新聞語彙調査の概略と語彙分析法試案

新聞語彙調査の概略と語彙分析法試案

林　四郎

新聞語彙調査の概略

電子計算機による語彙調査　国立国語研究所は，開所以来，一つの業務として書きことば資料による語彙調査を行なってきた。昭和24年度には，手始めとして1か月の新聞の小規模な調査を行ない，以後，婦人雑誌（昭和25年）総合雑誌（昭和29年），雑誌90種（昭和31年）と，一貫して推計学の方法により雑誌の用語を調査してきた。そのうち規模の最大なものは，雑誌90種で，サンプル延べ語数約53万語であった。しかし，現代語の基本語彙に関し，より有効な情報を得るためには，さらに大規模な語彙調査を比較的短期間内に仕上げる必要がある。それには，従来の人力とカードによる方法は，すでに限界に来たと思われるので，ここで，語彙調査の方法を電子計算機による機械処理にきりかえることにした。

漢字テレタイプによる入出力　電子計算機で語彙調査の作業を能率化するといっても，漢字を表記の主力とする日本語の現状においては，ローマ字だけで表記する欧米語とはちがって，電子計算機へ入力するまでの処理，および電子計算機からの出力の方法に大きな困難がある。かなあるいはローマ字を用いて入出力し，漢字を何らかの方法で他の記号に変換して扱う方法も考えられ，現にＮＨＫの放送文化研究所では，漢字を数字に変換して，話しことばの語彙調査を行なっているが，国語研究所の語彙調査では，漢字の調査に重点を置く必要があるので，直接漢字を入力する方法をまず考えなければならなかった。そこで，入出力には，漢字テレタイプを用いることにした。

漢テレ（以後「漢字テレタイプ」を略してこういう）は印刷電信機（テレプリンタ）の字母に漢字を入れたもので，現在新聞社が活字製版用及び遠隔地送信用に用いている機械である。漢テレは，文字を2進コードに変えて紙テープにせん孔するので，このテープは電子計算機への入力データにするこ

とができる。この点に着眼して，語彙調査の入力データ作成には漢テレを用いることにした。

電子計算機及び漢テレの設置　昭和40年度から，国語研究所へ電子計算機の借用と漢テレ購入の費用が認められたので，機種検討の結果，電子計算機は日立製作所の HITAC 3010 を使うことにし，漢テレは沖電気工業KKの機械を用いることにした。機種決定と機械設置までの経過は，国立国語研究所年報16，17にしるしてある。

漢テレは，新聞社で使っているものも，各社によって仕様がちがっており，統一された標準スタイルはない。特に電子計算機への入力データを作るためには，その機種との特別な関係があるので，国研のものは，HITAC 3010に合わせて，独自の仕様により設計した。その詳細は本書に松本が報告している。

調査対象　電子計算機による大量語彙調査の手始めとして，調査対象には新聞を選んだ。(新聞を選んだ理由は年報17にしるした。)母集団は昭和41年1月から12月31日までの，朝日，毎日，読売3紙の朝夕刊全紙面に含まれる文字及び記号である。標本の抽出は抽出比を1/60とし，面積を基準とするブロックサンプリングの方法で行なった。その詳細は，本書に田中と斎藤が報告している。

語彙調査の調査単位　分かち書きの行なわれていない文章を計算機に読ませて語をカウントするためには，あらかじめ人間が語の切れ目にしるしを入れておかなければならぬ。語の認め方には，いろいろな単位がありうる。研究所が今日まで行なってきた雑誌の語彙調査には2種類の単位が用いられてきた。一つは，α単位と称するもので，文節を基準とし，1文節内の自立語の部分を1語とするものである。いま一つはβ単位と称するもので，αをさらに細かく分解し，原則として二つの語源単位の結合をもって1語とするものである。

今回の新聞の調査では，まずα単位によってインプット・データを作り，アウトプットされたα単位語彙表の見出し語を，さらにβ単位に切るという，二段構えの作業をすることにした。はじめから単位を細かく切ることは，労が多く，しかも処理に不一致を生じやすいと考えたからである。α単

位の規定は『婦人雑誌の用語』19ページ以下にしるされているが，われわれは，この規定をそのまま用いることはせず，機械処理に適すように新たに規定を作った。したがって，α単位と完全に一致するわけではないので，以後われわれの単位を長単位という。

標本の大きさ　新聞を朝刊16ページ，夕刊8ページ，合計24ページとすると，1日の新聞紙面全体に含まれている文節の数は約8万になる。8万 × 3紙 × 360日 = 8,640万の計算で，3紙1年間の全紙面に含まれる文節数は8,600万余。その1/60の144万文節がわれわれの調査処理の対象となる標本である。

今回の調査では，自立語だけを対象とするのではなく，文節内の付属語もかぞえられるので，今回の調査処理の対象は，144万の長単位自立語に相当数の長単位付属語及び記号類が加わり，長単位として200万を越すことになると思われる。

語の所属層の分類　新聞記事は話題からいっても，文章の種類からいっても，多様である。調査の結果拾われた語が，どういう性格の集団に属しているかがわかることが必要なので，各語の属する文章を層に分けた。4つの角度から層別をしたが，そのうち三つは記事単位の区別であり，一つは，1記事内の位置上の区別である。

1)　文章の種類による区分(記事単位の分類)　G種層別

1. ニュース(ニュース価値のある事件を速報するもの)
2. ニュースの解説(独立した解説記事のほか，ニュース記事のあとに付属した「解説」や「注」も。)
3. 社説・コラム(ニュースの影響下にある論説，評論。寸描的描き方のものも含む)
4. ニュースに連なる特集記事(取材角度や編集法に特徴がある)
5. 特別読みもの(直接にはニュースと関係がないが，物を見る目に現時的問題意識を含んでいる場合が多い。"動物紳士録"，"東京むかしむかし"，"新人国記"等のようなもの)
6. 評論・論文(多く学芸欄にのるもの，時評的性格がある)
7. 実用知識読みもの(家庭欄に多い，ハウ・トゥーものなど)

8. 探訪ルポ(見て来たままを報ずる記事。インタビューなどによる。スポーツ欄に多い)
9. 長期ニュース展望(週間，月間，年間などのニュースをまとめたもの)
10. 記録・通知(株式，ラジオ・テレビ番組，催しもののお知らせなど，要件だけを簡潔にしるしたもの。天気予報，番組案内の文章も含む)
11. 紹介記事("時の人"のような特定人物の紹介。海外雑誌新聞の論調紹介など)
12. 読者の作文(投書式意見文や"ひととき"式随想作文など)
13. コミュニケーション(身上相談式な読者と社側との通信。読者の広場式な読者同士の通信)
14. 小説
15. 商業広告(商品や事業の広告)
16. 案内広告(いわゆる三行広告。大学入学案内・死亡通知などをふくむ)
17. 漫画

2) 話題による区分(記事単位の分類)　T種層別

1. 政治(国内政治)
2. 外交(日本の外交)
3. 経済(株式が主体，経済問題でも，第1面にあるものは概ね政治)
4. 労働
5. 社会(社会面にのるもので，内容は雑多)
6. 国際(第2面外電が主体，1面にも多い)
7. 文化(学芸，文化欄)
8. 地方(地方版)
9. スポーツ(碁・将棋を含む)
10. 婦人・家庭
11. 芸能・娯楽(ラジオ・テレビ番組，その解説)
12. 広告(Gの14から17までを合併したもの)

3) 署名態度による区分(記事単位の分類)　S種層別

1. 無署名記事(一般の記事)

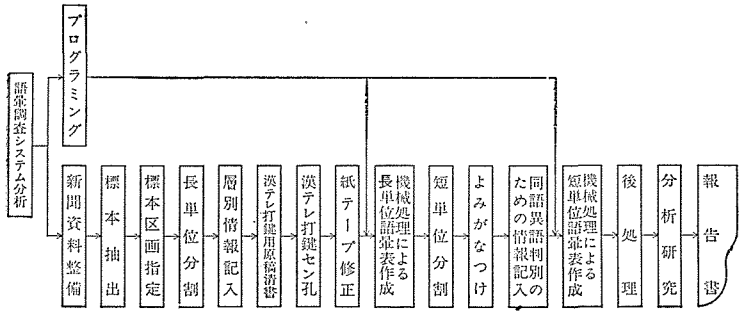
2. 通信社記事
3. 冒頭署名記事
4. 末尾署名記事1(外部者)
5. 末尾署名記事2(記者)
6. 末尾署名記事3(略称による。"Q" など)
7. 外電冒頭記名記事(「ニューヨーク十九日=小野寺正特派員」の如きもの)
8. 無署名だが社を代表する立場にある筆者による記事(社説,天声人語など)
9. 無署名で外部者たることが明らかなもの(天気予報など)
10. T 12に同じ

