

「データアーカイブとアジア研究」

2017年6月25日 13:00~16:00

会場：一橋大学西キャンパス本館 31 番教室

園田茂人

こんにちは。樫山セミナー「データアーカイブとアジア研究」を開始したいと思います。樫山奨学財団の冊子と本日の発表の資料をお手元に配布しておりますのでご確認ください。

まず、樫山奨学財団理事の波多野優子さんから開会のご挨拶をいただきます。

波多野優子

こんにちは。ご紹介ありがとうございます。樫山奨学財団の理事で選考委員を務めております波多野と申します。ではこの機会に簡単にご挨拶をさせていただきます。アジア政経学会理事長の丸川教授および副理事長の園田教授には、今回のセミナーを共同開催できて光栄に思っておりますこと、樫山奨学財団を代表してお伝えいたします。本日はあいにくの空模様にもかかわらず、ご参加いただきましてありがとうございます。お手元の資料に樫山奨学財団の冊子がございます。

中には当財団についての小冊子、40周年記念のパンフレット、昨年の第11回樫山純三賞のパンフレット、それから今年の樫山純三賞の募集要綱と応募用紙が入っております。当財団の奨学金プログラムが始まったのは1977年ですから、今年で41年目を迎えます。30周年を記念して創設された樫山純三賞は、アジア研究の優れた出版物に贈られており、今回が第12回となります。今回のセミナーにご出席されている末廣昭教授も選考委員でいらっしゃいます。

昨年4月から今年6月末までに日本で出版されたアジア研究の書籍が対象となります。締切りは6月末です。本日ここにいらっしゃる皆さんのような学者、研究者の方々による出版物がすでに多数推薦されております。恐らく皆さんは私たちが知らない優れた出版物をご存知かと思えます。優れた出版物がありましたら、ぜひご推薦ください。

園田茂人

どうもありがとうございました。では今回のセッションを開催しようと思った経緯を簡単にご説明します。

言うまでもありませんが、アジア研究をとりまく環境は変わりつつあり、定性的データを使用する研究者でも、ICT（情報通信技術）の発展から多大な恩恵を被るようになっていきます。例えば、チェックしたい文献はウェブサイトで簡単に入手することができます。このようなテクノロジーは、研究テーマや使用する方法論に直接的な、ときには間接的な影響を及ぼします。

もう一つの点は、さらに大きな変化です。すなわち、ローカルな問題のグローバル化です。アジア研究には多くの動きがあり、例えば中国の研究者の間ではいわゆる全球中国研究、つまり中国の問題を多くの点でよりグローバルな問題と捉える研究の可能性を議論する動きが出ています。問題によっては、非常に国際的、全世界的になりつつあるのです。

しかし本日は、これまでのアジア政経学会の年次全国大会では議論されていない側面に注目したいと思います。アジアで増えつつあるデータアーカイブプロジェクトについてです。もちろんアジア研究を進める方法は多数ありますが、従来の地域研究の専門家は、その多くが依然として観察、資料分析など定性的データに大きく頼っています。

好むと好まざるとにかかわらず、多くの政府や NGO が多大の資金と労力を費やしてアーカイブプロジェクトを実施、推進しています。今日は、独自のデータアーカイブを構築するこうした取り組みに携わった専門家を数名お招きしています。こうした取り組みの狙いは、専門家だけでなく学者や実務に携わる者が、一次データ、二次データに関わらず、それぞれの目的に合わせて利用できるようにすることです。

各発表者の方には、まずご自身の国のアーカイブプロジェクトに携わった経験をお話いただきたいので、それぞれのお名前と所属を簡単にご紹介します。最初にご発表いただくのは、ソウル大学社会学部の金 碩鎬教授です。その次は中国人民大学社会学部の王 衛東教授です。3 番目の発表者はデ・ラ・サール大学政治学部のロナルド・ホームズ教授です。ホームズ教授はパルス・アジアというフィリピンの非常に大きな調査会社の社長でもいらっしゃいます。最後の発表者は日本の統計数理研究所の吉野諒三教授です。

その他に 2 名の討論参加者をお招きしており、4 名の発表者の方が触れなかった問題のいくつかを取り上げていただきます。一人は、現在、政策研究大学院大学（GRIPS）の学長と教授を務めておられる田中明彦教授、もう一人は慶応大学法学部でメディア研究に携わる山本信人教授です。ご存知のように、この 2 名の討論参加者は当学会の重要なメンバーでもあります。

では金 碩鎬教授に携わられた韓国のプロジェクトをご説明いただきましょう。金先生、よろしく願いいたします。

金 碩鎬

ソウル大学の金 碩鎬です。私が携わる組織は Korean Social Science Data Archive（韓国社会科学データアーカイブ）と言いますが、普段は KOSSDA と呼んでいます。この発表ではソウル大学 KOSSDA の主な活動内容と歴史をお話したいと思います。

私たちは主にデータ収集、蓄積、保存、公開、教育を行っています。KOSSDA は韓国を代表するデータアーカイブの一つとして、さまざまな学術的イベントや方法論教育プログラムを通し、研究資料の収集、公開、普及を主導しています。

KOSSDA の創設にあたっては、ドイツの GESIS、米国の ICPSR、ノルウェーの CESSDA など複数の有名なデータアーカイブ組織を訪問しました。そしてデータアーカイブの歴史におけるそうした組織の経験を KOSSDA に活かそうと努めました。

KOSSDA は 1983 年に非営利の社会科学系図書館としてスタートしました。2003 年に調査データアーカイブを追加し、2006 年にはデジタルデータと文献アーカイブを統合しました。近年、KOSSDA ではユーザーに社会科学データを提供するとともに、社会科学を専門とする学者に貴重な文献を提供しています。また、韓国に関するデータと文献の総合オンラインサービスも提供しています。2015 年、KOSSDA の事務局はソウル大学のアジアセンターに移転しました。

KOSSDA では調査データ、統計表、インタビュー、ナラティブな歴史データ、文書、観察記録、および研究機関や個人が作成したその他のデータを収集しています。デジタルデータベースを立ち上げており、アーカイブのウェブサイトを通して研究目的で使用するデータを提供しています。言うてみれば、私たちは定量的データと定性的データを単一のシステムに統合しようとしているのです。

主な活動の中では、統計学の上級講座を含めた方法論ワークショップが国内で大変人気を博しています。毎年、他の大学から 1,000 人以上の学生がこのワークショップに参加しています。

2016 年、KOSSDA は 191 セットの調査データを取得し、そのうち 72 セットを使用してデータベースを作成しました。2017 年 2 月時点で、KOSSDA には 2,251 セットの調査データが蓄積されており、定性的データのデータベースを 211 提供しています。この図は、データセットの特定、選定からデータの公開に至る、データアーカイビングのワークフローを簡単に示したものです。

ご存知かもしれませんが、1990 年以降の韓国社会では、社会科学のデータラボをまとめる動きが起こっています。毎年、数千以上の社会科学データが作成されていますが、古いデータセットはデータの質に問題があるため、私たちのところでは受け入れていません。このように、社会科学データセットの中でも質の高いデータセットを見極めようとしているのです。

データ取得については、先ほど述べましたように、定性的データと定量的データの両方を収集しています。データは学者個人と研究機関から取得します。KOSSDA は、100 以上の政府系研究機関と 50 以上の大学で構成されるコンソーシアムのようなものです。その一方で、研究者個人に連絡をとることもあれば、National Research Foundation（国立研究振興財団）と協力することもあります。また、学術誌に掲載される論文を検索し、分析することで、KOSSDA にふさわしい質の高いデータを特定します。

KOSSDA のガバナンス基盤は各機関のデータ寄託者のコンソーシアムにあります。コンソーシアムの各機関はデータ取得に協力するだけでなく、KOSSDA の運営においてもその正式代表を通じて重要な役割を担います。現在、コンソーシアムには 111 を超える研究機関が参加しています。韓国で社会科学データを収集しているほぼすべての機関とも MoU を交わしています。

