

科研費データベースにおける研究課題と 実績報告との関連性について

○石塚隆男（亜細亜大）

1. はじめに

科学知識は研究活動により生み出され、研究論文として公表され、教育や後続の研究活動により普及していく。こうした科学知識のダイナミズムは研究者に限らず、広く興味のあることがらである。たとえば、田中耕一氏や白川英樹氏の例を出すまでもなく、毎年のノーベル賞受賞に関する新聞記事では成果の卓越さよりも新事実を発見したいきさつに多くの紙幅が割かれ、後々エピソードとして折に触れ、語り継がれている。

研究にはテーマ選定が全体の50%以上のウエイトを占めるほど重要であると言われているにも関わらず、たいした確信もなしに安易にテーマを選択し、苦勞することも多い。研究を始めるにあたってはテーマに関する何らかの問題提起とそれに対する作業仮説を立てることが必要となるが、論文作成まで機械的な流れ作業のように順調に研究が進むことはむしろ稀である。多くの研究は、実際にやってみて壁に突き当たり、そこで新しいアイデアが出たり、あるいは想定外の‘副産物’の方に針路変更し、それが論文につながっていくのが普通である。しかしながら、そうした紆余曲折の事実は論文には記載されず、論文は科学知識が当然のように導出されたかのように入れられ、余談を許さない。

ノーベル賞級の研究だけでなく、多くの研究者が行っている研究計画段階の課題と最終的な成果との関わりを明らかにすることは、単なる興味本位の問題ではなく、多くの研究活動にとって、‘他山の石’として参考になりうる‘メタ科学知識’である。こうした知識は、個別事例のミクロ的な研究過程とともに、全体的、マクロ的に見てどのような傾向にあるのか、あるいは分野による違いはあるのかなど今後の研究方向や科学技術政策を探る上でも役立つと考えられる。

そこで、本研究では研究計画段階における研究課題が最終的な研究実績や成果とどのような関係にあるのかを明らかにすることを目的とする。既に述べたように、研究論文からは初期の研究テーマや計画段階の課題を窺い知ることはできない。個々の論文は成果の一部を投稿する学会に合せて記述したものであり、初期の研究テーマと論文内容が一致しているかどうかは専門外の人間にはわからない。こうした状況を踏まえ、国立情報学研究所が提供する「科学研究費補助金採択課題・成果概要データベース」をもとに採択時の研究課題名と実績報告の文章との関連性について分析を行うことにした。

今回、工学領域の実績報告書データを解析したところ、発表論文数、配分額、研究概要等からいくつかの知見が得られたので報告する。

2. 科研費データベースについて

わが国では、政府による競争的研究資金による研究推進がなされ、その中で科学研究費補助金（科研費）は平成17年度で1,800億円を超える膨大な国家的経費である（遠藤(2005, p.3)）。その詳細は、遠藤（前掲）や毎年公刊される『文部科学白書』、『科研費ハンドブック』並びに日本学術振興会のホームページ等に述べられており、

改めて説明は要しないだろう。科研費による成果は、国立情報学研究所が作成・運営する『科学研究費補助金成果概要データベース』（通称：科研費データベース）により公開されている。科研費データベースでは、各採択課題について年度単位の研究実績報告書を検索・閲覧することができ、表1の項目が入力されている。

表1. 研究実績報告書の項目

報告年度, 研究期間, 研究課題番号, 研究課題名, 研究代表者, 研究機関, 研究機関番号, 研究分担者, 研究種目, 研究種目コード, 審査区分, 研究分野, 研究分野コード, キーワード, 研究概要, 発表文献, 工業所有権, 配分額
--

本研究で入手したい情報である研究計画段階の研究概要はデータベースに登録されておらず、研究課題名が唯一の手がかりである。また、発表文献数は、研究分野でかなりの差があり、人文社会分野では圧倒的に少ない。

3. 研究方法

研究計画段階における研究テーマと研究成果との関わりを明らかにするためには、ひとつひとつの研究においてこれらの対応関係にあるデータを入手する必要があるが、多分野にわたり相当規模のデータがアクセス可能な形で公表されているのは科研費データベースにおいて他にはない。

本研究では、科研費データベースの中から研究実績報告書の登録率が比較的高く、発表文献の数も多い工学分野を選択し、研究期間が2000年以降の研究実績報告書を対象とした。科研費データベースは、検索結果を画面で閲覧することを主目的に作成されているため、ある期間内のデータを一括でダウンロードする機能はない（というよりも、セキュリティの観点から手動以外のアクセスをシステムで禁止している）。このため、研究期間が1993～2002年の工学分野の検索を行い、1件ずつ研究実績報告書をテキストデータとして保存し、全部で1340件の研究実績報告書データを入手することができた。なお、各研究の期間は単年度のは少なく、大半が2～4年であり、原則として年単位に研究実績報告書を提出しなければならないとされている。しかし、各年度の研究実績報告書がすべて登録されていないため、欠測帳票が存在する。そこで今回は、各研究実績報告書を独立したデータとして扱い、複数年度にわたる研究実績報告書の連結処理は行っていない。

