

国際比較調査におけるレスポンス・スケールの等価性に関する研究(1)*

真 鍋 一 史**

1. はじめに

日本は、質問紙調査 (questionnaire survey) あるいは世論調査 (public opinion poll) については、いわゆる先進国であり、毎年おびただしい数の調査がさまざまな機関によって実施されている。これは、毎年、総理府がまとめて発表している『世論調査年鑑』を参照することで一目瞭然となる (Nishihira, 1983)。

このような質問紙調査あるいは世論調査の最近の新しい傾向の一つとして、同時に多くの異なる国 (地域) を対象とする大規模な国際比較調査 (large scale multi-national comparative surveys) の実施ということがあげられる。そして、その典型的な事例としてしばしば言及されるのが、米国ミシガン大学 Survey Research Center/ Center for Political Studies の Ronald Inglehart 教授が主宰する「世界価値観調査 (World Values Survey=WVS)」と、ドイツ・マンハイムの「世論調査の方法と分析のための研究センター (Zentrum für Umfragen, Methoden und Analysen=ZUMA)」と米国シカゴ大学の「全国世論調査研究センター (National Opinion Research Center=NORC)」が中心になり、そこに各国の大学・研究所・調査機関が加わり、設立された「国際社会調査プログラム (International Social Survey Programme=ISSP)」の二つである。

日本との関わり合いについては、まず、前者については、R. Inglehart 教授からの呼びかけに応じて (株)電通総研と (株)電通総研が共同で

1990年度の調査から参画している。筆者は1995年以来、共同研究の形で、この「世界価値観調査データ」の、①共同利用のための準備作業と、②二次的分析の試み、を進めてきている (真鍋ほか、1996、1997、1997、1999)。つぎに、後者については、NHK (放送文化研究所・世論調査部) が1992年度から正式のメンバーとして加盟しており、筆者はNHKからの依頼で1997年度から日本代表の一人として総会・質問紙原案作成委員会・研究発表会に参加するとともに、日本と各国との比較に焦点を合わせたデータ解析の作業を行ってきている (真鍋ほか、1997、1999、真鍋、1999、1999)。

さて、このような国際比較調査においては、比較のための測定 (measurement) の等価性 (equivalence) の検討がきわめて重要な課題となってくる。もちろん、測定の等価性という問題は、何も国際比較調査に固有な問題というわけではない。Kaplan (1964) が指摘しているように、測定というものの本質は、「標準化の原則 (the principle of standardization)」つまり「測定の単位は時間と空間を越えて一定でなければならない」という原則にある。このような原則を立てておかなければ、そもそも比較ということは成り立たない。質問紙法 (questionnaire method) にもとづく調査では、質問項目とその回答の選択肢をどの回答者にとっても同じ意味になるように作成するということが重要な課題となる。しかし、実際の調査においては、それぞれの回答者がそれらをまったく違った意味で受け取ってしまうということもすでに広く知られている。国際比較調査の場合には、このような比較ということについての根源的な問題に加えて、さらに Almond と

*キーワード：国際比較調査、レスポンス・スケール、等価性

** 関西学院大学社会学部教授

Verba (1963) の言葉を借りるならば、一つの国を扱っているときには無視できるたくさんの問題に直面しなければならなくなる。この点については、筆者(真鍋、1999)は、すでにそれらを、①変数——「被調査者の個人的属性や社会的環境に関する変数」と「被調査者の意見・態度・行動に関する変数」——の等価性の問題、②社会の変化の影響の問題、③翻訳の問題、④データ解析の問題、に分けて、それぞれについて詳細な検討を加えた。

たしかに、国際比較調査においては、測定の等価性をめぐって、さまざまな問題が検討されなければならない。しかし、それらのなかでも、調査の質問項目とその回答の選択肢の「翻訳」こそが最大の問題であるということに異論を唱える者はいないであろう。こうして、国際比較調査における翻訳の問題に関しては、さまざまな角度から研究が進められてきている。この領域の文献レビューについては、Janet A. Harkness ed., *Cross-Cultural Survey Equivalence*, ZUMA NACHRICHTEN SPEZIAL, No. 3, 1998 を参照されたいが、そこで紹介されている MINTS プロジェクト (*Research into Methodology of Intercultural Surveys*) は、この領域の新しい試みとして注目される。本稿は、この文献にもとづいて、MINTS プロジェクトにおけるレスポンス・スケールをめぐる議論を整理するとともに、この研究プロジェクトへの日本の参加の意義を提案するものである。