受け取ったデータは、そのまま保存するわけではありません。データを受け取り次第、データを処理するようにしています。データ処理とは、データクリーニングを行うという意味です。データに目を通し、データ内に論理的な矛盾がないかどうかをチェックします。可能であれば、元の質問票をチェックするようにしています。

データクリーニングとエラー修正の後、メタデータを作成します。メタデータはデータの紹介あるいはデータの要約です。これは私たちにとって最も重要な作業で、機密情報を保護するためにデータからすべての個人情報除去するよう努めています。定性的データには多くの個人情報が含まれ、個人の経験を辿れる場合さえあるため、こうした情報をデータから取り除こうと努めているわけです。

メタデータを作成する際には、ドイツの協力機関が提供する Data Documentation Initiative（DDI：データ文書化イニシアティブ）を使用します。これは社会調査の世

界におけるデータ紹介の一種の標準規格です。このところ使用しているのは DDI のバージョン 2.1 です。

政治と世論、労働と雇用に関するデータは韓国の社会調査で非常に人気があります。最も人気のあるデータは Korean General Social Survey (韓国総合社会調査) です。この 10 年で、この調査結果を利用した 700 本以上の論文が一流学術誌で発表されました。

KOSSDA のウェブサイトでは、ユーザーに英語のサービスを提供しており、外国人研究者向けに英語サイトも開設しています。またハーバード大学、スタンフォード大学、シカゴ大学の各図書館といった外国の機関とも MoU を交わしています。確か 2010 年からはこうした機関の会員にデータを無料提供している他、250 の調査データを英語に翻訳しました。実際、この 250 のデータは人気とデータの質を考慮して選定したものです。こうして、外国人ユーザーがデータアーカイブに自由にアクセスできるようになっています。

KOSSDA の個人会員は国内外を合わせて 1 万 4,000 人以上いる他、150 の機関も会員になっています。会員には年会費を払ってもらいます。事務局をソウル大学に移転してからは、データサービスを無料で一般公開できるよう、資金援助を求めています。しかし、適切な資金を見つけるのは実に困難です。データアーカイブプロジェクトは、実のところ、お金の問題なのです。

私たちは Nesstar も提供しています。Nesstar とは何か。それはオンライン分析プログラムです。今では、Nesstar を使うと単純な頻度分析から論理的回帰分析まで、データを直接分析することができます。Nesstar では OLS 回帰、ロジスティック回帰、カイ二乗検定、相関解析などもできます。これは無料です。Nesstar には、中でも質と人気が高い 250 ものデータをアップロードしました。

2005 年には、Korean Research Memory (KRM) プロジェクトが発足しました。Korean Research Foundation (国立研究振興財団) が KOSSDA と協力して KRM を支援しており、KRM プロジェクトのウェブサイトには数千を超えるデータが集まりました。KRM のウェブサイトには、多数の定性的・定量的データがあり、あらゆる論文と書籍のデジタルデータもアップロードされています。非常に大掛かりなプロジェクトなわけです。

KOSSDA は大規模な研究チームを組織し、2016 年の年次報告書『Social Trends in Korea』をまとめました。このプロジェクトは 2008 年に韓国統計庁とともに立ち上げたもので、この報告書は官僚や学者の間で人気の高い政府報告書となっています。この報告書には、近年の韓国社会の変化と動向が示されています。最近、やはり韓国統計庁と共同で Key National Indicators (主要国民指標) を作成しました。

KOSSDA には、教育・研修プログラムもあります。

KOSSDA では、方法論研修プログラムを提供しています。これは、ミシガン大学の ICPSR (Inter-University Consortium for Political and Social Science) が提供する Quantitative Methods of Social Research サマープログラムに準拠して開発されたものですが、韓国に特徴的な状況も考慮に入れています。そのため非常に人気があるのですが、講義の質は管理するように努めています。具体的に言うと、1 クラスの定員を

35 人に限定しており、事実、学生の間ではこの夏の教育プログラムに登録するための競争が激しくなっています。

KOSSDA では毎年、冬と夏のワークショップに分けて 4 つの方法論研修プログラムを運営し、講義と実習を並行して提供しています。一方、春と秋の短期講座は理論についての講義です。

また、韓国がよく知られた社会科学データを対象として、データフェアを開催しました。韓国社会には 55 以上の巨大なパネルデータがあり、私たちはパネルデータの作成者を結ぶ一種のネットワークのようなコンソーシアムを作ろうとしています。またそのネットワークを運営することで、データの質を改善するための経験を共有しやすくしています。今年は二つのデータフェアを開催しました。その他に、研究論文コンペティションもあります。

最後の点として、私たちは東アジア、いえ、アジアのデータアーカイブを対象とする国際的なネットワークづくりに取り組んでいます。去年は、日本、中国、台湾から 4 つの主要なデータアーカイブがソウルでの会議に参加しました。KOSSDA は、東京大学の SSJDA、中国人民大学の CNSDA、台湾中央研究院の SRDA という、3 つの国を代表するアーカイブとともに国際会議を開催し、アジアデータアーカイブ連合を立ち上げることを巡って議論しました。この会議の結果、4 つのアーカイブは Networks of Asian Social Science Data Archives (NASSDA) の設立に同意し、この活動の方向性が定まりました。

KOSSDA は 10 年以上にわたって量的成長に集中してきましたが、今後はデータアーカイブの質的発展を追求するつもりです。KOSSDA は欧米の主要なデータアーカイブセンターに倣って設立されましたが、規模と活動の点でははるかに及びません。私たちは、米国の ICPSR、英国の ESDS、ドイツの GESIS など、先進国のデータアーカイブからさらに多くを学ぶ必要があります。ご清聴ありがとうございました。

園田茂人

金先生、ありがとうございました。次は中国の王教授にお話しいただきます。

王 衛東

20 年前、中国の主な問題の一つにデータ不足がありました。しかし 2000 年以降、学術分野に対する資金援助が拡大したため、中国本土で入手できるデータセットは増加し、中国の科学者コミュニティはデータ共有の重要性と必要性を理解するに至りました。最も重要なことは、政府系の科学研究振興財団だけでなく中国政府もデータ共有の重要性を認識するようになったことです。

こうした歴史的状況を背景として、CNSDA（中国調査データアーカイブ）があります。CNSDA は設立当初から現在まで、中国本土で唯一の社会科学データアーカイブです。CNSDA は、中国本土で一番の政府基金である Natural Science Foundation (NSF) of China（国家自然科学基金委員会）から資金援助を受けています。資金援助は 2012 年に始まり、今年で終了する予定ですが、実は中国人民大学からも資金援助があります。二つの組織から並行して資金が出ているわけです。

人民大学の学長は、今回の資金援助が終了した後も、このデータアーカイブへの資金提供を継続すると宣言しています。しかし NSF もこのデータアーカイブへの資金援助を続けると思います。CNSDA プロジェクトの一環として、わずか 3 年前の 2014

年からオンラインでのデータアクセスサービスの提供を開始しました。CNSDA の歴史は韓国の KOSSDA に比べるとそれほど長くありません。

CNSDA の役割は、中国本土で実施されたあらゆる調査プロジェクトの原データと関連文書を取得し寄託することです。ちょうど中国で出版されるあらゆる書籍が北京の国家図書館に収蔵されるのと同じです。そのため、さらに調査データの収集と文書化を進め、オープンアクセスと関連サービスを学界に提供するつもりです。

主席研究員は袁 衛教授です。教授は人民大学の元副学長で、現在は当組織の所長です。私は次席研究員および CNSDA の所長代理を務めています。CNSDA の拠点は中国人民大学の National Survey Research Center（中国調査研究センター）に置かれています。

当チームは所長代理 1 名、というのは私ですが、DBA およびプログラマー4 名、UI および VI デザイナー2 名、データマネージャー6 名、それに多数のパートタイムアシスタントで構成されています。現在、20 人以上のアシスタントがいると思います。その大半は学部生、修士課程と博士課程の学生です。

