

レポート

EBPMのための研究プログラムの分析
(科研費を事例として)– Advanced Research Analysis in Keen-Keywords
Investigation –

第2 調査研究グループ 上席研究官 荒木 寛幸

概 要

証拠に基づく政策立案^{注1} (EBPM。エビデンス・ベースト・ポリシー・メイキング) を推進するために文部科学省科学技術・学術政策研究所 (以下「NISTEP」という) では Advanced Research Analysis in Keen-keywords Investigation システム (以下「ARAKI システム」という) を開発中である。本システムは研究助成プログラムの分析と可視化を目的としている。今回は科学研究費補助金^{注2} (以下「科研費」という) の分析を紹介する。2015年から2019年までの5年間の基盤研究 (A)・(B)・(C)^{注3}の研究種目で新規採択及び継続している課題について研究種目の状況及び役職に分類・集計し、分析を行った。

- ・基盤研究 (A) は採択件数、研究費の総額について 2015年から2019年までの5年間の変化はみられなかった。
- ・基盤研究 (B) や基盤研究 (C) は採択件数と配分された研究費は 2015年から2019年までの5年間で増加傾向であった。
- ・基盤研究 (A)・(B)・(C) の 2015年から2019年までの5年間において、教授の採択件数や研究費の総額は 2017年までは増加、以降減少傾向であった。
- ・基盤研究 (A)・(B)・(C) の 2015年から2019年までの5年間において、准教授や助教の役職では採択件数や研究費の総額は増加傾向であった。
- ・基盤研究 (A)・(B)・(C) の 2015年から2019年までの5年間において、1課題当たり (役職関係なし) の平均額は横ばいで 250万円前後であった。
- ・基盤研究 (A)・(B)・(C) の採択された課題名について、ライフサイエンス分野の語が多く観察された。なかでも「細胞」の頻出が顕著であった。

キーワード：エビデンス・ベースト・ポリシー・メイキング，政策立案，大学経営，産学官連携，ARAKI システム

注1 欧米諸国では、客観的な証拠に基づくエビデンス・ベーストでの政策立案への取組が比較的進んできたのに比べ、我が国では、これまで、統計の最大のユーザーである政府の政策立案において、統計や業務データなどが十分には活用されず、往々にしてエピソード・ベーストでの政策立案が行われているとの指摘がされてきた。我が国の経済社会構造が急速に変化する中、限られた資源を有効に活用し、国民により信頼される行政を展開するためには、政策部門が、統計等を積極的に利用して、証拠に基づく政策立案 (EBPM。エビデンス・ベースト・ポリシー・メイキング) を推進する必要がある。

〔統計改革推進会議最終取りまとめ (平成 29年 5月 19日統計改革推進会議決定) より抜粋〕

注2 科研費は、人文学、社会科学から自然科学までの全ての分野にわたり、基礎から応用までのあらゆる「学術研究」(研究者の自由な発想に基づく研究) を格段に発展させることを目的とする「競争的研究資金」であり、ピアレビュー (同業者 (peer) が審査すること (review) で、科研費においては、学術研究の場で切磋琢磨し「知の創造」の最前線を知る研究者が審査、評価するシステム) による審査を経て、独創的・先駆的な研究に対する助成を行うものである。

注3 一人又は複数の研究者が共同して行う独創的・先駆的な研究

(A) 3～5年間 2,000～5,000万円 (B) 3～5年間 500～2,000万円 (C) 3～5年間 500万円以下

1. はじめに

これまで、データを用いた解析には、データハンドリングのためのITスキル、そして統計的な知識の習得が必要とされてきた。より高度な解析を行うには、そのスキル習得のための学習コストは高い。また、政策立案のための資料に高度な解析結果を用いた場合には、政府の政策立案プロセスにおけるステークホルダー間の相互理解を要する。すなわち、高度な解析手法の理解はステークホルダー間でも重要なファクターであり、相互理解のためのコストが非常にかかる。

