

心理測定理論と家族的類似の概念

村 上 隆

1 問題の所在

質問紙による個人差測定尺度を用いた相関研究を考える。そのような研究では、複数の構成概念に対応した質問項目群が用意され、それらに対する反応を何らかの形で合成した尺度が作られる。次に、それらの尺度間の相関係数が計算されたり、被験者群間の平均差が検討される。それらの結果は、心理学的な構成概念 (construct) 間の関係や、構成概念の性質として解釈される。さらに、それらにもとづくさらに進んだ解析 (因果分析等) がなされることもある。

そこで問題になることは、構成された尺度がいかなる意味で当該の構成概念の測定値を与えることになるか、すなわち、尺度の妥当性の問題である。多くの場合、それらの構成概念を直接的に反映する外在基準は存在せず、基準関連妥当性の考え方は使えない。通常の手続きは、質問項目の適切性を複数の専門家に判断してもらうこと、得られたデータを、単純構造への回転をとまう因子分析にかけて項目を分類し、構成概念を複数の次元に分解することである。これらの方法によって検証されるものは、(恐らく不適切に) 内容的妥当性と因子的妥当性と呼ばれている。さらに尺度間の関係が、それらの間に想定される理論的予測と一致するならば、尺度の構成概念妥当性が (部分的に) 検証されたことになると思われる。

これらの手続きは十分な理論的基礎づけを欠くものであり、次のような疑問が生ずることはもっともである。たとえば、専門家の検討を経た項目群を分析した結果、2つ以上の因子が存在することが明らかになったとすれば、その段階でその構成概念は棄却されるべきなのではないか。また、理論的には高い値をとると予測される2つ尺度間の相関が実際には低い場合、それを理由に尺度の妥当性を否定するとすれば、理論がデータによって反証される可能性はありえないことになり、経験的な研究は無意味と化すのではないか。一方、相関が高いとすれば、それは2つの構成概念が実は同一であることを示すに過ぎないのではないか。さらに、質問項目全体を同時に因子分析しないのはなぜか。それによって見出され

る因子こそ構成概念そのものなのではないか。

本研究は、これらの批判を考慮しつつ、心理測定、特に個人差測定のための尺度構成の手続きと、さらにその根底にある個人差のイメージについて再検討することを意図したものである。特に、専門家による項目内容の判断と、領域ごとの因子分析という手続きは、適切に注意深く実行されれば、少なくとも研究の比較的初期の段階における validation の方法として正当である、というのが、一応 positive な結論である。

まず第一に、明確な外在基準が存在する場合を想定しながら、信頼性と妥当性とが究極的には矛盾した要請であること、その矛盾を解消するためには、心理測定尺度が1因子でなければならないことを示す (2節)。

次に、1因子の心理測定尺度を構成することが、現実問題としては不可能であることを示す (3節)。以上2つの節の結論は、古典的テスト理論における個人差の概念が、不十分なものであるということである。

その上で、心理学的構成概念が、本質的に多次元的なものとして想定されざるを得ないことを論じ、それとともに、心理尺度構成の基礎となる項目間相関と、因子分析の結果に影響を与えると考えられる諸要因について検討する (4節)。

最後に、そこまでの議論をもとに、妥当性の検討方法について考察するとともに、若干の提案を行い、今後の検討の方向について展望する (5節)。

2 信頼性と妥当性のディレンマ

まず、古典的テスト理論の主要部分を、必要な限りで振り返っておこう。証明は一切省くが、池田 (1973) 等により容易に確認することができよう。

古典的テスト理論では、テスト得点 X を、真の得点 T と無作為誤差 E の和に分解して考える。 E の平均値が0であること、および、 T と E が独立であること、(ともに、無作為性の定義と見なしうる)、すなわち、

$$\mu_X = 0 \quad (2-1)$$

$$\sigma_{TE} = 0 \quad (2-2)$$

を仮定する。ただし σ_{TE} は T と E の共分散を示す。 X , T , E の分散をそれぞれ, σ_X^2 , σ_T^2 , σ_E^2 , とすると, (2-1), (2-2) から,

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2 \quad (2-3)$$

が成立する。これにもとづいて, テスト得点の信頼性係数 ρ_X は,

$$\rho_X = \sigma_T^2 / \sigma_X^2 \quad (2-4)$$

と定義することができる。

一方, テストに対する適切な外在基準を Y とする。妥当性は, X と Y の相関の程度として定義される (基準関連妥当性。古典的テスト理論では, これがほぼ唯一の妥当性の概念である)。 X と Y の相関係数を ρ_{XY} と書くことにする。

