

温故知新：戦後間もない時期の社会調査再訪

前田 忠彦 データ科学研究系 准教授

1 はじめに：「日本人の読み書き能力調査」について

1948年に行われた「日本人の読み書き能力調査」（以下読み書き能力調査）について、その報告書（読み書き能力調査委員会、1951：以下「報告書」と略称）、を参照しながら、歴史的意義を考え、現代への教訓を考えている。その動機は、現代における「読み書き能力調査」の開発を目的とする科研費基盤研究A（課題番号19H00627「基礎教育を保障する社会の基盤となる日本語リテラシー調査の開発に向けた学際的研究」、研究代表野山広（国立国語研究所））に基づくプロジェクトに参画したことである。外国籍者の受入などが進むことが予想される今後の日本社会で、基礎的な日本語の読み書き能力調査を、正確に測定するための基盤を開発することが急務となっていることによる。このプロジェクトを端緒として、国立国語研究所（以下国語研）の研究者との共同研究を進めている（3、4節）。

2 「日本人の読み書き能力調査」の歴史的意義

読み書き能力調査は、台帳に基づく無作為抽出、標本設計は層化多段無作為抽出により行われた大規模な(n=16,820)学術的社会調査としては、戦後最も早期のものである。調査には後の所長林知己夫を初めとする何人もの統計数理研究所（以下統数研）のメンバーが参画した。この調査の歴史的意義としていくつものことがらを挙げる事ができよう。

- 社会調査史上では、戦後に導入された統計的な立場からの社会調査、特に確率的抽出法の標準的な設計（層化多段無作為抽出）の先駆けとも言える内容となったこと。統数研の研究者らもこの研究の一つのきっかけとして、標本調査法に関する研究を進展させたこと
- 統計解析の面では、この調査データが林の「数量化理論Ⅰ類」（実質的にはダミー変数を説明変数とした重回帰分析）の導入の素材を与えたこと
- 研究をめぐる環境という点では、統数研の研究者と国語研（上記調査実施後まもなく設立された）の研究者の間の共同調査・研究の先駆けとなったこと（この後、例えば「鶴岡市における言語調査」（1951年に第一次調査）のような社会言語学的調査が、両研究所の協力下で行われた）
- 社会的インパクトの面では、同調査の結果、日本人における読み書きの能力の高さが示され、当時連合軍側特に米国第一次教育使節団報告（1946年）の提言に端緒として検討されていた国語改革（日本語表記のローマ字化）が取りやめとなる資料を与えたこと。またその後しばしば、日本における非識字率の低さを示す根拠として引用される数値を提供したこと（この辺りの歴史的経緯については、より丁寧な考察が必要である）
- 統数研の社会調査グループの歴史としては、この調査や国語研との言語調査の経験を経て、1953年の「第1次 日本人の国民性調査」に結実したこと

さて、この先は、国立国語研究所の研究者たちとの共同研究の一部を紹介する。

3 ゲッシングを考慮した非識字者率の推定（前田・横山、2020）

読み書き能力調査はリテラシーを測定するための90問の出題に対する正答数が得点化されたものが主な測定内容となっており、この得点に性や年齢、教育歴、居住地（都市部か郡部か）といった社会的・属性的要因がどのように影響するかを調べている。

報告書内では、上記得点が0であった者、すなわち一問も正答できなかった者を完全非識字者（報告書内の用語が、現代的には差別的であるので、変更している）と定義し、この割合が1.7%と報告している（表1）。（恐らく）この数字が、国際的にも日本における非識字率の低さの根拠として、長らく利用・引用されてきたようである。逆の満点は4.4%である。

しかし、正答数0を非識字者とみなす立場は、全90問の出題のうち、4肢択一問題が19問、5肢択一問題が46問、のように「多肢選択」型の質問が65問含まれており、多肢択一問題では、「当て推量」（ゲッシング）による正答の可能性が否定できない点を無視しているとも言え、その意味では非識字者率を過小評価している可能性がある。

