

# 語彙理解尺度作成の試み

教育心理学研究室

芝 祐 順

## Construction of a Scale for Acquisition of Word Meanings

Sukeyori SHIBA

The purpose of the present study is to prepare a pool of vocabulary items which can be used for measuring broad range of ability and construct a scale for word meaning comprehension. After preliminary testing and item analysis, 11 forms of vocabulary tests appropriate from first grade through graduate school have been built and administered. Applying the latent trait model, item characteristic parameters were estimated in each test. Making use of the common items contained in each set of adjacent level tests, scale values of latent trait in different tests were equated. Thus, difficulty and discriminating power of all items in 11 tests were characterized on the common scale. Scores of subjects were estimated by the maximum likelihood method. Precision of estimates of each test is evaluated, at all levels of ability, in terms of information function. Possible applications of this item pool, for example, construction of a broad range tailored testing, are discussed.

- I 問題
- II 潜在特性モデル
- III テスト項目の作成
- IV テストの実施と結果
- V 尺度化の意義

### I 問題

語の意味を理解することは、ことばの理解や使用の前提である。その理解力を計量する方法を検討し、教育実践や研究のために使用することのできる尺度を構成するのが本研究の目的である。

これまで度々おこなわれてきた語彙調査も、理解できる語彙の量を計るということで、本研究と同様の目的をもったものである。しかし、本研究とはつぎの点で異っている。まず第1に、従来の語彙調査では語の意味を固定的に考え、各人にとってそれぞれの語は理解されているものとされていないものとに2分される。そして個人は語彙として何語かを獲得（理解）しているのだから、その語彙数を推定する試みがおこなわれてきた。そのため、理解していることの確認には比較的断定的な方法がもちいられる。たとえば、被調査者に語を示して知ってい

るかどうか尋ねるという方法（阪本<sup>1)</sup>、森岡<sup>2)</sup>）によって、その語が個人の語彙に含まれているという判定をしている。これに対し、本研究ではある語の意味について理解の状態は、原則的には連続的なもので、理解している状態と理解していない（知らない状態）とははっきり分けられるものではない、という考えを基礎においている。個々の語について各人の理解の程度はさまざまであって、“知っている語”の個数を数えることは必ずしも適切な方法でない。もちろん、日常的なものの名称や、基本的な動作をあらわす動詞などの場合には、全く知らないという状態と、よく理解しているという状態とが比較的明確にわかれていることがある。しかし、理解の水準が中間的ともいえる多くの語があり、どこまでの理解をもって知っているとするかによって、推定語彙数も変わってしまう。このような考えから本研究では語の意味の理解の測り方をこれまでの語彙調査の方法とは違って、意味理解のテストによっておこなう。個々の語の理解については連続的な次元を考え、テストの項目はその次元の一定水準に達した被験者とその水準には達していない被験者とを識別する。そして、被験者の語の意味理解力としては、多数の語の個々の連続的な次元とすべて相関の高い一つの共通な次元を想定し、これを語の意味の一般

的理解の水準をあらわす尺度とする。したがって、個人の語の理解力としては、どの語が習得されたかということではなく、語の意味の理解の一般的水準ほどの程度かを測ることになる。そのために、次にのべるような統計的モデルがもちいられる。

## II 潜在特性モデル

### 1. 潜在特性と項目特性曲線

本研究では語の意味の理解をあらわす尺度を構成するために潜在特性を仮定したテスト・モデル (Lord<sup>3)</sup>, 印東・鮫島<sup>4)</sup>, Lord and Novick<sup>5)</sup>) を採用する。このモデルの基本仮定はつぎのとおりである。

一つの項目  $j$  に関与する能力を変数  $y_j$  であらわす。能力  $y_j$  がある値を越えるとその被験者は項目  $j$  に正答することができ、その値以下の被験者は正答することができない。その値 (それを  $\gamma_j$  とあらわすことにする) は各項目によって定まっている。この能力変数  $y_j$  は各項目に固有のものであるが、これとは別に、すべての項目に共通の変数  $\theta$  を想定し、変数  $y_j$  と  $\theta$  との間に

$$(1) \quad y_j = \rho_j \theta + \sqrt{1 - \rho_j^2} e_j$$

という関係を仮定する。ここで、 $\rho_j$  は項目  $j$  に固有の定数であり、 $e_j$  は確率変数で規準正規分布  $N(0, 1)$  に従うものとする。そして、 $e_j$  は  $\theta$  とは独立で、また他の項目の  $e_k$  ともすべて独立な確率変数である。このような共通の変数  $\theta$  のことを潜在特性とよび、ここでは、その分布を正規分布  $N(0, 1)$  と仮定しておく。 $y_j$  が個々の語の意味を問う項目を解くのに必要な能力をあらわすのに対し、 $\theta$  はすべての語の意味理解にかかる一般的な能力をあらわす変数である。 $e_j$  は、したがって、項目  $j$  だけに固有な意味理解に関わる能力をあらわす。 $e_j$  の中には、 $\theta$  のように他の項目にも共通するような能力はいっさい含まない。このことから、 $\rho_j$  は項目  $j$  を解く能力  $y_j$  と潜在特性  $\theta$  との相関係数でもある。

以上の仮定から、すべての項目  $j=1, 2, \dots, n$  に対する能力  $y_j$  は多変量正規分布  $N(\mathbf{0}, \Sigma)$  に従うことになる。ここで、分散・共分散行列  $\Sigma$  は

$$\Sigma = \rho\rho' + D$$

によってあらわされる。ただし、 $\rho$  は各項目に関わる能力  $y_j$  と潜在特性  $\theta$  との相関係数  $\rho_1, \rho_2, \dots, \rho_n$  を成分とする列ベクトルをあらわし、 $D$  は、 $1 - \rho_1^2, 1 - \rho_2^2, \dots, 1 - \rho_n^2$  を対角成分とする対角行列をあらわす。

所与の  $\theta$  に対する  $y_j$  の分布は、平均  $\rho_j \theta$ 、分散  $1 - \rho_j^2$  の正規分布に従う。したがって、特性値  $\theta$  の被験者

が項目  $j$  に正答をあたえる確率 (特性値  $\theta$  の任意の被験者の  $y_j$  の値が  $\gamma_j$  以上である確率) は

$$(2) \quad P_j(\theta) = \frac{1}{\sqrt{2\pi} \sqrt{1 - \rho_j^2}} \int_{\gamma_j}^{\infty} \exp\left\{-\frac{(y_j - \rho_j \theta)^2}{2(1 - \rho_j^2)}\right\} dy_j$$

である。適宜、変数を変換するとこれは規準正規分布関数をもちいた次のような簡単な形であらわすことができる。

$$(3) \quad P_j(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_j(\theta - b_j)} e^{-\frac{t^2}{2}} dt \\ = \Phi[a_j(\theta - b_j)]$$

ここで、 $a_j$  は

$$(4) \quad a_j = \frac{\rho_j}{\sqrt{1 - \rho_j^2}}$$

$b_j$  は

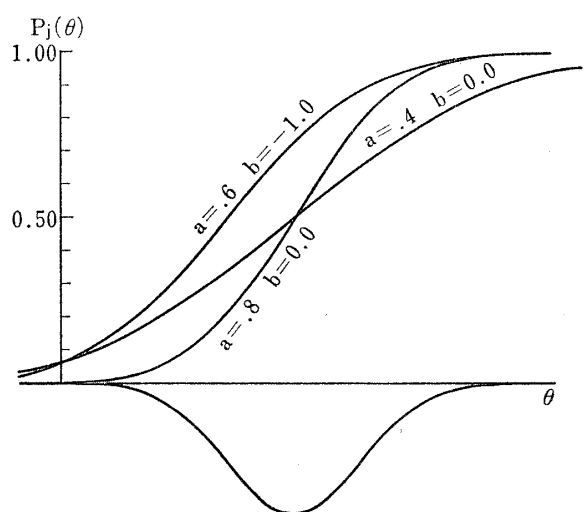
$$(5) \quad b_j = \frac{\gamma_j}{\rho_j} = -\frac{\sqrt{1 + a_j^2}}{a_j} \Phi^{-1}(\pi_j)$$