ここで, Y の信頼性係数を ρ_Y とすると, 次の希薄化の公式,

$$\rho_{XY} \leq \sqrt{\rho_X \rho_Y} \quad (2-5)$$

が成立する。これは, 信頼性が妥当性の必要条件である (X の信頼性が低いと, 必ず妥当性も低い) ことを示すと解される。

ここで考えるような質問紙調査では, テスト得点は複数の項目得点の総和として定義される。すなわち,

$$X = \sum_i x_i \quad (2-6)$$

それぞれの項目も, 真の得点 t_i と無作為誤差 e_i の和であると仮定し, 無作為誤差について, きわめて常識的な次の仮定をおこなう。

$$\mu_{e_i} = 0 \quad i = 1, \dots, n \quad (2-7)$$

$$\sigma_{t_i, e_j} = 0 \quad \begin{matrix} i = 1, \dots, n \\ j = 1, \dots, n \end{matrix} \quad (2-8)$$

$$\sigma_{e_i, e_j} = 0 \quad \begin{matrix} i = 1, \dots, n \\ j = 1, \dots, n \\ \text{ただし } i \neq j \end{matrix} \quad (2-9)$$

これらはそれぞれ, 無作為誤差の平均値は0であること, 真の得点と無作為誤差は無相関であること, 異なる項目の無作為誤差は相互に無相関であること, を意味している。また n は項目の数である。

さて, この仮定の下で,

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_i \sigma_i^2}{\sum_i \sum_j \sigma_i \sigma_j \rho_{ij}} \right) \quad (2-10)$$

は, Cronbach の α 係数と呼ばれる。ただし, σ_i ,

σ_i は項目の標準偏差, ρ_{ij} は項目間の相関係数である。これは, (2-5) ~ (2-7) の仮定の下で,

$$\rho \geq \alpha \quad (2-11)$$

すなわち, 信頼性係数の下限であることが証明される。

また, 項目と外在基準の相関係数を ρ_{iY} とすると

$$\rho_{XY} = \frac{\sum_i \sigma_i \rho_{iY}}{\sqrt{\sum_i \sum_j \sigma_i \sigma_j \rho_{ij}}} \quad (2-12)$$

となる。

式 (2-10) は, 他の条件が一定なら項目間相関係数 ρ_{ij} が大きいほど, テスト得点の信頼性が高くなることを意味している。ところが, (2-12) は, 他の条件が一定なら ρ_{ij} が小さいほどテスト得点の妥当性が高くなることを示している。このことは, テストの信頼性を高めることと妥当性を高めることが, 必ずしも両立しないことを意味している。しかも, このことは (2-5) と矛盾するように見える。

この「パラドックス」は, n 個の項目が弱平行測定 (池田, 1973) であるか, 少なくとも1因子であれば解消する。しかしそうでなければ, ρ_{XY} と α を両方とも最大化することは不可能で, 項目選択にはディレンマが生ずる。以下にそのことを示す。

モデルとして想定されるものはいくらでも複雑にすることができるが, 一応線型の範囲で考えるのが適当であろう (後に示すような超多次元モデルを, 強引に線型で近似していると考えればよい)。そこで, 次の多因子モデルを考える。すなわち, 項目得点が q 個の因子によつて,

$$x_i = \sum_m a_{im} f_m + e_i \quad i = 1, \dots, n \quad (2-13)$$

と表せると仮定する。ここで, f_m ($m = 1, \dots, q$) は因子得点, a_{im} は, 項目 i の第 m 因子への負荷であり,

$$\sigma_{f_m f_{m'}} = \begin{cases} 1 & m = m' \\ 0 & m \neq m' \end{cases} \quad (2-14)$$

$$\sigma_{f_m e_i} = 0 \quad (2-15)$$

$$\sigma_{e_i, e_{i'}} = 0 \quad i \neq i' \quad (2-16)$$

を満たすものとする。外在基準も同様に,

$$Y = \sum_m b_m f_m + e_Y \quad (2-17)$$

と表せるとする。 b_m は, 外在基準の負荷である。

このモデルの下では、

$$\sigma_x^2 = \sum_m (\sum_i a_{im})^2 + \sum_i \sigma_{e_i}^2 \quad (2-18)$$

$$\sigma_y^2 = \sum_m b_m^2 + \sigma_{e_y}^2 \quad (2-19)$$

$$\sigma_{xy} = \sum_m (\sum_i a_{im}) b_m \quad (2-20)$$

したがって、

$$\rho_{xy} = \frac{\sum_m (\sum_i a_{im}) b_m}{\sqrt{\sum_m (\sum_i a_{im})^2 + \sum_i \sigma_{e_i}^2} \sqrt{\sum_m b_m^2 + \sigma_{e_y}^2}} \quad (2-21)$$