4) 紙面上の位置による区分(1記事内の部分単位の分類)　P種層別

1. 見出し
2. 標題・欄名
3. リード
4. 本文
5. 情報源・署名
6. 表
7. 写真や図・表などの説明
8. T 12に同じ

機械処理プログラム　漢テレで作ったデータを電子計算機が処理して語彙表を作るためのプログラムについては,石綿の論文が,流れの概略をしるしている。このプログラムを作るに先立って,HITAC 3010 による,言語情報処理の,当研究所における最初の試みとして,漢テレによる用語総索引作成プログラムを作り,一部の文学作品や新聞記事を資料にして,作品の用例つき用語総索引を作った。本書,斎藤の論文は,これの報告である。木村,田中の論文は,それぞれ,機械処理プログラムの一部分の報告である。

作業の流れ　電子計算機による新聞語彙調査の作業は,次のような手順になっている。



新聞語彙分析法の研究

語のよく使われる度合の測りかた　語彙調査の大きな目的の一つは，基本語彙を求めることである。基本語彙の求め方には，いろいろな考え方があるが，大きな方向として，Basic English のように，意味の分析によって基本的意味を表わす語を求め，それらの組み合わせで表わせる事象の範囲から基本語の範囲を定めていく方法と，現実の発話や文章の用語を調べて，よく使われる語を統計的に求める方法と，二つのいきかたが考えられる。われわれが当面用いようとする方法は第二のものである。

語の使われる度合を統計的に求める方法は第1に使用度数をかぞえ，使用率を計算することであり，第2に語の所属層を単位にして出現の範囲をかぞえ，一定の基準によって幅の広さの指数を出すことである。そのほかにもありうるが，まず，この二つをたよりにする。

今回の新聞の調査では，語の所属層について，前述のように，G．T．S．Pと4種の層別を行なった。このうち，GとT，特にT(話題)は，語の出現の幅を測るときの尺度として，おもに利用しようとするものである。

出現の幅を重視するにともなって，使用度数の方も全体をいっしょにした総度数のほかに，各層内での使用度数を重視する考え方に傾く。各層内での使用度の高さを「深さ」と称することにする。

以下，語の使われる度合を測るための基本的な考え方は，使用度の深さと出現する層の幅の広さとを二つの観察次元としていくことである。すると，

下表のように，

出現幅＼使用度	深　い	浅　い
広　い	広くて深い	広くて浅い
狭　い	狭くて深い	狭くて浅い

1) 広くて深い語
2) 広くて浅い語
3) 狭くて深い語
4) 狭くて浅い語

の4群の語彙が得られることになる。広くて深い語は，文句なしに「よく使われる」と見てよい。反対に，狭くて浅い語は，「あまり使われない」語であろう。その中間にある二つの群が問題である。これらは，それぞれ「よく使われる」といえると思うが，その意味がちがうのである。広くて浅い語がもしあるなら，それは，その言語（われわれが課題にしているのは日本語）の基本的構造にかかわる語彙であろう。狭くて深い語は，特定層内での「よく使われる」語である。それがどんな層か，国民の言語生活に大事な意味をもつ層であるか否かによって，標準的に「よく使われる」といえるかどうかがきまる。

語彙調査の結果，得られた語を，このように，使われかたの性格によって群分けをしてみよう。扱うデータは，新聞語彙調査のサンプルに指定されたものの一部で，昭和41年12月に，それまでにできた入力データ長単位2万余語をインプットし，テスト的にアウトプットしたものである。

第1次語彙表の制約条件　長単位で切られた入力データを計算機で処理した結果，見出し語として立つものの中には，ふつうの語彙表の見出し語には出て来ない語形のものがある。その一つは，活用のある語の活用した形で終止形以外のものである。例えば「行った」という句からは，「行っ」と「た」とが拾われ「行っ」は「行く」とは別の見出し語に立てられる。もうひとつ異様なのは，付属語の連続である。「行きましょう」という句からは「行き」と「ましょう」が拾われ，「ましょう」は「ましょう」のままで見出し語に立つ。このような条件のもとで作られるのが，今回の作業において，計算機と漢テレが打ち出して来る第1次語彙表である。

こういう不便が除かれて，計算機の中で語形の整理がなされるようになるのは，計算機用言語学(Computational Linguistics)がずっと進み，その上に

立って自動処理プログラムが組まれるようになってからで，それは，まだまだ将来のことである。目下のところでは，pre-edit のときに見出し語形を記入しておくか，第1次アウトプットに中間処理の手を加えて再入力するかのどちらかしか方法がない。われわれは後者の方法をえらんでいるが，今ここにしるす語彙分析の試行は，第1次アウトプットに何の手も加えない，いわば変態的な語彙表を材料として行なったものであることをことわっておく。

第一次語彙表のアウトプット　12月テストランで作った語彙表は，次のような種類のものから成り立つ。

① 出典語彙表(*簡略五十音順に配列された見出し語のもとに，各語の出典ナンバーがしるされ，原文にもどれるようにしてあるもの。見出し語だけは漢テレで印字してある)

* 簡略五十音順については，田中の論文を参照されたい。以下これを配列順という。

② 度数順語彙表（総度数と順位とをしるし，度数順に見出し語だけを配列したもの。見出し語は漢テレで印字する）

③ 配列順T層別語彙表（簡略五十音順に配列された見出し語のもとに，総度数と，T_1～T_{12}各層内の度数とをしるしたもの。見出し語の印字は漢テレによる）

④ 配列順G層別語彙表（③と同形式でG_1～G_{17}までの各度数をしるしたもの）

以上4種の語彙表のうち，②③④を用いて語彙分析の試行を行なった。上位にどのような語があるかを一覧するために②の度数順語彙表の一部（度数5まで364語）を以下(表1)に示す。

表　1

度　数　順　語　彙　表

順位	度数	語
1	1,036	、
2	968	の
3	512	。
4	480	を
5	448	に
6	361	は
7	343	¶M○
8	325	が
9	304	て
10	296	と
11	290	「
12	280	」
13	266	た
14	264	・
15	234	で
16	208	—
17	159	『

18	155	」
19	104	0
20	97	も
21	89	ない
22	82	ニ
23	81	いる
24	68	ある
25	68	いう
26	65	から
27	65	こと
28	59	し
29	49	い
30	49	1
31	44	この
32	41	では
33	41	もの
34	40	ます
35	37	=
36	35	だ
37	35	など
38	34	なっ
39	33	その
40	33	…
41	31	他
42	28	より
43	27	には
44	27	30
45	26	これ
46	26	東京
47	25	れた
48	24	一
		いっ
		5
		2
		10
53	23	ため
		にも
		¶M○1
56	22	前
		へ
		や
59	21	か
		ところ
		まで
62	20	あり
		女
		¶M○2
65	19	でも
		なる
		20
		¶N○
69	18	あゝ
		歩
		7
		15
		3月
		00
75	17	後
		男
		東京都
		迄
		6
		3
		!
		¶10
		¶20
84	16	それ
		でき
		です
		み
		4
		～
		{
91	15	しかし
		人
		なり
		＜
95	14	上
		する
		普通
		2月
		＞
100	13	自分
		私
		政府
		たが
		日本
		ば
		方
		また
		れて
		30分
		¶M○3
111	12	歌
		午後
		考え
		受け
		だが
		中
		とっ
		ので
		への
		ました
		夜
		れる
		45
		25
		¶天○
126	11	商業
		対し
		ても
		とき
		ながら
		なく
		なら
		年
		よう
		55
		40
		／
		¶M○5
140	10	一部
		う

度数順層別語彙表の作成　広さと深さの関係を求める足がかりはT，Gの層別であるので，層別の度数順語彙表が必要であった。ところが分析法の考えがあとから追ったために，当初予定したプログラムには，層別語彙表は配列順だけが考えられており，度数順がなかった。そこで②と③④との照合によって度数順の層別語彙表を作った。②の度数順語彙表をもとにし，その語の配列順における位置をさがし求めて，層別度数を転記した。この配列順は，漢字については，頭1字の代表音訓で引くものだから，これを求める作業は，国語辞書を引くように簡単ではなかった。層別語彙表は，配列順よりも度数順の方がまず必要であることがわかったので，早速，度数順層別語彙表を作るプログラムを追加したが，そのオペレートは，今回の報告には間に合わなかった。よって，層別語彙表の配列順から度数順への作りかえは，すべて人為作業で行なった。サンプルとしてT層別度数順語彙表の上位52位まで(総度数24まで)を示したのが，表2である。