である。 $\gamma_j$  は項目  $j$  に正答をあたえることのできる  $y_j$  の限界値であり、これは項目  $j$  の全被験者による正答率  $\pi_j$  の正規分布逆関数  $\Phi^{-1}(\pi_j)$  によってあたえられる。

このように項目の正答確率を潜在特性の関数としてあらわしたものを、項目特性曲線とよんでいる。項目特性曲線は、語の意味の理解力を一つの仮定された特性  $\theta$  によってあらわしたとき、個々の項目に対する正答の条件つき確率をあらわす。

項目特性曲線を定める二つのパラメタ  $a_j, b_j$  のうち、 $b_j$  は曲線全体の位置をあらわすもので、項目  $j$  の困難度をあらわす。 $b_j$  が大きいときには項目特性曲線は右に

図 1 潜在特性  $\theta$  の分布と項目特性曲線の例



移動し、すべての  $\theta$  の値において正答確率が低くなる。これに対し、パラメタ  $a_j$  は曲線の勾配に関係している。曲線の変曲点  $\theta = b_j$  において勾配が最も大きくなるが、その値は  $a_j$  に比例している。 $a_j$  が大きいというこ

とは、一般的に  $\theta=b_j$  付近における  $\theta$  の変化に対して正答の確率の変化が著しいことを意味し、したがって  $a_j$  は項目の識別力をあらわすパラメタである。

## 2. パラメタ $a$ と $b$ の推定

パラメタ  $a_j$  は(4)のように、項目  $j$  に答える能力をあらわす変数  $y_j$  と、一般的意味理解能力をあらわす潜在特性  $\theta$  との相関係数  $\rho_j$  の関数としてあらわされている。実際のデータとしてあたえられるのは個人の  $y_j$  の値ではなく、それが  $\gamma_j$  を越えるか否かによって定まる正答誤答による項目得点  $(1, 0)$  である。もし、個人の潜在特性値  $\theta$  が知られるなら、これと項目得点との間の双列相関係数をもとめることによって  $\rho_j$  を推定することができる。しかし、この  $\theta$  も直接もとめることができないので、この双列相関係数の推定に因子分析的手法をもちいる。(1)にあたえたように各項目に対する能力  $y_j$  と共通の能力  $\theta$  の関係があらわされるなら、これはいわば一つだけの共通因子をもつ因子分析モデルにひとしい。そこで、項目得点の間の相関から 4 分相関係数によって項目に対する能力  $y_j, y_k$  間の相関係数を推定し、それによってあたえられる項目間相関行列の因子分析をおこなう。実際のデータからえられる相関行列が一つだけの共通因子で説明されるということはないが、近似的には主因子解によってもとめられる第 1 因子をもって潜在特性とすればよい。因子分析でもとめた第 1 因子負荷は第 1 因子と各項目の変数との相関係数をあらわすから、これはそのまま潜在特性  $\theta$  と各項目に関わる能力  $y_j$  との相関係数  $\rho_j$  の推定値となる。

パラメタ  $b_j$  は(5)に示したように、 $a_j$  の推定値と、 $\pi_j$  の推定値を必要とする。 $\pi_j$  の推定値には全被験者による項目  $j$  の正答率  $p_j$  をあてればよい。

以上のべた方法は、いわゆる heuristic method とよばれる方法<sup>6)</sup>である。このほか、最尤推定法によって項目特性を推定する方法<sup>7)</sup>や全被験者についての潜在特性  $\theta_i$  と項目特性  $a_j, b_j$  とを同時に推定する方法<sup>8)</sup>などもある。

## 3. 得点の理論分布

集団の潜在特性の分布と項目特性とが知られていれば、モデルに従って得点の理論分布をもとめることができる。この理論分布はモデルのあてはまりのよさの検証にも用いられるが、また実用上は新しく項目を組みかえて作ったテストの得点分布を予測することができて便利である。

$\theta$  を固定したときの得点  $x$  の確率は

$$(6) \quad f(x|\theta) = \sum_{v_x} \left[ \prod_{j=1}^n P_j(\theta)^{u_j} Q_j(\theta)^{1-u_j} \right]$$

によってあたえられる。ここで  $v_x$  は得点が  $x$  となるすべての応答パターンについてこの和をとることを意味する。 $u_j$  は項目  $j$  を正答とするとき 1、誤答とするとき 0 をとる 2 値変数である。すべての  $\theta$  にわたって  $\theta$  の確率密度を乗じた積分をとると得点分布を示す

$$(7) \quad f(x) = \int_{-\infty}^{\infty} f(x|\theta) g(\theta) d\theta$$

がえられる。これは  $\theta$  の簡単な関数としてあらわすことができないので、実際には数値的に算出する。

## 4. ロジスティック関数の利用

一般に正規分布関数はロジスティック関数によって、よく近似される。そこで、項目特性曲線  $P_j(\theta)$  をあらわすのにこれをもちいて

$$(8) \quad P_j(\theta) = \frac{1}{1 + \exp[-1.7a_j(\theta - b_j)]} \\ = \Psi[1.7a_j(\theta - b_j)]$$

とする。この関数は数学的取り扱いのうえで正規分布関数よりも便利なが多い。本研究では潜在特性モデルの説明には正規分布をもちいているが、モデルの展開や数値の計算には便宜上ロジスティック関数をもちいている。なお、はじめにあげた正規分布関数による項目特性曲線の場合にも、ロジスティック関数を用いた場合にも、パラメタは  $a, b$  の 2 個であったが、項目ごとの偶然正答の確率を考慮に入れた 3 パラメタの項目特性曲線をもちいるモデルも提案されている<sup>9)</sup>。その場合、項目特性曲線はそれぞれ

$$(9) \quad P_j(\theta) = c_j + (1 - c_j)\Phi[a_j(\theta - b_j)]$$

$$(10) \quad P_j(\theta) = c_j + (1 - c_j)\Psi[1.7a_j(\theta - b_j)]$$

となる。しかし、本研究ではこのモデルはもちいない。

## 5. 異った集団による尺度の等化

一つの集団データから項目パラメタを推定するとき、その集団における潜在特性  $\theta$  は標準正規分布  $N(0, 1)$  に従うものと仮定する。したがって同じ項目についても集団が異ると、推定されるパラメタの値は必ずしも同じにはならない。共通な項目を含む 2 つの版のテストを異なる集団に実施してえられたデータから、共通尺度上で扱うことのできる項目特性を推定する問題についてのべる。これは異った集団に異ったテストを実施してえられる尺度の等化の問題である。