さて、質問紙調査の諸項目は、さまざまな基準によってさまざまに分類されてきた。筆者(真鍋、1993)は、つぎのような分類の仕方を採用している。それは、まずそこで使用されている質問諸項目が *open-ended questions* (質問に対する決まった回答形式を定めず、被調査者に自由な表現形式で回答を求める形式) であるか、それとも *closed-ended questions* (質問に対してあらかじめ選択肢を用意して、そのなかから被調査者に回答を選ばせる形式) であるかを検討し、それらが後者である場合は、さらにそれらが *cafeteria questions* (次元の異なる選択肢をいくつか設ける方式) であるか、それとも *range questions* (同じ次元の異なるレベルの選択肢をいくつかを設ける方式)

であるかを検討するというものである。MINTS プロジェクトが、国際比較の可能性と問題性の検討のために取りあげるのは、まさにこのような *range question* である。そして、このような回答の選択肢のそれぞれのセットがレスポンス・スケール (*response scale*: 回答の尺度) と呼ばれるものである。

2. 調査におけるレスポンス・スケールの位置とその種類

MINTS プロジェクトは、ドイツ・マンハイムの ZUMA と米国シカゴ大学の NORC の共同研究として進められている「レスポンス・スケールの翻訳の可能性と問題性を探究するための実験的な試み」である。

調査の質問項目は、一般に、つぎのような部分から構成される。

- ①導入の部分。
- ②質問の部分。
- ③レスポンス・スケールと、たとえば「それぞれの枠内にレ印をつけてください」というような指示を与える部分。

そして、国際比較という視点を採らない調査については、上述の①②③のすべての部分について、すでにながりの研究の蓄積がある。たとえば、以下のような文献をあげることができる。

- ・質問への導入に関する研究 (Cannel et al., 1979; Schumann and Presser, 1981; Converse and Presser, 1994)。
- ・質問のワーディングに関する研究 (Hippler et al., 1987; Bradburn and Sudman, 1991; Converse and Presser, 1994; Sudman et al., 1996; Schwarz, 1996)。
- ・質問の長さに関する研究 (Payne, 1951; Cannel et al., 1979; Converse and Presser, 1994)。
- ・質問の順序に関する研究 (Schumann and Presser, 1981; Hippler et al., 1987; Converse and Presser, 1994; Wanke and Schwarz, 1997; Sudman et al. 1996)。

- ・レスポンス・スケールのデザインに関する研究 (Schumann and Presser, 1981; Presser and Schumann, 1980; Converse and Presser, 1994; Schwarz, 1996; Krosnick and Fabrigar, 1997)。
- ・レスポンス・スケールと質問項目との関連に関する研究 (Hippler et al., 1987; Schwarz et al., 1991; Schwarz and Hippler, 1991; Schwarz, 1996)。

ところが、国際比較調査については、そのような方法論的な検討はいまようやく始まったばかりである。では、その方法論的な検討の戦略としては、どこから手をつけるのが望ましいのであろうか。MINTS プロジェクトが、まず取りあげたのは、レスポンス・スケールである。そして、それも ISSP の英語とドイツ語の表現形式の検討から始めている。そこでは、一方で ISSP で最も普通に使われているレスポンス・スケールの表現形式、たとえば「そう思う／そう思わない」「賛成／反対」「重要である／重要でない」などが取りあげられるとともに (Davis, 1993)、他方で ISSP では用いられていないレスポンス・スケールの表現形式にも検討が加えられている (Smith, 1997)。

いうまでもなく、現在、質問紙調査においては、じつにさまざまなテーマ (トピックス) が取りあげられている。それは、B. Berelson (1970) の「欲しささえすればどのような種類の意見や行動に関するデータでも (たとえば、態度、価値、道徳様式から、育児法、宗教活動、家族行事、性行動、余暇活動、さらには生産意欲、幸福感まで) 質問紙調査という手段を用いて体系的に集めることが可能になった」という指摘に端的に表現されている。こうして、それぞれの質問紙調査ごとに、そこで扱われるテーマ (トピックス) はじつにさまざまであるということになる。ところが、そのようなテーマ (トピックス) の多様性とは裏腹に、それぞれの質問に対する回答の選択肢の部分——つまりレスポンス・スケールの部分——については、ある種の「標準化の方向」ともいうべき傾向が見られる。そして、このようにレスポンス・スケールについて、いったん「標準的なフォーマット

