

# キャリアオーバー効果をめぐる二つのジレンマ

山口 洋

## 〔抄録〕

本稿は、キャリアオーバー効果に関する二つのジレンマを指摘し、近年の海外での実証研究に基づいて解決の糸口を論じた。第一に、キャリアオーバー効果が意識項目間の相関に及ぼす影響は大きな問題（自己発生的妥当性）になりうるが、その観察は原理的に困難である。よってこの問題の核心に触れると思弁的な議論に陥り、観察が容易な効果に議論を絞ると重大な問題が無視される。過去の議論の混乱は一部このジレンマに起因する。第二に、キャリアオーバー効果を弱めるには関連項目間に緩衝項目を挿入せよとされるが、そうすると調査票がまとまりを欠く。このジレンマは「関連項目をまとめると必ず効果が強まる」「その効果は何であれ避けるべき」と仮定される限り抜け出せない。しかし、これらの仮定は近年の実証研究の知見によれば必ずしも真実ではない。最後に以上をまとめて今後の検討課題が示された。

**キーワード** 社会調査法、キャリアオーバー効果、自己発生的妥当性、緩衝項目

## 1. キャリアオーバー効果をめぐる二つのジレンマ

キャリアオーバー効果<sup>(1)</sup>(carry-over effect)は調査論で古くから知られ、今では社会調査の邦文概説書の多くが、その説明に紙幅を割いている。そこでの諸定義を総合すると「調査票・質問紙調査において、前の質問とその回答によって後の質問への回答が影響を受けること」となる。ただし近年の英語圏の文献では、ほぼ同義語として文脈効果(context effect)という言葉が用いられ、その一類型を特にキャリアオーバー効果と呼ぶことがある<sup>(2)</sup>。しかしキャリアオーバー効果という言葉の日本での普及状況をふまえ、本稿は、上記の定義でこの言葉を使うことにする。

また、質問の順序効果(question order effect)という言葉が類義語として用いられることも多い。しかし第2節で述べるように、質問の順序効果として顕在化しないキャリアオーバー効果もあるので、両者は区別されるべきである。また自記式質問紙では、質問群の回答順序が事実上回答者任せになり、後置の質問が前置の質問の回答に影響を与える可能性がある(井垣1968, Schwarz & Hippler 1995)。しかし、本稿の定義での「前」「後」を時間的に解釈すれば、この現象もキャリアオーバー効果の一種と理解できよう。

本稿の目的は、近年の海外での実証研究に基づき、キャリアオーバー効果に関して未だ完全な解決をみない二つのジレンマを指摘し、これらの解決の糸口を論じることである。第2

節および第3節でそれぞれ論じるように、この二つのジレンマは互いに関連する。最後の第4節では、これらの議論をまとめてキャリアオーバー効果をめぐる今後の検討課題を示す。

第一のジレンマ（第2節）は、キャリアオーバー効果には通常の観察方法（質問の順序効果の指摘）では捕捉困難な類型（意識項目間の相関効果）があり、そこが理論的には重大な問題（自己発生的妥当性）をはらんでいることから生じる。つまり、この問題の核心に触れようとするとき観察困難な部分だけに思弁的な議論に陥りやすく、逆に観察が容易な効果（比率・平均効果）に議論を絞ると重大な問題が無視されてしまう。このジレンマは、従来あまり指摘されておらず、キャリアオーバー効果を巡る過去の議論を混乱させている。今後の研究課題は、この直接観察し難い部分を、どう捉えるかにあると主張したい。

第二のジレンマ（第3節）は、キャリアオーバー効果を弱めるには、関連項目間に無関係な緩衝項目（buffer item）を挿入し調査票上に分散させよとされるが、そうすると調査票がまとまりを欠いてしまう問題である。このジレンマは邦文概説書で既に指摘されている（宝月他 1989：77頁）。「関連項目をまとめるとキャリアオーバー効果はどんなものであれ強まる」また「そうした効果は何であれ避けるべきである」と仮定される限りこのジレンマからは抜け出せない。しかし、これらの仮定は、近年の実証研究の知見に基づいて考えると、必ずしも真実ではない。その知見によれば、ある種の関連項目（類似した事柄を尋ねる諸項目）をまとめると、特定の効果（対比効果）は強まるが別の効果（同化効果）は弱まる。そして、強まる効果は必ずしも避けるべき種類のものではない。そして、このことは概説書の「関連項目をまとめよ」とする常識的示唆の、理論的・実証的裏付けの一つになるのではないか。一部の概説書は、キャリアオーバー効果を適切なものにすることを推奨している（盛山 2004：89頁）。近年の実証研究の知見は、この「適切なもの」が具体的に何なのかを探る手がかりになると主張したい。

