

# 国立国語研究所学術情報リポジトリ

On concentricity and similarity of vocabularies in news, editorials, and other newspaper texts

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: 村木, 新次郎, MURAKI, Shinjiro メールアドレス: 所属:
URL	<a href="https://doi.org/10.15084/00001019">https://doi.org/10.15084/00001019</a>

# 用語の集中度と共通度

——新聞各層間の比較において——

村 木 新 次 郎

この稿では、新聞の用語を二つの量的な側面から考察する。二つの側面というのは、一つは用語の集中度についてであり、他の一つは、用語の共通度についてである。

新聞には、いろいろな文章形態があるので、それぞれの角度からの区分が行なわれた。<sup>(注1)</sup>ここでは、「目的や動機」あるいは「表現活動の性格」によって分類された、16の「文種」の各層をとりあげ、それらを比較してみたい。16の区分とは、以下のものをいう。<sup>(注2)</sup>

- |         |           |           |
|---------|-----------|-----------|
| 01 ニュース | 02 ニュース解説 | 03 社説コラム  |
| 04 特集記事 | 05 特別記事   | 06 評論     |
| 07 実用読物 | 08 探訪ルポ   | 09 ニュース展望 |
| 10 記録通知 | 11 紹介記事   | 12 読者作文   |
| 13 相互通信 | 14 小説     | 15 商業広告   |
| 16 案内広告 |           |           |

「01」～「16」の数字は計算機処理のために用いたものであるが、表などで便宜上この数字を使うことがある。

さて、用語の集中度というのは、異なり語数の伸びと延べ語数の伸びとの関係を、使用率のたかいものについてや全体の分布についてみるものであり、そこでは個々の語を問題の対象から除外し、ただ用語の分布のみに着目する立場をとる。一方、用語の共通度というのは、二つの分野にそれぞれ使用されている用語がどの程度共通しているかをみるものである。

分析の対象として扱った用語は、新聞三紙一年分からエリア・サンプリング（抽出比1/60）で調査対象となった長単位語、約200万語である。

この調査の対象は、句読点、記号類なども用語として含まれていること、また、動詞、形容詞などの活用する語について、その変化形がひとつにまとめられていないこと、さらに、同じ表記で別の語であるもの（たとえば「米」のように、「コメ」「アメリカ」「メートル」など別々の語が同一の表記をとっているもの）の区分がなされていないことや、別の表記をとってはいるが、実際は同じ語であるもの（たとえば、「言葉」と「ことば」と「コトバ」）の統一がなされていないこと、などの点で従来からあった語彙表とは性格を異にするものである。それは、語彙論的<sup>(注3)</sup>な性格を欠く、いわば同表記（語彙）表、あるいは同形文字列表とでもいうべきものである。

語彙論的な語彙表でないことの不備は、語彙構造の細かい分析や他の語彙調査との比較などを試みる場合、それに十分にこたえられるものではないが、しかし、ここではそういった不備を承知しつつ、新聞の各層間の用語の量的構造を概観してみたい。

なお、この種の分析は、次のような前提があってはじめて成立つものと思う。この種の分析とは語彙の計量的分析という意味であるが、そこでは、調査対象となる各々の語がすべて等質であると考え、これが前提となる。ここで等質ということの意味は、「人間の生きがい」というときの〈人間〉も、「変わった人間」というときの〈人間〉も同じものとして扱うということであり、さらに、〈人間〉も〈の〉も〈生きがい〉も〈変わっ〉も〈た〉も同じレベルのものである、ということでもある。もちろん、語の認定の仕方や、〈人間〉〈生きがい〉と〈の〉〈た〉との間にレベルの違いを設けるということもありうるが、それもやはり上の前提のヴァリエーションでしかない。そういった前提のうえにはじめて統計的処理が意味をもってくる。語の使用率も分布曲線もすべてこの前提のもとではなしである。

ところで、私自身このような前提に疑いを持たないわけではないが、前提はいくまで前提であり、その前提のもとで何かがつかめ、それが用語の問題を把握する一助となるのではないかと考える。前提のたて方については、たとえば記号類や句読点は語と認めない方がよいとか、自立語のみを対象として分析したほうがよいとか、いろいろ立場によって変わってくる。ここでは、コンピュ

ータを利用して得た調査データを用いるので、人間向きの語の規定というよりは、機械向きの語の規定ともいえるものに従っての前提である。

ところで、この用語調査は、サンプリング理論を適用しているため、本来ならばそれをふまえて分析すべきであるが、この稿では、標本についてのみ考えていることをお断わりしておく。

注1) 話題・文種・情報源記載形式(無記名か記名入かなど)・記事の位置(見出し、標題、本文などの区別)の四点から立体的にとり扱われている。詳しくは、国研報告34. 林四郎氏『新聞語彙調査における層別とその意味』参照。

注2) 上の文献では17の文種に区分され、「漫画」が含まれているが、この層はデータ数が少ないため、ここでは省略した。なお我々は、注1の区分を層別とよんでいるため、各分野のことを層とよぶことがある。ところで層別区分に対する疑問がある。どこからどこまでが、どの層であるといった区分が、実際上大変むずかしい。それは、ある意味では、スペクトルのように連続するもので、どうしても、中間の性格のものが出てくる。ここでは記事単位によって処理されたデータを利用するわけであるが、そこにあらわれるエラー処理(層分けの不完全さ)は全体として互いに相殺しているものとみることとする。大量の調査には個々のデータのエラーは、ある程度、まぬがれえない。

注3) 仮に、語を「形態論的な語」「語彙論的な語」「意味論的な語」と区別するときの語彙論的な語彙表を指している。この区分に従えば、「形態論的な語」のリストは形態表、「意味論的な語」は意味表というべきで、語彙表とはっきり区別する必要がありそうだ。この稿で使用した用語表は、その意味では形態表にあたる。

## I 用語の集中度

この章では、個々の語がどういう語であるかという語の種類、性質などは問題としない。もっぱら語彙を総体的にみて、その集中度をはかる。

### (1) 使用率を規準にはかる方法

各々の分野で用いられている語を、使用率のたかいものからひくいものへと整理したとき、それはどのように分布するであろうか。さきに挙げた16の層と新聞全体について、ある一定の使用率以上の見出し語の数と、それらの見出し語の延べ語数の和がその層全体の延べ語数に占める割合とを調べてみた。表I 1.はその結果である。

表のPの欄に示した、それぞれの値以上の使用率をもった見出し語の数が、各層別に示してある。括弧内の数字は、その見出し語で延べ語数のどの程度を占めるかの割合(%)を表わしている。Pの値は便宜的に定めた。各層間を比較するのが目的なので、区切りやすい数字を適当に選んである。級間隔は分布の性質上一定ではない。もっと細かく区分することもできるが、層の数が多く煩雑になるのをさけて、大まかに7区分とした。

この表から、語の使用率の分布における各層間の似寄りや違いなどをおおむねつかむことができる。

たとえば、「ニュース(01)」と「ニュース解説(02)」と「社説コラム(03)」の三層をみてもみると、10パーミル以上のところを除いて、

ニュース < ニュース解説 < 社説コラム

という関係が成立する。「社説コラム」は「ニュース解説」より、「ニュース解説」は「ニュース」より用語が集中していることを示している。用語が集中しているというのは、全体の一部に偏しているという意味で用いている。

全体的にみて、極度に他のものから逸脱しているのは、「案内広告」、「商業広告」、「記録通知」である。「記録通知」は、経済欄の銘柄表やラジオ、テレビの番組表などを含んでいる関係上、ふつうの文章とは性格を異にする。また、「商業広告」はふつうの広告、「案内広告」はいわゆる三行広告を指し、これらの二層も他の分野に比べると異質なものである。これらはいずれも非文章形態の表現をとっている層だといえる。

この章では、個々の語を問題の対象外とする立場であるが、いま使用率が上位にある語と分布とを参考までにみてもみると、全体に使用率のきわめてたかい語というのは、「。」、「」、「の」「を」「た」「で」「こと」「もの」など句読点や助詞助動詞、それに形式名詞の類である。語を内容語(content words)と機能語(function words)に分けると、どれも機能語にあたるものである。それらは、個別に概念をもっているのではなく、文や文章の中で、句や文や語をつなぎ互いの関係を示す働きをしたり、形式名詞のようにそれ自体漠然とした概念しかもってなくて文脈の中でその働きが規定されるものであったりする。要するに、文章中に使用されてはじめてその役割をはたす語である。広告欄や