(standard format)」が確立されたならば、その後は、その同じフォーマットのスケールが繰り返し用いられることになるのである。

Davis (1993) は、1985年から1993年までの ISSP の質問諸項目を詳細にレビューしている。因みに、各年度ごとの調査テーマはつぎのとおりである。

1985年	「政府の役割 (第一回)」
1986年	「社会的ネットワークとサポート・システム (第一回)」
1987年	「社会的不平等 (第一回)」
1988年	「家族と性役割 (第一回)」
1989年	「仕事と余暇 (第一回)」
1990年	「政府の役割 (第二回)」
1991年	「宗教 (第一回)」
1992年	「社会的不平等 (第二回)」
1993年	「環境問題 (第一回)」

Davis によれば、このような七つの調査テーマについての九回の調査——「政府の役割」と「社会的不平等」についてはすでに二回目のいわゆる「反復調査」が実施されたので——において、それぞれ「そう思う／そう思わない」というフォーマットのスケールは92回、「重要である／重要でない」というフォーマットのスケールは23回、「認める／認めない」というフォーマットのスケールは22回、「賛成／反対」というフォーマットのスケールは11回、使用されているという。また米国シカゴ大学の NORC が毎年実施している「総合的社会調査 (GSS)」においても、ほぼ毎年同じフォーマットのスケールが繰り返し使用されている (Davis and Smith, 1992)。

以上から、MINTS プロジェクトでは、国際比較調査に関する方法論的な研究の出発点として、調査のレスポンス・スケールに焦点が当てられることになったのである。

3. アグリーメント・スケールの多様性とその影響

世界の国ぐにで実施されている主要な調査においては、必ずといっていいくらい「そう思う／そう思わない」というレスポンス・スケールが用い

られている。ここでは、このようなスケールを「アグリーメント (agreement : 同意) ・スケール」と呼ぶことにする。基本的には、それは「そう思う／そう思わない」というフォーマットで提示されるが、詳細に検討するならば、このスケールについても、いくつかのバリエーションが区別される。そして、ある特定の調査ではある特定のフォーマットが用いられ、また別の調査では別のフォーマットが使われるという傾向が見られるとともに、ここでもいったんある調査であるフォーマットが採用されると、しばらくはその同じフォーマットが継続して使用されるという傾向があることがわかる。

その具体的な例として、英語版の調査票からつぎのようないくつかのバリエーションがあげられる。

①回答者に対して読みあげる選択肢が「そう思う」あるいは「そう思わない」の二者択一の「強制選択 (forced choice) 式」のレスポンス・スケール。

Agree
Disagree
Don't know
No answer
Not applicable

Source: American General Social Survey (GSS), Cumulated Codebook, Q. 357 a, 1972 - 1993.

②4点スケールの「強制選択」式のレスポンス・スケール。つまり「そう思う」と「そう思わない」がそれぞれ「非常に」と「やや」に分けられ、しかし「どちらともいえない」という「中間の選択肢 (middle option)」は設定しない形式のスケール。

Strongly agree
Agree a little
Disagree a little
Strongly disagree
DK
NA

Source: British Social Attitudes (BSA), Cumulated Sourcebook. K-15 (1987/1989).

③7点スケールあるいは5点スケールの形式。ここでは「中間点 (middle point)」が設定されるとともに、「そう思う」と「そう思わない」がそれぞれその程度に応じて2段階あるいは3段階に分けられ、加えてつぎのイギリスのケースではその程度を表わす修飾語 (modifier) の位置が②のケースとくらべて逆 (後) になっている (その部分をイタリックにしている)。

<i>Agree strongly</i>
Agree
Neither agree nor disagree
Disagree
<i>Disagree strongly</i>
DK
NA

Source: British Social Attitudes (BSA), Cumulated Sourcebook. K-15 (1987/1989).

④「標準的な (standard)」ISSP のフォーマット (修飾語はイタリックにしている)。

<i>Strongly agree</i>
Agree
Neither agree nor disagree
Disagree
<i>Strongly disagree</i>
Can't choose, Don't know
NA, Refused

Source: ISSP 1993-GSS (USA) Q 542 A.

