

肯定的項目と否定的項目が社会調査データの分析に及ぼす影響：古典的テスト理論を用いた分析

EVALUATING THE INFLUENCE OF POSITIVELY AND NEGATIVELY WORDED ITEMS ON DATA ANALYSES IN SOCIAL SURVEYS: ANALYSING DATA WITH CLASSICAL TEST THEORY

澁谷泰秀¹ 渡部諭² 吉村治正³ 小久保温⁴

¹ 青森大学社会学部, ² 秋田県立大学総合科学教育研究センター, ³ 奈良大学社会学部,

⁴ 青森大学ソフトウェア情報学部

Abstract

The purpose of the study is to evaluate the effects of item wording on data analyses in social survey contexts. Two surveys, survey 1 and 2 conducted in 2011 and 2013 respectively, were compared in terms of the happiness scale and the quality of life scale. The happiness scale was composed of 5 items. All 5 items were worded negatively in the survey 1. The same scale, in terms of the construct and content of the items, was used in survey 2 with 3 positively framed items and two negatively framed items. The contents of the items were identical between the two scales. The only differences were in item expressions. The quality of life scale (QOL), composed of 6 sub-scales, was identical in the two surveys. The reliability coefficients of the happiness scale in both surveys were about the same. The structure of the happiness scale indicated a clear single factor according to factor analysis in both surveys, although there is a significant difference in their mean scores between the two surveys. A possible mechanism of how reversing negative items may have contributed to the significant mean difference was discussed. The difference in score distributions between the two surveys was characterized by the location of modes; mode for the survey 1 was 20, while mode for the survey 2 was 15. There was no clear evidence for supporting a systematic difference in a structure of correlations between happiness scale and QOL in the two surveys, although there was one pair of significantly different correlation coefficients between the two surveys. The possibility of differences in elements of correlation matrices having serious consequences in higher level analyses such as multiple regression analysis and structural equation modeling (SEM) was discussed.

Keyword: Item wording, positively worded items, negatively worded items, correlation matrix, structural equation modeling

1. イントロダクション

社会調査の中で最も頻繁に用いられる方法の一つに質問紙法がある。質問紙法は、複数の項目で構成される調査票を用いて回答者に関する情報や意見などを尋ねる方法である。回答者は、調査項目を解釈し、項目ごとに意思決定をして回答する。質問票を用いて行われる社会は、全ての回答者が質問票の項目を同一に解釈するとの前提に基づいて行われている。この前提は、社会調査において調査対象者の属性や行動を問う調査項目においては、問題となることは少ない（例えば、性別や誕

生年を尋ねる項目など）。しかし、態度や意見などの心理学的構成概念を測定する抽象的な質問内容の項目で構成される尺度においては、信頼性及び妥当性の確立が困難である場合がある。その為、調査項目の構築や調査票のデザインは、項目内容の誤解や回答票への誤記入などが無いようにして、妥当性及び信頼性を確保するために慎重に行う必要がある（Dillman, D. A., 1978; Biemer, P. P. & Lyberg, L. E., 2003 ; Liez, P. P., 2010）。

計量心理学の分野では、同一尺度の中に肯定的・否定的項目を混在させることは、一辺倒な回

答パターンに慣れて項目内容を熟読せずに前項と同様に応答するなどの応答バイアス (response bias) を避けるために必要とされてきた (Nunnally, 1978; Anastasi & Urbina, 1983; Spector, 1992)。肯定的項目と否定的項目を同一尺度内に混在させることが、回答者の応答に影響を及ぼさないとする根拠の一つに期待効用理論がある (von Neumann & Morgenstern, 1947)。期待効用理論によると、我々の意思決定 (判断) は、意思決定の対象となる事象の効用の期待値に基づいて決定される。効用の期待値は、判断をする際の各選択肢の効用とその効用の出現確率との積である。

回答者が社会調査項目に回答するプロセスでは、およそそのような計算が回答者の応答プロセスの中で無意識に生起していると捉えることが可能である。ここで重要な点は、回答者が具体的にどのようなプロセスで効用の期待値を得るののではなく、各選択肢の効用が調査項目で与えられた条件の内容に基づいて決定される点である。すなわち、項目が肯定的にワーディングされていても否定的にワーディングされていても、内容が同一であれば判断の基準となる効用に影響を及ぼさないと規定されている。

このように、意思決定研究においては、期待効用理論に基づいて数理的表現の一意性が前提とされていた。数理的表現の一意性は、意思決定問題において数理的な内容が同一であれば、質問表現が異なっても同一の反応が生起すると仮定する。この視点は社会調査項目の分析に応用され、設問の意味が同一であれば、表現が異なっても同一の項目であると判断されることが多かった。この考え方の延長上に、社会調査などで用いられる尺度が肯定的表現の項目と否定的表現の項目で構成されている場合、否定的項目を逆転させて、肯定的項目と加算して尺度得点とする方法が妥当であるとする考え方がある。しかし、Daniel Kahneman が 2002 年にノーベル経済学賞を受賞して以来注目を集めているフレーミング効果及びフ

レーミング効果の説明とされるプロスペクト理論は、回答者が項目を解釈して効用の期待値を予測するプロセスについて、期待効用理論とは異なり、項目の内容は同一であっても、項目のフレーム (肯定的・否定的) が異なる場合、回答者の意思決定は異なるとする立場をとっている (Tversky & Kahneman, 1981)。

1.1 フレーミング効果とプロスペクト理論

フレーミング効果とは、我々が意思決定する場合、提供される情報の客観的内容が同じであっても、情報提供のされ方の相違によって、意思決定者が構築する問題の捉え方 (心的構成 (フレーム)) が影響され、結果として生起される意思決定に相違が起こる現象である。フレーミング効果の代表的なものに、回答者にリスクを含む意思決定の選択を尋ねるリスク選択型フレーミング効果がある。Tversky & Kahneman (1981) が研究に用いたフレーミング効果は、リスク選択型フレーミング効果である。フレーミング効果の知見によると、肯定的にフレームされた項目と否定的にフレームされた項目では、回答者の反応が異なる場合があり、数理的に同一の意思決定問題でも、問題の表現の相違によって、次に示されるような選択肢の選好逆転が起る場合があることが明らかになっている。

Tversky & Kahneman (1981) は、下記のシナリオを回答者に提示し、回答者に、このシナリオに関する肯定的なフレーム項目或いは否定的なフレーム項目に回答することを求めた。

「人口 600 人の村に奇病が発生しました。もし対策を講じなければ村人全員が死亡します。そこで、厚生労働省によって 2 種類の対策が提案された」

[肯定的項目]

「対策 A : もしこの対策を採用すれば、200 人の村人が助かる」 (低リスク選択肢 ; Low Risk Option)
 「対策 B : もしこの対策を採用すれば、3分の1の確率で 600 人全員が助かり、3分の2の確率でだれも助からない」 (高リスク選択肢 ; High Risk

option)

[否定的項目]

「対策A：もしこの対策を採用すれば、400人の村人が死亡する」(低リスク選択肢; Low Risk Option)

「対策B：もしこの対策を採用すれば、3分の1

る特徴がある。しかし、損失領域では、凸関数となり、損失の初期の効用の減少は利得領域より更に激しく、その後は効用の減少はやや緩やかになるものの、利得領域と比較して効用の減少幅は大きい。プロスペクト理論によると、肯定的項目に

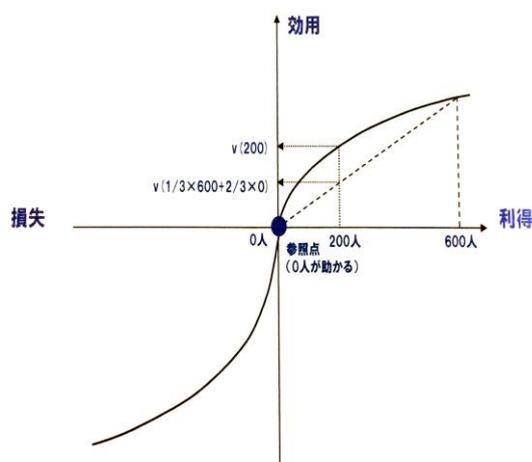


図1. プロスペクト理論 (肯定的項目)

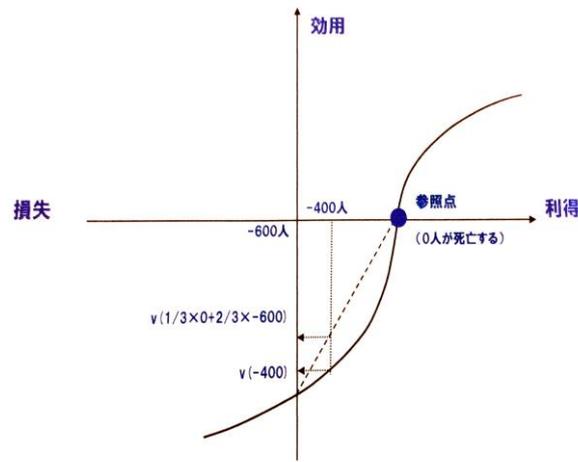


図2. プロスペクト理論 (否定的項目)

の確率でだれも死亡しないが、3分の2の確率で600人全員が死亡する」(高リスク選択肢; High Risk Option)

Tversky & Kahneman (1981) は、上記の心理実験などの結果から、肯定的フレーム項目ではリスク回避選択傾向が高くなり、否定的フレーム項目ではリスク志向選択傾向が高くなることを示した。これらの実験によって、項目の内容が同一であっても、項目のフレームが異なると、回答者は異なる意思決定をすることが実証された。更に、彼らはフレーミング効果が生起するメカニズムをプロスペクト理論で説明した。フレーミング効果の説明については、プロスペクト理論以外にも説明を試みる理論は発表されているが、プロスペクト理論が意思決定の分野では最も受け入れられている (Watanabe & Shibutani, 2010)。プロスペクト理論は、図1に示されるように利得領域において凹関数で示され、利得が増加する初期段階での効用の増加が著しく、その後は効用の増加が緩やかにな

回答する場合、項目の影響によって回答者の心的構成が肯定的フレームとなり、「助かる」という表現が利得を説明する参照点(原点)となる。図1に示されるように、回答者は「0人が助かる」を判断の参照点として効用を判断する。肯定的項目の low risk option では、確実に200人が助かる為、利得の200人に対応する効用関数の high risk option では、利得の600人に対応する効用と参照点を結んだ直線の三分の一の点(実線)の値が効用となる。一方、肯定的項目に対応する値が効用として認識されるため、high risk option と比較して low risk option の効用が高くなる。その為、low risk option の選択頻度が高くなるとするのがプロスペクト理論の説明である。損失領域においては、同様の論理に基づいて high risk option の効用が low risk option の効用と比較して高くなることから、high risk option の選択頻度が高くなることを説明できる(図2参照)。このような背景のもと、社会調査の分野においても、同一内容の設問でも項目

のワーディングが異なると、同一項目として分析することが適切ではないと考えられるようになってきた。

O’Muircheartaigh et al. (2000) によると、否定的項目は、肯定的項目と比較して項目信頼性が低いと報告している。Chen et al. (2007) の報告によると、否定的な項目は因子構造が肯定的項目とは異なる為、尺度の構成概念妥当性の低下を起因させる可能性があるとしている。Alexandrov (2010) は、社会調査の実証研究に基づいて、疫学などで用いられる 1 項目で構成される尺度を用いるとすれば、その項目は肯定的な項目であるべきであると報告している。Rose et al. (2011) は、精神保健サービスの評価尺度の因子分析結果から、項目のワーディング（肯定的・否定的）が因子構造に重要な影響を及ぼすことから、否定的フレーズを用いた項目を使用すべきでないとしている。

本研究の目的は、否定的項目の存在が社会調査で用いられる尺度の分析にどのような影響があるのかについて、項目得点を逆転することが幸福感得点の平均及び分布などの尺度自体に及ぼす影響を、実証データに基づき評価することである。更に、幸福感得点と生活の質の下位尺度との相関係数などに及ぼす影響を評価し、共分散構造分析などを用いて行う 2 次的分析に及ぼす影響などについて考察することである。

1.2 古典的テスト理論

本研究の分析に用いられる古典的テスト理論は、教育評価の分野及び社会・行動科学などで用いられる尺度の分析を行うために広く用いられている理論である。日本においては、これまでは唯一のテスト理論として広く用いられてきた。北米やヨーロッパ諸国などでは、1980 年代後半から項目反応理論が古典的テスト理論に代わり、項目分析や尺度推計に用いられる理論の主流となっている。

古典的テスト理論では、尺度得点は真の値と誤差（得点 = 真の値 + 誤差）で構成され、各項目

は独立した測定単位であるので、変数の推計値は測定単位を総計することで得ることが可能であると仮定している。この仮定に加えてパラレルテスト（複数の等価テスト；平均及び標準偏差の同一性が仮定される）の概念を導入することにより、古典的テスト理論を用いて尺度の信頼性を算出することができる。信頼性係数として広く使用されている Cronbach の α は、古典的テスト理論に基づいて規定された指標である。しかし、多くの計量心理学者は古典的テスト理論に決定的な理論的弱点があることを指摘してきた (Lord & Novick, 1968; Hambleton, Swaminathan, & Rogers, 1991)。

古典的テスト理論を用いて尺度及び項目を推計・評価する場合には、一般的に難易度指数（正解者の割合）や識別力指数（各項目と総合点との相関）が用いられている。これらの指標はテストを受ける被験者集団の能力・特性から影響を受ける。更に、テストによって推計された被験者の能力や特性は、テスト項目の難易度指数や識別力指数などの項目特性によって影響を受けるため、項目の特性と被験者集団の能力や特性の相互作用が推計値を左右するという計量理論としては致命的ともいえる弱点が存在する (Lord & Novick, 1968)。

これらの弱点を克服するために提唱されたのが項目反応理論である。項目反応理論は、日本においても、理論的優位性が認められており、最近は多くの場面で用いられ始めている。しかし、項目反応理論も万能の計量理論ではない。項目反応理論を用いる為には、理論モデルとデータとの間にモデルとデータの適合性が存在することが分析を進める前提となる為、全てのデータ分析に項目反応理論を用いることが出来るわけではない。本論では、モデルとデータの適合性の評価を必要としない古典的テスト理論を用いて尺度分析を行う。

2. 方法

本研究は、2012年に行われた調査1と2013年に行われた調査2の二つの社会調査の比較研究である。この2つの社会調査の全回答者の内、65歳以上であった204名(調査1)及び278名(調査2)を分析対象とした。両調査で回答者となった高齢者は、青森市の町内会などで活発に活動している高齢者を意図的に選択しているため、比較的健康で日常の活動レベルが高く、認知的に問題がないと考えられる個人であった。多くの調査協力者は、調査コーディネーターを通じて、事前に調査内容などについて知らされ、調査への協力の同意書にサインしていたが、実際の調査票への回答は、クラフトクラブや書道展覧会に出品するための活動集会及び町内会の運営会議などに参加した機会に行われた。

表1 幸福感尺度項目のワーディング

| 調査 1 | 調査 2 |
|--|--|
| Q11A: 人生を全体的に評価すると、自分は 恵まれていない と感じる | Q9A: 人生を全体的に評価すると、自分は 恵まれている と感じる |
| Q11J: 自分が不幸であると感じることがある | Q9J: 自分が不幸であると感じることがある |
| Q11B: 普段、幸せであると感じることは 少ない | Q9M: 普段、幸せであると感じることが 多い |
| Q11Y: 自分は「不幸な運命に生まれた人間だ」と感じる ことがある | Q9T: 自分は「不幸な運命に生まれた人間だ」と感じる ことがある |
| Q11C: 自分は、「生きていてよかった」と 感じることは少ない | Q9V: 自分は、「生きていてよかった」と 感じることが多い |

調査1ではQ11、調査2ではQ9が幸福感尺度として用いられた(表1)。二つの調査で共通に用いられた尺度は、幸福感尺度と生活の質尺度である(表2)。幸福感尺度には、表1に示されるように、調査1と調査2で異なるワーディング項目が用いられた(相違点は下線で示した)。各項目は、4肢選択のライカート型項目で、幸福感尺度項目の内容に「そう思う」と応答した場合4点、「ややそう

思う」と応答した場合3点、「あまりそう思わない」が2点、「そう思わない」は1点の得点が得られると設定し、総得点を算出した。生活の質尺度は、6つの下位尺度で構成され(総項目数は18項目)、全ての項目のワーディングは調査1と調査2で全く同一であった(表2)。

表2 生活の質尺度項目のワーディング

| |
|--|
| Q9B, Q11D: 私には、人並み以上の収入がある |
| Q9C, Q11E: 普段、自分は健康であると感じる |
| Q9D, Q11F: 私の家族の「きずな」は強い |
| Q9E, Q11G: 家族といるときには、心がなごむ |
| Q9F, Q11H: 友人と一緒に過ごす時間は楽しい |
| Q9G, Q11I: 私の自宅は美しい |
| Q9H, Q11K: 体調が悪く、日常生活に不便を感じる ことがある |
| Q9I, Q11L: 自宅には、自分が自由に使える 部屋がある |
| Q9K, Q11M: 私は、自分がしている仕事に 満足している |
| Q9N, Q11O: 自分のやりたい事にお金を十分 使うことができる |
| Q9O, Q11P: 私の友人関係は、自分の人生にとって大切な ものである |
| Q9P, Q11Q: 私は、自分と家族との関係に 満足している |
| Q9Q, Q11R: 私には親友と呼べる友人が いる |
| Q9R, Q11S: 私にとって仕事の 内容は楽しい |
| Q9S, Q11T: 私には、人並み以上の 財産(経済的な)がある |
| Q9U, Q11U: 私には、「自分のやりたい事」が できる体力がある |
| Q9W, Q11V: 私の自宅(部屋)には、 十分なスペース(広さ)がある |
| Q9X, Q11W: 私は、自分の仕事で 周囲の人に認められている |

項目ワーディングの影響の比較は、両研究で共通に用いられた5項目で構成された幸福感尺度及び生活の質尺度で行われた。幸福感尺度は、2つの研究で項目内容は同一であったが、表1に示されるように、ワーディングが5項目中3項目で異なっていた。調査1では幸福感尺度の全5項目が否定的表現であったのに対し、調査2では5項目

中 3 項目が肯定的な表現を用いた項目で、残りの 2 項目が否定的な表現を用いた項目であった。

古典的テスト理論を用いた分析として、①信頼性係数（クロンバックの α ）の比較、②因子分析を用いた項目の分散寄与率などの比較、③逆転項目を肯定的項目に合わせて逆転して算出した尺度の総合得点の分布比較、④総合得点と生活の質の下位尺度との相関マトリックスの評価などを行った。

本研究で用いた社会調査データの収集においては、身体的危険や人権を著しく侵害する危険のある実験や調査は行われなかったが、心理的外傷などを防ぎ、調査対象者の人権を尊重するため、個人情報保護法を遵守すると共に、全ての調査対象者に調査研究内容を説明し、調査への協力は任意であること及び何時いかなる理由においても、何ら不利益をこうむることなしに、自由に研究への協力を断ることが出来ることを説明した。また、上記の社会調査票の内容及び調査プロセスについては、青森大学社会学部の社会調査倫理規定委員会の承認を得た。

3. 結果と考察

3.1 回答者の特性

調査 1 と 2 における回答者の性別・平均年齢などの特性は、表 3 に示した。調査 2 における男性の割合が 43.9% (278 人中 122 人) で、調査 1 (204 人中 64 人) と比較して 12.5%高かった。既婚者の割合は、調査 2 (75.5%) において調査 1 (65.7%) と比較して若干高かった。両調査において、既婚者の割合は、男性が非常に高かった。調査 1 では、男性の既婚者が 85.9% (55 名)、調査 2 においては、91.0% (111 名) であった。教育レベルは、調査 1 で 23.5%が短大卒以上で、調査 2 と比較して若干高かった。教育レベルの特徴として、両調査において、女性の学歴が男性と比較して高い傾向が見られた；調査 1 においては、48 名の短大卒以上の学歴の回答者の内、30 名が女性で、調査 2 に

おいては、49 人の短大卒以上の内 31 名が女性であった。二つの調査間で回答者の特性には、若干の差はあるものの、極端に大きな差は見られなかった。本研究における回答者は、ランダムにサンプルされてはいなかったが、同一母集団からサンプルされている事、回答者の特性が大きく異なることから、項目のワーディングの影響を評価する比較分析が可能であると考えられる。

表 3 調査 1 及び 2 における回答者の特性

| 回答者の特性 | 調査 1 (n=204) | 調査 2 (n=278) |
|--------|--|---|
| 平均年齢 | 73.03 ($\sigma=6.01$) | 72.67 ($\sigma=5.94$) |
| 男性の割合 | 64 (31.4%) | 122 (43.9%) |
| 既婚者の割合 | 134 (65.7%) | 210 (75.5%) |
| 子供いる割合 | 191 (93.6%) | 259 (93.2%) |
| 教育レベル | 中卒 : 52 (25.5%) 高卒 : 94 (46.1%) 短大 : 48 (23.5%) 卒以上 | 中卒 : 66 (23.7%) 高卒 : 160 (57.6%) 短大 : 49 (17.6%) 卒以上 |

3.2 信頼性

古典的テスト理論を用いた信頼性分析では、全ての項目が否定的項目であった調査 1 の信頼性は .781 (α)、肯定的項目と否定的項目が混在する調査 2 の信頼性は .784 (α) であった。調査 2 の幸福感尺度は、肯定的項目と否定的項目の混在であったが、理論的には否定的項目を肯定的項目に合わせて逆転した場合であっても、その逆の場合であっても、クロンバックの α 係数は同値となる。調査 2 における項目の逆転においても同様に、肯定的・否定的項目のどちらの方向に項目得点を逆転させても α 係数は同値であった。肯定的項目と否定的項目とが混在した調査 2 の幸福感尺度の信頼性は、否定的フレーム項目のみで構築されている調査 1 の信頼性とほとんど差がなかった (.781 及び .784)。項目を除去した場合の α 係数は表 4 及び 5 に示した。両調査共に項目除去信頼性 (item deleted reliability) は、均一性が保たれており、肯定的項目と否定的項目が混在している調査 2 においても項目間の項目除去信頼性にほとんど差は見

られなかった。尺度の信頼性 (α) は健全な尺度構造の下では、項目除去信頼性と比較して高くなる。この点においては、両調査において、尺度の健全な信頼性構造が見られた。

表 4. 項目除去した場合の信頼性係数 (調査 1)

| Item | Item Mean and S.D. | | Item Deleted | |
|------|--------------------|----------|------------------------|----------------------|
| | Mean | σ | Item-Total Correlation | Coefficient α |
| Q11a | 1.711 | 0.865 | 0.502 | 0.763 |
| Q11b | 1.696 | 0.685 | 0.668 | 0.700 |
| Q11c | 1.534 | 0.732 | 0.635 | 0.717 |
| Q11j | 1.569 | 0.800 | 0.415 | 0.783 |
| Q11y | 1.451 | 0.711 | 0.584 | 0.731 |

表 5. 項目除去した場合の信頼性係数 (調査 2)

| Item | Item Mean and S.D. | | Item Deleted | |
|------|--------------------|----------|------------------------|----------------------|
| | Mean | σ | Item-Total Correlation | Coefficient α |
| Q9a | 2.152 | 0.760 | 0.482 | 0.767 |
| Q9j | 2.294 | 0.819 | 0.590 | 0.733 |
| Q9m | 2.152 | 0.789 | 0.580 | 0.737 |
| Q9t | 2.286 | 0.879 | 0.509 | 0.762 |
| Q9v | 2.327 | 0.785 | 0.645 | 0.715 |

3.3 因子構造

幸福感得点の因子分析においては、両調査データ共に明確な 1 因子構造を示したが、第一因子の分散寄与率は、調査 1 では 42.15% で、調査 2 では 41.14% であった。項目数が 5 項目と少ないにもかかわらず、両調査共に分散寄与率が 50% 未満であることは、項目内容が抽象的であることを考慮しても高い値ではない。全項目が否定的項目である調査 1 と項目のワーディングが混在している調査 2 の分散寄与率の差は 1.01% であった。この分散寄与率の差は、データ分析に実質的な影響を及ぼす程大きいものではなかった。各項目の第 1 因子への項目負荷量は、調査 1 では .61 から .78、調査 2 においては .55 から .73 であり、大きな相違はなかった。本研究で用いた尺度の項目数が 5 項目と少ない事、比較した 2 つの尺度は両尺度共に明確な 1 因子構造を示した事、各項目の第 1 因子への因子負荷量が似ていることから、分析結果として

因子構造に大きな相違が存在することを支持する証拠は見られなかった。項目ワーディングの混在が、尺度の妥当性などに及ぼす影響は、得点の平均、標準偏差、分布などを精査して評価する必要がある。

3.4 幸福感得点の分布

全項目が否定的表現である調査 1 では全項目を逆転し、肯定的表現と否定的表現が混在している調査 2 では否定的項目のみを逆転させて作成した幸福感得点の平均値は、調査 1 では 19.75 ($\sigma=4.00$)、調査 2 では 16.21 ($\sigma=2.96$) であり、調査 1 の方が有意に高かった ($p<.01$)。幸福感尺度の総点の差は 3.54 点で、統計学的に有意であるばかりでなく、他の変数との関連性の分析結果に実質的な影響が及ぶ可能性がある差であった。また、得点の分布には明確な相違が見られた (図 3、図 4)。調査 1 の幸福感得点の歪度は -1.11、調査 2 の歪度は -.867 となり、いずれも高い負の歪度を示した。

幸福感得点の分布は、2 つの調査間で明らかな相違があったが、最も大きな相違は幸福感得点の最頻値 (モード) の相違であった。否定的項目で構成された調査 1 の不幸福感尺度の 5 項目全てに「そう思わない」と応答して、幸福感得点としては最高得点の 20 点を得た回答者は全体の 25.0% である 51 人であった。一方、項目表現が混在していた調査 2 では、20 点満点を得た回答者は 13.4% の 36 名であったが、最高得点の 75% の得点である 15 点を獲得した回答者が最も多く、全体の 16.4% を占める 44 名であった (調査 1 では 46 名 (10.2%))。

幸福感得点の相違は、調査 1 と 2 における平均値に有意な差が存在したこと、幸福感の得点のモードに 2 つの調査間で大きな相違があったこと、尺度の因子構造に大きな相違が見られなかったことから、ネガティブ項目を逆転させ、総合得点を算出すると、ネガティブ項目に否定的に回答する傾向のため、尺度得点が高くなる傾向が存在したといえる。ネガティブ項目の影響は、得点平均に

影響を及ぼしたが、両調査における尺度の因子構造が非常に似ていることから、調査内における個人の相対的位置には有意な影響がないと予測されるため、尺度の妥当性を根本的に揺るがすものではないと考えられる。しかし、この影響は、重回帰分析や共分散構造分析などの高次の分析におけるパラメタ推計やモデルフィットなどに影響を及ぼす可能性がある。

3.5 相関構造に対する影響

相関構造は、幸福感尺度と生活の質の6つの下位尺度（家族関係、友人関係、収入の満足度、居住環境、健康、仕事）との相関マトリックスにおける相違は、他の関連変数との相関関係に強い影響を及ぼす可能性がある。最近の社会調査データの分析には、共分散構造モデリングなどの相関マ

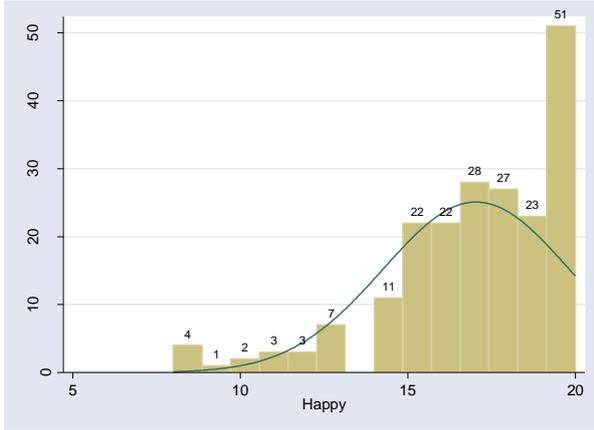


図3. 調査1の幸福感の得点分布

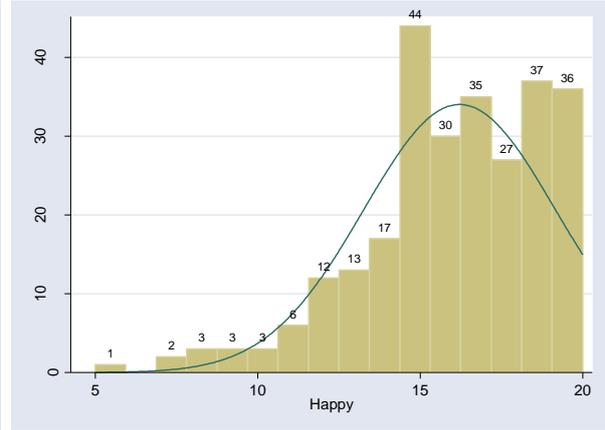


図4. 調査2の幸福感の得点分布

表6. 幸福感と生活の質の6下位尺度との相関係数（調査1）

| | Happy | Family | Friends | Income | Housing | Health | Work |
|---------|--------|--------|---------|--------|---------|--------|--------|
| Happy | 1.0000 | | | | | | |
| Family | 0.3988 | 1.0000 | | | | | |
| Friends | 0.4084 | 0.6785 | 1.0000 | | | | |
| Income | 0.5449 | 0.3980 | 0.4552 | 1.0000 | | | |
| Housing | 0.4295 | 0.2307 | 0.4804 | 0.3188 | 1.0000 | | |
| Health | 0.3869 | 0.3957 | 0.4106 | 0.3861 | 0.2787 | 1.0000 | |
| Work | 0.4124 | 0.2416 | 0.3030 | 0.4941 | 0.4102 | 0.2884 | 1.0000 |

表7. 幸福感と生活の質の6下位尺度との相関係数（調査2）

| | Happy | Family | Friends | Income | Housing | Health | Work |
|---------|--------|--------|---------|--------|---------|--------|--------|
| Happy | 1.0000 | | | | | | |
| Family | 0.5212 | 1.0000 | | | | | |
| Friends | 0.3376 | 0.2579 | 1.0000 | | | | |
| Income | 0.3149 | 0.2172 | 0.1130 | 1.0000 | | | |
| Housing | 0.3046 | 0.2775 | 0.1541 | 0.2627 | 1.0000 | | |
| Health | 0.4069 | 0.2252 | 0.1235 | 0.1222 | 0.1184 | 1.0000 | |
| Work | 0.5967 | 0.4934 | 0.3442 | 0.3043 | 0.4173 | 0.2695 | 1.0000 |

表 8. 幸福感と生活の質との相関係数の差 (調査 2 - 調査 1)

| | Happy | Family | Friends | Income | Housing | Health | Work |
|---------|----------------|----------------|----------------|----------------|----------------|----------------|--------|
| Happy | 0.0000 | | | | | | |
| Family | 0.1224 | 0.0000 | | | | | |
| Friends | <u>-0.0708</u> | <u>-0.4206</u> | 0.0000 | | | | |
| Income | <u>-0.2300</u> | <u>-0.1808</u> | <u>-0.3422</u> | 0.0000 | | | |
| Housing | <u>-0.1249</u> | <u>-0.0468</u> | <u>-0.3263</u> | <u>-0.0554</u> | 0.0000 | | |
| Health | 0.0200 | <u>-0.1705</u> | <u>-0.2871</u> | <u>-0.2639</u> | <u>-0.1603</u> | 1.0000 | |
| Work | 0.1843 | 0.2518 | <u>-0.0412</u> | <u>-0.1898</u> | 0.0071 | <u>-0.0188</u> | 1.0000 |

トリックスをデータとした二次的分析が頻繁に行われるようになってきている為、相関マトリックスに対する項目ワーディングの影響を評価した。幸福感尺度と生活の質の 6 つの下位尺度との相関マトリックスは、表 6 及び 7 に示した。表 8 は調査 2 の相関係数から調査 1 の相関係数を減じた差である。調査 1 における相関係数が調査 2 における相関係数と比較して高かったケース (下線及び赤字表示) は、幸福感と生活の質尺度の 6 つの下位尺度との相関係数値中 (6 相関係数)、3 ケースあり、その中の 1 ケースにおいて、相関係数の差が 0.20 以上であった。一方、調査 2 の相関係数が高かったケースも 3 ケースで、1 ケースが 0.20 以上であった。また、表 6 及び 7 に示された全ての相関係数は、有意であった ($H_0: r = 0$)。幸福感と生活の質尺度の下位尺度との相関係数の差を検定したところ、6 ペアの相関係数中、1 ペア (幸福感と収入) の差の出現確率が 5% 水準 ($p = .0127$: Bonferroni の多重比較調整を採用) で有意であったが、残りの 5 ペアの差には、有意性は確認されなかった。これらの分析結果から、調査 1 と調査 2 における幸福感と生活の質尺度の下位尺度との相関係数値の相違は大きなものではなかった。

確証的因子分析などの多変量解析を用いたデータ分析では、このような相関マトリックス同士の差の比較には、指定されたモデルで説明

される共分散部分を除去した残差部分の総和が 0 であるとする帰無仮説を χ^2 検定で検定する方法が一般的である。本論では、より単純に相関変数が及ぼす影響を明らかにするために、重回帰分析を例として相関マトリックスにおける係数の相違が重回帰係数に及ぼす影響を評価した。

幸福感を従属変数とし、収入と仕事関係を独立変数とした重回帰方程式 ((1) 参照) を用いて、収入と仕事関係が幸福感に及ぼす影響を調査 1 と 2 で比較した。独立変数として選択された収入は、両調査間で相関係数に唯一有意差が観察されている変数であり、仕事関係は、両調査で幸福感と高い相関を示した変数であった。ここで、 \hat{y} は幸福感の予測値、 X_1 は家族関係得点、 β_1 は家族関係の標準偏回帰係数、 X_2 は収入、 β_2 は収入の標準偏回帰係数、 a_0 は切片とする。

$$\hat{y} = \beta_1 X_1 + \beta_2 X_2 + a_0 \quad (1)$$

調査 1 のデータ分析結果は、重回帰方程式全体の予測効率を示す R^2 (決定係数; Adjusted R-squared) は .32、収入の標準偏回帰係数 (β_1) が .451 ($p = .000$)、仕事関連の標準偏回帰係数 (β_2) は .189 であった。調査 2 のデータの分析結果は、 R^2 が .32、収入の標準偏回帰係数 (β_1) が .154 ($p = .003$)、仕事関連の標準偏回帰係数 (β_2)

は.543 ($p=.000$)であった。独立変数群が回帰方程式モデルで、説明する従属変数の割合は、調査 1 及び調査 2 とともに 32%で、全く同値であった。両調査において、全ての回帰係数及び決定係数は、有意であった。

標準偏回帰係数は、収入においてはその差が.297 で比較的大きかった。この差は、相関係数が有意に異なることから、論理的に予測されることである。しかし、仕事関係の標準偏回帰係数の差は、相関係数レベルでは有意差が存在しなかったにもかかわらず、有意差が存在した収入と比較して更に大きい差 (.354) が得られた。標準偏回帰係数は相関マトリックスで示される相関係数から (2) 及び (3) の式で単純に計算することができる。

$$\beta_1 = \frac{r_{y1} - r_{y2} \cdot r_{12}}{1 - (r_{12})^2} \quad (2)$$

$$\beta_2 = \frac{r_{y2} - r_{y1} \cdot r_{12}}{1 - (r_{12})^2} \quad (3)$$

相関係数に有意な差がないにもかかわらず、回帰分析における標準偏回帰係数において、大きな相違が算出される現象は、独立変数同士の相関レベルで説明できる。相関係数の比較では、幸福感と生活の質の下位尺度 (6 下位尺度) との間の比較であったが、上記の方程式に示されるように、標準偏回帰係数の算出には、重回帰分析に用いられた独立変数同士の相関 (r_{12}) が必要となる。調査 1 における収入と仕事関係の相関 (r_{12}) は.4941 で、調査 2 においては.3043 であり、その差は有意であった ($p=.01$)。上記の式 (1) 及び (2) に基づき、 r_{y1} 及び r_{y2} が一定であると仮定すると、標準偏回帰係数は、独立変数同士の相関 (r_{12}) が 0 から 0.5 に向かって高くなると共に低くなり、 r_{12} が 0.5 から 1.0 に

向かって増加すると共に高くなる。従って、幸福感尺度に用いられた項目のワーディングの影響は、重回帰分析などの高次分析においては、独立変数として採用される変数との相関係数を通して当該尺度以外の分析に波及すると考えられる。

4. 結論

項目のワーディングが及ぼす影響に関する先行研究では、肯定的な項目で構成される尺度の中に小数の否定的な項目が混在すると、否定的な項目の信頼性が低下する傾向があること (O'Muicheartaigh et al., 2007)、否定的項目と肯定的項目を混在させると尺度の構成概念妥当性が低下すること (Chen et al., 2007) などが報告されている。本研究においては、幸福感尺度 5 項目中、3 項目のワーディングを変えたところ、クロンバックの α 係数 (内的整合性指標) に大きな相違は見られなかった。尺度項目数が多く (10 項目以上)、ワーディングの相違が多数項目に及ぶケースでは信頼性に大きな相違が生起する可能性があるが、本研究においては、先行研究で報告されたような相違は見られなかった。

信頼性の評価は、クロンバックの α を用いたが、正解が存在する学業試験のように項目間の内的整合性が高い事が前提となっている項目とは性質が異なり、抽象性が高い尺度項目の評価には試験 - 再試験信頼度 (test-retest reliability) を用いることも考えられたが、現実的に 2 回にわたって同一回答者に本研究のような調査を行うことは困難であった。Kano & Azuma (2003) は、SEM (structural equation modeling) を用いて正確に信頼性を測定する方法を紹介しているが、その報告によると α 係数は正確な信頼性より低い値を示す傾向にあるとしている。項目表現が混在していた調査 2 の信頼性

は.784 で、項目数が 5 項目であることを考慮すると低い値ではないが、正確に推計すれば報告した値より高い可能性がある。

調査 1 と調査 2 の分析結果は、項目のワーディングの相違によって幸福感尺度の妥当性を損なうとする論拠を提供するまでは至らなかった。しかし、本研究で用いた尺度の項目数は、Chen et al. (2007) の研究で用いられた尺度の項目数 (30 項目) の 6 分の 1 の項目数であること、更には項目の内容が抽象的であることなど、分析の背景の相違が影響している可能性がある。一般的な社会調査においては、30 項目にも及ぶ多数の項目で構成される尺度が使用されることは少なく、多くの場合、用いられる尺度は 5 項目程度である。本研究の分析結果によると、項目のワーディングを変化させることが、尺度の妥当性が直ちに大きな問題となるとは考えにくい。Liez (2010) は、項目の肯定性・否定性自体が因子負荷量の増減の原因となるため、尺度を構成する因子構造に重要な影響を及ぼすと報告しているが、項目数が限定された状況における否定的な項目の取り扱いについては、更に実証的データを蓄積する必要がある。

幸福感得点の分布については、両調査で負の歪度を示し、同一の分布形状を示したが、最頻値の相違が得点分布の大きな相違となった (図 3 及び 4)。幸福感とその関連変数との相関マトリックスの構築を考えると、調査 1 のように最も極端な値 (最高値である 20 点) がもっとも頻度が高い得点は、他の変数との相関係数は、相関係数が外れ値の影響を強く受けることから、特徴的な傾向を示すと考えられる。一方、肯定的項目と否定的項目が混在していた尺度 (調査 2) では、最頻値は 15 点であったため、外れ値の影響は調査 1 ほど強くはなかった。その結果、調査 1 では、幸福感が最高値であった

回答者が多くいたが、その回答者が必ずしも他の変数で高い値を示さなかったため、相関マトリックスは調査 2 ほど高い相関係数で構成されなかった。

調査 1 において、最頻値が最高点の 20 点であった大きな要因は、項目の性質にあった可能性がある。回答者が「あなたは不幸ですか」と尋ねられた時、多くの回答者は「あまりそう思わない」よりは「そう思わない」と応答しやすい。逆に、「あなたは幸せですか」と尋ねられた場合、「そう思う」と応答するより「ややそう思う」と回答しやすい。極端な否定的項目には、そうではないと強く反応し、極端な肯定的項目には、あまり強く反応しない傾向があると考えられる。これは、遠慮の文化のような国民性によるところがあるとの報告もある (前田, 1995)。否定的な項目には遠慮なく「自分は不幸ではない」と応答するが、肯定的な項目には遠慮して上から 2 番目の選択肢「自分はやや幸福である」に応答する。否定的な項目は逆転されて、最も極端な肯定的選択肢に応答したものとして得点が算出される。その結果、否定的な項目で構築された尺度得点の最頻値は理論的 maximum となる。本調査の分析結果から、明確にその傾向が観察された。

本研究の分析は、先行研究の知見と同様に否定的項目を逆転することは望ましくないことを示す結果であった。しかし、その理由は尺度の因子構造や信頼性に起因するものではなく、逆転項目の得点分布によって起こされると考えられる、関連変数との相関マトリックスへの影響である。また、社会調査の分析に頻繁に用いられている、重回帰分析、因子分析及び共分散構造分析は、相関マトリックスを基に推計されることから、尺度に用いられる項目のワーディングが及ぼす影響は、その尺度に留まらず、その尺度と関連する変数との相関係数に影響

し、社会調査の目的に関わるモデルの検証になどにも影響を及ぼす可能性がある。本研究においては、明確な相関構造の相違は確認されなかったが、この分析結果が、直ちに項目のワーディングが重回帰分析や共分散構造分析に影響を及ぼさないことを示すものではない論拠が示された。

項目数が少ない尺度を用いることが多い社会調査の実情に鑑み、少ない項目で構築される尺度における項目ワーディングの影響や、通常の日本語表現で否定的な言い回しをする表現を調査項目にする場合、肯定的に変換すべきか、などは更に研究が必要である。また、項目ワーディングが混在している調査票のデータ分析において、外れ値のスクリーニングが相関マトリックスの安定性に及ぼす影響、なども今後研究されるべきトピックである。

References

- Alexandrov, A. (2010). Characteristics of single-item measures in Likert scale format, *The electric journal of business research methods*, 8(1), 1-12, available online at ejbrm.com.
- Anastasi, A. & Urbina, S. (1997). *Psychological testing*, Upper Saddle River,
- Biemer, P.P. & Lyberg, L. E. (2003). *Introduction to Survey quality*, Hoboken, Wiley
- Chen, Y., Rendina-Gobioff, G., Dedrick, R. F., (2007). Detecting effects of positively worded items on a self-concept scale for third and sixth grade elementary students, *Paper presented at the annual meeting of the Florida Educational Research Association*, Tampa, FL. Nov.14-17.
- Dillman, D.A. (1978). *Mail and telephone surveys: the total design method*, Wiley & Sons.
- Embretson, S. E., and Reise, S.P. (2000). *Item Response Theory for Psychologists*, New Jersey, Lawrence Erlbaum Associates, Publishers.
- Hambleton, R.K., Swaminathan, H., and Rogers, H.J. (1991). *Foundation of Item Response Theory*. Newbury Park, CA.: Sage Publications.
- Kano, Y. & Azuma, Y. (2003). Use of SEM program to precisely measure scale reliability, *New development in psychometrics*, 141-148, Tokyo; Springer.
- Liez, P. P. (2010). Factor analysis and negatively worded items, *International journal of market research*, 52(2), 249-272.
- 前田忠彦 (1995). 日本人の満足感の構造とその規定因に関する因果モデル - 共分散構造分析の「日本人の国民性調査」への適応, *統計数理*, 43(1), 141 - 160.
- Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory (3rd ed.)*, New York, NY: McGraw-Hill.
- O'Muircheartaigh, C., Krosnick, J., & Helick, A. (2000). Middle alternatives, acquiescence, and the quality of questionnaires, *unpublished manuscript, available at harrisschool.uchicago.edu/about/publications/working-papers/pdf/wp_01_3.pdf*.
- Rose, G., Beale, I., Malone, J., Kinkead, S., and Higgin, J. (2011). Validity of a scale for consumer evaluation of mental health service delivery, *The international journal of person centered medicine*, 1(4), 733-740.
- Watanabe, S. & Shibutani, H. (2010). Aging and decision making: Differences in susceptibility to the risky-choice framing effect between older and younger adults in Japan, *Japanese Psychological Research*, 52, 3,163-174.
- Spector, P. E., (1992). *Summated rating scale construction; introduction*, Sage University paper series. Quantitative applications in the social sciences No.82, Newbury Park, CA: SAGE publications, Inc.
- Tversky, A., and Kahneman, D. (1981). The framing of decisions and the psychology of choice, *Science*, 211, 454-458.
- von Neumann, J., & Morgenstern, O. (1947). *Theory and games and economic behavior (2nd ed.)*, Princeton, NJ: Princeton University Press.

EVALUATING THE INFLUENCE OF POSITIVELY AND NEGATIVELY WORDED ITEMS ON DATA ANALYSES IN SOCIAL SURVEYS: ANALYSING DATA WITH CLASSICAL TEST THEORY

Hirohide SHIBUTANI¹ Satoshi WATANABE²
Harumasa YOSHIMURA³ Atsushi KOKUBO⁴

¹ Faculty of Sociology, Aomori University

² Research and Education Center for Comprehensive Science, Akita Prefectural University

³ Faculty of Sociology, University of Nara

⁴ Faculty of Software and Information Technology, Aomori University

要旨

本研究の目的は、社会調査における項目のワーディングの相違がデータの分析に及ぼす影響を評価することである。2011年と2013年に行われた調査1及び調査2において用いられた、幸福感尺度と生活の質尺度を対象に比較が行われた。幸福感尺度は5項目で構成され、5項目全てが否定的表現で構築されていた調査1と3項目が肯定的で残り2項目が否定的な項目で構成された尺度が用いられた調査2とを比較した。生活の質尺度は6つの下位尺度で構成されており、両調査でワーディングは同一であった。幸福感尺度の信頼性は、両調査でほぼ同一であったが、調査1得点平均は、調査2と比較して有意に高かった。この平均値の有意差には、項目の逆転操作が影響している可能性が指摘された。また、得点分布は、両調査において負の歪度を示したが、平均値が高かった調査1における最頻値は最高点の20点で、調査2の最頻値（15点）とは大きく異なっていた。幸福感尺度の因子構造には、両調査において大きな相違は見られなかった。項目ワーディングの影響は、重回帰分析や共分散構造分析などの高次分析において用いられる独立変数間の関係マトリックスを介して、当該尺度以外にも波及するメカニズムが示唆された。

キーワード：項目ワーディング 肯定的項目
否定的項目、相関マトリックス
共分散構造分析