

引用関係を応用した主題書誌拡張の試み

—IBSSとSSCIに基づいて—

○手塚 敬子

松井 幸子

岸田 和明

Enlarging Subject Bibliography through Applying
Citation Networks between Documents: Based on
the *IBSS* and the *SSCI*

○Keiko TEZUKA
Sachiko MATSUI
Kazuaki KISHIDA

ABSTRACT

The aim of this study was to find a method for enlarging the subject bibliography on "poverty," a multidisciplinary subject in social sciences. We investigated a method for retrieving related documents using semantic and citation retrieval methods. The sources of this study are the *International Bibliography of the Social Sciences (IBSS)* and the *Social Science Citation Index (SSCI)*. A set of 791 items on "poverty" (called Set P) were retrieved from the *IBSS* (including 129,568 items) by semantic searching using class numbers (on poverty), descriptors ("poverty"), and title words ("poverty" or "poor"). Another set of 3,255 items (called Set B) were found in the *SSCI* by citing the items in Set P. A third set of 133,896 items (called Set C) were cited by the items in Set B. A direct citation relationship between Set P and Set B, and a co-citation relationship between Set P and Set C were examined. Then, the subject bibliography on "poverty" was enlarged by combing the entire *IBSS* (including Set P) and Set B or Set C. The authors conclude that the method in this study is useful for finding out subjects relative to "poverty" inside the *IBSS*, and for retrieving documents on "poverty" outside the *IBSS*.

1. はじめに

社会が発展し複雑化するにつれ、社会科学においても各学問分野が専門化する一方、複数分野にわたる学際的なアプローチが必要となってきた。このような状況の中で、研究者が必要な文献を主題から検索することはますます困難になっている。そのひとつの解決策として、最近、情報検索の分野で注目を集めているのは、論文間の引用関係に基づいて関連文献を見出す方法である。これは、従来からの分類記号、ディスクリプタ、タイトルワード等では捉えきれない複雑で潜在的な主題間の関連を、引用リンクをたどることによって明らかにする方法である。

本研究では、国際的に関心が高まりつつある「貧困 (poverty)」²⁾という学際的な主題に着目し、「貧困」に関する主題書誌を作成するために、社会科学書誌と引用文献索引を用いて、主題検索と引用検索から関連文献を収集する手法を検討する。具体的には、社会科学書誌を用いた主題検索から得られる「貧困」に関する文献群を「貧困」に関する既存の主題書誌とみなし、この書誌を拡張するため、つまり、収録文献を増やすため、この書誌 (文献群) と引用関係にある文献群を引用文献索引から検索することを試みる。

2. 研究素材

社会科学分野における冊子体の年刊書誌に、経済学、政治学、社会学、文化人類学という社会科学の主要4分野を対象とする *International Bibliography of the Social Sciences (IBSS)*³⁾がある。*IBSS*は、UNESCOの援助のもとに、世界各国の専門家から寄せられる雑誌論文や単行書の書誌的データを収録する国際的な選択的書誌であり、1950年代の初頭に、

the International Committee for Social Science Documentation (現: the International Committee for Social Science Information and Documentation) によって編集刊行が開始された⁴⁾。1981年版からは4分野とも編集作業に計算機が導入され、機械可読形式のファイルが作成されている⁵⁾。なお、*IBSS*の編集刊行は、1989年(1987年版)以降、the London School of Economics の the British Library of Political and Economic Science に移行した⁶⁾。本研究では、*IBSS*の1981年版から1985年版までの約14万件の文献の書誌的データを収録する機械可読ファイル⁷⁾を用いる。

一方、社会科学分野における代表的な引用文献索引には、米国の Institute for Scientific Information (ISI) 社による *Social Science Citation Index (SSCI)*⁸⁾がある。*SSCI*は、1973年から編集刊行され、英語圏で出版される文献を中心に収録している⁹⁾。本研究では、冊子体の *SSCI*の1976年～1980年の累積版、1981年と1982年の年刊版、および学術情報センターの NACSIS-IR で提供されている Social SciSearch データベースを利用する。このデータベースでは、1983年から1995年までに出版された文献を検索する。1995年までの検索によって、文献の被引用の期間は最短で10年間となる。

3. 研究方法

本研究では、*IBSS*を用いた主題検索から得られる文献群を出発点として *SSCI*を検索し、引用検索による関連文献収集の手法を検討する。その手順は、以下のとおりである。