## 2．観察困難なキャリアオーバー効果の理論的重要性

### 2.1 比率・平均効果と相関効果

まずキャリアオーバー効果を論じる上での基本的な事柄をいくつか整理しよう。第一に、個人レベルの効果と集計レベルの効果を区別すべきである。キャリアオーバー効果は除去不能だといわれることがあるが（宝月他 1989：77頁、盛山 2004：89頁）、個人レベルの効果は文字通りそうかもしれない。しかし集計レベルの効果は、回答者の大多数が先行項目から同種の影響を受けてはじめて生じる。たとえ、ある調査の回答者全員において個人レベルの効果が生じたとしても、効果の具体的な性質が十人十色ならば、集計レベルでは効果が相殺され、何の効果も観察されない。調査論で主に問題となるのはこの集計レベルの効果である。そこで以下では特に断らない限り、この集計レベルの効果について論じる。個人レベルの効果については結論部分で改めて論じる。

第二に、主な集計レベルの効果には、比率・平均値に対する効果と、相関に対する効果がある。以下を「比率・平均効果」を「相関効果」と略称する。また、説明の便宜上、以下、影響を「与える」質問項目を文脈（context）項目、影響を「受ける」質問項目を対象

(target) 項目と呼ぼう。 は文脈項目の影響で対象項目の回答の比率・平均値が変化する現象であり、 は文脈項目の影響で対象項目の回答が変化し、その結果、文脈項目と対象項目との相関が上昇または低下する現象である。同一の文脈・対象項目のセットにおいて と が同時に観察されることもあるが、 が観察されても が観察されないケース(表1参照)、またはその逆(表2参照)が存在する。 の相関効果は、このように抽象的に定義すると遍在する事柄のように思われるが、その経験的な観察は に比べてはるかに難しい。何故なら、以下で論じるように、相関効果は「質問の順序効果」として顕在化しない場合があるからだ。

## 2.2 キャリーオーバー効果と質問の順序効果

集計レベルの効果の最もポピュラーな観察法は、質問の順序効果を指摘することである。質問の順序効果とは、質問の順序を変えた数種類の調査票を同質的な複数の回答者集団に割り当てたとき(スプリット・パロット法)、観察される回答傾向の違いである。典型例は、項目 AB をこの順に調査した結果と、BA の順に調査した結果との差異である。ただし自計式質問紙で、位置的に後の質問が前の質問に影響しうるときは、順序 AB, 順序 BA, A 単独, B 単独の4種類の調査票を比較する必要がある。この場合を考慮すると、以下の説明が極めて煩瑣になるため、特に断らない限り、後問が前問に影響しないことを前提に議論を進める。しかし、影響しうる場合でも、本稿の主張そのものに特に変更は無い。

表1 質問の順序効果：比率効果の観察例( は四分点相関係数)

順序 A	B	質問 B			順序 B	A	質問 B		
		はい	いいえ	計			はい	いいえ	計
質問 A	はい	20	10	30	質問 A	はい	20	30	50
	いいえ	30	40	70		いいえ	10	40	50
	計	50	50	100%		計	30	70	100%
= 0.218					= 0.218				

先述の比率・平均効果は、質問の順序効果を指摘することでほぼ確実に把握できる。例えば、順序 AB における B の比率・平均と、順序 BA におけるそれを比較し、これらが有意に違えば、A の B に対する効果の存在が確認され、同じならその不在が確認される。また観察された比率・平均の差の大きさで効果の強さも知ることができる。つまり比率・平均効果は質問の順序効果と同一視してよい。表1は比率効果が観察される仮想例である。質問 A の「はい」の割合は B を前置しないと 30%、前置すると 50% となり、B の「はい」も A を前置しないと 30%、前置すると 50% となる。これら 20% の差異がキャリーオーバー効果とみなせる。

一方、相関効果の場合、この方法で確実に把握できるのはいわば「氷山の一角」に過ぎない。相関効果も、順序 AB での AB 間の相関と、順序 BA でのそれが違えば効果の「存在」は確認できる。例えば表2の仮想例がそうである。順序 AB での両者の相関は 0.2、順序 BA では 0.6 であり、この 0.4 の差異が相関効果の存在を示してくれる。しかし確認されたのは項目 AB 相互の影響力の差異であって、影響力の絶対的な大きさではないことに注意したい。順序

ABでの0.2という相関も、項目AからBへのキャリアオーバー効果によってもたらされたものかもしれないからだ。こうして、表2で顕在化した効果は「氷山の一角」である可能性がある。仮に順序AB、順序BAそれぞれにおける相関が等しくても、効果の「不在」は確認できない。なぜなら相関の違いが観察されるのは、項目AからBへの影響力とBからAへの影響力が非対称なときだけであり、両方向の影響力が等しければ（両方向の影響力が共にゼロのときと同様）相関の違いは生じないからである。

表2 質問の順序効果：相関効果の観察例（ は四分点相関係数）

順序 A	B	質問 B			順序 B	A	質問 B		
		はい	いいえ	計			はい	いいえ	計
質問 A	はい	30	20	50	質問 A	はい	40	10	50
	いいえ	20	30	50		いいえ	10	40	50
	計	50	50	100%		計	50	50	100%
= 0.2					= 0.6				

