

〈依頼論文〉

# ここ 20 年間の社会学による 中南米地域研究への貢献度の潜在意味解析

Latent Semantic Analysis of Sociologists' Academic Contributions to  
Latin American Studies during the Last Twenty Years

東京大学大学院総合文化研究科 和田 毅

Takeshi WADA (The University of Tokyo)

---

## Abstract:

This paper presents a comparative analysis of the influences of sociology, economics, and political science over Latin American studies. It applies a natural language processing technique called “latent semantic analysis” to measure such influences and provides a unique systematic evidence of the relationship among academic disciplines. Latent semantic analysis uncovers shifting influences of the three disciplines over Latin American studies during the period between 1990 and 2012. This paper argues that the rise and fall of the influence of political science and economics can be explained by salient political and economic issues which Latin American countries had encountered at each historical juncture. It also claims that the unique patterns of sociological influence over Latin American studies reflect the distinctive characteristics of sociology as a discipline. Latent semantic analysis will open up a new field of research for the social scientists and area specialists.

---

## はじめに

中南米研究、アジア研究、アフリカ研究などの地域研究は、人類学、経済学、政治学、地理学、歴史学、文学、心理学など様々な学術分野の専門家が活躍する刺激的なフィールドである。当然のことながら、社会学の専門家もその理論的・方法論的な分析ツールを活かして地域研究に貢献してきた。この論文の主な目的

は、社会学が中南米地域研究に及ぼしてきた影響を、社会科学の主要なディシプリンである政治学と経済学の影響と比較検討することである。

1990 年以降今日までの 20 年間に、社会学が中南米地域研究に及ぼしてきた影響はどのように変化しているのだろうか？政治学や経済学の影響と比較して、どの程度の貢献をしてきたのだろうか？時代によって、貢献度は変化しているのだろうか？ディシプリン間の影響については、これまでほとんど分析されてこなかった。しかし、政治学者と経済学者が主体のラテン・アメリカ政経学会が発行する『ラテン・アメリカ論集』に、社会学が成し遂げてきた成果を政治学・経済学のそれと比較しながら報告することは、今後の研究動向やディシプリンを超えた共同研究の可能性を考察するためにも、有用な試みであり意義深いことだと思われる。

筆者は社会学を専門とし、中南米地域の社会運動や民衆抗議行動を研究しているが、社会学者であっても、社会学が過去 20 年間に中南米地域研究にどのような影響を及ぼしてきたかという問いについて報告するのは難しい。なぜなら、この問いの答えを明らかにするためには、「中南米地域研究とは何か？」「社会学とは何か？」「ひとつの学問分野が他の分野に及ぼす影響をどのようにして示すか？」という、3つの困難な問題を解決しなければならないからである。

『ラテン・アメリカ論集』の主な読者が中南米地域研究の専門家だという観点からすれば、第一の問題（「中南米地域研究とは何か？」）について、ここで愚考を展開する必要はないにしても、それ以外の二つの問題については議論する必要がある。そこで、まず次の第二節で、「社会学とは何か？」という問題について論じることにする。様々な論じ方がありうるが、この論文では、私のアメリカの大学院での学生・教員としての経験をもとに説明する。社会学者を育成するために、どのような訓練を大学院生たちに行っているのかを知ることによって、社会学とはどのような学問なのかを具体的に理解することができると思われるからである。

次に、第三節・第四節では、「ひとつの学問分野が他の分野に及ぼす影響をどのようにして示すか？」という課題について議論を進める。第三節では、既存の研究がどのようにこの課題を明らかにしようとしてきたかを考察し、その問題点を指摘する。第四節では、著者がこの問題点をどう乗り越えて、この課題に取り組もうとするのかを説明する。具体的には、自然言語処理学の手法を応用して、社会学で用いられている言葉や用語がどれだけ中南米地域研究でも用いられているかを計測することによって、その影響の度合いを理解しようというのが、この

論文のアプローチである。それ以降の節では、具体的な手法を説明し、結果を報告する。

## 社会学とは何か？

社会学とは何か？この難題を簡潔に説明するために、ここではアメリカの大学院で社会学者を育成する際にどのような教育を行っているのかを説明する。アメリカの大学院を取り上げる理由は、著者がコロンビア大学で社会学の訓練を受け、その後ミズーリ大学で社会学を教えていたからであるが、この論文で用いるデータがアメリカを中心とする学術雑誌であることからそれが適当だと考えられるからである。社会学の訓練の仕方は、国によって、大学によって、また、教員によっても異なるだろう。しかし、コロンビア大学とミズーリ大学での経験を通じて、多くの共通点があることも分かった。図1はその共通点を図式化したものである。

図1：アメリカにおける社会学の大学院教育のパターン例



この図は、大学院一年次から、どのような授業科目を履修し、試験を受けながら博士論文執筆へと進んでいくのかを示している。もちろん、学生によって履修科目は異なるし、履修順序も一様とは限らない。しかし、図を見ることによって、社会学訓練の大きな流れは把握できるであろう。

まず、入学した院生は、理論と方法論を二本の柱としたカリキュラムを手にし

る。左側にあるように、理論については、最初の学期に古典理論を履修し、続いて現代理論を学ぶ。多くの大学では、これらの理論の授業を履修した後、二年次以降に Qualifying Exam と呼ばれる試験を受けて合格しなければ先に進めない。理論の授業では、社会を分析する際の大きな切り口・見方・アプローチについて学ぶ。古典理論では、カール・マルクスやアダム・スミスの著書に始まり、エミール・デュルケーム、マックス・ウェーバー、ゲオルク・ジンメルなどを読む。誰の著作から現代理論が始まるかは、実際には古典理論の授業がどこまでカバーできるかによる側面も大きい。一般的には、現代理論の授業では、タルコット・パーソンズなどの機能主義論、フランクフルト学派らの批判理論、ジョージ・ハーバート・ミードや H・G・ブルーマーらのシンボリック相互作用論、アーヴィング・ゴッフマンのドラマツルギー論、ハロルド・ガーフィンケルのエスノメソドロロジーなどを読んだうえで、ニコラス・ルーマン、ユルゲン・ハーバマス、ピエール・ブルデューなどを読んで終わる場合が多いのではないだろうか。

理論同様に重視されるのが、方法論である。どのように証拠を解釈するか、その妥当性を判断するかといった一般的な方法論的問題を扱う「方法論基礎」に続き、あるいは並行して、量的分析の訓練が行われる。大学によって、どの程度の量的分析の授業を課すかは異なるようだが、初級とそれに続くものを必修としている場合が多い。それと並行して、歴史的文書を社会学研究に取り入れる手法を学ぶ歴史社会学、事例研究などのロジックを学ぶ比較社会学、口述歴史学、フィールド・リサーチなどの質的分析方法論の授業も提供されている。