テキスト保存された研究実績報告書データは、プログラムにより一括編集作業を行い、オリジナルの項目に加え、表2の各項目を追加し、CSV形式で保存した。

表2. データから新たに生成・追加した項目

研究期間区分： 単年度（11）、2年間の1年目（21）、2年間の2年目（22）、・・・4年間の4年目（44）
研究課題名を構成する名詞の数： 研究課題の具体性の度合いを表す
研究課題名を構成する名詞の中で研究概要中に含まれているものの数と割合： 課題と成果の異同を表す
研究概要の文章中の漢字の2字熟語： 研究成果の特性を表す
研究概要の各文の末尾の述語（ex. ～を・・・した。）： 研究概要の文章構造の解析に用いる
発表文献数： 当該年度の発表文献の数（ただし、投稿予定や掲載予定と書かれているものもカウントし、工業所有権も発表文献のひとつとしてカウントした）
当該年度配分額

研究期間区分は、各研究実績報告書が複数年度研究のどの時期のものかを特定するための項目である。本研究の目的は、採択時の研究課題と研究実績との関わりや異同を明らかにすることにあるため、研究課題名と研究実績としての研究概要との突合を行う必要から、研究課題名を構成する名詞数とそれらが研究概要中に含まれる数や割合を指標化した。研究概要は 800 字程度の文章で記述されており、その分野の専門用語が頻出する。しかし、漢字 2 字の熟語の中には多くの研究実績報告書に共通的に用いられているものが多く、研究概要の文脈を特徴づけていると考えられるため、それらを抽出した。発表文献数は研究成果を定量化した指標として用いた。

4. 結果

収集されたデータのプロフィールを示す。

表 3. 研究期間区分別件数

研究期間	頻度
1年	39
2年-1	211
2年-2	317
3年-1	155
3年-2	253
3年-3	348
4年-1	3
4年-2	5
4年-3	8
4年-4	1
総計	1340

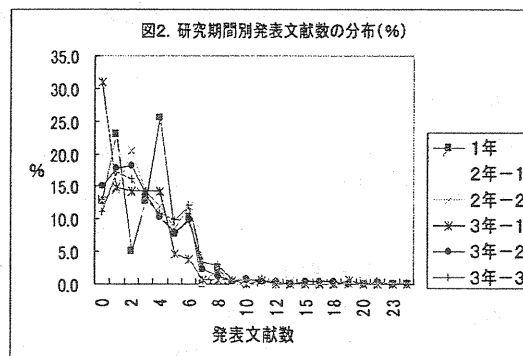
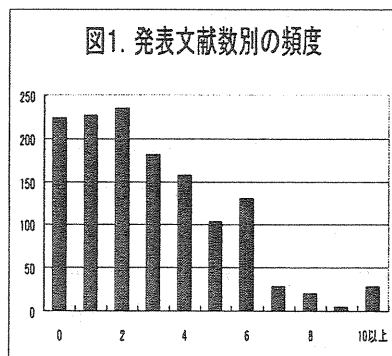


表 3 に、研究期間区分別データ件数を示す。本データは、悉皆調査ともサンプリング調査ともいえないが、2～3年にわたる研究が大半を占めていることがわかる。図 1 は、発表文献数別に研究実績報告書の件数を集計したものである。工学分野でも発表文献数が 0 件の研究が 1340 件中 200 件以上ある一方、10 本以上発表した研究も存在することがわかる。図 2 に、研究期間区分別の発表文献数の分布を示す。単年度研究でも発表文献数が 4 本のものが 25% あることがわかる。

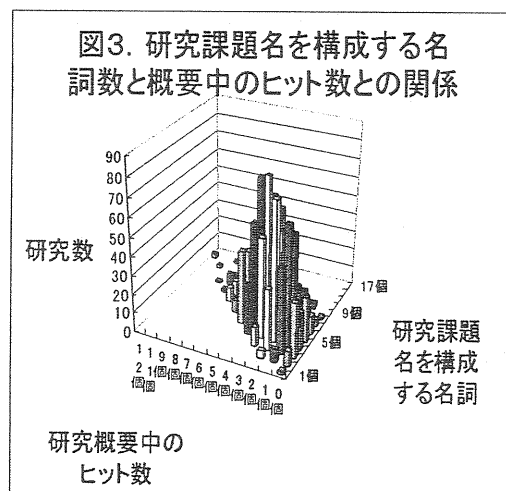


表 4. 研究期間区分別研究課題の概要ヒット割合階級別件数

研究期間区分	0	$0 \leq x \leq 0.3$	$0.3 < x \leq 0.6$	$0.6 < x \leq 0.9$	$0.9 < x \leq 1$	TOTAL
1年	2	8	12	12	5	39
2年-1	11	30	92	57	21	211
2年-2	25	56	128	78	30	317
3年-1	12	23	60	45	15	155
3年-2	22	40	102	79	10	253
3年-3	21	65	144	98	20	348
TOTAL	93	222	538	369	101	1323