### 2.3 意識項目間の相関効果の観察困難性

では、相関効果のうち質問の順序効果として顕在化しない部分を観察する他の手段は有るのか？答えは文脈・対象項目の質問内容によって異なる。まず文脈・対象項目の双方が意識項目だった場合を考える。すなわち調査票・質問紙での回答者本人の報告以外に観察手段の無い態度・意見・信念等を尋ねる質問項目である。こうした意識項目間の相関効果は、質問の順序効果として顕在化しない限り、厳密に言えば直接的観察は不可能である。

意識項目間の相関は、同一の回答者にそれらの意識項目を尋ねることではしか測定できない。よってそこでは常にキャリアオーバー効果の存在が疑われる。皮肉なことだが、この種の効果は、常にその存在が疑われるがゆえに、その存在／不在をはっきり確認できないのである。つまり、ある状況における効果の存在／不在を確認するには、効果が確実に存在しない状況との比較が必要になるが、意識項目間の相関効果の場合、後者の状況が設定できない。対照的に比率・平均効果の場合、効果不在の状況を実際に設定できる。例えば、項目Bのみの調査票ではBへの効果は不在であり、順序BAの場合も自計式質問紙の場合を除けばBへの効果は不在といえる。これらの状況と順序ABの場合とを比較すれば、AからBへの効果の存在／不在を確認できる。しかし意識項目間の相関効果の場合、項目AB間の相関が問題であるため、A単独やB単独の調査票は無意味だし、順序をどう変えても項目AB間の相関効果の可能性を排除できない。従って意識項目間の相関効果の直接的な観察は不可能なのである。

意識項目間の相関効果の近似的または間接的な観察（推定）手段としては、文脈項目と対象項目との間に緩衝項目を挿入し、調査票上に分散させる方法が考えられる。第3節で論じるようにこの方法は、キャリアオーバー効果を解消する手段としてむしろおなじみである。もし緩衝項目が効果を「解消する」手段として有効ならば、それと同じ理由で、効果を近似的に「確認」する手段としても有効だろう。つまり緩衝項目によって近似的に文脈項目不在（実際は対象項目と離れて存在）の調査票を作り出し、そこでの文脈・対象項目間の相関と、

両項目が近接している調査票での相関とを比較すればよい。この方法がうまくいけば相関効果の全貌がほぼ確認でき大変都合がよい。ただしこの方法は、緩衝項目によって効果が確実に軽減されるという保証が無い限り有効でない。その保証が無い場合、緩衝項目挿入時と非挿入時とで、文脈・対象項目間の相関が等しくても、相関効果が不在なのか、それは存在するが緩衝項目が無効なのか、の区別ができないからである。そこでこの問題は第3節で緩衝項目の機能を検討した後、改めて論じる。

一方、行動・事実項目がからむ相関効果は比較的容易に観察できよう。行動・事実項目の特色は、キャリアオーバー効果の影響を被りにくい場合があり(例えば勤務地、子供の有無等の基本属性)、原理的には観察・記録等の代替的手段でも測定可能なことである。例えば、ある種の基本属性項目と意識項目との間の相関効果の全貌は、質問の順序を逆転すれば顕在化するだろう。またもし文脈項目または対象項目が観察や記録で代替できるなら、それを利用した場合の相関と、調査票調査のみによる相関とを比較することで推定可能である。

#### 2.4 意識項目間の相関効果と自己発生的妥当性

キャリアオーバー効果の問題の厄介なところは、前節で述べた最も観察困難な部分(意識項目間の相関効果)が、他の部分に劣らず理論的に重要であることだ。意識項目間の相関によって何らかの仮説を検証するのは、心理学系の研究の常套手段であり、社会学や政治学でもかなり一般的な研究方法だからである。最悪の場合、キャリアオーバー効果によって意識項目間の相関が増大し、その結果、仮説の妥当性が見かけ上、立証されてしまうこともある。Feldman と Lynch (1988) はこの種の事態を総称して、自己発生的妥当性 (self-generated validity) と呼び、警告を発している<sup>(3)</sup>。彼らによれば、この事態が最も深刻になるのは、一つの信念・態度などを複数の評定尺度で測定する場合だという(前掲論文:427頁)。表3はこの種の評定尺度群の典型的な形式を示した。

表3 評定尺度群の例(番号に丸をする)

意見の文章	そう思う	まあ そう思う	あまり そう思 わない	そう思 わない
1, ~ ~ ~ ~	1	2	3	4
2, ~ ~ ~ ~	1	2	3	4
3, ~ ~ ~ ~	1	2	3	4
4, ~ ~ ~ ~	1	2	3	4

Feldman と Lynch (1988) によれば、この種の評定尺度群の各項目間の相関はキャリアオーバー効果の影響を最も被りやすく、通常、各項目間の相関は増大するという(前掲論文:424頁)。この種の評定尺度群は全項目を一括して尋ねるのが実施上便利だが、すると内容・形式とも非常に類似した質問群を連続して繰り返し尋ねることになる。彼らの理論によれば、この時キャリアオーバー効果は極大になるのである(前掲論文:427頁)。