エビデンス・ベースト・ポリシー・メイキング（以下「EBPM」という）を推進するためには、「報告者負担の軽減と統計業務・統計行政体制の見直し・業務効率化、基盤強化」が必要とされておりEBPM人材の確保・育成等に関する方針の策定が進められているが、人材育成には時間も必要であり、人材の確保は喫緊の課題である。

そこで、今回、誰でもがEBPM人材としてEBPMを推進するために、学習コストを低くすることで、ステークホルダー間の相互理解に係るコストも低くできる解析手法の検討を行った。

今回解析に用いるデータは、科学研究費助成事業データベース（以下「KAKEN」という）^{注4}に収録されている科研費採択データである。また、データ解析にはMicrosoft Power BI DesktopとKH Coder¹⁾及びR言語を用いた解析用のシステム、ARAKIシステムを構築した。Microsoft Power BI DesktopはExcelに似たユーザーインターフェースで集計及びその結果を可視化することができ、KH Coderは、グラフィカルなユーザーインターフェース（GUI）の操作でテキスト型（文章型）データを統計的に分析することが可能である。

2. 解析手法

解析に用いるデータはKAKENに収録されているデータ（2019年12月20日現在）のうち2015年から2019年までの5年間の基盤研究（A）・（B）・（C）の研究種目で新規採択及び継続している課題を対象としたデータについて分析を行った^{注5}。研究費は各年に配分された研究費の総額（直接経費＋間接

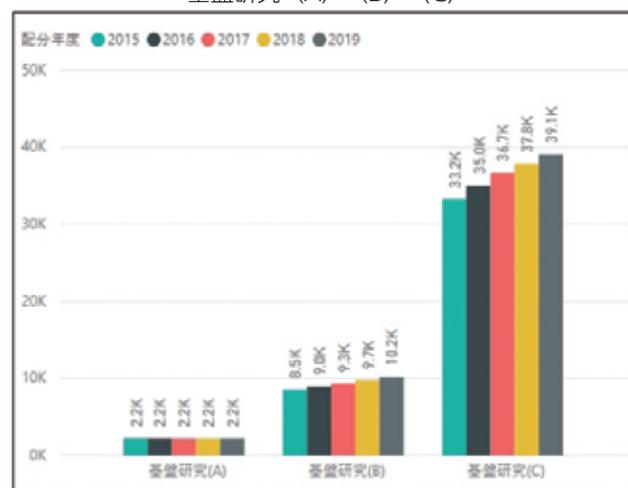
経費）で計算を行った。2015年から2019年までの新規及び継続課題の判別については研究費の配分年データを参考に、配分の実績がある課題を対象としたデータについて取り扱うことにした。KAKENには1,000を超える役職名が収録されている。そこで、本調査では2015年から2019年までの5年間の採択課題件数（基盤研究（A）・（B）・（C）の研究種目の合計）が1,000件を超える13の役職に絞った。KAKENに収録されているデータ（研究の当初採択時のデータ）を利用している本解析は、集計結果として得られている値が確定値ではない（実際の配分状況とは違う場合がある）ことに注意が必要である。

3. 基盤研究（A）・（B）・（C）それぞれにおける採択件数と研究費

基盤研究（A）の採択件数は2015年から2019年までの5年間の推移を見ると横ばいで、2015年は2,230件に対し2019年では2,229件であった。基盤研究（B）の採択件数は2015年から2019年までの5年間の推移を見ると増加傾向で、2015年は8,545件に対し2019年では10,186件であった。基盤研究（C）の採択件数は2015年から2019年までの5年間の推移を見ると増加傾向で、2015年は33,235件に対し2019年では39,057件であった（図表1）。

基盤研究（A）の採択課題の配分された研究費の総額は2015年から2019年までの5年間の推移を見ると横ばいで、2015年は243億円に対し2019年

図表1 科研費【採択件数】の推移（新規＋継続）
- 基盤研究（A）・（B）・（C） -



注4 科学研究費助成事業データベースは、文部科学省及び日本学術振興会が交付する科学研究費助成事業により行われた研究の当初採択時のデータ（採択課題）、研究成果の概要（研究実施状況報告書、研究実績報告書、研究成果報告書概要）、研究成果報告書及び自己評価報告書を収録したデータベースである。https://kaken.nii.ac.jp/ja/

