

……いんふおるむ(第44回)……

質問の仕方の違いによる回答への影響

—— カールトンとシューマンのレビュー(1) ——

Graham Kalton and Howard Schuman (1982): "The Effect of the Question on Survey Responses: A Review", Journal of the Royal Statistical Society, A(1982), 145. Part . PP42 - 73 より

翻案 鈴木 達三
(帝京平成大学教授)

一般に調査を企画・実施する場合、調査対象者の選定(標本企画)および質問項目の作成(質問↔回答形式の選択)調査票の構成が、基礎になる。

これまで、調査回答者の代表性については、ランダム・サンプルによるとして、我が国の世論調査ではランダム・サンプルに基づく統計的な調査方法が一般的であったが、最近では、調査環境その他の影響で、今までの方式ではうまく機能していないように見え、代表性の確保の点から問題のある?調査方式も出ているとして調査関係者の関心が高い。

一方、もう一つの基本要素である質問文の作り方や調査票の構成の方は、必要な情報の内容や質(正確さ、信頼性等)に関連しているから、これまでもいろいろ研究されている。しかし、問題の性質上、まだ研究が十分に進んでいるとはいえないように見える。この現状を考えるため、ここでは、KaltonとSchumanがこれまでのアメリカ、イギリスの研究の概要をまとめたものを紹介し参考としたい。

Graham Kalton and Howard Schuman (1982): "The Effect of the Question

on Survey Responses: A Review", [Journal of the Royal Statistical Society, A(1982), 145. Part . PP42 - 73]

1 はじめに

調査によって得られる回答は、質問文のいまわし(ワーディング)、質問↔回答形式、質問文の位置に影響される。この影響を検討するには、目的とする情報のタイプによって質問項目を分類する。よく利用される区分は事実・実態項目か意見項目かの区別であるがこの二つにすっきり区分されるわけではなく、回答者の知識をきく質問や理由をきく場合、仮想的なことをきく場合、この他評価を伴う事実・実態項目の質問(これは事実・実態的情報を得る目的であるが調査の仕方は回答者の評価(判断)にたよるので客観的というより回答者の主観によるもので意見項目に近い)などがある。ここでは、実態項目とそれ以外の項目(実態に関する評価を含む)の二分類で考える。また、実態項目とは他の資料から真偽の判定が可能なもので、意見項目はそれ以外としておく。

実態項目に関する妥当性の研究は数多く実

施されているが、それぞれの方法で、どのくらいの正確性を達成したかというレベルの問題にとどまっただけで、方法を変えたときどうなるか？という点、すなわち、回答の正確さに影響する要因の分析まではやられていない。

意見項目では、妥当性の確認は実態項目より困難で、回答の正確さは、理論上予期される他の項目との関係を満たすかどうかという回答構造の妥当性のみをチェックする。現在のところ、理論に合わないデータは、理論がよくないのか、あるいは測定 of 道具（質問文）の方に問題があるのか不明である。また、理論に合っているにもかかわらず、回答は確かかどうかわからない。したがって、意見項目（実態以外の項目）について、回答に影響する要因の研究は主に折半法（split - ballot）で研究されている。これは妥当性より内的一貫性に関することで、どれが一番よい質問形式かを定めることはできないが、折半法で、もし両者の回答が大きく異なれば、回答は質問形式に影響されるとして留意し、両者の回答の違いがそれ程なければ影響されないことになる。

「実態項目」と「それ以外の項目」の間の違いの二つ目は回答に影響する要因が異なると考えて研究される点である。実態項目では特に“定義に関すること”、“理解度に関すること”、“記憶に関すること”、および“回答が社会的に望ましい方向に偏る問題”などに重点があり、一方「実態以外の項目」ではいろいろな質問形式の影響、たとえば

質問文、回答肢のバランスの問題

中間項を提示するかどうか

回答肢を提示する順序

等に重点がおかれる。

まず実態項目についての研究を総括し、次に「実態以外の項目」についての研究をとり上げる。質問の順序や文脈・前後の関係につ

いての影響は両タイプの項目で考えられるので、後であらためてとり上げる。

また、質問項目の影響については主として面接調査の回答を考える。すなわち調査員が回答者に面接し、質問して、回答者の回答を調査員が記録する調査方式の回答を考える。

2. 実態項目の質問の影響

1) 定義の問題：実態に関する質問項目を構成するには、まず、収集すべき事実について正確な定義をする。ごく些細な（とみられる）定義の変更で重大な影響がみられる。この例として 失業・労働力の定義がよくあげられる。