さらに、この種の評定尺度群から(合計点を求めるなどして)一つの態度尺度が構成され

る場合、その態度尺度の信頼性は項目間の内的一貫性を基準に評価される。そして内的一貫性の指標（例えば 係数）は各項目間の相関に基づいて計算される。ところが Feldman と Lynch（1988）によれば、各項目間の相関は、キャリアオーバー効果によって不当に高い値を示す筈である。これが事実なら（第3節にみる通り一概にそうはいえない）、この種の評定尺度群の内的一貫性は、尺度群それ自体が生み出すキャリアオーバー効果により自動的に保証されてしまう。だとすると、この種の態度尺度の多くは、自らのバイアスによって自らの有効性を主張してきたことになる。

ただし評定尺度群の問題に限っていえば、Feldman と Lynch（1988）が描いたのは、理論上の「最悪のシナリオ」であって、彼ら自身が、実際に様々な評定尺度群を調べ、この事態を経験的に指摘したわけではなく、過去の研究例<sup>(4)</sup>を引用したわけでもない。これは認知心理学の様々な実験結果（しかも、その多くは調査票・質問紙調査の状況を念頭に置いていない）を接木することで導き出された理論的予想である。本稿は、その後の実証研究を踏まえ、第3節でこの理論的予想に修正を加えることになる。

## 2.5 キャリアオーバー効果をめぐる議論の混乱について

キャリアオーバー効果の類型として整合効果（consistency effect）と対比効果（contrast effect）がある（例は注5を参照）。整合効果は同化効果（assimilation effect）とも呼ばれる。これらは Schuman と Presser（1981）が論じて以来、最もよく議論されてきた類型である。ところが両概念の明快で説得力に富む操作的定義は未だに提案されていない。この区別の提案者である Schuman と Presser（1981：127-128頁）の説明によれば、整合効果は前の質問の回答に後の質問の回答が（前の質問が存在しないときよりも）接近することであり、対比効果は前の質問の回答から後の質問の回答が（同上）乖離することである。しかし「接近」および「乖離」が、比率や平均値の差の「縮小」「拡大」のことなのか、項目間の相関の「増大」「減少」のことなのかは不明確である。以来、様々な論者によって整合（同化）効果と対比効果という対概念が使われたが両効果の操作的定義は不明確なままだった。Moore（2002）はこの状態を「奇妙なこと（81頁）」とし、比率の差の縮小・拡大により両概念を操作的に定義した<sup>(5)</sup>。しかし両概念と相関の増減とを結びつける論者もあり（Tourangeau 1999:112頁）、Moore（2002）の操作的定義が今後全面的に採用されるとは思えない。相関効果を無視すると、前節でみたような重要な論点が抜け落ちてしまうからである。

このように、キャリアオーバー効果をめぐる議論がしばしば混乱するのは、経験的に観察困難な部分（意識項目間の相関効果）、すなわち操作的に定義することが困難な部分にかなりの関心が寄せられてきたからだと思われる。よって重要な概念が曖昧なまま放置されたり、逆に明快だが皮相的な定義がなされたりする。比率・平均効果に議論を限定すれば、Moore（2002）のように質問の順序効果として明快な定義ができて誤解も生じない。しかし、それでは Feldman と Lynch（1988）が提起するような意識項目間の相関効果の問題にアプローチできない。このジレンマにより、キャリアオーバー効果をめぐる議論は混乱しがちなように思われる。よって、今後の方法論的課題は、意識項目間の相関効果の観察（推定）方法を工夫することだと本稿は言いたい。

### 3. キャリーオーバー効果と緩衝項目

#### 3.1 緩衝項目とキャリーオーバー効果の強さ・方向性

文脈項目と対象項目の間に緩衝項目を挿入することは、キャリーオーバー効果を弱めるための常套手段とされる。また、この常套手段が実際に有効なら、既述の相関効果も、緩衝項目挿入時の文脈・対象項目間の相関と、挿入しないときのそれとを比較することでうまく観察できる。しかし緩衝項目の挿入により関連項目が調査票上に分散するため、調査票全体のまとまりはどうしても損なわれる。ここから回答意欲の低下といったデメリットが生じる危険もある。従って、関連項目を続けて尋ねた時と、緩衝項目を挿入した時とで、調査対象者の回答プロセスに実際にどんな違いが生じるのか、慎重に検討する必要がある。そこで、この点に関連する興味深い実験を行ってきた Schwarz らの研究を振り返ってみたい。

Wänke と Schwarz (1997) のレビューによれば、緩衝項目の機能に関する過去の実証研究の知見は様々である。定説通り緩衝項目はキャリーオーバー効果を有意に弱めるとする知見もあれば、弱める機能はあるが統計的に有意な程ではないとする知見もあり、弱める機能を全く持たなかったとする知見もある。いずれにせよ、緩衝項目によってキャリーオーバー効果が完全に解消されるわけではなさそうだ。また、緩衝項目の挿入によって、キャリーオーバー効果の方向性が、対比効果から同化(整合)効果に変化したとする知見も存在する(Ottati, Riggle, Wyer, Schwarz & Kuklinski 1989)。これは先ほどの Feldman と Lynch (1988) の理論的予想とも関連する非常に興味深い知見なので、やや詳しく紹介しよう。

