

国立国語研究所学術情報リポジトリ

Word counting and basic vocabulary

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: 林, 四郎, HAYASHI, Shiro メールアドレス: 所属:
URL	https://doi.org/10.15084/00001004

語彙調査と基本語彙

林 四 郎

これは、昭和44年2月15日、国立国語研究所創立二十周年記念講演会の席上で行なった講演をもとにして書いたものである。執筆している昭和45年9月においては、われわれは、すでに報告書『電子計算機による新聞の語彙調査』を世に送っており、これによって、いわゆる「1紙1年分」の長単位・短単位、両種の表を公表しているのであるが、講演時には、まだ長単位の表しかできていなかった。ここで材料にしている新聞語彙は、すべて長単位段階でのものである。

1. いろいろな基本語彙

「基本語彙」ということばは、いろいろな意味で使われる。われわれが今の新聞語彙調査を始めたころ、ひとりのマスコミ関係者が、何のために新聞の語彙調査をするのかと問うた。わたくしが「一つの目的は、基本語彙を求めることにある。」と答えたら、その人は、ふしぎそうな顔をして、「新聞で基本語彙なんか作っても、新聞記者は、わがままな連中ばかりだから、だれもそんなものに従いはしませんよ。無駄ではありませんか。」と反問した。この人においては、基本語彙は、「ことばを使う者が基本的に準拠すべき語彙」という意味でとらえられている。わたくしは、新聞語彙調査で基本語彙を求めるという時、そういう、価値観をこめた基本語彙を考えてはいなかったもので、その時はとまどったが、考えてみれば、基本語彙に関するこの人の理解は全く正しいものであって、とまどったわたくしの方がおかしいと言わなければならない。

そこでわたくしは、基本語彙をめぐるいろいろな考え方があるのは事実なのだから、いささか、その考え方を整理してみる必要があると感じた。わ

たくしは、大ざっぱに基本語彙といわれる概念をもうすこし細かく分けて、概略次のように定義しつつ、五つの概念を立ててみたい。

- (1) 基礎語彙 意味の論理的分析によって求められた半人工的な語彙
- (2) 基本語彙 特定目的のための「〇〇基本語彙」
- (3) 基準語彙 標準的社会人としての生活に必要な語彙
- (4) 基調語彙 特定作品の基調を作るのに働く語彙
- (5) 基幹語彙 ある語集団の基幹部として存在する語彙

1・1 基礎語彙

言語学習や言語使用の経済ということから、少数の有効な語を選びこれを有効に使うことによって、かなり広範囲のことまで言えるようにしようという考えが出て来る。この考えを組織的に進めて一つの言語体系を作ったのが、40年前の Basic English である。

C. K. Ogden は、1929年に、500語から成る Basic English (以下、B. E. と略記する。) の案を発表した。後に増補され、1932年には850語となった。これがB. E. の決定版である。B. E. は名詞600、形容詞150、動詞16、その他の語84から成り立つ。驚くべきは動詞が16しかないことであって、その16とは、

be, do, have, make, let, put, take, keep, get, come, go, give,
say, see, seem, send

である。日本語でいえば、「する」「なる」「ある」のような、ほとんど特定の意味内容をもたない補助動詞的なものばかりである。これだけの動詞でいろいろな動作や状態が言い表わせるとすれば、それは、前置詞や副詞を使って動詞を助けつつ、名詞とのさまざまな連語を作って活用するからにちがいない。

Ogdenは、I. A. Richards との共著『意味の意味』で示したように、言語の意味の開拓的な研究家であるから、意味を分析することによって、特殊な意味や複雑な意味を、一般的な意味や単純な意味の連合に還元することを考え、“The A B C of Basic English” という書に、B. E. の使い方を解説した。

B. E. にヒントを得て、いち早く、日本語について同方向の実験を試みたのが土居光知氏である。Ogdenの目的は、英語をもとにして、一つの国際語を

作り出すことにあったが、土居氏は国際語までは考えず、あくまでも日本の大衆のために基礎日本語を考えた。土居氏が最初にこれを発表したのは昭和8年(1933)で、Ogden が B. E. を 850 語にして発表した年の翌年であった。この時の基礎日本語の数は1000であったが、土居氏も、さらに考を重ねて、10年後の昭和18年(1943)に、100語をふやし、1100語にして、あらためて世に問うた。土居氏は基礎日本語を考案するについて、一つの日本語観をもっていた。それは、日本が明治維新によって新しい時代を迎え、急激に多量の外来文化をこなすのに漢語を用い、漢語によって新しい知識体系を作り上げたのであるが、その勢いはすでにきわまり、今は、明治十年代の漢語的知識体系から脱却して、新しいことばによる新しい知識体系を作るべき時期に来ているということであった。この場合、土居氏にとっては、知識体系が問題であったから、基礎日本語は、知識の伝達を目的とした一般の文章体において活用されるべきものであり、日常会話のことばや、漢語の多い文章は、基礎日本語適用の対象外とされた。会話には会話の慣習があるのであって、「奥様にお目にかかりとうございます。」を基礎日本語で「あなたの妻を見とうございます。」というのは無意味だという。

基礎日本語の選定に当っては、その語自身の使用頻度は考えに入れず、その語と他の基礎語との組み合わせによって、どれだけ広い範囲のことが言い表わせるかを重要視した。例えば、「見る」ことに関連して、「拝見」「お目にかかる」「お目にかける」「御覧」など、いろいろな敬語的表現が考えられるが、「知識を伝えることを目的とした文体」においては、そういう変化は無視することができるし、「覗く」「望む」「眺める」「にらむ」など、見方の変化には、「細い所から見る」「遠くを見る」「こわい目で見る」などのような、文脈からの言い替えを考えればよい。とすれば、見ることについては「見る」一語を基礎語と認めておけばよいと考えられる。

土居氏の考え方で、現在のわれわれの考え方からみて、やや奇異に見えることが一つある。それは、語の単位の認定のしかたである。われわれは、漢字二字で形成される漢語は、それ以上分割不可能な一語だと見ているが、土居氏はそう見ないで、漢字一字一字を元来のまま一語と見ているように思われる。例

えば、身体を表わす基礎語の候補に「からだ」と「体^{ない}」とを並べ、一般の常識的感覚に反して「体」の方を採用する。その理由は、「からだ」は人間の肉体のみを表わすが、「体」は広く具体的意味も抽象的意味も表わし、「肉体」「団体」「液体」など、多くの熟語を作りうるからだという。この辺にいささかふに落ちない点があるが、意味の分析と合成によって基礎語を割り出そうとする考え方はよくわかる。

Ogden も土居氏も、850語なり 1100 語なりの語それぞれを Basic Word、「基礎語」と呼ぶのであって、それらの語集団を Basic Vocabularyとか「基礎語彙」とか呼んではいけない。だから、ここで、これらを基礎語彙と呼ぶのはよろしくない。わたくしは、ここで、Ogden や土居氏の基礎語をそのまま基礎語彙と呼ぶのではなく、このような「基礎語」的な考え方で選ばれる語の集団を一般的に基礎語彙と名づけておこうと思う。「基礎語」と「基礎語彙」とでは、本当は、意味が非常に違わなければならない。「基礎語」は Ogden や土居氏が考えたように、英語なり日本語なり、一つの言語体系にもとづき、現実の言語体系に手を加えて作った半人工的モデル機構である。その範囲内で、普通のことは一応何でも言い表わせるはずの、一つの完結体であり、小宇宙である。ところが、「基礎語彙」といえば、どうもそういう意味にはなりにくい。ある言語体系の、語彙の面での基礎的部分をさすと思われる。例えばアメリカの言語学者 Swadesh が言語年代学という、言語史の推定法を考案し、これを服部四郎氏が日本に紹介した時、服部氏は「語彙統計学」または「基礎語彙統計学」の名を与えられた。また服部氏は『英語基礎語彙の研究』の書も公にしておられる。ここで言われる「基礎語彙」は、かなり、その名にふさわしいと思われる。基礎語彙統計学とは、発生系統上での類縁関係があると思われる二つの言語がいつごろ枝分かれしたかを推定するための統計学的技法であって、両言語の基礎語彙同士を年代によって比較し、一方の基礎語彙の中に他方の基礎語彙の中の語が何パーセント含まれているかを問題にする。ただしこういう基礎語彙の求め方について私は学んでいないから、くわしい紹介はできないが、服部氏は『日本語の系統』の中で、日本語で 450 語の基礎語彙を指定したことを述べ、将来修正しても 500 を越えることは極めてむずかしいだろうと

述べておられる。

私は、いわゆる基本語彙の最初に基礎語彙をかかげ、その内容に Ogden の基礎英語と土居氏の基礎日本語を当てた。しかも、「基礎語彙」の名はそれにはふさわしくないと、何だか筋の通らない言い方をしているのだが、それは、本稿で私が「これが基礎語彙だ。」というものをかかげる気がないために、基礎語彙をはっきり定義するだけの用意を欠いているからである。