以上の理論と方法論が社会学を志す院生たちが共通に履修する、いわば基礎科目である。これらを習得した後、もしくはこれらの履修と並行して、学生たちは各自が関心をもつ専門科目を選択していく。図の上段二段に記された小さな四角は専門科目である。スペースの都合から 12 の科目名しか記載できなかったが、多くの専門分野が存在し、それぞれを専門とする教員が担当する。これらの授業の履修を通じて、博士論文の指導教官を決め、論文のテーマを練り上げていく学生が多い。方法論か、もしくは履修した専門科目の中から、第 2、第 3 の Qualifying Exam を選んで受験し、これにパスするといよいよ博士論文の執筆に入る。

専門分野は、基礎科目である理論や方法論の上に成り立っているものである。しかし、それぞれの専門分野のなかでも基礎的文獻や基礎的方法論があり、また、独自の学術雑誌を刊行している場合<sup>1</sup> も多いため、ときにお互いが何をやっているのかを認識できない傾向も見受けられる。

それでは、どのような専門分野が存在するのだろうか。また、人気のある専門分野は何だろうか。表1は、アメリカ社会学会のセクションを、所属する研究者数の順に並べたものである。セクションとは、専門分野・関心領域が近い人々の交流を促すために学会が公式に承認している組織である。セクションに所属するためには、学生や研究者は所定の会費を学会費に加えて支払わなければならない、新しいセクションを作るためには、200名の学会員を集めなければならない<sup>2</sup>。2013年1月の時点で52ものセクションが存在している<sup>3</sup>。文化（1位）やセックスとジェンダー（2位）は1,000名以上が所属する人気専門分野であることが分かる。組織論（3位）、医療（4位）、社会階層論（人種、ジェンダー、階級）（5位）、政治社会学（6位）、社会運動論（7位）、経済社会学（8位）も多くの社会

表1：アメリカ社会学会（American Sociological Association）の各セクションに所属する研究者数からみた社会学の主要な研究分野（2012年）

No	ASA Section Name	N	No	ASA Section Name	N
1	Sociology of Culture	1,181	27	Labor and Labor Movements	435
2	Sex and Gender	1,119	28	Methodology	434
3	Organizations, Occupation & Work	996	29	Sociology of Mental Health	428
4	Medical Sociology	992	30	Sociology of Law	423
5	Race, Gender, and Class	942	31	Sociology of Development	421
6	Political Sociology	858	32	Political Economy of World-System	412
7	Collective Behavior/Social Movements	841	33	Children & Youth	400
8	Economic Sociology	823	34	Asia/Asian America	337
9	Sociology of Education	818	35	Social Practice & Public Sociology	333
10	Racial & Ethnic Minorities	811	36	Communication and Information	325
11	Theory	802	37	Latino/a Sociology	318
12	Teaching and Learning	801	38	Human Rights	315
13	Family	754	39	Consumers and Consumption	313
14	Comparative/Historical Sociology	710	40	Marxist Sociology	311
15	Social Psychology	692	41	Body and Embodiment	306
16	Inequality, Poverty & Mobility	671	42	Altruism, Morality and Social Solidarity	305
17	Sociology of Religion	643	43	Peace, War, & Social Conflict	275
18	Global/Transnational Sociology	627	44	Sociology of Emotions	267
19	Aging and the Life Course	620	45	Disability and Society	236
20	Community & Urban Sociology	619	46	Mathematical Sociology	220
21	Crime, Law, & Deviance	612	47	Alcohol, Drugs and Tobacco	200
22	International Migration	593	48	History of Sociology	194
23	Science, Knowledge & Technology	496	49	Evolution, Biology & Society	174
24	Environment & Technology	491	50	Rationality and Society	162
25	Sociology of Population	490	51	Animals and Society	149
26	Sociology of Sexualities	485	52	Ethnomethodology/Conversation	129

学者が集まっている。これらの専門分野がある一方で、すべての社会学者が履修しているはずの理論（11 位）と方法論（28 位）はそれほど順位が高くはない。

社会学の専門分野は多様であり、社会学者の多くは、それぞれが選択した専門分野で、ある意味独自に発展している理論や方法論を参照しながら研究を進めていく傾向がある。社会学が対象としている領域が多様であるため、このように専門領域が多様化・細分化されていくのは当然の帰結なのかもしれない。しかし、マクロ経済学やミクロ経済学という柱がある経済学や、政治理論、政治方法論、比較政治学、国際関係論といった中核のある政治学と比較すると、社会学はより細分化されたディシプリンとなり、ときに専門領域間のコミュニケーションが欠如しがちな側面があることは否定できない。

このような特性を持つ（少なくともアメリカの）社会学が、中南米地域研究にどのような影響を与えてきたのだろうか。多様化・細分化されているため、社会学全体としての影響を語るのは難しいのではないだろうか。次節では、既存の研究がどのように「影響」を説明してきたのかをみていく。

## ディシプリンの影響をどうすれば把握できるか？

社会学というひとつのディシプリンが中南米地域研究という学問分野に与えた影響を、いったいどのようにして把握し提示することができるのだろうか？これは難題であり、既存の報告は多くはないが（Tavares-dos-Santos and Baumgarten, 2005）、その数少ない報告例として、1998 年に国際社会学会から出版された“Sociology in Latin America”（Briceño-León and Sonntag, 1998）が挙げられる。

13 編の論文を含むこの報告書によると、独自のジャーナルの発行、学会の発足、諸大学での社会学科の設立時期などから鑑みて、中南米に社会学がディシプリンとして根付いたのは 1950 年代代という。深刻な社会問題と不平等という中南米地域独自の社会環境を背景に、機能主義・実証主義的な立場の学者でさえ政治や社会変革に強い関心をもち、プロフェッショナルな学者というよりも知識人という自覚を備えていたという。このため、研究テーマも、権威主義、民主主義、マージナリティ、従属、革命、搾取、解放、国内植民地主義など政治色の強いものが多く、後にブラジル大統領となる従属論者のカルドーゾ（Fernando Henrique Cardoso）は、この流れを汲む代表的な社会学者であった。このような形で発展した中南米の社会学は、政治問題に学術・科学的見識を結び付けることができたという強みを持つ反面、学術的な仕事がイデオロギー的な言説と混乱してしまう

傾向がみられるという。採択される理論も、社会変革や教育変革をめざす社会集団の行動を制約する構造を重視する構造主義的傾向が強く、また、ヨーロッパ式の思想や学術を輸入する、いわば「ヨーロッパ植民地主義」的な社会学から脱却して独自の社会学を作り出せるのかどうか重要だとする (Lander, 1998; Quijano, 1998)。