- 1) *IBSS*ファイルから重複レコードを除外し、検索対象レコードを確定する。
- 2) この *IBSS*ファイルを用い、従来の主題検索手法によって「貧困」に関する文献群 (P 集合) を検索する。
- 3) *SSCI*を用いて、P 集合中の各文献を引用している文献群 (B 集合)、および B 集合中の各文献が引用している文献群 (C 集合) を検索する。
- 4) P 集合と B 集合、*IBSS*ファイル全体と B 集合、および *IBSS*ファイル全体と C 集合中の同一文献の検出を行い、これらの同一文献からなる各集合を、*IBSS*レコードに含まれる主題情報を用いて分析する。一方、*IBSS*には含まれず、*SSCI*の B 集合のみに含まれる文献群 (S 集合) については、タイトルや収録誌を調査する。

なお、本研究での計算機処理は、図書館情報大学の HITAC M660K を利用し、PL/I でプログラムを作成して行った。また、文献集合間の同一レコードの検出には、各レコードの著者やタイトルなどの文字列の一部を切り出して「同定キー」を作成し、それらの比較によって同一レコードの組を同定する「PYAT 法」¹⁰⁾を用いた。

4. *IBSS*と *SSCI*からの「貧困」関連文献の抽出

4.1 *IBSS*の検索対象レコードの確定

本研究における *IBSS*ファイル中の検索対象レコードは、表1の手順で確定した。まず、検索対象文献の種類を雑誌論文と単行書に限定する(会議録等を除く)。次に、*IBSS*が対象とする社会科学4分野のそれぞれにおける重複レコードを除外し、さらに、異なる分野間での重複レコードも除く。この結果、*IBSS*ファイル中の検索対象レコードは、129,568件となった。

4.2 *IBSS*を用いた主題検索による文献群の検索 (P 集合)

4.1 節で確定した *IBSS* ファイルから「貧困」という主題を含む文献を分類記号 (4 分野から各 1 つずつ), ディスクリプタ (“poverty”), タイトルワード (“poverty” または “poor”) によって検索した。この結果, P 集合中の文献レコードは, 791 件になった。

4.3 *SSCI*を用いた引用検索による文献群の抽出 (B 集合, C 集合)

まず, P 集合中の各文献を引用している文献群 (B 集合) を, *SSCI* から検索した。この B 集合には同一文献が重複して含まれている可能性があるため, 次に, 「PYAT 法」を用いて重複レコードの組を検出した。その結果, B 集合中の文献レコードは, 異なり 3,255 件, 延べ 3,666 件となった。さらに, B 集合中の各文献 (異なり) が引用している文献群 (C 集合) を *SSCI* から検索したところ, C 集合中の文献レコードは, 延べ 133,896 件であった。この C 集合には P 集合が含まれており, P 集合中の文献レコードは C 集合中の文献レコードといわゆる「共引用関係」になっている (図 1 参照)。

4.4 P 集合と B 集合, *IBSS* ファイル全体と B 集合および C 集合からの同一文献の検出

B 集合および C 集合中の文献レコードを, *IBSS* にも含まれるものと *IBSS* には含まれないものとに分け, 前者に対しては *IBSS* レコードの主題情報を用いて検討するため, P 集合と B 集合, *IBSS* ファイル全体と B 集合, および *IBSS* ファイル全体と C 集合中の同一文献をそれぞれ検出した。

まず, P 集合 (791 件) と B 集合 (3,255 件) から「PYAT 法」によって同一レコードの組を 35 組検出した。この 35 件からなる文献集合を, 以下, 「PB 集合」(B 集合全体の 1.1%) と呼ぶ。次に, *IBSS* ファイル全体 (129,568 件) と B 集合中のレコードに「PYAT 法」を適用し, 得られた同一レコードの組 (232 組) から PB 集合に属する組と誤同定の組 (1 組) を除外して, 196 組の同一レコードの組を検出した。この 196 件からなる文献集合を, 以下, 「BI 集合」(B 集合全体の 6.0%) と呼ぶ。そして, B 集合中の文献レコードのうち, PB 集合にも BI 集合にも含まれない 3,024 件のレコードからなる文献集合を「S 集合」(同じく 92.9%) とする。このようにして, B 集合を, *IBSS* に含まれる PB 集合と BI 集合, および *IBSS* には含まれない S 集合の 3 つの文献集合に分けた (図 2 参照)。