1・2 基本語彙

「基本」という語には、確かに価値的な意味合いが感じられる。「何をするにも、まず則るべきもの」という感じである。スポーツにしても楽器演奏にしても、あらゆる技能には、基本技能といわれる部分があり、技能訓練の中で最も重要視される。だから、最初に述べたある人のように、「新聞基本語彙」ということばに、「新聞記者が記事を書く時にまず準拠すべき語彙」という価値的な意味を感じるのは、当然なことであった。私たち言語関係者は「基本語彙」ということばに慣れており、しかも、これを学問の対象には、あまり、して来なかったために、しばしば口にしながら、特に定義づけは、しないでできた。「基本技能」といえば、大変価値観がこもっていると感じられるが、「基本語彙」には、それほど価値観がこもっているとも感じなかった。しかし、ことばはやはり常識にそって使ったほうがよい。今後「基本語彙」という場合には、「基本」の本義に立って「何かの事を行なうために、第一段階でまず必要とされる語彙」という意味で使うことにしたい。従って、「何かの事」が何であるかによって、基本語彙の性格も内容も規定されることになる。「国語教育基本語彙」といえば、母国語教育の中で教育機関が責任をもってその意味用法に習熟させるべき語彙をさす。「外国人のための日本語基本語彙」なら、外国人が日本語を学習するのに、まず習得させるべき語彙をさす。同じ日本語について基本語彙を言っても、国語教育のためのものと外国人教育のためのものとは、量も質も違って来るだろう。外国人教育の場合でも、その母国語の種類によって違った日本語基本語彙が考えられるかもしれない。

基本語彙は、言語教育の中だけで考えられるわけではない。「数学教育のための基本語彙」などは、言語教育の場合よりも考えやすいだろう。すべて、理

科学は、用語の定義が正確・厳密になされなければならない世界であるから、用語教育に体系が必要で、各分野の学習のための基本語彙も作りやすいと考えられる。また、教育と関係のない所でも、基本語彙はいくらでも考えられる。「新聞記事を書くための基本語彙」「放送スクリプト作成のための基本語彙」「児童読物を書くための基本語彙」など、さまざまな要求があろう。

基本語彙は、特定の目的が見定められてのち、そのための「○○基本語彙」として制定さるべきものである。無限定に「日本語基本語彙」というようなものは、考えることがむずかしい。

1・3 基準語彙

「基本語彙」を上記のように考えると、Ogden の基礎英語も土居氏の基礎日本語も、いずれも基本語彙の一種であることがわかる。両者に共通することは、「実用目的の文章において」ということで、前者では「外国人のため」に重点がかかり、後者では「母国語の一般大衆のため」に重点がかかっているが、ともに教育基本語彙であることは確かである。そして、内容的に、両種の基礎語に共通していることは、語の選択に、論理的な方法がとられていることである。

ここで、三番目にかかげる「基準語彙」も、一種の基本語彙であり、やはり教育基本語彙であろうが、語の選び方で、基礎語とは全く違った考え方がなされる。基礎語では、それで作られる文が世間通用の文とは多少違って、やや不自然で、ぎこちないものになることを覚悟している。それに反して、基準語彙では、あくまでも世間でふつうに行なわれていることばや言い回しを尊重する。ある人がある社会にはいて生活するについて、その社会で最も普通に行なわれていることばや言い方に慣れさせようとする。その社会での特殊な用や、さらに細分化した社会内での用を弁ずることはできないが、一般的な生活は不自由なくできる——というために必要な語彙が、その社会での基準語彙である。

基準語彙は、ある個人の論理的思考によって組み立てられるものでなく、あくまでも実社会での使用実態に即して求められなければならない。従って、語数も千語内外というような少数ですむとは思われない。

基準語彙を選ぶには、二つの観点からの実態調査が必要である。一つは、その社会でどんな語がどのくらい使われているかの調査、もう一つは、その社会に属する人々の間で、それらの語がどのくらい理解されているかの調査である。第一の調査として、語彙調査が始まった。今日までに、語彙調査で最も大きな仕事をしたのは、アメリカの教育心理学者 E. L. Thorndike、およびその協同者たちである。Thorndikeの語彙調査は、全く基準語彙を求めるために行なわれた。日本での語彙調査は、国立国語研究所のものが代表的であるが、後でも述べるように、国語研究所の語彙調査は、雑誌とか新聞とかいうように、その都度単一種の文献について調査しており、一回の調査で、すぐ基準語彙が求められるようにはなっていないので、その点、Thorndikeの語彙調査とは根本的に性格がちがっている。第二の、理解語彙の調査では、わたしたちに近い所に適例がある。阪本一郎氏の調査、国立国語研究所で森岡健二氏が行なった調査、文部省国語課（現在、文化庁所属）での調査などである。

Thorndikeは、Irving Lorge, Michael West などの有力な協力者を得て、1921年以来、4回の大調査を行なった。第1回は聖書や古典など41種の文献について450万語を調べ、その結果から1万語の語彙表を発表した。第2回は児童読物について450万語、第3回は雑誌について、これも450万語を調べた。以上3回の調査では、各語は出現するごとに単純に各1回と数えられていた。例えば mood ということばは、「気分」をさす場合と文法用語での「法」をさす場合とでは全く別の語と考えられるが、その区別はしないで、こみで扱われていた。それでは調べる意味がうすいので、4回目の調査では、mood のような同形異語を区別してかぞえるのは勿論のこと、1語の中の意味のちがいも区別し、それぞれの意味での出現回数を調べた。例えば game という名詞には、「子どもの遊戯」「競技」「勝負」「運動会（複合形で）」の四つのおもな意味がある。調査で得られた game(s) は638個あったが、そのうち38%は競技の意、23%が勝負、9%が子供の遊戯、8%が運動会の意味であった。それ以外は教育的には取り上げるに足りない雑例であった。——というように調査した。このようなかぞえかたを Semantic Count という。この4回目の調査は500万語について行なわれた。Thorndike らは、この4回の調査結果に教育的

配慮を加え、3万語の語彙表を作って公表した。この語彙はそれ以後の英語教育に大きな貢献をした。Thorndikeは何種類もの英語辞典を作ったが、それらの辞書の収録語には、以上の語彙調査に基づいた、語の重要度の評定が20階級に分けて、施してある。

阪本一郎氏は、幼児から成人に至るまでの語彙の発達を調べた点で先駆的の功労がある。昭和13年の調査だが、氏の著『読みと作文の心理』（昭30刊）によってその結果を見ると、満6歳から20歳に至る児童・青年約4万人の調査からの推定値が次のように記されている。

満年齢	保有語彙量
6	5,661
7	6,700
8	7,971
9	10,276
10	13,878
11	19,326
12	25,668
13	31,240
14	36,229
15	40,462
16	43,919
17	46,440
18	47,829
19	48,267
20	48,336

これを今日の教育段階に当てて大まかにまとめてみると、

小学校入学時	6歳	6,000語
小学校卒業時	11歳	20,000語
中学校卒業時	14歳	36,000語
高等学校卒業時	17歳	46,000語
成人期	20歳	48,000語

のようになる。

ここでいう保有語彙とは、ごく条件のゆるいもので、「このことばを知っているか」と聞かれて「知っている」と答えるという程度のもので、理解語彙ともいわれるもので、使用語彙ではない。

これに類する調査で戦後のものとして、昭和25年に国立国語研究所で森岡健二氏が行なったものがある。東京の高校1年生15名について、竹原スタンダードと英辞典の見出し語からサンプリングした刺激語によって理解語彙を調べた結果では、最高3万6千、最低2万3千で、平均保有語彙量は約3万と報告されている。

阪本氏や森岡氏の調査から、日本人の成人の理解語彙量は大体4万語程度であろうと推察される。使用語彙は調査法がむずかしく、まだ調査例に接していないので、日本人の平均使用語彙量がどのくらいのものか、いっこうにわからないが、理解語彙量よりずっと少ないことは確かだろう。4万語よりずっと内

輪のところで何かの境目になるような語彙の区画が見出せないものだろうか。使用語彙を調べるのとは違った所で、それに何かのヒントを与える数字をさがしてみよう。

語彙調査をしてみてわかることはいろいろあるが、その中でいちばん普通に役に立つのは、異なり語数と延べ語数との関係についての情報である。採集された語を頻度順に並べて頻度数を累積させていくと、何語までで頻度総合計の何パーセントを占めるかがわかる。これは、大量の調査をすれば、何を資料にして調べても、大体似た結果になる。研究所のこれまでの雑誌の語彙調査の代表として雑誌九十種の調査結果を見ると、下のようになっている。すなわち、

100語で	32.9%
1,000	60.5
3,000	75.3
5,000	81.7
10,000	91.7
40,000	100

全体で4万語あったのであるが、千語で60%を越え、5千語で80%を越え、1万語では90%を越えるわけである。4万語のうち、上位1万語で90%以上を占め、残り3万語はわずかに8.3%をまかなうにすぎないということは、始めて知った時は、やはり一つの驚きである。