表 I 1. 用語の使用率の分布

P(%)	50	20	10	5	2	1	0.5
01	2 (12.3)	8 (28.8)	13 (36.5)	19 (40.1)	35 (45.6)	62 (49.2)	114 (52.6)
02	2 (12.4)	9 (31.7)	11 (34.9)	23 (42.8)	37 (46.7)	65 (50.5)	142 (55.8)
03	2 (13.3)	9 (33.1)	12 (37.5)	21 (43.7)	37 (48.5)	68 (52.8)	129 (56.9)
04	2 (12.6)	9 (32.2)	11 (35.3)	23 (43.9)	34 (47.2)	64 (51.4)	129 (55.8)
05	2 (12.6)	9 (32.3)	11 (35.7)	23 (43.9)	34 (47.5)	66 (51.9)	121 (55.4)
06	2 (12.6)	9 (32.4)	12 (36.7)	20 (43.2)	36 (48.4)	69 (53.1)	137 (57.7)
07	2 (12.2)	8 (29.9)	11 (34.3)	24 (42.4)	37 (46.3)	75 (51.3)	133 (55.1)
08	1 (5.2)	8 (25.7)	16 (38.9)	19 (40.6)	40 (47.6)	63 (50.6)	123 (54.9)
09	2 (12.8)	8 (29.6)	13 (36.9)	17 (39.7)	35 (45.0)	67 (49.4)	— —
10	1 (6.6)	6 (20.2)	11 (28.4)	20 (34.9)	42 (42.3)	63 (45.3)	143 (50.4)
11	2 (12.5)	9 (31.1)	13 (36.8)	22 (42.9)	34 (46.6)	55 (49.5)	108 (53.2)
12	2 (11.4)	8 (29.0)	11 (33.9)	22 (41.4)	37 (46.4)	72 (51.5)	138 (56.0)
13	1 (6.4)	8 (28.1)	10 (31.2)	23 (40.6)	42 (46.7)	77 (51.4)	156 (57.5)
14	1 (6.8)	9 (33.1)	13 (39.6)	20 (44.1)	35 (48.8)	67 (53.3)	145 (58.9)
15	0 (0.0)	6 (16.9)	11 (24.3)	19 (30.2)	31 (33.9)	57 (37.6)	108 (41.0)
16	0 (0.0)	4 (13.6)	5 (15.5)	12 (20.2)	49 (32.1)	109 (40.7)	221 (48.4)
全体	0 (0.0)	6 (16.9)	16 (30.9)	20 (33.7)	32 (37.4)	60 (41.4)	134 (46.5)

その使用率までの累加異なり語数 (括弧内は層全体の延べ語数に対する割合%)

記録通知の層ではそういった語の使用率がひくくなり、数字や記号類の使用率とならぶ。そのために50パーミル、20パーミルの使用率をこえる語というの

は、ふつうの文章からなる他の層より少ない。

各層をまとめた新聞の語彙全体の使用率の分布で、きわめてたかい使用率をもつ語が少ないのは、各層の上位にある語が相殺されているからであろう。

このように比較的使用率のたかい語がどのように分布しているかを知ることによって、ある程度、語彙の集中度がわかる。

## (2) 使用順位を規準にはかる方法

使用率のたかい順に語をならべ、ある一定の順位までの累加延べ語数とその分野に含まれる全体の延べ語数との比を集中度と考えることもできる。フランスのP. ギローは上位50語までと全体の延べ語数との比をもって集中度として(注4)いる。それを式にあらわせば次のようになる。

$$C = \frac{\sum_{i=1}^{50} f_i}{\sum_{i=1}^n f_i}$$

(ただし、 $n$ : 異なり語数, 使用率( $i$ )  $\geq$  使用率( $i+1$ )とする)

ギローの考案したこの式が、ただちに日本語の場合に妥当であるかどうかは検討してみる必要がある。また、使用するデータが語彙論的な語彙表でないものであるため、いくぶんこの式を修正する必要もありそうだ。そういった点の考察は保留して、とりあえずこの式を多少延長した形で各層間の語彙の集中度を測定してみた。その結果は表 I 2. に示してある。

ところで、この場合、延べ語数の違いが気になる。ちなみに、延べ語数の多い「ニュース」「記録通知」「案内広告」などの層の集中度が他よりひくい。比率を問題にしているのであるから、実測数の多少にかかわらず、少なく調べたものと多く調べたものが、縮図と拡大の関係を呈しているといえるであろうか。しかし、そうとはいいがたい。異なり語には、おのずと限界があるはずだと思われる。調査対象をどんどんふやしても、異なり語がふえない状態があるはずである。そのことを考えると、やはり延べ語数の違いは問題点となる。異なり語の限界がどのあたりにあるかは興味ある問題であるが、今後の課題とする。それは当然、その分野によって違うはずである。

延べ語数の差異による障害はあるが、表から判断すると、各層間の集中度の

表12. 用語の集中

(比率はいずれも%)

	延べ 語数	異なり 語数	百分比	上位 50語	上位 100語	上位 200語	上位 500語	上位 1000語
ニュース	452,876	66,900	23.02	47.85	51.83	55.85	61.89	66.63
ニュース解説	29,217	7,985	1.48	48.78	53.36	58.11	65.11	71.23
社説コラム	57,554	12,818	2.93	50.63	55.30	59.76	66.26	71.76
特集記事	53,427	12,913	2.72	49.76	54.13	58.60	64.92	70.31
特別読物	101,851	21,798	5.18	50.00	54.31	58.61	64.46	69.25
評論	45,000	10,936	2.29	50.72	55.54	60.24	66.76	72.11
実用読物	41,844	10,272	2.13	48.35	53.22	57.80	64.70	70.70
探訪ルポ	80,950	16,458	4.11	49.15	53.64	58.03	64.67	70.26
ニュース展望	5,184	2,070	0.26	47.41	52.33	58.68	69.71	79.36
記録通知	426,804	53,519	21.69	43.74	48.00	52.92	59.59	64.58
紹介記事	75,580	18,793	3.84	48.98	52.75	56.60	62.33	67.31
読者作文	31,192	8,099	1.59	48.65	53.85	58.73	66.01	72.25
相互通信	28,538	7,624	1.45	48.11	53.48	58.92	66.16	72.47
小説	34,754	8,347	1.77	51.29	56.11	61.28	68.17	74.08
商業広告	200,576	45,952	10.19	36.79	40.58	44.36	50.42	56.09
案内広告	299,651	42,255	15.23	32.31	39.72	49.81	57.66	65.32
その他	2,577		0.12					
全体	1,967,552		100.00	40.22	44.58	49.06	54.88	59.94

関係は次のようになる。「>」は左辺が右辺よりいくぶん集中度のたかいことを示し、「>>」は、その差のかなり大きいことをあらわす。

さきの式の一般式である

$$\frac{\sum_{i=1}^s f_i}{\sum_{i=1}^n f_i}$$

において、

(1) s=50のとき

小説>評論・社説コラム>特別読物・特集記事>探訪ルポ・紹介記事・  
 ニュース解説・読者作文>実用読物・相互通信・ニュース・ニュース展  
望>>記録通知>商業広告>案内広告

(2) s=100のとき

小説>評論・社説コラム>特別読物・特集記事・読者作文・探訪ルポ・  
相互通信・ニュース解説・実用読物>紹介記事・ニュース展望>ニュー



ス≫記録通知≫商業広告>案内広告

(3) S=200のとき

小説>評論・社説コラム>相互通信・読者作文・ニュース展望・特別読物・特集記事>ニュース解説・探訪ルポ・実用読物>紹介記事・ニュース≫記録通知≫案内広告≫商業広告

上の結果をみるかぎり、大部分のものは、上位50語であれ、100語であれ、200語であれ、この範囲であれば集中度の大小関係は大きく変化しないようである。その例外は「ニュース展望」「相互通信」「案内広告」などであろう。

「ニュース展望」は、上の図で次第に集中度のたかいほうに移行する。この層は、他の層に比べて標本延べ語数が少なく、使用率の分布そのものが安定していないからであろう。この層は、「社説コラム」に近い文章であると予想され、図のもっと左にきてしかるべきものである。上位500語までの集中度測定の結果によると、「ニュース展望」が一番左にきて、そのあと「小説」「評論」「社説コラム」と続く。上の図での、この層の移動は、やはりデータが少量であるためであろうと思われる。

「相互通信」も次第に集中度のたかいほうに移行している。これは、「ニュース展望」の次に標本延べ語数が少ない分野であるが、このあたりの使用率は安定していると思われ、「相互通信」のこの変化はこの層の特徴になるであろう。「案内広告」は「商業広告」との大小関係が、途中で入れかわる。こういったことは、あとでとりあげる分布曲線を描くことでより明らかになる。

以上、ギローの集中度の公式を延長させて利用してみた。その結果、各層の語彙の集中の度合を整理してみると次のようになろう。もちろん、新聞記事という範囲内での相対的な規定である。

集中度がややたかい層……………小説 評論 社説コラム

集中度がふつうの層……………特別読物 特集記事 読者作文 相互通信  
探訪ルポ 紹介記事 ニュース展望 ニュース解説 実用読物

集中度がややひくい層……………ニュース

集中度がかなりひくい層……………記録通信 商業広告 案内広告

これらのグループ分けは、集中度のたかいものがより文章的であり、反対に集中度のひくいものがより非文章的であるといえそうである。この範囲内では語彙の集中度の高低と文章的であるかそうでないかの二つの尺度の関係が強いようである。このことが一般性をもつかどうかはまだわからない。句読点やいわゆる付属語も含んでいることが以上の結果を左右していると思われる。