⑤標準的なISSPのレスポンス・スケールのオーストラリア版。ISSPとの相違点は、オーストラリア調査が「郵送調査 (mail survey)」で行なわれているところから、それがより回答者の目を引くように工夫されているということであろう。

To begin with we have some questions about (topic). Do you agree or disagree... (topic)

Yes !!	Strongly agree
Yes	Agree
??	Neither agree nor disagree
No	No Disagree
No!!	No!! Strongly
—	(Can't choose)

Please circle a ward

a. text first item	Yes!!	Yes	??	No	No!!	—
b. text second item	Yes!!	Yes	??	No	No!!	—
c. text third item	Yes!!	Yes	??	No	No!!	—
d. text fourth item	Yes!!	Yes	??	No	No!!	—
Source ISSP ! 988Australis Q 1 .	Yes!!	Yes	??	No	No!!	—

さて、レスポンス・スケールのフォーマットにおける以上にあげたような差異は、被調査者の回答の傾向にどのような影響を与えるであろうか。この点については、認知科学的な調査方法論の研究の系譜が注目される。そこでは、以下のような諸知見が示されている（文献レビューについては Schwarz, 1996を参照されたい）。

- ・ 回答者は調査の質問文を理解するためにレスポンス・スケールを利用し、また逆にレスポンス・スケールを理解するために質問文を利用する。
- ・ 「同じ」内容の質問に対する回答の集合的な分布 (distribution)——それぞれの選択肢を選ぶ回答者の%——が、そこに提示されるレスポンス・スケールの性質 (characteristics) によって異なるものとなる。
- ・ レスポンス・スケールのそれぞれの選択肢に言語的なラベル (verbal labels) や数値的なラベル (numeric labels) が付けられているかどうかということが、それぞれがどのようなラベルであるかということとともに、被調査者の回答の傾向に影響を与える。

こうして、調査で用いられるレスポンス・スケールの等価性の問題とともに、そうしたレスポ

ンス・スケールのデザインのもたらす影響という問題は、国際比較調査においてさらに増幅されたものとなる。それは、一つにはそのようなレスポンス・スケールがほかの言語へ「翻訳」されなければならないからである (Harkness, 1993, 1996; Van de Vijver and Leung, 1997)。

4. レスポンス・スケールの強度の測定

レスポンス・スケールの研究における最初の目標は、「そう思う／そう思わない」「賛成／反対」「重要である／重要でない」などのそれぞれのレスポンス・スケールのいくつかの選択肢、つまりそれぞれの次元におけるいくつかの「程度 (degree)」の段階をどのような表現 (expression) 形式で表わすかということである。国際比較調査の場合はしばらく措くとして、単一文化における (mono-cultural context) 調査の場合は、すでにそれらレスポンス・スケールの「強度 (strength)」を測定するアプローチが開発されてきている。それは、ある特定のレスポンス・スケールにおけるそれぞれの段階を表現する用語 (terms) について、その強度を回答者に判断させるというものである。このアプローチにはつぎの三つの標準的な方法がある。

- ①それぞれの用語を、「より弱い」から「より

強い」へ、あるいは「より少ない」から「より多い」へとといった一つの連続体 (continuum) の上にランクづけるといふ「序列法」である (Spector, 1976)。いうまでもなく、この方法ではそれぞれの用語の相対的な位置が示されるにとどまり、それら用語間の絶対的な強度あるいは距離が知られるわけではない。

②それぞれの用語を、普通は10点から20点といった一つの数値的なスケール (numerical scale) の上に評定していくという「評定法」である (Wildt and Mazis, 1978; Worcester and Burns, 1975; Myers and Warner, 1968; Cliff, 1959; Jones and Thurstone, 1955; Mittelstadt, 1971)。この方法によるならば、それら用語間の絶対的な強度あるいは距離を知ることができ、こうして「等現間隔尺度 (equal-appearing interval scale=Thurstone scale)」の作成が可能となる。この方法では、「意味差判別尺度 (semantic differential scale)」(Osgood et al., 1957) がまさにそうであるが、アルファベットのスケール、あるいはラベル付けられていないスペース、棧、箱などのスケールを用いることもできる。文字やスペースが数値に変換 (transform) されるからである。

③それぞれの用語を「比例 (率) 尺度 (ratio scale)」の上に位置づける「大量測定法」が開発されてきた (Wegener, 1991; Hougland et al., 1992)。この技法では、調査者(ときには回答者)が特定の準拠となる用語 (a reference term) にある任意の数値を与え、回答者にその基本的な用語との比例の関係でほかの用語を評定してもらうが、普通はその評定は、たとえば線でいえば、その数と長さというような二つのモードによってなされる。このようにして作成されたスケールでは、人びとの集合体の反応を全体として測定することもできるし、また個々人の反応を個別に測定することもできる。このスケールは、それぞれの用語の評定を限られた点数のなかから判断しなければならない数値的なスケールの場合にくらべて、よりの確かな評定を可能にするものといわれる。