成人の理解語彙が約4万と見積られることと、雑誌九十種の語彙調査で見出された語が約4万であることとの間には、別に必然的関係があるわけではないが、また、全く関係がないともいえないように思われる。現在各社で出版している小型国語辞典の収録語彙は大体五、六万語である。私たちのところで現在進行している新聞の語彙調査で、1紙1年分に当たる量からのサンプル94万語について、異なり語数を短単位に切って整理すると、約4万7千、その中から、記号や無意味な数字などを除くと約3万語になる。こういったところから推定して、私たちの身の回りでふつうに使われている語は、精々かき集めて4万程度かと思われる。してみれば、理解語彙が約4万というのは、大体私たちの周囲にあって、出会っても驚かない語の数を語っているのではあるまいか。そう思えば、理解語彙の4万と語彙調査から得られた語数の4万とは、結局、関係があるということになるかもしれない。

それを関係ありと見れば、さらに一步憶測を進めて、被調査語彙4万のうち上位1万で90%以上をまかなうという事実は、その1万が、私たちの理解語彙

の内側にある使用語彙の範囲と、おぼろげに対応するかもしれないという仮説を立ててみたくなる。

文化庁国語課での理解語彙の調査は、昭和32年以来、一部の小中学生を対象に続けられており、今までに7冊の報告書が出ている。（文部省、国語シリーズ 41, 42, 51, 52, 58, 59, 63, 『児童・生徒の語い力の調査』）この調査では、1万4千余語を調査の土台に使い、その中で、中学生の理解を基準とした基準語に考え得るものとして6,874語が挙げられている。

1. 4 基調語彙

基調語彙ということばを、私は、特定の作品を構成する語彙のうちのある部分と呼ぶ名として用いたい。一作品の語彙調査をすれば、どういう調査をしても必ず多く見出される語が、そこでもまず見出されることは言うまでもないが、そのほかに、その作品において特によく用いられ、その作品に独自の調子を与えていると思われる語の群れが見出されるだろう。そのような、ある作品に独自の存在であり、それゆえ、その作品に特徴を与える働きをする語群を、その作品の基調語彙と呼ぶことができる。源氏物語の語彙調査をした寿岳章子氏は、見出された語彙と作品との関係を考えて、テーマ語彙を名づける一群の語を指定している。これなどに、基調語彙の好例を見ることができる。文学作品の研究や、文体論の研究には、基調語彙の探究が大きな力を発揮するであろう。

1. 5 基幹語彙

この基幹語彙が、私がここで言おうとする主眼点である。

基礎語彙から基準語彙までは、いずれも、指定する一群の語によって、何か大きな働きをさせようとする功利性を、土台にもっている。基幹語彙は、こういう功利性を全く除き去ったところに成立する概念である。

語彙調査は、あくまでも、そこに在るものを調べ、それについて何かがわかるだけのことで、調べなかったもののことは、何もわからない。私たちは、今、昭和41年の朝日・毎日・読売、三紙の語彙を調べている。だから、そこからわかることは、昭和41年の三紙についてだけである。昭和40年のことも42年のこともわからないし、昭和41年でも、三紙以外の新聞のことは、わからな

い。だから、このような語彙調査から、直ちに、何の基本語彙も、基準語彙も、求めることはできない。求めることができるのは、基幹語彙である。

基幹語彙ということばは、まだ聞いたことがない。私の造語のつもりである。基幹語彙とは、ある語集団の中に、その集団の骨格のような部分として、その集団をささえる基幹的部分として、現に存在する、語の部分集団を呼ぶ。基幹語彙が基礎語彙から基準語彙までのものと大いに違うところは、要請によって空に描き出される仮設的存在ではなく、現に実在する実体であることである。では、どのような実体か、それは、以下の実例によって明らかにしていこう。

2. 基幹語彙の求め方のテスト

国語研究所は、これまでに何度も語彙調査を行なって来た。その結果のほんの一部を使って、基幹語彙を求める手続きをたどってみたい。基幹語彙についての私の考え方は至って単純であって、一つの語彙調査をたてに見るとともに、いくつもの語彙調査を横に見て、どこでもよく使われている語ほど、全資料内で、それだけ基幹的存在だと考えるのである。

研究所のこれまでの語彙調査は、試験的なものを除いて5種類ある。

- 1) 明治初期文献の語彙調査（明治10～20年）
- 2) 婦人雑誌の語彙調査（昭和25年）
- 3) 総合雑誌の語彙調査（昭和28～29年）
- 4) 雑誌九十種の語彙調査（昭和31年）
- 5) 新聞の語彙調査（昭和41年）

五番目の新聞の調査は、電子計算機を用いて現在進行中のもので、規模は最も大きい、まだ最終結果が出ていない。最近、三分の一の量を処理し、ある形での語彙表を得たので、これを材料に使う。四番目の雑誌九十種の調査は、それまでの調査の中で最も規模が大きく統計処理の方法は、すべてを通じて最も整備されたものであるが、雑誌を広くもうらしている、今このテストでは、婦人雑誌と総合雑誌とに、性格の違う二つのものを代表させ、中間的になる九十種調査は使わないことにする。

明治初期文献の調査は、明治10～11年の郵便報知新聞の調査を中心とし、それに『近世事情』『国体新論』『ミル経済論』等、硬い文体の文献23種、『読売新聞』『東京絵入新聞』『交易問答』『安愚楽鍋』という軟かい文体の新聞や読み物を加えた調査である。（以下、これを明治調査という。）明治調査は、あとからいろいろつけ足した寄せ集めの調査で、単一の調査ではないから、そこで得られた語彙を全体として度数順に並べることができない。そこで、明治調査については、私が以前近代語研究室長をしていたころに、専ら出典文献の幅（range）をたよりに、順位づけに準じたグループ分けを、上位についてだけ施したものがあるので、これを順位づけの代用として用いる。

婦人雑誌の調査は『主婦之友』を主体にし、『婦人生活』で補充した調査であるが、語彙表は『主婦之友』によって順位づけられているので、それをそのまま使う。総合雑誌の調査は、『世界』『中央公論』などを調べたものである。

明治調査、婦人雑誌調査、総合雑誌調査、新聞調査、この4種の語彙調査結果をそれぞれ度数順に並べ、それぞれ、100番目ごとに区切りながら500番目まで採る。500に意味があるのではない。ひとりで机上で簡単に扱える数として、500ずつ、計2000の単語に限ってみた。さて、これら各500語は、お互いにダブリがある。ダブリかたの甚だしいものほど使用の幅の広い語であり、ダブリった中で等級の高いものほど基幹度が高いものと考ええる。

まず、ダブリの度合が最も甚だしく、4種調査ことごとくにまたがっているもの、つまり、4種の語彙表でことごとく500位以内にあるものを基幹度最高の語と認める。これが48語あった。この48語をさらに等級分けするために、100位以内の語に1点、101～200位の語に2点、201～300位の語に3点、301～400位の語に4点、401～500位の語に5点を与え、合計点数の低い順に並べると、4種調査でことごとく1点のものが4点で最高位に位置づけられ、以下、 $5 \times 4 = 20$ で、20点まで分布することになる。実際は次のとおりであった。

4点 いう、こと、この、これ、その、ため、とき、ところ、なる、また、
もの、よる

5点 思う、それ

6点 今、前

7点 以上, うち, 同じ, 対す, 見る

8点 ある, 手, 出来る, 場合

9点 あと, 家, 後, ここ, さらに, しかし, ほか, わたくし

10点 行く, 初め, 目, 中

11点 持つ

12点 仕事, 十分 (充分), 次, とともに, なお

13点 気, 開く

15点 一般, 形, 強い

これを仮りに三つのグループに分けてみよう。4点のものを最高第1群, 5

	最高第1群 (4種とも100位以内)	最高第2群	最高第3群
名 詞	こと (事) もの (者, 物) とき (時) ところ (処, 所) ため (為)	手 いま (今) まえ (前) うち (内, 中) 以上 場合	目 (め) 家 (いえ) 形 (かたち) 仕事 一般 はじめ (初, 始) つき (次) なか (中) あと (後, 跡) 後 (ご) ほか 気
(代 名 詞 指 示 語)	これ この その	それ	ここ わたくし
動 詞	いう (言) なる (成) よる (依, 拠) 〔よって〕	ある (有, 在) 思う 見る できる 対する 〔一対して〕	もつ (持) いく・ゆく (行) 開く
形 容 詞		同じ	強い
(副 詞 接 続 詞)	また		なお, さら (に) ともに, しかし, 十 (充) 分

点から8点までのものを最高第2群、以下を最高第3群とする。そして、品詞ごとにまとめてみると、次のようになる。(この際、表記のちがいは、すべて無視する。)

最高第1群には12語が属しているが、これらの語を見ていると、私は、どうしても、ここに私たちの認識や思考の活動の源泉とも言うべく、精神活動を煎じつめて原型にまで戻したような、極めて根源的な言語形式を見出さずにはいられない。名詞の「こと」「もの」「とき」「ところ」「ため」は、いずれも形式名詞で、意味内容はほとんど無く、詞の中で最も辞に近い存在であり、その点では、個性のない、つまらない語のようでもある。こ・そ・あ・どの指示語や接続詞が辞的な詞であることはいうまでもない。動詞の「いう」「なる」「よる」も、ふつうの動詞とはちがう。「いう」は、ほとんどが、物を言う意味の「言う」ではなくて「AというB」の形で使われたもの。「なる」は「～になる」「～となる」「～くなる」のように使われた補助動詞的なもの。「よる」は多くが「～によって」で、複合して助詞化したものである。つまり、この3動詞も辞的な性格の語である。してみると、以上12語は、いずれも辞的な詞だということができ、それゆえに、かくも流布の度合いが高いのだといえる。