Ottati 他(1989)は、著者のうち二人が所属するイリノイ大学の政治学専攻の学生約1000人を被験者とし、表現の自由や雇用機会均等などの政策「一般」に対する賛否が、その前に「特殊」な集団に対するこれらの政策への賛否を尋ねることで、どう変化するかを実験した。例えば、前にPTAの表現の自由に対する賛否を質問した場合(賛成多)と、KKK(クークラックスクラン)の表現の自由に対する賛否を質問した場合(反対多)とで、表現の自由「一般」に対する賛否がどう異なるか、といった類の実験が、政策や集団を様々に変化させながら行われたのである。

このように質問の順序が特殊質問から一般質問へと進む場合、可能性としては同化効果と対比効果の双方が起こりうる。回答者が政策一般の中に「PTA」または「KKK」に対する政策を含めて考えれば、政策一般への賛成度は「KKK」の後よりも「PTA」の後で高くなるだろう(同化効果)。逆に回答者が「一般」から「PTA」や「KKK」を除外し、これらと「一般」とを対比するなら、政策一般への賛成度は「PTA」の後よりも「KKK」の後で高くなるだろう(対比効果)。実験の焦点は「PTA」「KKK」等の特殊質問の直後に一般質問を行う場合と、特殊・一般質問間に8つの緩衝項目を挿入した場合とで、効果の方向がどう異なるかだった。

実験の結果、緩衝項目を挿入すると政策一般への賛成度は、低好感度集団(KKK等)よりも高好感度集団(PTA等)の質問の後で高くなり、同化効果が生じていた。一方、緩衝項目を挿入しない場合、政策一般への賛成度は逆に高好感度集団よりも低好感度集団の質問の後で高くなり、対比効果が生じていた。こうして、緩衝項目はキャリーオーバー効果を解消せず、その方向性を逆転させたのである。

### 3.2 キャリアオーバー効果と会話の規範

何故、緩衝項目を挿入しないと対比効果が生じ、挿入すると同化効果が生じたのか？ Schwarzらの解釈によれば、それは類似した質問が連続して行われる場合、同じ話の繰り返し（冗長性）を避けようとする会話の規範（conversational norm）が作用し、回答者は前問とは違った角度から後問を解釈し、それに従って調査者に新たな情報を提供しようとするからだという（Sudman, Bradburn & Schwarz 1996, Wänke & Schwarz 1997）。一方、文脈項目の後に無関係な緩衝項目が挿入されると、冗長性回避の規範は作動しない。緩衝項目の挿入により、一度、会話の流れが断ち切られるため、文脈項目と対象項目は、回答者にとって、一つの会話の流れの中の質問とはみなされない。両項目は、むしろ別の会話の流れの中の、単なる類似した質問とみなされ、回答者は、文脈項目に回答した際に頭に浮かんだ情報をそのまま使って対象項目に回答する。こうして同化効果が生じる。

この解釈を裏付けるのが Strack, Schwarz および Wänke (1991 : 120 頁 : 研究 2) による実験である。これは緩衝項目の実験ではないが、調査における冗長性回避の規範の実証例として興味深い。ここでは、幸福感と生活満足感という極めて類似した質問項目間の相関効果が問題となった。被験者は Wänke が所属するマンハイム大学の学生 45 名であった。ある被験者群には、この 2 項目が別々の質問紙で質問された。すなわち幸福感の項目は最初の質問紙の末尾で提示され、生活満足感の項目は次の質問紙の冒頭で提示された。この二つの質問紙は、字体やヘッダを変えてあり、別々の調査であることが強調されている。この時、両項目の相関は、0.96 と非常に高かった。両項目はほぼ同じ意味に解され、同様に回答されたのである。一方、別の被験者群には、この 2 項目が「次にあなたの生活について二つお伺いします」という導入文をつけて、同じ質問紙の末尾に並べられた。すると両項目の相関は 0.75 に低下した。Schwarz らの解釈によれば、前者で冗長性の規範が働かず、後者でそれが働いたためである。日常会話になぞらえれば、「違う人」に同じ話をするのはかまわないが、「同じ人」に同じ話を繰り返すのはためらわれたということになる。

ただし、この実験に対しては、「二つお伺いします」という導入文で 2 項目の違いが強調されたのではないかと、との解釈もありうる。しかし「類似した項目を別々に尋ねるよりも、まとめて質問した方が、互いの相関が低くなることもある」という知見の持つ意味は大きい。次節でこの点について敷衍する。

### 3.3 似た項目をまとめて問うことの意味

以上の議論を踏まえて、改めて Feldman と Lynch (1988) の所説を再検討してみよう。彼らが描いた「最悪のシナリオ」の一つは、キャリアオーバー効果により、一つの態度や信念を測定するための評定尺度群において、高度の内的一貫性が「自己発生」することであった。つまり、この種の評定尺度群は、内容・形式の似た質問群をまとめて繰り返し尋ねるので、極めて大きなキャリアオーバー効果を生み出し、その結果、各項目間の相関は一般に「強まる」とされた。

しかし、Schwarz らの実験によれば、類似した質問群が、調査票の一箇所に（導入文で括られるなどして）まとめられるときの方が、「冗長性回避の規範」が作動しやすく、各項目の二