注5 図表中の単位はbn：10億、M：100万、K：千

では247億円であった。基盤研究（B）の採択課題の配分された研究費の総額は2015年から2019年までの5年間の推移を見ると増加傾向で、2015年は412億円に対し2019年では500億円であった。基盤研究（C）の採択課題の配分された研究費の総額は2015年から2019年までの5年間の推移を見ると増加傾向で、2015年は480億円に対し2019年では507億円であった（図表2）。

基盤研究（A）の採択課題の配分された研究費の平均額は2015年から2019年までの5年間の推移を見ると横ばいで、おおむね1,100万前後を行き来している。また、中央値も1,000万円前後で大きな変化はなかった。基盤研究（B）の採択課題の配分された研究費の平均額は2015年から2019年までの5年間の推移を見ると横ばいで、各年490万円程度であった。中央値は460万円程度で変化がなかった。基盤研究（C）の採択課題の配分された研究費の平均額は2015年から2019年までの5年間の推移を見ると減少傾向

向で、2015年は144万円に対し2019年では130万円であった。中央値は各年で平均値とおおむね同じであった。また、最大値は2015年は455万円に対し2019年では403万円であった（図表3）。

基盤研究（A）は採択件数と配分された研究費は2015年から2019年までの5年間での変化は少なく、基盤研究（B）は採択件数と配分された研究費は2015年から2019年までの5年間で増加傾向であり、その配分の平均額には変化がみられなかった。基盤研究（C）は採択件数と配分された研究費は2015年から2019年までの5年間で増加傾向であるが、その配分の平均額は減少傾向であった。特に最大値は減少傾向で変化が大きかった。

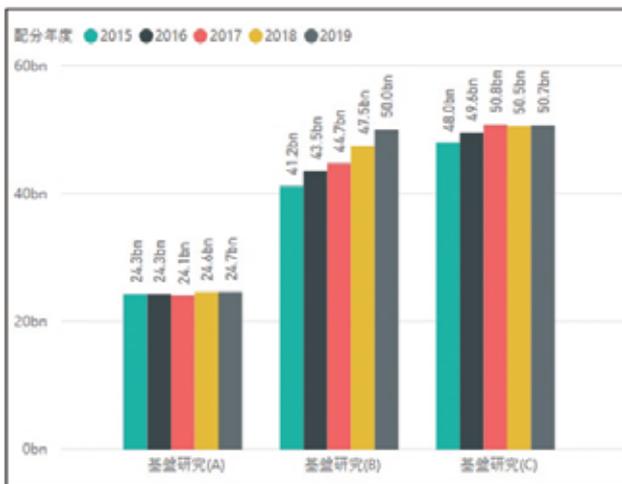
4. 基盤研究（A）・（B）・（C）合算における役職ごとの採択件数と研究費

ここでは、基盤研究（A）・（B）・（C）を合算した採択件数と研究費の総計を役職別に集計した。

2015年から2019年までの5年間における採択件数の推移を役職別に見ると、教授は2017年までは増加、以降減少傾向で、2017年は20,738件に対し2019年では20,099件であった。准教授は増加傾向で、2015年は10,599件に対し2019年では13,244件であった。助教は増加傾向で、2015年は2,891件に対し2019年では5,222件であった（図表4）。

2015年から2019年までの5年間における採択課題の配分研究費・総額の推移を役職別に見ると、教授は2017年までは増加、以降減少傾向で、2017年は619億円に対し2019年では592億円であった。准教授は増加傾向で、2015年は229億円に対し2019年では283億円であった。助教は増加傾向