この他、簡単にみえる世帯（家）の部屋数のような項目も 定義されたことを回答者に伝える過程で問題が生じる。イギリスの国勢調査の質の研究（1966年）で6人に1人は部屋数で不正確な回答をしていることが分った。これは、「国勢調査のこまかい定義」を無視して「回答者の考える定義」によって回答していることが原因であるとしている（Wales, Gray, and Gee (1966)）。

2) 理解の問題：「調査上の定義を回答者に伝える」伝達過程の他の面は、回答者が質問されたことは何であるか、何が適切な回答であるかということを完全に理解していることを確かめることである。回答者は質問に含まれる概念や枠組み（の範囲）を一応の水準で理解する必要がある。（チャンネルとカーン（1968））。しかし、簡単な質問でも、たとえば、“夕方のテレビでニュースをみる割合は？”と質問されて、この意味のわかる人はほとんどいなかった。また、“平日の夕方、番組の合間”のCMをいつもみるか？の質問で

の“平日”の意味もはっきりしなかった (Speak (1967), Belson (1968))。

3) 記憶誤差の問題：実態項目の質問に正しく回答するには回答者は回答に必要な情報を入手する必要がある。

質問が過去のことであれば、回答者は記憶から取り出しうることが必要である。

記憶をよみがえらせるには、呼び戻す時間の長さ、呼び戻す記憶が回答者にとって特別にはっきりしているかどうかに関連する。

よく知られているのは、実際より最近のこととして記憶している (テレスコーピング・エラー) 誤差 (Sudman と Bradburn (1973, 74))。

記憶を呼び戻せない (リコール・ロス) とテレスコーピングは逆の作用をする。すなわち、記憶を呼び戻せない誤差は過小報告になり、実際より最近のこととして記憶している誤差は過大報告となるからである。この二つの誤差の大きさは情報に必要な期間の長さによる、期間が長ければ、リコール・ロスは小、テレスコーピングは大、期間が短かければ、逆にテレスコーピングは大、リコール・ロスは小になる。このテレスコーピング・エラーの解消をはかる方法として「くり返し調査」の方法がある。また、記憶誤差をなくすために、記録による、一定方式で助けを出す、日記方式にする、等が利用されるが一長一短である。

4) 回答が社会的に望ましい方向に偏る誤差：回答者は自分の (本来の) 回答が社会的にみてより好ましいと考える回答によって妨げられる傾向がある。たとえば、“投票したか？”で“投票した”の回答率は実際よりいつも高い (Parry, Crossley

(1950), Traugott と Katosh (1979))。

もし、ある事柄が、“ためらったり”、あまり正直に回答したくないようなとき、回答者は回答するのに圧迫感をもつかもしれないし、本来の回答よりも、より社会に受け入れられるようにして回答することになる。このような敏感な情報 (回答) を引き出すために、よく知られた方法は、より個人的にする。...たとえば収入をきく質問では、回答カードを利用して、該当する収入額のコード番号を回答してもらおう (秘密投票法) 等である (Barton (1958)) はこれらをまとめている。

これに関連して、最近、ランダムイズド・レスポンス法が開発された。この方法は、回答者は回答する質問項目 (本来知りたい敏感な (センシティブな) 質問とそうでない普通の質問との二つ) からランダムに質問を選択し、調査員にはどちらの質問に回答したのかわからないようになっている。これで回答者のプライバシーは守られ、より一層真実の回答が得られると期待される。ワーナー (Warner (1965)) が開発してから、いろいろ理論面の研究はあるが、実際面では少ない。実際面で共通に得られたことは、この方法が回答者によく理解され受け入れられていることである。やり方で困惑するのは20人に1人ほどで、8人に7人はランダム化器具 (調査員に分からないようにカードをめくってどちらの質問に回答するかを決めるカードなどのこと) の使用をよく理解している (Locander et al. (1976))。ランダム化回答法により、従来の方法より回答率が改善されている。ランダム化回答法が最も社会的にみて望ましくない行動の過小報告を減らすのに有効 (たと

えば、破産宣告を受けた、飲酒運転をして罰金をとられた等)である(Locander et. al.).