CNSDA では、すべてのソフトウェアにオープンソースソフトウェアを使用する方針をとっています。使用するデータベースは MySQL で、スクリプト言語は PHP、ウェブサーバーは APACHE です。システム全体はオープンソースの Linux システムを基盤としています。

データアーカイブ構築の最も重要な点は、データをいかにして取得するかです。実際、プロジェクトが発足した 2012 年には、中国本土で利用できるデータソースを求めて、出版された学術誌、書籍、報告書をすべて検索しました。1980 年から 2012 年までに中国本土で実施された調査プロジェクトとして、712 件が見つかりました。

こうしたデータのうち、286 のデータは中国政府が、197 のデータは国内の大学と研究機関が、127 のデータは企業が、100 余りのデータは海外の研究者と機関が所有しています。

これら初期シードデータセットを使用してデータアーカイブを構築するにあたり、中国人民大学の National Survey Research Center（中国調査データセンター）は主要な調査プロジェクトの権利を手放すことに同意しました。ここには、Chinese General Social Survey（中国総合社会調査）、China Education Panel Survey（中国教育追跡調査）、China Religion Survey（中国宗教調査）、Chinese Longitudinal Aging Social Survey（中国老年社会追跡調査）、China Employer-employee Matched Data Survey（中国雇用主・被雇用者マッチングデータ調査）、China Enterprise Human Resource Survey（中国企業人事調査）、College Students Panel Survey（首都大学生成長追跡調査）が含まれます。この7つのデータセットは、いずれもまず CNSDA で公表されました。そしてここがこのデータセットをダウンロードできる唯一の場となっています。

他の機関が収集した700を超えるデータセットについては、当データアーカイブにデータを保存する許可を得るため、さまざまな戦略を用いました。中国政府が所有するデータセットについては、官庁を一つずつ訪問し、私たちが何をしたいかを説明し、当データアーカイブにデータを保管するよう説得を試みましたが、今のところほとんど成功しておらず、わずか1件か2件について許可を得ただけです。これは本当の話です。国内の大学や研究機関については、メールを送ったりウェブサイトか

ら番号を見つけて電話をかけたりしました。肯定的な回答もありましたが、それほどうまくいってはいません。それから、あちこちでデータセットを公表することについて話しています。

サーバールームは、クラウドストレージに移転しました。中国のクラウドサービス企業であるアリババから、当データアーカイブに無料でクラウドサービスを提供してもらおうという同意を得たのです。これは大変助かります。

これまでに、122 の調査データセットが検索またはダウンロードできるようになっており、そうしたデータと文書の大半が当データアーカイブに保管されています。しかし中には、当データアーカイブへのリンク情報を有するのみのデータセットもあります。私が所長代理に就任してからは、登録手続きを簡素化しようと努めました。今では、データを使用したいと思ったら、オンラインで登録するだけで1分以内にデータをダウンロードできます。

この他に、オンラインで Q&A 形式の自律的ナレッジシステムも提供しています。これはユーザーにとって非常に役立つものですし、メール、電話コンサルティングサービスもあります。また、ユーザーにジオコード情報など限定的なデータサービスも提供しています。ユーザーが私たちのデータとセンサスデータおよびマクロ有効データをマッチングさせるのを助けることができます。非常に大切なのは、当データアーカイブで提供するすべてのサービスが完全に無料だということです。

これまでのところ、CNSDA に登録したユーザー数は 3 万人を超えており、昨年末時点でのダウンロード数は 20 万回を超えました。ダウンロードされたデータセットの上位 5 位は、Chinese General Social Survey (中国総合社会調査)、Chinese Private Entrepreneur Survey (中国私営企業家調査)、China Education Panel Survey (中国教育追跡調査)、College Students Panel Survey (首都大学生成長追跡調査)、China Longitudinal Aging Social Survey (中国老年社会追跡調査) となっています。最初の二つは中国政府機関から提供されたもので、私はこれらの調査プロジェクトで主任研究員を務めました。

当データアーカイブのユーザーはさまざまな分野に属しています。これまでで、200 人以上のユーザーを数える分野は 11 あり、経済学、経営科学、社会学がその上位 3 分野となっています。その後、統計学、教育学、ジャーナリズム学、政治学、法学、哲学、心理学、民俗・人類学の 8 分野が続きます。

海外のユーザーが利用するデータは 22% を超え、100 人以上のユーザーが登録している国または地域は 8 つあります。中国本土に続くのは米国で、ユーザー数は 2,000 人を超えます。その次が英国と日本です。日本のユーザーは 500 人以上です。これらユーザーがどのような人なのかはわかりませんが、データベースをチェックして知ることはできます。次が韓国、香港、台湾、ドイツなどです。

園田茂人

どうもありがとうございました。次はホームズ教授にお話しいただきましょう。

ロナルド・ホームズ

金教授と王教授の発表をお聞きして、両国がはるかに進んでいることに感心しています。フィリピンで使えるサーバーや設備の数を考えると、韓国と中国はデータアーカイビングの面では先進国の部類に入り、かたや私たちは開発途上国のようです。

中国と韓国ではいずれも大学が拠点となっていることをうらやましく思います。フィリピンの調査研究には長い歴史があります。日本でいつ調査機関が発足したのかは知りませんが、フィリピンでは 1950 年代前半に発足しており、東南アジアで調査機関が設立された最初の国かもしれません。しかし実際には、この歴史には中断があります。再び表舞台に出てくるのは 1985 年、民間の非営利・無党派の組織である Social Weather Stations (SWS) が設立されてからです。私が率いるパルス・アジアは比較的新しい組織で、1999 年の設立です。

王教授はセンターが被るリソースの制約について話されましたが、私たちの制約は、資金を自ら賄っていることもあり、大学や政府に頼れないことです。政府が調査結果に不満な場合はなおさらです。事実、現大統領のもとでは特に、調査結果に不満を持っています。

SWS の担当者に問い合わせたところ、SWS には 566 のデータセットがあり、それは 276 の全国規模のデータセット、つまり全国調査に分かれています。SWS で実施した最大の調査は 2016 年に行った出口調査で、5 万 7,000 人の回答者から得たサンプルがあります。SWS が実施する調査の標準的な規模はパルス・アジアとほぼ同じで、全国 1,200 人となっています。また 290 の地域調査があります。

SWS の全国調査のうち 5 件は比較調査を行ったもので、World Values Surveys (世界価値観調査) と Asian Barometer (アジア・バロメーター) が含まれます。中には、私たちが「便乗質問」と呼ぶ、比較調査グループがそのような全国調査を利用した全国調査もあります。実施される調査の中に、ISSP や Comparative Study of Electoral Systems (選挙制度の効果の国際比較調査) が独自の設問をいくつか持ち込んだモジュールがあるのです。SWS は 3 カ月ごとに定期的な調査を実施し、そのデータはコーネル大学のローパーセンター、ミシガン大学の ICPSR に寄託されます。SWS は近々、I-ASSIST というもう一つの国際的なデータアーカイブに加わる予定です。

私たちの組織は新しいため、わずか 251 の定量的データセットしかありません。独自の定性的調査は約 13 に留まります。そのため、調査の大半は定量的調査です。そのうち 65 は全国調査で、残りは地域調査です。また、3 カ月ごとの定期調査も行っており、調査データはコーネル大学のローパーセンターに寄託されます。この点で、当組織は中国や韓国の場合とは違います。いまだにかつての宗主国、つまり米国に頼る必要があるのです。私たちのところではデータアーカイビング保存を担うフルタイムのスタッフが不足しているため、米国の手を借りています。

定期的に尋ねている質問はもっと不変なこと、つまり生活の質です。参加者に、自分の生活の質がこの数年で向上したか悪化したか、そして今後数年で生活の質が向上すると思うか悪化すると思うかを尋ねます。このデータセットだけを見ると、フィリピン人はいつも、自分たちの生活の質はこの数年で悪化したか、今後数年では向上するだろう、と答えることに気づくでしょう。要するに、フィリピン人は生活面では失敗しているのですが、本質的に楽観的なのです。