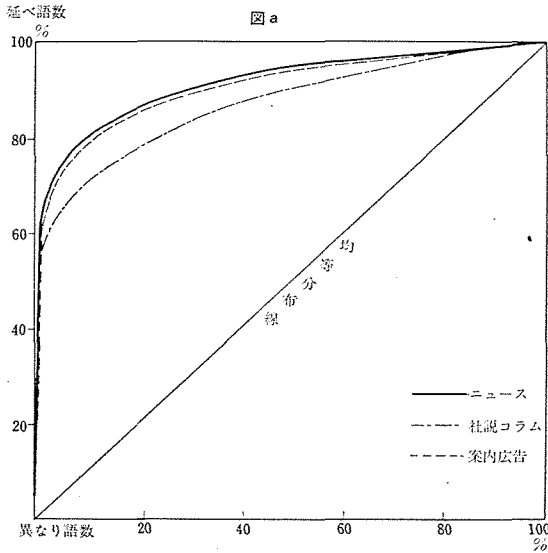
「小説」で集中度がたかいのは、あるいは、固定した個人の執筆によるせいだ。「小説」では、人名地名なども定まったものしか使われない。これは、延べ語数を小さくし、ひいては集中の度合をたかくする。また、執筆者が一定していることは、表記のゆれも少なくしていると予想される。これも、集中度をたかめる要因であろうか。

### (3) 分布曲線（ローレンツ曲線）による方法

(1) や (2) の方法では、比較的使用率のたかい語のみを扱ってきた。上位語の部分で語彙の集中性をみたのであるが、全体の語彙の分布から語彙の集中性を測ることはできないものか。ローレンツ曲線は、経済学の分野で元来所得の分布状態を比較する方法として考案されたものであるが、言語現象の分析にも利用されているようである。<sup>(注5)</sup> ここでも、異なり語数の伸びと延べ語数の伸びとの関係をみるのに、ローレンツ曲線が応用できそうである。おのおの

表 I 3.

異なり語累積比率	0.5%	1%	2%	5%	10%	50%
ニュース	59.25	66.02	68.72	75.59	80.87	
社説コラム	49.04	56.82	61.41	68.19	73.83	88.86
特集記事	51.42	55.77	60.28	66.84	72.53	87.92
特別読物	54.81	59.15	63.77	69.77	75.19	89.30
評論	51.40	57.80	60.85	66.72	72.86	87.85
探訪ルポ	52.39	56.78	61.55	68.63	74.74	89.84
記録通知	55.11	60.06	65.12	74.66	81.73	
紹介記事	52.42	56.24	60.44	66.83	72.36	87.57
小説	50.03	54.77	59.97	66.74	72.48	87.99
案内広告	47.94	55.74	63.46	73.41	80.34	92.95



の伸びとは累積度数のパーセンテージのふえ方という意味である。使用率がすべての語について同じなら、図 a の対角線である、均等分布線に一致する。そうでない場合は、弓状の曲線を描く。ふつうは計測値の小さいものから大きいものへと示されることが多く、曲線は対角線より下側にくるのであるが、ここでは、さきの使用率の分布のときと同じように、使用率の大なるものから小さなものへと移行させたので、図のような対角線の上側に弓形の曲線ができる。そうした場合、この曲線と対角線とで囲まれた面積をもって分布の集中度とすることができる。この面積が大きいほど集中の度合いが強いことは当然のことである。面積は、積分によって求めることが可能であるが、この場合だと、シンプソンの近似式を変形して、その値を求めてもいいだろう。集中度  $C$  は次式によって求められる。

$$C = \frac{1}{6} \{2y_1 + y_2 + 2y_3\} - 2.5$$

$y_1, y_2, y_3$  は、異なり語数のそれぞれ 25%、50%、75% の延べ語数の百分比を 100 で割った値である。対角線と弓形でできる面積（すなわち集中度）は、対角線と他の二辺とによってできる直角三角形より大きくなることはなく、矩形の面積を 1 とすると、集中度  $C$  は、0 と 0.5 の間の値をとる。

表 I 3. は主なものの10層について異なり語のそれぞれの百分比 (0.5, 1, 2, 5, 10, 50%) に対して延べ語の百分比のいくらにあたるかを計算したものである。異なり語で、その百分比が  $i$  番目と  $(i+1)$  番目の間にきたときは、 $i$  番目までの延べ語数の和に、 $i$  番目と  $(i+1)$  番目の語の延べ語数を比例配分し、それを加えて計算した。

この方法で、次の五層について測定した結果を示す。

評論	0.313
案内広告	0.318
特別読物	0.322
社説コラム	0.350
ニュース	0.358

調べた五層に関しては、「評論」が一番集中の度合が小さく「ニュース」のそれが一番大きい。

使用率のたかいものも、きわめてひくいものも集中度の値に関係してくるのであるから、この方法による測定は、より精緻だと一往はいえる。ところで、上の数値と、(1) や (2) で得られた結果とをあわせて考えてみると、「ニュース」の場合、その性質上一回きりしか使われない人名地名などが多く、それが原因で (3) の集中度の値がたかいのだと思う。「評論」は、(1)(2) の結果より、上位語にかなり集中していることがわかるが、上位語を除いたものは使用率が均一化しているとみられる。

ともかく一般に、上位語だけでみた集中度と全体の分布での集中度とは別のものである。前者は、集中度というよりは、上位語の反復度ともいうべきものである。後者は、すべて同じ使用率であることからどの程度かたよっているかを測るものであり、そのかたよりは、上位語のかたよりと、頻度数のきわめてひくいものの多少によって決められる。図 a のようなローレンツ曲線を得るのはふつつ大変やっかいなことなので、上位語へのかたよりを (2) の方法でもとめ、さらに頻度数 1 や 2 の全体に占める割合をもとめ、二つをあわせて計量化し、(3) にあてはまる集中度を得ることもできよう。

参考までに各層の度数 1, 度数 2 であった標本語数を表 I 4. に掲げる。

表 I 4.

	延べ語数	異なり語数	標本度数 1	標本度数 2
ニュース	452,876	66,900	44,666	8,820
ニュース解説	29,217	7,985	5,710	1,129
社説コラム	57,554	12,818	8,984	1,729
特集記事	53,427	12,913	9,266	1,660
特別読物	101,851	21,798	15,498	2,755
評論	45,000	10,936	7,957	1,401
実用読物	41,844	10,272	7,132	1,421
探訪ルポ	80,950	16,458	11,114	2,236
ニュース展望	5,184	2,070	1,607	254
記録通知	426,804	53,519	34,045	6,851
紹介記事	75,580	18,793	13,744	2,362
読者作文	31,192	8,099	5,804	1,033
相互通信	28,538	7,624	5,526	963
小説	34,754	8,347	5,983	1,066
商業広告	200,576	45,952	30,726	6,469
案内広告	299,651	42,255	27,311	5,553

表 I 5. 語彙の豊かさ

ニュース	99.40
ニュース解説	46.72
社説コラム	53.43
特集記事	55.87
特別読物	68.30
評論	51.56
実用読物	50.20
探訪ルポ	57.84
ニュース展望	28.75
記録通知	81.92
紹介記事	68.36
読者作文	45.86
相互通信	45.13
小説	44.78
商業広告	102.60
案内広告	94.56
全体	

これに関連して、延べ語数と異なり語数とをもって語彙の豊かさ (R) をはかる、ギローの次の式に、今回の調査結果を応用してみた。

$$R = \frac{\text{異なり語数}}{\sqrt{\text{延べ語数}}}$$

この式によって、各層の語彙の豊かさ (R) をもとめた結果が表 I 5. である。これによると、「商業広告」「ニュース」「案内広告」などの分野が、語彙が豊かであることを示し、反対に、「ニュース展望」が豊かでないことを示している。ところで、前者の三層は、共通して、その標本述べ語数が多く、後者はそれが他より少ない。この延べ語数による違いによって起こった誤差が表 I 5. の結果に含まれるのではないかと思う。語彙の豊かさをはかる尺度として、ユールの「K特性値」などもあり、今後検討してみたいと思っている。

注 4) この式の分母は、ギローの場合、 $2N$  と示され、 $N$  は most fots の数をあらわしている。総語数は、その倍であるとされている。

注 5) 宮島達夫氏が「<かかり>の量的性質」(『現代雑誌九十種の用語用字第三分冊』所収) の中で、ローレンス曲線の考え方にもとづいて、<うけ>の集中度の測定を、試みている。

また、野村雅昭氏も、「用語用字調査における用語と用字の関係についての実

験(その1)」(『LDD4』所収)の中で、宮島氏にならって漢字の使用が、特定の語に集中するか、あるいは数多くの語に均等に用いられるかを数量化している。

## II 用語の共通度

いくつかの分野があって、それぞれの分野の中に使用されている用語が、どの程度共通しているかを測定する。それは、用語の共通度、あるいは用語の類似度といえようが、このことが、ただちに分野(あるいは一般に作品)の共通度、類似度とみることはできない。ただ、作品の似寄りの一つの大きな要因として、用語の共通度をあげることはできよう。

