

○山下 泰弘, 小林 信一 (電気通信大学)

## 1 はじめに

本研究は、新しい学際分野である環境科学分野のコンコダンスに関する分析を行い、さらに実際の論文における専門用語使用の特徴とその時間変化を分析する。以上の分析を通して、新しい学際分野である環境科学分野における専門用語使用の特性を明らかにし、さらには専門用語使用の時間変化に関する分析を通じて環境科学分野におけるトピックの変化を明らかにすることを目的とする。

専門用語は、研究者間のコミュニケーションを円滑にし、R&D活動を推進するための重要なツールである。それゆえ、専門用語の制定は、当該分野や関連分野の発展のためには欠かせない事業である。特に異なる分野の研究者同士のコミュニケーションを可能にするためにこの事業は重要である。

また同時に専門用語はR&D活動の動向を映し出す鏡でもある。R&D活動の成果物に対して、必ずそれに対応する語彙が与えられる。通常は把握が困難なR&D活動の実態を、専門用語使用の変化を通して見ることにより、客観的に捉えることができる場合もある。従来、R&D活動の把握のためには、citation indexingなどの論文単位の分析が主流であったが、本研究は論文中に踏み込んだ語彙単位の分析であり、新たな視点からの分析が可能である。

環境科学は、多数の分野と一部を共有する形で発展しており、他分野の語彙が非常に多く用いられている。そのため、環境科学分野の専門用語に関する調査を行う場合、環境用語のみならず関連分野の用語についての分析が要求される。本研究では、関連分野として学術用語集の各分野を取り上げる。

## 2 研究の概要

本研究では、コンコダンスレベルの分析と論文中における専門用語使用の分析を行う。論文中における専門用語使用の分析は、全体的な分析と専門用語使用の時間変化の分析の2種類の分析を試みる。

### 2.1 使用するデータ

環境科学分野の論文および環境用語辞書は、スキャナーで読み込んだものを人手により修正して使用する。論文は、通時的な分析を行うためにソースの種類が揃っているほうがよいので、1種類の雑誌から抽出して使用している。本研究で用いるデータは次の通りである。

#### (1)環境科学分野の論文

論文は、雑誌「環境情報科学」(環境情報科学センター) 1巻1号から21巻4号まで21年分の各号の巻頭論文(計84件)を使用した。

#### (2)環境用語辞書

環境科学分野の辞書には、「環境科学辞典」(東京化学同人)の索引から、見出し語部分だけを抜き出して使用している。

#### (3)関連分野の辞書

すでに電子化されている25分野の「学術用語集」(I P C発行、収録語彙数129,543語)から和語の重複を削除したものをを用いる。

## 2.2 分析の方針

分析は次のような方針に沿って行う。

### (1)分析は複合語レベルで行う

本研究では複合語レベルで語彙を扱い、語彙の切り出しの簡略化をはかった。テキストからの語彙の切り出しは、文を字種ごとに切り分ける方法を採用し、テキストから漢字とカタカナの文字列を抽出する方法を採用した。最終的には、人手による語彙の切り直しを行っている。

### (2)文字は漢字とカタカナのみを取り扱う

スキャナーで読み込んだテキストはひらがなの読み取り誤りが多い。また、人手による修正の際にもひらがなや英数字の誤りは見落としやすい。ひらがなや英数字を含む専門用語は量的にあまり多くないので、本研究では扱わないこととした。