このような議論は、中南米地域の社会学の発展の記述としてはスタンダードなものだと考えられる。しかし、あくまでも中南米の社会学者の研究のありかたや社会への貢献・影響を記したものであって、中南米地域以外の社会学者をも含むディシプリン全体の貢献・影響を語るものではない。また、社会学の影響を相対化するために、他の学術分野の影響と比較しようという視点もない。さらに、社会学の過去を振り返る際に、報告者自身の専門知識、学術的背景、個人的な見解に依拠してしまっている。もちろん、このテーマについて適任の研究者が執筆を依頼されているのであれば、その研究者の専門知識に依拠したレビューが悪いわけではない。しかし、社会学のように幅広く豊かなディシプリンが、中南米地域の多様な研究テーマに与えてきた影響を、ひとりの学者がバランスよく総括的にレビューすることは不可能であろう。例えば教育の社会学のように (Albornoz, 1998)、自分が得意とする社会学の専門分野に焦点を当ててしまい、不得手な専門分野における社会学の貢献・影響については触れずに終わってしまう可能性が高い。さらに、社会学が与える影響の度合いを、代表的な社会学者による著作によって示すなど、印象論の範囲を越えることができず、他のディシプリンとの相対比較などは不可能である。

この論文は、これまでの数少ないレビュー論文とは全く異なるアプローチを用いてこの問題に取り組む。特定の学者の知識に依存するのではなく、系統的にデータを処理していくことで、社会学が中南米地域研究に及ぼした影響を政治学と経済学のそれと比較しながら実証分析していく。

多様な専門分野に細分化されている社会学の影響をどう捉えたらよいのか。既存の研究報告のように一部の専門分野の影響にだけ注目するのではなく、社会学全体の影響というものを捉えることはできるのだろうか。この論文は、以下のアプローチを用いることによってこの問題を解決しようと試みる。それは、社会学全般が中南米地域研究に及ぼしてきた影響を測定するために、中南米地域研究の学術雑誌論文にどれだけ社会学の考え方や思想 (理論) や方法が入り込んでいるかを観察するというアプローチである。このアプローチの特徴は、社会学や中南米地域研究という学術領域を、それぞれの領域の専門家が投稿する学術雑誌に掲

載された論文で代表させる点にある。もちろん、この手法には、著書、学会発表、シンポジウムでの発表など、学術雑誌以外の形式による研究発表を考慮に入れないという欠点がある。しかし、研究者が凌ぎを削って投稿をする対象であるいわゆる「トップジャーナル」に掲載されている論文にこそ、各学術領域に代表的な（そして最も優れて先端的な）考え方・思想・方法論・問題関心などが反映されていると考えられる。

では、どのようにして社会学に代表的な考え方・思想・方法論・問題関心などを把握し、それらが中南米地域研究論文に「入り込んでいるか」を調べることができるのか。論文を一からすべて読み、代表的な考え方・思想・方法論・問題関心などを把握するというのは現実的な方法ではない。この論文では、学術雑誌論文に用いられている言葉や単語を、近年革命的な進歩を遂げている自然言語処理学の技術を応用して分析することで、この問題に対処する。詳細については後述するが、「自然言語」とは、コンピュータ言語に対して、人間が使っている言語を指し、「自然言語処理学」とは、話し言葉（音声）や書き言葉（オンラインの文書）を扱う主として工学系情報処理分野の学問のことである。自然言語処理（Natural Language Processing）は、Computational Linguistics や Corpus Linguistics などとも呼ばれ、言語情報をデータ（コーパスという）として扱い、これに統計・プログラミングの手法を適用して、データに内在する特徴をつかみ、様々な生活の場面に応用しようとする学問である。

学術雑誌論文に用いられている言葉や単語は、多様で豊かである。しかし、学術分野ごとに主要な概念・理論・手法・専門用語が異なるため、論文の中で用いられる言葉や単語の使用頻度には、当然違いがあるはずである。この論文では、学術雑誌論文で使用された言葉の頻度分布に着眼する。もし、社会学雑誌の言葉の頻度分布が中南米地域研究の言葉の頻度分布とほぼ一致したとすれば、二つの学術分野に大きな違いは存在しないと主張する大きな根拠となろう。もし、これらの二つの学術分野の言葉の頻度分布に相関関係が全く認められない場合には、二つの学問は全く関連のない研究をしている（つまり、相互の影響は全くない）と判断できるだろう。このように、論文で使用される言葉の使用頻度の相関度もしくは類似度を測ることによって、学問領域間の影響・関連を示すことができると考える。

もちろん、影響という概念には方向性が含まれる。言葉の使用頻度パターンが似ていた場合、それが、社会学が中南米地域研究に及ぼした影響を意味するのか、それともその逆なのか、それとも双方向の影響があるのかまでは、この手法では



判断できない。私見では、社会学的な訓練を受けていない中南米地域研究者が社会学のトップジャーナルに論文を掲載する例は記憶にないため、この論文のテーマに関しては、社会学から中南米地域研究への影響だと考えてもいいように思うが、影響力の方向性を論じるには限界がある手法を用いるという点を強調したうえで議論を進めたい。

表2は、この研究のために収集した学術雑誌の一覧である。中南米地域研究を代表する学術雑誌として、10種類を選択した。中南米の一部地域を主に扱う雑誌 (*Luso-Brazilian Review*, *Mexican Studies/Estudios Mexicanos*, *Caribbean Studies*) も研究対象とした。雑誌選択の基準として、Thomson Reuter による 5-Year Impact

表2：データとして用いた学術雑誌一覧

学術雑誌タイトル	収集年 <sup>*1</sup>	5YIF <sup>*2</sup>	論文数 <sup>*3</sup>
ラテン・アメリカ地域研究			
Journal of Latin American Studies	1990-2012	0.953	1036
Latin American Politics and Society/Journal of Interamerican Studies and World Affairs <sup>*4</sup>	1990-2012	0.731	1042
Latin American Research Review	1990-2012	0.442	1640
Latin American Perspectives	1990-2012	0.422	1118
Bulletin of Latin American Research	1990-2012		1171
Luso-Brazilian Review	1990-2012		659
Mexican Studies/Estudios Mexicanos	1990-2012		477
Caribbean Studies	1990-2012		186
European Review of Latin American and Caribbean Studies	1990-2012		280
Canadian Journal of Latin American & Caribbean Studies	1990-2011		311
社会学			
American Sociological Review	1990-2012	5.840	2376
American Journal of Sociology	1990-2012	5.113	1788
政治学			
American Political Science Review	1990-2012	3.759	1117
American Journal of Political Science	1990-2012	3.941	1301
経済学			
American Economic Review	1990-2012	4.076	4509
Econometrica	1990-2012	4.700	1607

\*1 2012年は10月5日現在でEBSCOデータベースから入手可能だった論文のみ。

\*2 5YIFは、5-Year Impact Factorの略で、引用情報から各学術雑誌の重要度を測ったものである。情報源は、2012年10月31日時点のThomson Reuterによる。

\*3 論文数はダウンロードした時点のもので、重複や書評を含む。最終的な分析ではこれらの重複や書評は除外した。

\*4 Journal of Interamerican Studies and World Affairsは2001年からLatin American Politics and Societyと名称を変更した。

Factor を用いる方針であったが、地域研究雑誌の場合にはその値が計算されていないものも多くあるため、これに依拠はしなかった。Web of Science など様々なオンラインデータベースを試してみたが、中南米地域研究雑誌については、EBSCO が多くを提供しているため便利であったため、このデータベースを選んだ。このデータベースが提供している雑誌で、かつ 1990 年から 2010 年代までの論文が手に入るものを選んだ。