調査では、政党を支持するかどうかも尋ねました。興味深いことに、政党を支持すると答えた人はわずか 10% となっています。なぜフィリピン人は政党を支持しないのでしょうか。それはアメリカ人が、政党をあまり重要としない政治システムを作ったからです！我が国の政党はただ、アメリカの政治学者が指摘するように、内部

で動員されているだけです。その一方、選挙は任意ですが、80%のフィリピン人情報提供者が選挙に参加し、選挙は民主主義を推進するといまだに信じています。

フィリピンは民主主義の国で、私たちが質問票に記載する質問は世界価値観調査やアジア・バロメーターにはありません。時には、自分にとって民主主義は何を意味するかを定義してもらおうよう、自由回答式の質問をすることもあります。最も多い一般的な回答は「雇用の確保」や「賃金の上昇」で、民主主義とは何の関係もありませんが、民主主義の成果と言ってよいのかもしれませんが。

園田茂人

ありがとうございました。次は吉野教授です。

吉野諒三

私は統計数理研究所に勤めております。当研究所にデータアーカイブはありませんが、長年にわたって社会調査を行ってきました。

当研究所は第二次世界大戦中に設立されました。ですから、戦時中に何を行っていたかは想像がつくと思います。第二次大戦後、日本は米軍に占領されました。占領下で当研究所は、戦後の民主主義の発展に向けた世論調査を実施するため、政府統計と科学的方法論を再編するという新たな使命を帯びました。

中国人民大学の首席研究員、袁教授は、1990年代に当研究所の客員教授でいらっしゃいました。当時、私たちは日本にデータアーカイブを構築するということを議論していたのですが、袁教授はそのアイデアを北京に持ち帰って早速ご自身の研究センターを開設され、中国版の National Character Survey (国民性調査) を実施されたのです。Gallop Korea の朴社長も 1970年代と 80年代に当研究所を訪問されました。朴社長は調査方法論を学ばれ、韓国に帰国して Gallop Korea を立ち上げました。ですから、私たちはさまざまな形でつながっているのです。

第二次世界大戦後、当研究所は政府統計と世論調査の再編を支援し、世論調査の方法論を用いて日本人の国民性調査 (Japanese National Character Survey) を開始しました。これは 60年以上続きました。皆さんは米国の General Social Survey (総合的社会調査) やヨーロッパの Eurobarometer (ユーロバロメーター) をご存知かもしれませんが、そうした調査はいずれも、1953年に始まった日本人の国民性調査にヒントを得て始まったものです。ですから、これが世界中の似たような調査に影響を与えたオリジナルなのです。

諸外国ではデータを一般公開したため、世界中に知られるようになりました。しかし当研究所ではまだデータセットを一般公開していません。もちろん、多数の概略報告書を出版し、世界中に流通させていますが、十分に認知されているとは言えません。

1970年頃には、いくつかの調査を比較調査に拡大しました。私は約 30年にわたってこの調査に携わっているのですが、こちらがその調査の一部をリストにしたものです。この 10年から 15年は、アジアのいくつかの国も対象にしています。

質問票で繰り返し使っている質問の例を一つご紹介しましょう。それは、「生まれ変わるとしたら、男と女のどちらに生まれたいですか?」というものです。60年以上にわたり、日本人男性の答えには変化がありません。約 90%の日本人男性が「ま

た男に生まれてほしい」と答え、「次は女に生まれてほしい」と答える人は 5、6%に留まります。それに対して、こちらは女性の回答です。1953 年には 64%の日本人女性が「次は男に生まれてほしい」と答え、27%が「また女に生まれてほしい」と答えていましたが、これは着々と変化し、1993 年頃にはこの数字がすっかり逆転しました。

私たちの仕事は、国際比較調査の研究です。使用する言語と調査の方法論は、国によってそれぞれ異なります。日本では住民あるいは有権者のほぼ完全なリストが使用できるため、それを使うことで理想的な無作為抽出を行うことができます。しかし、他の国では、米国やヨーロッパ諸国でさえも、有権者リストがないためにそれができません。無作為抽出が困難なために、米国の大統領選挙と英国の EU 離脱を問う国民投票では予想が大きく外れました。

詳しいことは省きますが、国際比較にあたって、私たちは Cultural Link Analysis（連鎖的比較調査分析論）または Cultural Manifold Analysis（文化多様体解析）という、各国を比較するための科学的パラダイムを開発しました。初めから全く異なる二つの国を比較しても、あまり意味はないでしょう。それで私たちはまず、共通する側面を持ちながら異なる側面もある二つの国またはグループを比較します。

約 20 年前、EU は大変うまくいっていました。当時、アジア諸国の統合というアイデアについて話すと、アジアはあまりにも多様で統一など考えられないと批判する人がいました。しかし EU における多くの問題を目の当たりにしている今、多様なアジア諸国のゆるやかな統合を目指す方が、はるかに優れた方法だと考えてもよいかもしれません。

過去の経時的調査では、日本人の基本的な特質を特定しました。ここでは 1 点だけ強調させてください。対人関係に対する姿勢は過去半世紀以上にわたって、もしかするともっと長い間、変化していません。対人関係に対する姿勢の基本的な側面は、どの国でも長年にわたって変わらない可能性が高いでしょう。当研究所のウェブサイトではさらに興味深い調査結果がご覧いただけます。

私たちは、文化、歴史、宗教、政治経済に関する社会的制度の違い、および言語や調査方法論の違いを踏まえて、調査回答の国際比較を行う難しさを認識しています。人々の回答には、本人の意見と、性差、民族、性格などに固有の一般的な回答傾向とが入り混じっています。日本人は極端な回答を避ける傾向にあり、中間の回答カテゴリーまたは「わかりません」という回答を好みますが、フランス人は批判的あるいは否定的なカテゴリーを選ぶ傾向があり、インド人は肯定的なカテゴリーを選ぶ傾向にある、という具合です。

各国の調査方法論には、その経済的、歴史的、政治的、社会的な条件が現れます。調査データの分析を実施する前に、調査方法論の違いにもっと注意を払うべきです。データ分析の際、単一の指標や指針は信頼できないため、多次元データ分析を利用し、多数の国において多数の質問項目から得られたデータを合わせ、位相的關係を調べるべきです。

当研究所の調査データは、世界の相互理解や世界平和、そして世界の繁栄に向けたアジア研究に役立つと考えています。しかし「エビデンスに基づくアジア研究」を促進するためには、経時的かつ国際的な調査を数十年にわたり継続するためのシステムを確立する必要があります。しかし、データ収集とデータアーカイビングは現時点では十分ではありません。

抽出方法、言語、経済的・政治的制度を含め調査条件が大きく異なることから、国際比較にはさらに注意を払わなくてはなりません。アジア的なゆるやかな統合を研究することで、政策決定に対する新たな視点が得られるかもしれません。政治学者がこのような問題にアプローチすることを期待しています。

園田茂人

吉野先生、ありがとうございました。

では、コメントセッションに移る前に、4名の発表者の方全員に簡単な質問をしたいと思います。一つは、皆さんが直面している事実や傾向の中で何が最も困難だと思いますか、という質問です。発表者の中には金銭的な問題に触れた方もいらっしゃいましたが、私自身の経験から言うと、私たちが直面する可能性のある最大の課題はデータの有効利用です。

東京大学東洋文化研究所では、2003年から2008年にかけて田中教授と猪口教授がアジア・バロメーター・プロジェクトに力を入れていました。しかし私の目から見た最大の課題の一つは、たとえデータがあっても、そのデータを使って利益を得られるユーザーを見つけることは大変困難だ、ということです。データを作成し、まとめ、仕上げた本人がその重要性を十分に認識していても、そのデータを使い、データ分析に基づいて新しいタイプの論文を書く有望な若手学者を見つけることは、決して容易なことではありません。

二つ目は、一つ目の質問とも関連しますが、この困難を克服する打開策はどこにあるか、という質問です。発表者の中には、公開の重要性や国際会議に触れた方もいますが、他に困難を克服する方法を考えたことがあるのでしょうか。