さて、作品Aと作品Bの用語の共通度を測る方法としては、かつて『計量国語学』誌上で次のような式が示された。

$$C_{AB} = \frac{x}{a+b-x} \quad (\text{注7}) \quad (a, b: \text{各作品の見出し語の数}, x: \text{二つの作品に共通な見出し語の数})$$

$$C_{AB} = \frac{x}{\sqrt{ab}} \quad (\text{注8}) \quad (\text{同上})$$

これらの式は、二つの作品の長さの違いによって、大きな影響を受ける。宮島達夫氏はこういった式を検討された末、次のような、使用率により用語の共通度を測る方法を示された。<sup>(注9)</sup>

$$C_{AB} = \sum_i \min [P_i(A), P_i(B)] \dots \dots \dots (\text{IIa})$$

この式の意味するところは、ある語が、作品Aにおいては、その使用率が  $P_i(A)$ 、作品Bにおいては、その使用率が  $P_i(B)$  であるとき、その小さいほうの値をすべての見出し語についてもとめ、それらの総和が二つの作品の共通度であるということになる。そして、これは次式のように、表わすこともできる。(詳しくは『国語学』82を参照されたい)

$$C_{AB} = 1 - \frac{1}{2} \sum (|P_i(A) - P_i(B)|) \dots \dots \dots (\text{IIb})$$

IIa式やIIb式で表わされる、この作品の共通度  $C_{AB}$  のとりうる値は、0と1の間で、二つの作品に全く共通な語がみられない場合に0、各語の使用率



がすべての見出し語について、一致する場合に1という値をとる。実際には、 $C_{AB}$  の値が1に近づくほど共通の度合いがたかく、0に近づくほどその度合いがひくいということになる。

表Ⅱ1.に示したものは、宮島氏の示された測定方法により、各分野間の用語の共通度を調べた結果である。測定に使用した用語表は、はじめに断わったように同表記表あるいは同形文字列表ともいうべきリストであって、語彙論的な意味での用語表のリストではない。だから、語彙論的な、いわゆる語彙表にもとづき、同じ測定を試みるなら、ここに示した数値と、いくらか違った値を示すに違いない。しかし、新聞のデータは、雑誌などに比べれば比較的表記が案定していると予想されるので、一方に、同表記で別語という問題をはらんではいないが、各分野間の関係を大まかに知ることにはたいした差し障りはないであろう。本来ならば、標本度数の小さいものは、この種の分析に利用しにくいだが、標本どおしの比較ということで度数にこだわらず、すべての異なり語を対象に計算してみた。

表は、各層間の共通度測定結果のうち、もっともその値のたかかった「社説コラム～特集記事」を左肩にし、「社説コラム」との共通度のたかいものからひくいものへの順に、配列してある。この配列により、一般に表の右にいくほど、また下にいくほど共通度の値が小さくなる傾向をもつ。すべてがそうだというわけではない。中には順序が入れかわるものもある。

たとえば、「特集記事」と他の分野との共通度の値はどれも、上のものが下よりも大きく、「社説コラム」と他の層との共通度の関係と全くパラレルであるが、「特別読物」は、それと平行の関係を呈さないようである。

共通度のたかい組合せとしては、次のようなものがあげられる。

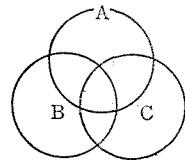
社説コラム～特集記事	.708
特集記事 ～特別読物	.706
特別読物 ～評論	.706
社説コラム～特別読物	.700
社説コラム～評論	.694

これとは反対に、「案内広告」「記録通知」「商業広告」は、どの層との共



通度もその値はひくい。「商業広告」はどの層との共通度もすべて0.3から0.4までの値をとり語彙の面で他の分野との距離がほとんど等しいことを知らせる。「案内広告」は、他との共通度で極めてひくい値を示し、この層独特の用語で構成されていることを、これらの数字は物語る。ただ、「商業広告」とだけは、0.36とこの層にとっては相対的にたかいは、広告という共通面があるからだろう。

この表は、全体を眺めることによって次のようなこともわかる。たとえば、「特別読物」や「評論」は、「社説コラム」「特集記事」との共通度がたかく、次の「ニュース解説」「ニュース」との共通度よりも「紹介記事」「読者作文」との共通度のほうがいくぶんたかい。これらの関係を図示すれば、右のようになる。(A=社説コラム or 特集記事, B=ニュース解説 or ニュース, C=紹介記事 or 読者作文, と考えればよい。)



また、一つ一つの分野について、他の分野との共通度の算術平均値をもとめ、そのたかい順に並べると次のようになる。

特別読物	.563	読者作文	.534	ニュース展望	.454
特集記事	.550	ニュース解説	.527	商業広告	.351
社説コラム	.544	実用読物	.522	記録通知	.233
評論	.543	相互通信	.521	案内広告	.164
紹介記事	.538	小説	.492		
ニュース	.538	探訪ルポ	.485		

この数値じたい特別意味をもつものではないと思うが、どの分野でも用いられている語を使っていれば、この値がたかくなるといえよう。「特別読物」や「特集記事」は、この値が他より比較的たかく、はばの広い(どの層にも出現しやすい)語が多く用いられていると考えられよう。「案内広告」はきわめてその値が小さく、その層に特有な語で表現されていることを示している。「記録通知」や「商業広告」も「案内広告」ほどではないが、この傾向がつよい。

この共通度は、類似度をはかる方法で、広くいろいろな方面に活用できるこ

とが、さきの文献で示されている。私は、共通度をはかる汎用のプログラムを作成し、電子計算機によって以上の結果をもとめた。データである用語のリストなどは磁気テープに収められているものであり、今後随時、所望のデータをとりだして同じような測定結果を短時間に入手できる。たとえば、他の区分による層別の類似関係をみるとか、句読点や記号類を除いたリストで各層間の類似関係をみるとか、語種別あるいは品詞別に应用するなど、いろいろ可能である。同じ区分による層分けで、記号などを含んだもの、それを省いたもの、語種別によるもの、品詞別によるもの、助詞や助動詞の使われ方などの類似の度合をはかり、それを総合的に眺めれば、分野ごとの語彙構成の性格や特徴を知る重要な手がかりになると思う。

また、いわゆる自立語にかぎられるが、各分野に用いられた語を、カテゴリーで分類して類似関係をみることもできる。分類の方法として、いろいろ考えられようが、たとえば、『分類語彙表』によって、グループを単位とし、その上で各分野間の共通度をみるのも一つの方法かと思う。電子計算機に『分類語彙表』を収録し、調査した用語とそれを照合させてグループに分ければ、このような規準でも類似関係をもとめることができる。

私は、調査資料をかえて、ニュースと社説とコラムの三つの分野について、それぞれの体言のみをひろい出し、この方法を試みた。これは、体言の類似度をはかることで、話題の類似をみようとしたわけである。この試みで調査対象としたのは、昭和47年4月に出た朝日新聞の朝刊で、ニュースは、この新聞の第一面の最初の記事、社説は社説欄の記事、コラムは天声人語であり、各々の分野で対象となる全文から、文を単位として150文をランダムに抽出し、その150文に含まれる体言を分類した。表II.2.はその結果であり、これに共通度の式をあてはめて得た値は、次に示すとおりである。

ニュース～社説	0.83
社説　～コラム	0.81
ニュース～コラム	0.78

つまり、ニュースと社説との類似度がもっともたかく、ニュースとコラムとの類似度がもっともひくい。これは、調査の対象とした三つの分野の話題の類

表II2.

( )内は百分比

	ニユー	社説	コラム		ニユー	社説	コラム			
抽象的関係	1.10	86 (5.21)	120 (8.24)	61 (6.35)	人間活動の主体	1.20	75 (4.55)	58 (3.98)	36 (3.75)	人間
	.11	58 (3.52)	53 (3.64)	28 (2.92)		.21	5 (0.30)	—	3 (0.31)	家族など
	.12	7 (0.42)	11 (0.75)	5 (0.52)		.22	4 (0.24)	4 (0.27)	2 (0.21)	相手・友など
	.13	32 (1.94)	52 (3.57)	19 (1.98)		.23	17 (1.03)	30 (2.06)	20 (2.08)	国民・階級
	.14	8 (0.48)	6 (0.41)	5 (0.52)		.24	91 (5.52)	26 (1.78)	41 (4.27)	職業など
	.15	65 (3.94)	67 (4.60)	31 (3.23)		.25	80 (4.85)	86 (5.90)	61 (6.35)	国家など
	.16	150 (9.09)	131 (8.99)	94 (9.79)		.26	45 (2.73)	36 (2.47)	18 (1.88)	社会
	.17	93 (5.64)	67 (4.60)	31 (3.23)		.27	67 (4.60)	52 (3.57)	25 (2.60)	機関(政府)
	.18	9 (0.55)	10 (0.69)	8 (0.83)		.28	96 (5.82)	23 (1.58)	12 (1.25)	団体
	.19	73 (4.54)	65 (4.46)	64 (6.67)						量・数・程度