ニュアンスの違いが回答に反映される。要するに、似た項目を並べると項目間の相関が「弱まる」ことがあり、このとき内の一貫性は上昇せずに低下する。このように Feldman と Lynch (1988) および Schwarz らは、似た項目をまとめて問う際に、それぞれ方向性の異なる効果を予測している。そして予測を支持する証拠を提出しているのは Schwarz らの方である。

ところで、似た項目をまとめて問うことで各項目間の相関が「強まる」ことの危険性は、Feldman と Lynch (1988) が既に指摘しているが、Schwarz らがいうように会話の規範によって「弱まる」としたら果たしてそれは「危険な現象」といえるのか？各項目の内容や調査目的によってケースバイケースといわざるを得ない。しかし、一つの構成概念を測定する評定尺度群の場合、それはむしろ「望ましい現象」のように思ってしまうのは筆者だけだろうか。

そもそも、こうした評定尺度群を作成・使用する調査者も、回答者が全項目に同様に回答すること(相関が1となること)を期待してはいない。多数の尺度項目を用意する理由の一つは、単一の項目では、構成概念の内容領域を網羅できないためである。とすれば、回答者が一つ一つの項目のニュアンスの微妙な違いを意識して回答するのは、内容領域の網羅という目的からして比較的真っ当な回答プロセスといえよう。一部の調査法の概説書では、適切なキャリーオーバー効果ならば利用せよといった示唆(盛山 2004: 89 頁)がなされることもあるが、「会話の規範」によるキャリーオーバー効果は、利用できるものの一つに数えられそうな気がするのである。

こう考えると、調査法の概説書で「関連項目はまとめよ」とされることの意義の一つがみえてくる。一口に「関連項目」といっても大きく分けて二種類ある。第一は、互いに関連するが明らかに別々の事柄を尋ねる場合<sup>(6)</sup>であり、第二は、同じ事柄の諸側面や、よく似た同種の事柄を尋ねる場合である。前者はともかく、後者の場合、関連項目をまとめて尋ねることで、項目間の微妙な差異を際立たせることができる。よく似た双子の兄弟でも一緒に並んでいると区別できることがあるように、似た質問項目を列挙する目的の一つは、似たもの同士の微妙な差異を際立たせる為なのである。

とはいえ Schwarz らのいう「冗長性回避の規範」が実際の多様な調査場面で常に作動するとは限らない。Schwarz らの被験者はいずれも実験者が所属する大学の学生であり、しかも Ottati ら(1989)は政治学専攻の学生に政策課題について質問している。よって回答意欲や調査者に対する信頼感はかなり高かったものと想像できる。私見によれば「冗長性回避の規範」は、回答者に一定の精神的作業を要求するから、回答意欲が低いと作動しないのではないか。また回答者が冗長性回避の規範に従うのは「調査者も同じ規範に従っている」と信頼できるときだろう。つまり「似た質問ばかりする」と感じるか「似た質問を敢えてするのは、実は別の事がききたいからだろう」と感じるかは調査者への信頼度による(学力テストで同様の事態に直面したときの学生の反応を想起して欲しい)。また類似した質問を二つだけ並べる場合は冗長性回避の規範が作動するとしても、ある種の心理測定のように数十個に及ぶ類似した質問群を構成する場合、最後までこの規範がうまく作動するだろうか。質問群の後半では回答意欲や信頼感が低下し、各質問のニュアンスの違いが回答に反映されなくなるのではないかと<sup>(7)</sup>。こうした点も含めて、似た項目をまとめて問うことで生じる効果については、今後、さらなる実証研究が必要であろう。

いずれにせよ、緩衝項目を挿入することで、あらゆるタイプのキャリアオーバー効果が弱まるとはいえないことが明らかとなった。Schwarzらの実験によれば、緩衝項目を挿入することによって、対比効果は弱まるが、そのかわり同化効果が強まる。そして調査目的によっては、対比効果よりもむしろ同化効果の方が由々しき問題を引き起こすであろう。こうして緩衝項目の有効性に疑問符がつくならば、当然ながら、調査票のまとまりを犠牲にしてまで関連項目を分散させる必要は無い。したがって、本稿が論じた第二のジレンマは、調査によって決して宿命的なものとはいえないのである。

#### 4．今後の課題

最後に以上の議論をまとめ、今後の課題を2点示しておきたい。第一に、意識項目間の相関効果の（間接的な）観察手段を工夫することが必要である。第2節で述べた通りこの種の効果は直接的な観察が不可能なため、比率・平均効果に比べ、その存在や性質がよく分かっていない。相関効果の近似的な観察方法として、関連項目間に緩衝項目を挿入した場合と、しない場合とを比較する方法があり、実際に試みられた例もある（注4を参照）。この方法は、プリテストとして比較の手軽に実施できる。しかし第3節でみたように、緩衝項目を挿入することで、キャリアオーバー効果が完全に解消されるとは限らないため、その有効性には限界がある。

同種の方法としては、関連項目を充分時間を空けて（例えば日を変えて）測定した場合と、続けて尋ねた場合とを比較することも考えられる。この方法は、プリテスト法としては労力のいる方法であり、しかも信頼性評価のための再テスト法（test-retest method）と同様のジレンマを抱えている。すなわち、どれだけ時間を空ければ前回の質問の記憶が薄れるのか不明確なことと、時間を空けすぎて意見や態度の正味の変化が起きると有効でなくなることである。他にも色々な方法が考えられるが、質問の順序を変えたり、緩衝項目を挿入したりする方法に比べると、かなりの手間がかかることを覚悟すべきだろう。

発想を変えて、個人レベルの効果を直接観察する方法も試みられてよい。つまり個人個人がどんな思考プロセスを経て各質問に答えているのかを、質的なインタビューによって直接把握するのである（例えばSudman, Bradburn & Schwarz 1996 : 15-54頁）。この種の観察方法によって、集計レベルで顕在化しない個人レベルの効果が観察可能になる。それなりに労力のいる方法だが、量的調査のプリテストとしてこの種の質的調査を行うことは、調査手順として至極真っ当であろう。

第二に、関連項目間に緩衝項目を挿入することの意義、逆にいえば、関連項目を一箇所ですべてまとめて質問することの意義について、改めて考え直す必要がある。第3節でみた通り、緩衝項目の挿入は、キャリアオーバー効果を弱める手段として推奨されてきたが、この手段は「関連項目をまとめて分かりやすい調査票にせよ」という調査法の基本と矛盾する。しかも過去の実証研究によれば、緩衝項目を挿入してもキャリアオーバー効果が弱まらないことが多い。また緩衝項目の挿入・非挿入により、効果に質的な転換が起きるとする研究（Ottati 他 1989）もある。しかも、この研究で明らかにされた非挿入の場合の効果（対比効果）は、調

査目的に必ずしも悪影響を及ぼすものではなく、回答者が類似する諸項目の一つ一つの意味を慎重に見分けている証拠とすらいえるのである。こう考えると、緩衝項目の挿入を、キャリアオーバー効果を弱める手段として無条件に推奨するわけにはいかない。

そもそも「関連項目をまとめよ」という常識的示唆は、当然ながら、そうすることで回答プロセスに何らかの変化が生じることが前提となっている。従って、ある種の「適切な」キャリアオーバー効果が暗黙のうちに仮定されているとみるべきである。つまり、この示唆には「関連項目をまとめると、何らかのキャリアオーバー効果によって、質の良い回答が得られる筈だ」という考え方が織り込まれている。今後の課題は、こうした暗黙の理論を明文化し、実証的な調査法研究によって検証していくことであろう。

「重要なことは、キャリア・オーバー効果をなくすことではなく、適切なものにすることである」といわれる(盛山 2004: 89 頁)。しかし概説書の指摘はここまでで、具体的にどんなキャリアオーバー効果なら「適切」といえ、その種の効果はどんな場合に生じるのかまでは明らかにされていない。今後の調査法研究では、この点が具体的に明らかにされねばならない。本稿の第3節でレビューした一連の研究は、こうした問題点に答えていくための一つの糸口になると思われる。

#### 〔注〕

- (1) 「もちこし効果」といった訳語をあてることもある(安田・原 1982, 平松 1998)。
- (2) こうした用語法の例としては Tourangeau 他(1989)を参照。ここでは、キャリアオーバー効果という言葉が2.5節で述べる整合効果の類義語として用いられている。
- (3) 自己発生的妥当性の概念には、ある行動の意図の測定によって、その行動が促され、意図と行動とを関連づける研究者の仮説が(見かけ上)立証される現象も含まれる。例えば、意図の測定の予測的妥当性(Carmines & Zeller 1979: 訳書 14 頁)を評価する基準は、行動そのものであるが、意図の測定によって行動が促されるとすれば、予測的妥当性は自己発生することになる。最悪のシナリオは、その結果、「意図の測定は行動の測定の代理手段として有効だ」といった主張がまかり通ることである。
- (4) Schuman と Presser (1981: 52-54 頁)でいくつかの先行研究がレビューされている。そこでは、Feldman と Lynch (1988)の「最悪のシナリオ」は支持されていない。いずれの研究も、緩衝項目挿入時と非挿入時とで、関連項目間の相関の違いを指摘できなかったからである。しかし、これらの結果は「最悪のシナリオ」を支持していないが、却下しているわけでもない。なぜなら、効果そのものが不在だったのか、緩衝項目が効果を弱める機能を果たさなかったのか、判断し難いからだ(第3節参照)。
- (5) Moore (2002)は、自らの定義による整合効果および対比効果の実例として1997年のGallupによる調査結果を提示する。同年9月の調査によれば、民主党(当時)の「クリントンは誠実で信頼に値すると思うか」「ゴアは(以下同文)」との質問を続けて行ったところ、「クリントン」に関する「はい」の割合は、先に尋ねると50%、後に尋ねると57%だったのに対し、「ゴア」に関する「はい」は先に尋ねると68%、後に尋ねると60%となった。「クリントン」と「ゴア」の支持率の差は、それぞれが先に尋ねられた場合(18%)よりも、後で尋ねられた場合(3%)の方が縮小している。Moore (2002)はこのことをもって、整合効果としている。両者は同僚と考えられたため、続けて質問することで両者の支持率は接近したのだと解釈できる。また、同年3月の調査では、同じ質問を共和党(当時)の「ギングリッチ」と「ドール」について、やはり続けて行った。その結果、「ギングリッチ」に関する「はい」は先に尋ねると41%、後で尋ねると33%となり、「ドール」に対する「はい」は先に尋ねると60%、後で尋ねると64%となった。両者の支持率の差は、それぞれが先に尋ねられた場合(19%)よりも後で尋ねられた場合(31%)の方が拡大しており、Moore (2002)はこのことをもって、対比効果としている。両者は比較の対象として考えられたた