金碩鎬

先ほど述べましたように、最大の困難は資金調達ですが、幸い、私たちのところはあるコングロマリットが所有する文化団体から支援を受けています。年間約50万米ドルの支援がありますが、当機関の運営にはさらに50万米ドルが必要です。そこで政府機関や企業と協力する中でさらなる財政的支援を求めているのです。

ユーザーについてですが、データアーカイブの目的は質の高いデータを公開し、質の高い論文の発表を支援することです。私の発表で、教育プログラムとデータフェアのネットワークをご紹介しました。しかしその鍵は、あるいはこれを課題と呼んでもよいのですが、高度な統計を実行し学習する中で相手に私たちのデータを使ってもらえるかどうかです。私たちは論文コンペティションのプログラムも運営しており、これも当機関のデータ利用者を増やす効果的な取り組みの一つだと思っています。

私たちが抱える第三の課題は、関連機関との連携です。

例えば、韓国総合社会調査(KGSS)について言うと、KGSSとの協力は非常に効果的でしたが、その理由は、韓国の石賢浩教授とシカゴ大学のトム・スミス教授が私たちのパートナーだからです。実は、トム・スミスは私の師で、彼と私でKGSSを立ち上げました。KGSSを確立して間もなく、私たちは2006年にKorean Social Science Data Archive(韓国社会科学データアーカイブ)を立ち上げました。

他の政府機関と大学の学者からも、価値と質の高いデータを入手しなくてはなりません。問題は、相手の関心がすべて異なることです。そのため、相手が当機関に協力してくれる方法を見つけなくてはなりません。最近私たちが、パネルデータを収集する機関にそのデータ作成の経験とノウハウを共有してもらうためのネットワークづくりに努めているのはそのためです。

データ検索については、韓国の各機関がそれぞれ独自のデータアーカイブを構築しようとするという問題もあります。政府機関を訪問すると、必ずと言ってよいほど、独自のデータアーカイブを作るつもりだと聞かされるのですが、本当の問題は、そのような姿勢では各データアーカイブのユーザー数が減ってしまうということです。

私たちとしては、彼らが独自のデータアーカイブを構築するのを支援する戦略をとっています。通信事業者のKTからはサーバー用に大容量クラウドストレージの無料提供を受けており、そのサーバーを通してウェブサイトの一部を共有することができます。また、彼らを支援したり、この10年間にKOSSDAに蓄積された技術や経験を提供したりすることにも努めています。

回答の最後に、米国の総合的社会調査（USGSS）の例を紹介させてください。

USGSSは、シカゴ大学の全国世論調査研究センター（National Opinion Research Center：NORC）が収集しているもので、そのデータのユーザーは年間1万人を超えます。さらに、500本以上の論文が一流学術誌に発表されています。NORCでは、そのウェブサイトからユーザーが欲しいと思うデータをダウンロードできるようにする戦略をとっています。さらに興味深いのは、学界の潜在的なユーザーにUSGSSを提供できるウェブサイトがさらに5つあるということです。ミシガン大学のICPSR、コーネル大学のローパーセンター、ドイツのGESIS、そしてインドのSRDAもUSGSSを一般に提供する権利を有しています。

ですから私は、データが、特に大規模なデータセットが収集されたらすぐに公開し、一般の、社会の所有物にすべきだと考えています。

園田茂人

韓国のケースは、多くの機関が先頭に立とうと競争していることを示唆していますが、これは本当に矛盾した状況ですね。それが、金先生も示唆されている懸念の一つです。

王 衛東

中国は韓国と似た状況にあります。

中国には、ナンバーワンになりたがっている学術調査機関が多数あります。しかし私が思うに、これが最大の課題ではありません。中国にとって最大の問題は、中国政府に対し、持っているデータを公開してもらうよう説得するのが非常に難しいということです。実際、先ほどお話ししましたように、中国政府は中国本土では最大のデータ保持者です。私自身がとった統計によると、中国政府は40%以上のデータコンテンツあるいは調査データを所有しています。マクロ有効データや中国政府のみが所有しているセンサスデータだけでこの通りです。

現在の中国政府は、データ公開の重要性を理解していると思います。現在の李克強首相は、政府機関にデータを公開するよう再三命じていますが、おかしなことに彼

らは命令に従いません。首相が政府機関にデータを公開するよう何度も命じているのに、中国政府当局にはいまだにデータを公開する気がないのです。彼らはデータを公開しても、何か得られるものがあるとは思っていません。

もう一つの課題は、異なる機関どうしではなく、異なる分野どうしの競争です。「経済帝国主義」はどここの国でも見られる普遍的な現象に違いありません。経済学者は因果モデルを過度に強調し、横断的なデータよりもパネルデータや時系列データを偏重します。実は、国家自然科学基金委員会と Social Science Foundation of China（国家社会科学基金）は経済学者に牛耳られています。そのため最近では、こうした基金の審査委員会は、横断的なデータ調査よりパネルデータ調査に割く資金を増やすべきだと考えるようになりました。

おかしな考え方ですが、いずれにしても中国人民大学は今後も社会調査プロジェクトを実施していきます。2010年以降、中国本土では、北京大学の Chinese Family Dynamic Panel Study（中国家庭追跡調査）、中国人民大学の Child's Health and The Aging Study（児童の健康と老年研究）、中国教育追跡調査など、多数の学術調査プロジェクトが立ち上げられています。中国総合社会調査は、おそらく中国で唯一の横断的な大規模社会調査と考えられますが、この調査からはさまざまな社会的問題に関する多彩な変数が得られます。先週、幸運にも卒業生の一人が当大学に多額の寄付をしてくれたため、中国総合社会調査を比較的スムーズに継続することができるようになりました。

園田茂人

ありがとうございました。各国が抱える困難の説明を伺うと、さまざまな国の固有性がわかるのではないかと思います。王教授が中国の「経済帝国主義」に言及されましたので、私自身のエピソードをご紹介します。私はこのセミナーの準備に奮闘する中で、アジアのさまざまな分野で実施されている調査プロジェクトを探していました。しかし、アジアで共通のデータセットを作成することによってアジアの研究者を結び付けようとしている経済学者はなかなかいない、ということがわかりました。しかし、社会学者と政治学者はむしろ熱心にこうしたことを行っています。ホームズ先生は政治学者ですが、何かご意見はありますか？

ロナルド・ホームズ

Social Weather Stations は1人の経済学者が設立した組織で、長年にわたって貧困を自己評価する質問を実施してきました。幸い、フィリピンには王教授がおっしゃるような「経済帝国主義」はありません。

フィリピンの第一の課題は、定性的研究と定量的研究を組み合わせた社会調査に取り組むためのリソースと協力的な組織が十分でないということです。二つを組み合わせれば、いくつかの社会問題や社会現象を掘り下げることができるのですが。残念ながら、フィリピンでは私たちと協力し、データを取り扱うことができるような組織はあまり多くありません。

私は、調査の質問票を長い時間かけて作成する度に、他の国で実施された調査項目を調べます。しかし中国や日本、韓国における調査項目はなかなか見つかりません。というのは、調査の質問票がそれぞれの国の言語で書かれているからです。それで多くの場合は純粋な研究、ミシガン大学に目を向けることとなります。韓国の NASSDA に、少なくとも東南アジアの国々でアクセスが可能になるよう、こうした質問票のいくつかを英語に翻訳する支援をしてもらえれば、とても助かります。

園田茂人

統計数理研究所については、長い歴史を持つという点では特別だと思いますが、やはりアーカイブセンターを持っています。研究所が注意深く国際比較を推進してきた中で、研究所や教授個人が直面している困難や課題についてお聞きしたいと思います。

吉野諒三

お話ししたように、当研究所では原データはまだ一般公開していません。しかし過去20年、もしかすると30年ほどにわたって、必ずしも当研究所においてではありませんが、ケルン大学のGESISやコーネル大学のローパーセンターなど大規模な機関においてデータアーカイブの構築に取り組んでいます。私たちの調査チームの規模ははるかに大きいのですが、現在、研究所には全部で約45人が所属しているものの、社会調査プロジェクトに加わるメンバーは2、3人しかいません。私たちは今もなお大規模な調査プロジェクトを進めていますが、現在はビッグデータの時代になっているので、大変に困難です。エンジニアなら問題なく多額の研究資金を獲得できますが。