5) 記憶誤差や理解を改善する方策:

ミシガン大学のキャネル(Cannell)らは記憶誤差問題やセンシティブな質問に対する新しい方法を開発してきた。研究は主に健康問題について回答の質を改善することに向けられ、初期には回答者の理解を高め、回答者が何をすべきかを適切に指示する必要から、それに合う方法を考えた。

その一つは 質問文の長さ を長くする方法である。面接調査員の面接行動が長い質問では回答は長く、短い質問では短くなることから、これまでの調査常識からみて問題ではあるが、長い質問がより妥当なこともあるのではないかと考えて実験した。この長い質問は本質的には質問を二度くり返すようなやり方であるが、結果は長い質問の回答は特別長くはならず多くの健康問題を引き出すことができた。

調査票が長い質問ばかりで構成されていると回答者は退屈するので、長い質問と短い質問を混ぜたものを実施したところ、長い質問も短い質問も健康問題の回答をより多く引き出せた(Cannell et. al (1977, 1981))。

このことから長い質問をすると、回答者の理解を深め、回答者にその質問は重要だと思わせ、回答者の関心をそそるように働き、その結果、より妥当な回答が得られるということだろうと考えた。通常よくいわれる“質問は短く”というのは不正確な表現で“質問は単純に”というべきである。実際長い質問による困難

さはその長さにあるのではなく、その複雑さによるといえる。

またキャネル(Cannell)らは、回答内容を改善する目的で、

回答者に対する指示を質問票に組入れる。

回答者の回答の仕方がよかったかどうかを、回答者にわかるように伝える。

正しく回答すると公約しサインさせる。等の実験をした。こうすることによって1)回答者は自分のやるべきことは何かがわかり、2)回答者の回答行動がよくなったかどうか回答者に伝わり、3)回答者が良心的に回答することに同意するようにすすめられて納得すれば、その同意したことにこだわりを感じて正しく回答するように努めると期待される。

実験では回答者の5%ほどが(サイン)協力を拒否したが、電話調査では口約束であるが何も問題なかった。

これらの方法の有効性を検討する実験で、どの方法も回答内容を改善するようにみえ、三つの方法すべてを利用すると結果は最も良かった。この方法をとると高学歴層に反発されるのではと懸念されたがそれもなかった。

表 1

項目	実験群	対照群
前日のテレビ視聴	86%	66%
昨日の平均ラジオ視聴時間	2.5時間	1.5時間
昨日の社説(編)を読む率	38%	55%
3カ月間の読書(本)平均冊数	2.9冊	5.3冊

テレビ・ラジオの視聴に関する実験調査で、この三つの方法すべてを用いた群とそうでない対照群と比較すると、実験群では通常過小報告されがちな行動ではよ

り多く、通常過大報告されがちな事柄では、より少なくなった（表1参照）。

これらの実験結果は、この方法が回答内容を改善するのに有効である可能性があることを示している。

しかし、まだ未熟な段階であり、通常の調査に一般的に利用できるというわけにはいかない。質問票の抜本的改変や調査員の訓練等の問題がある。

3. 実態項目以外の質問項目の質問の影響

実態以外の項目でも実態項目同様、まず、測定すべきものの概念を構成するのが第一段階である。しかし実態項目以外の項目は実態項目に比べ、測定すべきものの概念化はより抽象的で、正確に定義することは困難であるし、操作化するのも困難である。

ただ一つの項目で、ものの本質をあらわすことは不可能で、いくつかの指標が必要になり、これらは相互に重なり合っているため、それらの組み合わせで測定したいものに接近できるようにする。

態度測定においては、態度の次元の概念化と操作化は密接に関連している。概念化の段階で、その次元の操作化に用いる個別の質問項目の選択がきまる。そして、その項目自体によって、その次元の定義がはっきりしてくる。結局、態度の側面は、それを測定するために利用している質問項目の組に依存して定義されることになる。