しかし、それだけではない。辞ならば、文法形式を表現するための語として、語彙論的な意味を表わさない語として、英語でいう function word として片づけてしまえばそれまでなのだが、この12語は、それとはちがう。これらはやはり、意味内容がある。「こと」や「もの」は、すべての認識の対象を、そのどちらかで言い表わし得ることばだし、「とき」「ところ」は、すべてのものごとがその中にしか存在し得ない時間と空間を言い表わしている。

第2群・第3群を見よう。第2群に「ある」があるが、これは本当は第1群に入るべきことばである。明治調査の対象が大部分文語文であるために、そこでは文語の「あり」が最高ランクにあり、口語の「ある」は401～500位の中にある。それで第1群からはずれて第2群に入ったのだから、「あり」を「ある」と同一性質のものと見なせば、もちろん第1群に入ることになる。そのように扱うべきである。第2群・第3群の名詞を見ると、「手」「目」「家」

「形」などの語があるが、これらは「もの」の最も身近な具現なのではあるまいか。「仕事」は「こと」の具現である。「一般」もそう見られるかもしれない。「いま」「まえ」「あと」「後」「うち」「なか」などは、「とき」または「ところ」の具現である。

こうして見てくると、第1群を最高抽象レベルでの認識の原型として、以下、次第にそれが特殊化し、具体化し、分化して、たくさんのことばを生んで行くような、発生系統的な一つの姿がここに見出されるような気がするのである。それを勇敢に私の哲学で位置づけてみると、次のような表ができる。

存在実体を認知する行為				
{	認知された対象を表す・	こと、もの	形, 手, 目, 家, 仕事, (一般)	
	対象の存在を表わす・	ある	もつ	
	対象の変化を表わす・	なる	いく, 開く	
存在形式を把握する行為				
{	様式をとらえる {	時間として・	とき	いま, 前, あと, 後
		空間として・	ところ	なか, うち, (前)
{	関係をとらえる {	同異関係・	いう	同じ
		因果・依存関係・	ため よる (より, よって)	
		序列関係・	はじめ, つぎ	
		比較・程度関係・	以上, 対し (さらに), 充分, 強い	
		集合・領域関係・	ほか, とともに, (以上) (うち) (なか)	
{	条件をとらえる・	場合, (とき), 一般		
存在の延長や拡大を認知する行為・		また	なお, さらに, しかし	
存在に対処する行為				
{	自他の区別・		わたくし	
	方向づけ一記号代置・	これ, この, その	ここ, それ	
判断する行為・		思う, 見る, (同じ)	できる, 気	

認知の対象と、その存在および変化を表わす「こと」「もの」, 「ある」, 「なる」によって,

〔ナニ〕は〔ナニ〕である。 〔ナニ〕が〔ナニ〕になる。

という認識が成り立つ。これは、私たちの状況把握の最初だといえる。これに「とき」「ところ」が加わると,

〔ナニ〕が〔ドコ〕にある。 〔ナニ〕が〔イツ〕〔ナニ〕になる。

の表現ができる。対象, 存在, 変化, 時間, 空間の認識は, 思考の始まりを約

束する。これに関係把握がそうと、思考に論理性が加わる。「AというB」の形は、AとBの同一認識（Identification）を表わす。同一認識は「AはBである。」がその完結した表現だが、「AというB」の形式が、それを随所に成り立たせる。「AはBのため」は、AがBに依存したり、BがAの原因や理由であったり目的であったりする関係を表わす。「AはBによる」「BによってA」はAがBに強く依存することを表わす。因果・依存の関係把握は事実の中に法則を見出したり、異なる事象間の関連を見出すのに極めて重要な拠り所となる。「はじめ」「つぎ」によって表わされる序列関係の把握は、混乱した事象を整頓するのにまず手がかりを与える。「以上」「対し」「充分」「強い」を比較・程度関係という概念でくくったのは、やや無理な点もあるが、「以上」がAとBとの上下関係を表わすこと、「対し」がAとBとの対比関係を表わすことには異論があるまい。「充分」はいわゆる程度副詞で、比較とは違った発想による程度表現をする。「強い」がどういう用法でこれだけ上位に位置するのか、調べてみないとわからないが、便利な強調用語なのではあるまいか。「ほか」「ともに」には、ある集合を作って、その領域の内か外かを問題にする意識がある。こういう集合とか領域とかの概念は数学的な考え方に発展していくもので、見のがされやすいが、意外に重要でかつ高度な論理の基礎概念である。「場合」は条件設定になくはならぬことば、「一般」は抽象の所産で、これがないと、法則の記述ができない。

「また」を先頭にした、「なお」「さらに」「しかし」等の接続詞は、一つの認識を終り、そこで終りきらないで、次へ認識が延びていくことを表わすことばで、私たちはこれを知的領域をひろげていくための武器として使う。

自己をとらえる「わたくし」は、自他の区別の表われであり、「こ」「その」つくことばは、「これ」「それ」が代名詞と呼ばれるように、すべてのものを共通の記号に置きかえる働きをするが、それは自己を離れて客観的にするのでなく、すべてを自己との関係でとらえることによって、単純な記号代置を可能にするのである。代名詞や指示語は、自己が対象に身をもって対処する手段としてあるところに、認識の原型の一つとして意味がある。

「思う」以下、一群の動詞は判断を担当する語としてとらえられよう。

「気」という一字だけを見ていると、何でこの語がこんなに高い所に位置するのか、そして、どうしてこの語を動詞に分類したのか、ふしぎな感じがしようが、これは「気体」の意味の「気」ではなくて「心」の意味の「気」であり、「気がする」「気がつく」「気をつける」「気をもむ」「気になる」「気に入る」など、連語で一つの動詞相当句を作るものである。そう考えてみると、

	明治を含む3種間のまたがり (62語)	現代だけの3種間のまたがり (56語)
名 詞	・まま, ほど, 風(ふう), 方(ほう), 元(もと), 通り ・日(ひ), 頃, 毎日 ・上, 間(あいだ), 他(た), 余り, 前後, 全体, 場所 ・ひと, 人間, 男, 女, 子, 身, 姿, 心, 口, 声, 力 ・国, 政府, 会社, 生活, 自由 ・山, 水, 地	・方(かた), 点, わけ, 程度 ・時間 ・下, 中心 ・顔 ・必要
数 詞		・一, 二, 三, 四, 五, 六, 七, 八, 九, 十, 十一, 二十, 一つ, 二つ 二回
代 名 詞 指 示 語	・われわれ, 彼 ・なに, いずれ	・自分, あなた, こう, そう
動 詞	・行なう, 起きる, 掛かる, おく, 切る ・作る, 買う, 取る ・聞く, 知る ・帰る, 当たる	・する, いる, おる ・はいる, 入れる, 出す, しまう, 着く, つける ・上げる, 下さる, 受ける, 得(え)る ・来る ・考える
形 容 詞		・ない, 大きい, 多い, よい, 早い, 新しい
副 詞	・必らず, もちろん, ただ ・すこし, たくさん, 最も ・それぞれ ・すでに, 今度 ・例えば	・まず, すぐ, もう, まだ ・特に, すべて
接 続 詞		・そして
連 体 詞	・或る	

「氣」で作られる連語動詞の何と多いことか。これらの連語を判断行為の表現と見ていかどうかは保障の限りでないが、「思う」の分化した形をここに見ることは許されるかと思う。

全体に、この表の分類は私の独断によるものであり、知的遊びの所産であるから、厳密な意味はどこにももっていない。おぼろげに、こんな構造が想像できるのではないかという程度の提出物である。

4種の500語間のダブリを、さらに条件をゆるめて観察してみよう。4種のうち3種にまたがる語は、118あったが、その3種を、明治を含む3種と現代だけの3種とに分けてリストしてみよう。明治を含む3種間にまたがるものは62語、現代だけの3種にまたがるものは56語であった。

品詞別した中を、黒丸でくくったのは、意味のうえで何となく同類と思われるものを集めたのだが、あくまでも「何となく」で、精密なものではない。明治を含む3種間にまたがるものは、明治初期から今日に至るまで変わらずに、国語の基幹部を構成している語と見られ、現代だけの3種にまたがるものは、それだけ現代性の強い基幹語彙だと見ることができだろう。この表で目立つことは、数詞と形容詞とが現代だけの方に片寄って在ることだ。数詞がばかにきれいに現代の方にそろっているのはどういうわけだろう。数詞の語形が現代において安定したということか、数値の使用度が特に現代に高いということか、はっきりした説明法は、思いつかない。形容詞が現代だけにある理由は明瞭だ。さきの「あり」と同じことで、文語から口語への移行行きを示している。これと同じことは、動詞の「する」「おる」にもいえる。文体の変遷がなければ、「する」のごときは全く国語の基幹語中の基幹語でなければならない。