しかし私たちの研究によると、第二次大戦前に米国で行われたビッグデータに基づく大統領選挙の予想は外れています。そのため私たちは、予想には小規模でも質の高い統計データを重視するようになりました。それなのに最近、エンジニアたちは口を開けばビッグデータ、ビッグデータとばかり言っています。当研究所の日本人の国民性調査は60年以上続きましたが、調査を開始する度に、経時的調査の実施だけを名目として政府に資金を要請することはできません。それで、資金援助を得るために、まったく新しいプロジェクトをスタートさせるかのように振る舞っていました。

研究者の皆さんは、時系列比較を行うための経時的調査の重要性を理解または評価していると思います。しかし国民性調査に関しては、お役所からは「ただ国民性調査を実施するだけではよくない」と言われます。お役所の典型的な姿勢です。そのため、調査を実施するための資金獲得は大変です。

最近の日本の若手官僚には、私たちが行う調査の意義が理解できません。このところ世界中でエビデンスに基づく政策決定が話題になっています。ある若手官僚とこの問題について話したことがあります。彼は私たちの調査を支援することに対して悲観的な見方を示しました。その一方、米国、ヨーロッパ諸国、その他の国々から多くの研究者が日本語も知らずに日本にやってくる、いくつかの統計報告を使わせてほしいと言います。これは世界中の政治学者にとって自然な態度です。日本の官僚がなぜこれほど違うのか、不思議でなりません。

園田茂人

ありがとうございました。そうすると、いくつかの問題は資金に関係したのですが、他の問題は政府部門による資金援助を含め、支援に関するものですね。国によって問題はそれぞれですが、それでも改めてこのプロジェクトは、研究のためにデータを共有できる公の場を作ろうという試みと見なす必要があります。

扱いにくい問題にお答えいただき、ありがとうございました。次に、これまでの議論に対し、田中教授からコメントをいただきたいと思います。

田中明彦

園田先生、それから4名のパネリストの方には、消化しきれないほど多くの情報を提供していただき、ありがとうございました。少々宣伝を用意してきましたのですが、それは省略して、今日の議論に関連すると思う話題から始めたいと思います。

今日の発表の話題は、主に調査データと調査データアーカイブです。しかしまず、私たちは現在、さまざまな形態のデータがあふれる事態に直面しているということをお話したいと思います。

今や経済学者は、World Development Indicators (WDI: 世界開発指標) を丸ごと、クリック一つでダウンロードして利用することができます。WDI 指標は、264カ国について1960年から2016年までの1,504の指標を189メガバイトのCSVファイル一つに収めたもので、同様のデータがさまざまな国際機関から利用できます。IMFからはInternational Finance Statistics (国際金融統計) と Direction of Trade Statistics (貿易統計) という3ギガバイトの、やはり巨大なCSVファイルが提供されています。国内経済と統計のデータも多数あります。

経済学については、本日の会場となっている一橋大学が、開発途上国の長期的な経済統計を収集していることで昔から知られています。この他、現在では多数の調査データが保存されています。World Value Surveys (世界価値観調査) と世界中の実にさまざまなバロメーター、および国内調査およびメディア調査が保存されているのです。

こうした数値データに加え、現在では多くのテキストデータもあります。新聞社のウェブサイトを開覧すると、19世紀からの、あるいは20世紀ほぼ全体に及ぶ新聞記事のデータセットがすべて得られるでしょう。私たちの「The World and Japan (世界と日本)」というささやかなデータベースでも8,000を超える文書を提供しており、国際的な研究に役立ちます。日本の国会のウェブサイトを開覧すると、本会議と委員会を含め、1945年から現在までの国会審議録を検索することができます。社会的、経済的、歴史的な分析に役立つ画像データもたくさんあります。現在、アジア歴史資料センターでは、戦前の日本に遡る歴史的に重要な多くの記録を、オリジナル文書の画像もあわせてまとめているところです。ですから、私たちはデータがあふれる時代を生きているのだと思います。

地域専門家と社会学者にとっての課題は、データをどう使うかということです。それこそが、園田先生がこの場に私たちを集めた動機だと思います。ただ、矛盾しているのは、いわゆるデータの不在という問題があることです。少なくとも若い人たちの間では、データがインターネット上に存在しなければ、まったく存在しないものと見なされる傾向があります。言い換えると、文書や数値データをインターネットにアップロードしなければ、全世界から無視されてしまうのです。

吉野教授がお話しされた、日本の国民性調査の窮状に似たところがあると思います。私としては、統計数理研究所に次回の日本人の国民性調査データを原データとともに一般公開していただきたいと切に願っています。インターネット上に置かなければ、誰も注意を払いません。

同様に、日本のニュースメディア企業が実施した無数の調査のように、一般公開されていないというだけの理由で十分に利用されていないデータが大量にあります。各新聞社は毎月、毎週のように調査を実施しますが、一般市民はそのデータを利用

させてもらえません。原データはすべて隠され、限られた人だけが、例えば安倍首相の支持率などのデータを利用できます。言い換えると、皆さんはどのようなタイプの人がどのような理由で首相を支持しているのかを知るためのデータにはアクセスできないということです。

資金が途絶えたためにデータが欠けてしまうこともあります。例えば、猪口先生と私、そして園田さんが一生懸命に取り組んだアジア・バロメーター・プロジェクトですが、この調査では 2008 年でデータ収集は終わり、その後はまったくデータ収集が行われていません。貴重な調査データがあっても打ち切られ、インターネット上から強制的に消されてしまうかもしれません。このような場合、こうしたデータは実質的に存在しなくなるということです。

次の問題は、私が整ったデータや乱雑なデータと呼ぶものの問題です。

通常、テキストデータと画像データの最も重要な用途と目的は、自分が伝えたいものを見つけることです。自分のモノグラフの中で、誰かが悪者だったことを証明するために、自分を正しいとする文書を見つけたら、それで十分でしょう。しかし、大規模なデータを使用し始めると、乱雑なデータと整ったデータの問題に遭遇する可能性があります。トルストイは『アンナ・カレーニナ』の中で、「幸福な家族はみな似ているが、不幸な家族はそれぞれの形で不幸」だと説明しました。後にウィックハンとグロールドマンが『R for Data Science』（2017年）の中でこのメッセージをもじって、「整ったデータセットはみな似ているが、乱雑なデータセットはそれぞれの形で乱雑である」と述べました。大規模なデータセットを分析する事業に乗り出すよう学生その他の人々を納得させられる問題の一つは、この乱雑さから来るのだと思います。

通常、コードブックを読み通すことは苦痛です。単一の調査データを扱っている限り、コードブックに慣れることは可能で、最終的にはデータセットを利用することができます。しかし単一の調査を超え、世界銀行などから提供される社会経済的情報をあわせてデータを分析すると、カテゴリーがまったく異なる二つの乱雑なデータを扱わなければならなくなります。時には、国の名前すらデータセットごとに異なることがあるのです！

世界開発指標は、例えば国ごとの ISO コードを使っているが、いくつかの国におけるその使用は IMF とは異なります。世界銀行のオフィスは IMF の隣にあるにもかかわらず、両者は小さな違いを生み、それがどんどん積み重なった結果、両方のデータを同時に使うことはほぼ不可能になります。ウィックハンが言うように、データ分析にかかる時間の 80% は、データのクリーニングと準備のプロセスに費やされます。首尾よくクリーニングが終わり、整ったデータが用意できたら、残りの 20% の時間で強力な統計プログラムを使って分析できますが、多くの場合、この 80% の部分が問題になります。

どうすればよいのでしょうか。わかりませんが、私たちにできそうなことをいくつか提示したいと思います。特に個性的でも独創的でもありませんが、調査データのフォーマットにも、統計データのフォーマットにも、一貫性を持たせるべきだと思います。

