

認知の能動性に関する予備研究

北林 行雄, 郭 立新, 黒田 靖子,
加藤 ジェーン, 川田 勉, 中山 剛

A Preliminary Study on Activity of Cognition

Yukio Kitabayashi, Lisheng Guo, Yasuko Kuroda
Jien Kato, Tsutomu Kawata and Takeshi Nakayama

Cognitive property of Japanese surnames having peculiar intimacy to each of Japanese as an individual label was investigated. The frequencies of 5778 surnames taken from directory of a society including 30,000 members were investigated. Twenty subjects assessed the intimacy of surnames sampled from above mentioned 5778 surnames. Average scale values of intimacy for sampled surnames were found to be proportional to logarithms of frequencies. As a result of factor analytic study of intimacy, two factors were extracted. Quantitative relationship among frequencies and those factors was examined.

Key words : active cognition, Japanese surnames, frequency, intimacy scales,
factor analysis

1 ま え が き

人間による外界の認知は、機械認識と異なり、認知が能動的であることが特徴である。テレビカメラやマイクロホン等を通じた機械的な情報の取り込みは、全く無選択的であるのに対し、視覚や聴覚による認知は極めて選択的であり、認知を行なう人間にとって必要のない情報は無視される。この現象は、聴覚の世界では、古くからカクテルパーティ効果として知られ、その機構の解明が試みられてきたが、まだ完全には明らかでない。この効果は、パーティなど、周囲で大勢の人の話し声がする場合でも、誰かが自分の名前を云うと聞こえたり、離れたところで話している人に注意を集中するとその人の話を聞くことができるといった聴覚における選択的認知に関するものである。また、視覚の世界でも、眼の網膜の解像度の特性は均一でなく、中心か (fovia) と呼ばれる部分を中心に視角にして+3度程度のごく狭い範囲だけが解像度特性が高く、それをとりまく周辺部は解像度が低い替わりに動きに対して敏感であるといわれている。したがって、人間は外界の視覚的情報を得ようとする場合には、見ようとする対象に頭の運動と眼球運動を組み合わせ、かつ、レンズとして作用する水晶体の焦点を調整して、中心か上に像を結ぶような、能動的な活動が必要となる。

近年、人工知能の世界で、人間の認知機構を計算機に取り入れ、計算機による外界の認識を柔軟なものにしようとする研究が行なわれている。我々も、こうした研究の一貫として、人間の姓のもつ認

知的特徴に着目し、これを材料として認知の能動性の一貫を解明しようと考えた。姓は個人のラベルとしての特性を持ち、その認知は特定の個人にとって特別な意味を持っている。すなわち、自分の姓は非常に高度な親近性を持ち、カクテルパーティ効果に代表されるように、認知される度も他の人の姓よりはるかに大である。したがって、姓の持つ特性を明らかにし、これと認知特性との関係を調べることにより、人間の認知特性の側面を解明できるものとする。その研究の第一歩として、本報告では、日本人の姓の出現頻度を調べ、同時に、個々の人間にとっての姓の親近度の特性を明らかにし、頻度と親近度の関係付けを行なった。これをベースとして姓の認知特性の研究を進めたい。

2 研究目的

姓の属性として考えられる出現頻度を調査すると共に、個々の被験者にとっての姓の親近度の評価を行ない、親近度を構成する因子を求めて、これと頻度との関係を明らかにする。

3 姓の出現頻度の調査

実験に使用する日本人の姓のデータは全国規模の約3万人の名簿から収集した [1, 2]。丹羽の研究によれば日本人の姓の数は世界中で最も多く、約14万に及ぶ [3] と云う。ただし、出現頻度の高いものから5000で全戸数の92%, 100で37%となる。今回収集したデータは姓の数が5778種あり、姓ベストテンも過去の調査とほぼ同一で、データベースとして妥当なものと考えられる。

出現頻度の高い姓から順にクラスタ分けし、クラスタ毎の出現頻度と含まれる姓の数の関係を図1に示す。最初にリニアスケールで均等配分を試みたが、全体の姓の99%が出現頻度が最も低いクラスタ5に集中したので、対数スケールでクラスタ配分を行い、各クラスタに分散するようにした。