EBSCO だけに情報源を限定した理由は、このデータベースは他のデータベースよりも、各論文に付与する「主題 (Subject Terms)」「類語 (Thesaurus)」「著者によるキーワード (Author-Supplied-Keyword)」が充実していたためである。論文のアイデア・概念・理論・方法論などを把握する一つの手法として、データベースが提供するこれらの情報を利用することが考えられる。EBSCO はそれ以外の論文データベース (Web of Science など) よりこの点が優れていた。さらに、複数のデータベースを利用しないメリットとして、データ収集の手順をパターン化できることも挙げられる。大量の情報を処理することを考えると、これは大きな利点であった。

「社会学」の学術領域を代表する雑誌として、*American Sociological Review* と *American Journal of Sociology* というトップジャーナルを選んだ。5-Year Impact Factor からみても社会学の頂点の雑誌だということができる。また、「政治学」については、*American Political Science Review* と *American Journal of Political Science* を、「経済学」については、*American Economic Review* と *Econometrica* の二誌を選んだ。社会学の場合も、政治学や経済学の場合も、トップジャーナルを見分けるために、5-Year Impact Factor を参考にしたが、政治学や経済学の場合にはいくつかの候補があったため、同僚の政治学者や経済学者たちの判断も同った。また、選択のもう一つの基準として、それぞれの学問の総合的な学術雑誌を選んだ。たとえば 5-Year Impact Factor が高い雑誌であっても、社会学の *Social Networks* のように、専門分野に特化した雑誌は今回の分析対象からは除外した。

このような雑誌の選択方法には問題もある。既に述べたが、雑誌論文以外の研究発表は対象から外れてしまうこと、EBSCO データベースが提供していない学術雑誌がデータとして含まれないこと、英語中心の学術雑誌のみが対象となってしまうことなどが挙げられよう。また、トップジャーナルとして別の雑誌を取り上げるべきだという批判もあるかもしれない。このような問題点があるにもかかわらず、この研究が採択する手法は、既存の研究報告の限界を超えて、社会学そして社会科学の地域研究への影響を可視化する斬新なアプローチであり、

将来の研究の新たな方向性を開くものだと考える。次節では、自然言語処理学の様々な技術のなかからこの研究で用いる「潜在意味解析法」について説明する。

## 潜在意味解析

潜在意味解析法—“latent semantic analysis” (Landauer, 1999) もしくは “latent semantic indexing” (Manning and Schütze, 1999, p. 554) —とは、「テキストを集めた大きなコーパスに統計的計算を適用し、言葉の文脈・用法による意味を抽出する理論と方法である」(Landauer et al., 1998, p. 259)。日常生活の中で、人は、話し言葉や書き言葉の豊かな表現から無意識のうちに類義語を想像し潜在的な意味、つまり、直接明確には発言・表現されてはいない意味を連想している。例えば、「病院」や「メディカルセンター」という表現を新聞記事に見つけた際には、「健康」だとか「医者」などの表記されてはいない語を連想できるだろうし、「イチロー」、「ジーター」、「ダルビッシュ」という表現からは、メジャーリーグの野球を想像できるだろう。「沢穂希」、「佐々木則夫」、「なでしこ」という表現からは、メジャーリーグよりも女子サッカーを思い浮かべることができるだろう。それは、潜在的な意味、つまり、言葉の意味連関の構造を理解しているからだと考えられる。この意味連関の構造——言葉の同義性や多義性——を抽出して、実生活に役立てることが、潜在意味解析法の目的の一つである。

この潜在意味解析法の特徴と利点を説明するため、1990年から2012年までの23年間に、社会科学の三大領域である社会学、政治学、経済学が中南米地域研究に及ぼした影響の分析例をとりあげる。実際の分析過程を具体的に示しながら研究方法を説明したほうが理解しやすいからである。

各ディシプリンの学術雑誌論文に用いられている言葉は、文学ほどではないにしろ、多様で豊かである。しかし、それぞれのディシプリンにおいて特有の概念、独自の理論、特異な学術用語が生み出され発展してきているため、学術雑誌論文に用いられているすべての言葉の頻度分布を求めれば、それはディシプリンによって大きく異なるものと予想される。経済学者が論文を書く際に用いる言葉のパターンは、社会学者のそれとは当然大きく異なっているであろう。

この研究の革新的な考えはこの点に着目したところにある。もし、社会学、経済学、政治学が学術論文の中で用いる言葉の用法の分布パターンに明確な違いがあるとすれば、中南米地域研究の学術雑誌で用いられている言葉の分布パターンを3つのディシプリンのそれと比較することによって、中南米地域研究が、三大

分野の中ではどの分野にもっとも近いのか（つまり、影響されているのか）を測ることができるのではないだろうか。もし、中南米地域研究で用いられている言葉の分布パターンが、経済学や政治学よりも社会学のパターンにより近いものであれば、それは社会学の影響が他のふたつのディシプリンよりも強いことを示すひとつの根拠になるのではないだろうか。もちろん、この方法自体は、ふたつの学問領域（例えば、社会学と中南米地域研究）間の影響について因果の方向性を示唆するものではなく、あくまでも類似性を示しているだけであることは、既に述べたとおりである。

これまでに、Landauer, Laham, and Derr (2004) が潜在意味解析法を用いて生物学の専門分野 (biochemistry, medical sciences, neurobiology, cell biology, genetics, immunology, biophysics, evolution など) の類似性を可視化する研究を行っているが、ディシプリン間の類似性や相違性を解析しようとするものは、著者の知る限りこの研究が初めてである。

では、具体的にどのように潜在意味解析法を実行すればよいのだろうか？最初のステップは、分析の単位となる「文書 (Document)」と「語 (Term)」を決めることである。文書は、語・言葉の分布を比較する単位となるため、ここでは中南米地域研究、社会学、経済学、政治学の4つのディシプリンを文書とする。つまり、表2の学術雑誌に含まれるすべての論文をそれぞれの学術分野ごとにまとめた4つの大きな文書が存在すると考えればよい。

語の選択肢としては、(A) 論文で用いられたすべての語を分析に用いる、(B) 論文の要旨 (abstract) に使われている語を用いる、などが考えられる。ここでは、EBSCO データベースで入手した論文情報の中の主題項目 (Subject Term; Thesaurus Term; Author-Supplied Subject Term) を分析の単位として用いることにした。その理由は、主題項目に記されている言葉は論文の中核的な考えを端的に示していると考えられるからである。表2の学術雑誌リストに掲載されている雑誌論文の主題項目情報をすべて収集し、その頻度を数える。ここでは、分析方法を分かりやすく説明することに主眼を置いているので、最も頻繁に出現した10の主題項目のみを分析に用いることにする。つまり、10の主題項目がここでの分析の Terms に相当する。