次に, *IBSS* ファイル全体と C 集合中のレコードから同一レコードの組を検出する。ただし, P 集合中の文献は「貧困」に関する既存の書誌とみなされているので, 書誌の拡張には無関係である。そこで, まず, PB 集合, BI 集合, S 集合のそれぞれが引用している文献の集合 (C 集合の部分集合) から, P 集合にも含まれる文献レコード群 (それぞれ, 「P(PB) 集合」, 「P(BI) 集合」, 「P(S) 集合」と呼ぶ) を除外した。そして, 残った C 集合中のそれぞれの 3 つのレコード群に「PYAT 法」を適用して重複レコードを除外したうえで, これらの 3 つのレコード群と *IBSS* ファイル全体との間で同一レコードの組の検出を試みた。ただし, これらの同一レコードの組とみなされた中で誤同定の可能性が高い組を除くため, 今回は, *IBSS* レコードの件数が 1 件のもののみを以下の検討の対象とすることにした。それらは, CI(PB) 集合 (61 件), CI(BI) 集合 (389 件), CI(S) 集合 (3,255 件) であった (図 3 参照)。

5. 主題書誌拡張の手法の検討

ここでは, 主題書誌 (P 集合) の拡張の手法を, *IBSS* に含まれる文献集合を用いる場合と,

IBSSに含まれない文献集合を用いる場合に分けて検討する。前者はBI集合、CI(PB)集合、CI(BI)集合、CI(S)集合の4つの集合であり、後者はS集合である。

5.1 引用検索による文献がIBSSにも含まれている場合

IBSS ファイル中の文献レコードには、主題情報として、そのレコードが属する分野の分類表による分類記号、およびIBSSのシソーラスである *Thematic List of Descriptors*¹¹⁾によって分類記号と対応付けられたディスクリプタが付与されている。そこで、IBSSには含まれているが、P集合には含まれていない文献レコードの集合のうち、P集合と直接的な引用の関係にあるBI集合、P集合と共引用の関係にあるCI(PB)集合、CI(BI)集合、およびCI(S)集合中の文献レコードに付与されている分類記号とディスクリプタを調査して、「貧困」という主題の周辺領域の析出を試みた。

(1)分類記号の調査とその結果

まず、BI集合、CI(PB)集合、CI(BI)集合、CI(S)集合の4つの集合中、分類記号を持つ文献レコードの件数と4分野別の分類記号の件数を示すと、表2のとおりである。なお、分類記号は1分野ごとに複数付与されていることもあるため、表2では延べ件数を示した。

IBSSでは収録対象の4分野でそれぞれ異なる階層分類表を用いているので、次に、表2で示した分類記号の件数を、各分野の分類表の最上位の階層である大分類項目別に集計した。その結果、各分類記号が属している大分類項目は、各分野で10項目から15項目にわたっていた。表3に社会学の場合を示す。

また、各分野の分類記号の約半数は、その分野の大分類項目の上位2項目までから4項目までに含まれていた。これらの大分類項目は、文化人類学分野では「社会組織と社会関係」と「形態学的基礎」、経済学分野では「生産組織」、「経済活動」、「公共経済」および「生産(商品とサービス)」、政治学分野では「政策過程」、「政府と行政」および「政治生活」、そして社会学分野では「人口、家族、民族」、「社会問題、社会サービス、社会事業」および「労働」であった。これらは、「貧困」の周辺に存在する主題領域といえよう。

(2)ディスクリプタの調査とその結果

分類記号の調査の場合と同じ4つの集合に含まれる、文献レコードの件数とディスクリプタの分野別の件数を示すと、表4のとおりである。ディスクリプタの場合、複数付与されていることが多いので、まず、延べ件数で集計したところ、6,144件であった。

次に、これらのディスクリプタを出現頻度順に示すと表5のとおりであり、4つの集合から得られたディスクリプタの異なり件数は、1,752件であることが分かる。これは、*Thematic List of Descriptors*中の全ディスクリプタ6,486件の27.0%を占める。このうち、最も多く出現したディスクリプタである“Income Distribution”(61回出現)が、延べ6,144件のディスクリプタ中に占める割合は1.0%である。また、異なり1,752件の半数近くの825件が調査対象の4つの集合のいずれかで1回しか出現していない。分類記号の場合は少数の大分類項目に文献が集中していたが(表3参照)、ディスクリプタの場合、多種多様のディスクリプタが「貧困」関連文献に付与されていることが分かった。

