

研究者数の国際比較に関する一考察 —論文データベースによる主要国の研究者数推計—

阪 彩香, ○桑原輝隆 (文科省・科学技術政策研)

1. はじめに

科学技術政策を検討し、また我が国と諸外国を比較する上での基礎情報として、研究者数は最も基本的なデータの一つである。OECDによる主要国の全研究者数及び大学研究者数を表1に示す。

ここから明らかとなるのは、我が国の研究者数が欧米に比べてかなり多く見えることである。即ち、日本を100とした場合に、アメリカの全研究者は2倍程度、大学研究者はほぼ同数となっている。我が国の研究者統計では、1996年より大学研究者についてもFTEが導入されており、この点の相違は解消されているはずであるが、研究大学だけで100を超えるといわれるアメリカの大学研究者数が日本の大学研究者数と同程度であるとは考えにくい。

また、今日の科学技術政策においては、広い意味における効率を向上させていくことが重視されるようになっている。このような評価を行う場合に研究者数は例えば論文数などのアウトプットに対する“分母”として用いられる。諸外国との比較等のベンチマーキングを行う上でも“分母”となる研究者数のデータを分析し、必要なら統計等の改善を検討していくことは重要な課題である。

本稿は、このような問題意識に立ち、研究者数の実像について国際比較を行うための一つの指標として、論文データベースに現れる著者を国別に計測し、その結果の分析を試みるものである。

2. 論文データベースによる研究者数推計手法

推計に用いたデータはThomson ScientificのWeb of Science (以下WoSという。)のデータである。1981年から2005年までの25年間に於いて、WoSに収録されるレコードのうち、article, note, letterおよびreviewの4つ(以下単に「論文」と呼ぶ)を分析の対象とした。

論文の書誌情報のうち、本分析に使用したのは、著者名、所属機関—国名である。なお、所属機関についてデータベースでは大学名のほかに学部など内部組織名なども与えられているが、ここでは大学名など組織

全体を示すもののみを取り上げている。従って、たとえば東京大学理学部と工学部は区別していない。

多くの論文において、著者、所属機関は複数であり、このような場合データベース上は各著者と所属機関の対応関係は特定できない。すなわち、

(参考例)

著者: A_1, A_2, A_3, A_4 機関: O_1, O_2, O_3

著者: A_1, A_3, A_5 機関: O_1, O_4

の場合、著者の所属機関は特定できない。

そこで、以下の研究者数推計アルゴリズムを考案した。

- (1) ある年(複数年を対象とする場合もある)の全レコードを当初リストとする。
- (2) 当初リスト中で、機関数が1のレコードについては、(複数の)著者の所属を特定することができるので、このように決定される著者—機関の組み合わせのリスト(確定リスト)を作成する。また、これらのレコードを当初リストから除く(第1次未確定リスト)。
- (3) 第1次未確定リストのレコードをチェックし、確定リストに存在する著者—機関の組み合わせが存在する場合、当該著者を第1次未確定リストから除く(機関は除かない)。これが第2次未確定リストとなる。
- (4) 第2次未確定リストのレコードを対象として、著者—機関のすべての組み合わせとその確率 P を求める。参考例の場合、 $P(A_1, O_1) = 5/6$ 、 $P(A_1, O_2) = 1/3$ などとなる。
- (5) これらについて、著者名で名寄せを行い、各組み合わせの確率を算出する。例えば、 $P(A_x, O_1) = 2.5$ 、 $P(A_x, O_2) = 1.3$ 、 $P(A_x, O_3) = 0.8$ のようにある著者につて確率が最大値をとる組織が単一の場合には、 A_1 の所属機関を O_1 と決定する。なお、最大値をとるのが複数機関になる場合は未決定のままとする。これによって定まった著者—機関のリストを準確定リストと呼ぶ。

(6) (3)の場合と同様に、第2次未確定リストから準確定リストにある著者を除く。残りが最終未確定リストとなる。

(7) 確定リストと準確定リストをマージして決定リストを作成し、国別の著者数を算出する。

このようにして所属機関（国）が決定された研究者数は表2のようになる。以降これらを「論文研究者数」と呼び、表1に示されるOECD統計の研究者数を「統計研究者数」と呼ぶ。

表2のデータから明らかなように、この処理の場合各国ともに論文研究者数が最大となるのは1995年であり、以降はその数は減少している。これは、このアルゴリズムによる研究者の所属の決定率が80年代に比べ近年では低下するためである。表3は、このアルゴリズムで著者のうちどの程度の所属が決定できたかを示している。単年処理を行った場合、1981年から1985年のそれぞれの処理結果の平均を見ると、論文著者延べ数約135万のうち92%が決定されるが、2001年から2005年の平均では73%まで低下する。これは、多数の著者及び多数の機関に関わる論文が増加して、著者の所属が確定する単独機関論文の比率が低下しているためである。実際、表3によれば、著者所属が決定できる単独機関論文は、80年代前半に92%あったが、最近では62%に低下している。また、80年代前半延べ135万人の研究者が50万件の論文を書いていたのに対し、2000年代前半には延べ346万人の著者が84万件の論文を書いている。即ち、1論文あたりの平均著者数は2.7人から4.1人に増加している。

3. 複数年データの処理による論文研究者数

このように、単年データの場合、最近のデータで決定率がかなり低下するため、例えば2001～2005年の5年間をデータ処理の対象とする場合の試算を行った。単年の場合に比べて5年間のデータを見ればある著者が単独機関の論文を書くケースが多く含まれると考えられるためである。以下、例えば5年間のデータを対象とする場合を5年集計という。試算においては、2年か

ら5年までの各集計を行ったが、ここでは5年集計の結果を表2の中段に示す。同様に、表3の下段は5年集計の場合の決定状況である。

単年集計の場合、2001～2005年の平均で決定率は73%であるが、2001・2005の5年集計では85%まで向上する。

次に、表2の主要国の研究者数について、単年集計と5年集計の結果を比較する。5年集計による論文研究者数は、概して単年集計結果の3.5倍程度であり80年代以降大きな変動は見られない。ただ、日本のみこの値が3.0程度と小さいことが特徴となっている。5年集計が単年集計より大きくなる要因としては次が考えられる。

- 1) 研究者のライフサイクルによる変化。毎年研究者集団には若い層が新規参入し、同時に高齢層がリタイアしていく。このような変化は単年集計に比べて5年集計ではより大きくなる。
- 2) 毎年論文を書かない著者が多いと、単年集計ではその一部しか捕捉されない。これに対して、5年集計ではかなり捕捉されると考えられる。
- 3) 対象期間のうちに研究者の所属組織が変わった場合、研究者一機関の組み合わせによりカウントすることから一人の研究者が多重カウントされることになり、論文研究者数が増加する。特に、論文著者の中には相当数の大学院生等が含まれており、5年間という機関では所属が変わって多重カウントされる可能性が高い。
- 4) 同様に、研究者が機関間を移動しなくても、所属機関以外に他の機関に兼務するようになった場合も、多重カウントが発生する。この可能性も5年集計の場合より大きくなる。

ただ、これらの要因からは、日本の比率が他国に比べて低い理由は見いだしにくく、今後さらに分析が必要である。

4. 主要国間の論文研究者数の比較

5年集計の各国の論文研究者数を比較する。決定率

が年代により変化することから時間軸での増減を議論することはできない。そこで、各対象年の日本の論文研究者数を100として、他国の論文研究者数を指数化したものが表4である。

さらに、5年集計で決定された著者の所属機関名に”Univ “が含まれるものを抽出し、その数を大学論文研究者数とした。大学の中には東京工業大学やMITのように名称に”Univ”を含まないものがあるが、ここではこれらは処理対象としていない。

まず全論文研究者数で、80年代前半アメリカは日本の5倍以上、イギリス、ドイツ、フランスも日本の数を上回っていた。しかし、2001-2005ではアメリカは日本の3.7倍にとどまり、また、日本の論文研究者数はイギリス、ドイツ、フランスのすべてを上回っている。このことは、この20年間に於いて日本の（英語論文を書く）研究者が欧米を上回るペースで増加したことを示すと考えられる。

大学に所属する論文研究者数を比較すると、日本とアメリカの関係は全論文研究者数の場合と同様である。日本は80年代後半以降イギリス、ドイツを上回っており、またフランスは一貫して日本の半分以下という特徴がある。

各国の全論文研究者に占める大学論文研究者の比率の推移を見ると、①各国ともこの20年間でその比率が増大している、②最近では、日本、アメリカ、ドイツで60%程度を占め、イギリスは55%とやや低く、フランスは31%と低い、という特徴が認められる。

表1 各国の統計研究者数（OECD）

区分	対象期間	研究者数(OECD)					日本を100とした場合				
		アメリカ	日本	イギリス	ドイツ	フランス	アメリカ	日本	イギリス	ドイツ	フランス
全研究者数	1981～1985の平均	749,320	431,004	128,400	136,715	93,744	173.9	100.0	29.8	31.7	21.8
	1986～1990の平均	919,388	535,829	134,200	171,009	114,769	171.6	100.0	25.0	31.9	21.4
	1991～1995の平均	1,010,475	638,823	133,535	236,499	143,566	158.2	100.0	20.9	37.0	22.5
	1996～2000の平均	1,236,697	640,427	149,346	243,252	159,558	193.1	100.0	23.3	38.0	24.9
	2001～2005の平均	1,327,362	668,745		266,380	185,527	198.5	100.0		39.8	27.7
大学研究者数	1981～1985の平均	99,360	174,977	25,000	32,857	34,068	56.8	100.0	14.3	18.8	19.5
	1986～1990の平均	136,188	200,232	26,400	37,741	38,145	68.0	100.0	13.2	18.8	19.1
	1991～1995の平均	164,875	228,839	36,200	63,303	49,202	72.0	100.0	15.8	27.7	21.5
	1996～2000の平均	182,329	175,654	47,891	66,314	56,819	103.8	100.0	27.3	37.8	32.3
	2001～2005の平均		180,150		69,166	63,462		100.0		38.4	35.2

これら論文研究者数の各国の比較の状況と表1のOECD統計による統計研究者数を比較すると以下のような相違点が認められる。

- 1) 論文生産の主体である大学についてみると、アメリカの大学論文研究者数は、日本、ドイツ、イギリスの約4倍程度であるのに対して、統計大学研究者数では日本の約1倍、ドイツ、イギリスの約4倍となっている。特に日本の乖離が著しい。またアメリカをフランスと比較すると大学論文研究者数では約9倍であるのに対して大学統計研究者数では約3倍となっている。フランスについては、大学研究者の統計の取り方あるいは大学研究者が論文を執筆する際の所属機関の記述など（英語名称にUnivが含まれない大学が多い可能性もある）に独特の点があると推察される。
- 2) 全論文研究者数では、アメリカは日本、イギリス、ドイツの約4倍であるのに対して、統計研究者数では日本の2倍、イギリスの8倍、ドイツの5倍となっている。このことは、これらの国において大学以外の国立研究機関や企業の研究者数の統計の取り方にある種の相違が存在する可能性を示唆する。また人数の大きい企業研究者が論文生産にどの程度のエフォートを割くかについての国情の相違も反映されていると考えられる。

表2 論文データベースによる論文研究者数推計

単年集計 による主 要国研究 者数推計	集計期間	アメリカ	日本	イギリス	ドイツ	フランス	全世界
	1981～1985平均	253,638	54,410	61,198	56,553	50,180	794,326
	1986～1990平均	287,508	73,808	70,500	64,847	59,958	945,501
	1991～1995平均	336,116	97,653	83,896	74,100	70,188	1,113,391
	1996～2000平均	310,828	99,161	81,974	78,464	68,264	1,108,712
	2001～2005平均	297,420	91,068	72,670	73,352	57,610	1,106,811
5年集計 による主 要国研究 者数推計	集計期間	アメリカ	日本	イギリス	ドイツ	フランス	全世界
	1981-1985	873,276	166,526	215,058	198,078	182,646	2,769,283
	1986-1990	984,022	228,928	252,668	229,387	218,821	3,301,592
	1991-1995	1,167,545	304,509	304,128	270,720	261,154	3,976,653
	1996-2000	1,079,130	299,086	296,952	274,706	255,397	3,869,716
	2001-2005	999,952	267,167	258,564	249,890	210,391	3,717,390
5年集計/ 単年集計	集計期間	アメリカ	日本	イギリス	ドイツ	フランス	全世界
	1981-1985	3.4	3.1	3.5	3.5	3.6	3.5
	1986-1990	3.4	3.1	3.6	3.5	3.6	3.5
	1991-1995	3.5	3.1	3.6	3.7	3.7	3.6
	1996-2000	3.5	3.0	3.6	3.5	3.7	3.5
	2001-2005	3.4	2.9	3.6	3.4	3.7	3.4