第2のステップは、表3のような「語—文書表 (Term-Document table)」を作成することである。表3は10の語 (主題項目) がそれぞれの文書 (ディシプリン) に何回出現しているかを示したものである。語—文書表は語の分布パターンを表し、ディシプリンごとにどのように主題項目が使い分けられているかを示してい

表 3：潜在意味解析：語—文書表

語	中南米 地域研究	経済学	政治学	社会学
econometrics	0	314	1	0
economic aspects	20	217	17	8
economics	18	629	15	14
latin america	194	0	9	2
mathematical models	0	242	10	5
mexico	298	7	5	3
political science	70	27	252	20
politics & government	243	11	137	5
research	21	178	46	44
united states	122	138	546	185

る。学術分野によって用語の使い方が違うのが一目瞭然である。“Mexico,” “politics & government,” や “Latin America” といった語は中南米地域研究の論文で頻繁に用いられているのに対し、社会学の論文では “United States” や “research” が最も多く登場する主題項目となっている（これはアメリカ合衆国の社会学雑誌をデータに用いていることに起因している）。経済学者は “econometrics” や “mathematical models” を語らずにいられないのに対し、中南米地域研究の専門家は一度もこの主題項目が付けられるような研究をしていない（もちろん、選択された雑誌においてであるが）。

さらに、この表は語の共起 (co-occurrence) パターンも表している。例えば、“United States” という主題項目は政治学と社会学の文書で頻繁に登場している。ふたつの文書間で共起状態の語が多いほど、それらの文書の類似性が高いと判断できる。一方で、“econometrics” という主題項目は経済学では頻繁にみられるが社会学では一度も使われていない、いわばゼロ共起状態 (zero co-occurrence) にある。共起状態の語が少ないほど、ふたつの文書の類似性は低いと結論付けることができよう。

この表を眺めるだけでも面白いが、それは、語数や文書数が少ないからである。語数や文書数が多くなると表を見ただけで意味の連関を把握するのは困難になる。潜在意味解析法は、語数や文書数が多い場合でも、語の共起情報を活用して、意味の近い語群を検出し、その潜在的意味を表す次元を抽出することができる。つまり、多くの語と文書で（多次元で）構成されている雑多な情報を、意味の構造を把握することによって、より少ない次元（潜在意味空間）で効率的に要

約するのが潜在意味解析法なのである。第 3 のステップは、要因分析（特異値分解とも呼ばれる）という統計手法を適用して、潜在意味の次元を抽出することである。

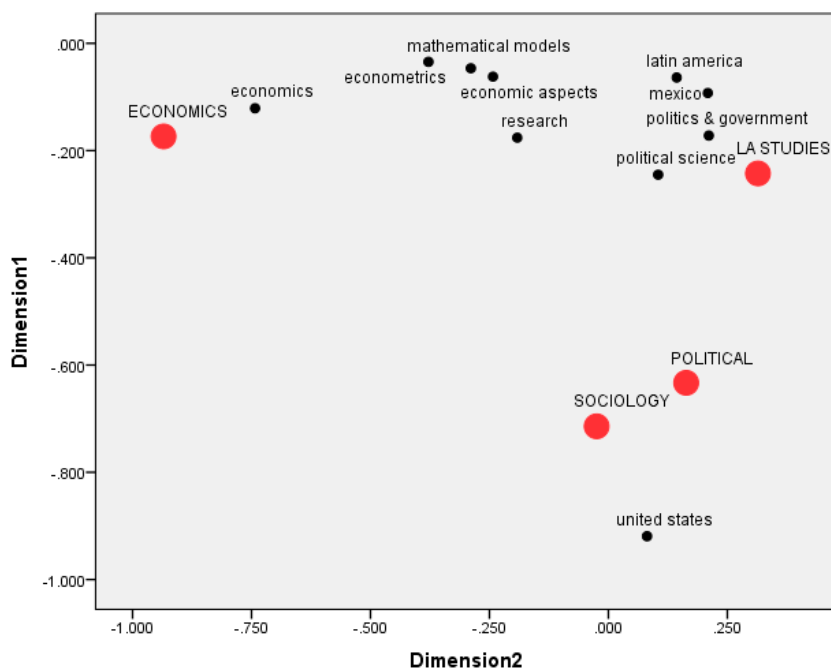
要因分析は、語一文書の共起情報を入力データとして用いる。具体的には、各語（主題項目）がそれぞれの文書（ディシプリン）に出現する相対頻度（%）を「観測変数」とし、これを入力データとする。4 種類の文書があるため観測変数も 4 つ存在する。要因分析では、観測変数は、観測することはできない潜在的な要因（ここでは潜在的意味）によって生じた結果として観測されるものであり、いくつかの潜在要因の線形結合として表すことができると想定している（Kim and Mueller, 1978, p. 8）。要因分析は、このような潜在的な要因を、統計的計算を用いて数値化する。つまり、直接観測できない要因を、それに関連する複数の観測変数を用いて可視化する手法なのである。

図 2 は要因分析の結果である。4 つの観測変数から 2 つの要因（潜在意味次元）が抽出された。そのふたつの要因を X 軸、Y 軸として潜在意味空間を図式化したものである。この空間に、10 語（主題項目）と 4 文書（ディシプリン）を、それぞれの要因との相関値によってプロットしている。ふたつの潜在的意味の次元で構成される空間に、語と文書を投影することで、意味の近い語群、内容が類似している文書群、そして文書に頻繁に用いられる語が一目瞭然になる。

まず、ふたつの潜在意味とは何かを推察するために、語の位置をみてみよう。第 1 の要因・潜在意味の次元（Dimension1）である Y 軸の大小の値をとる語を調べてみると、“United States”だけが特に小さい値をとっていることが分かる。つまり、アメリカ合衆国についての研究かどうかという次元だといえる。第 2 の要因・潜在意味を示す X 軸の分布をみると、“economics,” “econometrics,” “mathematical models,” “economic aspects”などがマイナスの領域に位置し、それ以外の主題項目と離れていることから、第 2 の潜在意味は経済の研究かどうかを示す次元だといえるだろう。

次に大きな丸で表示されている文書の位置を見てみよう。社会学は経済学や中南米地域研究（LA STUDIES）よりも政治学（POLITICAL）に近い位置を占めていることから、政治学にもっとも内容が似ていることが分かる。中南米地域研究は、第 2 の要因・潜在意味（Dimension2）では経済学よりも社会学や政治学に近い位置にあるが、第 1 の要因・潜在意味（Dimension1）では経済学のほうに近い。経済学と社会学・政治学双方と似ている側面があることから、両方のディシプリンから影響を受けていることを示しているものと考えられる。