であり、また、

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_m \sum_i a_{im}^2 + \sum_i \sigma_{e_i}^2}{\sum_m (\sum_i a_{im})^2 + \sum_i \sigma_{e_i}^2} \right) \quad (2-22)$$

となる。そこで1因子の場合について考えると、

$$\begin{aligned} \rho_{xy} &= \frac{(\sum_i a_{i1}) b_1}{\sqrt{(\sum_i a_{i1})^2 + \sum_i \sigma_{e_i}^2} \sqrt{b_1^2 + \sigma_{e_y}^2}} \\ &= \frac{1}{\sqrt{1 + \sum_i \sigma_{e_i}^2 / (\sum_i a_{i1})^2} \sqrt{1 + \sigma_{e_y}^2 / b_1^2}} \end{aligned} \quad (2-23)$$

$$\alpha = \frac{n}{n-1} \left(1 - \frac{\sum_i a_{i1}^2 + \sum_i \sigma_{e_i}^2}{(\sum_i a_{i1})^2 + \sum_i \sigma_{e_i}^2} \right) \quad (2-24)$$

だから、この場合は、 ρ_{xy} も α も、ともに項目の負荷の増加関数となり、パラドックスは生じない。

しかし、因子数が2以上の場合には、 a_{im} が b_m と比例しないかぎり、 a_{im} の大きい項目は、 α は増加させるものの ρ_{xy} は減少させる。

こうして、古典的テスト理論において、信頼性と妥当性がともに高まるための条件は、テストが1因子であることである。このことは、1因子性、あるいは平行測定条件を満たすテストを作成することに努力の大部分が費やされ、妥当性の向上にはあまり力が注がれないという結果を生み出したのではないとも思われる。テストが1因子である場合には、信頼性の向上が、自動的に妥当性の向上を意味するからである。

たとえば、Gulliksen (1950, p.88) は、「妥当性は、テストの固定的かつ単一の特性と見なすことはできない。あるテストの新しい利用方法がもくろまれたときには、新しい妥当性係数が決定されなければならない。また、テストの使用が継続されるならば、妥当性係数は定期的

に再決定されなければならない」と述べている。ここで、テストの項目を入れ換えて妥当性を改善する、という方向が示唆されていない点に注目すべきである。このGulliksenの議論は、外在基準は複数の因子は反映するが、テストそのものは1因子である、という事態を想定していると考えれば、整合的に理解できる。

3 1因子の不可能性

このように考えると、実際の測定尺度においてそれが1次元であることを確認することは、極めて重要である。しかしそのことは実際問題として不可能に近い。

1つの尺度に対応する項目群が1因子であるとは、どの2つの項目についても、

$$\rho_{ij} = a_i a_j \quad (3-1)$$

とあらわせることである(先の(2-13)はデータを基準化しないで考えていたが、ここでは基準化して相関モデルで考える。結論に本質的な影響はない)。

この条件を満たすためには、一部の項目間相関が高すぎてもいけない。たとえば、 n 個の項目については、すべて(3-1)が満たされているとしよう。そこにたとえば、項目 n と極めて相関の高い $n+1$ 番目が追加されたとする。項目 n と $n+1$ の相関の説明のために、 a_n は大きくしなければならないが、それは他の $n-1$ 個の項目との相関係数と整合しないことになる。

これは決して特殊な事態を想定しているわけではない。質問紙調査では、ある項目を、わずかにワーディングを変えるだけで、別の項目を作ることができる。もとの項目と新しい項目とを、同一の質問紙の中に共存させれば、ここで述べたような事態となる。

平行測定の仮定の下では、2つの項目間相関が1にならない理由は、もっぱら誤差の存在のせいになる。もしそうなら、いかに見掛けの似た項目であっても、そこだけ異常に相関が高まることはありえないはずである。しかし、実際に項目間相関には、項目間の意味的な類似—非類似が大きく影響しており、項目相互間の距離をすべて一定に保つことができない限り、平行測定はありえない。