測定する項目の組のささいな変更でも結果が大きく影響を受けることがあるのは実態項目の場合と同様である。

このように、比較的軽い変更や、研究する概念の定義がどちらかといえばあいまいなことに対しても、実態以外の項目の回答は敏感に影響されることから、経験のある分析家は

このような質問に対する回答分布をあまり真剣に考えすぎないように用心している。そのかわりに何らかの関連性分析に注意を集中している。たとえば、各層グループ別の回答分布の差を比較する等のことをする。このような分析法はたとえ、異なった質問形式が回答結果に大きな差を示しても質問形式による影響は各層グループ間の対比によって打消されると仮定して正当化されている。シューマンとプレッサー（Schuman, Presser (1977)）はこの仮説を“形式に抵抗する関連”（の仮説）“form resistant correlation”と名付けた。後に示すようにこの仮定は多くの場合よくあうが、いつも必ずよいとはいえない。

実態以外の質問項目を作成するとき、質問項目の企画者は質問の形式について、数多くの選択決定をしなければならない（このときどうするのがよいか？）。

これらの選択決定について、主に意見項目に関連することを対象にして、とった方策がどのような影響を与えるかということを簡潔にまとめておく。

a) D K (don't know) の取り扱い

事実・実態項目の回答で“D K.”は必要な情報が得られなかったことを示す。質問には一つの回答があるのに回答者はそれを示せなかった。しかし、実態以外の質問では、“D K.”は異なった解釈ができる。すなわち、回答者はその質問に本当に意見がないということ（Schuman, Presser (1977)）。

通常、意見（の質問）項目のD Kを認めるかどうかの取り扱いは、事実・実態項目の場合と同様である。D K回答は回答者には示されていない、調査員は回答者が自分から“分からない”と申し出たときだけ、このD Kカテゴリを用いるように指示されて

いる。

このため、本来はDKの人が、“何らかの(特定)の回答をしなければ”と負担感をもつような危険がある。これは実態項目でも同様であるが、実態以外の項目の方が大きいと考えられる。

シューマンとプレッサー(1980)は回答者は“質問内容に含まれている事柄”は知らないにもかかわらず、ある意見項目の回答肢の一つを選択した例示を二つあげている。この両方とも回答者は、ほとんど誰も知らないような法律施行に関する見解を質問されて、30%の人が意見を述べた(例を示した)。ビショップ(Bishop et al. (1980))からも全くの作り事に関する意見について同様の報告をしている。

意見のない人を選別するために、いろいろの方法(フィルター質問等)が用いられる。

その一つは、回答肢の一つとして“意見なし”の選択あるいはフィルターを回答者に示す——回答選択肢の中の一つとして、はっきり“意見なし”を回答者に示す(これは擬似フィルター)——ことである。

シューマンとプレッサーの実験では、この項を入れると、前の実験で意見を述べる率は10%以下に低くなった。

より強い方法は、質問の前に

“.....についてのご意見をお持ちですか?”と前置きの質問をする方法である(完全フィルター)。

シューマンとプレッサー(1978)はこのフィルター方式の影響について実験し、完全フィルター方式のとき、通常の方法にくらべDKの比率は20~25%ふえることを示した。ビショップらも同様の結果を得た。そして、一般に、皆がよく知っているような事柄についての質問では、DKの増加は

少なく、あまり知られていない事柄のときは(増加は)多くなった。

シューマンとプレッサーの実験で、質問形式のちがいによる本質的な結果への影響はどちらかといえば予想外のことだった。

通常分析でするように、DK(回答者)を除いて回答分布を求めると、多くの場合比例配分になり、形式に影響されていなかった。また、意見調査の回答と属性項目との関係も影響されなかったが、意見項目の回答相互の関連は質問の方式により大きな差のある場合もみられた。

b) 自由回答形式か回答選択(肢)形式か

回答者が回答するとき、“いくつかの回答選択肢から選択するためのリストを用意するか”、“回答者に自由に回答をさせるか”二種類ある。

回答選択形式の長所は、回答を標準化しており、取り扱いが便利。

短所(自由回答の長所)は、

回答選択肢が限定されていて、必ずしも回答者の望む本来の回答と一致しないこと

示された回答選択肢に回答者が影響され、正しい回答が得られないこと

があげられる。

自由回答がよく利用される場合は、主に回答が本質的に質的なものであり

回答が各人によってまちまちで多岐にわたるようなとき

である。

これらの状況は、動機をきく質問の場合、主な(あるいはすべての)理由をきくとき、最も(いくつかの)大事な要因をきくときなどに生じている。このような場合、質問項目を作成するとき自由回答法にするか、選