表5に示した上位18位までのディスクリプタを文献集合別・分野別で見ると、4つの集合のうちレコード件数が最も多いCI(S)集合では、4分野にわたって出現しているディスクリプタが4件、3分野が11件、2分野が3件である。このことは、「貧困」の周辺に存在する主

題も、「貧困」と同様に、学際的な主題であることを示している。

なお、表5からは延べ6,144件のディスクリプタのうち、地域細目を持つディスクリプタが3,912件(63.7%)を占めていることも分かる。このことから、「貧困」に関連する文献は、地域的な属性を持つ傾向があると考えられる。

5.2 引用検索による文献が *IBSS*には含まれていない場合

*IBSS*には含まれないS集合中の3,024件の文献レコードの場合、その文献タイトル中に“poverty”か“poor”を含むものが292件(S集合全体の9.7%)存在した。これらの文献の掲載誌を調査したところ、掲載誌は156誌であり、そのうちの104誌(66.7%)が*IBSS*の収録対象誌であった。これらの104誌には、219件の文献(全292件中の75.0%)が収録されていた。*IBSS*に収録される文献は、社会科学分野の専門家によって選択されているため、これらの219件の文献は、何らかの理由で選択から漏れてしまったものであると考えられる。したがって、これらは「貧困」に関する主題書誌の拡張に利用できる文献の候補といえよう。

一方、*IBSS*の収録対象誌ではない52誌(全156誌中の33.3%)のうち、*SCI*から選択的に*SSCI*に収録されていた5誌を除いた47誌が掲載している68件の文献(全292件中の23.3%)も、文献タイトル中に“poverty”か“poor”を含んでいることから、主題書誌の拡張に利用できると考えられる。

6. おわりに

本研究では、社会科学分野を対象として、「貧困」という学際的な主題を選び、選択的書誌*IBSS*と引用文献索引*SSCI*を利用し、従来からの主題検索によって得られた文献群を出発点とする引用検索によって、その主題の関連文献を収集する手法を検討した。その結果、主題検索で得られた文献791件と直接的な引用関係を持つ文献が約3,000件検索され、共引用関係にある文献は約130,000件も存在することが分かった。これらは、従来の主題からの検索手法では得られない「貧困」関連文献である可能性を持っている。

これらのうち、*IBSS*にも含まれている文献については、*IBSS*の主題情報を利用して、「貧困」の周辺に存在する主題領域を見だし、これらの領域に含まれるディスクリプタ群も明らかにした。また、*IBSS*に含まれていない文献については、その文献タイトルから「貧困」そのものを扱っていると考えられる文献を選んだが、さらに「貧困」の関連文献を選ぶには、今回の調査で得られた、*IBSS*に含まれる文献に付与されていた分類記号やディスクリプタを、利用することができる。

ただし、共引用関係となる文献の組は非常に多くなるので、今後は重み付けを導入したりして、文献を絞り込む方法を考える必要がある。また、今回は、引用索引から得られた文献が実際に「貧困」に適合しているかどうかを十分には調査しなかった。これについては、近年、「適合」の概念が多様化しており、安易に適合・不適合の判定を行なうことは避けるべきだと考えたからである。これも今後の検討課題としたい。

謝辞

本研究を進めるにあたって、一橋大学経済研究所日本経済統計情報センターの松田芳郎教授から貴重な助言を得た。記して謝意を表する。

参照文献

- 1) Harter, Stephen P. Psychological relevance and Information Science. *Journal of the American Society for Information Science*. Vol.43, No.9, p.602-615(1992)
- 2) 西川潤. 貧困. 新版. 東京, 岩波書店, 1994, 79p. (岩波ブックレット No. 347)
- 3) International Committee for Social Science Information and Documentation. *International Bibliography of Sociology, 1951-*. London, Tavistock Publications, 1952-. (International Bibliography of the Social Sciences) (他の3分野についても同様に刊行されている)
- 4) UNESCO. The International Committee for Social Sciences Documentation : ten years of activity. *International Social Science Journal*. Vol.14, No.1, p.177-191(1962)
- 5) Meyriat, Jean. "An attempt to international scientific cooperation : forty years contributing to social science information" . *Social Science Information and Documentation in Changing World, Papers Presented at Symposium and Seminar, Tokyo, Japan, 1992-09*, p.144-146.
- 6) Brindley, Lynne J. "The role of the IBSS in the field of social science information retrieval" . *Social Science Information and Documentation in Changing World, Papers Presented at Symposium and Seminar, Tokyo, Japan, 1992-09*, p.23.
- 7) 岸田和明, 松井幸子. 国際協力による学際的書誌データベースの品質と特性 : IBSS データベースによる社会科学分野の事例研究. *図書館情報大学研究報告*. Vol.12, No.2, p.68-78 (1993)
- 8) Institute for Scientific Information. *Social Science Citation Index, 1973-*. Philadelphia, Institute for Scientific Information, 1973-.
- 9) Institute for Scientific Information. *Social Science Citation Index: guide and lists of source publications, 1985*. Philadelphia, Institute for Scientific Information, 1985, p.120-125.
- 10) 松井幸子. 社会科学文献データベースの構造解析. 東京, 丸善, 1992, p.176-177. [初出 : 書誌情報データベースの統合について : 重複文献同定の同定子 (PYAT システムの設計) . *図書館短期大学紀要*. No.14, p.113-139(1977)]
- 11) International Committee for Social Science Information and Documentation. *Thematic List of Descriptors : Sociology*. London, Routledge, 1989. 475p. (他の3分野についても同様に刊行されている)