1因子性についてもほぼ同じことが言える。すなわち、ただやみくもに内容の似た項目を集めても1因子のテストができるわけではなく、1因子の項目は相互に適切な距離を保っていなければならない。これは極めて微妙な性質であるが、このことは、従来あまり指摘されてこなかったように思われる。このことから、少なくとも次の2つの教訓が得られる。

(1) 項目—合計点相関にもとづく項目分析の手続きは、

外在基準との相関については無視して、1次元の尺度を構成する方法として見ても、不十分、あるいは不適切である。

(2) 項目間構成は、同時に実施、分析される他の項目のコンテキストの中で定まるものであり、(たとえば) 1因子性は項目自体の持つ性質ではない。

要するに、1因子の測定尺度を構成することは、現実問題としては不可能なのである。

4 特性に関する別の見方 ——家族的類似のモデル

しかしながら、一つの構成概念を単一の次元と対応させる必要はない。構成概念は Wittgenstein (1953) の意味における「家族的類似」に基づいて成立しているにすぎないと考えられる。いささか晦渋な Wittgenstein 自身の代わりに、ここではブルア (1988) の解説を引用しよう。

「一つの概念に属するものの実例を調べてみると、我々はしばしば、それらを一まとめにするための手掛かりになるような、成員全てに共通する性質なるものは存在しないことに気付く。その代わりに見出されるものは、時に、全面的な類似であり、時に、細部における類似であるという具合に『互いに重なりあったり、交差し合ったりしている複雑な網の目なのである』。…家族の成員は互いに似通っているが、それは全員が、例えばだんごっ鼻といった、ただ一つの性質を共有しているからとは限らない。似通っているという印象は、例えば娘は父親譲りの顎を持ち、他方その息子は彼女と同じ鼻を持つが髪の毛の色は叔父と同じであるという風な事実に由来する方が多いのである」(p.47)

前節で述べたような、1因子の不可能性は、1つの構成概念に対応すると想定される質問項目が、ここで言う1つの家族のようなものであると想定すれば、矛盾なく理解できる。すなわち、一つの構成概念に対応する質問項目のすべてに共通するような意味的特徴は存在せず、それらは多くの特徴を相互に部分的に共有しているにすぎない。また、項目の背景にある意味的特徴の数は、事実上無限であろう。その結果、あらゆる構成概念は、境界がはっきりしないあいまいなものとならざるを得ない。式(2-13)は、このような超多次元モデルを、線型近似したものと考えることができる(図1)。

ところで、ほぼ同様のことが、個体(被験者)の側についても言える。ある特性の測定に関して、適切な母集団を明確に定義することは難しい。あらゆる年齢段階のすべての人間を母集団と定義すれば、ほとんどすべての能力に関する特性は、1因子に近く見えてくることにな

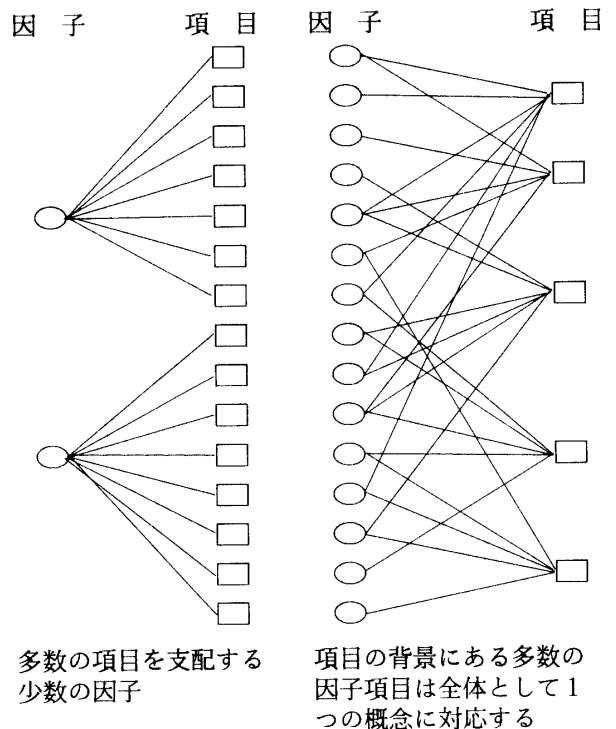


図1 心理測定における従来の考え方(左)とここでの考え方(右)