表 5. 表 2 の特化係数

研究期間区分	0	$0 \leq x \leq 0.3$	$0.3 < x \leq 0.6$	$0.6 < x \leq 0.9$	$0.9 < x \leq 1$
1年	0.73	1.22	0.76	1.10	1.68
2年-1	0.74	0.85	1.07	0.97	1.30
2年-2	1.12	1.05	0.99	0.88	1.24
3年-1	1.10	0.88	0.95	1.04	1.27
3年-2	1.24	0.94	0.99	1.12	0.52
3年-3	0.86	1.11	1.02	1.01	0.75

図3は、研究課題名を構成する名詞が研究概要にいくつ含まれているかを集計した結果である。研究課題名を構成する名詞数により当該研究課題が具体的か抽象的かを定量化し、研究概要中にそれらの名詞が含まれている数や割合を調べることで申請時の研究課題と研究実績との関係を示した。表4は、研究期間区分別に含まれている課題の名詞の割合を集計し、表5はその特化係数を求めた結果である。研究概要中に課題中の名詞が100%含まれている研究が101件ある一方、0個のものが93件あった。

表6、7は、発表文献数を特徴づける単語を研究概要から抽出した結果である。発表文献数6本以上のデータは211件存在し、全体(1340件)の約16%に相当する。各熟語が出現したデータ数と比率を算出し、比率の差の検定により5%有意($|Z| \geq 1.96$)となった2字熟語を抽出した。図4は、配分額と発表文献数の関係を散布図で示したものである。単年度に100万円前後の助成を受けた研究が発表文献数のバラツキが大きいことがわかる。

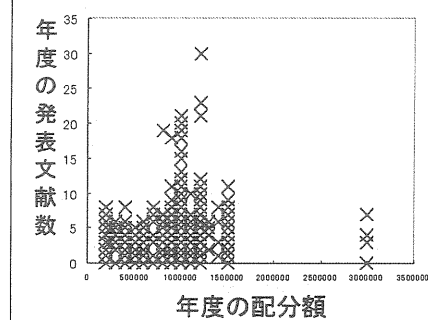
表6. 発表文献数が6本以上の研究概要に有意な2字熟語

熟語	6本以上%	Z
起因	29.49	3.18
試作	31.91	2.95
処理	26.76	2.45
考慮	23.61	2.42
設置	30.00	2.41
低下	24.53	2.35
適用	22.62	2.26
困難	23.39	2.20
拡張	26.92	2.15
領域	24.69	2.12
達成	24.66	2.01
提案	20.88	2.00
導出	27.50	1.99

表7. 発表文献数0本の研究概要に有意な2字熟語

熟語	0本の比率	Z
予定	32.79	4.42
計測	29.17	2.72
今後	25.00	2.57
現在	24.31	2.52
反応	23.75	2.21
部分	28.89	2.13
分離	28.89	2.13
顕著	27.12	2.07
示唆	10.00	-1.99
困難	9.68	-2.04
起因	7.69	-2.10
可能	12.32	-2.14
変換	6.06	-2.29
低減	5.08	-2.37
予測	6.25	-2.47
考案	2.27	-2.56
電流	0.00	-2.83
導出	0.00	-2.83
掲載	0.00	-3.13

図4. 年度の配分額と発表文献数の関係



5. 考察並びに今後の課題

本研究では、研究過程のダイナミズムもメタ科学知識と考へ、特に研究計画段階の課題やテーマと実績や成果との関わりや異同を明らかにするために科研費データベースを用いて自然言語解析手法を用いた分析を行った。いくつかの知見は得られたが、今回用いた指標からは明確な結論を導くことはできなかった。今回はマクロ的な定量化を行ったにすぎず、個々の研究課題と概要との関係は個別に精査する必要がある。また、複数年度にまたがる研究の扱いについても、トータルの発表文献数や配分額を用いて検討する必要があると考える。

参考文献

遠藤啓／編著(2005)『科研費の解説』ぎょうせい

「科学研究費補助金データベース」：<http://seika.nii.ac.jp>

日本学術振興会「科研費による成果」：<http://www.jsps.go.jp/j-grantsinaid/index.html>