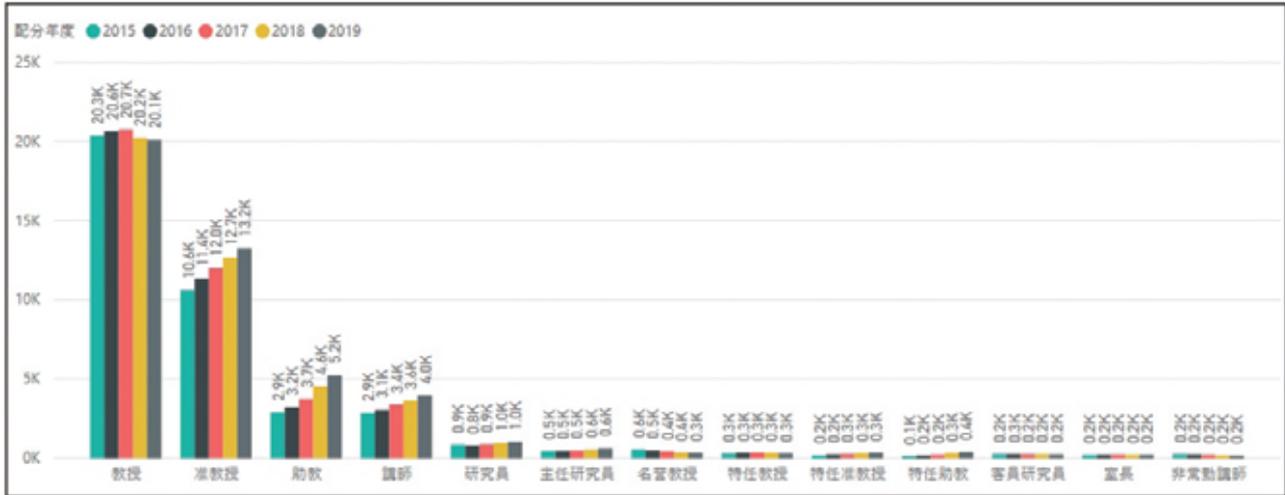
図表2 科研費【研究費総額】の推移（新規+継続）
- 基盤研究（A）・（B）・（C） -



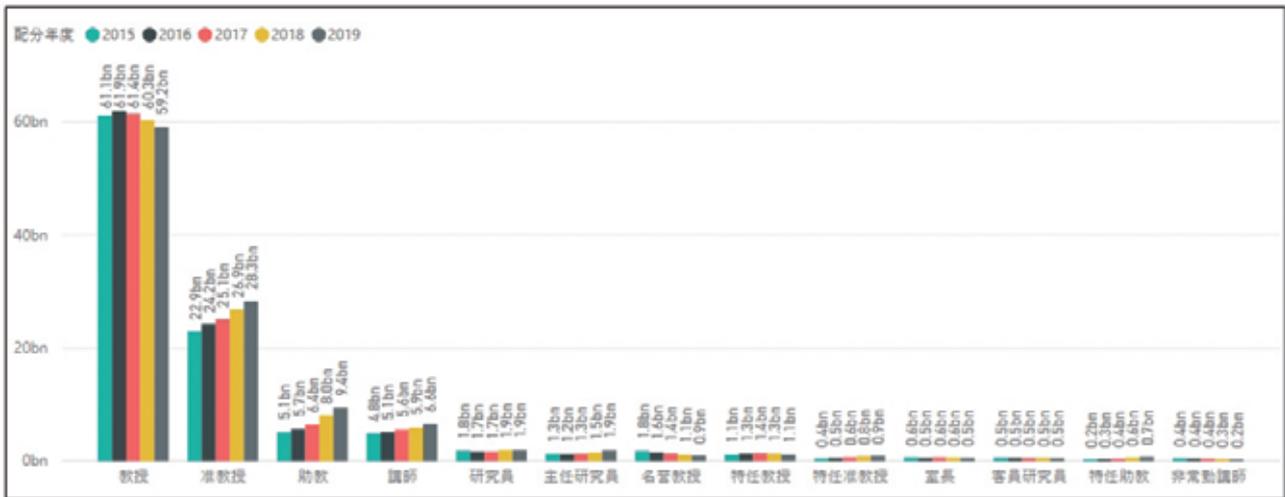
図表3 研究種別による配分研究費の各年における統計量（新規+継続） - 基盤研究（A）・（B）・（C） -

