

# 日本語コンテンツにおける科学用語の 共起ネットワーク分析

Cooccurrence network analysis of scientific terms using Japanese language contents

坂谷内 勝（国立教育政策研究所）

saka@nier.go.jp

Sakayauchi Masaru, National Institute for Educational Policy Research

## 【要約】

日本の数学・理科・情報の教科書索引から科学用語を収集し、科学用語に関する様々な分析を試みた。本稿では、複数の教科で共通に使われている科学用語の実態について、さらに、日本語コンテンツにおける科学用語の共起ネットワーク分析について報告する。分析に使用した日本語コンテンツは、読み物（理科系、文科系）と映画台本である。科学用語の出現頻度は、理科系読み物、文科系読み物、映画台本の順に多いことが明らかになった。

## 1. 研究の目的

本研究は、数学・理科・情報の教科書と学習指導要領に出現する科学用語を基本にして、科学用語の共起ネットワークを作成し、このネットワークと多種多様なテキストデータとの関連分析を行い、科学用語からみた科学教育の構造を把握し、科学教育の指導・学習に貢献する知見を明らかにすることを目的とする。

これまで、中学校・高等学校の「数学」「理科」「情報」の教科書の用語を分析し、用語の使用頻度の集計結果や用語の適切な使用方法等について報告されている。例えば情報の教科書では、頻出回数が多い用語は、「情報」、「インターネット」、「通信」、「データ」、・・・という順になっているという報告<sup>(1)</sup>や、化学の教科書で使用されている「イオン式」は不適切な用語で「イオンの化学式」が適切な用語であるという提案<sup>(2)</sup>がある。

しかし、多くの科学用語の分析は一つの教科・科目の教科書の中で調べており、科学用語が使用されている複数の教科・科目全体を対象とした俯瞰的な研究は少ない。ある学年のある科目で出現する科学用語が他の学年の他の科目で出現している事例がどのくらいあるのか、共通に出現する科学用語がある場合、出現の順序はどちらの科目が先なのかなどを明らかにすることは、学校の教師にとって有益な情報である。

2つの科学用語が文章中に近い距離で出現する頻度が多いとき、この2つの科学用語には共起関係が存在し、共起関係（リンク）と科学用語（ノード）を図式化すると科学用語の共起ネットワークを描くことができる。このネットワークを様々な観点で分析することによって、科学教育の全体的構造への示唆、教師への指導助言、生徒への学習支援に貢献できると考えられる。また、科学用語を指導する日本語教師や、科学用語を学ぶ日本語学習者にも貢献する知見が得られると考える。