表　2

	順位	総度数	1 政治	2 外交	3 経済	4 労働	5 社会	6 国際	7 文化	8 地方	9 スポーツ	10 婦・家	11 芸・娯	12 広告
、	1	1,036	116	15	111	6	274	48	118		64	97	41	146
の	2	968	122	18	60	10	207	61	118		49	59	39	225
。	3	512	68	1	45	2	150	17	71		41	56	2	59
を	4	480	74	11	41	6	111	25	48		33	51	2	78
に	5	448	60	9	31	3	102	24	63		22	44	3	87
は	6	361	52	3	32	1	92	19	37		26	30	3	66
¶M○	7	343			83		7		4		95		32	122
が	8	325	45		30	1	116	13	38		25	30		27
て	9	304	40	5	22	1	91	18	41		18	32	2	34
と	10	296	40	9	18	3	66	20	39		17	22	5	57
〔	11	290	16	2	8	1	31	8	6		36	3	13	166
〕	12	280	16	2	8	1	31	8	5		36	3	13	157
た	13	266	35	4	26	3	96	13	27		17	21		24
・	14	264	2		7	1	9		3		6	3	56	177
で	15	235	35	2	18	1	51	14	28		13	32	2	38
—	16	208	14		144		11		11		11	1		16
「	17	159	17	1	2	1	41	14	18		1	6	43	15
」	18	155	19	1	2	1	37	14	19		1	6	43	12
0	19	104			54						6		44	
も	20	97	13	3	11		38		14		2	8		8

ない	21	89	18	2	10		35	3	6		6	6		3
〃	22	82	10	6	6		18	2	2	13	2	2		21
いる	23	81	17		7	1	24	9	18		2	1	1	1
ある	24	68	17		4		13	4	5		5	12		8
いう	25	68	13	3	7		14	5	13		2	6		5
から	26	65	14		8		20	3	10		4	1		5
こと	27	65	12	2	6		12	6	8		2	7		10
し	28	59	15	1	3	1	14	3	7		3	6		6
い	29	49	6	1	5		24	5	2		1	2		3
1	30	49			26						9		1	13
この	31	44	14		3		9	3	4		4	2		5
では	32	41	3	1	4		16	3	4		2	5		3
もの	33	41	8		6		10	2	5		2	3		5
ます	34	40					4		2			8		26
=	35	37	2	2	3		2	5			10	2		11
だ	36	35	8	1	3		10	1	4			4	1	3
など	37	35	9	2	8		7	1	4		1		1	2
なっ	38	34	4		4		15	1	3		2	2		3
その	39	33	5		1		6	2	3		2	6		8
…	40	33					1					2		30
他	41	31	1				1						27	2
より	42	28	5		2		1		1		1			18
には	43	27	3		2		10	1	7		1	1		2
30	44	27											21	6
これ	45	26	6		2		5	4	4		2	1		2
東京	46	26	1			1	7		2		4	1		10
れた	47	25	1		1		10	3	2		1	2		5
一	48	24	3				3					1		17
いっ	49	24	2		4		11	1	5		1			
5	50	24			8						2		3	11
2	51	24			12						1		2	9
10	52	24			5								15	4

各層内での度数分布調べと段階分け　表2を横に見れば，各語が各層にどのように分布しているかがわかる。例えば，最高度数の「、」(読点)は全体で1036あり，その内訳は，政治記事に116，外交記事に15…… そして広告に146となっている。次にこの表をたてに見ると，各層の中で，何という語が何回出現したかがわかる。例えば，政治記事の中では，「、」が116，「の」が122，「。」(句点)が68……という具合である。政治から広告まで各層を上

から下へたどってみると，大体において，下へ行くほど数がへっているけれども，細かく見れば，必ずしも順番どおりではない。すなわち，各語は，層によって，出現のしかたにちがいがあるわけである。各語が，各層の中で出現している度数が多い方か少ない方か，どの辺の位置を占めているかをつかみやすくするために，各層内での語の度数分布を調べ，便宜上5段階に区分した。その結果は，例えば政治記事でいえば表3のようになる。

度数1が他と比較にならなく多いが，これは長単位の切り方のために，数字の組み合わせなどがやたらにたくさん出て来るためである。これからの考察をするのに，そのような無意味な数字列などを相手にしてもしかたがないので，以下の分析では，総度数1の語を対象外とする。表3の政治記事内での度数1の685語のうち，総度数1の語が518あるので，これを除くと残りは167となる。このようにして総度数1の語を除外し，総度数2以上の語について，上記の処理をした結果，TとGの層別において，表4.1，4.2のような段階区分を得た。

表 3　政治記事内での度数分布と段階分け

度数	語数	段階
122	1	5
116	1	
74	1	4
68	1	
60	1	
52	1	
45	1	
40	2	
35	2	
19	1	
18	1	
17	3	
16	2	
15	1	3
14	3	
13	2	
12	2	
10	1	
9	1	
8	2	
7	1	
6	6	2
5	11	
4	12	
3	32	
2	102	1
1	685	

表 4.1

	T1		T2		T3		T4		T5		T6		T7		T8		T9		T10		T11		T12	
段階	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数
5	122 116	2	18 15	2	144 111	2	10	6	274 150	3	61 48	2	118	2	13 12	2	95 64	2	97	1	56	1	225 122	6
4	74 35	9	11 9	3	83 20	13	6	2	116 35	12	25 17	6	71 37	7	11	2	49 33	5	59 44	4	44 32	6	87 21	12
3	19 7	20	6 4	3	18 6	24	3	3	31 7	30	14 4	18	28 6	15	10 3	6	26 9	13	32 12	7	27 8	16	18 9	27
2	6 3	61	3 2	14	5 3	50	2	4	6 3	104	3 2	49	5 3	42	2	8	8 3	51	8 3	42	7 3	25	8 3	211
1	2 1	269	1	52	2 1	295	1	26	2 1	477	1	106	2 1	268	1	19	2 1	241	2 1	228	2 1	97	2 1	569
計		361		74		384		36		626		181		334		37		312		282		145		825

表 4.2

	G1		G2		G3		G4		G5		G6		G7		G8		G9		G10		G11		G12		G13		G14		G15		G16	
段階	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数	度数	語数
5	375 353	2	26	1	8 7	2	38 37	2	102	1	54 47	2	43 33	2	34 30	3	16 7	4	197 98	4	14	1	51 30	4	45 26	3	9	2	169	1	124 90	4
4	191 101	8	16 12	6	5 4	2	23 17	6	59 7	16	32 6	15	23 20	3	19 5	10	5 4	5	66 15	20	11 10	2	26 7	9	21 6	11	6 4	6	82 22	15	55 8	20
3	76 10	30	9 3	12	3	2	11 3	18	6 3	27	5 3	23	13 3	13	4 3	16	3	5	13 7	35	8 5	7	6 3	33	5 3	17	3	4	18 4	43	7 5	35

2		$\frac{9}{3}$	167	2	14	2	11	2	36	2	52	2	51	2	31	2	30	2	15	$\frac{6}{3}$	85	$\frac{4}{2}$	14	2	36	2	32	2	7	$\frac{3}{2}$	178	$\frac{4}{3}$	81
1		$\frac{2}{1}$	574	1	79	1	23	1	106	1	145	1	102	1	103	1	97	1	30	$\frac{2}{1}$	381	1	47	1	121	1	115	1	27	1	285	$\frac{2}{1}$	333
計			781		112		40		168		241		193		152		156		59		525		71		203		178		46		522		473

この段階区分に従って各層内で度数順に配列した語彙表を示すと，次のようになる。（Tについて，段階2までを示す。）

表 5　話題(T)層別による各層頻度段階別，度数順語彙表(段階2まで)

段階	T1 政治		T2 外交		T3 経済		T4 労働		T5 社会		T6 国際		T7 文化		T8 地方		T9 スポーツ		T10 婦・家		T11 芸・娯		T12 広告	
5	の	122	の	18	—	144	の	10	、	274	の	61	、	118	、、	13	¶M○	95	、	97	・	56	の	225
	、	116	、	15	、	111			の	207	、	48	の	118	普通	13	、	64					・	177
									。	150													〔	166
																							〕	157
																							、	146
																							¶M○	122
4	を	74	を	11	¶M○	83	、	6	が	116	を	25	。	71	女	11	の	49	の	59	0	44	に	87
	・	68	に	9	の	60	を	6	をに	111	に	24	に	63	〔	11	。	41	。	56	「	43	を	78
	には	60	と	9	0	54			に	102	と	20	を	48			〔	36	を	51	」	43	は	66
	はが	52			。	45			たは	96	は	19	て	41			〕	36	に	44	、	41	。	59
	がて	45			を	41			は	92	て	18	と	39			を	33			の	39	と	57
	てと	40			はに	32			て	91	。	17	が	38							¶M○	32	で	38
	とた	40			にが	31			と	66			は	37									て	34
	た	35			が	30			で	51													…	30

で 35
」 19
ない 18
「 17
いる 17
ある 17
〔 16
〕 16
しから 15
この 14
— 14
も 13
いう 13
こと 12
ました 12
ニ 10
など 9
の 8
も 8

3

"
て
た

た 26
1 26
¶M○1 23
て 22
¶M○2 20
と 18
で 18
¶M○3 13
2 12
も 11
¶M○5 11
ない 10
同新 10
4 9
〔 8
〕 8
から 8
など 8
5 8
・ 7
いる 7
いう 7
145 7