ろう。それに対して被験者のある属性をもつものだけに限定すれば、全般的に項目間相関は低下し、多数の因子が見出されることになろう。しかし、多数の属性を考慮して、被験者を限定していけば、際限のないことになってしまうであろう。また、考慮されるあらゆる属性に関して同じであるような被験者は、属性の数を増すにつれて、どんどん減少していつてしまう。要するに、個体を、内部的に等質なグループに分割するための、自然な根拠は存在しない。

しかしながら、母集団を決定しない限り、測定尺度を確定することはできない。ここには、個人間の分散なしには、なにごとにも測定不可能であるという、心理測定における根本的な問題が横たわっている。

これは、共通因子の数を1と想定する Spearman の2因子説と、Thurstone の多因子説との、基本的対立点の一つでもあった。

このように考えることの意義の一つは、問題の構想と測定方法(研究方法)の検討の段階で、研究者の考慮の及ぶ範囲には常に限界がある、という事実の認識であると思われる。研究者は当然、項目全体を完全に一樣なものとして構想するのではなく、幾つかの下位概念の存在を想定していることが多い。被験者についても、性別、年齢等、必要な属性を、標本抽出、あるいは分析の単位として考慮している。しかし、それらが重要な要因(因

子)を網羅しているという保証はない。データが収集された段階で、その変動を説明するためには、より重要な要因が発見される(あるいは、発見できずに、「誤差」として処理されざるを得ない)ことが多い。

したがって、それぞれの部分の因子分析によって導出される次元は常に、使用されている項目と、データ収集の対象となった被験者に依存した、暫定的なものにとどまる。特に、得られたデータの因子数を、厳密に決定しようとすることは意味がない。統計的な議論としては、被験者の数を十分大きくすれば、有意な因子数は、まず間違いなく $n-1$ となるはずである。どんな2つの項目の対を考えても、その2つの項目だけが共有している特徴が存在するはずだからである。また、因子寄与の大きさや、個々の因子負荷の数値の大小を詳しく比較することも、ほとんど無意味である。

しかしながら、因子分析そのものが無意味と化すわけではない。因子分析の結果を、現象の忠実な描写であると考えことを止めさえすれば、因子分析の結果から、心理学的構成概念について、多くの知識が得られる。

家族的類似のモデルによれば、1つの構成概念は、無数の次元の「束」として認識せざるを得ない。それは必ずしも、1つの次元を構成するとは限らないばかりか、先程の因子数の議論からすれば、使用された項目は $n-1$ 次元以下の空間に収めることはできないであろう。しかしながら、現実の個人差は、研究者が1つの構成概念として構想しているような領域の範囲内ですら、超多次元的な広がりを持っているものと思われる。そこに、あえて少数の次元を設定していくのが因子分析の役割であり、これはソシュールの一般言語学における言語の機能にも似ている。丸山(1981)によれば、

「ソシュール以前は、コトバは表現でしかなく、すでに言語以前からカテゴリー化されている事物や、言語以前から存在する純粹觀念を指し示す道具と考えられていたが、ソシュール以後の考え方では、コトバは《表現》であると同時に《意味》であり、これが逆に、それ自体は混沌たるカオスの如き連続体に反映して現実を非連続化し、概念化するということになる。」(p. 120)。

構成概念は、客観的実在として、測定以前に存在しているのではない。個人差は「カオスのごとき」超多次元の中にあり、因子分析はその中の少数の次元として、個人差を「概念化」する。

「我々の生活世界は、コトバを知る以前からきちんと区分され、分類されているのではない。それぞれの言語のもつ単語が、既成の概念や事物の名づけをするのではなく、その正反対に、コトバがあってはじめて概念が生まれるのである。」(pp. 117~118)。

ソシュールは、言語による世界の分節化が、まったく「自然的基礎」をもたず、「恣意的」であるとする。

「日本語の「木」は、机などを作っている材料であれば、庭の青々とした樹木でもあるが、英語では前者が wood、後者が tree である…。それでは、材木の意味の「木」と wood が完全に重なりあう概念であるかという、これもそうはいかない。wood には、「森」という意味も含まれているからである。」(p. 118)

因子は、項目が別のものに入れ換えられたり、被験者が変わったりすれば、変化する。

最後に、我々日本人が、紫、藍、青、緑、黄、橙、赤の七色から構成されているとする虹の色が、英語では6色、ショナ語では3色、サンゴ語やバッサ語では2色とされるという指摘がある(p. 118)。これは、因子数の決定に関して、純粋に客観的な方法がありえないことと、まさしく平行関係にあるように思われる。