研究種目	各年度配分額	各年度配分額のカウン	各年度配分額の平均	各年度配分額の中央値	各年度配分額の最大値	各年度配分額の最小値	各年度配分額の標準偏差
基盤研究(A)							
2015	24.3bn	2.2K	10.89M	10.01M	37.96M	0.39M	4.80M
2016	24.3bn	2.2K	10.88M	9.88M	39.91M	0.39M	5.00M
2017	24.1bn	2.2K	10.84M	9.88M	41.60M	0.78M	5.08M
2018	24.6bn	2.2K	11.19M	10.01M	42.77M	0.91M	5.50M
2019	24.7bn	2.2K	11.06M	10.14M	45.50M	0.65M	5.13M
基盤研究(B)							
2015	41.2bn	8.5K	4.82M	4.55M	17.03M	0.39M	2.10M
2016	43.5bn	9.0K	4.86M	4.55M	16.51M	0.13M	2.21M
2017	44.7bn	9.3K	4.82M	4.55M	17.55M	0.13M	2.17M
2018	47.5bn	9.7K	4.90M	4.68M	16.64M	0.26M	2.23M
2019	50.0bn	10.2K	4.91M	4.55M	17.42M	0.26M	2.22M
基盤研究(C)							
2015	48.0bn	33.2K	1.44M	1.43M	4.55M	0.13M	0.57M
2016	49.6bn	35.0K	1.42M	1.43M	4.55M	0.13M	0.56M
2017	50.8bn	36.7K	1.38M	1.43M	4.42M	0.13M	0.54M
2018	50.5bn	37.8K	1.34M	1.30M	4.16M	0.13M	0.51M
2019	50.7bn	39.1K	1.30M	1.30M	4.03M	0.13M	0.51M

図表 4 役職別採択課題【件数】-基盤研究 (A)・(B)・(C) -



図表 5 役職別採択課題【配分研究費・総額】-基盤研究 (A)・(B)・(C) -



で、2015年は51億円に対し2019年では94億円であった(図表5)。

2015年から2019年までにおける採択課題の研究費の平均額は2015年から2019年までの5年間の推移を見ると教授は横ばいで、2015年は300万円に対し2019年では294万円であった。准教授は横ばいで、2015年は216万円に対し2019年では213万円であった。助教は横ばいで、2015年は178万円に対し2019年では181万円であった(図表6)。また、1課題当たりの平均額は2015年は258万円に対し2019年では243万円であった。