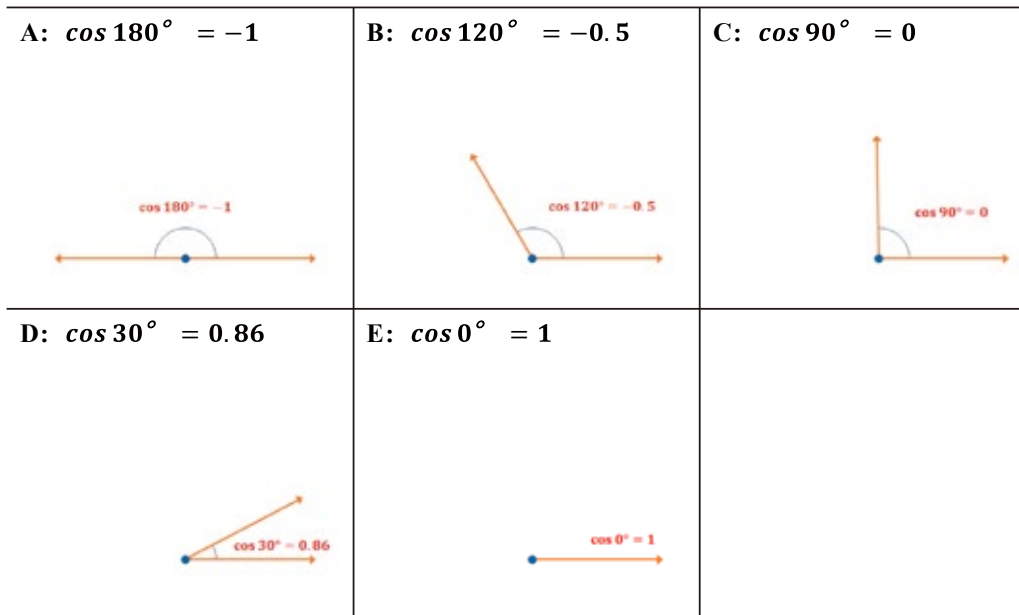
図 2：潜在意味解析：抽出された二つの次元(潜在意味)による語と文書の散布図



次に、語と文書の関係に着目してみると、経済学（文書）に近い位置を占める語は“economics,” “econometrics,” “mathematical models,” “economic aspects” などであるのに対し、社会学と政治学の近辺には、“United States” が位置するだけであった。アメリカの学術雑誌をデータとして利用しているため、社会学や政治学の論文の圧倒的多数がアメリカ合衆国についてであるために、このような結果となったのであろう。さて、中南米地域研究の近くには、当然のことながら、“Latin America” や “Mexico” が位置しているが、これら以外に “politics & government” や “political science” が位置している。政治を示すこれらの語が、政治学よりも中南米地域研究の近くに位置していることは興味深いが、おそらく政治学の学術雑誌掲載論文ではこれらの主題項目はあまりにも当たり前なので用いられない傾向があるからだと思う。

このように、潜在意味空間のグラフを描いて分析する方法はとても直観的で理解しやすいが、これが可能なのは潜在意味がふたつしかなく語数も限られているためである。3つ以上の次元の場合、図2のような単純なグラフでは記述できなくなる。また、2次元のグラフであっても、語数が増えるに従って判読不能なグ

図 3：潜在意味解析におけるコサイン類似度



ラフになってしまうだろう。では、語数（主題項目）が数百、数千、数万に増え、文書数も多くなり、潜在意味も 3 つ以上抽出された場合は、どうすればいいのだろうか？

潜在意味解析は、たとえ語数や文書数が大量にある場合でも、また、多次元の潜在意味空間が構成されている場合でも、潜在意味の構造（要するに、①語と語、②文書と文書、そして③語と文書の意味空間での近さ）を、コサイン類似度を使って特定することができる。図 3 はコサイン類似度を説明したものである。矢印は多次元潜在意味空間に占めるひとつの文書（もしくは語）の位置を表しているとする。A では、ふたつの文書（たとえば、社会学と中南米地域研究）の矢印が全く逆方向に向いている。このような場合、ふたつの文書は完全に異なる内容だということになる。ふたつの矢印で構成する角度は 180 度となり、コサイン類似度は最小値  $-1$  をとる。ふたつの文書が完全に同一の場合、E のようにふたつの矢印は完全に重なり、その角度は 0 度、コサイン類似度は最大値  $+1$  になる。コサイン類似度は、 $-1$  と  $+1$  の間の値を取り、大きい値ほど類似度が高くなる。

表 4 は 4 つの文書（ディシプリン）のコサイン類似度を示したものである。中南米地域研究にもっとも類似しているディシプリンは政治学 (0.789) であり、もっ



表 4：4 つのディシプリン（文書）のコサイン類似度

ディシプリン	中南米 地域研究	経済学	政治学	社会学
中南米地域研究	1.000	-0.667	0.789	0.583
経済学	-0.667	1.000	-0.069	0.216
政治学	0.789	-0.069	1.000	0.959
社会学	0.583	0.216	0.959	1.000

とも異なるディシプリンは経済学（-0.667）となった。図 2 から予想されるように、社会学は経済学（0.216）や中南米地域研究（0.583）よりも政治学（0.959）に近いことが分かる。スペースの都合から省略するが、語（主題項目）の潜在意味空間上の距離も、コサイン類似度を用いて示すことができる<sup>4</sup>。

潜在意味解析による文書間・語句間の類似性の分析結果はどの程度信用できるのだろうか？既存の研究によると、潜在意味解析の結果と、人が実際に文章を読んで同様な判断を行った結果との間に統計的に有意な差はなかったという (Wolfe and Goldman, 2003, McCarthy et al., 2007, p. 111)。類似度を数値化し、人間ではとても読めない大量の文章を解析することができる潜在意味解析法の利点は明らかである。

さて、次節ではこの研究の主な問いの結果を明らかにする。1990 年以降今日に至るまで社会学者は中南米地域研究にどのような影響を与えたのだろうか？その影響は政治学や経済学のそれと比べて大きかったのだろうか？

## 中南米地域研究と三大社会科学ディシプリンの時系列分析

前節の潜在意味解析では、分析手法を紹介することが目的だったため、EBSCO データベースから収集した論文に最も頻繁に出現した 10 の主題項目だけを分析に利用した。ここでは、出現頻度が 10 未満の主題項目を除いたすべての主題項目を分析に利用する。出現頻度 10 未満の主題項目を除外する理由は、あまり使われていない主題項目は、複数の文書間の共起状態を作り出す可能性は非常に小さい上、そもそも項目自体が重要でないと考えられるからである。前節では 1990 年から 2012 年までのデータをすべてまとめて分析したが、ここでは時系列変化を捉えるために年単位の分析を行う。