因子、すなわち、個人差の量的次元は、あらかじめ客観的に存在しているのではなく、それはデータから、因子分析という手法を通じて定義される。特に単純構造化は、項目のクラスターとして、近似的に概念(次元)の境界を明確化する効果をもつ。もちろん単純構造化は、必ずしも「自然的基礎」を持つものではない。それぞれの項目は、同時に多数の因子に負荷しているのであり、どの因子に負荷するかは、項目ごとにさまざま、というのが「自然な」状態であろうと考えられるからである。しかしながら、結果は究極的には日常言語に近いもので説明されなければならないから、言語そのものと類似の構造をもつことになる。これはいわば必然である。

ただし、因子分析の結果は、言語のように全く恣意的なものではあり得ない。因子分析の最適化基準は、同一のデータに関しては(因子数、回転方法等の選択による違いを除けば)、通常、一義的な解を生み出す。

こうした議論は、結局のところ因子分析による因子の認識論的位置、といった論点に行き着く。因子は実在するのか、あるいは、単なる数学的便宜にすぎないのか、という議論は古くからあった。この節の議論は、因子の実在性に疑問を投げ掛けるものである。しかし、一方、因子を完全に数学上の虚構とするものでもない。因子は、研究者のもっている現象に関する理論的枠組みと、まったく独立ではありえないからである。

結局、単一の構成概念に対応すると考えられた項目群を因子分析して、複数の因子を取り出すことは、必ずしも構成概念の否定につながらないということが、この節の議論の結論である。ただし、あくまでも因子分析の結果を、事実の忠実な描写とは見ないこと、さらに、事実の近似としてもなお、因子分析の結果は、少なくとも以

下のような諸条件の影響を受けることを認識した上での話しであるが。

- 1) 項目の選択。研究者自身が持っている（多くは暗黙裡の）理論的枠組みにもとづく選択の効果。
- 2) 被験者のレンジ。多くの場合、経済的にやむをえない事情もあるが、やはり、研究者自身が持っている、問題意識にも依拠する。
- 3) 分析方法の選択。あるいは、因子分析における因子数の選択。
- 4) 被験者の属する文化あるいは時代等の効果。被験者の特性を形成する諸要因の布置の違いは、項目間相関を著しく変容させる可能性がある。

このうち、1)～3)は主として方法論上の問題であるが、4)は事実レベルの問題であり、ここでは少し区別して扱っておく方がよいかもしれない。

ところで最後に、同一の構成概念に対応する項目から、2つ以上の因子を取り出すことについて、消極的ながら1つのメリットをあげておこう。少なくとも、2つ以上の因子得点間の、他の構成概念に対応する尺度との相関が一様に高いものでないことは、validationとして有効である。なぜなら、それにもとづいて測定値が内容と無関係な反応の構え（あるいは方法分散）だけからなっていることを否定することができるからである。

5 妥当性検討の方法について

こうして2つ以上の構成概念間の関係を検討するための研究を開始したとしよう。質問紙の作成にあたっては、一つの構成概念に対して一つの項目の集合（当然複数の項目を含む）を対応させることになる。前節の議論にもとづき、それらはいずれも1次元である必要はなく、（暫定的な）因子間の相関が、構成概念の関係を与えることになる。

ところが、構成概念の成立の根拠を家族的類似に求めるとすると、そこにおける一つの問題点は、構成概念が「無限に開ききってしまう」（ブルア、1988）ことである。すなわち、複数の構成概念に対応する項目が内容的に重複することによって、見掛け上の相関関係が生じてしまう可能性である。実際、項目(1)と項目(2)は、何らかの共通の特徴をもつであろう。項目(2)と項目(3)は別の特徴を共有するであろう。こういった連鎖は無限に続き得るであろうから、結局あらゆる項目は、単一の構成概念にかかわる、ということになりかねない。これが、概念が無限に開ききった状態である。すべてが1つの構成概念というのは極端としても、実際的な問題としては、構成概念Aの範囲が拡大する結果として、構成概念Bの領域と重なってしまう、という事態は当然起こりうるで

あろう。

実はこれは、概念の成立の根拠を家族的類似に求めることに對する、最もありふれた批判でもあった。しかしブルア（1988）によれば、Wittgenstein の理論には、このことを防止する仕掛けが用意されていると言う。