等化の方法としてはいく通りかが考えられるが、ここでは、まず 2 つの版に共通に含まれる項目をもちいて、

パラメタ  $a_j$  を等化し、つぎにその結果をもちいてパラメタ  $b_j$  を等化する方法をもちいる。一つの共通項目における正答の確率は、集団が異っても同等の潜在特性に対しては等しい値をとるはずである。したがって、各集団において同じ理解力の水準をあらわす潜在特性値をそれぞれ  $\theta, \theta^*$  とし、また、各集団における共通項目  $j$  の特性をそれぞれ  $P_j(\theta), P_j^*(\theta^*)$  とすると

$$(11) \quad P_j(\theta) = P_j^*(\theta^*)$$

となる。したがって、2つの集団における項目特性  $a_j, b_j$  および  $a_j^*, b_j^*$  などの間には(3)より

$$(12) \quad a_j(\theta - b_j) = a_j^*(\theta^* - b_j^*)$$

という関係がなりたつ。これを書きかえて、対応する2つの尺度上の値の関係をもとめると

$$(13) \quad \theta = k\theta^* + d$$

となる。ここで

$$(14) \quad k = \frac{a_j^*}{a_j}$$

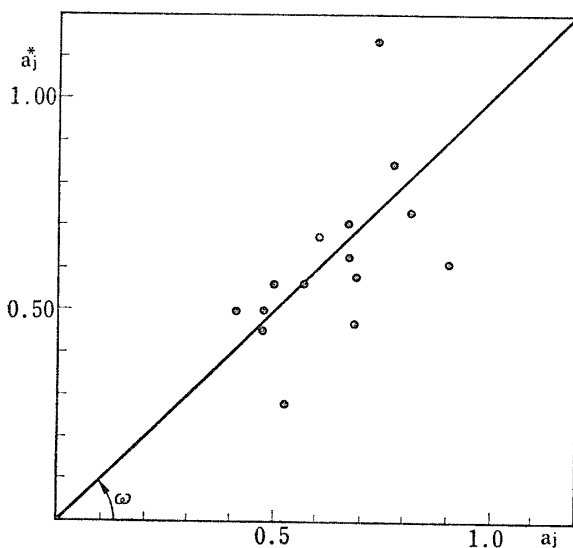
$$(15) \quad d = b_j - kb_j^*$$

である。こうして、2つの尺度の関係を示す定数  $k$  と  $d$  が、共通項目の特性をあらわすパラメタからもとめられる。

ただし、実際のデータではさまざまな誤差を含むため、2つ以上の項目からもとめられるその値が、それぞれ必ずしも一致しない。そのために、複数の共通項目の結果をあわせてもっともよい  $k$  と  $d$  の値を推定する必要がある。

パラメタ  $a$  については、すべての共通項目の間で等しい  $k$  の値が得られるためには、 $a_j$  と  $a_j^*$  とを2軸にとったグラフ(図2)において、すべての項目が原点を通る勾配  $k$  の直線上に配列されなければならない。実際に

図2 共通項目のパラメタ  $a$  (J1版とA6版)



えられた値では図2に示すように直線上にはのらない。これらに対して最もよい勾配の値を推定するには、その重心を通るように直線をひく方法や、各点からの距離の2乗和を最小にするように直線をひく方法が考えられる。いずれの場合も直線は原点を通らなければならない。重心の場合には、 $k$  の推定値は

$$(16) \quad k = \frac{\sum_{j=1}^m a_j^*}{\sum_{j=1}^m a_j}$$

によってあたえられる。

また最小2乗による場合、原点を通る直線で、各点からの距離の2乗和を最小とするものは、これらの点の主軸にあたる。したがって、このような直線の勾配をもとめる問題は主軸への回転角をもとめる問題である。 $m$  この点の座標が  $(a_1, a_1^*), (a_2, a_2^*), \dots, (a_m, a_m^*)$  であるとき主軸への回転角は

$$\omega = \frac{1}{2} \tan^{-1} \alpha$$

$$\alpha = \frac{2 \sum_{j=1}^m a_j a_j^*}{\sum_{j=1}^m (a_j^2 - a_j'^2)}$$

となることが知られている<sup>10)</sup>。したがって直線の勾配は

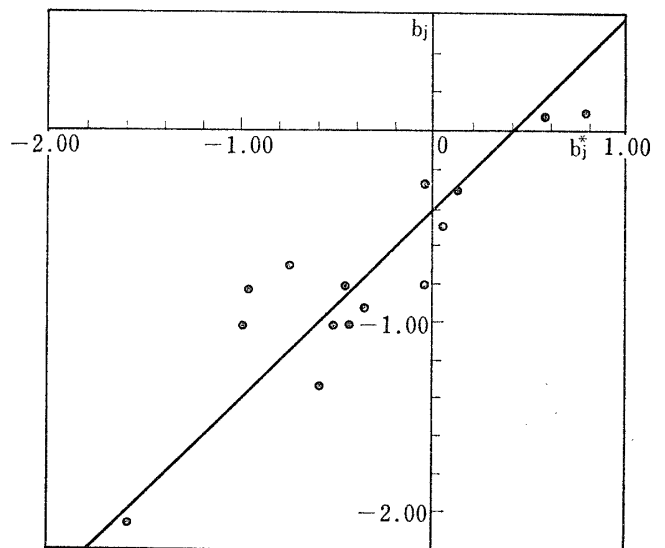
$$k = \tan \omega$$

$$= \frac{\sqrt{1 + \alpha^2} - 1}{\alpha}$$

となる。

つぎに  $k$  があたえられたとして、定数  $d$  の推定をおこなう。2つの集団における項目困難度を2軸にとり、描かれた各点よりの距離の2乗和が最小となる直線をもと

図3 共通項目のパラメタ  $b$  (J1版とA6版)



める。ただし直線の勾配は  $k$  とする。その結果から  
 (17) 
$$d = \bar{b} - k\bar{b}^*$$

となる。ただし、 $\bar{b}, \bar{b}^*$  はそれぞれ共通項目の平均

$$\bar{b} = \frac{1}{m} \sum b_j$$

$$\bar{b}^* = \frac{1}{m} \sum b_j^*$$

である。

## 6. 潜在特性値 $\theta$ の推定

一般に特性値  $\theta$  のものが、 $n$  項目に対し応答パターン  $v$  を示す確率は

$$(18) \quad P(v|\theta) = \prod_{j=1}^n P_j(\theta)^{u_{ij}} [1 - P_j(\theta)]^{1-u_{ij}}$$

である。 $u_{ij}$  は項目  $j$  に正答のとき 1、誤答のときを 0 とする確率変数で、 $v$  はそれらによる応答パターンをあらわす。ここで応答パターン  $v$  の方を固定してこの確率を  $\theta$  の関数とみなしたとき、これはいわゆる尤度関数となる。

ここでもちいる  $\theta$  の推定法は、被験者の応答パターンについてこの尤度の値を最も大きくするような  $\theta$  の値をもとめるもので、いわゆる最尤推定法である。最尤推定値を数値的にもとめるには、まず尤度関数の対数を取り、これを  $\theta$  に関して微分し、

$$(19) \quad \frac{\partial \log L}{\partial \theta} = 0$$

とおいて解をもとめる。それは

$$(20) \quad \sum_{j=1}^n a_j P_j(\hat{\theta}) = \sum_{j=1}^n a_j u_{ij}$$

となる。これを  $\theta$  について簡単な表現にすることはできないので、 $u_{ij}$  としてあたえられる実際の応答パターンから、数値計算によって  $\theta$  の推定値  $\hat{\theta}$  をもとめる。