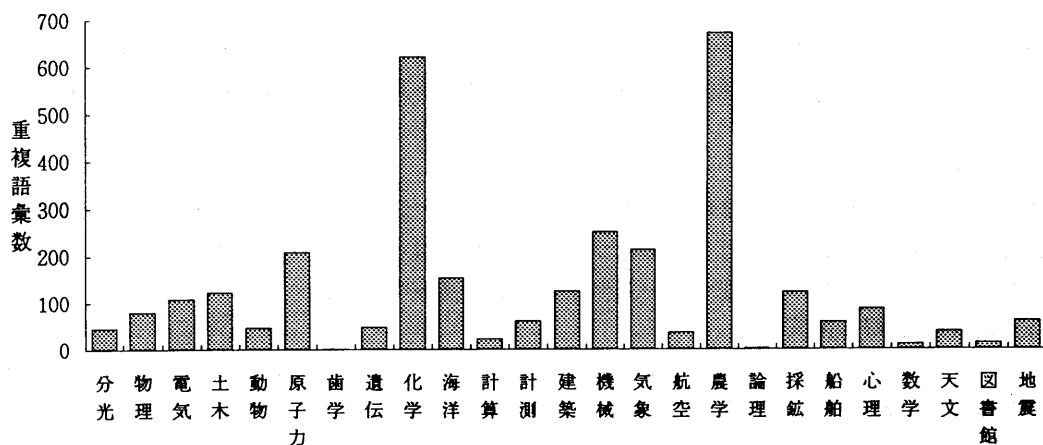
## 3 コンコーダンスから見た環境科学分野

本章では、学術用語と環境用語の2種類の辞書データを照合・比較する。それを通じて、環境科学分野の学術分野内での相対的な位置付けを調査する。

このようなコンコーダンスレベルの分析は、語彙の最小単位である語基レベルでの分析 (Ishii 1987) や対訳辞書を用いた分析 (科学技術情報研究所 1993) など、精密な分析がこれまでも行われているが、環境科学を対象として他分野との比較を行った研究は見られない。

まず、環境科学と学術用語集各分野の重複する語彙数を示す (図1)。環境科学と重複する語彙数(以下重複語彙数と呼ぶ) は多いのは化学と農学である。一方、歯学、論理学、数学、図書館学は殆ど重複が見られず、環境科学との関係は深くない。

図1 環境科学と学術用語各分野の重複語彙数



前述の分析において、環境科学と他分野の重複語彙数を示したが、重複しない語彙数については考慮しなかったため、2つの分野間の距離をこの分析結果から知ることはできない。そこで、次に、環境科学と他分野の親近・包含関係について分析する。ここでは、環境分野 ( $i$ ) と学術用語の各分野 ( $j$ ) につ

いて次の量 ( $X_{ij}$ ,  $Y_{ij}$ ) を求め, x-y 空間上にプロットすることにより, 環境科学と他分野の互いの親近・包含関係を分析する.

$$X_{ij} = \frac{f_{ij}}{N_i}, Y_{ij} = \frac{f_{ij}}{N_j}$$

点 ( $X_{ij}$ ,  $Y_{ij}$ ) は分野  $i$ ,  $j$  の距離が小さくなるほど原点から遠くなる. また, X軸に近づくほど分野  $j$  が分野  $i$  よりもより上位になり, Y軸に近づく程分野  $i$  が分野  $j$  よりもより上位になる. この結果を図2に示す.