そのほか、右側の語群の中で、「点」「程度」「時間」「必要」「考える」「もう」「特に」「そして」などのことばに、そう言われれば確かに現代性が感じられるが、それ以外では、左側の語と右側の語との間に、そう目立った性格の違いは感じられない。

またがりの条件をさらにゆるくしてみよう。2種間のまたがりについて、これも、明治を含む2種と現代だけの2種とに分けて示すと、次のようになる。

こちらの表になると、左側（明治を含む）の語群と右側（現代だけ）の語群

との間に、かなりはっきりした違いが見え、現代の方に、それらしいローカル性が見られるではないか。準体助詞の「の」とは、「いものにえたのご存じないか。」の下線の方の「の」で、形式名詞「こと」に近いもの、この語は確かに現代語の文体を特徴づけるに足る。

	明治を含む2種間のまたがり (76語)	現代だけの2種間のまたがり (89語)
名 詞	・はず, よう (様), 同様, 限り, 実, 名, 色, 数, 図, からだ (体), 書 (しょ) ・のち, 一時, 今日, 今年, 時代, 当時, さき, 月, 最初 ・右, 左, 本 (もと), 皆 ・両, 油, 米, 金 (かね), 金 (きん) ・国家, 権, 全国, 外国, 人々, 人民, 土地, 主人, 夫 (おっと), 新聞	・の (準体助詞), 間 (かん), 面 ・日 (にち), 昔, 朝, 現在, 最近, 中 (ちゅう), 年, 昭和 ・頭, 白 ・世界, 国民, 社会, 学校, 大学, 学生, 代表, 家庭, 子供, 人 (じん) ・日本, アメリカ, 米国, ソ連, 中国, 東京 ・関係, 理由, 結果, 相当, 一方, 普通, 事件, 問題, 計画, 方法 ・意見, 気持, 感じ, 態度 ・映画
コ・ソ・ア・ド	・これら, どこ, こんな	・だれ, どう, そんな, どんな
動 詞	・なす, あり, 致す ・待つ, 向かう, 入 (い) る, 経 (た) つ, 続く, 結ぶ, 去る, 絶 (裁) つ ・好む, 飲む, 用いる, 引く ・示す, 読む, 呼ぶ, 論ず	・与える, やる, もらう, くれる ・始める, 立てる, つづける, 掛ける, 出る, 変わる, つく, 合 (会) う ・歩く, 働く, 使う, 加える ・見える, 認める, わかる, 言える, 決める, 調べる, 書 (描) く
形 容 詞		・高い, 長い, 深い, 小さい, 少ない, 正しい, わるい
形容動詞	・大切	・明らかな, 非常, いっしょ
副 詞	・あるいは, 決して, 殆ど, 全く, 少々, もし ・即ち, 以て, よく ・はじめて, 以来	・ちょっと, やはり, しかも ・いつ, いろいろ
接 続 詞		・および
連 体 詞	・いわゆる, いかなる, 同 (どう)	・各, 約

右側の名詞の下半分（「世界」以下）の語は、極めてはっきりと、現代共通の話題や、現代的な問題把握法を示している。左側の「国家」や「人民」や「主人」に対して、右側には「世界」「社会」，「国民」，「家庭」の語がある。右側の「学校」「大学」「学生」は、ヤング・パワーの時代を示すものであろうか。国名の「アメリカ」「米国」「ソ連」「中国」は、あまりにもはっきりと現代の国際感覚を示している。「関係」「結果」「問題」「方法」などは、ほとんど私たちの物の見方や考え方を規定していることばで、もし、これらのことばを封じられたら、現代の教養ある人々は、ものを考えるのに、大きな支障をきたすにちがいない。

以上、4種の語彙調査の上位500語について、2種以上のグループにまたがって存在する語を取り出してみた。これは、多少とも幅広く用いられた語をさがしたのである。次に、反対に、ある所に片寄って使われた語をさがしてみよう。どれか1種類の調査では100位以内に入っているが、他の3種では、いずれも500位以内に見えないという語をより出してみると、次のようになる。

明治調査で100位以内にあり、他で500位以内でないもの（30語）

- （名 詞） 法，故（ゆえ），世
- （動 詞） 能ふ，出づ，曰く，受く，失ふ，及ぶ，従ふ，す（為），付く，謀る，用ふ，許
- （形 容 詞） 多し，無し，かくのごとし
- （副 詞） 未だ，大いに，凡そ，却て，実に，総て，遂に，常に，自から
- （接 続 詞） 而して
- （連 体 詞） 彼の，我が

婦人雑誌調査で100位以内にあり、他で500位以内でないもの（25語）

- （名 詞） 昼，夕，表，裏，幅，中央，芯，作り方，袖，衿，鎖，細編，縫代，身頃，前身頃，後身頃
- （数 詞） 一因，二目，一杯，一センチ，一・五
- （動 詞） いただく，合わせる，編む，縫う

総合雑誌調査で100位以内にあり、他で500位以内でないもの(22語)

(名 詞) 政治, 経済, 戦争, 平和, 主義, 労働, 僕

(造語要素) 月, 会, 者, 人, 第, 党, 的

(接 頭 語) 御

(接 尾 語) さん(敬称), たち, ら(等)

(助 詞 的) 於て

(数 詞) 千, 萬

(動 詞) される

新聞調査で100位以内にあり、他で500位以内でないもの(20語)

(名 詞) 午前, 午後, 東京都, 面接, 面談, 優遇, 不問, 社保, 委託, 株

(略 記) 給, 歴, 持, 通, 住, 上, 歩, 建, 完, 同, 新

明治調査だけに特徴的な語というのは、ほとんど全部が、文語文——それも、漢文書き下し式の硬い文語文——の文体をささえている語である。婦人雑誌だけに特徴的なのは、全く裁縫用語、その中にわずかに「一杯」「いただく」など料理用語がまじる。「昼」「夕」は料理的なものか、あるいはもっと幅のある生活用語的なものか、よくわからないが、後者のように感じられる。総合雑誌だけに特徴的な語には、名詞に「政治」「経済」以下、まことに政治経済的、文化評論的な用語が見られる。ここにだけ、「第」「者」「的」など造語要素がずらりと並んでいるが、これは、内容の問題ではなく、この時の調査から、語認定の単位にβ単位と称する短かい単位を採用したための結果である。新聞調査だけに特徴的な語には、一目でそれとわかる案内広告専用の語が並んだ。これらのことばが新聞用語の端的な特徴を示すと言ったら、おかしいことになる。そういう結果しか出せない調査だったら、困った調査だということになるが、そんな心配は、いらない。心配のいらないわけを、次に示そう。

3. 新聞語彙調査から求められる新聞基幹語彙

今までの記述で、「基幹語彙」が、調査された言語資料の中で、幅広く用いられ、かつ、どの方面でも高い頻度で用いられている語の群れをさしていることがわかったであろう。「基本語彙」が無限定には言えないように、基幹語彙も無限定には言えない。ただし、限定する条件の性格が、基本語彙と基幹語彙とでは、まるで違う。基本語彙の場合は、目的と用途とが限定の条件であったが、基幹語彙の限定条件は語の属する集団の範囲である。基本語彙の条件は幾分心理的であるが、基幹語彙の条件は専ら物理的である。昭和25年1年間の「主婦之友」に記載されている語であるとか、A社の小学校国語教科書全学年に含まれる語であるとかいうふうに、語集団の範囲がはっきりしさえすれば、その中から基幹語彙を見出すことができる。あとは、基幹度の測定法をどう立て、基幹度の評価水準をどこに置くかという技術的問題が残るだけである。

私は、どの方面にも厚く行きわたっている語ほど基幹度の高い語だとする考え方をとるから、ある語集団から基幹語彙を求めるためには、その語集団が何らかの意味で方面に分割できなければならない。その分割原理は、なるべくその語集団の存立をささえる構造の組成原理にそっていることが望ましい。統計学における層別と同じことである。構造を無視して集団成員の数で等量に分割していく方法もあるが、ここではその方法をとらない。

このたびの新聞語彙調査では、データを、四種の分類原理によって区分している。このことは、この『電子計算機による国語研究』報告シリーズの第2集（昭44）の論文「新聞語彙調査における層別とその意味」にくわしく記したので、ここに重ねては述べないが、四種とは、話題別（12区分）、文種別（17区分）、位置別（8区分）、署名別（10区分）である。この区分を、基幹語彙探求の手がかりに使いたい。四種のどれも役に立つが、常識的に、用語は最も話題に左右されやすいものと思われるから、ここでは、話題別の12区分を使うことにする。これらの区分を、私たちは便宜上「層別」と呼んできたので、サンプリング法における「層別」とは多少意味がちがうが、以下、ここでも層別と称することがある。

新聞語彙調査の層別区分を使ってデータの範囲内での新聞基幹語彙を求める方法については、上記報告書シリーズ第1集（昭43）の論文「新聞語彙調査の

概略と語彙分析法試案」に述べたので、これも、重ねて詳説はしない。ただし、そこでは、私はまだ「基幹語彙」ということばは用いていない。その書の33ページで「日本語の基幹をなす語」「基幹的形式語」などの語を用いているにとどまる。