潜在特性モデルの一つの特徴は、この  $\theta$  の推定法にある。ふつうのテスト得点などでは正答項目数を直接もちいるため、一つのテストに新たな項目を加えたり、あるいは減らしたりして項目数を変えたときには、得点をそのまま比較することができない。ところがここにのべた最尤推定法によると、潜在特性  $\theta$  は、原則的に任意の数の項目への応答から推定されるもので、項目の数に直接関係するものではない。したがって項目の選択は自由である。共通の尺度上でのパラメタの値  $a, b$  の知られている項目であれば、その中から任意に選んだ項目をもちいて  $\theta$  を推定すれば、同じ尺度上の値がえられる。このことは、すべての被験者に同じ一組の項目を与えるふつうのテストの場合とはちがって、一人一人の被験者に異なった項目を選んで与える特殊な形式のテスト、たとえばプログラム化されたテストなどの場合に極めて重要な利点

となる。

## 7. 情報関数によるテストの評価

最尤推定法による推定値  $\theta$  の標本分布は、標本数（この場合は項目数にあたる）が大きいために、漸近的に正規分布に従うことが知られている。その標本分布の分散をもって、推定値がどの程度真の  $\theta$  に近いものであるかを評価するのにもちいる。すなわち分散が小さければ、実際にえられた推定値  $\hat{\theta}$  が真の  $\theta$  に近いことが期待されるわけで、分散が小さい程テストのもたらす情報は大きい。推定値  $\hat{\theta}$  の漸近的な標本分布は  $N(\theta, 1/I(\theta))$  になる。その平均は真の  $\theta$ 、分散は情報関数とよばれる量  $I(\theta)$  の逆数である。情報関数は尤度関数  $L$  をもちいて

$$(21) \quad I(\theta) = E \left\{ \left( \frac{\partial \log L}{\partial \theta} \right)^2 \right\}$$

とあらわされる。1 項目について、この情報関数をもとめると

$$(22) \quad I(\theta) = 1.7^2 a_j^2 P_j(\theta) [1 - P_j(\theta)]$$

となる。 $n$  項目のテストでは

$$(23) \quad I(\theta) = 1.7^2 \sum_{j=1}^n a_j^2 P_j(\theta) [1 - P_j(\theta)]$$

となる。

## III テスト項目の作成

### 1. 語の選択

日本語の語彙の中から、いくつかの語を選んでテストの項目を作るが、これによってえられる結果の解釈や利用のためには、語の選択の基準を明らかにしておかなければならない。まず、日本語の語彙全体を母集団として、そこから任意標本をとりだすということは実際に日本語の語彙の母集団を把握することができないので、不可能である。ここでは、そのような任意標本をとるということせず、基本的と思われる語を重点的に選ぶことにする。しかし、どの語を基本とするかということ、いわゆる一般的な基本語彙の基準そのものも唯一ではなく、いずれもこのテストの目的からみてそのまま適しているとはいえない。

基本的な語彙を定める観点の主なものにもとめてみるとつぎのとおりである<sup>11)12)</sup>。

まず第 1 は語の使用の実態にもとづくものである。使用度数の高いものほど基本的であるとする。この場合には使用度数の推定のための調査のように、客観的手続きによって基本語彙を定めることができる。

第 2 は、言語習得の順序にもとづくものである。発達

的にみて、早期に獲得されるものほど基本的であるとする。この場合、“獲得”の定義やそれを確認する方法について困難が伴うが、適当な調査によって実際に語彙の基本性を示すことができよう。

第3は特定言語の中での役割にもとづくものである。構造的に中枢部分にあるものほど基本的であるとする。この場合、どのような構造としてとらえるかによって、中枢となるものが変りうるということや、重要な役割りを果していることをどう計量化して基本性の順位を定めるかが問題となる。

第4は、特別の目的のために設定するもので、教育基本語彙とか外国人のための日本語基本語彙などがその例である。その目的のためにもっとも効果的な語を基本として選びますが、ときによっては上記の3つの観点のいずれかにもとづくことになる。

このような観点はいずれもこのテストの目的に関連するものであるが、中でも第2の観点がもっとも重要である。しかし、具体的に語を選ぶとなると、そのための利用できる資料はほとんどない。この点は、第1の観点についても、第3の観点についても同様である。結局、実際の語を選ぶときの資料としては、各社の検定教科書、教育基本語彙<sup>13)</sup>、辞書などがもちいられた。その際、語の使用の実態に注意し、また、教育を目的とした語彙を重視する。特別な専門用語は除くこととし、日常の言語生活に重要と思われる語を選んだ。

## 2. 意味の理解の問い方

選択肢形式のテストであるという制約から意味の理解の問い方にも限界があるため、つぎのような方法がとられた。

同義語やいいかえによるもの：意味の理解をきくためにとりあげられた主題の語に同義語がある場合には、その両者の意味の一致を認めることのできるもののみが、正しい選択肢を選ぶように項目をつくる。しかし、一般に同義語といっても意味が全く同じわけではない。いいかえにしても全く意味を同じにすることはできない。そこで、テストでは意味の最も近いものを選択するように教示する。それでも、いくつかの選択肢の中で主題の語に意味の最も近いものがどれであるのかを判断するのは必ずしも容易ではないであろう。数語の間の意味の近さを相互に比較するとき、いわば意味の空間が多次元的で、その距離の定義によって近さが一義的に定まらないこともある。このようなことを考慮して意味の近さについては、理解しているものにとっては直観的判断が容易にできるように選択肢の組み合わせを作る必要がある。

絵によるもの：年齢が低い場合には、同義語やいいかえによって理解を問うことは困難である。その場合には絵による選択肢を利用する。

反対語によるもの：形容詞、副詞などの理解を問うのに、年齢の低い場合には同義語やいいかえが困難であり、また絵による表現も難しいことが多い。そのときには反対語を選ばせることによって間接的に意味の理解を確認する。この場合には反対関係そのものの理解が前提であり、対にされる語の理解が容易なものでなければならない。

擬態語によるもの：日本語には擬態語が多く、意味伝達の上でも重要な役割を果しているが、その意味をいいかえて表現するとかえって難しくなることが多い。そこで、擬態語については、簡単な文をあたえてその文とのつながりの自然さを判断させることによって、間接的な擬態語の意味の理解の程度を測定するという方法をとる。

## 3. 項目の形式

項目はすべて5肢択一の形式に統一する。

〔例〕 ときおり

- A いつも
- B むかし
- C きゅうに
- D ときどき
- E ゆっくり

このように原則として主題の語を幹の部分におき、選択肢には正しい肢を1こ、誤りの肢を4こ設ける。選択肢を作るときの原則的事項はつぎのとおりである。

- i 選択肢は正誤いずれのものも、主題の語より意味の理解が容易なものであること。そして、一つの項目内では、その程度をできるだけそろえること。
- ii 誤りの肢には、主題の語の誤った理解としてありがちなものを選ぶこと。あるいは、主題の語を全く理解していないものが、その意味であろうと推測し易いものを選ぶこと。
- iii 字数が少なく、その意味を読みとるのに多くの時間を要しないもの。一つの項目内ではできるだけ字数を揃えること。
- iv 項目の識別力を高める働きのあるものを選ぶこと。これは統計的解析によって確認されることである。のちにのべる項目分析によって、識別力の低い項目は除くか修正する。

なお、例外的にはあるが、主題の語を選択肢におき、幹の部分に説明をおくことがある。これは主として

意味の説明が長くなるような場合である。なお、反対語や擬態語をもちいる項目においても、上の原則はほぼそのまま適用される。

#### 4. 項目の作成と分析

項目の作成の過程は、原案の作成のあと、予備テストの実施、項目分析、修正の作業が循環的くりかえされる。原案の作成には主題の語の選出と選択肢の作成が含まれるが、この段階は多数の協力者によっておこなわれた。予備テストは1973年6月から1975年11月までの間に随時おこなわれた。項目分析には、選択肢別状況を明らかにするGP分析<sup>14)15)</sup>がおこなわれた。各項目について次のような選択状況の資料にもとづいて修正すべき点の確認がおこなわれた。