図2 分野の親近・包含関係

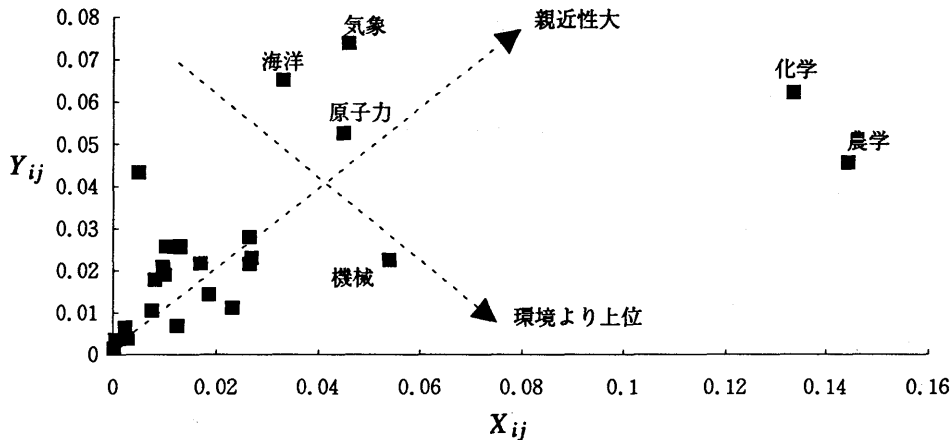


図2から農学・化学が, 環境科学と近い関係にあり, かつ環境科学をより包含する関係にあることがわかる. また, 学術用語の各分野で環境科学と関連を持つのは農学・化学の他, 気象学・海洋学・原子力工学などであり, 環境科学より上位に位置するのは機械工学, 下位に位置するのは原子力工学・気象学・海洋学であることがわかる.

#### 4 環境科学分野の専門用語使用の特徴

本章では, 環境科学分野の論文で使用されている語彙に関する分析を行う. これにより, 学際分野の専門用語使用の特徴についての知見を得るとともに, 専門用語使用の特徴の変化から環境科学分野におけるトピックの変化に関する分析を行う.

##### 4.1 全体的な分析

ここでは, 収集した84件の論文に学術用語や環境用語がどの程度用いられているかを分析する.

図3に論文中の語彙の出現頻度を示す(他分野との重複を含む). コンコーダンスレベルの分析(図1)では, 化学用語が農学と並んで多く環境科学と重複していたが, 実際の論文における出現頻度は農学と比べて少ない. 一方, 心理学用語はコンコーダンスレベルでは環境科学との重複は大きくはないが(環境科学との重複語彙数は87), 論文における出現頻度は農学に次いで大きい. これは化学用語があまり一般的ではない語彙が多いのに対して, 心理学用語は「図」, 「人」等かなり一般的な語彙が多いためと考えられる. しかし, 心理学用語には「アセスメント」, 「モニタリング」等専門的な意味で

用いられており、かつ環境用語ではない語彙も数多く存在する。また、他分野と重複しない語彙（以下独自語彙と呼ぶ）の頻度では農学を上回っており、実際の用語使用の面では環境科学と強い関係を持つと言える。環境用語は、化学用語と同様あまり一般的ではない語彙が多く、「環境」以外の語彙は全て頻度200以下となっている。数学用語は、心理学用語と同様に、コンコーダンスレベルでの環境科学との重複が少ないにもかかわらず、論文中での出現頻度が高い。これは、数学分野では「以上」「増加」などの一般的な語彙を意味を限定して用いているものが多いことと、数式やグラフを扱う際に数学用語が多く用いられることによる（例えば、「数式」、「縦軸」等）。天文学用語、図書館学用語も数学に次いで頻度が高いが、それらはいずれも専門的な意味で用いられていないものが殆どである（「出」、「約」等）。特に天文学用語として抽出されたものの中には、一般的な和語と単漢字の専門用語を混同しているものも多く見られる（例えば和語「出る」と天文学用語「出」）。これは、本研究ではひらがなを扱っていないために生ずる誤りであり、今後改善する必要がある。