これら三つの方法のなかで、MINTS プロジェクトにとっては、②の「評定法 (rating ap-

proach)」が最も適切なものと判断された。それは、一方で①の「序列法 (ranking method)」では、さまざまな用語を言語の違いを越えて (たとえば英語とドイツ語というように) 対応させるために必要な数値的な正確さといったことを確保することができないからであり、また他方で③の「大量測定法 (magnitude measurement technique)」はその取扱いがあまりにも面倒であり、また一般の人びとが簡単に回答できるようなものでもなく、さらにその技法が達成しようとするいわば超精密性 (extra precision) のようなものは MINTS プロジェクトにはとくに必要ないものといえるからである。

さて、「評定法」は、これまでさまざまな用語を評定するために用いられてきた。MINTS プロジェクトでとくに関心が示されたのは、一般的な「よい—わるい」あるいは「肯定 (positive) —否定 (negative)」といった次元でさまざまな用語を評定する技法と、それらの次元の修飾語 (modifiers) の強度を評定する技法である (Worcester and Burns, 1975; Wildt and Mazis, 1978)。これらと同じ線上で、確率的な表現のステートメントの評定の研究 (Lichtenstein and Newman, 1967; Wallsten et al., 1986)、頻度を表わす用語の評定の研究 (Simpson, 1944; Spector, 1976; Schaeffer, 1991; O'Muircheartaigh et al., 1993)、世論調査の結果の%の報告の用語の評定の研究 (Crespi, 1981)、などがある。

これらの研究では、つぎのような諸知見が確認されている。

- ・このような方法論的な調査の被調査者——多くの場合米国の大学生である——は、その調査の課題である「評定」の作業を十分にこなすことができる。
- ・「評定法」と「序列法」の結果は、それぞれのテーマ (あるいはトピックス) と調査対象者の違い (大学生以外の場合もある) を越えて、きわめて類似したものとなる。
- ・テストと再テストではほぼ同じ結果が得られるところから、その測定の信頼性 (reliability) は高いといえる (一般に、測定において、反復測定しても同じ結果が得られる場合、その

測定の信頼性は高いといわれる)。

- ・「評定」の処理あるいは手続に若干の差異があっても、それぞれの結果は十分に比較が可能である。
- ・頻度を表現する用語 (Schaeffer, 1991; Bradburn and Sudman, 1979) や、特別の修飾語 (qualifiers) などについては、さらに別の考察が必要である。

つぎに、「強度 (intensity)」を評定する技法であるが、これはレスポンス・スケールで用いられる用語 (scale terms) とその回答の修飾語 (qualifiers) の「強度」の分析を、さまざまなレスポンス・スケールへの回答の分布 (distribution) の比較にもとづいて試みようとするものである (Smith, 1979; Lauman et al., 1994)。この技法は、具体的には、さらにつぎの二つに区別される。

(i) 異なる回答者を用いる実験デザイン (an across respondents design) : これは、無作為に選ばれた二つの回答者のグループに、いくつかのレスポンス・スケールに答えてもらって、その結果を分析するというものである (Clogg, 1982; 1984)。

(ii) 同じ回答者を用いる実験デザイン (a within subjects design) : これは、同じ回答者のグループに、異なるレスポンス・スケールを用いた同じ内容の質問項目に対して、二回あるいはそれ以上答えてもらって、その結果を分析するというものである。

最後に、このような回答の分布を分析する技法の利点と欠点についても、検討しておかなければならない。

まず、利点としては、この技法が回答者に新奇な特別な課題を要求するというものではなく、ごく普通の質問紙調査の方式で課題に答えてもらう——つまり、あるテーマあるいはトピックスをめぐる質問項目に対して、標準的・典型的なレスポンス・スケールを用いて答えてもらう——ものである、ということがあげられる。このような質問紙調査の技法は、現在では、多くの一般の人びとにとって、すでに多かれ少なかれ「なじみ」のあるものとなっているからである。