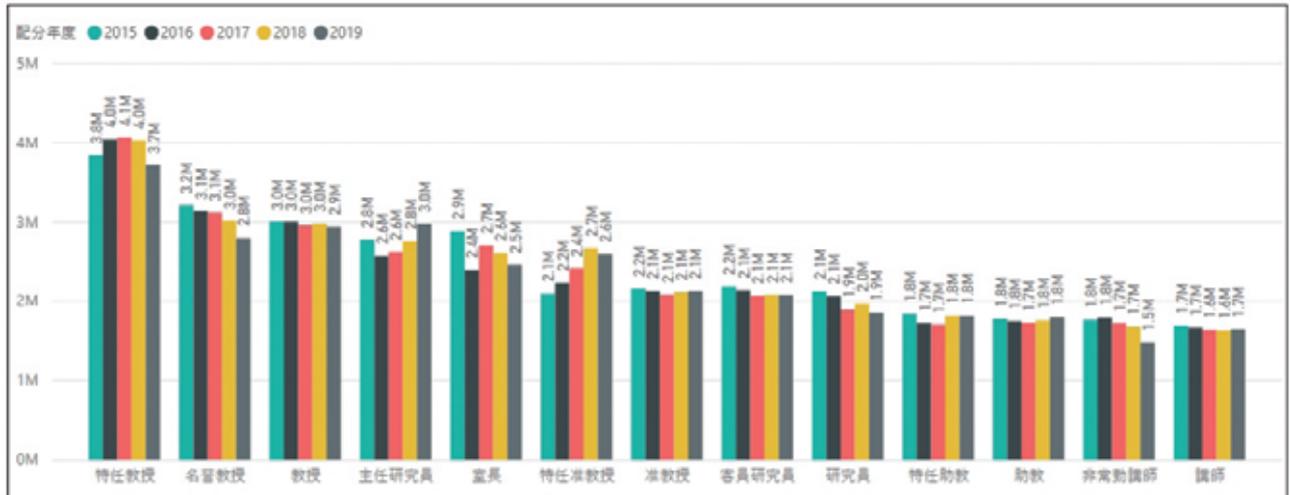
5. 基盤研究 (A)・(B)・(C) における採択課題名の分析

ここでは、2015年から2019年までにおける基盤

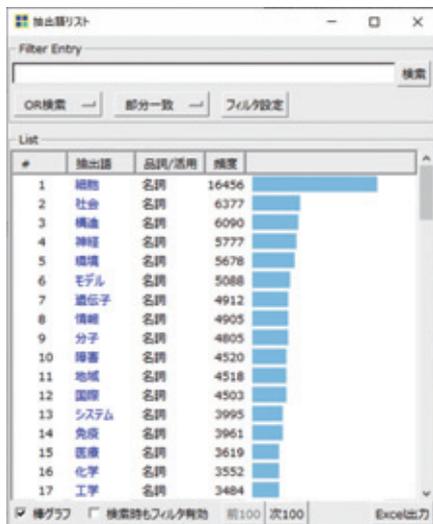
研究 (A)・(B)・(C) に採択された課題名の分析を行った。課題名はテキスト型(文章型)データであるため、解析のためにその文章中の単語を抽出し、その単語の関係性について統計的に分析する。KH Coderを利用する際には、形態素解析器を選択することができるが、本分析ではMeCab^{注6}を利用し単語を抽出、さらに、品詞を名詞に絞った。その集計結果(上位17語)が図表7である。また、採択された課題名の特徴を可視化するために、共起ネットワーク分析を行った。その結果が図表8である。KH Coderにおける共起ネットワーク分析では、文章中に出現する単語同士の関係性、すなわち語と語の共起の程度を測定するためにJaccard係数で数値化しており、語と語の共起の程度が高い(一定値以上の)語と語は線でつながっている。また、サブグラフ検出としてmodularityによる語の共起ネットワークのクラスタリングが行わ

注6 MeCabは京都大学情報学研究科-日本電信電話株式会社コミュニケーション科学基礎研究所共同研究ユニットプロジェクトを通じて開発されたオープンソース形態素解析エンジンである。https://taku910.github.io/mecab/

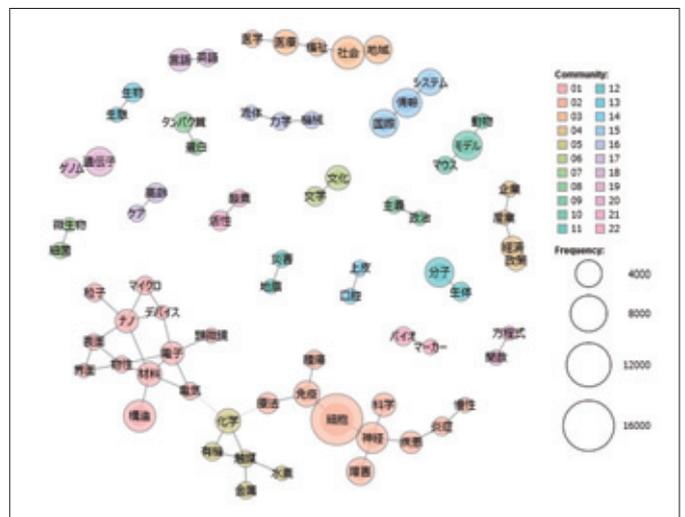
図表 6 役職別採択課題【配分研究費・平均額】 - 基盤研究 (A)・(B)・(C) -



図表 7 抽出語リスト (KH Coder)



図表 8 採択課題名の共起ネットワーク分析 (KH Coder)



れており、ここでは Community として 22 に分類された。飽くまでも統計的な分析による数値的な分類結果であり、Community ごとに意味が付加されているわけではなく、解釈する際の参考であることに注意したい。