図3 論文中の語彙の出現頻度

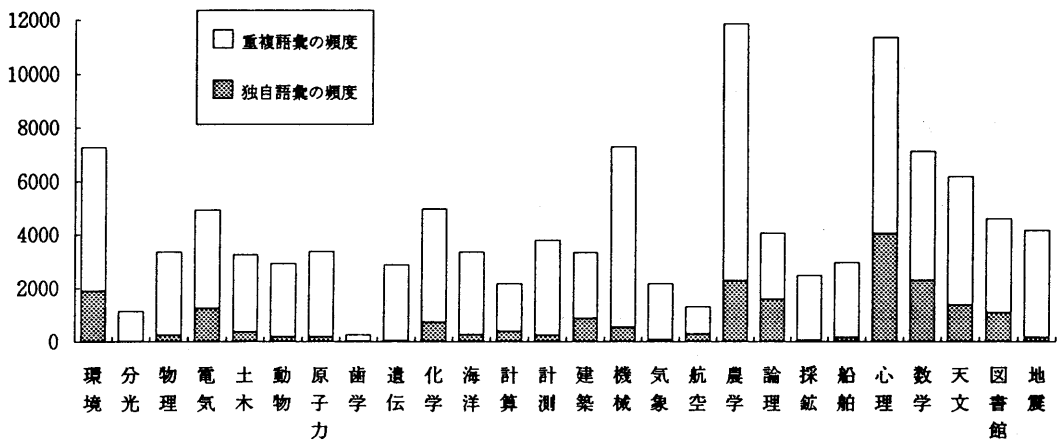


表1 各分野の高出現頻度語彙（括弧内の数字は論文に出現する語彙数）

順位	環境科学(801)		化学(154)		農学(1251)		心理学(570)	
	語彙	頻度	語彙	頻度	語彙	頻度	語彙	頻度
1	環境	750	水	291	環境	750	環境	750
2	人口	187	エネルギー	151	結果	228	必要	695
3	資源	176	汚染	118	計画	209	人	570
4	公害	169	形成	101	発生	204	図	391
5	廃棄物	167	二酸化炭素	90	人口	187	人間	375
6	環境庁	146	基	89	廃棄物	167	評価	239
7	健康	106	基準	80	調査	149	計画	209
8	環境保全	98	展開	79	大気	148	意味	205
9	生態系	95	分析	73	機能	144	発生	204
10	環境教育	94	空気	71	保全	129	人口	187

#### 4.2 環境科学分野におけるトピックの変化

学術分野におけるトピックは、その分野の研究のフロンティアの変化に伴い、常に変化している。特に、環境科学のような社会と密接なつながりを持つ分野のトピックは、そのときの社会情勢にも大きく影響される。学際分野の成熟する過程や、研究活動のダイナミクスについての知見を得ることを目的として、ここでは時系列に沿った分析を行う。ここでは、論文を第1期をVol 1～Vol 5(1972～1976)、第2期をVol 6～Vol 10(1977～1981)、第3期をVol 11～Vol 15(1982～1986)、第4期をVol 16～Vol 20(1987～1991)のように5年ごとの4期に分けて、各時期における環境科学分野におけるトピックの変化に関する分析を行う。

まず、最も高頻度で用いられている環境用語である「環境」という語彙の用法の変化を分析し、環境科学におけるトピックがどのように変化してきたかを分析する。

「環境」という語を含む語彙数は、第1期には106であったが、第4期には205とほぼ倍増している(図4)。また、「環境」という語を含む語彙の頻度は、第1期には554であったが、第4期には946まで増加している(図5)。一方、「環境」という語彙の頻度は、第1期から第4期までの各期においてそれぞれ146、224、183、176であり、単調な増加傾向は示していない。すなわち、「環境」という語を含む語彙の頻度の増加は、複合語の増加に依るものである。環境科学の確立して間もない時期には「環境」という語彙によりトピックの中心が存在していたが、研究の進展により多くの複合語が生じ、それらに中心が移行したためと考えられる。