- i 正答の比率が極端に偏っていることはないか。一つの目安として正答率.2以下の項目、.9以上の項目は再検討される。
- ii 一つの正答の枝が、他の4つの誤答の枝に対して十分な識別力を示しているか。この識別力の指標として総得点と項目得点との点双列相関係数がもちいられる。そのときの一つの目安としてこの係数が.25以下のものはその原因を検討し、さらに高くなるように修正する。
- iii 誤答の枝には、いずれも適度の数の選択があるか。全く選択のない枝はほとんど機能していないので修正する。
- iv 誤答の枝の選択傾向はいずれも総得点と負の相関の傾向を示しているか。

このような項目分析を経て、最終的には表1に示した11版に構成された。(この他幼児用として2版がある。これは個人テスト用のものである。) 共通尺度を作成するため各版は、それぞれ隣接する版との間に共通項目をもっている。(表1のうち共通項目数の欄参照)

### IV テストの実施と結果

尺度作成のために準備された11版のテストは1976年3月から同年5月までの間に実施された。各版ごとの被験者数は表1に示したとおりである。ここにもちいられた被験者はいずれも任意に選び協力を得たものであって、何らかの母集団を想定してその標本を選んだものではない。本研究では各版に相応した能力水準の異った集団が得られれば十分である。

表1 各版の被験者数と項目数

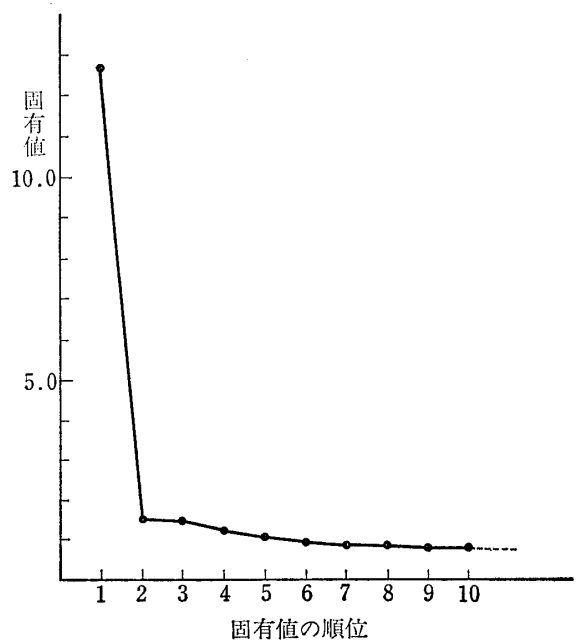
版名	被験者数	項目数	共通項目数
A1	小学校1年 650名	36(30)	10( 5)
A2	2年 650名	34(26)	16(14)
A3	3年 546名	40(35)	8( 6)
A4	4年 617名	40(37)	16(14)
A5	5年 599名	48(42)	16(16)
A6	6年 412名	56(54)	16(15)
J1	1年 614名	56(55)	20(19)
J2	3年 758名	60(58)	20(15)
S1	1年 924名	60(53)	20(18)
S2	2年 759名	60(52)	20(18)
U	大学生 740名	57(54)	

注 カッコ内はパラメタを推定することが可能であった項目数。学年表示はすべて1976年3月現在。

#### 1. 項目パラメタの推定

項目パラメタの推定のためには、項目間相関行列をもとめ、これを因子分析して主因子解の因子負荷を算出する。この場合、4分相関係数の近似計算の制約から、正答率がとくに偏っている項目が除外される<sup>16)</sup>。共通性の推定値としては、相関行列の各行の相関係数の最大値をもちいている。まず、固有値の分布をみてみると、図4

図4 項目間相関行列の固有値 (J1版)



のようになる。第1固有値が、第2以下の固有値に比べて大きな割合を占めており、項目間相関をもたらす因子として1つだけの潜在特性を仮定するがほぼ妥当であるといえよう。ここに示したのはJ1版の例であるが各版とも同様の傾向がみられる。表2には一例としてJ1版の項目の一部について因子負荷と正答率を示してある。これらをもちいて(4)、(5)からパラメタ $a, b$ を推定することができる。それらは表2の各欄に示されている。

表2 項目の正答率、因子負荷とパラメタ $a_j, b_j$   
(J1版よりの15項目)

主題の語	正答率	因子負荷	$a_j$	$b_j$
晩夏	.49	.56	.67	.07
ひときわ	.66	.51	.60	-.82
かたをならべる	.71	.55	.66	-1.01
口実	.68	.59	.73	-.80
敬意をはらう	.57	.57	.69	-.30
やさきに	.64	.43	.48	-.80
音さた	.49	.45	.50	.09
架空の	.90	.63	.81	-2.05
あくまでも	.67	.43	.47	-1.01
鼻をあかす	.62	.56	.68	-.52
発端	.81	.67	.90	-1.33
紀行	.63	.46	.52	-.71
月並	.67	.48	.55	-.93
不慮の	.38	.46	.51	.66
服用	.57	.61	.77	-.28

度がえられる。今回は便宜的に中学1年生に実施したJ1版の結果によってえられる $\theta$ の分布を $N(0, 1)$ とし、他の版の尺度をこれに等化するようにしている。各版別に算出した結果を、共通の尺度上に表現するための定数を表3に示す。共通の尺度上の潜在特性値 $\theta_c$ は版ごとにもとめた尺度値から

$$\theta_c = k\theta + d$$

によって換算される。また、項目特性 $a, b$ についても共通の尺度上であらわすには

$$a_c = \frac{1}{k}a, \quad b_c = kb + d$$

とすればよい。

表3 各版の尺度値を共通尺度に変換するための定数

版	$k$	$d$
A1	.69	-3.67
A2	.72	-3.32
A3	.81	-2.58
A4	.92	-1.66
A5	.85	-1.05
A6	1.00	-.41
J1	1.00	0.0
J2	1.39	1.84
S1	1.09	3.33
S2	1.10	3.99
U	1.74	5.89