に
と
た

「 41
も 38
」 37
ない 35
〕 31
〔 31
いる 24
い 24
から 20
" 18
では 16
なつ 15
いう 14
し 14
ある 13
こと 12
— 11
いつ 11
もの 10
だ 10
には 10
れた 10

で 14
「 14
」 14
が 13
た 13
いる 9
〔 8
〕 8
インド 7
こと 6
いう 5
い 5
＝ 5
ある 4
これ 4
ところ 4
行わ 4
米国 4

で 28
た 27
「 19
」 18
いる 18
も 14
いう 13
— 11
から 10
こと 8
し 7
には 7
〔 6
ない 6
や 6

商業 10
男 7
100 6
150 4
50 3
250 3
＝ 10
¶10 10
1 9
6 9
¶20 9

は 26
が 25
に 22
て 18
と 17
た 17
で 13
—— 11
＝ 10
¶10 10
1 9
6 9
¶20 9

て 32
で 32
は 30
が 30
と 22
た 21
ある 12

他 27
30 21
¶N○ 19
10 15
20 15
〔 13
〕 13
00 12
¶天○ 12
歌 10
05 10
45 9
25 8
40 8
35 8

より 18
3月 18
— 17
迄 17
— 16
～ 16
「 15
東京都 15
！ 15
＜ 14
2月 14
1 13
まで 13
＞ 13
」 12
＝ 11
5 11
歩 11
／ 11

が 27
ます 26
た 24
" 21

7 7 7 7 7 7 7 7 7 7 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
4 4 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
7 3 方 写真 共 こんな 女子 完備 経験 履歴書 し 30 へ 00 年齢 車 宅地 電話 委託 応募 願書受付 高卒 山林 社員

ーク ュッ ミジ
9 ミュージには5 50 8 映画 英語 ポー スツ ニュー ス 化学 海外 アワー パレード
5 5 5 5 5 4 4 4 4 4 4 4 4 4 3 3 3 3 3

5 5 5 5 5 4 4 4 4 4 4 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
では ありつ あつと 結婚でした だへか 女それ 受け 中音楽 すぐころ 材料 竹馬 テープ [ ] ・ ものところ

6 第二日 単複 1時からこの東京ほか 三〇大 早塩原 石幡 八幡 〇 1分し 7 3 なりがたのは 11時 四

5 5 5 5 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
交渉 当地 米政府 対し ものの そのめ たれそ す午後 夜 うのは よるれて られた八 示発 して とは 北爆停止 満足

3 3 3 3 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
この では だ など これ がて れて でのな らつい たのは 窓口 好 三 小包 ・つ なそのか ところ 日本 中 点 あと きょう

4 4 4 4 4 4 4 4 4 4 4 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
へしかし 人 なりて れ受付が だ夜れいす られ若まや ところ とでもるつす なあで分いまり 自ついおき わい 行た

50 大蔵省 日興 買い 48 大和 38 140 126 90 ¶M○ 10 しこの =だ 7 府 政 でのながら 100 国債 新しい 動き

2 2

ネ協 ジュブ ー定 責任 設備
5 5 5 5 5 5 4 4 4 4 4 4 4 4 4 4 4 3 3 3 3
問題 あなた 行監委 小麦粉 内閣 平年分 なつ ところ それ でき { よう 方針 社会党 機能強化 江田 昭和41年分 秘書団 では には 一 にも

4 4 4 4 4 4 4 4 4 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3

新宿 6
送料 6
当社 6
郵送 6
有 6
ＫＫ 6
100円 6
200 6
いう 5
からの 5
このの 5
もれた 5
でも 5
男 5
4 5
日本 5
へのる 5
れる割 5
七口 5
西さい 5
下車 5
下可 5

後 3
なり 3
自分 3
ながら 3
時 3
ません 3
限り 3

10時 3
全日本 3
一〇〇 3
一一〇 3
互い 3
山本 3
四五 3
話し合 3
松田 3
選手権 3
戦 3
端歩 3
塚田 3
適 3
ナシ 3
200 3

く 3
ように 3
やっと 3
エンジ 3
2 著者 3
順番 3
切手 3

近 2
ミ 2
シ 2
ル 2
カー 2
題 2
関 2
する 2
共 2
同 2
言 2
討し 2
検 2
サイゴ 2
ン 2
四条件 2
出発 2
成果 2
それほ 2
そど 2
タイ 2
日本時 2
間 2
ハノイ 2
反応 2
唯一ニ 2
ワ 2

4 もう 3
4 事件 3
4 二人 3
4 三丁目 3
4 四人 3
4 十二月 3
4 女性 3
4 やつ 3
4 恵美子 3
4 さん 3
4 私大 3
4 少年 3
4 審議 3
4 みとめ 3
4 Y子 3
3 一 3
3 かが 3
3 たば 3
3 方 3
3 ま た 3
3 考えるき 3
3 れとき 3

70 3
62 3
43 3
および 3
三井山 3
修正 3
トヨタ 3
ナイロ 3
ン 3
売り 3
免税 3
51 3
60 3
66 3
110 3
164 3
137 3

3 へ 3
3 やまで 3
3 でもば 3
3 へのと 3
3 きまい 3
3 おり 3
3 場合 3
3 ては 3
3 こんど 3
3 対する 3
3 はずら 3
3 知認め 3
3 中小企 3
3 業約 3
3 協 3
3 僚 3
3 閣会 3
3 議行 3
3 現で 3
3 実行る 3
3 き十七日 3

試験日	5
主任教授	5
株式会社	5
渋谷区	5
十	5
地下鉄	5
入学案内	5
→	5
10	4
ため	4
にも	4
かま	4
した	4
8	4
以上	4
家庭	4
今年	4
社	4
社会	4
そう	4
建築	4

くなよう	3
なよう	3
のは	3
点	3
れい	3
いい	3
見	3
次	3
大きな	3
とも	3
二	3
られて	3
ノ	3
各	3
住所	3
のでは	3
まだ	3
あまり	3
意見	3
いわ	3
計	3
生れ	3
多く	3

政務	3
総評	3
二日	3
本気	3
EEC	3
4293	3

出来る	4
坪	4
院長	4
資格	4
神田	4
推薦	4
セールスマン	4
選択	4
専攻科	4
待遇	4
短期大学	4
追っ	4
通知す	4
定価	4
訴	4
オカ	4
マカオの竜	4
無指定	4
試験	4
利息	4
TEL	4
1日	4

新日社	3
朝聞書	3
調か	3
と定	3
予うと	3
よっか	3
わ味	3
意産	3
共彦君	3
克ろ	3
ごリン	3
サラマン	3
一浦	3
三割	3
三治省	3
自わ	3
使印	3
指民	3
自か	3
しる会	3
都十日	3
二十七日	3
二	

1月	4
15日	4
10万円	4
←	4
ない	3
い	3
では	3
だ	3
な っ	3
女	3
なる	3
後	3
できる	3
する	3
40	3
よる	3
午前	3
『 』	3
場合	3
二	3
ません	3
英語	3
高校	3
木下さん	3
硫酸	3
5時	3

国債	3
ノ	3
よく	3
月	3
制度	3
研究	3
港区	3
コース	3
大学	3
代表	3
著	3
目	3
バス	3
豊島区	3
本社	3
5分	3
悪徳	3
案内	3
宇佐美	3
営業	3
お申込み	3
科	3
額面	3

学部	3
学科	3
含	3
店	3
元金	3
現代	3
建	3
香水	3
高	3
工	3
妻	3
済	3
最大	3
支給	3
資料	3
書類	3
選考	3
賞与	3
昇給	3
職種	3
進呈	3
新兵隊	3
やくざ	3
すべて	3

雪	3
総額	3
大学院	3
中野	3
中央区	3
貯蓄	3
長篇小説	3
坪価	3
呈	3
デニム	3
都電	3
答える	3
東宝	3
南町	3
入学	3
バ停	3
不問	3
別荘地	3
変る	3
又	3
面接	3
面接日	3
ロードジム	3

證書	ぢ
3分	ぢ
6日	ぢ
10日	ぢ
30歳	ぢ
14日	ぢ
10分	ぢ
5年	ぢ
120錠	ぢ
26日	ぢ
342	ぢ
2年	ぢ
1000cc	ぢ
13日	ぢ
28日	ぢ
4日	ぢ
50円	ぢ
007サンダーボール作戦	ぢ