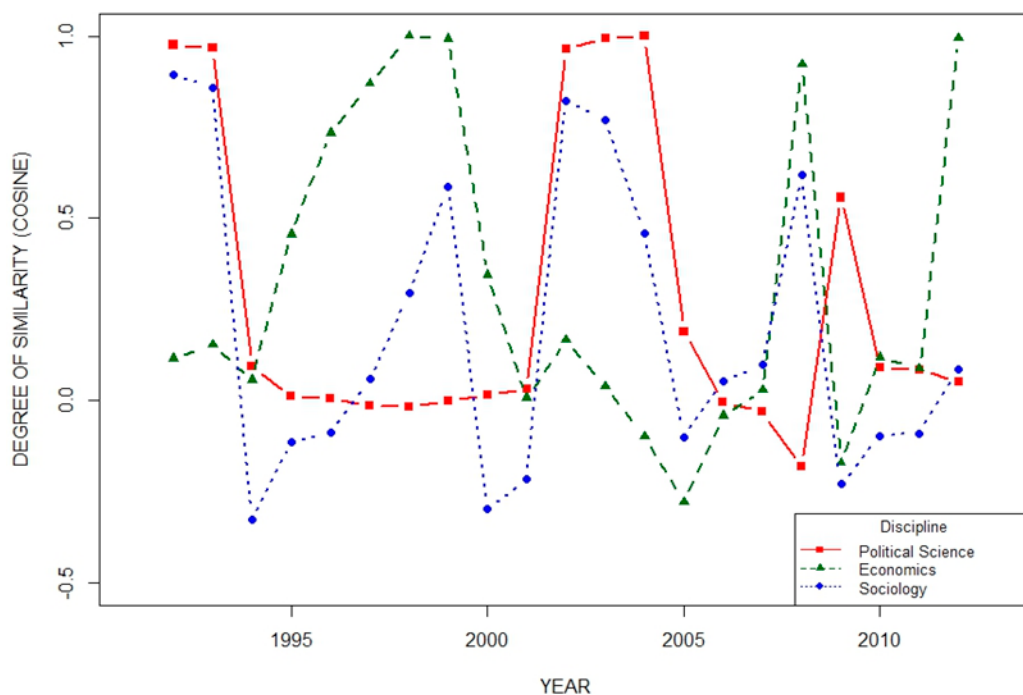
図 4 は、3 つの社会科学ディシプリンと中南米地域研究との類似性を比較したものである。株価のように上下の変動が激しい値の動向・トレンドを捉えるために頻繁に用いられる移動平均法の考え方を応用し、過去 3 年分のデータを 1 年ずつ移動させながら類似度を計算したものを Y 軸に表示している。

まず、政治学と中南米地域研究のコサイン類似度の歴史的変遷を見てみよう。1990 年代初頭までは政治学が中南米地域研究のけん引役となっていたようである。この時期のコサイン類似度が +1 に近いことから、政治学と中南米地域研究の主題項目のパターンは非常に近かったことが分かる。つまり、中南米地域研究の学術雑誌掲載論文と政治学の論文は同様なテーマを扱っていたということである。1990 年代初頭は、民主化やその深化というトピックが政治学者や中南米地域研究者の間で大きな関心と呼んでいた時期であった。これに対し、その後の 1990 年代は、これらふたつの学問分野の類似度が低下する時期が長く続く。2002 年から 2004 年になって類似度が再び上昇する。2000 年代前半に政治学が中南米地域研究にて再び主要な役割を果たすようになったのは、おそらく、“Pink Tide” と呼ばれる中南米地域における左翼政治・政治家の台頭や先住民運動の隆盛が、大きな関心を集めたことによると思われる。ベネズエラのチャベスが大統領になったのは 1999 年、ブラジル労働党のルラが大統領選挙に勝利したのは 2002 年であった。

次に、経済学をみてみよう。経済学は、1990 年代後半には中南米地域研究との類似度が最も高かった学問である。1994 年以降、次第にその中南米地域研究における影響力を増大している。この時期の中南米は、グローバル化や新自由主義経済に関連する多くの経済問題が表出していたことは周知のとおりである。1994 年の北米自由貿易協定 (NAFTA) の発効、1995 年のメキシコ通貨危機、1997 年のアジア通貨危機を端緒とするグローバル規模での金融危機、1999 年から 2002 年にかけてのアルゼンチン経済破綻など数多くの問題が注目を集めた。その後は、2008 年まで経済学と中南米地域研究との類似度が低迷する時期が続き、中南米地域研究における主要な関心が経済から政治へと移っていることが分かる。

潜在意味解析から明らかになったことは、政治学と経済学が中南米地域研究における主導的な位置を交互に占めてきたことである。そのある意味「支配的な」時期を考慮すると、地域が抱える主要な課題と密接に連動しているように思われる。地域の問題を解決する答えを提供する学問として、政治学や経済学が期待されていたとも言えることができるかもしれない。それでは、社会学はどうであろう

図 4：中南米地域研究と社会科学の 3 つのディシプリンの類似性の比較



か？社会学は、中南米地域研究のなかで最も重要なディシプリンの地位を獲得したことはなかったという結果が出た。図 4 をよくみると、2006 年と 2007 年に 3 つのディシプリンの中ではもっとも高い類似度を記録してはいる。しかし、これらの 2 年においてさえ、そのコサイン類似度はほぼ 0 であり、中南米地域研究にほとんど貢献していなかったことが分かる。これらの年のように、3 つのディシプリンすべての類似度が小さい年には、歴史学、文学、人類学など他のディシプリンの研究が中南米地域研究において主要であったと推測される。

図 4 に表れている社会学の特徴は、1992 年から 1994 年にかけては政治学と並行した動きをし、1994 年から 2000 年にかけては経済学と、そして 2001 年から 2005 年には再び政治学と、さらに 2005 年から 2012 年にかけては再び経済学と同様な動きをしている点である。この社会学特有の動きは、社会学というディシプリンの固有の特徴を反映したものだと考えられる。「社会学とは何か？」の節で説明したように、社会学とは様々な社会的、政治的、経済的、文化的現象を扱う非常に間口の広い学問である。多数の専門分野に細分化され、2013 年 1 月時点でアメリカ社会学会には 52 の「セクション」が存在し、その多くが専門分野に特化していることはすでに述べたとおりである。

扱う対象・現象が広範にわたることや、多種多様な専門分野が並立していることによって、社会学は多様な社会、政治、経済、文化の諸問題に対処できる（しようとする）学問となっている。このため、1992年から1994年や2002年から2004年のように、中南米地域にて政治問題が最大の関心事となり、それによって政治学者が地域研究の中で重要な役割を果たしていた時期には、政治社会学、社会運動、戦争と平和、社会抗争などを専門分野とする社会学者も政治学者同様に貢献していたために、社会学の類似度の線が政治学の線と並行した動きをしていると考えられる。1996年から1999年や2008年のように、経済問題が顕著となり、経済学者が地域研究の中で活躍した時期には、今度は経済社会学、組織論、不平等、貧困、グローバル社会学などを専門分野にする社会学者がこれらの経済問題を取り上げて貢献していたのであろう。このように、社会の抱える様々な課題に対処できるディシプリンとしての社会学の柔軟さを、潜在意味解析が捉えているのである。