「ウィットゲンシュタインの理論は単なる類似の理論ではなく、家族的類似の理論なのである。…類似の他に、血のつながりの役割を果たしている何か他の要因がなければならない。…『家族間の類似』から『家族内の類似』を区別するものがあるはずなのだ。第一に、似ているという判断は受け入れられたある範囲の範例の周りに集中している。…第二に、類似は、常に或る特定の言語ゲームの文脈の中で判断される。」（pp.50～51）。そして、「ウィットゲンシュタインは、言語の使用者は一連の慣習化された実践の中で訓練を受けるということを想定して、（開き切った構造という混沌を）回避しようとするのである。」（p. 51）。

要するにこれは、データにもとづいて解決できる問題ではなく、理論的な問題である。たとえば、ラザーズフェルド（1984）は、次のような例をあげる。「ある有名な研究は、権威主義的パーソナリティを持っている人びとは、少数民族を差別しがちであることを示した。…むしろ少数民族への攻撃を独立の概念であると考えたよりも、むしろ権威主義的パーソナリティを示す指標群の一部として含めることができたであろう。そのような問題に関する決定は、主に実質的および理論的配慮からなされなければならない。」

彼は、この後で、因子分析の結果がその判断のための補助手段となりうることを示唆している。しかし、「フォーマルな分析は実質的な調査研究の内容を完全には指示できない。」

筆者は、構成概念の「開かれすぎ」を回避するためにこそ、専門家による項目内容の判断が用いられるべきであると考え。ただしこの際の判断は、通常なされる各項目が当該の構成概念Aの測定に適切かどうかだけでなく、構成概念Bに誤って属させられないことがないかどうかという観点からもなされなければならない。

ここで言う専門家は、研究者自身と理論的文脈を共有する者、つまり彼と「特定の言語ゲーム」を共有する者であり、彼と同じ「一連の慣習化された実践の中で訓練を受け」た者である。彼の判断は、単なる直観的印象ではなく、研究全体を支える理論的文脈がここに反映されることになる。

この手順によって、アプリオリな命題と、エンピリカルな命題が一応区別され、トートロジーや循環を避けることができる。すなわち、ここでの研究の文脈の中で

は、構成概念間（項目集合間）の相関は、項目内容の重なり（トートロジー）にもとづくのではなく、経験的事実を指すことになる。別の観点からすれば、事前に各集合（領域）が幾つの次元からなるかについての仮説は明確でなくてもよいが、複数の領域が融合する（そういう形で理論的前提が崩れる）可能性はないものとする。したがって、予想された相関が認められない場合、あるいは予想しないところに相関が出現した場合、その理由を尺度（あるいは項目）の妥当性の欠如にではなく、理論の不完全性に帰することができる。すなわち、理論がデータによって修正を受ける一線を、ある程度明確にすることができる。

これはあくまでも原理的な議論であって、実際問題としては、そうまくはいかないことが多いであろう。たとえば、複数の専門家間で意見が一致しないことは十分ありうる。一方、われわれが想定したように、項目の特徴の次元は事実上無限であるとするならば、極めて重要な特徴が専門家によっても見落とされることも大いにありうるであろう。

また、分析の結果から、判断が疑問に付されることもありうる。たとえば、(2-3)の不等式が成立しない場合、すなわち、いわゆる希薄化の修正で相関係数が1を上回る場合がそうである。もちろん、今や1次元モデル（正確には平行測定モデル）にはこだわっていないから、これはただちにシステム全体を崩壊させるわけではないが、構成概念間の特徴の重複を疑わせる。

このあたりの、理論とデータの関係については、なお検討すべきことが多い。

もう少し別の角度から、4節の超多次元モデルの含意を考えてみよう。少なくとも、次のようないささか否定的な認識は必然的である。すなわち、因子分析の結果、あるいは、(2-4)の合計点の形で定義した尺度の性質は、4節の条件(1)~(4)が「たまたま」現在あるような布置をとっていたから成立していたものにすぎず、それらは、何らかの実在する次元の近似値とは考えられない。また、最善の尺度を一義的に定めるような、基準も存在しない。

このことから、ここで考察しているような相関研究において、「進歩」、「研究の積み上げ」といった概念も、実験研究のそれとは違ってくることになるだろう。たとえば、一般に相関係数に関して、推移律は成立しない。たとえば、 $r_{AB} > 0$, $r_{BC} > 0$ であっても、 $r_{AC} < 0$ となることがあり得る。実際、 B を統制した A と C の間の偏相関係数 $r_{AC.B}$ が-1と1の間にある、という条件から、 r_{AC} の下限は、 $r_{AB}r_{BC} - \sqrt{1 - r_{AB}^2} \sqrt{1 - r_{BC}^2}$ であり、 r_{AB} と r_{BC} がともに0.7に達していても、 r_{AC} は