広さと深さの調査　また，表2にもどる。表3，4できまった各語の段階を，段階点数と呼ぶことにする。表2の層別度数を，段階点数に置きかえると表6ができる。

表　6

	順位	総度数	T 1	T 2	T 3	T 4	T 5	T 6	T 7	T 8	T 9	T10	T11	T12	層 数	合計点
、	1	1036	5	5	5	4	5	5	5	—	5	5	4	5	11	53
の	2	968	5	5	4	5	5	5	5	—	4	4	4	5	11	51
。	3	512	4	1	4	2	5	4	4	—	4	4	1	4	11	37
を	4	480	4	4	4	4	4	4	4	—	4	4	1	4	11	41
に	5	448	4	4	4	3	4	4	4	—	3	4	2	4	11	40
は	6	361	4	2	4	1	4	4	4	—	3	3	2	4	11	35
¶M○	7	343	—	—	4	—	3	—	2	—	5	—	4	5	6	23
が	8	325	4	—	4	1	4	3	4	—	3	3	—	4	9	30
て	9	304	4	3	4	1	4	4	4	—	3	3	1	4	11	35
と	10	296	4	4	3	3	4	4	4	—	3	3	2	4	11	38
〔	11	290	3	2	3	1	3	3	3	—	4	2	3	5	11	32
〕	12	280	3	2	3	1	3	3	2	—	4	2	3	5	11	31

表6の右端2欄は新しく加えられたものだが，層数とは，各語が度数の多少にかかわらず存在している層の数をかぞえたものである。だから，層数は，Tに関していえば1から12までの数となる。例えば，「、」はT8を除いて他のすべてにあるから層数は11となる。合計点は，段階点数を合計したもの。層数は出現の幅を表わし，合計点は，幅の同じものの中での使用度の深さの度合を示す。

層数と合計点とによって，広さと深さを測り，広くて深い語から狭くて浅い語にいたるまで，語の群分けをしてみよう。最初に示したように「広い」「狭い」と「深い」「浅い」を掛け合わせれば四つの群になるが，今は，それぞれの次元に「中位」を入れて，九つの群に分けることにする。層数と合計点との関係を考えると，層数12の語において，合計点の最大値は12×5で，60であり，最小値は12×1で12であるから，その値は60から12までの整数である。同様に層数11の語の合計点は55から11まで，層数10のものは50から10までというふうになる。それぞれの中を，「深い」「中位」「浅い」の三部に分ける。分け方は，平均3.5から1.5までを「中位」とし，平均3.5をこえるものを「深い」，平均1.5未満のものを「浅い」とする。層数10でいえば，50から36までが「深い」，35から15までが「中位」，14から10までが「浅い」となる。広さの方は，層数12から9までを「広い」，8から4までを「中位」，3から1までを「狭い」とした。こうしてできる広さと深さの理論的配置は表7のようになる。

表 7

	層数	深い	中位	浅い
広い	12	60 — 43	42 — 18	17 — 12
	11	55 — 39	38 — 17	16 — 11
	10	50 — 36	35 — 15	14 — 10
	9	45 — 32	31 — 14	13 — 9
中位	8	40 — 29	28 — 12	11 — 8
	7	35 — 25	24 — 11	10 — 7
	6	30 — 22	21 — 9	8 — 6
	5	25 — 18	17 — 8	7 — 5
	4	20 — 15	14 — 6	5 — 4

狭い	3	15 — 11	10 — 5	4 — 3
	2	10 — 8	7 — 3	2
	1	5 — 4	3 — 2	1

表6の最右端の合計点により，表7のわくづけに従って各語を分配したものが表8である。G層別について同様の処置を施したものが表9である。ただしこの表には，総度数6以上の287語だけがのっている。したがって，幅の狭い語で度数が6以上ならば，どうしても浅くはならないので，狭くて浅い群には一つも語が登記されていないのである。度数1までの語をこの表にのせれば，度数1は全部狭くて浅い群にはいるのだから，この群の語数がけたちがいに大きくなるわけである。

表8

語の使われかたの広さと深さとの関係

総度数6以上の語（287）の層数別，頻度段階点数別一覧表（話題層別）

深さ / 広さ 層数	1 深い	2 中位	3 浅い
A 広い (53語) 11	53、 51の 41を 40に	38と 37。 35は 35て 32「 31」 31で 26「 26」	
10		34た 26〃 21し 20いる	
9		30が 24、 23ない 23こと 22いう 17では 17だ 17など 16い	
8		20も 20ある 19から 18もの 17この 16＝ 14なっ 14その 13には 13これ 13や 12れた	11それ 10する 9う 8ても
7		21— 13ところ 12東京 12か 11ため 11へ	10でも 10なる 9ので 8なく 7のが
B 中位 (152語) 6	23¶M○	11にも 10いっ 9より 9前 9まで 9あり 9あっ 9男 9なり	8¶10 8¶20 8しかし 8自分 8日本 8ば 8また 8への 8ながら 8よう 7私 7だが 7とっ 7なら 7れ 6との 6目 6られる
5		8でき 8のは 8たが	7み 7人 7れて 7考え 7受け 7中 7夜 7れる 7とき 7よる 7点 6き 6ほか 6あと 6たい 6時 6ては 6次 5まったく 5間

			5十分 5一番 5だった 5だと
4		111 9ます 95 9女 8— 82 87 8後 8ました 850 76 73 7です 7政府 7午後 6他 6上 6方 6つい 6午前	5写真 5たのは 5問題 5¶30 5いい 5きょう 5見 5三〇 5大きな 5とも 5もう 5近く 5限り 5高校 5こんど 5対する 5たり 5はず 5よく 4対し 4多い 4夫 4一人 4一日 4うえ 4何 4だけ 4とは 4なお 4わけ
3		100 710 7{ 7普通 6… 64 520 515 5東京都 545 525 555 5一部 59 5いま 5おり 58 5場合 5二 5米国 5ません 5られて 5150	4二〇 411 4年齢 4六 4あなた 4音楽 4結婚 4四 4七割 4すぐ 4でした 4方針 4られ 4以上 4映画 4家庭 4今年 4事件 4社 4スポーツ 4そう 4二人 4ように 312
C 2 狭い (82語)		530 5歩 500 530分 5歌 540 5行わ 4！ 4＜ 4＞ 4商業 4年 4共 4こんな 411時 4インド 4『 4』 4ついて 4170 4英語 4国債 4ノ 3女子 3100 3A 318日 3車 3西口 3宅地 3値上げ 3調べ 3電話 3八〇 3不況 3下さい 3社会	
1	4¶MO1 4¶MO2	3¶NO1 33月 3迄 3～ 32月 3¶MO3 3¶天O 3／ 3¶MO5 3代 3同新 305 335 3晴着 3145 3¶MO4 2株 2千代田区 2男子 2電 2募集 2100名 2完備 2経験 2履歴書 2委託 2応募 2願書受付 2熊谷 2高卒 2作家 2山林 2社員 2新宿 2送料 2当社 2必要経費 2郵送 2有 2KK 2100円 2200 2¶再O	

表9

語の使われかたの広さ深さとの関係

総度数6以上の語(287)の，層数別，頻度段階点数別一覧表(文種層別)

広さ＼深さ 層数	1 深い	2 中位	3 浅い
A 広い (53語) 16	76, 73の 66。 61を 61に 60は 58と	52た 52で 51が	
15	53て	40「 39」 36から	
14		38[38] 32いう	
13		32ない 31し 23など	18なる
12		35も 30こと 27もの	
11		28・ 27ある 24この 24— 23いる 22では 19これ	
10		22い 21だ 20その 16あっ	14ため 14か 12ば
9		17ます 16なっ 15れた 14ところ	13なり 10する
8		17や 16より 15へ	11日本 11とっ 10み 10方 9への 9う
B 中位 (152語) 7		22¶M○ 19〃 19東京 16にも 15には 13あり 13でも 11前 11考え	10でき 10私 10れて 10なく 10なら 8たが 8とき 7しかし
6		15= 12いっ 12まで 12それ 12人 11— 11後 10自分 10だが 9れる 9ながら 9との	8また 8ても 8たのは 8点 8はず 7おり 7まったく 7れ 7いい 7すぐ 7目 6次 6られる
5		14 1 11 5 10中 9男 9 3 9上 9つい 8¶20 8受け 8ので 8よう 8こんな	7¶10 7写真 7き 7のが 7ほか 7 8 7たい 7結婚 7時 7場合 7ません 6夜 6間 6四 6多い 6とも 6夫 6うえ 6近く 6こんど 6だけ 5一日
4		10 2 10です 9女 9 7 9東京都 9 4 9政府 8 6 8のは 8問題 6一部 6あと 6年齢 6見 6大きな 6二 6もう 6一番 6何 6とは	5よる 5¶30 5午前 5十分 5ては 5られ 5一人 5家庭 5社 5社会 5対する 5だと 5わけ 4高校
3		8ました 7… 7他 7年 6！ 6{ 6< 6 9 6共 6そう	4対し 4行わ 4三○ 4ついて 4られて 4下さい