手塚敬子 : 東京工業大学附属図書館情報サービス課参考調査掛

Keiko TEZUKA: Information Service Section, Tokyo Institute of Technology Library

松井幸子 : 図書館情報大学図書館情報学部教授

Sachiko MATSUI: Professor, University of Library and Information Science

岸田和明 : 駿河台大学文化情報学部助教授

Kazuaki KISHIDA: Associate Professor, Surugadai University

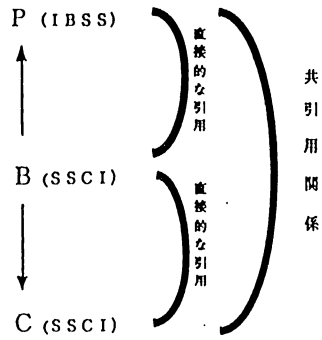


図1 P集合, B集合, C集合の引用関係

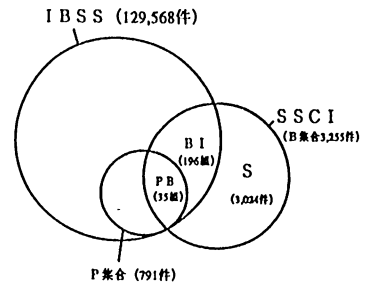


図2 IBSSとSSCI (B集合) の同定結果

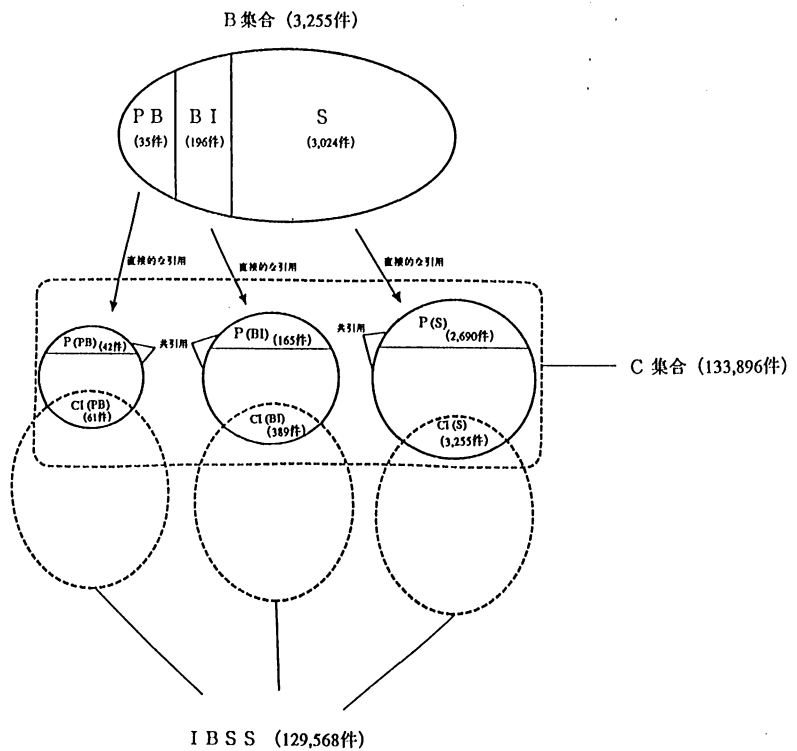


図3 IBSSとSSCI (C集合) の同定結果

