また，そこに用いられた表外字は，

異なりで 613字

延べで 1647字

(この表外字の延べ字数が漢字全体の延べ字数に対する割合は，まだ計算されていないが，おそらく3%前後と予想される。昭和28年度の総合雑誌の調査では3.17%であった。)

3・1 表外字の表

用いられた表外字613は，次の表の通りである。(配列はほぼ康熙字典の順による。数字は部首の画数)

- | | | | | | | |
|--|------------------|----------------|-----------------|-----------------|-----------------|-----------------|
| | 云 | 僕 | 僂 | 此 | 或 | 於 |
| | $\frac{70}{108}$ | $\frac{5}{26}$ | $\frac{17}{36}$ | $\frac{5}{787}$ | $\frac{23}{90}$ | $\frac{8}{131}$ |
- 1 乃之乎 乞
 - 2 亘 亨 仆仔仙伊伎侈俟俯棒俺倚倣偵僂傘
備傲僅僕僂儲儼 冑 冥 凌凜 凭凱 函
剗劓劉 勃 勿 卿 厄厠 叛
 - 3 叩只吃吊吞吻吾呂呆呵咳咳咽哨哩唇唵啄啼喉
喋喘喧喰鳴嘗嘘唾嘲噲噲噎噎嚙囊 囧 坐垣埃
埒堀堆界塚墳塵墟塾壻 壺 夙夥 夷奈套奠
奢 妄妾姦娼嫌嬉嬌 宋宏宥寐寔 尖尙 尤
尨 尻屢 屯 岡崎崖嵌嵐嵩巖 巴巷 巾
庄庇庖廚廠 廻 廿 弗弼彌 彥
 - 4 忽怨恰恣悄悉惚惚惧惹愕愴愴 滙慄慧慾慾憑憚
儻 戊戌或戚戟戮戴 戾扉 托把拋披抹拭拳

按挑挫挽挾捉捏抄据捲捻捷捺掠掩揄揆揅揅
 搔摺撒撥撰擡攪 斬斯 於 且昂昆晋晦智曝
 曰曹曾 朋 朴杉李杜杞枉枕砗柑柏柢柵梅
 柴栓栖栗桎桶桿梁梏梨棍棉棚棟棧倚楓楯楷檟
 榭榭檣櫓 此 歿殆殞 設毅 毫 汜汁汎汰
 汲沁沫沙泄泡泥瀉洩洪洲浚浩涎涯涵淀淋淳淵
 渙渦漑洩溜溝溢瀉滓滔漱潑渴潰濤濡瀕灘 灯
 灼炬焚煉熙煽熄熊燧 爺 俎爽 牌牒 牙
 牢牽犀 狗狙狡猪猫猿獅

5 茲 琉瑞瑣 瓦甌瓶 甚 甥 畠崎畿疇 疏
 疵痕瘠癩 皿孟盃盞 眉眺睦睥睨瞭瞰 砌
 砥硯碍碼磯 祀崇祠祿禎禰 秦稀稼稽 穿窟
 窠窺竈 颯

6 笠筍筭筮筴篇簞簫籥籠 秆糞 紐紗紘絃綜綻
 絢綺繩繫辮纏 罐 懼羅 翳 而 聯聳 肘
 肚肢脇腎膝膿臀臂臆臘 白 絃艘 芥芦芹
 苋苑苜茱芒莉莫菱荀萎葡葵葦菟蒔蒙蒲蒼蓋蔭
 蓮蔭蔽蕩薩藁藤藪蘇蘭 虎 蛙蛛蛤蜂蜘蛛
 螢蟬蟹蟻蠅 袍袖袴裡裔襖襟 霸羈

7 視訊註詮誰誹諒諜諤諦諱謂謗謬謳譏讐讖讚
 貌 貪貫貼賭贅 趨 趾跌蹉蹠蹙躁躅躅 軀
 軋輔輯輻轟轆 辣 辰 迥迄迦這逝逞逢遙遜
 遮邁 那郁 醬醞醪 采

8 鈞鈞鉞鍋鍬鍵鐘鑽鑿 閔閻闖 阪阿隅隙隴
 雀雖 雱 靡

9 靴鞞鞞 韓 頁頃頑頤頰頻顎頰顛 餃餅餌餐
 饗 10 馳馴駁馱駟駒 骸 髡髻鬚 鬩

11 鮭鯖鯛鯨鯉鱈鱗 鳩鴨鵝鵬鶴鷗鷹 鹵 鹿
 麵 13 鼠 14 齊 17 龍

3・2 表外字を用いた語の表

使用度数が5以上に及ぶ表外字74種について、その字の用いられた語を示すと、次の表の通りである。

○この表の漢字の配列は、字の度数順とし、度数の等しいものの間では康熙字典の順によった。

○同じ字を用いた語の配列は、その字の音を用いたものを先に、訓を用いたものをあととし、さらに、一字めに用いたもの、二字めに用いたもの……に分け、それぞれ五十音順とした。