		5普通 525 520 5いま 511 511時 5六 5あなた 5七割 5車 5でした 5電話 5米国 5限り 5今年 5だった 5たり 5なお 5二人 51 5ように	4社員 4よく 312
C 2 狭い (82語)	8歩	70 610 620 615 600 6迄 62月 5～ 540 5代 5株 5千代田区 5電 5募集 5『 5』 5不況 5熊谷 4歌 445 555 450 4／ 4女子 4男子 4経験 4宅地 418日 4西口 4履歴書 4英語 4応募 4国債 4新宿 4送料 4当社 4郵送 4100円 3インド 3A 3音楽 3きょう 3調べ 3八○ 3方針 3170 3150 3以上 3映画 3スポーツ	
1	5¶M○1 4¶M○2 4¶N○1 43月 4100名	330分 3¶M○3 3¶天○2 3商業 3¶M○5 3同新 305 3100 335 3完備 3145 3¶M○4 3委託 3願書受付 3高卒 3作家 3山林 3必要経費 3有 3KK 2晴着 2値上げ 2事件 2200 2¶再○	

いま，表9によって，広くて深い語にどんなものがあるかと見ると，次の8語である。

、 の 。を に は と て

句読点とテニヲハがきれいにそろっている。句読点は，国語の文章の表記に絶対欠かすことのできないものであり，テニヲハは日本語の構造の根幹をなす語であることが，これでよくわかる。

次に幅が広くて深さ中位のもの33語を語類によって分けてみると，次のようになる。

〔助　詞〕 で が から など も では や より へ

〔助動詞〕 た ない だ ます れた

〔指示語〕 この これ その

〔動　詞〕いう　し　ある　いる　い　あっ　なっ

〔名　詞〕こと　もの　ところ

〔記　号〕「　」〔　〕・(ナカグロ)　—(横線)

助動詞では，単純な判断を表わすものがここにあり，動詞では，形式動詞，名詞では形式名詞が集まっている。すなわち，詞の中で辞に近い性質をもった形式的な語が，辞とともにここに位置する。

広くて浅い語は12語で次のとおりである。

なる　ため　か　ば　なり　する　日本　とっ　み　方　へ　の　う

やはり，形式語が大部分であるが，「日本」という語がここにはいっているのは，新聞語彙の特徴を示すものであろう。

以上の事実から，日本語の基幹をなす語に，重要さの度合で序列をつけてみると，表9の上部の配置から，大体のところ次のようになろう。

1　テニヲハ　の　を　に　は　と　て　が　も　から　より　へ

2　単純な文末判断辞　た　だ　ない

3　コ・ソのつく指示語　この　これ　その

4　形式動詞　する　いう　ある　いる　なる

5　形式名詞　こと　もの　ところ　ため

表9で，幅の中位の欄を見ると，「深い」には1語もない。かなりの幅をもって出現し，かつ，どこにも深く存在するという語はなかったわけである。中位の幅で，深さの「中位」と「浅い」とをいっしょにながめると，層数7と6のところに，「にも」「には」「まで」「それ」「あり」「いっ」「おり」「とき」「点」など，基幹的形式語のつづきが見られるが，そのほかにも注意すべきことがある。「ても」「でも」「ながら」「たが」「しかし」「また」など，接続助詞または接続詞の類があること，「まったく」「すぐ」のような副詞があること。また，「日本」につながる性格の語として「東京」「前」「後」「人」「私」「自分」などの語があることなどである。「日本」や「東京」がこの資料の範囲内における新聞語彙的なものであることはすぐ感じられるが「前」「後」「人」「私」「自分」などが新聞語彙的なものか，一般基本語彙的なものかは，これだけではわからない。しかし，層数5，4，3の欄を見る

と，「男」「女」「結婚」「東京都」「政府」「問題」「社会」「年齢」「家庭」「米国」「社員」など，人事や社会現象を表わすことばが見え，これらが新聞語彙的なものであることには疑いがあるまい。

さらに，非常にはっきりしているのは，幅の狭いところに属する語彙である。「深い」と「中位」とを合わせて，一目見てわかることは，ここにある語の大半が案内広告欄に属するものであることだ。求人関係の語が多いこと，また，「3月」「2月」という語があることから見てこの入力データが2・3月の新聞記事であることもすぐわかる。また「¶ M○1」は「①」のような表記をさしているのだが，これらの特殊マークがどの層に属するかは表5を見ればわかる。

層による用語の特徴　広さと深さの調査から基本語彙を求めようとする試みが，層による用語の特徴の一端を明らかにするのであるが，また別の方向から，層と用語との関係をさぐってみよう。

もし，各語が層とは無関係で所属層の性格に左右されることがないのならば，語の各層への散らばりかたは，大まかに言ってどの語も大体同じ割り合いになるはずである。それが実際はどうであるか，ある語はある層に特に多いとか特に少ないとかいう現象があるだろうか。それを調べるためにまず各語を層に対して中性と考えた場合の層別理論度数を算出してみる。理論度数は各層の層別総度数の延べの総語数に対する比率を算出し，その比率によって，各語の総度数を比例配分すれば出せる。T層別についていえば，層別総度数とその比率は次のようになっている。

	T 1	T 2	T 3	T 4	T 5	T 6	T 7	T 8	T 9	T 10	T 11	T 12	計
層別総度数	2,080	225	2,276	93	3,846	773	1,693	150	1,476	1,288	1,064	5,552	20,516
層間の語の分布の比率	10.1	1.1	11.1	0.5	18.8	3.8	8.3	0.7	7.2	6.3	5.2	27.1	100

だから，度数1036の| 、|について，層別理論度数を作れば，1036に10.1以下の係数を掛けて次の表の上欄の数値が得られる。そして下欄の実度数と比べてみる。

	T 1	T 2	T 3	T 4	T 5	T 6	T 7	T 8	T 9	T 10	T 11	T 12	計
理論度数	105	11	115	5	195	39	86	7	75	65	54	281	1,036
実度数	116	15	111	6	274	48	118	—	64	97	41	146	1,036

理論度数と実度数とが非常に近いものもあり，やや離れたものもある。その離れかたに統計学的基準を設けて，標準からの逸脱度を測れば，逸脱している所に，その語とその層との特殊な関係を見ることができる。逸脱度を測るためには，品質管理の方法や，有意差の検定などの統計学的方法を適用すべきであるが，今は仮りに全く便宜的に次のような基準を設けて大きな逸脱を拾い出してみた。

片方の数が	0 または 1	で，他方の数が	10 以上	ならば，逸脱あり。
	2 から 5 まで		その 5 倍以上	
	6 から 9 まで		その 3 倍以上	
	10 から 19 まで		その 2 倍以上	
	20 以上		その 1.5 倍以上	

この基準でいくと，「読点」の層別分布で逸脱があるのは，T12 の理論度数 281 に対する実度数 146 だけである。このように，実度数が理論度数より少ない方へ逸脱しているのは，その語のその層に対する消極的特徴と見ることができる。すなわち，読点は，広告欄には少ないという特徴があるわけである。これに対して，実度数が理論度数に対し多い方へ逸脱していれば，その語はその層に多いという積極的特徴を認めるのである。

だが，この方法は，度数の高い語についてしか適用できない。小さな数を比例配分してみても意味がないからである。そこで，この場合は，度数 35 以上の語 37 箇についてだけ，理論度数を算出した。表 10 と表 11 は，T，Gそれぞれの層別について，理論度数と実度数との比較を行なったものである。表の各欄の左上が理論度数，右下が実度数で，▲は積極的特徴，△は消極的特徴の認められるところである。

表 10　話題（T）層別による度数分布，理論度数と実度数との比較

顕著な差
△　理＞実
▲　理＜実

No.	語	総度数	T1 政治 10.1	T2 外交 1.1	T3 経済 11.1	T4 労働 0.5	T5 社会 18.8	T6 国際 3.8	T7 文化 8.3	T8 地方 0.7	T9 スポーツ 7.2	T10 婦・家 6.3	T11 芸・娯 5.2	T12 広告 27.1	全体 延語数 20.516
1	、（読点）	1036	105 116	11 15	115 111	5 6	195 274	39 48	86 118	7 —	75 64	65 97	54 41	281 △ 146	理論 度数
2	の	968	98 122	11 18	108 △ 60	5 10	182 207	37 ▲ 61	80 118	7 —	70 △ 49	61 59	50 39	262 225	実度数
3	。（句点）	512	52 68	6 1	57 45	3 2	96 ▲ 150	19 17	42 ▲ 71	4 —	37 41	32 56	27 △ 2	139 △ 59	
4	を	480	48 ▲ 74	5 11	53 41	2 6	90 111	18 25	40 48	3 —	35 33	30 ▲ 51	25 △ 2	130 △ 78	
5	に	448	45 60	5 9	50 △ 31	2 3	85 102	17 24	37 ▲ 63	3 —	32 22	28 ▲ 44	23 △ 3	121 87	
6	は	361	35 52	4 3	39 32	2 1	66 92	13 19	29 37	2 —	25 26	22 30	18 △ 3	95 66	
7	¶M○	343	35 △ —	4 —	38 ▲ 83	2 —	65 △ 7	13 △ —	29 △ 4	2 —	25 ▲ 95	22 △ —	18 32	93 122	
8	が	325	33 45	4 —	36 30	2 1	61 ▲ 116	12 13	27 38	2 —	23 25	20 30	17 △ —	88 △ 27	
9	て	304	31 40	3 5	34 △ 22	2 1	57 ▲ 91	12 18	25 ▲ 41	2 —	22 18	19 32	16 △ 2	82 △ 34	
10	と	296	30 40	3 9	33 18	1 3	56 66	11 20	25 ▲ 39	2 —	21 17	19 22	15 5	80 57	
11	「	290	29 16	3 2	32 △ 8	1 1	55 △ 31	12 8	24 6	2 —	21 ▲ 36	18 △ 3	15 13	79 ▲ 166	