## 2. 共通尺度の作成

異った版の中に含まれる共通項目をもちいて順次の版の間の等化を行うことによって、すべての版に共通の尺

もともと、各版の被験者集団について、潜在特性値 $\theta$ の分布を $N(0, 1)$ と仮定して項目特性などをもとめた

図5 共通尺度上における各版の被験者群の $\theta$ の理論分布

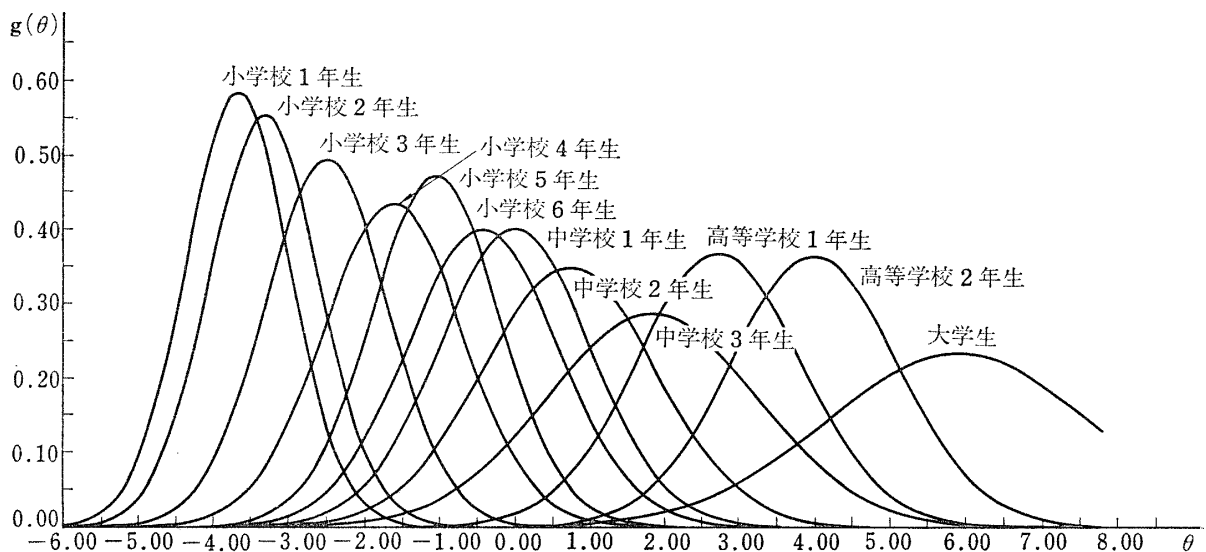
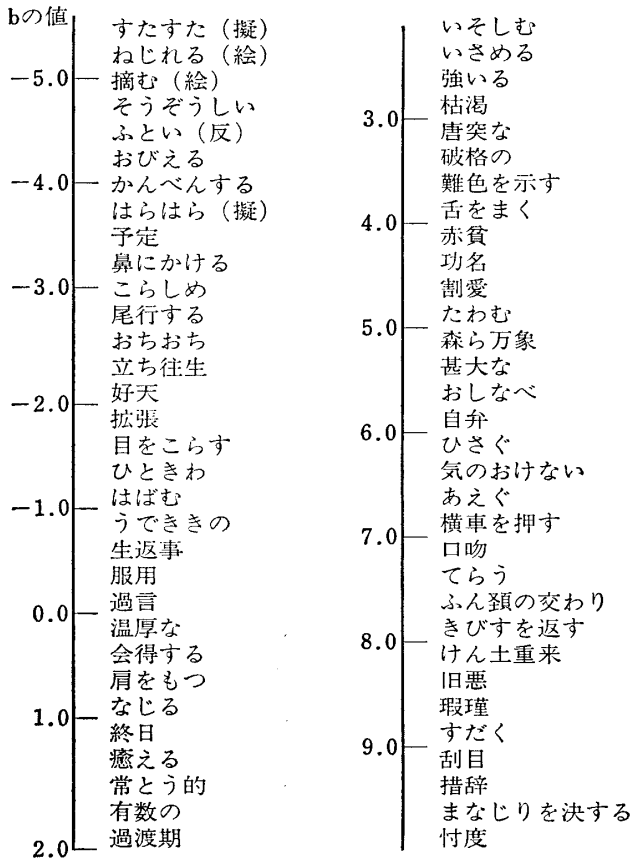




図6 いろいろの困難度の項目の主題の語の見本  
(左側の目盛は共通尺度)



が、共通尺度に変換をすることによって、各版の分布はその平均と分散が変化する。変換後の $\theta$ の分布はそれぞれ正規分布 $N(d, k^2)$ となる。このようにしてもとめられた各版の被験者の $\theta$ の分布を共通尺度上に描いたものが図5である。各版の被験者はさきにものべたように、各学年の任意標本というものではないので、ここに示された結果をもって $\theta$ の分布の発達の変化の一般的傾向とすることはできないが、その概略についてはうかがうことができる。

図6にはこのような換算によってえられた共通尺度上に、 $b$ の値の大きさにしたがって易しいものから難しいものまで適宜選んだ項目(主題の語)が例示してある。

### 3. 情報関数

各版の測定用具としての評価のために、情報関数を計算した。その結果は図7に示してある。各版がそれぞれどのような $\theta$ の値の範囲において効果的な測定をするのかがわかる。たとえば、A6版についていえば、これは共通尺度上の $\theta$ の値でほぼ-1.0を中心にしたある範囲においてとくに情報の量が大きい。そして、情報の量のピークの値は他のどの版よりも高いことがわかる。

### 4. 観測された得点分布とその理論分布

さきにものべたように、項目パラメタが知られていると

図7 各版の情報曲線

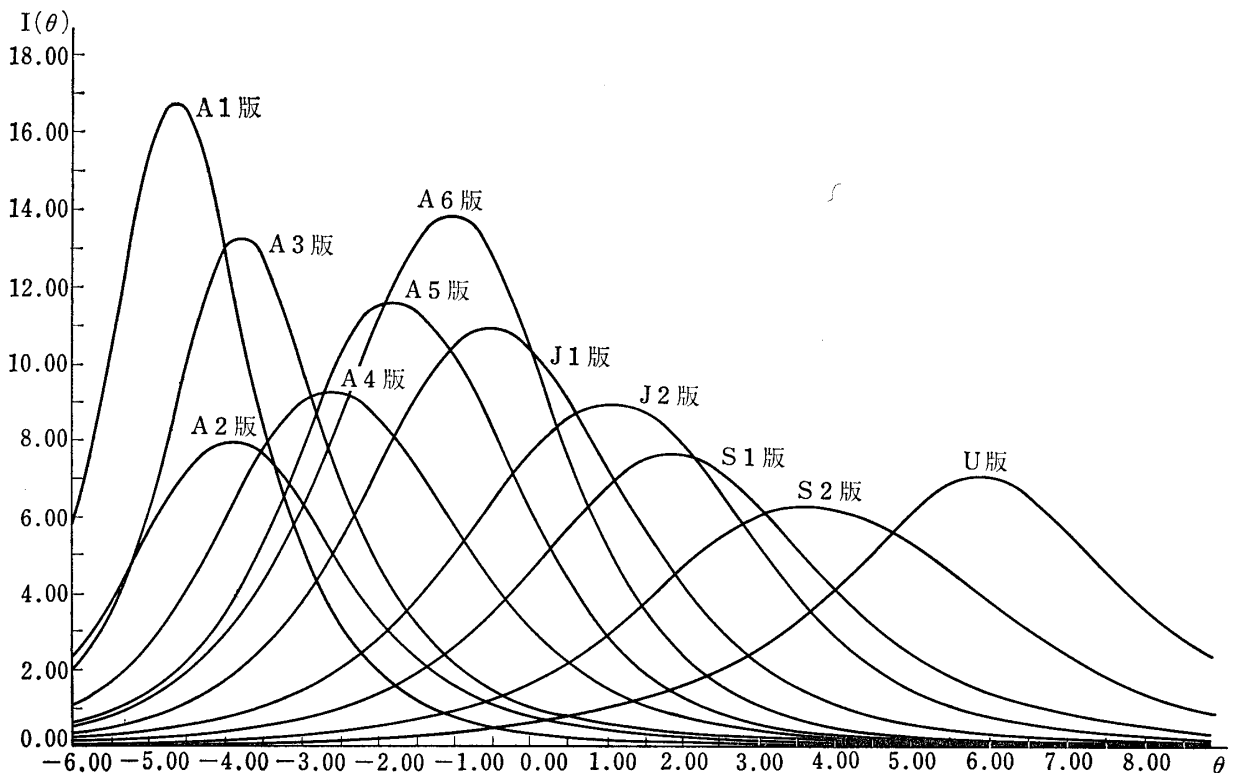
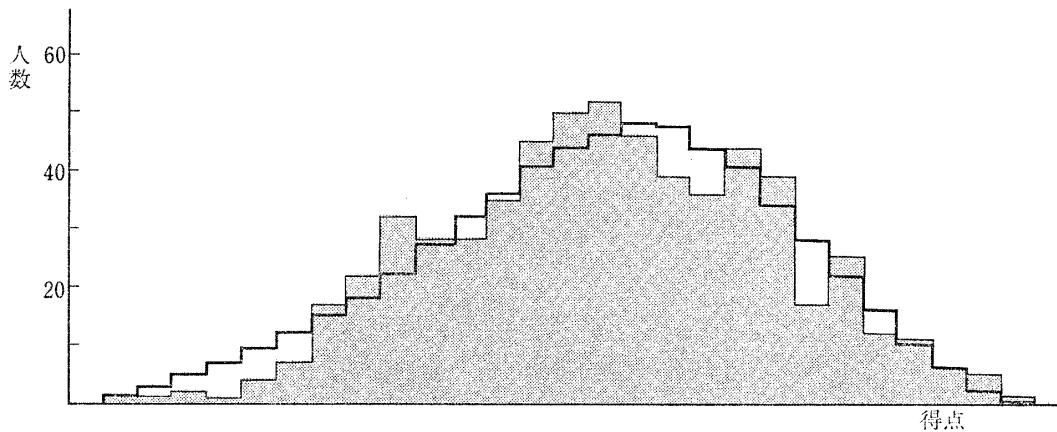