	ニユー	社説	コラム		ニユー	社説	コラム			
人間活動—精神および行為	1.30	181 (10.97)	193 (13.25)	75 (7.81)	生産物および用具物品	1.40	2 (0.12)	4 (0.27)	2 (0.21)	物品
	.31	90 (5.45)	50 (3.43)	40 (4.17)		.41	8 (0.48)	3 (0.21)	1 (0.10)	資材
	.32	—	6 (0.41)	3 (0.31)		.42	2 (0.12)	3 (0.21)		衣食住
	.33	17 (1.03)	27 (1.85)	33 (3.44)		.43	3 (0.18)	4 (0.27)	19 (1.98)	住
	.34	22 (1.33)	26 (1.78)	12 (1.25)		.44	2 (0.12)	8 (0.55)	1 (0.10)	道
	.35	62 (3.76)	53 (3.64)	30 (3.13)		.45	10 (0.61)	5 (0.34)	5 (0.52)	具
	.36	76 (4.61)	43 (2.95)	20 (2.08)		.46	23 (1.39)	3 (0.21)	8 (0.83)	機械など
	.37	38 (2.30)	53 (3.64)	27 (2.81)		.47	1 (0.06)	4 (0.27)	7 (0.73)	地類
	.38	23 (1.39)	47 (3.23)	22 (2.29)						仕事

	ニユー	社説	コラム		ニユー	社説	コラム			
自然物および	1.50	1 (0.06)	5 (0.34)	4 (0.42)	自然現象	.56	4 (0.24)	3 (0.21)	13 (1.35)	動物
	.51	8 (0.48)	3 (0.21)	19 (1.98)		.57	5 (0.30)	4 (0.27)	15 (1.56)	からだ
	.52	2 (0.12)	3 (0.21)	6 (0.63)		.58	6 (0.36)	7 (0.48)	10 (1.04)	生命
	.55	3 (0.18)	3 (0.21)	25 (2.60)						植物

似性をものごとが持っているだろう。分野間での話題の類似度をはかるのに、こういったやり方も有効ではないかと思う。このようにグループごとに分類し、類似関係を測定すると、個々の語についてやるよりもとめる数値はたかくなる。話題による区分（T層）でこれを試みれば、相互関係がはっきりと示されると想像する。

ところで、新聞の用語調査が設計され、それが進行して、その結果得られたデータをその用語規定にしたがい、層別情報を利用して各分野間の用語の共通度をみてきたわけであるが、これらの結果を次のように、役立てることができる。共通度のたかいものをまとめた形でいくつかのグループに分けることができ、このような語彙表を使っていろいろ分析をすすめていくのに、より大きな分類（層分け）にしたがったほうが好都合な場合もある。また、多数の分野を少数の分野のグループに分けることは、全体の語彙構造を知るのに効果的であり、次回から同じような調査をくりかえすときの能率化にもつながるだろう。

注6) 共通度と類似度とは厳密には異なる概念だと思うが、ここではほとんど区別していない。ただ、共通度は全く同じものについて、類似度は、似ているものについての度合という意味で区別していることもある。

注7) 水谷静夫氏「語イの共通度について」『計量国語学』7号

注8) 安本美典・本多正久両氏「言語系統論のための一つのこころみ」『計量国語学』47号、両氏は、 $\approx$ 相伴指数 $\approx$ と名づけている。

注9) 宮島達夫氏「語いの類似度」『国語学』82。本稿の「用語の共通度」に関する部分は、この文献に負うところが大きい。

## 付表

以下の表は、「層別用語表」ともいうべきもので、以上の分析に使用したものの一部である。話題による区分（T層）の用語表は別に語彙表として出されるが、文種による区分も活用できるものと考え、使用率の充分たかいもののみ、ここに掲載した。標本使用率のたかい順に配列し、いずれの層も0.5パーミルでうちきってある。ここでは、相対的尺度である使用率のみを示しており、標本使用度数、精度などは省略した。（なお、「ニュース展望(09)」は、標本数が小さいため、表を割愛した）

この表の作成には、桜井敏子さんの協力をえた。

順位	見出し語	標本 使用率						
ニュース (01)								
1	、	65.280	40	1	1.744		ても	.644
	の	57.280		つ	1.711		米	.635
	を	34.329		い	1.700		国	"
	に	32.989		な	1.585		発	.633
	・	28.809		だ	1.545		た	.631
	は	25.163		が	1.497		と	.629
	が	22.339		そ	1.494		ろ	"
	て	21.321		の	1.439		な	.624
10	た	19.899	50	れ	1.433	90	ど	"
	と	18.475		て	1.282		の	.622
	で	17.742		一	1.278		調	.613
	〔	10.905		午	1.269		べ	.611
	〕	10.837		後	1.234		な	.607
	い	7.536		東	1.207		ら	.598
	ら	5.992		京	1.205		さ	.589
	「	5.957		ま	1.112		ら	"
	」	5.953		た	1.057		年	.582
	㊦MO	5.149		だ	1.044		へ	.578
20	し	5.085	60	午	1.033	100	の	"
	も	4.882		前	1.026		4	.574
	こ	4.703		と	1.011		な	.571
	と	4.623		の	1.006		く	.560
	あ	4.352		す	1.000		う	.556
	る	4.186		る	.967		ち	.554
	0	4.186		れ	"		そ	"
	こ	3.775		2	.914		れ	.538
	の	3.696		て	.911		よ	.536
	い	3.245		は	.850	110	り	"
	い	2.877		ま	.832		た	.527
	う	2.744		で	.830		い	.518
	た	2.744	70	ま	.821		あ	.516
	め	2.660		み	.801		と	.514
30	で	2.660		か	.797		あ	.510
	は	2.530		か	.788		と	"
	の	2.278		ら	.748		ほ	.527
	な	2.250		あ	.717		か	.518
	=	2.236		っ	.708		結	.516
	に	2.150		対	.684		果	.514
	は	1.965		し	.664		で	.514
	れ	1.910		あ	.662		き	.514
	た	1.910		り	.653		日	.510
	・	1.832		あ	.646		中	"
	一	1.832	80	り				
	こ	1.823		日				
	れ			本				
				に				
				も				
				問				
				題				
				し				
				か				
				し				
				は				
				ば				
				で				
				も				
				だ				
				が				
				や				
				う				
				現				
				在				

順位	見出し語	標本 使用率
ニュース解説 (02)		
1	、	64.209
	の	59.383
	を	35.595
	に	34.089
	・	31.214
	は	25.841

	が	22.931		たが	1.300		う	.718
	て	22.623	50	3	"		だろ	"
	と	20.912		また	"		=	"
10	で	17.079		だ	1.266		自民党	"
	た	15.059		一	"		一	"
	〔	9.583		にも	1.232		結果	.684
	〕	9.480		とも	1.197		たと	"
	いる	8.009		それ	1.163		なければ	"
	ある	7.324		ても	"		いま	.650
	こと	6.503		日本	"	100	中国	"
	し	6.195		なら	1.129		わけ	"
	”	6.023	60	や	"		こう	"
	も	5.510		なる	1.095		とっ	"
20	から	5.168		ば	"		昨年	"
	「	5.133		か	1.061		たい	"
	」	"		あ	"		ての	"
	ない	5.099		っ	"		とは	"
	いう	4.209		考え	1.026		開か	.616
	この	3.559		対する	.958		場	"
	もの	3.559		よる	"	110	よう	"
	では	3.319		ず	"		な	"
	0	3.148		対し	.924		一方	"
	0	2.977	70	問題	"		ように	"
	㊦MO	2.840		でき	"		国民	.581
30	その	2.703		き	.889		中	"
	ため	2.669		でも	"		よ	"
	な	2.361		れ	"		っ	"
	い	2.327		ま	"		点	"
	には	2.293		まで	"		より	"
	は	2.293		へ	"		㊦20	"
	れた	2.224		現在	.885		5	"
	・	2.190		政府	"	120	間	.547
	など	2.053		家庭	"		ソ連	"
	これ	1.985	80	なく	"		約	"
	1	1.882		あり	"		おり	"
40	つい	1.848		のは	"		社会党	"
	する	1.779		ところ	"		必要	"
	み	1.642		米	.821		ある	"
	2	1.608		国	.787		での	"
	しかし	1.574		だ	"		㊦10	"
	し	"		が	"		およ	"
	れる	"		ら	"		び	"
	の	1.471		れる	"	130	などの	"
	て	1.437		で	"		以上	"
	て	1.369	90	きる	"		解	.513
				前	752		説	
				よう	"			
				方	"			
				針	"			