12	」	280	28	16	3	2	31	△8	1	1	53	△31	12	8	23	5	2	—	20	▲36	18	△3	15	13	76	▲157
13	た	266	27	35	3	4	30	26	1	3	50	▲96	11	13	22	27	2	—	19	17	17	21	14	△—	72	△24
14	•（ナカグロ）	264	27	△2	3	—	29	△7	1	1	50	△9	11	△—	22	△3	2	—	19	△6	17	△3	14	▲56	72	▲177
15	で	234	24	35	3	2	25	18	1	1	44	51	10	14	19	28	2	—	17	13	15	▲32	12	△2	63	△38
16	—（ハイフン）	208	21	14	2	—	23	▲144	1	—	39	11	9	—	17	11	1	—	15	11	13	△1	11	△—	56	△16
17	「	159	16	17	2	1	18	△2	1	1	30	41	7	14	13	18	1	—	11	△1	10	6	8	▲43	43	△15
18	」	155	16	19	1	1	17	△2	1	1	29	37	7	14	13	19	1	—	11	△1	10	6	8	▲43	42	△12
19	0	104	11	△—	1	—	12	▲54	1	—	20	△—	4	—	9	—	1	—	7	6	7	—	5	▲44	28	△—
20	も	97	10	13	1	3	11	11	0	—	18	▲38	4	—	8	14	1	—	7	2	6	8	5	—	26	△8
21	ない	89	9	18	1	2	10	10	—	—	17	▲35	3	3	7	6	1	—	6	6	6	6	5	—	24	△3
22	〃	82	8	10	1	6	9	6	—	—	15	18	3	2	7	2	1	▲13	6	2	5	2	4	—	22	21
23	いる	81	8	17	1	—	9	7	—	1	15	24	3	9	7	18	1	—	6	2	5	1	4	1	22	△1
24	ある	68	7	17	1	—	7	4	—	—	13	13	3	4	6	5	—	—	5	5	4	12	4	—	18	8
25	いう	68	7	13	1	3	7	7	—	—	13	14	3	5	6	13	—	—	5	2	4	6	4	—	18	5

26	から		65	7	14	1	—	7	8	—	—	12	20	2	3	5	10	—	—	5	4	4	1	3	—	18	5
27	こと		65	7	12	1	2	7	6	—	—	12	12	2	6	5	8	—	—	5	2	4	7	3	—	18	10
28	し		59	6	15	1	1	7	3	—	1	11	14	2	3	5	7	—	—	4	3	4	6	3	—	16	6
29	い		49	5	6	1	1	5	5	—	—	9	24	2	5	4	2	—	—	4	1	3	2	3	—	13	3
30	1		49	5	—	1	—	5	▲26	—	—	9	—	2	—	4	—	—	—	4	9	3	—	3	1	13	13
31	この		44	4	14	—	—	5	3	—	—	8	9	2	3	4	4	—	—	3	4	3	2	2	—	12	5
32	では		41	4	3	—	1	5	4	—	—	8	16	2	3	3	4	—	—	3	2	3	5	2	—	11	3
33	もの		41	4	8	—	—	5	6	—	—	8	10	2	2	3	5	—	—	3	2	3	3	2	—	11	5
34	ます		40	4	—	—	—	4	—	—	—	8	4	2	—	3	2	—	—	3	—	3	8	2	—	11	▲26
35	=		37	4	2	—	2	4	3	—	—	7	2	1	5	3	—	—	—	3	10	2	2	2	—	10	11
36	だ		35	4	8	—	1	4	3	—	—	7	10	1	1	3	4	—	—	3	—	2	4	2	1	10	3
37	など		35	4	9	—	2	4	8	—	—	7	7	1	1	3	4	—	—	3	1	2	—	2	1	10	△2

表11　文種層(G)別による度数分布，理論度数と実度数との比較

顕著な差
△理＞実
▲理＜実

No.	語	総度数	G1 ニュース 27.6	G2 ニュ解説 1.9	G3 社・コ 0.5	G4 ニュ特記 3.0	G5 特読み 5.8	G6 評論 3.9	G7 実読み 2.7	G8 探ルポ 2.7	G9 長期ニュ 1.1	G10 記録 16.0	G11 紹介記 1.1	G12 読作文 3.8	G13 コミュニケ 3.0	G14 小説 0.6	G15 商広 13.7	G16 案広 12.7
1	、	1036	287 375	20 26	5 7	31 37	60▲ 102	10 47	28▲ 43	28 31	11 16	166△ 103	11 7	39 51	31 45	6 9	143△ 82	132△ 55
2	の	968	267 353	18 14	5 3	29 38	56 59	33 54	26 33	26 34	11 13	155△ 66	11 14	37 36	29 26	6 3	134 169	123△ 53
3	。	512	141 179	10 13	3 5	15 22	30▲ 51	20▲ 32	14 20	14▲ 30	5 5	82△ 27	5 8	19 32	15 29	3 5	70△ 41	65△ 13
4	を	480	133 191	9 13	2 8	14 21	28 36	19 23	13 23	13 14	5 5	77△ 9	5△ 8	18 30	14 21	3 1	66 65	61△ 12
5	に	448	124 147	9 15	2 4	13 23	26 39	17 27	12 22	12 17	5 3	72△ 19	5 4	17 26	13 15	3 4	61 65	57△ 18
6	は	361	100 128	7 9	2 1	11 20	21 23	14 17	10 12	10 19	4 4	58△ 23	4 10	14 18	11 11	2 6	49 45	46△ 15
7	¶M○	343	95△ 19	7 —	2 —	10△ —	20△ 2	13△ —	9 —	9 2	4 —	55▲ 197	4 1	13△ —	10△ —	2 —	47 32	44▲ 90
8	が	325	90 131	6 16	2 1	10 10	19▲ 39	13 10	9 13	9 14	4 5	52△ 15	4 4	12 19	10▲ 21	2 1	45△ 25	41△ 1
9	て	304	84 113	6 12	2 2	9 17	18▲ 42	12 13	8 13	8 12	3 —	49△ 8	3 6	12 17	9 15	2 2	42△ 24	39△ 8
10	と	296	82 101	6 8	1 2	9 20	17 19	12 21	8 10	8 7	3 4	47△ 17	3 6	11 12	9 12	2 5	41 45	38△ 7

11	「	290	80	6	59	1	—	9	1	17△	2	11△	3	8	2	8	5	3	2	3	46	3	11	2	9	5	2	1	—	40▲	—	37▲	104
12	」	280	77	5	59	1	—	8	1	16△	2	11△	3	8	2	8	4	3	2	3	45	3	11	2	8	5	2	1	—	38▲	—	36▲	97
13	た	266	74▲	5	131	1	7	8	2	15	11	10	26	7	2	7	11	9	3	9	43▲	7	10	2	8	8	2	1	11	36△	3	34△	4
14	•	264	73△	5	9	1	1	8	1	15△	—	10	2	7	1	3	—	1	3	1	42	—	10	2	7	6	1	1	14	36	—	17	53
15	で	234	65	4	76	1	7	7	1	14	7	9	17	8	1	6	18	6	11	2	37△	11	9	5	8	13	6	1	1	32	1	25	124
16	—	208	57△	4	30	1	—	6	1	12	—	8	3	6	1	6	10	6	1	2	33△	—	9	5	8	7	1	1	1	29△	1	4	11
17	「	159	44	3	45	1	—	5	1	9	3	6	22	4	9	4	2	4	2	2	25▲	1	6	3	6	6	5	1	1	22△	9	20△	3
17	」	155	43	3	48	1	—	5	1	9	3	6	19	4	9	4	2	4	2	2	25▲	1	6	3	6	6	5	1	1	21△	6	20△	3
18	0	104	29▲	2	6	1	—	3	1	6	—	6	4	—	3	3	—	3	1	1	17▲	—	6	3	4	—	3	1	1	14△	—	13△	3
19	ゐ	97	27	2	33	—	—	3	2	7	10	5	8	4	11	2	1	5	7	1	16	7	10	5	6	3	5	2	1	13	—	12△	6
20	たい	89	24	2	35	6	—	3	—	5	6	5	12	3	4	2	1	7	2	1	14△	1	3	1	5	4	2	1	1	12△	1	10△	2
21	ゃ	82	23	2	32	8	—	2	—	6	—	2	3	2	2	2	—	2	1	2	13	1	4	3	3	—	2	1	—	11	18	10	10
22	い、そ	81	22▲	2	48	1	—	2	—	7	5	6	8	5	2	2	1	2	1	1	13	3	4	1	4	1	1	—	—	11△	1	10△	1