- め、続けて質問することで両者の支持率は乖離したのだと解釈できる。
- (6) 例えば「海外からの出稼ぎ労働者の流入は労働者の賃金を引き下げるといわれていますが、あなたはこの意見に賛成ですか、それとも反対ですか」「海外からの出稼ぎ労働者が近年増加していますが、あなたはこのことについてどう思いますか」(宝月他 1989 : 77 頁)といったタイプの関連質問である。両者は別々の事柄を尋ねる質問であるが、前者は後者に答える際の根拠になりうる。
- (7) この見方を支持するような実験結果が存在する。Nowles はある人格を測定するための 30 項目に及ぶ評定尺度群 (Internal-External Locus of Control scale) において、各質問が必ず 1 ~ 30 番目の各順番にくるように、質問の順序を変えて 30 種類の調査票を用意し、120 人を対象にスプリットバロット実験を行っている (Nowles 他 1992)。その結果、項目信頼性 (ある項目とその他の項目との相関に基づく) は、その項目が調査票に現れる順番が後になればなるほど、高まる傾向 (順番 × 項目信頼性の相関係数  $r=.510$ ) にあった。質問群の後半では、各質問のニュアンスの違いは、回答にあまり反映されなくなるようである。

## 〔文献〕

- Carmines, E. G., and R. A. Zeller, 1979, *Reliability and Validity Assessment*. Sage. (訳書:カーミン・ツェラー著, 水野欽司・野嶋栄一郎訳, 1983, 『テストの信頼性と妥当性』朝倉書店)
- Feldman, J.M., and J.G. Lynch, Jr., 1988, Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior. *Journal of Applied Psychology* 73:421-435.
- 平松貞実, 1998, 『世論調査で社会が読めるか: 事例による社会調査入門』, 新曜社。
- 宝月誠・中道實・田中滋・中野正大, 1989, 『社会調査』, 有斐閣。
- 井垣章二, 1968, 『社会調査入門』, ミネルヴァ書房。
- Knowles, E. S., M. C. Coker, D. A. Cook, S. R. Diercks, M. E. Irwin, E. J. Lundeen, J. W. Neville, and, M. E. Sibicky, 1992, Order Effects within Personality Measures. In Schwarz, N. and S. Sudman (Eds.), *Context Effects in Social and Psychological Research*, pp.221-247. Springer-Verlag.
- Moore, D. W., 2002, Measuring New Types of Question-Order Effects. *Public Opinion Quarterly* 66:80-91.
- Ottati, V. C., R. S. Wyer, Jr., E. J. Riggel, N. Schwarz, and J. Kuklinski, 1989, Cognitive and Affective Bases of Opinion Survey Responses. *Journal of Personality and Social Psychology* 57:404-415.
- Schuman, H., and S. Presser, 1981, *Questions & Answers in Attitude Surveys*. Academic Press.
- Schwarz, N., and H. J. Hippler, 1995, Subsequent Questions May Influence Answers to Preceding Questions in Mail Surveys. *Public Opinion Quarterly* 59:93-97.
- 盛山和夫, 2004, 『社会調査法入門』, 有斐閣。
- Strack, F., N. Schwarz, and M. Wänke, 1991, Semantic and Pragmatic Aspects of Context Effect in Social and Psychological Research. *Social Cognition* 9:111-125.
- Sudman, S., N. M. Bradburn, and N. Schwarz, 1996, *Thinking about Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass.
- Tourangeau, R., 1999, Context Effects on Answers to Attitude Questions. In Sirken, M. G., D. J. Hermann, S. Schechter, N. Schwarz, J. M. Tanur, and R. Tourangeau (Eds.), *Cognition and survey research*, pp. 111-131. John Wiley & Sons, Inc.
- Tourangeau, R., K. A. Rasinski, N. Bradburn, and R. D'Andrade, 1989, Carryover Effects in Attitude Surveys. *Public Opinion Quarterly* 53:495-524.
- Wänke, M., and N. Schwarz, 1997, Reducing Question Order Effects: The Operation of Buffer Items. In Lyberg, L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (Eds.), *Survey Measurement and Process Quality*, pp. 115-140. Wiley.
- 安田三郎・原純輔, 1982, 『社会調査ハンドブック: 第3版』, 有斐閣。

(やまぐち よう 現代社会学科)

2005年4月27日受理