図 8 J1版の観測度数分布（灰色部分）と理論度数分布。得点は2点刻みにとってある。



きには、被験者の $\theta$ の分布を仮定することによって、得点分布を理論値としてもとめることができる。図8にはJ1版について、観測得点とその理論分布が示されている。理論分布がかなりよく一致していることがわかる。得点分布の形は潜在特性 $\theta$ の分布形にも依存しているが、またテストの中に含まれる項目の特性にも依存している。J1版の得点分布がかなり扁平になっているのは、この版に含まれる項目の中に識別力の高い項目が多いことによる。

## V 尺度化の意義

### 1. 尺度の意味づけ

本尺度を構成するためにもちいられた被験者集団は、さきにものべたように特定の母集団を代表する標本ではない。各年齢層の被験者についてはそれぞれ潜在特性の分布が正規分布をするということが仮定されただけであって、統計的な意味で各年齢層を代表するものとはみなされていない。被験者として異った年齢層を適宜もちいたのは、広範囲にわたる困難度の項目を統計的に解析して尺度を構成するためであって、年齢そのものは尺度化に全くもちいられていない。この尺度は、多数の項目に対する多数の被験者のデータから、一つのモデルにもとづいて構成されたものであり、語の理解力の個人差の記述に利用しうるが、その尺度値は発達の意味などをもつものではない。そのような意味づけは、この尺度にもとづいて、むしろこれから行わなければならない。

個人の理解力を同年齢の理解力の分布にてらして評価することがあるが、そのためには年齢規準を定めなければならない。年齢規準を厳格に定めるには、各年齢を代表する正しい標本を抽出する必要がある。ふつう、標準化されたテストでは何らかの形でこれが実行されているが、満足すべき標本をもちいるためにはかなり困難が伴

う。

このような標準化の意義を否定するものではないが、多くの困難を押しつけて標準化をおこなったとしても、実際利用上必ずしも有効であるとは限らないことや、時間的変化によって情報が役立たなくなる点などに疑問が残る。そこで、尺度の意味づけのために、もっと素朴な方法をとることも考えられる。一般的な母集団を想定しそれを推定するために任意標本をとることを唯一の方法とすることをやめ、いかなる被験者集団でも、それを具体的な情報として尺度上に位置づけ、それらの情報を蓄積することによって尺度値の解釈の範囲をひろげていこうというものである。特定の個人や集団の理解力をこの尺度上であらわし、その情報を蓄積し、利用がある個人や集団について理解力の推定値をえたとき、これと比較できるように提供すれば、さまざまな角度から解釈することもできよう。地域、年齢、時代を異にする情報もそれなりの関係において有用な情報となる。本研究において、まったく恣意的に各学年の被験者が選ばれているのは、このような情報蓄積の第1歩である。

### 2. 自由編集のテストと項目ライブラリ

本研究でもちいられた各版のテストは、尺度構成のために仮に編集されたもので、この尺度の特徴は、むしろこのような固定的な版をはなれて、各項目の特性を明らかにしたことにある。すべての項目について、同じ尺度上における困難度と識別力のパラメタが推定されているので、これをもちいて利用者が利用目的にもっともよく適した項目を選んでテストを編集することができる。新しく編集したテストをもちいて、被験者の潜在特性を推定する方法はさきにものべたと全く同じ方法であり、えられる $\theta$ の値は、この研究にもちいられた固定版の場合と全く同じ尺度上のものである。したがって、自由に項目をえらんで編集したテストとはいえ、そのまま固定的な

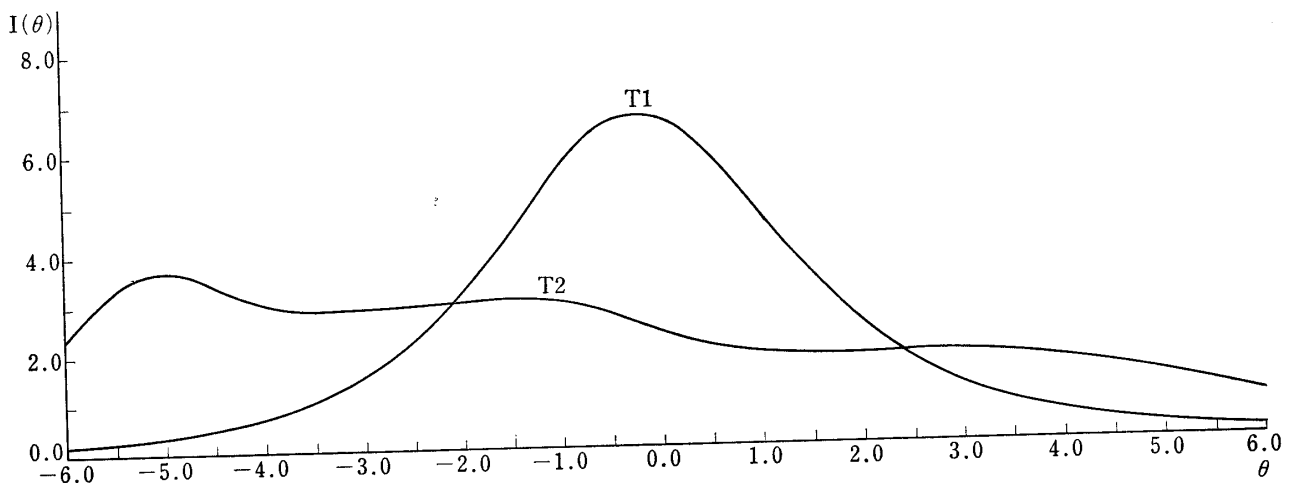
版の結果と全く同じように利用することができる。

このことは実用上いろいろの利点となる。まず、テスト時間の制限やテスト結果の精度への要求の違いに応じて、テストの項目数を増減することができる。また、被験者の理解力の分布に応じて、項目を選択すれば、いろいろの要求に応じたテストを構成することができる。たとえば被験者の理解力がある狭い範囲に分布しているときには、その分布の範囲において情報量の高い項目を集中的に選ぶことによって、精度の高い測定をすることができる。これに対して、理解力の分布が広い範囲にわたっている被験者集団が予想されるときには、広い範囲においてテストの識別力が高くなるように、いろいろの困難度の項目を選ばなければならない。これらの事情をよく説明するのが情報曲線である。例として、 $\theta$ の値が0.0あたりの狭い範囲で情報量の多くなるテスト(T1)と、 $\theta$ の値が-5.0から5.0にわたる広い範囲でできるだけ情報量が多くなるようなテスト(T2)のため表4のような項目を30ずつ選んだ。そして、これらのテストについて情報曲線を計算し、図9に示した。図にみられるとおりに、T1はたしかに $\theta = -.25$ のあたりを中心