データを扱う人全員が、まずはそれぞれの分野で、データの分析力が向上するよう学生を指導するべきだと思います。おそらく、調査データから始めるべきでしょう。

しかしその後は、単一の調査データの使用を超え、調査データをさまざまな組織から提供される社会経済的データと組み合わせるかもしれません。私たちがアジア・バロメーター・プロジェクトに取り組んでいたとき、多くの著者が論文を執筆するためにアジア・バロメーターのデータ分析を求められましたが、残念ながら、アジア・バロメーターのデータを他の組織が提供するデータと組み合わせることは、どちらかと言うと彼らにとって困難な作業でした。

多くの場合、調査データは他の社会経済的データと組み合わせた方が有効利用できるはずです。皆さんが日本人に、幸せか不幸せかを尋ねたとしましょう。彼らは、「そうですね、あまり幸せではありません」と回答するかもしれません。しかし同じ質問をタイ人にした場合、その答えは「私は幸せです」となるかもしれません。タイ人の回答パターンは日本人の回答パターンとは異なる可能性があり、こうした違いの理由は何かを調査することで、いくつかの文化的変数や社会経済的要因の重要性が示唆されるかもしれないと思います。こうした要因はこの調査データそのものからは得られないでしょう。つまり各国の開発庁や政府が提供するデータを追い求めるべきなのです。

4番目の問題は、このテキストデータベースを用いて何を行おうとしているかです。大部分のテキストデータベースは、正しいテキストを見つけるという、一つの具体的な目的で使用されます。何百もの首相のスピーチ、何百もの外相のスピーチなど、そうしたテキストを、何らかの大きな社会的・経済的動向を語るための本当に重要なデータの源として使ってはどうでしょうか。私が思うに、テキストの定量化は将来的発展にとっての課題です。最近では、コンピュータープログラムが次第に日本語や中国語の形態素解析をできるようになっています。以前は、日本語や中国語については言葉を構成するものを判断し、文章を区切ることが難しかったため、言語を分析するのは困難でした。しかし今では形態素プログラムの能力がかなり向上しています。そのため、テキストの定量分析ができるのです。

要するに、ビッグデータの混沌とした巨大なまとまりから整ったデータを作り出すことは私たちの課題です。ご清聴ありがとうございました。

園田茂人

田中先生、ありがとうございました。次に山本先生から議論に対するコメントを伺いたいと思います。

山本信人

私は質志向の研究者であって、量志向の研究者ではないので、なぜここでコメントを求められるのか不思議に思っております。私は自分を地域専門家、特にインドネシアの専門家だと思っています。こうしたあらゆる調査データ分析やビッグデータの分析には詳しくありませんが、それでも私の学生の中には現在こうした新しいタイプの研究に興味を持っている者もいます。そのため今日ここでは、地域専門家の視点から基本的な問いを提示したいと思います。

吉野教授が、最近の官僚は調査データを軽視しているとおっしゃいましたが、調査データはそれでも公共政策や学術分野で役に立つものです。そのため、ホームズ教授が発表で提案されたように、方法論を改善すること、特に説明能力を向上させることが重要です。

東南アジア研究の文脈では、調査データ研究は 21 世紀に急速に発展しました。これは政治的または社会政治的な状況と関係があります。それはフィリピンの民主化によって 1980 年代半ばから始まり、韓国、台湾、インドネシアがその後に続きます。政治的・社会的状況が変化する中で、調査データ分析の重要性と人気が高まりました。これは外国の政府と外国の支援によって加速されました。

調査データ分析の主要研究分野の一つに選挙の研究があります。インドネシアの場合、日本と米国の政府援助のおかげで調査機関が設立され、世論調査が実施されています。以前は一つか二つだった調査機関が、今では数十カ所になっています。こうした機関は取引を巡って競争し、顧客の利益のために仕事をします。時には、顧客に有利なようにデータを操作することさえあります。政党によっては自分たちのデータ分析調査機関を所有し、自分たちにとって非常に必要な、ただし一般や研究者にとっては無用かもしれないデータを常に提供しています。

この新たな傾向によって、現在私たちは一、私たちというのは地域専門家のことですが、科学者やメーカー、データ調査の設計者と協力する方向に進んでいます。しかし、このような新しい世代の学者はなかなかいません。インドネシアで研究を進める人は、1990 年代や 21 世紀初めに米国や英国で教育を受けた後、民主化を受けて帰国し、自ら事業を始めています。ただ残念なことに、彼らはお金のために仕事をするのです。教授職には就いていません。インドネシアでも外国でも教授職が見つからないため、利益を上げる事業をスタートさせました。非常に高い技能を備えた学者だということに、ある意味、悲しいことです。

大部分のデータ収集と分析は国ごとに実施されているため、国際比較調査の実施は困難です。彼らは国の目的のためだけに仕事をしており、アジア・バロメーターやユーロバロメーターなどの場合を除き、国際比較調査を行うのは難しいのです。また、データを分析するには、ある国を専門とする人、またはその国に詳しい人が必要です。そうでなければ国を比較することは非常に難しくなります。

3 点目は、インドネシアでは政治的目的とビジネス上の目的から、調査研究が大きなビジネスになっていることです。知り合いの研究者たちは、いつも方法論を話題にするのですが、彼らの言語は私の言語とはまったく異なります。たいていの場合、彼らがどのような言語について話しているのか、私には理解できません。しかし、認識論の問題は置き去りです。このような方法で、人はあるテーマや環境をどう認識するのでしょうか。昔から地域専門家は、自分が研究対象としている国をどうすれば理解できるか、と問いかけてきました。これが、従来地域専門家が研究上の問いにアプローチする方法です。しかし、調査データを分析する人々はこれを何年も見過ごしてきたに違いありません。

発表者の方にお尋ねしたいことがあります。一つ目は国際比較調査の実施方法、二つ目は調査の信頼性についてです。

将来、データへのオープンアクセスは、特にインフラが貧弱な開発途上国の学術環境に身を置く者にとって、非常に重要になると思います。フィリピンとインドネシアの場合、若者はインターネットを自由に使えません。新しいモデルを使いこなす高度なスキルもありません。装備一式がなくてはどんな研究もできないため、こうしたあまり進んでいない国の若者と学者がデータに容易にアクセスできるようになれば、新たな研究領域が開けるでしょう。

また、データベースの作り方やデータベースの分析方法に関するオンラインフォーラムを用意することができれば、やはり新たな研究領域が生まれるでしょう。これは、Facebook のようなソーシャルメディアについて私がすでに始めていることです。人々はフォーラムを作って政治的、社会的、経済的な問題について話し合っています。インフラはすでにあります。私たちの仕事は、新たなアイデアをもって議論できるようなデータを提供することです。これがアジア研究の将来の方向性を提示するのではないかと思います。以上です。

園田茂人

山本先生、ありがとうございました。では、4名の発表者の方に、当学会の2名の教授がコメントした内容にお答えいただきたいと思います。

田中教授は、トライアングレーション手法が非常に難しくても重要だということをお話されました。山本教授は国際協力の重要性をお話されました。お二方のコメントにお答えいただけますでしょうか。

吉野諒三

田中先生、ありがとうございました。データ公開を望まれるということですが、ぜひほしいと思います。先生がおっしゃるトライアングラー戦略、というか…

田中明彦

トライアングレーション手法ですね。

吉野諒三

複数の異なるデータセットを組み合わせることで、新しい種類の分析ができるようになる、というのは本当です。これは、日本でデータアーカイブを構築するよう人々を説得するために重要なことで、それは私が何度も強調している通りです。しかし現実には、Google ネットワークで得られるようなソーシャルビッグデータを使用する場合、個人識別の度合いが高くなるため、非常に慎重になる必要があります。当研究所のデータとある種のデータを組み合わせることで、別の新たな問題が生まれる可能性があるため、それについては非常に慎重にならざるをえません。