○固有名詞には*印をつけ、さらに、人名には(人)、地名には(地)の注記をした。

○その語が、その表外字を用いないでも書かれている時は、「その語の他の表記の度数」の欄に、その種類と度数を示した。その表外字にあたる部分が他の漢字である時はその漢字を示し、ひらがな書きはひ、かたかな書きはカとした。

番号	漢字	度数	音 訓	その字を用いた語	(A)その 度 数	(B) その語の 他の表記の度数	(A) + (B) その語の総度数
1	云	70	ウン いう	云々	1	—	1
				云う	65	言 72. ひ942	1079
				云いかえる	2	言 2. ひ 2	6
				云い放つ	1	—	1
				云い古す	1	—	1
2	頃	44	ころ	頃	33	ひ 8	41
				頃合い	1	—	1
				今頃	1	—	1
				この頃	3	ひ 3	6
				近頃	2	ひ 1	3
				手頃	1	—	1
				年頃	1	—	1
				中頃	1	—	1
				日頃	1	—	1
3	藤	37	トウ	*藤村(人)	2	—	2

番号	漢字	度数	音訓	その字を用いた語	(A) その度数	(B) その語の 他の表記の度数	(A) + (B) その語の総度数
				* 阿 藤 (人)	1	—	1
				* 安 藤 (人)	8	—	8
				* 伊 藤 (人)	10	—	10
				* 加 藤 (人)	2	—	2
				* 斎 藤 (人)	2	—	2
				* 佐 藤 (人)	5	—	5
			ふ じ	* 藤 崎 (人)	1	—	1
				* 藤 田 (人)	1	—	1
				* 藤 原 (人)	5	—	5
4	誰	29	だ れ	誰	29	ひ 7	36
5	或	23	あ る	或	14	ひ 34	48
			あるいは	或いは (人)	9	ひ 33	42
6	伊	21	イ	* 伊 井 (人)	1	—	1
				* 伊 作 (人)	5	—	5
				* 伊 勢 (地)	1	—	1
				* 伊 藤 (人)	10	—	10
				* 伊 吹 (人)	3	—	3
			だ て	伊 達	1	—	1
7	嫌	21	ケ ン	嫌 悪	1	—	1
				嫌 忌	1	—	1
				嫌 疑	1	—	1
				機 嫌	4	ひ 1	5
			い や	嫌	12	ひ2 カ2	16
			きらい	嫌 い	1	—	1
			きらう	嫌 う	1	—	1
8	僕	20	ボ ク	僕	20	ひ22 カ1	43
9	喰	16	く う	喰 う	2	食7 喫1	11
				喰い込む	2	ひ1	2
				喰違い	1	食 1	2

番号	漢字	度数	音訓	その字を用いた語	(A) その度数	(B) その語の 他の表記の度数	(A) + (B) その語の総度数
10	杉	16	くえる たべる すぎ	喰違う	1	—	1
				喰いとめる	1	—	1
				喰える	2	食 3	5
				喰べる	7	食 2	9
				* 杉 (人) 姓	12	—	12
				* 杉 (人) 名	3	—	3
11	洲	16	シュウ ス	* 杉 作 (人)	1	—	1
				洲	1	—	1
				* 欧 洲 (地)	3	州 1	4
				* 満 洲 "	9	州 1	10
				* 洲 崎 (人)	1	—	1
12	乃	14	ナ イ ノ	* 八重洲 (地)	2	—	2
				乃至	11	ひ 2	13
				* 乃 木 (人)	1	—	1
				* 志 乃 (人)	1	—	1
				* 花乃家	1	—	1
13	岡	14	お か	* 岡 倉 (人)	2	—	2
				* 岡 崎 (人)	3	—	3
				* 岡 田 (人)	1	—	1
				* 岡 林 (人)	1	—	1
				* 市 岡 (人)	1	—	1
				* 菊 岡 (人)	2	—	2
				* 北 岡 (人)	1	—	1
				* 静 岡 (地)	1	—	1
				* 福 岡 (地)	1	—	1
				* 安 岡 (人)	1	—	1
14	頁	14	ぺえじ	頁	14	—	14
15	勿	12	モ チ モ ツ	勿 論	11	ひ 22	33
				勿 体	1	—	1