## おわりに

この論文では、社会科学の三大ディシプリンが中南米地域研究に及ぼしてきた影響力を測るといふ、極めて独創的な分析を提示した。社会学、政治学、そして経済学が、それぞれのディシプリンの特徴と歴史的背景によって、独特な影響を中南米地域研究に及ぼしていたことを可視化することができた。潜在意味解析という自然言語処理技術を応用した分析は、まだ予備的分析の段階にあることは否めない。しかし、学問分野間の重要性・類似性・貢献度・影響力に関して、実証データを系統的に収集・解析したうえで議論するという、従来はほとんどみられなかった類の研究分野を開拓する潜在性を提示できたとは思う。

それでは、次のステップは何であろうか。まず、第一に、語と文書の分析であろう。政治学や経済学が主流であった時代の主な主題項目はどのようなものであったか？政治学と社会学が共有していたテーマは何であって、それぞれが独自に扱っていたトピックは何であったのか？語の分析を組み入れることで、時代背景とディシプリンごとの研究テーマの関係をより詳細に知ることができよう。

第二に、文書の類似性をディシプリンというレベルで比較するのではなく、学術雑誌レベルで比較するというのも興味深い。地域研究雑誌ごとに、取り扱うディシプリンの優先度や政治経済問題への対応の迅速性が異なると考えられ、潜在意味解析はこれを比較研究する手法となりうる。

第三に、時代をもっとさかのぼり、例えば第二次世界大戦以降の半世紀以上を分析することも意義のあることだと思われる。冷戦、キューバ革命、軍事政権・権威主義の台頭、累積債務危機、民主化など、中南米地域を左右した様々な時代背景が、どれほど地域研究に影響を及ぼしてきたのかを読み取ることが可能になる。

最後に、この論文では EBSCO データベースの主題項目を「語」として扱ったが、論文全体もしくは論文要旨に登場するすべてのワードを「語」として分析することが考えられる。主題項目を付与するのは EBSCO データベースの作成者や論文著者であるため、一貫性のあるキーワード作成が為されているとは言い難い。論文内容を数少ない主題項目で代表しているわけであるが、その正確性や妥当性を全ての論文について保証することはできない。この問題点を考慮すれば、著者自身が書いた要旨や論文そのものをデータとして扱う方が、より説得力があるだろう。その場合、“a,” “the,” “with,” “and” など、どのような文書でも多用される語（自然言語処理では stop word と呼ばれる）が文書間の共起状態を作り出してしまう問題を解決する必要がある。研究内容を表すワードを分析に用いて、それ以外のワードを除外したほうが、高い精度の分析結果を得られるはずで、そのため工夫が必要になる。また、“Latin America” のような語を、“Latin” と “America” の 2 語として扱っては意味が異なってしまうため、複数の連続する語によって意味を成す言葉をひとつの語として認識させるための優れたコロケーション分析 (collocation analysis) のプログラムも開発する必要がある。

このように、潜在意味解析は、中南米地域研究者にとっても、各ディシプリンの専門家にとっても、自らの研究分野についての新たな研究を開拓するものになると考えられる。

### 〈付記〉

本論文はラテン・アメリカ政経学会第 49 回全国大会（2012 年 11 月 10 日）での特別講演を元に行っている。この研究のためのデータ収集に尽力した出川永さん（東京大学大学院法学政治学研究科博士課程）に感謝の意を表したい。

---

注記

<sup>1</sup> 一部の例だけ取り上げると、*American Journal of Criminal Justice* (#21 犯罪、法、逸脱)、*Children & Society* (#33 子供 & 若者)、*Gender & Society* (#2 性とジェンダー)、*Population Research and Policy Review* (#25 人口の社会学)、*Sociological Methodology* (#28 方法論) などである (数字は後掲の表 1 の専門分野の番号)。詳細は、アメリカ社会学会発行の *Publishing Options: An Author's Guide to Journals* のなかの Index of Journals by Area of Interest を参照のこと (American Sociological Association, 2011, pp. 244-277)。

<sup>2</sup> American Sociological Association, “How to Form a New Section” (<http://www.asanet.org/sections/HowtoFormaNewSection.cfm>) . 2013 年 1 月 5 日アクセス。

<sup>3</sup> アメリカ社会学会のセクションに関する情報は、学会ホームページから入手した ([www.asanet.org/sections/Final2012Counts.cfm](http://www.asanet.org/sections/Final2012Counts.cfm), 2013 年 1 月 4 日アクセス)。

<sup>4</sup> 10 の語 (主題項目) のコサイン類似度を示した表は、著者から直接入手可能である。

参考文献

ALBORNOZ, O. 1998. Past and future of sociological analysis in Latin America: the case of sociology of education. In: BRICE O-LE N, R. & SONNTAG, H. R. (eds.) *Sociology in Latin America*. Madrid: International Sociological Association (Proceedings of the ISA Regional Conference for Latin America).

AMERICAN SOCIOLOGICAL ASSOCIATION 2011. *Publishing options: an author's guide to journals*. Washington, D.C.: American Sociological Association.

BRICE O-LE N, R. & SONNTAG, H. R. (eds.) 1998. *Sociology in Latin America*, Madrid: International Sociological Association (Proceedings of the ISA Regional Conference for Latin America).

KIM, J.-O. & MUELLER, C. W. 1978. *Factor analysis: statistical methods and practical issues*, Beverly Hills, Calif., Sage Publications.

LANDAUER, T. K. 1999. Latent semantic analysis: A theory of the psychology of language and mind. *Discourse Processes*, 27, 303-310.

LANDAUER, T. K., FOLTZ, P. W. & LAHAM, D. 1998. An introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284.

LANDAUER, T. K., LAHAM, D. & DERR, M. 2004. From paragraph to graph: Latent semantic analysis for information visualization. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5214-5219.

LANDER, E. 1998. Eurocentrism and colonialism in the Latinamerican social thought. In: BRICE O-LE N, R. & SONNTAG, H. R. (eds.) *Sociology in Latin America*. Madrid: International Sociological Association (Proceedings of the ISA Regional Conference for Latin America).

MANNING, C. D. & SCH TZE, H. 1999. *Foundations of statistical natural language processing*, Cambridge, Mass., MIT Press.

MCCARTHY, P. M., BRINER, S. W., RUS, V. & MCNAMARA, D. S. 2007. Textual signatures:

- Identifying text-types using latent semantic analysis to measure the cohesion of text structures. *In: KAO, A. & POTEET, S. R. (eds.) Natural language processing and text mining.* London: Springer.
- QUIJANO, A. 1998. The colonial nature of power and Latin America's cultural experience. *In: BRICE O-LE N, R. & SONNTAG, H. R. (eds.) Sociology in Latin America.* Madrid: International Sociological Association (Proceedings of the ISA Regional Conference for Latin America).
- TAVARES-DOS-SANTOS, J. V. & BAUMGARTEN, M. 2005. Contribuições da Sociologia na América Latina à imaginação sociológica: análise, crítica e compromisso social. *Sociologias*, 7, 178-243.
- WOLFE, M. B. W. & GOLDMAN, S. R. 2003. Use of latent semantic analysis for predicting psychological phenomena: Two issues and proposed solutions. *Behavior Research Methods Instruments & Computers*, 35, 22-31.