その論文で材料に使ったデータは、私たちの電子計算機による最初のテスト・ランのアウトプット、延べ2万語をまかなう、異なる7500余語、その中の度数6以上の287語という、データともいえない少数の語であった。今度材料に使うのは、今回の語彙調査の全データの三分の一に当たる長単位延べ68万をまかなう、異なる長単位約10万、そのうち度数10以上の6411語、その中から、さらに、記号類や無意味な数字を除いた5417語（長単位）である。

材料の数は大分ふえたが、これもまだ正式に使える材料ではない。私たちの調査単位の「長単位」「短単位」については、さきの報告書『電子計算機による新聞の語彙調査』にくわしい解説があるので、ここでは述べないが、要するに、長単位とは、記事の中の文章を文節に切り、文節の中を自立語の部分と付属語の部分とに切断したものである。例えば、次のような文は、

新五カ年計画／の／もと／では／、／小麦輸入／は／考え／られない／。
と切れ、切ったままで、何の手も加えられないから、「新五カ年計画」や「小麦輸入」のような複合した自立語はもちろんのこと、「では」「られない」のような付属語の連続もそのまま1語とかぞえられるし、「考え」は動詞の活用した形も名詞といっしょにされてしまい、動詞の「考え」が「考える」に合算されることがない。これらは、後に短単位で処理されるときに、かなり救われて、もっと扱いやすくなるが、今は長単位のままのデータを用いる。従って、語の数を正面から問題にされると、困ることがたくさんある。前の論文の時よりも、データの数は一躍的にふえたものの、データの性質は、この通りであるから、この基幹語彙も、個々の語を問題にしているのではなく、やはり、基幹語彙の求め方を問題にしている。その点ではこの論文も、前の論文と同じ性格のものである。

話題別による12区分を手がかりにし、上述した長単位、度数10以上の5417語の中から基幹語彙を求めてみよう。

話題分野区分			異なり数	延べ数	延べ%	話題分野区分			異なり数	延べ数	延べ%
1	政	治	11,311	48,674	7.2	8	地	方	3,545	9,523	1.4
2	外	交	2,382	7,169	1.1	9	ス	ポーツ	9,870	47,815	7.0
3	経	済	12,129	83,550	12.3	10	婦	人家庭	6,910	25,896	3.8
4	労	働	1,731	4,569	0.7	11	芸	能	15,760	68,714	10.1
5	社	会	20,966	93,889	13.8	12	広	告	40,424	192,738	28.4
6	国	際	10,705	48,466	7.1	全 体			101,081	679,342	100
7	文	化	12,452	48,339	7.1						

〔層別の度数分布と層別度数による6階級区分〕

使用度数	政治	外交	経済	労働	社会	国際	文化	地方	スポーツ	婦人家庭	芸能	広告
500以上	9		7		12	10	9	1	9	7	9	20
499～400	1	1	2		4				1		1	13
399～300	3		1		2	2	3			1		8
299～200	5	3	2	1	5	5	5	2	1	2	3	27
199～100	9	5	8	4	22	12	10	6	5	9	9	74
99～80	5		5	2	11	4	4		8	1	6	32
79～60	15	1	4	2	25	11	14	1	12	7	11	62
59～40	19	5	18	1	38	36	24	5	24	11	27	126
39～30	29	2	27	2	65	24	19	2	21	19	31	125
29～20	67	4	66	5	130	68	62	12	62	28	80	306
19～15	74	13	115	2	123	76	66	11	71	24	66	254
14～10	167	22	632	11	275	155	128	19	127	80	197	587
9	50	6	62	1	112	54	46	4	40	25	51	130
8	64	6	81	4	126	65	60	8	55	39	65	119
7	82	9	79	11	134	91	67	11	63	43	69	123
6	124	23	81	14	197	107	111	21	64	57	82	162
5	149	34	113	12	243	123	164	26	94	98	110	209
4	218	61	158	22	333	168	217	60	124	126	171	232
3	312	96	211	61	444	285	350	97	199	232	273	344
2	462	173	367	152	548	403	621	252	378	391	495	466
1	771	581	641	496	681	731	916	743	698	811	784	646
計	2605	1045	2680	803	3530	2430	2896	1285	2056	2011	2541	4065

まず、68万長単位全体での層別語数を、延べ語数と異なり語数とについて、一覧する。

広告が最大の集団で、社会、経済、芸能がこれに次ぎ、地方、外交、労働は極度に小さい集団である。

次に、度数10以上の5417語に範囲を限り、それらの語の12層各層内での度数分布がどうなっているかを見よう。度数の大きい方は段階にくくりながら、それぞれの度数をもつ語が何語あるかを一覧表にする。

政治欄では、度数 500 以上の語が 9 個、度数 499 ～ 400 の語が 1 個……となっていて、同欄内での度数 2 の語は 462 個、度数 1 の語は 771 個あるわけである。そして、政治欄だけで見ると、その中に現われた語の異なりの数は2605である。労働記事などは量が少ないから、いちばん度数の多い語でも度数 200 台である。

この表をたてに見ると、どの層にも、5本の線による仕切りが入っており、これによって6つのグループに区切られている。各層に属する語を度数段階によって6階級に分けたのである。層ごとに度数分布状況が違うから、区分のしかたも、それに従って違えてある。政治記事の層では、度数 100 以上を最高グループとし、これを階級 6 とする。度数99から30までが次のグループで、これを階級 5 とする。度数29から15までを階級 4，度数14から 7 までを階級 3，度数 6 から 3 までを階級 2，度数 2 と 1 を階級 1 とする。外交記事の層なら、度数60以上を階級 6 とし、度数 1 だけを階級 1 とするわけである。

上の表を階級別の表に作り変えて、各階級に属する語の数を一覧表にすると、次のようになる。

〔層別度数階級と各階級所属語数〕

層別度数階級	政治	外交	経済	労働	社会	国際	文化	地方	スポーツ	婦人家庭	芸能	広告
6	27	10	29	9	45	29	31	15	24	20	28	41
5	68	46	111	26	139	75	119	48	57	89	69	195
4	141	44	115	37	253	144	194	66	133	105	146	251
3	363	191	632	83	513	274	284	157	359	363	382	560
2	803	173	416	152	907	774	731	252	417	623	636	1126
1	1233	581	1377	492	1673	1134	1537	743	1076	811	1279	1897
計	2605	1045	2680	803	3530	2430	2896	1285	2056	2011	2541	4065

各語に、自分の階級の数をそのまま点数として与えると、階級 6 の語は 6 点となり、階級 1 の語は 1 点となる。そうすると、語彙表中の各語は、層ごとに点をもつわけで、その点の最高は 6，最低は 0（ある語がある層に存在しなけ

れば、その層での点は0)である。従って、語ごとに点数を合計すると、最高は、12層ことごとくで6点をとった場合の72点、最低はどこか1層だけに存在し、その点が1である場合の1点である。(実際には、総度数10以上の語ばかりを材料にしているから、1点の語はあり得ない。)

語がいくつの層にわたって出現するかによって、語が使われる幅の「広さ」を評定する。4段階に分け、12層から10層までを「極めて広い」、9層から7層までを「かなり広い」、6層から3層までを「中位」、2層と1層を「狭い」とする。広さに対して、各層の中での出現頻度を、語の使われ方の「深さ」とし、層内度数から導き出した層内階級点数の合計で深さを評定する。層数12の語についていえば、72点から37点までを「深い」、36点から25点までを「中位」、24点から12点までを「浅い」とする。広さに伴っての深さの評定基準を下の左側の表のように与える。(基準点数の割り出し方は、語の分布とにらみ合わせながら、適当に、ある係数を設けたのであるが、述べれば、いたずらにくどくなるから述べない。)右側の表は、5417語が、「極めて広く」て「深

		階級合計点数の評定基準			〔広さと深さのかけ合わせによる 12区画と所属語数〕					
またがる層の数		1. 深い	2. 中位	3. 浅い	深 さ			計		
					1. 深い	2. 中位	3. 浅い			
A 極めて広い	12	72—37	36—25	24—12	広	A 極めて広い	A 1	A 2	A 3	589
	11	66—34	33—23	22—11			162	229	198	
	10	60—31	30—21	20—10						
B かなり広い	9	54—27	26—16	15—9	さ	B かなり広い	B 1	B 2	B 3	1181
	8	48—24	23—14	13—8			10	405	766	
	7	42—21	20—12	11—7						
C 中位	6	36—13	12—10	9—6		C 中位	C 1	C 2	C 3	2123
	5	30—11	10—9	8—5			213	580	1330	
	4	24—9	8—7	6—4						
D 狭い	3	18—7	6—5	4, 3		D 狭い	D 1	D 2	D 3	1524
	2	12—5	4	3, 2			987	409	128	
	1	6—3	2	1						
										5417