番号	漢字	度数	音訓	その字を用いた語	(A) その度数	(B) 他の表記の度数	(A) + (B) その語の総度数
16	廻	12	まわす まわり まわる	眺め廻す	1	—	1
				廻り	2	回 1 ひ 2	5
				空廻り	2	—	2
				廻る	3	回 1 ひ 3	7
				歩き廻る	2	—	2
				上廻る	1	—	1
				出廻る	1	—	1
17	彌	12	や	* 彌市 (人)	1	—	1
				* 彌右衛門 (人)	1	—	1
				* 彌四郎 (人)	1	—	1
				* 彌平 (人)	7	—	7
				* 彌平太 (人)	1	—	1
				* 乙彌 (人)	1	—	1
18	筈	12	はす	筈	12	ひ 18	30
19	崎	11	サキ	* 岡崎 (人)	3	—	3
				* 川崎 (人)	2	—	2
				* 川崎 (地)	1	—	1
				* 洲崎 (人)	1	—	1
				* 高崎 (地)	2	—	2
				* 長崎 (地)	1	—	1
20	猿	9	エソ さる	猿類	1	—	1
				猿	7	カ 5	12
				* 猿若	1	—	1
21	紆	9	きろめえとる	紆	9	—	9
22	貰	9	もらう もらえる	貰う	8	ひ 15	23
				貰える	1	ひ 3	4
23	於	8	おく	於ける	8	ひ 123	131
24	曾	8	ソ	* 曾禰 (人)	1	—	1

番号	漢字	度数	音訓	その字を用いた語	(A) その度数	(B) 他の表記の度数	(A) + (B) その語の総度数
			ゾウ	未曾有	1	—	1
			かつて	曾て	4	嘗1 ひ9	14
25	殆	8	タイ	危殆	1	—	1
			ほとんど	殆んど	7	ひ8	15
26	眺	8	ながめる	眺める	7	—	7
				眺め廻す	1	—	1
27	讚 <small>(讚)</small>	8	サン	讚辞	1	—	1
				讚美	3	—	3
				賞讚(讚)	3	—	3
				礼讚	1	—	1
28	韓	8	カン	*韓(人)	1	—	1
				*韓国(地)	7	—	7
29	鳩	8	はと	*鳩自	2	—	2
				*鳩山(人)	6	—	6
30	椅	7	イ	椅子	7	—	7
31	廳	7	トン	廳	7	噸1 カ4	12
32	註	7	チュウ	註	7	—	7
33	靴	7	カ	製靴	1	—	1
			くつ	靴	5	—	5
				短靴	1	—	1
34	顛	7	テン	顛倒	1	—	1
				顛覆	5	—	5
				顛末	1	—	1
35	駄	7	タ	下駄	1	—	1
			ダ	駄目	1	カ4 ひ5	10
				*千駄ヶ谷(地)	1	—	1
				無駄	4	ひ1	5
36	僅	6	キン	僅々	1	—	1
			わずか	僅か	5	ひ8	13

番号	漢字	度数	音訓	その字を用いた語	(A) その度数	(B) その語の 他の表記の度数	(A) + (B) その語の総度数	
37	劃	6	カ ク	劃する	1	—	1	
				計劃	4	画 49	53	
				参劃	1	画 1	2	
38	坐	6	ザ すわる	連坐	1	座 1	2	
				坐る	4	座 1	5	
				押し坐らせる	1	—	1	
				*堺 (地)	6	—	6	
39	堺	6	さかい	*堺 (地)	6	—	6	
40	屯	6	ト ン	駐屯	6	—	6	
41	廿	6	にじゅう	廿	6	三+58 二}29 二〇12	105	
42	掩	6	エン おおう	掩蔽	4	—	4	
				掩う	2	—	2	
43	棚	6	た な	棚	5	カ 1	6	
				棚上げ	1	—	1	
44	此	6	シ この ここ これ	彼此	1	—	1	
				此	3	ひ 438	441	
				此処	1	ひ 52	53	
				此	1	之2 ひ 290	293	
45	熊	6	く ま	熊	4	—	4	
				*熊 (人)	1	—	1	
				熊打ち	1	—	1	
				羅漢	1	—	1	
46	羅	6	ラ	羅漢	1	—	1	
				羅紗	1	—	1	
				羅列	1	—	1	
				一帳羅	1	—	1	
				網羅	1	—	1	
				しらぎ	*新羅 (地)	1	—	1
				蜜柑	3	—	3	
47	蜜	6	ミ ツ	蜜柑	3	—	3	
				蜜蜂	3	—	3	
48	袴	6	コ	袴下	1	—	1	