表1 報告書にみる読み書き能力調査の正答数得点の分布

code	19	18	17	16	15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	00
点	90	89	84	79	74	69	64	59	54	49	44	39	34	29	24	19	14	09	04	00
数	90	85	80	75	70	65	60	55	50	45	40	35	30	25	20	15	10	05	01	00
割合	3.72	3.17	3.24	2.99	2.75	2.59	2.43	2.27	2.11	1.95	1.79	1.63	1.47	1.31	1.15	0.99	0.83	0.67	0.51	0.35
累積	3.72	6.89	10.13	13.37	16.61	19.85	23.09	26.33	29.57	32.81	36.05	39.29	42.53	45.77	49.01	52.25	55.49	58.73	61.97	65.21
非識	4.4	26.7	48.0	69.3	90.6	111.9	133.2	154.5	175.8	197.1	218.4	239.7	261.0	282.3	303.6	324.9	346.2	367.5	388.8	410.1
率	4.9	29.8	53.2	76.6	99.9	123.3	146.7	170.1	193.5	216.9	240.3	263.7	287.1	310.5	333.9	357.3	380.7	404.1	427.5	450.9
合計	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820	16820

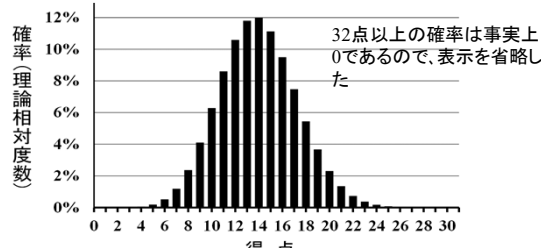


図1 複合二項分布によるゲッシングによる得点の理論分布

各問のゲッシングによる正答確率は 1/(選択肢の数)であり、（そのような調査参加者が実際にいたかどうかはともかくとして）完全な非識字者が多肢選択型の65問全てにゲッシング回答をするという「最も極端な想定」下での得点分布は、複合2項分布により評価できよう。実際にこの設定下で、得点分布を求めたものが図1である。ゲッシングでも平均的には14点弱の得点をとる事ができる。累積確率が95%を超えるのは20点であり、統計的検定になぞらえると、確実に非識字者（によるゲッシング）でないと言える得点は20点以上、逆に19点以下は非識字者の疑いを拭えないことになる。報告書によると19点以下の割合は5.5%弱である。

4 非識字者率に生年が与える影響（横山・前田他、2021、一部改変）

上記のようにゲッシングを考慮した“非識字者”がどの程度いたのかを報告書の集計表からある程度再現することができる。正答数得点19点以下を非識字者の可能性がある者と判断した場合の非識字者率を、市部居住者と郡部居住者別に、生年毎に検討したものが図2である。集計表に基づいて非識字者率を生年と市郡別を説明変数としたロジスティック回帰分析によれば、どちらの説明変数の効果も有意となる。

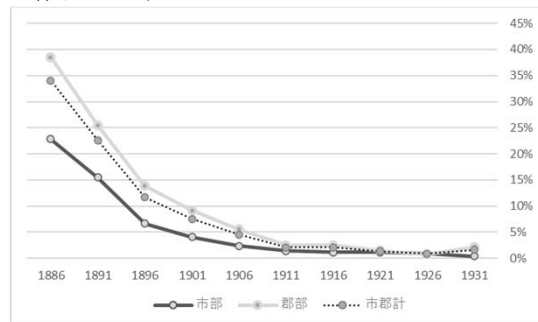


図2. 市部・郡部別の非識字者率の各生年における割合

「読み書き能力調査」の設計から実施にかけては、社会調査の礎となる様々な工夫が詰め込まれ、現代でも学ぶところが多い。

文献

- [1] 前田忠彦・横山詔一 (2020) 複合二項分布を利用したゲッシングによる正答分布の社会調査への示唆 第69回数理社会学会
- [2] 横山詔一・前田忠彦・高田智和・相澤正夫・野山 広・福永由佳・朝日祥之 (2021) 日本人の読み書き能力1948年調査の非識字者率における生年の影響, 日本語学会2021年度春季大会
- [3] 読み書き能力調査委員会 (1951) 日本人の読み書き能力 東京大学出版部