い」A1から、「狭く」て「浅い」D3までの12区画にどう分布するかを示したものである。

この表は、広さの段階が中間的であるB、Cに属する語は、とにかく「浅い」ものばかりであって、中間的な広さをもちつつ「深い」語は非常に少ないことを示している。

右側の表で、A1、A2、B1に属する語が基幹度の高い語であることは、疑う余地がない。A3の、浅くても幅広く存在するものは、やはり基幹度が高いと見るべきだろうし、B2の実力も相当高く評価してよいと思われるので、A1からB2までの5区画に属する語は、今回の資料の範囲内で文句なしの基幹語彙と指定することができよう。それらの語を、以下に示す。

新 聞 基 幹 語 彙

(昭和41年3紙の場合)

1) 極めて幅が広く、深さも深いもの(A1)

〔名 詞〕 こと もの ため とき ところ 方 点 わけ ほど ◆ 前 以上
ほか 中 次 一方 上 他 あと 中心 はじめ 間 後 ◆ いま
午前 午後 昨年 夜 現在 最近 ◆ 私 人 手 話 問題 場合
考え 結果 必要 政府 世界 日 一部 ◆ 東京 日本 昭和
アメリカ

〔数 詞〕 一 二 三

〔コソアド〕 この その これ それ どう

〔動 詞〕 いる ある いう なる する つく よる いく できる
対する 出る 開く かける くる 見る とる

〔形 容 詞〕 ない(助動詞も含む) 多い 同じ 強い いい

〔連 体 詞〕 大きな 約 同

〔副 詞〕 さらに よく とくに

〔接 続 詞〕 また しかし

2) 極めて幅が広く、深さ中位のもの(A2)

〔名 詞〕 なか 通り 時 右 左 下 北 中央 まま 限り 程度 ◆ 一時

ことし 朝 今後 将来 今回 つぎ 年 水 金 同日 時代 春
 秋 夏 最後 時間 当時 ◆ 意見 予定 意味 頭 道 全国 自分
 発表 仕事 感じ 気 半 動き 声 内容 姿 態度 形 計画 普通
 原因 車 代表 調べ 土地 国 関係 方針 一般 理由 顔 社会
 各地 建設 高校 立場 例 調査 方法 期待 名 ◆ 大阪 東京都
 ソ連 米 ◆ われわれ

〔数 詞〕 四 六

〔コソアド〕 ここ そう そこ こんな どこ これら

〔動 詞〕 いえる みる 出す 行なう まとめる 認める 受ける 思う つける
 もつ 持つ あげる 発表する やる 決める 加える くれる ふえる
 得る 語る 来る 知る 関する 求める わかる 集める 進める

〔形容詞〕 近い ほしい 新しい 大きい 若い 早い

〔連体詞〕 さる (に) おける

〔副 詞〕 もう かなり それぞれ すべて いずれ もっとも 最も こんど
 これまで あまり ともに もと もちろん ただ ほとんど やはり
 はじめて つまり

〔接統詞〕 なお そして しかも

3) 極めて幅が広く、深さは浅いもの (A3)

〔名 詞〕 者 以下 以外 ごろ ◆ 別 全部 全体 延長 共通 発展 途中
 ◆ 第一 過去 時期 現状 末 第一回 今年 毎年 季節 ◆ 効果
 当然 現地 準備 条件 戦後 会長 最大 希望 足 影響 人たち
 傾向 わずか 変化 性格 成功 経験 人気 直接 教育 新聞 目標
 町 質問 農業 可能性 危険 山 状態 成果 収入 あて 結論
 参加 場 不安 グループ 資金 不満 教え 特徴 カギ 運動 調整
 関心 向上 実現 恐れ 数字 不足 訪れ

〔動 詞〕 続ける 違う 入れる 行く 出る すぎる 迎える しまう はいる
 注目する しれる つくる 設ける かける 見せる 聞く 使う おく
 立つ やめる 含める 成功する 示す 終る 許す 残す 合う
 生かす 捨てる 立てる 起きる つづく

〔形容詞〕 少ない 悪い 高い 長い 激しい むずかしい

〔連体詞〕 各 いわゆる 去る このような

〔副詞〕 再び はっきり とにかく あるいは これから いつも とても
なかなか それだけ その後 それほど 具体的に たとえ

〔接統詞〕 そこで それに

4) かなり幅が広く、深さが深いもの (B1)

〔名詞〕 写真 男 女 家 映画 女性 ◆ 彼

〔副詞〕 すでに すぐ

5) かなり幅が広く、深さ中位のもの (B2)

〔名詞〕 以降 場所 はず ほう のち 先 面 以来 今 内 際 幅 初
最初 以内 以前 毎 日 月 南 東 最高 前後 付近 身 色 単位
部分 差 始め 高さ 昔 年 令 大幅 名前 うえ あす ◆ 発
学生 解説 生活 バス 大学 人間 会社 テレビ 自民党 社会党
今日 事件 子ども 予算 経済 わが国 みんな 見通し 会議 学
校 答え 事故 愛 技術 協力 数 正午 努力 来年 交渉 研究
従来 施設 相手 開発 気持 国鉄 事情 解決 総会 方向 利益
言葉 各国 現実 対象 考え方 課題 胸 生徒 電話 雨 風 国
民 地 自由 子 家庭 朝日 心 機構 海 計 歴史 住宅 責任
裏 平均 妻 自動車 海外 主 政治 戦争 対策 平和 企業 主
婦 人々 政策 記録 成績 主張 重要 担当 目的 ことば 主催
勉強 対立 人口 女子 男子 店 ニュース 白 英語 話題 歓迎
社 急 先生 地下鉄 夫 交通 夢 決定 隣 工場 審議 本 両
国 デパート サービス 批判 委員会 中小企業 幹部 国会 作家
資料 自宅 道路 メーカー 経営 死 病気 材料 スポーツ 話し
合い 家族 意向 工事 双方 記者会見 商品 検査 若干 警官
職業 増 地域 伝統 空気 行動 姿勢 情勢 提案 原則 流れ
連続 卒業 委員長 改善 仲間 名前 強化 相当 検討 鳥 都市
大幅 注意 あらわれ 当局 ◆ わたし あなた 彼ら 彼女 だれ
何 ◆ 米国 インドネシア アジア フランス インド ヨーロッパ
ベトナム 韓国 日本人 ソ連側 沖縄 パリ ニューヨーク モスク

ワ 北京 名古屋 千葉 静岡 北海道 京都 新潟 横浜 札幌 広島
島 新宿 銀座 大田区 世田谷区 神田 新宿区 上野 渋谷区 埼玉
玉県 大蔵省 外務省 佐藤首相 朝日新聞

〔数 詞〕 五 七 八 九

〔コソアド〕 こう そんな あれ どの どんな

〔動 詞〕 考える 応じる 通じる 与える めぐる 待つ 帰る 伝える わたる
作る 生れる 続く 減る 打つ 主張する 生きる たずねる 述べる
はかる 高める いえる

〔形 容 詞〕 きびしい 美しい

〔連 体 詞〕 小さな そういう わが

〔副 詞〕 初めて 少し なぜ たとえば ちょっと ほぼ 結局 このほど
きわめて 全く まったく どうか むしろ 明らかに 特に やっと
一応 今度 いろいろ やや そのまま 逆に たくさん ようやく
もし

〔接 続 詞〕 および ところが または

A1のグループに属する語は、前に見た4種の語彙調査の最高重複群に属する語と大変よく似ている。基幹語中の基幹語ということになると、どうしても、こういう顔ぶれになるのだろう。それにしても、この最高基幹語彙の中に「アメリカ」という語が入って来るのは是非もない。

各グループの語彙を、ずっとながめていくと、4種の調査からの表で、4種にまたがる語から、3種にまたがる語、2種にまたがる語へとひろがっていったのと同様な発展の方向がたどられる。それは、思考のわくぐみを作っている抽象的な語から具体的な話題を盛る語へとひろがっていく過程である。これらの語について、これ以上考察を加える知恵は、今のところ、私にないが、とにかく、昭和41年の新聞紙上に組み上げられていた認識の世界が、その中枢部をここに見せているように思われる。

最後に、基幹語彙と著しく違う性格をもつ語、「狭くて深い」語群を見よう。D1に属する語は1000近くあるわけであるが、その中でも特にくせの強いもの、1層だけに現われてそこでは極めて多く用いられている語を取り出して

みる。

新聞の中で幅狭く、かつ深く用いられた語＜D 1 の一部＞

〔案内広告欄だけで極めて多く用いられた語〕

給 歴 有 住 完 通 優遇 社保 不問 面談 含 高給 代 見習 賞与
委細 寮 履歴書 昇給 交通費 浴 二萬 住込 急募 賃 来社 環 数名
週休 日給 又 敷 ホカ 若干名 送 乞 交費 求 初給 陽 営業社員 月収
祭 パマ 営業部員 經理 左記 手取 東口

〔スポーツ欄だけで多く用いられた語〕

安打 三振 本塁 四死 巨 大鵬

〔経済欄だけで多く用いられた語〕

ダウ 綿糸 人絹糸 日産自 三井山 前日比 三菱電 前期 九州電 三菱重
本田技 郵船

〔社会記事だけで多く用いられた語〕

同署

4 種調査の比較の最後で、新聞調査で 100 位以内にあり、他では 500 以内にはないものが 20 語あり、ほとんどが案内広告欄の語と思われた。あの 20 語のほとんどは、この D 1 に属するものである。それらは案内広告欄だけに頻出し、他には一切現われないものである。案内広告欄のサンプル量が多いために、全体としてかくも上位を占める結果となったのだが、それらは、もちろん、基幹語彙探求の過程で自然に排除され、この特殊領域に掃き寄せられた。