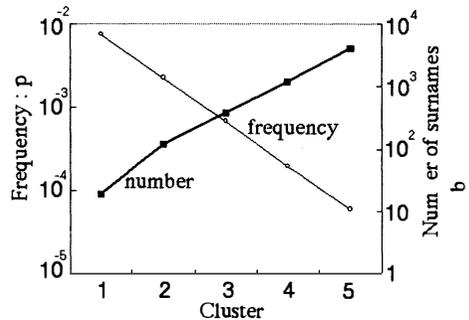


図1：クラスタ毎の出現頻度と姓の数

4 姓の親近度の評価

4.1 評価対象

後述の姓の読みを表す文字列を音節列に変換した刺激の聴取実験に使用する姓を評価対象とした。これはクラスタ1, 3, 5から選んだ78の姓と被験者が所属する学科の教官5名の姓および被験者自身の姓の計84の姓で構成される。

4.2 実験方法

84の姓に対して20名の大学生に以下の5段階の評価尺度による評価を行わせた。

- 5 非常に親近性を感じず。
- 4 やや親近性を感じず。
- 3 どちらでもない。

- 2 あまり親近性を感じない。
- 1 全く親近性を感じない。

なお、事前に提示した姓が被験者にとって好き嫌いで判断するのではなく、親近性すなわち身近なものであるか否かの判断であるとの教示を与えた。

4.3 実験結果

結果を図2に示す。横軸は姓の出現頻度の対数値であり、縦軸は抽出した84個の姓について被験者20名が回答した親近度の平均値である。この図から姓の出現頻度と親近度には正の相関関係があることが分かる。また、図2中に表記するようなクラスタによる3つのかたまりが見られる。図に見るように高頻度のもほど親近度が高いことが分かる。なお、図で、白点で表示したものは被験者が所属する学科の教官の姓で、頻度とは別の要因で親近度が高いと考えられる。図中の直線は教官の姓を除いた姓の親近度の被験者平均値と出現頻度の関係を次式で近似したものである。

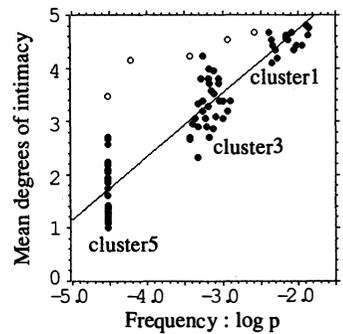


図2：姓の頻度と親近度の関係

$$s = 7.15 + 1.223 \log_{10} p \quad (1)$$

4.4 親近度の因子分析

4.4.1 因子分析モデル

図2で見られた関係を更に詳細に解明するために因子分析 [4] を行った。 s_{ij} を被験者 i の姓 j に対する親近度評点とすると、これは次のような因子構造を持つと考えられる。

$$s_{ij} = a_{i1}x_{1j} + a_{i2}x_{2j} + \dots + a_{ik}x_{kj} + \dots + a_{im}x_{mj} + \varepsilon \quad (2)$$

ここで、

$$\begin{aligned} i &= 1, 2, \dots, n \\ j &= 1, 2, \dots, N \\ k &= 1, 2, \dots, m \end{aligned}$$

で、 i は被験者番号、 j は姓の番号、 k は因子である。 a_{ik} は被験者 i の因子 k に与える重み、すなわち因子負荷量を表し、 x_{kj} は j 番目の姓が因子 k を含む度合い、すなわち因子得点である。 ε は m 個の因子で説明できない独自因子の分散である。このモデルに対応する因子分析は被験者相互間の姓 j に対する親近度評点 s_{ij} の相互相関係数行列を求め、これを因子分析することによって実現される。この際、因子数は相関行列の固有値の大きさが1以上となる基準で決定した。その結果、2因子が抽出された。因子負荷量の推定には主因子分析法を使い、バリマックス法により直交回転を行った。直交回転後の因子負荷量をプロットしたものを図3に示し、直交回転後の因子得点を図4に示す。