120	まず	〃	30	から	4.810	70	き	.955
	国	.555		この	3.968		どう	.936
	MO	〃		では	3.800		中国	〃
	れ	.538		その	3.144		ところ	.917
	いい	〃		には	2.976		とき	.880
	最近	〃		い	2.882		まで	.861
	のが	〃		もの	2.508		ソ連	.824
	ぬ	.521		だ	2.489		いま	〃
	くる	〃		か	2.283		ように	〃
	対し	〃		一	2.115		との	.805
	とき	.503		これ	2.059		とき	〃
	現在	〃		なっ	1.984		日本	〃
中	〃	ため	1.872	わけ	〃			
一方	〃	なる	1.759	とは	〃			
これら	〃	MO	1.722	ので	〃			
こう	〃	れた	1.703	あり	.786			
順位	見出し語	標本 使用率	40	れる	1.666	考え	.767	
特集記事		(04)		しかし	1.647	...	〃	
10	1	,	66.109	や	1.647	こう	.749	
		の	59.408	する	1.572	昨年	.730	
		を	34.589	つい	1.535	れ	〃	
		を	34.365	ば	1.460	ず	.711	
		・	31.070	ても	1.441	現在	〃	
		には	26.092	など	1.423	場合	〃	
		は	24.426	それ	1.385	一つ	〃	
		が	23.752	問題	1.385	国	.674	
		て	22.498	れて	1.310	う	.674	
		と	22.498	でも	1.310	=	.655	
		で	15.535	にも	1.273	たと	〃	
		た	15.030	なく	1.217	必要	〃	
20		いる	9.658	いっ	1.217	とも	.636	
		”	8.554	ては	1.217	対し	〃	
		「	8.067	たが	1.217	対する	〃	
		」	8.048	また	1.160	政府	〃	
		ない	7.449	み	1.160	よう	〃	
		ある	6.588	なら	1.086	ような	.618	
		〔	6.476	・	1.067	以上	〃	
		し	6.383	0	1.048	への	〃	
		も	6.364	あ	〃	のが	〃	
		〕	6.308	のは	1.029	ことし	〃	
		こと	6.177	だが	1.011	へ	〃	
		いう	6.102	よる	.992	なり	.599	



110	とくに	"	20	らい	6.559	60	たが	1.129			
	国民	"		こと	6.235		私	"			
	米國	"		”	5.901		人	1.080			
	などの	.580		[	5.832		き	1.031			
	強い	"		し	"		とき	1.021			
	人	"		]	5.783		ところ	"			
	点	"		から	4.546		なく	"			
	出	"		その	3.849		まで	.962			
	できる	.562		では	3.535		いま	"			
	より	"		この	3.495		では	.913			
	だから	.543		い	3.417		あり	"			
	おり	.524		もの	3.093		でき	.903			
120	みる	"	30	には	3.063	70	…	.844			
	とっ	"		なつ	2.572		れ	.835			
	だと	"		だ	2.405		つい	.815			
	さらに	"		や	2.150		へ	.805			
	よっ	"		あつ	2.032		ので	"			
	同じ	.505		それ	1.826		問題	.776			
	いわ	"		れて	1.757		だった	.756			
	たい	"		・	1.728		み	"			
	こんど	"		—	1.718		よる	.697			
	はじめ	"		こそ	1.708		東京	.687			
	順位	見出し語		標本 使用率	40		0	1.699	80	いい	"
	特別読物 (05)			れた			"	ような		.677	
1	、	69.317	でも	1.669		中	"				
	の	57.024	か	1.640		ように	"				
	・	37.898	なる	1.600		とは	.668				
	に	31.252	ため	1.581		こう	"				
	を	31.065	いっ	"		どう	.658				
	が	25.400	日本	1.541		わけ	"				
	て	25.331	ても	"		とっ	.648				
	は	24.605	ば	1.483		以上	.628				
	と	21.149	する	1.355		のが	.619				
	た	17.437	だが	1.286		できる	"				
	で	16.573	だが	1.267	間	.609					
	」	8.238	など	1.267	だから	.599					
10	いる	8.080	れる	"	ず	"					
	「	8.041	にも	1.227	への	.589					
	ある	7.864	なら	1.198	アメリカ	"					
	も	7.491	また	"	=	"					
	いう	6.588	しかし	1.188	なり	.579					
			のは	1.168	たと	"					
					100						

110	だと	.569	20	し	6.689	70	ような	1.200	
	より	.560		も	6.467		いつ	1.178	
	現在	"		その	4.889		たが	1.156	
	出	"		から	4.422		中	1.111	
	よつ	.550		[	4.156		ところ	"	
	昨年	.540		]	"		れる	1.089	
	前	"		この	4.067		一	"	
	ここ	"		もの	4.044		つい	1.044	
	多い	"		には	3.467		だが	1.022	
	思う	.530		では	3.156		まで	.978	
	だろう	.520		"	3.044		そして	.933	
	くる	"		30	それ		3.022	問題	"
	いわ	"		い	2.711		こう	.911	
	みる	"		や	2.667		とは	.844	
そう	"	これ	2.289	う	"				
同じ	.511	なっ	2.178	とき	"				
見	"	あっ	2.133	考え	"				
たら	"	なる	2.000	れ	"				
120	場合	"	だ	1.956	一つ	.822			
	とも	"	ば	1.933	あろ	"			
		.491	日本	1.733	など	.800			
順位	見出し語	標本 使用率	40	でも	1.711	いい	"		
評論・論文(06)				。	1.667	だろう	"		
1	、	68.356		する	1.644	思う	.778		
	の	57.644		なく	1.622	いま	"		
	・	36.333		なら	"	そう	"		
	に	33.178		れて	1.600	ず	"		
	を	31.956		にも	1.556	への	.756		
	て	26.511		私	1.489	でき	.733		
	は	24.333		ため	1.467	90	ように	"	
	が	24.289		れた	1.444	との	"		
	と	22.133	50	ても	"	よつ	.733		
10	で	17.000		か	"	知識人	.711		
	た	14.600		ので	"	人間	"		
	ある	10.911		のは	"	ての	.689		
	」	9.400		また	1.422	いえ	"		
	「	9.333		しかし	1.378	なり	"		
	いる	9.022		あり	1.356	より	"		
	こと	8.089		11MO	1.289	とも	"		
	いう	7.844		人	"	100	方	"	
	ない	7.644	60	き	1.222	自分	.667		
				ては	1.200	アメリカ	"		

110	よく	.667	10	に	30.637	50	これ	1.481	
	わけ	.644		て	27.005		よい	1.409	
	み	"		が	24.256		ば	"	
	たら	"		は	22.679		だ	"	
	へ	"		と	18.640		なり	1.386	
	どう	.622		で	14.816		ように	"	
	そこ	"		た	10.491		み	1.362	
	ぬ	"		〔	8.173		れて	"	
	なか	"		〕	8.077		など	1.338	
	…	"		も	7.862		よく	"	
120	点	.600	20	いる	7.217	60	人	1.314	
	話	"		ます	5.902		か	1.290	
	ながら	"		し	5.759		それ	1.218	
	現在	"		「	5.663		また	1.194	
	まず	"		」	5.639		ので	1.147	
	できる	.578		”	5.616		いっ	"	
	だった	"		ある	5.377		中	"	
	作品	"		から	"		たら	1.123	
	たと	.556		いう	5.257		なら	"	
	間	"		こと	5.233		ため	"	
130	見	"	30	もの	4.636	70	場合	1.099	
	とつ	.533		ない	4.492		たり	1.075	
	一	"		や	4.445		にも	1.051	
	だと	"		では	3.154		まで	"	
	以上	"		には	3.106		れる	"	
	さらに	"		その	2.891		のは	1.027	
	あるいは	"		この	2.461		なく	1.003	
	のが	"		い	2.413		多い	"	
	意味	"		なっ	2.342		…	"	
	場合	"		なる	2.246		でき	.955	
140	たい	.511	40	なる	2.222	80	つけ	"	
	くる	"		です	2.198		「10	"	
	たのは	"		・	2.150		方	"	
	国民	"		あり	1.959		できる	.908	
	よう	"		でも	1.744		き	.860	
	順位	見出し語		標本 使用率	とき		1.696	のが	.812
	実用読物 (07)			こ	"		しかし	"	
	1	、		70.189	入れ		1.648	ました	.788
		の		51.476	一		1.625	大きじ	.764
		・		41.224	ても		1.577	たい	"
	を	31.521	する	1.505	つい	"			
			ところ	1.481	だが	"			

			とっ					
			よっ	だから				
90	同じ	.740			.501		たが	2.211
	いい	"			"		だ	2.161
	へ	"			"		白	2.075
	上	"			"		黒	2.038
	1120	.716	順位	見出し語	標本	40	なっ	1.988
	水	.693	探訪ルポ (08)				には	1.964
	あっ	"	1	,	51.550		6	1.717
	加え	"		の	40.988		その	1.680
	かけ	"		0	36.837		…	1.519
	ては	.669		・	33.403		これ	1.470
	ません	.645		を	26.102		れた	1.445
100	おく	"		は	23.211		7	1.383
	よう	"		に	23.075		もの	1.272
	問題	"		が	21.680	50	こ	1.247
	ほか	621		て	18.381		東京	"
	家庭	"	10	]	18.208		巨人	1.222
	前	"		た	18.134		まで	1.210
	東京	.597		[	18.122		だった	1.198
	ような	"		と	17.220		それ	1.185
	日本	"		で	15.330		8	1.173
	より	"		1	13.613		だが	1.136
110	材料	.573		11MO	13.057		打	1.124
	れた	"		2	6.485		ば	1.087
	子	.549		も	5.954	60	ところ	1.050
	いわ	"		]	5.052		へ	1.037
	前後	"	20	[	4.916		あっ	"
	気	"		3	4.867		いっ	"
	出	"		から	4.632		前	.988
	とも	"		一	4.410		三	"
	手	"		いる	4.212		れて	.975
	もう	"		4	4.187		中	"
120	ず	"		"	4.002		9	.963
	わけ	"		ない	3.891		なる	"
	などの	.525		ある	3.570	70	や	.938
	考え	"		し	3.409		あと	"
	値段	"	30	いう	3.236		か	"
	まず	"		・	2.865		中日	.901
	だけ	"		この	2.730		でも	"
	1130	"		こと	2.717		ため	.889
	てから	.501		い	2.655		投	.877
	塩	"		では	2.421		ても	.864
130	母親	"		5	2.248		日本	.852