図4 「環境」を含む語彙数の変化

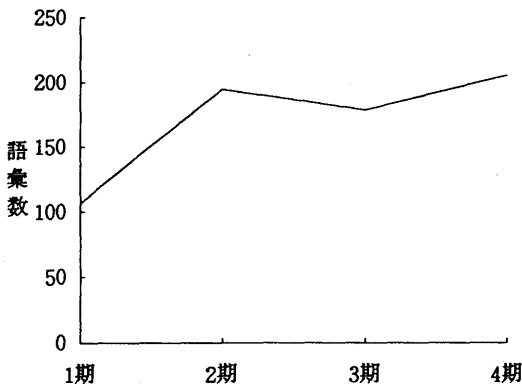
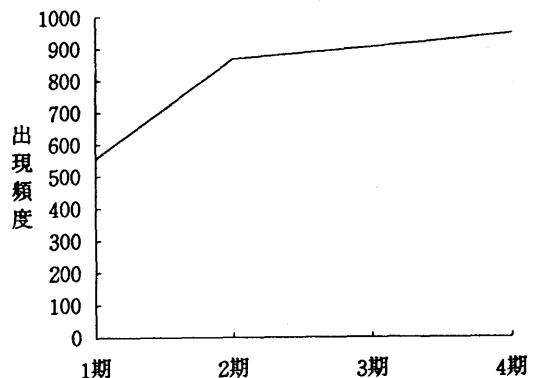
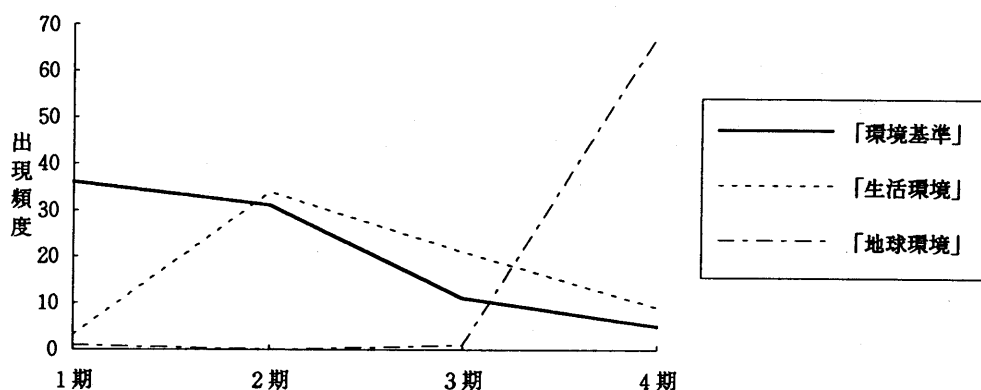


図5 「環境」を含む語彙の出現頻度の変化



次に、環境科学分野におけるトピックの変化を、特定の語彙の頻度変化より分析する(図6)。環境科学成立初期の公害問題の多発していた第1期には、「環境基準」という語彙が高頻度で用いられているが、漸減傾向を示し、最近では殆ど用いられていない。第2期・第3期には、「生活環境」という語彙が多く出現している。第4期になると、最近の環境科学の大きなトピックである地球環境に関する諸問題がクローズアップされるようになったため、それまでほとんど使われなかった「地球環境」という語彙が高頻度で出現している。

図6 高出現頻度語の変化



## 5. まとめ

本研究では、新しい学際分野である環境科学分野の論文の専門用語使用の特徴を分析した。その結果、環境科学分野の位置付けやトピックの変化に関する知見を得たが、同時に次のような課題を残した。

### (1) データの充実

本研究では「環境情報科学」という限られたデータのみを用いた。今後は環境科学の他の雑誌などを含めるなどしてより一般的な分析を進める必要がある。

### (2) 語基レベルの分析

本研究では、技術的問題等により、複合語を単語と同様に扱っている。今後は複合語を語基レベルに分解するなど方法により、より厳密に分析する必要がある。

### (3) 対象分野の充実

本研究では学術用語制定分野のみを関連分野として取り上げた。しかし、実際には環境科学はその他の分野（例えば医学・生理学など）とも関連を持っている。これらの分野と環境科学がどのような関係をもっているかの調査を行う必要がある。

## 文献

- [1] 科学技術情報研究所, "先端産業分野における専門用語の電子辞書データベース化の調査研究", データベース振興センター(1993).
- [2] Ishii, Masahiko, "Economy in Japanese Scientific Terminology", Terminology and Knowledge Engineering, pp123-136, INDEKS Verlag Frankfurt /M(1987).
- [3] 春山暁美他, "環境分野における検索用語共通化の試み", ドキュメンテーション研究, 33(7), pp319~327(1982).