4. 基本語彙について

語彙調査から直接に求められるのは基幹語彙だけだ、というのが、本稿の主旨である。「基本語彙」とよく言われるが、基本語彙は目的が定まらなければ考えることができないというのも、それとともに強調したいことである。

基本語彙は、いろいろな専門領域でそれぞれ求められるだろうが、いちばん一般的には、何といても、初等教育における国語教育基本語彙と、外国人の日本語学習のための教育基本語彙とが、まず考えられるだろう。そのどちらにしても、それを考案するには、次のような段取りが考えられるだろう。

(1) いくつかの語彙調査を行ない、それぞれの資料内で基幹語彙を求める。それぞれの基幹語彙を比較して、各基幹語彙からのまた基幹語彙を求める。これを土台にして、これに論理的操作を加える。論理的操作とは、同意語、類義語、対義語、反意語などの系列を作って、評価を加えていくことである。例えば、前節の新聞基幹語彙で、「北」という語はA2に属し、「南」「東」はB2に属している。A2とB2だから重みが違うが、教育基本語彙として、重要度に差がつけられるとは思えない。また、これらの中に「西」は見当たらないが、だからといって、「北」「南」「東」を基本語と認めた基準が「西」を基本語と認めないですむとは思えない。語彙調査がどうあろうとも、これは加えなければならない。

基幹語彙は closed な世界でのみ求められるが、基本語彙となれば、open な性格をもって来るから、外側からのわくよりも、内側からの発展を重視しなければならない。

(2) 語彙調査データを用いて、基礎語彙の試作をしてみる必要がある。例えば、上記B1に「女」「女性」の語があり、B2に「女子」の語があって、基幹語彙の中に「女」「女性」「女子」の3語がそろったわけであるが、これらはともに必要であるか否か。1語で他を代用したら、無理なく代用が成立するだろうか。無理があるとすれば、どこに無理が出るだろうか。「女」を用いる場合には「女の人」という連語を必要とするが、「人」はA1で頑張っているから心配がないとか。いろいろな角度からの用語経済学を試みる必要がある。

(3) 語彙調査結果の基幹語彙から、層別への洞察を働かして、基準語彙へ発展させてみる必要がある。新聞記事の世界が話題の性質によっての12層に別れると見るのが妥当だとしても、その12層が私たちの生活のもつ諸面を充分におおっているという保障はない。例えば、冠婚葬祭のもつ重要さは新聞の層別には反映していないだろう。そんなことが、ほかにもいろいろあるだろう。私た

ちの生活の構造についての余程賢明な考察がないと、役に立つ基準語彙は案出できないのである。

そのほか、教育基本語彙を考えるには、表記法との関係が大きくあり、漢字にどれだけの働きをさせるかによって、語彙の性格が大きく左右される。

語構造論、品詞論、語種論等の立場からの考察も絶対に必要である。現代日本語の中で、漢語が、個々の語については嫌われ、敬遠される面があっても、全体として大きな力を持ち続けているのは、その造語力・語結合力のゆえである。長い概念を瞬時に一語にまとめて名詞を作り、自由に文の主語に立たせ得る能力は、到底和語の太刀打ちできないところである。語構造論は構文論へも発展する。

すぐれた基本語彙を考案するには、すぐれた科学とすぐれた哲学とが手を結ばなければならない。

追記　くれぐれもことわっておくが、この基幹語彙論は、あくまでも、考え方と方法とを論じたものである。新聞基幹語彙の実例として提出した語彙表の各語は、上に述べた「長単位」作業の所産であるから、個々に見れば、おかしなことがたくさんある。例えばA2の名詞のあとのほうに「米」とあるが、これは「米国」をさす「米」も、「こめ」と読むべき「米」も、いっしょになっている。それから、この語彙表の語数を克明にかぞえたら、さきにかかげた12区画内の数と合わないだろう。これは、そんな長単位を多少修正して扱っているところがあるからである。

参 考 文 献

- | | |
|---------------|---|
| Bergsland, K. | On the Validity of Glottochronology (Current Anthropology |
| and H. Vogt | 3) 1962 |
| 寿岳章子 | 「源氏物語基礎語彙の構成」(『計量国語学』No.41) 1967—7 |
| 垣内松三 | 『基本語彙学上』(文学社) 1938 |
| 国立国語研究所 | 「義務教育終了者に対する語彙調査の試み」(『国立国語研究所 |

- 年報』2, 秀英出版) 1951
- 『語彙調査——現代新聞用語の一例——』(国立国語研究所資料集2) (秀英出版) 1952
- 『現代語の語彙調査 婦人雑誌の用語』(国立国語研究所報告4) (秀英出版) 1953
- 『現代語の語彙調査 総合雑誌の用語 前編』(国立国語研究所報告12) (秀英出版) 1957
- 『現代語の語彙調査 総合雑誌の用語 後編』(国立国語研究所報告13) (秀英出版) 1958
- 『明治初期の新聞の用語』(国立国語研究所報告15) (秀英出版) 1959
- 『総合雑誌の用語』(国立国語研究所報告19) (秀英出版) 1960
- 『同音語の研究』(国立国語研究所報告20) (秀英出版) 1961
- 『現代雑誌九十種の用語用字(第一分冊) 総記および語彙表』(国立国語研究所報告21) (秀英出版) 1962
- 『明治時代語の調査研究』(『国立国語研究所年報』14, 秀英出版) 1963
- 『現代雑誌九十種の用語用字(第二分冊) 漢字表』(国立国語研究所報告22) (秀英出版) 1963
- 『現代雑誌九十種の用語用字(第三分冊) 分析』(国立国語研究所報告25) (秀英出版) 1964
- 『分類語彙表』(国立国語研究所資料集6) (秀英出版) 1964
- 『類義語の研究』(国立国語研究所報告28) (秀英出版) 1965
- 『電子計算機による国語研究』(国立国語研究所報告31) (秀英出版) 1968
- 『電子計算機による国語研究Ⅱ——新聞の用語用字調査の処理組織』(国立国語研究所報告34) (秀英出版) 1969
- 『電子計算機による新聞の語彙調査』(国立国語研究所報告37) (秀英出版) 1970
- 国際文化振興会 『日本語基本語彙』(国際文化振興会) 1944
- 土居光知 『基礎日本語』(六星館) 1933
- 『日本語の姿』(改造社) 1943
- 服部四郎 『日本語の系統』(岩波書店) 1959
- 『基礎語彙統計学について』(「計量国語学」No. 15) 1960—12
- 『ふたたび基礎語彙統計学について』(「計量国語学」No. 17) 1961—7
- 『英語基礎語彙の研究』(三省堂) 1968
- 林四郎 A Trial Research in Semi-Automatc Abstracting by

- 文部省 『国語シリーズ41 児童・生徒の語い力の調査 準備調査(昭和32年度)〔第1分冊〕』(明治図書)1960
『国語シリーズ42 児童・生徒の語い力の調査 準備調査(昭和32年度)〔第2分冊〕』(明治図書)1960
『国語シリーズ51 児童・生徒の語い力の調査 本調査(昭和35年度)〔小学校第6学年〕』(光風出版)1962
『国語シリーズ52 児童・生徒の語い力の調査 本調査(昭和34年度)〔小学校第4学年〕』(光風出版)1963
『国語シリーズ58 児童・生徒の語い力の調査 本調査(昭和35年度)〔中学校第3学年〕』(教育図書)1964
『国語シリーズ59 児童・生徒の語い力の調査 低学年の学習語(昭和37年度)』(教育図書)1964
『国語シリーズ63 児童・生徒の語い力の調査 本調査(昭和36年度)〔中学校第1学年〕』(大日本図書)1967
- Ogden, C. K. 『ベーシックのABC』(The ABC of Basic English) (研究社)1936
(高田力訳)
- Ogden, Richards. 『意味の意味』(The Meaning of Meaning) (興文社)1936,
(石橋幸太郎訳) (改訳, ぺりかん社)1967
- 阪本一郎 『日本語基本語彙 幼年之部』(明治図書)1943
『読みと作文の心理』(牧書店)1955
『教育基本語彙』(牧書店)1958
- Swadesh, M. Diffusional Cumulation and Archaic Residue as Historical Explanation (Southern Journal of Anthropology 7) 1951
(Language in Culture and Society 1964 に再収)
Basic Vocabulary : Glottochronology (mimeographed) 1951
Towards Greater Accuracy in Lexicostatistic Dating (International Journal of American Linguistics 21) 1955
Glottochronology (Readings in Anthropology) (Thomas Y. Crowell Co.) 1959
- Thorndike, E. L. Century Junior Dictionary (Scott, Foresman and Company) 1935
Century Senior Dictionary (Scott, Foresman and Company) 1941
- Thorndike, E. L. and I. Lorge The Teacher's Word Book of 30,000 Words (Bureau of Publications, Teachers College, Columbia University) 1944