共起ネットワーク分析はこのままでは解釈が難しく、有識者とディスカッションを行うことで意味のある可視化となる点も注意したい。

図表 8 を詳しく見ると、最もよく出現する語は、「細胞」であり、採択課題名では「免疫」や「神経」とともに出現することが多い。これらは Community01 に分類されておりライフサイエンス分野でよく利用される語が多い。また、「神経」が多くの単語とつながっていることから中心的な語となっていることがわかる。「ナノ」や「電子」が中心的な語である Community02 は材料分野でよく利用される語が出現しており、なかでも「構造」の語が多く出現している。Community03 は「触媒」が中心の語であり、化学

分野の語が多い。なかでも「化学」は Community01 の「療法」、Community02 の「電気」とも共起の関係にある、言わば分野横断としての懸け橋となるような語としての特徴がある。全体的にライフサイエンス分野の語が多くみられ、Community も「遺伝子」や「分子」、「医療」、「モデル」「動物」などに分かれている。特徴的なのは「地震」、「災害」であろう。この語が含まれている採択課題が Community としても現れている。

6. まとめ

ARAKI システムによる 2015 年から 2019 年までの 5 年間の基盤研究 (A)・(B)・(C) の採択課題の傾向について分析を行った。研究種目としての傾向としては、基盤研究 (A) は採択件数、研究費の総額について 2015 年から 2019 年までの 5 年間の変化はみられなかった。基盤研究 (B) や基盤研究 (C) は

採択件数と配分された研究費は 2015 年から 2019 年までの 5 年間で増加傾向であった。

基盤研究 (A)・(B)・(C) を合算した採択件数と研究費の総計を役職別に集計した結果、2015 年から 2019 年までの 5 年間に於いて、教授の採択件数や研究費の総額は減少傾向であった。准教授や助教の役職では採択件数や研究費の総額は増加傾向であったが、平均額は教授、准教授、助教それぞれで横ばいであった。また、基盤研究 (A)・(B)・(C) を合算した総計の 1 課題当たり (役職関係なし) の平均額は横ばいで 250 万円前後であったが、教授は 250 万円以上の平均額であるのに対し、准教授、助教ともに平均額は 250 万円以下であった。

基盤研究 (A)・(B)・(C) の採択された課題名について、ライフサイエンス分野の語が多く観察された。なかでも「細胞」の頻出が顕著であった。2015 年から 2019 年までの 5 年間で採択された課題名に含まれる語で特徴的であったのは「地震」、「災害」で、日本における未曾有の震災においては、科学技術による解決や予防が望まれているということであろう。

高度な統計解析には専門のソフトウェアの操作が必要で、命令文を入力して実行する方式が主であるために学習コストが高かった。今回、共起ネットワーク分析を行うために利用した KH Coder は GUI によ

る操作で分析ができるため、学習コストも低く、手軽に高度な分析ができるツールと言えるだろう。注意すべきは、社会調査データを分析するために制作されているという点で、今回のような研究費の分析での利用には、形態素解析のための専門用語辞書の用意やサブグラフ検出に関して研究費の分析にあった解析方法の探索が必要で現状のままの利用には限界があり、課題も残る。目的に合う解析を行うためには、目的を明確にし、解析の論理を構築した上で KH Coder の可視化にも利用されている R 言語等そのものを利用した統計的な解析が必要である。

上記の分析結果は、飽くまでも ARAKI システムで得られるほんの一部の結果にすぎない。このシステムを利用すれば、研究分析のディスカッションを行いながら研究助成プログラムにおける研究の状況を確認することが可能であり、目的が明確になればその場で解析結果を得ることができる。NISTEP では、このシステムを利用した分析により科学技術分野の技術俯瞰 MAP の作成なども行っている。また、このシステムは EBPM を推進するために開発を進めており、既に省内及び他府省庁からの相談については随時対応している。今後は、大学の運営や大学経営、産学官連携のための研究分析にも活用できないか探索したい。

参考文献・資料

- 1) 樋口耕一. テキスト型データの計量的分析 —2 つのアプローチの峻別と統合—『理論と方法』. 数理社会学会. 2004, 19(1), p.101-115.