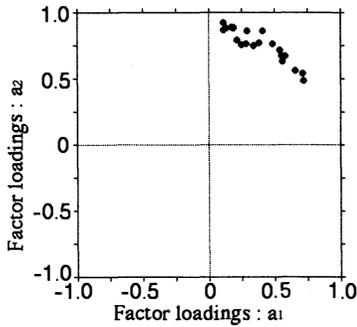


図3：直交回転後の第1—第2因子負荷量の布置

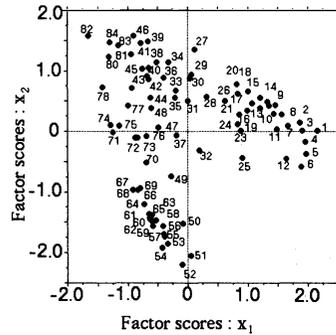


図4：直交回転後の第1—第2因子得点の布置

4.4.2 因子構造の分析

図3からすべての因子負荷量が第一象限にある。因子得点の布置を表す図4において点の番号1が被験者自身の姓であり、横軸が高い正の値を示す各点は佐藤、鈴木、田中など日本人で頻度の高い姓である。一方縦軸の特徴としては負の大きな値を示す姓は [めら](図中の52), [うざぎ](51), [しらそう](54), [せんた](53)等であり、日本人の姓に一般的な漢字表現が簡単に思いつかないものである。ところで、図4の直交回転で明らかに逆V字型の構造が見られるが、因子得点が両因子軸上に重複して値を取っているため因子の意味付けが明確な形で行い得ない。そこで、それぞれの因子軸に対して該当する因子得点が最大となり、その他の因子に対して最小となるような単純構造を得るための斜交回転を行った。

斜交回転した時の因子負荷量を図5に、因子得点を図6に示す。なお、斜交回転における相反軸の回転角は32度である。図5から分かるように因子負荷量が第1因子、第2因子ともほぼ0.5以上であり、個々の被験者が両因子に積極的な重みを与えていると云える [4]。また、図6は図4よりかなり見やすい単純構造になっている。横軸が第1因子得点に対応するが、正の最も高い値を示すプロット1が被験者名に該当し、2が [たなか], 3が [すずき], 7が [さとう], 8が [やまもと] で、それぞれ高頻度の姓である。また、4, 5, 6はいずれも担当研究室の教官の姓で、後述するように今回目標刺激に選んだものである。そして、値が小さくなるにつれて、低頻度の姓、または被験者に馴染みの薄い姓が現れている。従って、この第1因子は身辺性に関する因子と解釈する。

一方、縦軸は第2因子を表すが、原点から離れた、負の値が最も大きい所にかたまりが見られる。図6中のプロット52~65の点の集団である。これらの姓は上記のように [めら], [しらそう], [せんた], [しのみ], [ねだ] 等で代表される音を聴いてすぐに日本語の漢字を当てはめ難い姓である。漢字は表意文字であるから、漢字に変換し難いという事は意味のある組み合わせが作り難いことである。その一方、原点に近い点にある姓は84 [つかだ], 83 [ひらさわ], 82 [みずはら], 80 [おのでら] 等で、それぞれ [塚田], [平沢], [水原], [小野寺] と漢字にすぐ変換することができる。従って、第2因子は意味を表す尺度と考え、ここでは意味性の因子と解釈した。以下の記述では第1因子を身辺性と呼び、第2因子を意味性と呼ぶことにする。なお、斜交回転は直交回転とは異なり、因子間が完全独立ではなく、相互の角度32度の余弦、すなわち0.85の相関があることになる。図6の場合、身辺性の高い姓には漢字に変換しやすい意味な姓が多く見られること、また、漢字に変換し難い姓は馴染みが薄い傾向があることから、身辺性と意味性が完全独立でないことが分かる。