つぎに、欠点としては、つぎのような点があげられる。

- ・質問紙調査の形式をとるところから、どうしても一度に限られた数のレスポンス・スケールしか使えない。
- ・統計学的には相対的に多くの被調査者 (回答者) を必要とする。
- ・レスポンス・スケールの差異にもとづく分布の型の変化という問題を検討する場合には、そのための前提として「真の分布 (true distribution)」という考え方とそのモデルが不可欠となるが、それはかなりやっかいな課題といわなければならない。因みに、このような方法は、「判定と決定 (judge and decision)」 (Schwarz, 1996) や「ファセット・セオリー (Facet Theory)」 (Borg, 1994; Borg and Groenen, 1997) に関する最近の文献とはまったく別の線上にあるものといえる。

さて、以上において、「評定法」に関してかなり詳細に検討してきた。このような検討から、つぎのような結論が導かれる。つまり、MINTS プロジェクトのような国際比較を目標とする実験的な研究にとっては、このような「評定法」はきわめて有効なものであるということである。それは、この方法によって、一方でさまざまな質問と回答のレスポンス・スケールの用語を0点から20点までの21点スケールで回答者に評定してもらうことが可能になるとともに、他方でこのような操作をおしてひとまず比較のための出発点に必要な強度についての数量化された得点 (quantified intensity scores) の獲得が可能になるからにはかならない。こうして MINTS プロジェクトにおいては、米国とドイツのレスポンス・スケールの比較のために「評定法」が採用されることになったのである。

しかし、米国とドイツの比較だからこそ、このような「評定法」が有効であったともいえる。つまり、この両国の比較においては、それらが同じ西欧社会の国であるところから、ある程度の文化の共有化を前提とすることが可能であった。しか

し、世界の別の地域や国ぐにを比較の対象に取り上げようとするならば、つぎのようなさまざまな問題が出てくることになる。

- ・それぞれの国（地域）の文化的背景からするならば、「評定」の点数についても、縁起のいい数と、縁起のわるい数ということがかかわってくる。
- ・それぞれの国（地域）で、教育をはじめとして——たとえば成績の評価のスケールなど——、さまざまな領域ですでに確立している標準的な——したがって定着した——「評定スケール」というものがある。
- ・一桁の数以外の数——たとえば MINTS プロジェクトでは0点から20点までの21点スケールを用いている——を用いて、回答者に評定作業をしてもらうという課題そのものに対する「なじみ」の程度が、国（地域）によって大きく異なる。

もっとも、このような問題は、一般的にはある程度まで予測が可能であるにしても、より具体的な個別の問題の発見のためには、やはりそのために特定の国（地域）を比較の対象に取りあげる国際比較・共同研究がどうしても必要になってくる。現在、筆者は、日本の質問紙調査で多用されているレスポンス・スケールの分析を、ドイツ・マンハイム ZUMA の Peter Ph. Mohler 教授、Janet A. Harkness 博士と共同研究の形で進めているが、この共同研究の意義と目標はまさにこのような点にあるといえるのである。

<文献>

- Almond, G. A. and Verba, S. (1963). *The Civic Culture: Political Attitudes and Democracy in Five Nations*. New Jersey: Princeton University Press.
- Berelson, B., 福岡住男訳 (1970). 「世論研究の系譜」『KYOWA AD-REVIEW』51. 協和広告.
- Borg, I. (1994). Evolving Notions of Facet Theory. In: I. Borg and P. Ph. Mohler(eds.), *Trends and Perspectives in Empirical Social Research* (pp. 178-200). Berlin; New York: de Gruyter.
- Borg, I. and Groenen, P. (1997). *Modern Multidimensional Scaling: Theory and Applications*. New York: Springer.
- Bradburn, N.M. and Sudman, S. (1979). *Improving Interview Method and Questionnaire Design*. San Francisco: Jossey-Bass.
- Bradburn, N.M. and Sudman, S. (1991). The Current Status of Questionnaire Design. In: P. P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz and S. Sudman (eds.), *Measurement Errors in Surveys* (pp. 29-40). New York: John Wiley & Sons.
- Cannel, Ch. F., Oksenberg, L. and Converse, J. M. (1979). *Field Experiments in Health Reporting 1971-1977*. ISR Research Report Series. Ann Arbor: ISR.
- Cliff, N. (1959). Adverbs as Multipliers. *Psychological Review* 66: 27-44.
- Clogg, C. C. (1982). Using Association Models in Sociological Research: Some Examples. *American Journal of Sociology* 88: 114-134.
- Clogg, C. C. (1984). Some Statistical Models for Analyzing Why Surveys Disagree. In: Ch. F. Turner and E. Martin(eds.), *Surveying Subjective Phenomena* Vol. 2. New York: Russel Sage.
- Converse, J.M. and Presser, S. (1994). Survey Questins: Handcrafting the Standardized Questionnaire. In: M.S. Lewis-Beck (ed.), *Research Practice* (pp. 89-162). London: Sage/Toppan.
- Crespi, L. P. (1981). *Semantic Guidelines to Better Survey Reportage*. Office of Research, International Communication Agency, Memorandum.
- Davis, J. A. (1993). Memorandum to the ISSP, Chicago: National Opinion Research Center (mimeo).
- Davis, J. A. and Smith, T. W. (1992). *The NORC General Social Survey: A User's Guide*. Newbury Park: Sage.
- Harkness, J. A. (1993). Mountains and Molehills-Equivalence in Cross-Cultural Surveys: the case of response scales. Paper presented at the annual meeting of the American Association for Public Opinion Research, St Charles, IL.
- Harkness, J. A. (1996). The (Re) Presentation of Self in Everyday Questionnaires. Paper Presented at the International Sociological Association Conference on Social Science Methodology, Colchester.
- Hippler, H.-J., Schwarz, N. and Sudman, S. (1987). *Social Information Processing and Survey Methodology*. Heidelberg: Springer.