24	ある	68	19	29	1	2	—	—	2	2	4	5	3	7	2	—	2	—	1	—	12△	2	1	2	3	6	2	5	—	—	11	7	9	1
25	いう	68	19	22	1	3	—	1	2	7	4	3	3	12	2	1	2	3	1	1	12△	2	1	—	3	6	2	2	—	1	11	4	9	—
26	から	65	18	27	1	3	—	1	2	3	4	2	3	6	2	2	2	3	1	3	10	4	1	2	2	3	2	1	—	4	9	1	8	—
27	こと	65	18	20	1	2	—	2	2	5	4	6	3	3	2	—	2	4	1	—	10△	2	1	—	2	6	2	5	—	—	9	9	8	1
28	し	59	16	20	1	2	—	2	2	4	3	7	3	5	2	2	2	—	1	—	9	2	1	1	2	3	2	5	—	—	8	5	8	1
29	い	49	16	25	1	3	—	—	1	2	3	4	2	—	1	1	1	—	1	2	8	1	1	—	2	6	1	2	—	—	7	3	6	—
30	1	49	16	8	1	—	—	—	1	—	3	—	2	—	1	—	1	—	1	—	8▲	26	1	—	2	—	1	2	—	—	7	3	6	10
31	この	44	12	20	1	—	—	3	1	—	3	2	2	2	1	1	1	2	—	1	7	2	—	—	2	3	1	3	—	—	6	5	6	—
32	では	41	11	17	1	1	—	—	1	1	2	4	2	—	1	1	1	2	—	—	7	5	—	—	2	3	1	4	—	1	6	2	5	2
33	もの	41	11	14	1	—	—	—	1	2	2	2	2	4	1	1	1	3	—	3	7	1	—	—	2	4	1	2	—	—	6	4	5	1
34	ます	40	10△	—	1	—	—	—	1	—	2	1	2	—	1	6	1	1	—	—	6	1	—	—	2	4	1	1	—	1	5△	22	5	3
35	=	37	10	12	1	—	—	—	1	1	2	—	1	—	1	2	1	—	—	—	6	11	—	—	1	—	1	—	—	—	5	6	5	5
36	だ	35	10	9	1	—	—	—	1	2	2	5	1	7	1	2	1	4	—	1	6	2	—	—	1	—	1	—	—	—	5	2	4	1
37	など	35	10	12	1	1	—	1	1	4	2	5	1	3	1	1	1	1	—	2	6	1	—	1	1	1	1	—	—	—	5	2	4	1

表10，表11において，逸脱の認められた箇所だけを抜き出し，積極的特徴と消極的特徴とに分けて，語と層の関係を見よう。表12は，語を語類に分けて，語と層の特徴関係を整理したものである。

表 12

	語	話題(T)層別 ▲	話題(T)層別 △	文種(G)層別 ▲	文種(G)層別 △
句読点	。	社，文	芸，広	特読，評，ル	記，商，案
	、	——	広，	特読，実読	記，商，案
文章記号	・	広	政，経，社，国，文，ス，婦	案	ニ，特読
	「，」	——	経，ス，広	記	商，案
	〃	地	——	——	——
	〔，〕	ス，広	経，社，婦	商，案	特読，評
	—	——	婦，芸，広	記	ニ，商，案
特殊マーク	¶M○	経，ス	政，社，国，文，婦	記，案	ニ，ニ特，特読，評読作，コ
	0	経	政，社，広	記	ニ，商，案
	1	経	——	記	——
テニヲハ	が	社	芸，広	特読，コ	記，商，案
	の	国	ス	——	記，案
	に	文，婦	経，芸	——	記，案
	を	政，婦	芸，広	——	記，商，案
	て	社，文	経，芸，広	特読	記，商，案
	と	文	——	——	記，案
	は	——	——	——	記，案
	も	社	広	——	案
	で	婦	芸，広	——	記，案
	など	——	広	——	——
文末辞	た	社	芸，広	ニ	記，商，案
	ない	社	広	——	記，商，案
	ます	広	——	商	ニ
形式語	いる	——	広	ニ	商，案
	ある	——	——	——	記
	いう	——	——	——	記
	こと	——	——	——	記

この表から，日本語の構造と文体との関係について，なにがしかのことが見られると思うので，読めるだけのことを読んでみようと思う。

1. 句読点は，記録と広告に少ない。記録とは，ラジオ・テレビ番組や株式欄のように，ことがらや数量だけが並んでいる欄であるから，文章の形式上，句読点がいらないのであろう。広告のうち，案内広告は，1字1字が金であるから，句読点を省く習慣がある。このように句読点は一般的にいって，日本語の文章のあるところに必ずあるわけであるが，記録や案内広告のように，特殊な地帯では，例外的に，その必要度が少ない。

2. 句読点または読点がT別では社会記事と文化記事に，G別では読みもの記事や評論，ルポなどに多いことになっているが，これが何を意味するかよくわからない。読点の数の多少は個人の書きぐせにもよることであり，文構造を考えるうえにあまり大きな意味はなさそうだが，句点の数は文の数を意味するので，句点の数が一定の語数に対して多ければ，その文章の文の平均の長さが短かいわけである。特別読物や評論，ルポなどの記事が平均して文が短かいことになるのであろうか。もっと調べてみなければわからない。

3. テニヲハ類についても，句読点と同様のことがいえる。一貫して記録形式と案内広告の記事に少なく一部のテニヲハは商業広告にも少ない。T別の方で芸能欄に少ない傾向があり，経済欄にもややその傾向があるのは，芸能欄のうちのラジオ・テレビ番組，経済欄の中の株式欄がそうさせているものと思われる。

4. テニヲハについて，積極的特徴は，消極的特徴ほどに明瞭ではないが，T別において，社会記事，文化記事，婦人家庭記事がしばしばあげられている。文章にテニヲハがたくさん使ってあるということは，それが日本語らしい曲折を多分に含んだ文章であり，ぶっきらぼうでない文章であることを意味するものと思われる。だから，知らせることを本位とするよりも読ませることを本位とする文章がこれらの記事に多いということになるのではあるまいか。G別では，テニヲハに積極的特徴はあまりないが「が」と「て」だけが特別読物に多くなっているのは，「読みもの」ということに何か関係があるかどうか，まだ，わからない。

5. 形式動詞や形式名詞が記録類に少ないのは，それらの語が形式語であってテニヲハや助動詞に近いことを示していよう。

6. 「た」「ない」「ます」という文末辞には，興味ある事実が見られる。「た」と「ない」は記録や広告に少なく，社会記事に多い。社会記事に多いのは，それが事件のてんまつを叙することが多いからだろうか。「た」がニュース記事に多いのは，ニュースが過去に起こった事実を報道するからであろう。「ます」がニュース記事に少なく，商業広告に多いのは，全くさもありなんと思われる。案内広告と商業広告とは，しばしば同列に並んでいたが，「ます」の使用においては，同列でないのがおもしろい。

7. 記号類は，その記号の性格によってそれぞれの特徴をもつようであるが，この調査ではいろいろな記号を，便宜的に一つの記号に統一して扱ったりしているから，この結果では何ともいえない。

あ と が き

2万語のテストラン結果で行なった語彙分析の試行は，以上のとおりである。非常に少ない語数で行なったものであるから，この結果自身は問題ではなく，方法だけが問題である。層別を足がかりにして新聞資料の範囲内での基本語彙を求めること，層別に特徴のある語彙を求めること，この二つを行なってみた。このほか，語彙統計の上での基本的作業がいろいろ行なわれなければならないが，それらと併せて，この二つの方向での分析を今後の大量のデータについて施してみたいと思う。

おわりに，この時のデータを，その後作った統計プログラムによって，計算機で処理した結果の総度数と度数分布を示しておく。

出現度数	異なり語数	出現度数	異なり語数
500以上	3	9	16
499～400	2	8	20
399～300	4	7	47
299～200	7	6	53
199～100	3	5	77
99～ 80	4	4	165
79～ 60	4	3	310
59～ 40	7	2	933
39～ 30	7	1	5,782
29～ 20	23	計	7,554
19～ 15	30	延べ語数	20,496
14～ 10	57		

異なり語数7,554のうち，度数1のものが5,782という多数を占めている。これは，従来の語彙調査結果に比べて，著しく多い量である。こういう結果になったのは，長単位と新聞記事のためであり，その中の非常に多くの部分を株式欄その他に出ていた数量などが占めている。また，延べ語数20,496は35ページの表に示した手集計による延べ語数20,516とちがっている。

計算機の計算の方が正しいことはいうまでもないが，ここでは，行なった作業の結果をすべてそのまましるしたので，あえてちがったままにしておいた。