園田茂人

異なるデータセットを組み合わせる場合の否定的な側面をご指摘いただき、ありがとうございました。

金 碩鎬

異なるタイプのデータを組み合わせることについて、そのプロセスのすばらしい面を強調したいと思います。社会学者である私たちは、常にデータを使って社会現象を説明しようとしています。個人情報を保護できさえすれば、例えば構造的な特徴、個人の特徴や特性などに基づく個々の従属変数を説明する際に、階層的線形モデリングを使うことができます。

KOSSDA では地方自治体に関する情報を提供しようとしています。韓国には 234 の地方自治体があり、ある地方自治体における大卒の割合や移民の割合など、地方自治体の特徴や構造的な特性をユーザーに提供しようとしています。

異なる種類のデータを組み合わせることに関しての 2 点目は、私たちは韓国の Twitter ユーザーのあらゆる情報を持っているということです。実は当機関は、韓国

の Twitter Inc.からこのような情報の提供を受けている他、ソーシャルネットワークの分析を専門とする Cyram と契約を交わしています。彼らはネットワーク分析のために、UCINET や Pajek のような NetMiner を考案しました。それで、私たちが行うのは、社会調査のために個々のユーザーを特定することです。ビッグデータと調査データを合わせて使えるよう、Twitter のテキストから得る情報と実際の調査における回答を組み合わせようとしているわけです。

ロナルド・ホームズ

私は国際比較調査についてコメントしたいと思います。

国際比較調査を行う項目は多数あり、例えば、人々がどちらかと言えばポピュリストなのかリベラリストなのか、などがわかりました。私たちが試したこのモジュールは、ある国の国民がポピュリストかリベラリストかをテストするために欧米で開発されたものです。設問を翻訳するのは一苦勞でした。というのは、フィリピンでは多様性の概念が「自由」を意味するからです。他のアジアの人々にそれが理解できるか、そして私たちの経験を多少なりとも反映するようにこれを翻訳できるのかどうかはわかりません。

欧米では、多様性が話題になるとき、人はそれが何を意味するのかを理解しています。しかし多くのアジア諸国では、多様性の概念は差異を意味し、あまりにも抽象的なため、その意味するところについて合意できないかもしれません。国際比較調査では、なんとか合意できる質問を増やしていかなければなりません。調査の設問のいくつかは、必ずしも相互に対応していない可能性があります。フィリピンで理解する内容がインドネシアで理解される内容とは対応しないでしょう。背景と経験が異なるからです。

王 衛東

私も、異なる種類のデータを組み合わせる方法について例を挙げたいと思います。

中国教育追跡調査のサンプル数は2万を超えており、国民を代表するサンプルになっています。この調査では、調査データと生徒の医療記録データを組み合わせられるよう、調査対象となる生徒の親、教師、校長に協力を要請しました。

同意を得るのは大変でしたが、生徒の 95%近くが入院に関する情報を送ってくれました。また、学生の質問票データを、人柄、意欲、自己効力感を含む心理テストデータと組み合わせました。

このようにデータを組み合わせる目的の一つは、中国で次第に幅を利かせつつある近視眼的な姿勢を避けることでした。調査を実施するため、各生徒にリストバンドをプレゼントしました。このリストバンドを利用すると、彼らがどこにしようとその移動と居場所に関する情報を継続的に収集することができます。また、生徒の行動を分析するため、WeChat や QQ におけるソーシャルメディアデータも提供してもらおうと頼みました。

もう一つの例は、国民全員の銀行口座を有する中国人民銀行の場合です。私たちはこのデータを基に研究を行い、ランダムに選んだサンプルに質問票を送りました。口座情報を除くすべての個人情報情報は消去し、プライバシーを維持しました。

園田茂人

ありがとうございました。では全体での議論を開始しましょう。客席からも質問を募りたいと思います。何か質問はありますか？

益尾知佐子

九州大学の益尾知佐子と申します。中国の外交政策を専門としており、普段は定量的データではなく、定性的データを扱っております。将来的に中国の研究者の方と協力できないかと考えていまして、王教授に2点質問したいと思います。

1点目として、将来の国際的な合同研究について、どのような条件でならば、中国政府が中国チームと外国のグループとの協力に対して許可を与えたとお考えですか？もう1点として、そうした関係者や政府とどの程度の協力あるいは情報交換を行っていますか？

王 衛東

一つ目の質問については、実際のところ非常に難しいです。中国人ならわかるように、国際的な合同研究を実施したいと思うなら、政府機関から許可を得なくてはなりません。我が国のシステムでは、許可を申請する際、ある種の許可証があれば効果的ですが、これまでのところまったく許可が下りていません。中国人民大学の私の研究室の壁には許可証が掛かっていますが、利用できたことはありません。とりわけ重要なのは、外国人に対する資金は得られないことです。そのため、中国で資金を得たいなら、協力相手になりそうな人で、自分と共通の関心を持つ人を見つけようとするのです。そうすればこのようなソーシャルネットワークを利用することができます。これが私からの助言です。

二つ目の質問については、中国ではほぼ不可能だと思います。私たちにとってさえ、ますます難しくなっています。ですから、将来何ができるかはわかりませんが、現在の政府は状況を変えすぎたと思います。

園田茂人

他に質問はありますか？

大庭三枝

東京理科大学の大庭三枝と申します。ASEANを含めたアジアの地域機構の発展について研究しています。そこでホームズ教授に伺いたいことがあります。まず、データセットを獲得する、または構築するために、他のASEAN加盟諸国における研究機関や大学と国際協力を行うプロジェクトはありますか？次に、このような国際協力プロジェクトがない場合、東南アジア諸国における社会学のデータセット構築の状況をご存知ですか？

ロナルド・ホームズ

一つ目の質問についてですが、オーストラリア国立大学を拠点として、金権政治に注目するプロジェクトを行いました。ここには実験的な調査も含まれます。調査は、マレーシアでは調査機関のMerdeka Center（ムルデカ・センター）が、インドネシアではLSI（Lembaga Survei Indonesia）が、フィリピンではパルス・アジアが実施しました。タイでも実施する予定でしたが、クーデターが起こってしまいました。これが協力の例です。

東南アジアはかなり多様なので、東南アジアの広い範囲における調査機関との協力は行っていません。国名は挙げませんが、ある ASEAN 加盟国に滞在している間つけられていると感じました。それで思うに、結局、私は無事に帰国したのですが。

Social Weather Stations は、先ほどお話ししたように国立台湾大学におけるアジア・バロメーター、ISSP、世界価値観調査、選挙研究調査を含めた国際協力プロジェクトに関わってきました。私は SWS と仕事をしていたので、私たちは協力関係を維持しています。私たちが行ったいくつかの研究は、必ずしも東南アジアの機関とはなく、むしろ多国間機関で行ったものです。ASEAN は、データ生成に関しては最も協力的でない機構です。

園田茂人

2013年に東京大学と早稲田大学の学部生がアジア学生調査（Asian Student Survey）の第二波調査を行ったのですが、その調査の興味深い結果の一つに、ASEAN の学生は米国と北東アジアに対して比較的肯定的な見方をしている一方、近隣諸国に対してはどちらかというと否定的な見方をしている、ということがあります。

王 衛東

中国における国際協力について、付け加えたいことがあります。

中国におけるいくつかの大規模な社会調査プロジェクトについては、海外の研究者にオープンモジュールの機会を提供することが容易なはずで、ある設問が中国研究にとって非常に重要だと考えるなら、その提案を中国総合社会調査、中国家庭追跡調査、中国教育追跡調査に提出すればよいのです。英語で提案を送るのです。その提案が受け入れられれば、データが入手できます。お金を払う必要はありません。中国政府がフィールドワークのコストを負担します。それが最も簡単な方法です。

園田茂人

王先生、すばらしい情報をありがとうございます。

大変充実した議論ができたと思います。本セミナーは録音しており、書き起こす予定ですので、この話題に関心をお持ちの方は本日の議論からインスピレーションを得られることと思います。

これで檜山セミナーを終了いたします。ご清聴、ご参加いただき、誠にありがとうございました。