- Houglund, J. G., Jhonson, T. P. and Wolf, J. G. (1992). A Fairly Common Ambiguity: Comparing rating and approval measures of public opinion. *Sociological Focus* 25: 257–271.
- Jones, L. V. and Thurstone, L. L. (1955). The Psychophysics of Semantics: An Experimental Investigation. *Journal of Applied Psychology* 39: 31–36.
- Kaplan, A. (1964). *The Conduct of Inquiry*. San Francisco: Chadler.
- Krosnick, J. A. and Fabrigar, L.A. (1997). Designing Rating Scales for Effective Measurement in Surveys. In: L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewim (eds.), *Survey Measurement and Process Quality* (pp. 141–164). New York: John Wiley & Sons.
- Laumann, E. O., Gagnon, J. H., Michael, R.T. and Michaels, S. (1994). *The Social Organization of Sexuality: Sexual Practices in the United States*. Chicago: University of Chicago Press.
- Lichtenstein, S. and Newman, J. R. (1967). Empirical Scaling of Common Verbal Phrases Associated with Numerical Probabilities, *Psychon. Sci.* 9: 563–564.
- 真鍋一史 (1993). 『社会・世論調査のデータ解析』慶應通信.
- 真鍋一史、栗田真樹、劉志明、加藤敬子、李鐘煥(1996). 「R. イングルハート (R.Inglehart) の『世界価値観調査 (World Values Survey) データ』の二次分析のための準備作業」『関西学院大学社会学部紀要』75: 67–82.
- 真鍋一史、栗田真樹、加藤敬子 (1997). 「R. イングルハート (R. Inglehart) の『世界価値観調査 (World Values Survey) データ』の二次分析のための準備作業 (2)」『関西学院大学社会学部紀要』76: 67–82.
- 真鍋一史、R. Inglehart (1997). 「Well-being の構造の国際比較」『関西学院大学社会学部紀要』78: 187–203.
- 真鍋一史、小野寺典子 (1997). 「家庭・仕事・結婚についての意識と行動の国際比較 (I) (II)」『放送研究と調査』557: 22–41, 558: 46–65.
- 真鍋一史 (1999). 「ナショナル・アイデンティティの構造——ISSP 国際比較調査のデータ解析——」『関西学院大学社会学部紀要』82: 145–156.
- 真鍋一史 (1999). 「質問紙法にもとづく国際比較調査の現状と課題」『国際協力論集』(神戸大学大学院国際協力研究科) 第7巻2号.
- 真鍋一史、小野寺典子 (1999). 「国への帰属意識——探索的データ解析の試み」『放送文化と調査』577: 44–63.
- Mittelstädt R. A. (1971). Semantic Properties of Selected Evaluative Adjectives: Other Evidence. *Journal of Marketing Research* 8: 236–237.
- Myers, J. H. and Warner, W. G. (1968). Semantic Properties of Selected Evaluation Adjectives. *Journal of Marketing Research* 5: 409–412.
- Nishihira, S. (1983). Political Opinion Polling in Japan. In: R.M. Worcester(ed.), *Political Opinion Polling: An International Review* (pp. 152–168). London: The Macmillan.
- O’Muircheartaigh, C. A. Gaskell, G.D. and Wright, D. B. (1993). The Impact of Intensifiers. *Public Opinion Quarterly* 57: 552–565.
- Osgood, Ch. E., Suci, G. J. and Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Urbana, IL: University of Illinois Press.
- Payne, S.L. (1951). *The Arts of Asking Questions*. Princeton/NJ: Princeton University Press.
- Presser, S. and Schumann, H. (1980). The measurement of a Middle Position in Attitude Surveys. *Public Opinion Quarterly* 46: 70–85.
- Schaeffer, N. C. (1991). Hardly Ever or Constantly?: Group Comparisons Using Vague Quantifiers. *Public Opinion Quarterly* 55: 395–423.
- Schumann, H. and Presser, S. (1981). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording and Context*. New York: Academic Press.
- Schwarz, N. (1996). *Cognition and Communication: Judgmental Biases, Research Methods, and the Logic of Conversation*, Mahwah: Lawrence Erlbaum.
- Schwarz, N. and Hippler, H.-J. (1991). Response Alternatives: The impact of their choice and presentation order. In: P. P. Biemer, R. M. Groves, L. Lyberg, N. A. Mathiowetz & S. Sudman (eds.), *Measurement Errors in Surveys* (pp. 41–56). New York: John Wiley & Sons.
- Schwarz, N., Knauper, B., Hippler, H.-J., Noelle-Neumann, E. and Clark, L. (1991). Rating Scales: Numeric values may change the meanings of scale labels. *Public Opinion Quarterly* 55: 570–582.
- Simpson, R. H. (1944). The Specific Meanings of Certain Terms Indicating Differing Degrees of Frequency. *Quarterly Journal of Speech* 30: 328–330.
- Smith, T. W. (1979). Happiness: Time trends, seasonal variations, inter-survey differences, and other mysteries. *Social Psychology Quarterly* 42: 18–30.