共起ネットワークの例を図1に示す。これは、日本国憲法に出現する語を KH Coder<sup>(3)</sup> を使って描いたものである。

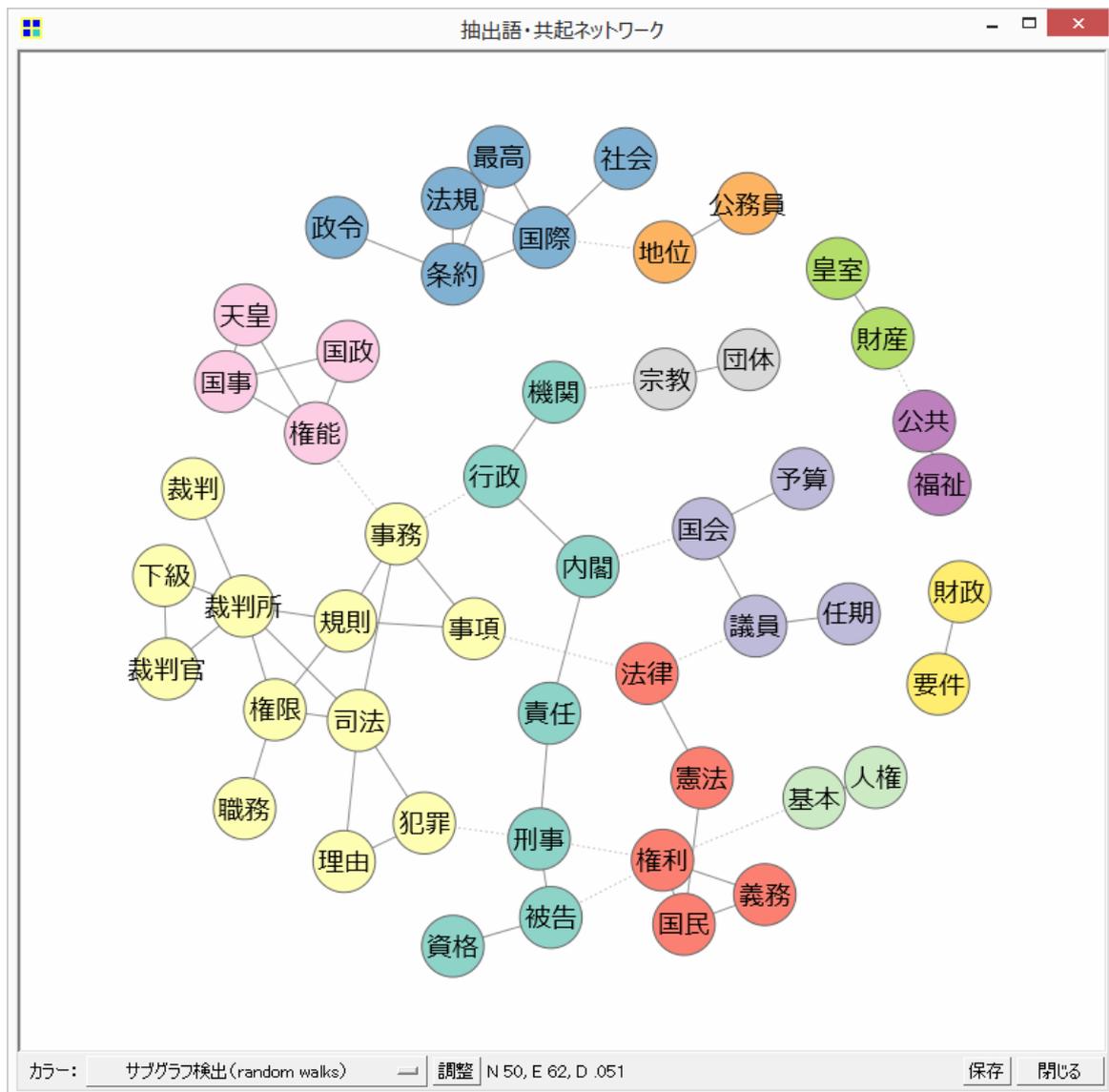


図1 日本国憲法の共起ネットワーク

## 2. 科学用語とは

科学用語は日本の教科書の索引から収集した。調査した教科書は、中学校と高等学校の数学、中学校理科、高等学校物理・化学・生物・地学、中学校技術・家庭、高等学校情報の教科書である。収集した索引の総数は 3552 件である。なお、教科書は、平成 13～14 年に文部科学省検定済のもので、今から約 15 年前のものである。

重複しない科学用語（異なり語数）は 3054 件である。異なる教科・科目で共通に出現する科学用語を調べると、4 科目以上で出現する科学用語は存在しなく、3 科目で出現する科学用語は 11 件存在する（表 1）。本稿では、このとき収集した科学用語を用いて、さらなる分析結果を報告する。

表1 3科目で出現する科学用語

| 科学用語        | 教科・科目    |
|-------------|----------|
| 電池、弾性、回路    | 理科／技術／物理 |
| 紫外線、位置エネルギー | 理科／物理／地学 |
| 光合成、単細胞生物、核 | 理科／生物／地学 |
| 光ファイバー      | 理科／物理／情報 |
| アンモニア       | 理科／化学／生物 |
| 周波数         | 技術／物理／情報 |

2科目で共通に出現する科学用語は、215件で、中学校「理科」と他の科目に出現する科学用語が最も多く131件存在する。相手の科目は、多い順に地学(44件)、物理(36件)、化学(27件)、生物(22件)である。中学校と高等学校の「理科」は、密接な繋がりがあるので、当然の結果といえる。

(1) 中学校「理科」と高等学校「地学」に出現する科学用語

P波、S波、オゾン層、マグニチュード、マグマ、温室効果、温暖前線、化石、化石燃料、火山岩、火成岩、火星、寒冷前線、金星、銀河、銀河系、恒星、黒点、酸性雨、示相化石、小惑星、食物連鎖、深成岩、震央、震源、震度、星団、石基、前線、太陽、太陽系、堆積岩、大気汚染、断層、地層、土星、等圧線、被子植物、偏西風、北極星、木星、隆起、両生類、惑星

(2) 中学校「理科」と高等学校「物理」に出現する科学用語

オームの法則、コイル、モーター、運動エネルギー、音源、火力発電、慣性の法則、虚像、原子力発電、作用点、磁界、質量、実像、瞬間の速さ、振動数、振幅、絶縁体、全反射、弾性エネルギー、抵抗、電圧、電圧計、電気抵抗、電磁誘導、電流、電流計、等速直線運動、導体、入射角、熱量、発電機、反射の法則、反射角、平均の速さ、誘導電流、力学的エネルギー

(3) 中学校「理科」と高等学校「化学」に出現する科学用語

アルカリ性、オゾン、ドルトン、プラスチック、ボルタ、塩、塩素、化学式、化学反応、化学反応式、化学変化、化合物、原子、混合物、再結晶、酸、酸化銀、酸性、酸素、蒸留、単体、炭酸水素ナトリウム、窒素、電気分解、二酸化炭素、二酸化硫黄、分子

(4) 中学校「理科」と高等学校「生物」に出現する科学用語

遺伝、遺伝子、液胞、感覚器官、気孔、減数分裂、細胞質、細胞壁、細胞膜、師管、受精、受精卵、生殖、精細胞、染色体、体細胞分裂、道管、反射、無性生殖、有性生殖、葉緑体、卵細胞

### 3. 文字数の少ない科学用語と多い科学用語

各教科・科目の科学用語の平均文字数は4.4文字である。長い科学用語を持つのは、情報5.6文字、技術・家庭5.3文字、短い科学用語を持つのは、理科3.6文字、生物3.9文字、数学3.9文字(中学校と高等学校全体)である。