択肢法にするか検討する必要にせまられる。

質問形式の研究の一環として、シューマンとプレッサー（1979）は既存の質問項目を利用し、自由回答形式と選択肢形式を比較する実験をした。両方の回答分布は大きな差があった。

とり上げた質問は、仕事の価値についてで自由回答形式：“次の質問は仕事に関することです、人びとは仕事について異なることを望みますが、あなたが、仕事について最も好ましい（と思う）ことは何ですか？”

回答選択形式：“次の質問は仕事に関することです。このカードを見て下さい。あなたが仕事について最も好ましいと思うことは、このカードの中のどれですか、おっしゃって下さい？”

カードの項目	回答分布
1. 高収入.....	(12.4%)
2. 失業する危険がない.....	(7.2%)
3. 働く時間は短く、自由時間は一杯ある.....	(3.0%)
4. 昇進の機会がある.....	(17.2%)
5. 仕事が重要で、やりとげたと感じるの持てること.....	(59.1%)
どれも選択しない.....	(1.0%)

選択肢法では“どれも選択しない”は1%であったが、自由回答形式では60%近くの人が、この5つのカテゴリ以外の回答をした。余分に作ったコードの主なものは
楽しいあるいは楽しめる仕事 (15.4%)
仕事をする環境(状態)..... (14.9%)
満足(仕事が好き)..... (17.0%)
である。5つのカテゴリのうち、自由回答法と選択肢法でよく似ていたのは、はじめの

2つ(1と2)で、3番目は自由回答の%がくやや低、後の2つ(4と5)は自由回答の%は非常に低かった。

このような大きな差異を説明する一つは「形式の変更によるのではなく、用意された回答選択肢が不適切だった」というものである。これにしたいが、シューマンとプレッサーは自由回答形式の結果から、回答選択肢を作り直して実験した。

しかし、この改良版でも回答選択形式では自由回答の58%しかカバーできなかった。

その上、両方の回答結果で共通の5つの選択肢の%の差も残った。この回答分布の差の他に、両形式は属性項目との関連に影響することがわかった。検討の結果、自由回答形式と、適切に構成された回答選択形式の質問で、異なる回答結果が得られた場合、回答選択形式の回答の方が回答者を分類するとき、また、その質問の回答と他の変数との関係を記述する場合に、より妥当性がある場合があることを示唆する間接的な証拠を示した。

c) 対立する意見の有無

回答者に、ある事柄についての意見をきくとき、質問作成者は質問の仕方として、対立する意見の示し方に選択の余地がある。一方の極は、対立する意見を全く示さないで単に“Xに賛成ですか？”(A)とする。もう一方の極は、実質的な対立選択肢をはっきりと示す。たとえば、

“(XあるいはY)に賛成ですか？”(B)である。この中間的なものは“賛成”と“反対”の対立項できく形式で

“Xに賛成ですか、反対ですか”(C)の形である。

折半法による多くの実験により、一般的

にいて(A)と(C)の比較でほとんど差はない((A) (C))。しかし、実質的な対立項が示されるか示されないかでは大きな差がみられた(いつもではないが)。

多くの実験により、対立するものを含めることは新しい項目の持込みになることが示される(Hedges(1979))。

そうであれば、二つの質問は見かけ上よく似ていて比較できるようにみえるが、結果は大きくちらばるので注意が必要になる。

d) 回答者が質問の内容によらず“賛成”する傾向

多くの態度測定法では質問の組について“賛成”か“賛成しない”かを質問する。この方法の問題点として、質問内容によらず“賛成”と回答する傾向があること...、これは心理学の方で“賛成回答の偏り”といわれ注目されてきた。心理テストではこの影響は重要性をもたないというが、社会調査では、問題であることが示され、シューマンとプレッサー(1981)は、これについて実験をした。実験では、“賛成する偏り”のあることを分析した。詳しくはここでは集約できないが、その大きさを例示すると、
A：この国で、犯罪や法律を守らないことは、社会環境のせいというより個人の責任である
B：この国で、犯罪や法律を守らないことは、個人のせいというより社会環境のせいである