- Smith, T. W. (1997). Improving Cross-National Survey Research by Measuring the Intensity of Response Categories. GSS Cross-National Report No. 17. Chicago: National Opinion Research Center-NORC (mimeo).
- Spector, P. E., (1976). Choosing Response Categories for Summated Rating Scales. *Journal of Applied Psychology* 61: 374–375.
- Stone, L. and Campbell, J. (1984). The Use and Misuse of Surveys in International Development: An Experiment from Nepal. *Human Organization* 43: 30–37.
- Sudman, S., Bradburn, N. M. and Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. San Francisco: Jossey Bass.
- Van de Vijver, F. J. R. and Leung, K. (1997). *Methods and Data Analysis for Cross-Cultural Research*. Newbury Park/CA: Sage.
- Wänke, M. and Schwarz, N. (1997). Reducing Question Order Effects: The operation of buffer items. In: L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, and D. Trewin (eds.), *Survey Measurement and Process Quality* (pp. 115–140). New York: Jhon Wiley & Sons.
- Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R. and Forsyth, B. (1986). Measuring the Vague Meanings of Probability Terms. *Journal of Experimental Psychology* 115, 348–365.
- Wegener, B. (ed.) (1991). *Social Attitudes and Psychophysical Measurement*. Hillsdale / NJ : Erlbaum.
- Wildt A. R. and Mazis, M. B. (1978). Determinants of Scale Response: Label vs. Position. *Journal of Marketing Research* 15: 261–267.
- Worcester, R. M. and Burns, T. R. (1975). A Statistical Examination of the Relative Precision of Verbal Scales. *Journal of the Market Research Society* 17: 181–197.

A Study of Equivalence of Expressions from Response Scales Used in Cross-National Survey Research (1)

ABSTRACT

The purpose of this paper is to review studies on equivalence and non-equivalence of translated instruments in cross-national survey researches, which is based upon a literature survey in Janet A. Harkness ed., “Cross-Cultural Survey Equivalence,” ZUMA NACHRICHTEN SPEZIAL, No. 3, 1998.

The first section of this paper deals with the findings of empirical research on expressions from response scales used in cross-national survey research.

The second section reviews the MINTS project (Research into Methodology of Intercultural Surveys) which was carried out as a comparative and collaborative research of ZUMA (Zentrum für Umfragen, Methoden und Analysen, Mannheim, Germany) and NORC (the National Opinion Research Center, University of Chicago, USA). The MINTS project investigated expressions used in English and German ISSP (the International Social Survey Programme) response scales. The project is the first step in a research program aimed at exploring the limits and possibilities of translation with respect to response scales.

The third section suggests the meaning and merit for Japanese scholars to join this MINTS project.

Key words: cross-national survey research, response scale, equivalence