番号	漢字	度数	音訓	その字を用いた語	(A) その度数	(B) その語の 他の表記の度数	(A) + (B) その語の総度数	
49	輯	6	はかま	袴	5	—	5	
			シュウ	特輯	3	—	3	
				編輯	3	集 5	8	
50	鹿	6	ロク	*鹿鳴	1	—	1	
			か	*鹿島(地)	1	—	1	
				馬鹿	4	莫迦 1 カ 2 ひ 2	9	
51	俺	5	おれ	俺	5	カ 1 ひ 4	10	
52	儲	5	もうかる	儲かる	2	—	2	
			もうけ	大儲け	1	—	1	
				金儲け	2	—	2	
53	吾	5	ゴ	*吾郎(人)	1	—	1	
				*安吾(人)	1	—	1	
				*宗吾(人)	1	—	1	
			われ	吾	1	我 2 ひ 4	7	
				吾々	1	我 5 ひ 3	69	
54	壺	5	つぼ	壺	4	—	4	
				骨壺	1	—	1	
55	尻	5	しり	尻	1	—	1	
				尻あがり	1	—	1	
				尻おし	1	—	1	
				尻拭い	1	—	1	
				言葉尻	1	—	1	
56	巴	5	バ	*巴里(地)	5	カ 4	9	
57	彦	5	ひこ	*彦三郎(人)	1	—	1	
				*彦三	1	—	1	
				*輝彦	1	—	1	
				*幹彦	1	—	1	
				*光彦	1	—	1	
58	戟	5	ゲキ	刺戟	5	—	5	

番号	漢字	度数	音訓	その字を用いた語	(A) その度数	(B) その語の 他の表記の度数	(A) + (B) その語の総度数
59	戴	5	タイ	戴冠	2	—	2
				頂戴	3	—	3
60	戻	5	もどす	とり戻す	1	—	1
				引き戻す	2	—	2
				取り戻せる	1	—	1
			もどる	戻る	1	ひ 1	2
61	扉	5	とびら	扉	5	—	5
62	拭	5	シヨク	払拭	1	—	1
				尻拭い	1	—	1
				拭う	1	—	1
				拭く	2	—	2
63	捲	5	まく	捲き込む	5	巻 1 ひ 1	7
64	斯	5	シ	斯国	1	—	1
				斯よう	1	ひ 2	3
				斯る	1	ひ 6	7
				斯く	1	ひ 9	10
				*波斯(地)	1	カ 1	2
65	溢	5	イツ	溢血	1	—	1
				溢れる	3	—	3
				溢れ出る	1	—	1
66	瓶	5	ビン	瓶	3	—	3
				鉄瓶	1	—	1
				釣瓶	1	—	1
67	莫	5	バ(カ)	莫迦	1	馬鹿 4 力 2 ひ 2	9
				莫大	4	—	4
68	蒙	5	モウ	*蒙古(地)	1	—	1
				啓蒙	2	—	2
				*満蒙(地)	1	—	1
			こうむる	蒙むる	1	ひ 1	2

番号	漢字	度数	音訓	その字を用いた語	(A) その度数	(B) 他の表記の度数	(A) + (B) その語の総度数
69	迎	5	たどる	迎る 迎り歩く	4 1	ひ 2 —	6 1
70	逢	5	あう	逢う 出逢ふ	4 1	会 19 1 合 11 会 1 ひ 9	44 2
71	鍋	5	なべ	鍋 *鍋山(人)	1 4	— —	1 4
72	罐	5	カン	罐	5	—	5
73	隅	5	すみ	隅 片隅	1 4	— —	1 4
74	雰	5	フン	雰囲気	5	—	5

4 語彙の総量の推定

4・0 総合雑誌の一年分の本文は、どれほどの数の異なる語で書いてあるか。延べ語数(総使用度数)が標本値から推定出来るのだから、異なり語数についても同様な推定値が得られないものか。異なり語数の性質上、延べ語数のようには簡単でないが、一往の推定法を見出したので、中間報告をする。結果をまず示せば、その語彙の総量は約二万語と推定された。

4・1 推定法のアイデア

標本における異なり語数を知っても、それだけでは調査対象の語彙の総量を推定する事は出来ない。それは使用度数(延べ語数)がたとえば2倍になっても異なり語数は2倍近くはならず、それより少ない語数となるからである。つまり異なり語数が増える速度は次第に鈍くなる訳である。一方もし延べ語数 n が増すに連れ異なり語数 k がどう変るかが函数で表現出来れば、すなわち $k=f(n)$ になるような函数 $f(n)$ が見つかれば、これを使って n に一年分の本文の延べ語数を代入することにより、語彙の総量が推定される。

そこで、ランダムに抜いた延べ n_1 語についてまず整理し、異なり語数が k_1 である事を確かめる。次に n_2 語($n_2 > n_1$)までを整理した結果から k_2 ($k_2 > k_1$)という値を得る。こうして観測値(n_i, k_i)の組を記録して、それらの観測点



(の近く)を通るような函数を求める事が、この問題を解く第一段階である。幸な事に、われわれの得たデータは第1図に見る通り適当な曲線で相当よく近似出来そうである。

4・2 推定に使う函数を定める事

一般に、ある大きな言語集団のそれほど短くはない時期の言語表現を考えれば、そこに使っている延べ語数 n は無量大だとしてよい。しかしその場合も異なり語数 k には限りがある、そうやたらに沢山の語が使われる訳ではない。この限界の値を L とする。従って

前提 i $n \rightarrow \infty$ の時 $k \rightarrow L$.

一方、 n の変域の全体にわたってよく近似する函数が求められたとすれば、その函数は言語理論上自明な次の二つの条件をも満たさなくてはならない。すなわち

前提 ii $n = 0$ の時 $k = 0$;

前提 iii $n = 1$ の時 $k = 1$.

今第1図をながめて、第一次近似として次の想定をする：延べ語数で n までを調べた時の k の増加率 $\frac{dk}{dn}$ が、語彙総量の L から k を引いた値に比例する。この仮説は適当な比例定数 a を選ぶ事によって式(1)のように表わせる：

$$(1) \quad \frac{dk}{dn} = a(L-k).$$

微分方程式(1)を変数分離法で解くと、

$$(2) \quad k = L(1 - e^{-a(n-b)}), \quad \text{ただし } b \text{ も定数.}$$

さて e は自然対数の底 2.71828... であり、 a は定数であるから、 $e^{-a} = c$ と置こう。そうすると式(2)は

$$(2') \quad k = L(1 - c^{n-b}).$$

と表わされる。この式(2)、同じ事であるが式(2')が、前提 i を満たす事は明らかである。条件 ii, iii を満たすようにすれば、定数 a, b にもある制限がつくが、語彙の総量の推定のためにはさほど重大でないから、こゝでは詳しくは述べない。

式(2)または(2')がわれわれのデータの相当よい近似函数である事は、次のようにして確かめられる。延べ語数がある特定の値 n に取った時の異なり語数を、改めて k_n と書こう。つまり

$$k_n = L(1 - e^{-a(n-b)}) = L(1 - c^{n-b})$$

こゝで n を任意の実数(実際には整数)とすれば、延べ語数が $(n+h)$ の時の異なり語数 $k_{n+h} = L(1 - c^{(n+h)-b})$ は、代数的な変形によって

$$(3) \quad k_{n+h} = c^h k_n + L(1 - c^h).$$

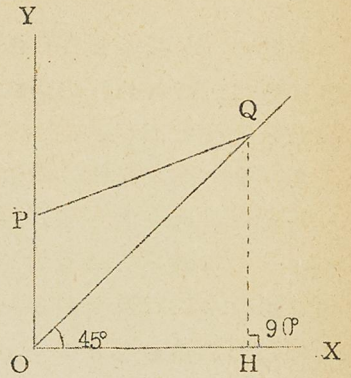
(26)

もし c を常に一定の値に固定して置けば c^h は定数 C_1 となり、従って $\angle(1 - c^h)$ も定数 C_2 となる。この時式(3)は

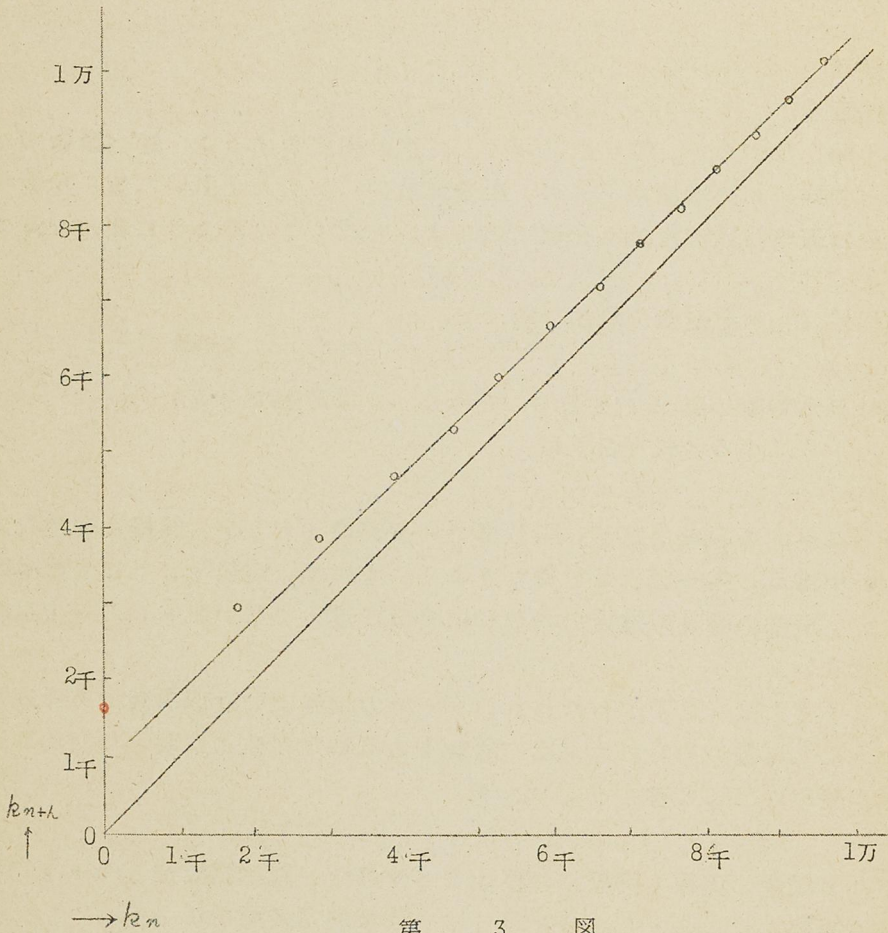
$$k_{n+h} = C_1 k_n + C_2$$

であるから、点 (k_n, k_{n+h}) を順々にプロットして行けば、その点は第2図に示すような、ある直線 PQ に乗る。

観測点を使ってこのような定差図(第3図)を描いて見ると、第四番目の観測点以降は、事実、きわめてよく直線をなしている。この事は、先の仮説が、延べ語数の余り小さくない範囲では、実



第 2 図



第 3 図

際の結果とよく合う事を示すものである。たゞいゝの小さなあたりで観測点がこの直線から傾向的にそれているとしても、当面の問題は \angle を推定する事であるから、 n の大きな部分が直線でよく近似されさえすれば、一向にさしつかえない。何となれば \angle は n のきわめて大きな値—— $n=\infty$ と見なせるような場合に対応する n の値だからである。

4・3 \angle の推定値

立返って第2図で、原点 O を通り軸 OX と 45° をなす直線 OQ と先の PQ との交点を Q とすれば、 \angle の値は QH (同じ事になるが OH) の座標の値で求められる。

そこで第3図を視察して第四ないし第十四番の観測点の近くを通るような直線を求める。このグラフから数学的な操作によって \angle 、 b 、 c の近似値が計算出来る。その値から出発し、最小二乗解を求めれば、 \angle の一層よい近似値が導けるのである。実際には、

$$\angle \text{の推定値} = 19450 \text{ 語} \qquad \text{推定量} \angle \text{の標準誤差} = 2740 \text{ 語}$$

となった。

4・4 結論および付記

上に述べた方法により、現在までに整理を終えた延べ語数約六万の標本値を用いて、総合雑誌一年分の本文の語彙の総量を求めると、それは19450語、高24830語(信頼度95%で)と推定された。

なおこの値を算出するのに使ったデータは、検査を一回しか通していないので、再検によって数値がある程度動くであろう。また観測誤差に関する十分なインフォメーションを得ていないので、精度もさほどよくはない。これらは今後更に精密に調べる事によって改良されるであろう。推定値を算出する基礎とした函数についても改良の余地はある。更にこの推定法には外挿を用いたので、なお検討を重ねるべきであろう。しかし上述の所は、今までほとんど不可能として放置されて来た問題に対し、第一次近似を与えたという意味で、こゝに報告する次第である。詳細は別の機会に発表したい。