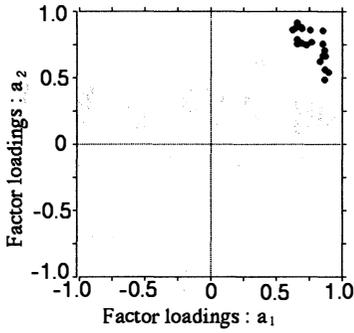


図5：斜交回転後の第1－第2因子負荷量の布置

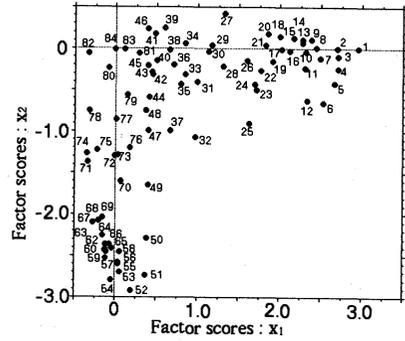


図6：斜交回転後の第1－第2因子得点の布置

5 考 察

身辺性尺度（第1因子得点）と頻度の関係を図7に、意味性尺度（第2因子得点）と頻度の関係を図8に示す。頻度と第1因子得点との間の関係を概略的に把握するために図7の曲線に示すように2次の多項式近似を行った。第1因子得点を x_1 で表し、頻度 p の対数を ρ で表すと、 x_1 と ρ の関係は次式で近似することが出来る。

$$x_1 = 6.85 + 2.773\rho + 0.277\rho^2 \quad (3)$$

$$\rho = \log_{10} p \quad (4)$$

一方、図8では多項式近似曲線は上に凸の曲線となり、低頻度の姓ほど ρ に対する第2因子得点 x_2 の変化が大きく、高頻度の姓の第2因子得点の増加は飽和する傾向を示している。このことは中頻度以上の姓は概して意味性が大であること、低頻度の姓は分散が大きいが、意味性が低い姓が多いことを意味する。図8における第2因子得点 x_2 と頻度の対数 ρ の関係は次式で近似される。

$$x_2 = -2.498 - 2.157\rho - 0.454\rho^2 \quad (5)$$

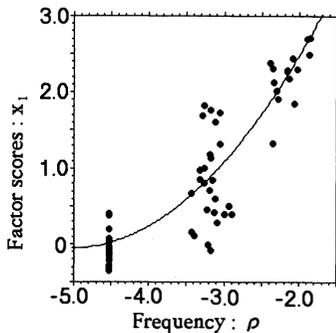


図7：頻度と第1因子得点の関係

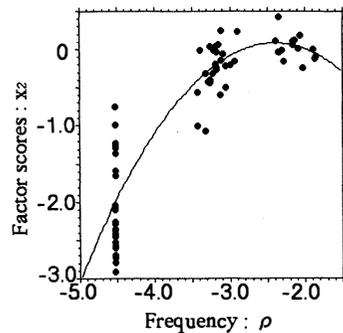


図8：頻度と第2因子得点の関係

6 む す び

1. 全国規模の約3万人の名簿から日本人の姓の出現頻度を調査した。その結果、5778の姓が得られた。これは日本の全戸数の90%以上を網羅するものと考えられる。姓のベストテンも、過去のさらに大規模な調査とほぼ同一で、データベースとして妥当なものと考えられる。
2. 出現頻度が大、中、小の姓のクラスタから選んだ78の姓と、被験者が所属する学科の教官5名の姓および被験者自身の姓の計84の姓について、20名の学生に親近度の5段階評価を行なわせた。その結果、頻度の対数と親近度の被験者の平均値の間には正の相関関係が見られた。
3. 親近度の被験者間相互相関係数行列を因子分析した結果、2因子が得られた。因子軸上での姓の分布から、これらは身辺性の因子と意味性の因子と解釈された。
4. 姓の出現頻度の対数とこれらの因子の因子得点との間には、それぞれ2次の多項式で近似される関係が見られた。

今後は、姓の出現頻度および親近度の因子と姓の認知率との関係を求め、認知の能動性に、これらの要因がどのように関与するかを求める研究を進めたい。

参 考 文 献

- [1] 情報処理学会編，“情報処理学会会員名簿”，
- [2] 北林行雄，黒田靖子，中山剛，“名前埋め込み文字列の聴覚認知実験”，平8北陸連大，B-30，1996。
- [3] 丹羽基二，“日本姓氏大辞典”，角川書店，1985。
- [4] A. L. コムリー，“因子分析入門”，サイエンス社，1979。