この二つの項目に“賛成”か“賛成しない”かをきくと、項目Aに“賛成”ならば、項目Bには“賛成しない”となる筈であるのに、実際はAの“賛成”：59.6%、Bの“賛成でない”：43.2%となった。

e) 中間的項目

回答者は、ある意見に対する見解をきかれたとき、中間的意見あるいは中立の回答を選択したくなることがある。このとき

-) 中間項をはっきり回答者に示すか
-) 回答者が中間的回答をしたときだけ記録するか
-) 中間項をなるべく回答させないようにするか

等がある。

回答者に中間項をはっきり示すと中(間・立)的な回答がふえる(Kalton et .al(1980)の%の増加は15~49%、PresserとSchuman(1980)では10~20%)。

プレッサーとシューマンはそれ以前の“pro”(賛)、“neutral”(中立、中間)と“anti”(反)の3項目の実験も含め、中間項を示すと中間項の回答はふえるが、これは両極の回答から比例的にくるので、“pro”(賛)と“anti”(反)のバランスは変化しない(中間項を提示するかどうかに影響されない)。しかし、カールトンらの実験ではいくつかの場合こうならなかった。他の変数との関連についてはほとんど何もわかっていない。このような質問形式の変更による実質的な影響についてみると、無批判に“form resistant correlation”の仮説を信賴しているのは危険であろうと思う。

f) 回答選択肢の提示の順

回答選択形式の質問項目は選択肢の提示の順序によって影響される。この提示の順について、二つの方式を分けて考える。

- A：自記式、あるいは面接のとき回答カードを利用(印刷した形式の場合)
- B：会話により選択肢を伝える。面接のときつぎつぎに読み上げる(面接および電話調査)

結果として、

A : 印刷形式のカードを提示したとき、最初の選択肢が多少多く好まれる傾向がみられた (Belson (1966)、Quinn and Belson (1969))。

B : 言葉により伝えるとき、ラグとキャントリル (RuggとCantril (1944)) は最後に伝えた選択肢が好まれる例を示したが、ペイン (Payne (1951)) はいくつかの例で順序効果は無視できることを示している。

カールトン (kalton et al. (1978)) は言葉で伝える順序をいろいろ変えて、四つの簡単な質問で実験している。

このすべての実験で、最初に示した回答選択肢が好まれるという結果であった。影響ということからみると極めて小 (差は2%程度) で、サンプリング誤差程度のものであった。

これまでは“実態項目”の質問に関連する調査法の問題点と“実態以外の項目”の質問に関連する問題点とを分けて述べてきた。これは、この二つのタイプの質問のワーディングや形式についての研究上の差をうまく反映させるため便宜的に仕分けした区分である。

したがって、一方でのべた問題点が他方では全く無いということではない。

これらについての事例および質問の順などの影響については、またあらためて報告する。

Belson, W. A. (1966). The effects of reversing the presentation order of verbal rating scales. *Journal of Advertising Research*, 6, 30 - 37.

Bradburn, N. M., Sudman, S., & Associates. (1979). *Improving interview method*

and questionnaire design: Response effects to threatening questions in survey research. San Francisco: Jossey-Bass.

Cantril, H. (1944). *Gauging public opinion*. Princeton, NJ: Princeton University Press.

Kahn, R. L., & Cannell, C. F. (1957). *The dynamics of interviewing*. New York: Wiley.

Kalton, G., Collins, M., & Brook, L. (1978). Experiments in wording opinion questions. *Journal of the Royal Statistical Society (Series C)* 27, 149-161.

Locander, W., Sudman, S., Bradburn, N. (1976). An investigation of interview method, threat, and response distortion. *Journal of the American Statistical Association*, 71, 269-275

Payne, S. L. (1951). *The art of asking questions*: Princeton: Princeton University Press.

Quinn, S. B., & Belson W. A. (1969). The effects of reversing the order of presentation of verbal rating scales in survey interviews. London: Survey Research Centre.

Rugg, D., & Cantril, H. (1944). The wording of questions. In H. Cantril, *Gauging public opinion* (pp. 23-50). Princeton, NJ., Princeton University Press.

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments in question form, wording, and context*. New York: Academic Press.

Sudman, S., Bradburn, N. M. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine.