

CGSI レポート

日本学術振興会グローバル学術情報センター
JSPS Center for Global Science Information

第1号（平成26年10月31日）

「CGSI レポート」の刊行に際して

平成25年4月、独立行政法人日本学術振興会グローバル学術情報センター（以下「センター」といいます。）が創設されました。センターは、その主な業務として本会諸事業の情報の収集・蓄積を行うとともに、国内外の学術振興機関の事業の実施状況などの情報を収集しています。さらにそれらの情報を分析し、その結果を本会学術システム研究センターに提供するとともに、事業担当部署へ提案し、事業改善に活用することを目的としています。

これに関連して、センターでは収集・蓄積・分析した情報の一部を外部に紹介することにより、センターの活動についての理解を深めていただくために「CGSI レポート」を刊行することとしました。本レポートには、センターの活動として行った科学研究費助成事業等の分析結果や海外の学術振興機関の情報等を掲載し、不定期に刊行して参ります。

本レポートが、読者の皆様の本会諸事業及び海外の学術事情の理解を深めるための有益な情報源となりますことを祈念し、刊行の言葉といたします。

日本学術振興会グローバル学術情報センター所長 前島 信

Scopus 収録論文における科研費成果論文の分析

1. はじめに—本分析の目的と対象

学術研究は、人文学・社会科学から自然科学までの全ての分野をカバーして行われ、そこから生まれる成果は多様である。このため、成果を一律の基準で定量的に評価することは容易ではなく、また、評価したとしても、多様な学術研究の限られた側面のみを見るに留まることも事実である。一方、学術論文は多くの研究分野の研究者にとって研究成果を発表する極めて重要な手段であり、それが多くの他の研究者に読まれ、また、引用されることはその研究の価値を測る上で一つの指標ともなり得る。

センターではその業務の一環として、今回エルゼビア社の文献データベース **Scopus** を用いて、研究者の自由な発想に基づいて行われる学術研究を支援する最大の競争的資金である科学研究費助成事業（以下「科研費」という。）により助成された研究成果の論文の被引用状況についての分析を行った。

科研費による支援を受けた研究代表者は、毎年度、研究実績報告書を提出し、また、研究期間

終了後には研究成果報告書を提出する¹。これらの報告書には、科研費の研究成果として出版された論文等が記載されている。そして、これらのデータは、国立情報学研究所が作成・公開しているデータベース「KAKEN」²において公表されている。

センターでは、エルゼビア社にこのKAKENに収録された文献情報と同社のScopusに収録された文献情報のマッチング、及びマッチングした文献情報のうち、指定した項目による取りまとめを委託した³。Scopusは、エルゼビア社が作成・保有する世界最大級の学術文献データベースで、科学・技術・医学・社会科学・人文科学の各分野にわたる世界5,000社以上の出版社の21,000誌以上（平成26年5月現在）のジャーナルに掲載された論文等の情報を収録している。エルゼビア社は、収録の対象とするジャーナル等の選定は外部の専門家集団によるコンテンツ選定・諮問委員会を設置し、同社の収録ポリシーに基づき行っている。収録誌は大半が英文誌であり、出版者の国別で見ても欧米諸国が中心で、我が国で出版されたジャーナルで収録された数は800誌に満たない。従って、本レポートに記された分析結果は、科研費の成果のうち日本語でまとめられたものは含まず、Scopusに収録された英語論文という限定的なデータに基づくことに注意が必要である。

なお、このような学術文献データベースは、Scopusの他にトムソン・ロイター社が作成・保有するWeb of Science（以下「WoS」という）がある。WoSにはScience Citation Indexをはじめとする6つのデータベースが含まれており、その収録ジャーナル数は計約12,000誌である。これまで、WoSを用いた我が国の論文を通じた学術研究活動の分析はいくつか行われているが、本分析においてはできるだけ科研費による成果の全体像を把握するために、収録対象のジャーナル数の多いScopusを用いた。

本レポートでは、このKAKENとScopusのマッチング論文のデータについて、被引用数やトップ10%論文割合等の項目を、科研費の種目別、年齢別といった観点から、マッチングしなかった論文との比較を含めた分析を行った⁴。

なお、一般に研究成果が論文として出版されるまでには一定の時間を要するため、科研費の成果論文がその研究期間終了後に出版されることもある。しかし、KAKENには研究実績報告書提出の後に受理または出版された論文は含まれていないため、これらの論文は分析の対象となっていないことに注意が必要である。

KAKENとScopusのマッチングの手順は巻末の「備考」において説明したが、このマッチングは完全ではなく、例えば科研費による成果として挙げられ、Scopusに収録されている論文数にはある程度の誤差があり、本レポートの分析もそうしたデータに基づいていることを予め承知いただきたい。

2. 一論文あたりの被引用数の推移

論文の被引用データは、その論文がどの程度の影響力を持つものかを理解するひとつの手掛かりとなる。図1は、著者の所属機関の所在地に日本を含むScopus収録論文（以下「日本Scopus

¹ 平成26年度の使用ルールによる。

² <http://kaken.nii.ac.jp/>

³ Scopusのデータは随時更新されているが、今回の調査では2013年5月1日時点のデータを用いた。

⁴ KAKENのデータは会計年度ごとに収録されるが、論文の出版年は暦年で記入されており、Scopusのデータも暦年で管理されているため、以下の分析では暦年を使用した。

論文」という。)、KAKEN の中で日本 Scopus 論文とマッチした論文 (以下「科研費 - Scopus 論文」という。)、日本 Scopus 論文から科研費 - Scopus 論文を除いた論文 (以下「非科研費 - Scopus 論文」という。)) のそれぞれについて、一論文あたりの被引用数の比較を行ったものである。なお、被引用数は、時間が経過するにつれ累積的に増加するため、出版された時期が新しいほど少なくなる。

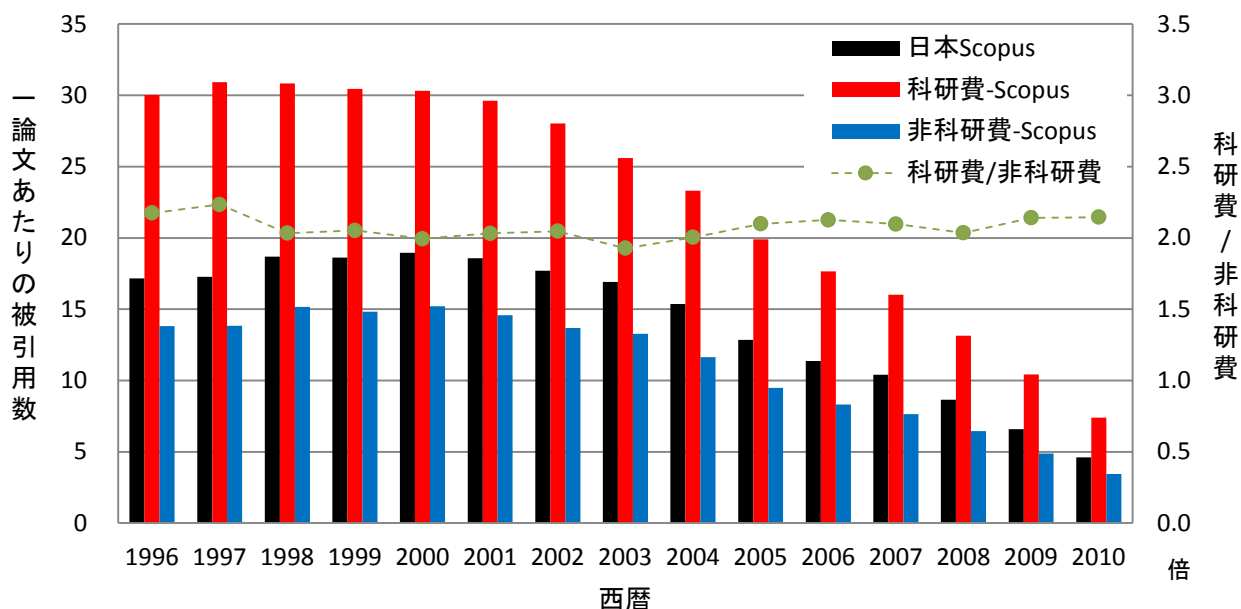


図1. 日本 Scopus 論文、科研費 - Scopus 論文、非科研費 - Scopus 論文の一論文あたりの被引用数の推移

分析対象期間を通して、科研費 - Scopus 論文の一論文あたりの被引用数は、日本 Scopus 論文に対しては概ね 1.5~1.6 倍である。また、非科研費 - Scopus 論文に対しては、概ね 2 倍程度であり、いずれの比較においても科研費 - Scopus 論文が優位であった。

3. 被引用数上位 10%、及び上位 1%論文の割合

エルゼビア社は、独自の研究分野の区分 (ASJC の 27 中分類) における被引用数上位 10%及び上位 1%に該当する論文を抜き出したデータを作成している。図2は日本 Scopus 論文、科研費 - Scopus 論文及び非科研費 - Scopus 論文にかかる被引用数上位 10%及び上位 1%の割合を示したものである。この割合がそれぞれ期待値である 10%、1%よりも大きいかまたは小さいかによって、高被引用論文の割合が Scopus 全体の平均を上回っているか下回っているかを見ることができる。

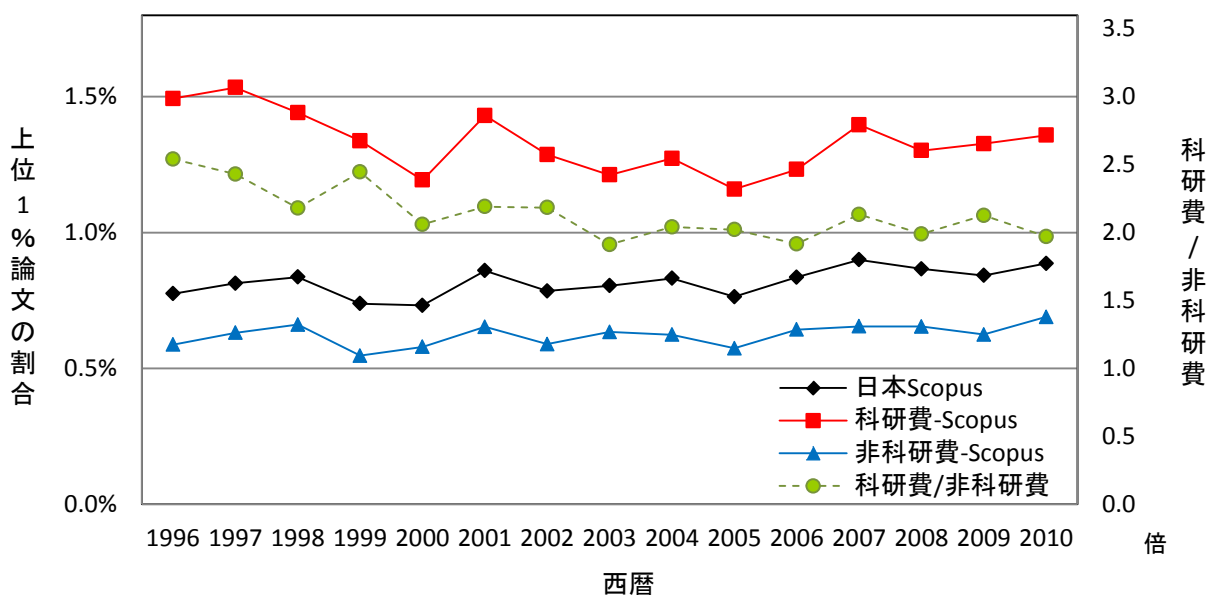
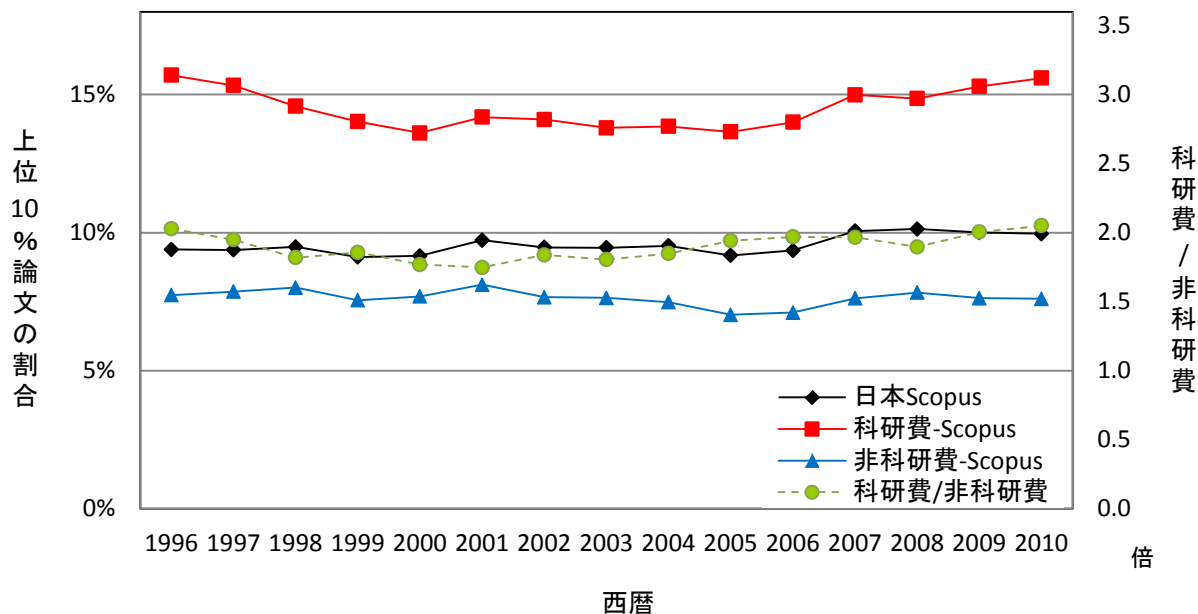


図2. 日本 Scopus 論文、科研費 - Scopus 論文、非科研費 - Scopus 論文の被引用数上位 10%論文の割合（上図）及び上位 1%論文の割合（下図）の推移

科研費 - Scopus 論文における上位 10%論文の割合は 1998～2006 年の間は 13～14%であったが、2007 年以降は 1996～1997 年同様、15%前後となっている。一方、日本 Scopus 論文は、9～10%台、また、非科研費 - Scopus 論文は、1998 年、2001 年の 8%台を除き、7%台となっている。科研費 - Scopus 論文と非科研費 - Scopus 論文を比べると、1998～2004 年は 1.8 倍前後で推移しているが、それ以外は 2 倍程度となっている。

上位 1%論文の割合については、科研費 - Scopus 論文は、近年概ね 1.2～1.5%の範囲を推移しており、非科研費 - Scopus 論文と比較すると概ね 2 倍程度である。

以上のように、科研費 - Scopus 論文は上位 10%論文及び 1%論文の割合のいずれで見ても、日本 Scopus 論文及び非科研費 - Scopus 論文より優位にあるといえる。

4. 科研費 - Scopus 論文の種目別の傾向

科研費には、特別推進研究、基盤研究、挑戦的萌芽研究、若手研究等の性格の異なる種目があり、それぞれ配分額、申請資格等が異なり、また、各課題に参加する研究者（研究分担者等）の構成も異なる。このため、種目により論文の被引用数の傾向にも違いがあることが推測される。図 3 は、主な種目の上位 10%論文及び上位 1%論文の割合について、2006 年から 2008 年の 3 年間の平均をとったものである⁵。

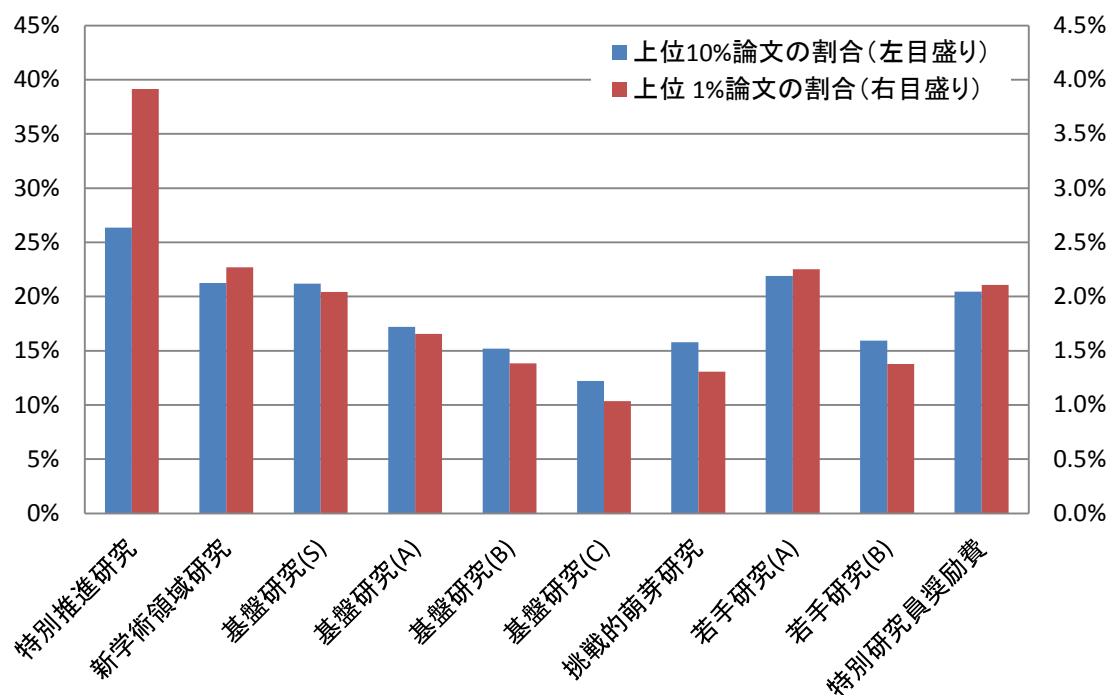


図 3. 主な種目別にみた科研費 - Scopus論文の被引用数上位 10%及び上位 1%論文の割合 (2006~2008 年)⁶

上位 10%論文、上位 1%論文の割合のいずれにおいても、全種目で Scopus 平均(それぞれ 10%、1%)を上回っている。また、大型種目や若手種目、そして特別研究員奨励費で割合が高い傾向が見られる。

【論文被引用に見るファンディングの支援効果について】

科研費の研究支援効果がどの程度かを測りたい場合、同じ研究課題を科研費で支援した場合としなかった場合で研究業績を比較すればよいが、これは当然不可能である。それに代わる一つの

⁵ 特に上位 1%論文数は絶対数が少なく、年ごとのばらつきが大きいため、3年の累計値を使用した。

⁶ 本分析では、新学術領域研究・挑戦的萌芽研究はそれぞれ特定領域研究・萌芽研究のデータを含む。

方法として、他の科研費に類似した研究資金を受けた場合との比較が考えられる。しかし、内外のファンディングエージェンシーの中でこのように論文被引用のデータに基づく分析を行った機関は極めて限られている。

その限られた事例のひとつに、英国の大学等に競争的研究資金を配分するリサーチカウンシルのひとつである医学研究会議（Medical Research Council: MRC）により行われた支援成果論文の被引用数のデータの分析がある。これは、MRCが支援を行った研究者に対して発表した論文等を報告させ、それとWoSとのマッチングを行い、どの程度の支援の成果があったかを定量的に示そうとする試みである⁷。トムソン・ロイター社による分野に関する規格化が行われたnormalised citation impact（nci）と呼ばれる被引用指標⁸を用いて、WoSに収録されたMRCが支援を行った研究成果論文と英国全体の医学分野の研究論文との比較を行ったところ、英国全体のnciが1.39であるのに対し、MRCの支援を受けた研究者（対象数3,841人）による研究成果のnciは2.16と、英国全体の値を大きく上回っており、MRCによる支援が英国の医学研究分野の論文の生産性に貢献していることが示されている⁹。

この指標と、本稿で示す被引用数は異なる指標であり、分析対象の母集団の特性も異なるため、直接比較することは必ずしも適当ではないが、両者は比較優位性を定量的に表しているという点で同じであることから、MRCのnciと、科研費の一論文あたりの被引用数の当該国（英国または日本）の値と当該プログラムの値及びその比を下表に示した。科研費の値としては、2006年～2008年の累計値を用いて、全種目の平均に加え、MRCのグラントの規模に比較的近い種目である基盤研究(S)と基盤研究(A)も併せて示した¹⁰。

	指標	当該プログラムの値 (A)	当該国の平均値との比較	
			当該国の平均値 (B)	A/B
MRC グラント	nci	2.16	1.39	1.55
科研費全体	一論文あたりの被引用数	15.63	10.15	1.54
基盤研究(S)	一論文あたりの被引用数	21.40		2.11
基盤研究(A)	一論文あたりの被引用数	17.50		1.72

MRCは、比較的少数の規模の大きいグラントを通して資金配分を行っており、2006年度～2008年度の間の新規採択件数は計996件、それらの全採択期間に対する平均配分額は59万ポンド（約1億円）である¹¹。

⁷ MRCは、e-Valという支援対象研究者が利用するオンラインシステムを用い、過去に支援を行った研究者に対し、論文等の支援の成果の報告を求めた。

⁸ nciは、当該年に引用されることが期待される値を1とし、その数値が高くなる程、より多くの引用が行われた論文であることを示す。

⁹ MRC「Economic Impact Report 2011/12」

¹⁰ 基盤研究(S)の2006～2008年度の単年度の課題あたりの配分額は1793万円である（5年間の採択の場合、計8964万円）。また、基盤研究(A)は、同期間の単年度の課題あたりの配分額は980万円（3年間の採択の場合、計2940万円）である。

¹¹ MRC Annual Report and Accounts 2006/07、同2007/08、同2008/09

これに対し科研費は、比較的小規模なものから大規模なものまで様々な種目により多数の研究課題を支援しており、1 課題あたりの平均配分額も MRC の規模に比べ遥かに少ない。科研費の中でも比較的規模の大きい種目は、基盤研究 (S) (総額の上限は、2006 年度採択までは 1 億円、2007 年度以降の採択は 2 億円) 及び基盤研究 (A) (総額の上限は 5,000 万円) であるが、これらの種目の一論文あたりの被引用数を見ると、MRC のグラントに見劣りしない支援の効果が確認できる。

5. 研究代表者の年齢層別の分析

科研費の研究代表者の年齢に着目した分析も行った。具体的には、科研費 - Scopus 論文について、科研費の研究者代表者の申請資格登録情報を用い、主な種目について、年齢層別に分析を行った。

図 4 は、各種目における科研費 - Scopus 論文について、研究代表者の年齢層別の被引用数上位 10%論文の割合を示したものである。ただし、年齢層別・種目別の論文数が 20 件以下のものは不定性が大きいため図から除外した。

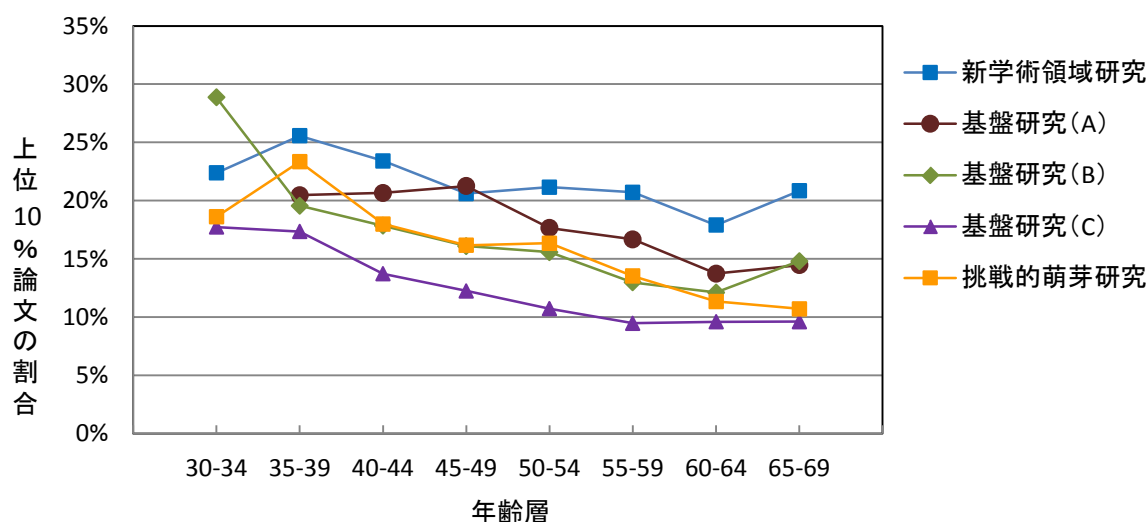


図 4. 主な種目の各年齢層における被引用数上位 10%論文の割合 (2006-2008 年)

ほぼすべての種目及び年齢層で、上位 10%論文の割合は平均 (期待値) の 10%より高い値をとっているが、それぞれの種目において研究代表者の年齢層により値に差があることも示された。この要因や背景を明らかにするためには、更に分析を行う必要がある。

備考 : KAKEN 収録論文と Scopus とのマッチングについて

ここでは、エルゼビア社に委託した KAKEN と Scopus とのマッチングで本レポートでいう「科研費 - Scopus 論文」を同定した手順について説明する。KAKEN には 1996 年度から 2010 年度の間科研費の成果として発表された文献として約 320 万件が収録されているが、そのうち、複数の課題の成果として収録されたもの等、重複分を除いた数は 1,607,734 件であった。そしてこ

の中から英語以外の言語で書かれた論文（主に和文論文）を除外し、980,303 件の英語論文を抽出した。

この KAKEN 英語論文と Scopus に収録された論文のうち、著者の所属機関が日本の機関である論文（本レポートでいう日本 Scopus 論文）とを、後述する手法によりマッチングを行った。その結果、一致した 442,119 件を科研費 - Scopus 論文とした。科研費 - Scopus 論文は、日本 Scopus 論文のおよそ 3 割である。

このマッチング作業は、エルゼビア社が被引用数のカウントに用いている技術を利用し、論文タイトルの完全一致検索を基本として、著者名、出版年、ジャーナル名、DOI、ISSN といった情報を用いて行った。タイトルの完全一致を基本としたため、KAKEN にある複数の別の論文を同一論文としたり、KAKEN にある論文を異なる Scopus 収録論文とマッチングしたりする可能性は小さいことが確認されている。しかし、KAKEN に収録された情報はほとんどが紙媒体の実績報告書から作成されたものであるため、同一論文であっても論文タイトルが一致しない場合もある。このため、重複の除去ができなかったり、マッチングから漏れたりしてしまうことも考えられる。KAKEN にある論文と Scopus 収録論文とのマッチングに漏れがある場合は、先に示した科研費 - Scopus 論文数及び非科研費 - Scopus 論文数に修正が加えられなければならない。この点に関して、エルゼビア社から納品されたデータを用いて、少数のサンプルによる検証をセンターで行ったが、その結果では実際にこうした重複やマッチングの漏れの存在が確認された。従って本レポートの科研費 - Scopus 論文数等にもこうした誤差が含まれることに注意が必要である。ただし、平均被引用数や上位 10%、上位 1%論文の割合など、本レポートに示した分析の主な結果に関しては、誤差による影響は少ないと考えている。

独立行政法人日本学術振興会グローバル学術情報センター

CGSI レポート 第1号

平成 26 年 10 月 31 日発行

本レポートに関する問い合わせ先：

〒102-0083 東京都千代田区麹町 5-3-1 麹町ビジネスセンター 電話 03-3263-1971（遠藤）