負になることがある。 r_{AB} と r_{BC} がさほど小さくなくても、 $r_{AC} > 0$ となるのは、 A , B , C が同一の因子に高い負荷を持つ場合である。しかし、そうでなければ、 $r_{AB} > 0$, $r_{BC} > 0$ であっても、 A と C は、 B の中の別の成分と相関している可能性がある。

こうして、相関研究においては、一部分ずつ順に固めていく、という研究方法があまり効率の良いものとは言えず、できるかぎり全体を同時的に取り扱い、分析にあたっても、全体の同時的分析が望ましいことになる。この点に関して、筆者の最近の研究（村上，1987等）は、その方向を目指している。

このような困難に遭遇する理由は、相関研究において我々が、人間とその文化という巨大システムを、まともに関与にしなければならぬからであるとも考えられる。実験研究では、不十分ながらそのシステムの一部を切り取ってアプローチすることができる。他の部分は、無作為化によって考慮の外に置くことができるからである。しかし、相関研究では、テーマ自体はいかに小さいものであっても、あらゆる部分から影響をこうむる可能性が否定できなくなる。

筆者がここまで目指してきたものは、なるべく柔らかい心理測定のイメージなのであるが、なお、4節の(4)にあげたような、ダイナミックな側面への言及は不足していることを認めざるを得ない。つまり、「超多次元」モデルにおいて、因子を相互に独立な属性の集合として、いわば平面的に見てしまうことは、特性全体を、極めてスタティックにとらえることにつながっているように思われる。この点も今後の課題である。

文 献

- ブルア（戸田由和訳）1988 ウィトゲンシュタイン：知識の社会理論 勁草書房（Bloor, D. 1983 *Wittgenstein, A Social Theory of Knowledge*. Macmillan Press.）
- Gulliksen, H. 1950 *Theory of Mental Tests*. Lawrence Erlbaum Associates.
- 池田 央 1973 心理学的研究法7 テストⅡ 東京大学出版会
- ラザーズフェルド（西田春彦・高沢健次・奥川櫻豊彦訳）1984 質的分析法 岩波書店（Lazarfeld, P. F. 1972 *Qualitative Analysis: Historical and Critical Essays*. Allyn and Bacon.）
- 丸山圭三郎 1981 ソシユールの思想 岩波書店
- 村上 隆 1987 複数の変数集合の主成分と正準変量

名古屋大学教育学部紀要——教育心理学科——, 34,
235—253.

Wittgenstein, L. 1953 *Philosophical Investigations*. Basil Blackwell.

(1989年7月31日 受稿)

ABSTRACT

Theory of Psychological Measurement and the Concept of Family Resemblances

Takashi MURAKAMI

The concept of unidimensionality is a central notion of classical test theory. As is well known, there is a dilemma between reliability and validity of a test because it is impossible to maximize both the coefficient alpha, an increasing function of variance of the test score, and the validity coefficient, a decreasing function of it simultaneously. But the dilemma is diminished if scores of all the items of the test are explained by a single factor in the basic model of factor analysis. Under the condition, it is demonstrated that both the correlation of an item with another item in the test and the correlation of the item with any external criterion are proportional to the factor loading of the item. In a way, the assumption of unidimensionality guarantees the legitimacy of ordinary item analysis technique.

The condition of unidimensionality, however, can be seldom met in real data, and is easily violated by addition of only one item or slight change of sample. But use of a single test score as a composite of items can be justified on the basis of family resemblances in Wittgenstein's sense. Assume that each test item is characterized by a set of features, and that there exists some overlapping of features between sets although no feature is common to all the items of a test. This network of similarities formed by overlapping of features defines family resemblances and makes it possible to regard a set of items as a representation of a single concept.

Great indeterminacy beyond the reliability-validity dilemma is introduced to test construction process by repealing the unidimensionality. For example, a test can be decomposed into any number of subtests, and the number cannot be determined uniquely because the number of features may exceed the number of items.

This reconceptualization of psychological dimension leads to the re-examination of the process of content validation. As the boundary of universe of no psychological concept based on family resemblances can be drawn definitely, no two experts might agree with the judgment of appropriateness of items. Rather, experts' judgments about exclusiveness of sets of features characterizing two item groups, which are assumed to correspond to different concepts, would be necessary in order to prevent the domain of a concept from expanding unlimitedly.