表3 論文著者の所属の決定率

対象年(単年集計の平均)	レコード数 a	著者数 (延べ) b	確定レコード 数 c	確定著者数 (延べ) d	準確定著者 数(延べ) e	決定著者 数(延べ) d+e	確定レ コード比率 c/a	確定 著者比率 d/b	準確定 著者比率 e/b	決定著者 比率 (d+e)/b
1981-1985平均	504,997	1,350,178	465,943	1,225,500	18,275	1,243,775	92.3	90.8	1.3	92.2
1986-1990平均	563,373	1,691,875	503,424	1,483,090	34,620	1,517,710	89.4	87.7	2.0	89.8
1991-1995平均	647,014	2,177,580	549,947	1,781,350	75,572	1,856,923	85.1	82.0	3.4	85.4
1996-2000平均	742,157	2,780,069	539,173	1,950,659	212,196	2,162,854	72.8	70.4	7.5	78.0
2001-2005平均	840,633	3,461,954	520,229	2,123,874	389,850	2,513,724	62.0	61.5	11.2	72.7
対象年(5年集計)	レコード数 a	著者数 (延べ) b	確定レコード 数 c	確定著者数 (延べ) d	準確定著者 数(延べ) e	決定著者 数(延べ) d+e	確定レ コード比率 c/a	確定 著者比率 d/b	準確定 著者比率 e/b	決定著者 比率 (d+e)/b
1981-1985	2,524,984	6,750,891	2,385,603	6,374,126	77,264	6,451,390	94.5	94.4	1.1	95.6
1986-1990	2,816,865	8,459,375	2,601,737	7,824,334	147,804	7,972,138	92.4	92.5	1.7	94.2
1991-1995	3,235,069	10,887,902	2,879,012	9,632,556	345,850	9,978,406	89.0	88.5	3.2	91.6
1996-2000	3,710,785	13,900,346	2,979,609	11,348,939	862,926	12,211,865	80.3	81.6	6.2	87.9
2001-2005	4,203,163	17,309,769	3,021,530	13,051,723	1,600,419	14,652,142	71.9	75.4	9.2	84.6

表4

区分	対象期間	論文研究者数(5年集計)					日本を100とした場合				
		アメリカ	日本	イギリス	ドイツ	フランス	アメリカ	日本	イギリス	ドイツ	フランス
全論文 研究者 数 a	1981-1985	873,276	166,526	215,058	198,078	182,646	524.4	100.0	129.1	118.9	109.7
	1986-1990	984,022	228,928	252,668	229,387	218,821	429.8	100.0	110.4	100.2	95.6
	1991-1995	1,167,545	304,509	304,128	270,720	261,154	383.4	100.0	99.9	88.9	85.8
	1996-2000	1,079,130	299,086	296,952	274,706	255,397	360.8	100.0	99.3	91.8	85.4
	2001-2005	999,952	267,167	258,564	249,890	210,391	374.3	100.0	96.8	93.5	78.7
大学論 文研究 者数 b	1981-1985	463,078	91,239	85,428	96,191	40,579	507.5	100.0	93.6	105.4	44.5
	1986-1990	547,118	119,134	98,595	116,107	49,274	459.2	100.0	82.8	97.5	41.4
	1991-1995	636,276	156,580	132,300	147,479	63,084	406.4	100.0	84.5	94.2	40.3
	1996-2000	613,280	166,626	148,787	165,121	67,901	368.1	100.0	89.3	99.1	40.8
	2001-2005	597,365	165,022	143,158	158,386	65,635	362.0	100.0	86.8	96.0	39.8
大学論 文研究 者の比 率 b/a	1981-1985	53.0	54.8	39.7	48.6	22.2					
	1986-1990	55.6	52.0	39.0	50.6	22.5					
	1991-1995	54.5	51.4	43.5	54.5	24.2					
	1996-2000	56.8	55.7	50.1	60.1	26.6					
	2001-2005	59.7	61.8	55.4	63.4	31.2					