	なら	.852		手	.506		6	2.413	
80	れ	.840		㊦MO...	"		55	2.324	
	しかし	"		だろう	"		同新	2.174	
	午後	.827		順位	見出し語	標本 使用率	40	ほか	2.170
	左	"		記録・通知(10)				㊦10	2.158
	右	.778		1	㊦MO	65.566		35	2.062
	早大	.765			—	45.553		05	1.989
	にも	.753			0	26.928		㊦20	1.886
	四	.741			〔	21.989		から	1.837
	とき	.716			〕	21.806		7	1.835
	ず	"			,	20.635		8	1.778
90	広島	.704			の	18.873		12	1.673
	安打	"			1	18.086		25	1.511
	点	.691			・	16.565	50	9	1.464
	出	"			「	14.145		=	1.406
	いい	"	10		」	14.030		㊦30	1.296
	大洋	.679			2	9.543		…	1.258
	き	"			・	9.428		A	1.244
	する	.667			㊦NO	8.660		同	"
	また	.654			30	7.141		56	1.181
	なく	.642			00	6.162		曇	1.169
100	安	"			を	5.956		も	1.157
	二	.630			に	5.902		東京	1.150
	一	"			は	5.890	60	いる	1.143
	など	"			3	5.712		晴	1.047
	本塁打	.617	20		15	4.923		"	1.033
	優勝	.605			が	4.766		11	1.012
	ここ	"			10	4.531		㊦映○	.965
	なかっ	"			と	4.510		無	.902
	よく	.580			他	4.295		60	.896
	あげ	"			㊦天○	4.248		カラー	.883
110	ては	.568			4	3.999		午後	.862
	ので	"			て	3.824	70	58	"
	み	"			20	3.728		もちあい	.851
	のは	"			5	3.634		歌	.836
	西鉄	.555	30		40	3.489		65	.783
	でき	.543			45	3.435		70	.780
	打つ	"			で	3.388		スポーツ	.757
	いま	"			は	3.041		ゲスト	"
	たのは	.518			㊦再○	2.926		75	"
	あり	"			50	2.875		54	.733
120	㊦14○	.506						㊦14○	.724

80	53	.724	130	6.00	.562	20	”	6.960
	52	.715		120	.560		も	6.324
	7.00	.710		66	”		から	5.782
	80	”		前	.558		いう	5.517
	38	.703		13	”		ある	5.504
	51	.701		76	.555		し	5.345
	39	.698		48	.548		こと	4.102
	し	.694		46	.546		¶M○	4.035
	後	.691		34	”		ない	3.903
	90	.670		49	.534		この	3.691
90	71	”	105	.532	・	3.559		
	63	”	いう	.527	その	2.951		
	72	.668	140	”	もの	2.792		
	73	”	7:00	”	では	2.686		
	57	”	150	.525	い	2.554		
	きょう	.663	なっ	.522	や	2.435		
	100	.658	64	.520	なっ	2.342		
	ある	.656	43	.518	には	2.091		
	42	.654	32	.511	=	1.971		
	風	.651	95	.508	だ	1.866		
100	雨	.647	140	¶7○	”	一	1.694	
	36	.637		78	.504	など	1.614	
	解説	.626		{	.501	でも	1.548	
	まで	.623		順位 見出し語 標本 使用率		れて	1.522	
	85	.621		紹介記事 (11)		き	”	
	77	.619		1	,	れる	1.508	
	67	.616			の	これ	1.402	
	61	”			・	れた	1.376	
	¶6○	.614			を	それ	”	
	59	”			に	だが	1.336	
110	110	.609		て	日本	1.310		
	18	.600		は	まで	1.297		
	37	.593		が	ため	1.178		
	この	.590		と	あっ	1.164		
	130	”	10	で	…	1.125		
	68	”		た	ば	1.111		
	47	.588		た	たが	1.058		
	こと	.581		」	ほか	1.032		
	41	”		「	にも	1.019		
	ない	.574		〔	人	.979		
!	.569	〕		へ	.966			
120	¶声○	”		いる	なる	”		

60	東京	.953		ことし	.542		では	2.821
	か	"		のが	"		この	2.725
	する	.939		目	.529		0	2.597
	とき	.913		昨年	"		には	2.308
	ところ	.886		より	.516		なっ	2.052
	ても	.847		昭和	"		ました	2.020
	いま	.834		ながら	"		や	1.924
	のは	.820		などを	.503		人	1.892
	なく	"					また	1.795
	作品	.794		順位	見出し語	標本 使用率	あっ	1.731
いっ	"		読者の作文 (12)			それ	1.699	
70	ます	.781	1	,	60.080		なる	"
	また	.767		の	53.475		なく	"
	しかし	.754		・	38.119		か	1.667
	後	"		を	34.624		東京都	"
	夜	.714		に	30.489		あり	"
	れ	.701		て	29.559		でも	1.635
	出	"		が	22.634		する	1.603
	ては	"		は	21.288		ば	"
	いい	"		と	18.402		ても	1.539
	前	.688	10	た	15.645		れて	1.475
80	中	.675		で	15.068		まで	"
	あり	"		も	8.496		思う	1.443
	ように	.662		こと	7.951		ため	1.411
	私	"		いる	7.502		ので	1.379
	です	.648		ある	7.213		だ	1.346
	なら	.635		ない	7.053		れた	"
	見	"		し	6.572		です	"
	なり	"		〔	6.540		にも	"
	つい	"		〕	6.444		なら	"
	アメリカ	.609	20	」	6.027		これ	1.314
90	でき	.582		「	5.995		き	1.250
	よく	.569		から	5.065		いっ	"
	とも	"		・	4.745		一	1.218
	み	"		い	4.617		へ	"
	との	"		いう	4.585		つい	1.186
	だった	.556		私	4.104		たい	1.154
	自分	"		その	4.039		中	1.090
	たい	"		∥MO	3.879		ず	"
	ので	"		”	3.398		見	1.026
	たと	"	30	もの	3.078		思い	"
100	ような	.542		ます	3.046		だが	.994

80	よう	.994	120	学生	.577	20	も	7.253	
	ところ	"		しまっ	"		し	7.218	
	いい	.962		やっ	"		・	6.552	
	東京	"		一つ	"		一	6.342	
	…	"		ながら	"		いる	5.886	
	とき	.930		ません	"		ない	"	
	など	"		感じ	"		ない	5.186	
	時	.898		1	"		」	4.660	
	ような	"		会社員	.545		から	4.555	
	でき	"		多い	"		「	"	
90	ほしい	"	130	よく	"	30	ある	"	
	たが	.866		どう	"		いう	3.819	
	れる	"		母	"		です	3.784	
	考え	.834		前	"		あり	3.539	
	自分	"		のが	"		もの	"	
	なり	"		たり	"		では	3.258	
	のは	.801		との	"		その	3.188	
	れ	"		そう	"		この	2.628	
	方	"		くる	.513		には	2.593	
	ては	.769		たと	"		私	2.382	
100	み	"	10	子供	"	40	人	"	
	気	.705		出	"		か	2.347	
	主婦	"		かけ	"		なっ	2.312	
	問題	"		順位	見出し語		標本 使用率	あなた	2.277
	できる	"		相互通信 (13)			ません	2.242	
	日本	"		1	,		64.054	ました	2.172
	一人	.673			の		49.022	1	1.962
	たら	"			・		38.369	れて	"
	しかし	.641			を		30.100	ため	1.857
	目	"			を		27.647	や	1.822
110	日	"		て	27.051	ば	1.752		
	より	"		に	22.846	これ	1.717		
	場合	"		は	21.445	なり	"		
	家	.609		が	17.310	0	"		
	持つ	"		と	13.946	ても	1.576		
	ように	"	10	で	9.531	また	1.541		
	仕事	"		」	9.285	思い	1.506		
	とも	"		「	8.269	方	"		
	間	"		MO	7.849	なる	"		
	手	"		た	7.498	それ	1.401		
思つ	"		ます	7.428	へ	"			
心	"		こと		でも	"			



60	なら	1.331	110	ください	.805	150	三	.560
	にも	"		どう	"		つけ	"
	いっ	1.261		考え	"		のです	"
	中	1.191		現在	"		お答え	.525
	=	"		ず	.770		時	"
	する	1.156		自分	"		よる	"
	ように	"		親	"		あと	"
	など	"		結婚	.735		投書	"
	のは	"		ですが	"		子	"
	いい	1.121		ぬ	"		とっ	"
70	"	"	出	.700	一	"		
	たら	1.086	よう	"	同じ	"		
	ますが	1.051	できる	"	以上	"		
	まで	"	への	"				
	家	1.016	問題	"	順位	見出し語	標本 使用率	
	たい	"	多い	.665	小説(14)			
	き	"	より	"	1	,	68.078	
	ついで	"	しかし	"		.	45.145	
	では	"	東京都	"		の	42.671	
	2	.981	だと	"		て	33.118	
80	とは	"	ながら	"		に	32.025	
	でき	"	だけ	.630		を	30.816	
	でしょう	.946	3	"		は	29.867	
	彼	"	一	"		た	26.788	
	下さい	.911	必要	"		が	22.501	
	前	"	よく	"	10	と	18.990	
	夫	"	娘	"		」	15.739	
	あっ	.876	ご主人	.595		「	15.566	
	場合	"	だ	"		で	15.106	
	よい	"	れ	"		も	8.833	
90	東京	"	子ども	"		こと	6.848	
	ような	"	手	"		いる	6.215	
	なく	"	ほしい	"		し	6.100	
	とき	"	ノ	"		ない	5.898	
	おり	.840	み	"		い	5.466	
	たり	"	仕事	"	20	ある	5.294	
	れる	"	いわ	.560		いう	4.661	
	ところ	"	他	"		から	4.488	
	気	.805	いえ	"		その	4.200	
	思う	"	主婦	"		〔	3.826	
100	れた	"	今	"		…	3.740	
		"	のが	"		〕	3.711	

30	それ	2.848	70	つい	.949	120	事	.690
	あっ	2.819		たら	.920		ず	"
	か	2.675		れた	"		者	"
	だ	2.618		彼	"		ところ	"
	には	2.560		ぬ	"		ため	"
	この	2.532		言う	.891		だった	"
	なっ	2.301		たと	"		行っ	.661
	もの	2.273		・	"		です	"
	では	2.129		私	"		乃木	"
	へ	1.841		人	"		宇吉	.633
40	来	1.812	80	まだ	"	130	聞い	"
	そう	1.783		思う	.863		言葉	.604
	たが	1.755		ては	"		だろう	"
	言っ	1.697		方	"		おれ	"
	いっ	"		どう	"		しまっ	"
	ます	1.668		のを	"		間	"
	自分	1.640		なかった	"		うと	"
	時	1.582		とは	"		こう	"
	や	"		そんな	.834		え	.575
	一	1.553		まで	"		だから	"
50	なる	1.496	90	なく	"	140	れる	"
	ように	"		ので	"		思い	"
	ても	1.467		なかっ	"		せ	"
	一鬼	1.438		いい	.805		ません	"
	見	1.323		な	"		彼女	"
	する	"		考え	"		眼	.546
	でも	1.266		いま	"		日	"
	？	1.237		あの	"		よく	"
	れて	"		画	.776		藤沢	"
	顔	1.208		くれ	.748		み	"
60	また	"	100	中	"	140	だと	.517
	これ	"		話	"		来る	"
	しかし	1.179		のが	"		女	"
	ば	"		たので	"		ました	"
	なら	1.179		ながら	"		とも	"
	にも	1.150		男	.719			
	のは	1.122		"	"			
	何	1.064		出	"			
	もう	"		なり	"			
	あり	1.007		き	.690			
思っ	"	だが	"					
気	.949	ような	"					
			110					
				順位	見出し語	標本 使用率		
				商業広告 (15)				
				1	の	43.983		
					・	28.842		
					〔	25.173		
					〕	24.878		

10	↑MO	24.220	50	あり	1.251	90	か	.583
	,	22.206		他	1.246		電	.568
	を	19.035		電話	1.207		千代田区	"
	に	16.986		TEL	1.162		なる	.563
	・	16.662		ました	1.157		?	"
	は	10.769		3	1.127		なら	"
	が	10.011		こと	1.122		歴	"
	…	9.313		や	1.117		ほか	"
	で	8.909		には	1.092		できる	.558
	と	8.580		いる	1.077		中	"
20	て	8.361	にも	1.007	する	"		
	/	6.407	など	.982	前	.553		
	ます	6.058	株式会社	.977	新しい	.543		
	一	6.028	もの	"	優遇	"		
	た	5.529	でも	.967	中央区	.538		
	!	4.821	あなた	.937	なっ	.523		
	です	3.769	迄	.907	6	"		
	<	3.605	代	.887	れて	"		
	>	3.555	下さい	.882	なく	.519		
	も	3.495	株	.853	でき	.509		
30	から	3.231	年	"	順位	見出し語	標本 使用率	
	=	2.932	ません	.828	案内広告 (16)			
	一	2.518		.808	1	[	44.381	
	「	2.493	銀座	.793		]	44.171	
	東京	2.453	ため	.788		↑MO	24.067	
	”	2.383	いう	.753		・	23.557	
	まで	2.159	世界	.743		の	18.611	
	ください	1.959	日本	.703		,	9.280	
	より	1.949	5	.693		歩	7.949	
	↑YO	1.880	日	.668		迄	6.954	
40	し	1.875	では	.658	10	一	5.963	
	東京都	1.730	以上	.643		可	5.800	
	1	1.565	新宿	.633		歴	5.529	
	～	1.546	これ	.623		給	5.346	
	へ	1.516	100	"		に	4.895	
	ない	1.511	4	.613		年	4.879	
	この	1.466	有	"		有	4.461	
	2	1.446	上	.608		委託	4.401	
	ある	1.376	れた	.598		一	4.391	
	その	1.321	可	.593		株	4.375	
方	"	だ	.588		優遇	4.361		
い	1.281	発売	"					

20	建	4.231		仲介	1.642		来社	1.041
	上	3.727		私道	1.615		写真	1.038
	を	3.607		委細	1.592		三万	1.031
	～	3.460		年齢	1.575		男女	1.015
	完	3.407		見習	1.572		急募	1.011
	社保	3.393		履歴書	"		日	1.005
	6	3.387		寮	1.565	110	月	.994
	持	3.313		道	1.555		10分	.991
	面接	3.237	70	高卒	1.525		五	.988
	二	3.207		…	1.522		株式会社	.978
30	面	3.173		交通費	1.495		毎日	.968
	は	3.063		昇給	1.482		二回	.941
	住	3.060		経験者	1.475		総	.931
	舎	"		給与	1.465		1	.924
	以上	3.040		四	1.462		経	.918
	面談	2.900		と	"		位	"
	不問	2.873		台	"	120	30	.908
	売	2.836		て	1.462		事務員	.904
	45	2.743	80	代	1.442		中央区	"
	通	2.723		国電	1.422		代表	.891
40	持参	2.613		女子	1.408		電話	.878
	方	2.516		資格	1.362		{	"
	KK	2.489		新宿	1.308		午後	.874
	より	2.446		2	1.295		千代田区	.851
	ます	2.316		男子	1.285		池袋	.841
	・	2.302		二万	1.258		4	.834
	及	2.262		環	1.248	130	あり	.828
	三	2.179		和	1.245		休	.821
	募集	2.139	90	付	1.228		まで	.818
	経験	2.122		六	1.215		面接日	.814
50	残	1.985		住込	1.211		細	.804
	頭	1.972		当社	1.178		位迄	"
	他	1.965		も	1.175		銀座	.794
	3	1.952		社員	1.131		10時	"
	浴	"		電	"		363	.788
	支給	1.946		東京都	1.121		25歳	.784
	待遇	1.902		秀	1.115	140	港区	.768
	完備	1.872		下車	1.108		職種	.761
	で	1.805	100	勤務	1.101		下さい	.758
	高給	1.789		車	1.065		送	"
60	前	1.755		が	1.061		通勤	.751
	賞与	1.682		五年	1.051		25	.748

	保	.748		10名	.594
	!	"		七	.591
	5分	"	190	984	"
	神田	.744		ある	"
150	営業	.721		三年	.587
	新築	.718		20	"
	週休	.714		交費	.571
	日給	"		広	.567
	8	.711		運転手	"
	五分	.707		乞	"
	店	"		事務	.564
	又	.704		5名	"
	日本橋	.697	200	徒歩	"
	数名	.694		た	"
160	です	"		歓迎	.561
	8分	.691		全額	"
	敷	"		初給	.551
	要	.684		応募	"
	規定	.677		地	.547
	"	.674		社	.541
	八	"		新宿区	"
	渋谷	.671		後	.537
	南	.664	210	ホカ	"
	陽	"		自由	.534
170	5	.661		貸与	"
	東京	.657		387	.527
	30歳	"		年令	.514
	共	.651		借	.511
	午前	.647		築	"
	洋	"		従業員	.507
	宅地	"		急	"
	即決	.634		バス	.504
	即	.627	220	月収	.501
	学歴	.624		手当	"
180	へ	"			
	郵送	.621			
	通知	.617			
	本社	"			
	庭	"			
	賃	.607			
	10	"			
	若干名	.594			