表4 情報量を比較するために編集された2つのテスト。T1は困難度パラメタ**b**が狭い範囲に分布しているもので、T2は**b**が広い範囲に分布しているもの。

テスト	項目数	a		b	
		平均	標準偏差	平均	標準偏差
T1	30	.54	.17	-.15	.38
T2	30	.62	.26	.00	3.59

図9 項目困難度がほぼひとしい30項目からなるテスト(T1)と、困難度が広範囲に分布する30項目からなるテスト(T2)の情報曲線



に、比較的狭い範囲において情報量が高い。これに対し、T2の情報曲線は全体的にかなり低いが、たしかに $\theta = -6.0$ から $\theta = 5.0$ あたりまでの範囲にわたってひろく測定するのにふさわしい情報曲線を示している。情報量が低いのは項目数の限界によるもので、情報関数の値を広い範囲にわたって高くするためには、項目の識別力の値がいずれも同等であるなら、項目数を多くする必要がある。このことは情報曲線の下面積が

$$(24) \quad A = \int_{-\infty}^{\infty} I(\theta) d\theta = 1.7 \sum_{j=1}^n a_j$$

のように、項目の識別力パラメタの和に比例することからも明らかである。

もし、被験者の $\theta$ について事前にある程度情報があれば、それを利用し、そのあたりにおいて情報関数の値が高くなるような項目ばかりを選択して与えれば、比較的少数の項目で精度の高い測定をすることができる。それを実際におこなうために、いわゆるプログラム化されたテストのいろいろの試みがおこなわれている<sup>17)18)</sup>。

本研究で特性の明らかになった356項目は易しいものからかなり難しいものまでを含み、一応これだけあればふつうの測定には十分であろう。しかし、さらに精度の高い測定をしたり、あるいは、何版かの平行テストを必要とすることもあろう。また、日本語の語彙の大きさからみると、これはまことに微々たるものであり、もっと多くの語を項目化してライブラリに加えることが望ましい。本研究のために用意された項目は、この種の項目ライブラリの実現のための最小限の見本のようなもので

ある。項目を新たに加えるには、随時予備的テストを実施し、ある程度の条件（困難度や識別力についての）が整ったところで、項目パラメタを求めるためのテストを実施すればよい。そのときの尺度の等化は、原則的に本研究でおこなったと同様の方法によることになる。

これまでの議論ではふれなかった重要な問題がある。それは項目特性の標本分布の問題である。被験者の標本抽出についてはすでにくりかえしのべたように、標準集団となる母集団を想定する必要はないので問題はないとしたが、項目特性の算出という点からいえばここにもちいられた被験者の数はけっして十分とはいえない。そのため項目特性の推定値もそれほど安定したものとはいえない。このことは、たとえば図2や図3をみれば明らかである。もし、モデルがよく適合し、項目特性の標本誤差が十分小さければ、これら図における $a$ や $b$ の分布は一直線上に乗るはずである。実際にはかなりそれからはずれている。今後、データが蓄積されたときにそれらを利用してより大きな被験者数による項目特性の再推定をする必要がある。つねに最善の推定値になるようデータの累積的利用のシステムを作ることは今後の課題である。

#### 付記

本研究は数年にわたり、多数の学校の協力を得ておこなわれたため、非常に多くの方々のお世話になりました。御協力いただいた学校は次のとおりです。

和泉小学校、浜田山小学校、沓掛小学校、松庵小学校、済美小学校、永福小学校、西田小学校、杉並第四小学校、杉並第六小学校、桃井第一小学校、桃井第二小学校、井荻小学校(以上12校は杉並区立)、鷺宮小学校(中野区立)、松溪中学校、天沼中学校、高井戸中学校(以上3校は杉並区立)、聖霊女子短期大学付属中・高校、東京大学教育学部付属中・高校、杉並高校、荻窪高校、豊多摩高校(以上3校は東京都立)、K大学、T大学。

御関係の先生方、生徒の皆さんに心から感謝します。

項目原案の作成には、言語研究グループの撫尾知信、大嶋百合子、栗山容子、斎藤こずゑ、関戸百合子、武井澄江、辰野俊子、秦野悦子、三宅なほみの諸氏の御協力をいただき、絵による項目の描画については小番賀津子氏にお世話になりました。また塗師斌助手および野口裕

之、南風原朝和両氏からは、面倒な計算に関して有益な助言と協力をいただきました。野田章子氏には項目の作成をはじめ研究の計画と実行全般にわたって多大の協力をいただきました。このほか、お世話になった多数の方々に、この機会にあらためてお礼申しあげる次第です。

なお、本研究に関連して昭和51年度科学研究費の補助を受けました。

#### 注

- 1) 阪本一郎 読みと作文の心理 牧書店, 1955.
- 2) 森岡健二 義務教育終了者に対する語彙調査の試み, 国研年報2, 秀英出版, 1951.
- 3) Lord, F. M., A theory of test scores. Psychometric Monograph. No. 7, 1952.
- 4) 印東太郎・鮫島史子 LIS 推測因子測定法—non verbal—, 日本文化科学社, 1962.
- 5) Lord, F. M. and M. R. Novick, Statistical Theories of Mental Test Scores. Addison-Wesley, 1968.
- 6) Lord, F. M. 1952 op. cit.
- 7) Bock, R. D. and M. Lieberman, Fitting a response model for  $N$  dichotomously scored items. Psychometrika, Vol. 35, No. 2, 179—198.
- 8) Birnbaum, A., Some latent trait models and their use in inferring an examinee's ability. In Lord and Novick, 1968, op. cit.
- 9) Ibid.
- 10) 芝 祐順 因子分析法 東大出版会 1972.
- 11) 林 四郎 語彙調査と基本語彙 電子計算機による国語研究Ⅲ 国研報告39 秀英出版 1971.
- 12) 真田信治 基本語彙・基礎語彙 岩波講座日本語 岩波書店 1977.
- 13) 阪本一郎 教育基本語彙 牧書店 1958.
- 14) Guiford, J. P., Psychometric Method, Mc Graw Hill 1954 秋重義治監訳 精神測定法 培風館 1959.
- 15) 芝 祐順 項目分析 肥田野直編 テストⅠ 東大出版会 1972.
- 16) Kirk, D. B., On the numerical approximation of bivariate normal (tetrachoric) correlation coefficient. Psychometrika, Vol. 38, 259—268, 1973.
- 17) Weiss, D. J., Strategies of adaptive ability measurement. Research Report 74—5, Psychometric Methods Program. Dept. of Psychology., Univ. of Minnesota. 1974.
- 18) 芝 祐順・野口裕之・南風原朝和 語彙理解力測定のための多層適応形テスト 教育心理学研究 第26巻 第4号 (予定) 1978.