最も文字数が少ない科学用語は1文字から成るもので、35件存在する。「B」(情報)と「g」(物理)を除いて全て漢字1文字である。1文字の科学用語は、「B、g、雨、雲、円、塩、核、逆、銀、月、鍵、弦、光、綱、行、項、酸、軸、髓、節、雪、霜、像、族、鉄、銅、虹、波、苗、腹、霧、目、卵、列、

胚」である。

最も文字数が多い科学用語は16文字から成るもので、「インターネットサービスプロバイダ」(情報)、「マルチメディアプレゼンテーション」(情報)、「エッジワース・カイパーベルト天体」(地学)の3つが存在する。情報の科学用語が長いのは、本来英語の用語を外来語としてカタカナで表記しているためである。10文字以上の科学用語は147件で、その内、情報は約半数(70件)を占める。13文字以上の情報の科学用語を以下に列挙する。

インターネットサービスプロバイダ, マルチメディアプレゼンテーション, データ管理を中心とするシステム, ハイパーテキスト転送プロトコル, プレゼンテーションソフトウェア, アプリケーションソフトウェア, インターネット・オークション, デジタルオーディオディスク, カーナビゲーションシステム, フォトレタッチソフトウェア, フォトレタッチソフトウェア<sup>(注)</sup>, ホームセキュリティシステム, リレーショナルデータベース, リレーショナルデータモデル, ローカルエリアネットワーク, 非対称デジタル加入者回線  
(注: 表記のゆれは入力ミスと思われる。)

#### 4. 身の周りで使用されている科学用語の実態

科学用語と日常生活の関連性を明らかにするために、身の周りの読み物(理科系<sup>(4)</sup>、文科系<sup>(5)</sup>)と映画台本<sup>(6)</sup>で使用されている科学用語を分析する。計量テキスト分析は、KH coder<sup>(3)</sup>を使用して行った。表2は、それぞれの総抽出語数、異なり語数、科学用語数とその割合(%)である。科学用語の出現頻度は、理科系読み物、文科系読み物、映画台本の順に多いことが明らかになった。特に、映画台本の中には科学用語がほとんど出現していない。

表2 総抽出語数、異なり語数、科学用語数とその割合(%)

|         | 理科系読み物 | 文科系読み物 | 映画の台詞  |
|---------|--------|--------|--------|
| 総抽出語数   | 65,607 | 61,395 | 14,822 |
| 異なり語数   | 4,791  | 5,507  | 1,816  |
| 科学用語数   | 360    | 85     | 11     |
| 科学用語(%) | 7.51%  | 1.54%  | 0.61%  |

理科系読み物「化学とんち問答」に頻出する科学用語は、多い順に、酸(理科、化学)559回、塩(理科、化学)240回、分子(理科、化学)195回、アルコール(化学)137回、原子(理科、化学)137回である。酸を含む科学用語は、多い順に、硫酸(91回)、酢酸(74回)、硝酸(29回)、脂肪酸(16回)、ギ酸(16回)、サリチル酸(13回)、濃硫酸(7回)、無水酢酸(6回)、希硫酸(4回)、テレフタル酸(2回)、熱濃硫酸(1回)、希塩酸(1回)、強酸(1回)である。

「硫酸」の関連語(共起する語上位12位)を表3に、硫酸及び関連語の共起ネットワーク(一部)を図2に示す。共起ネットワーク中、「硫酸」の近くに「ミスト」、「スモッグ」等の関連語が描かれているのは、原文に、「光化学スモッグの中に、酸性スモッグ、硫酸ミストと呼ばれる酸性のものがある。これが酸性雨の原因となって、森を枯らしたり、大理石の像を溶かしたりする。」という文章があるためと考えられる。

表3 「硫酸」の関連語（共起する語）

|    | 抽出語    | 品詞  | 全体          | 共起         |
|----|--------|-----|-------------|------------|
| 1  | 酸      | タグ  | 422 (0.147) | 28 (0.359) |
| 2  | 水      | 名詞C | 218 (0.076) | 14 (0.179) |
| 3  | アルミニウム | タグ  | 44 (0.015)  | 14 (0.179) |
| 4  | 塩      | タグ  | 175 (0.061) | 11 (0.141) |
| 5  | 紙      | 名詞C | 36 (0.013)  | 10 (0.128) |
| 6  | 性      | 名詞C | 72 (0.025)  | 9 (0.115)  |
| 7  | 作る     | 動詞  | 130 (0.045) | 9 (0.115)  |
| 8  | 使う     | 動詞  | 142 (0.049) | 8 (0.103)  |
| 9  | 銅      | タグ  | 45 (0.016)  | 7 (0.090)  |
| 10 | SO     | 名詞  | 22 (0.008)  | 7 (0.090)  |
| 11 | ナトリウム  | タグ  | 37 (0.013)  | 7 (0.090)  |
| 12 | 加水     | 名詞  | 16 (0.006)  | 6 (0.077)  |

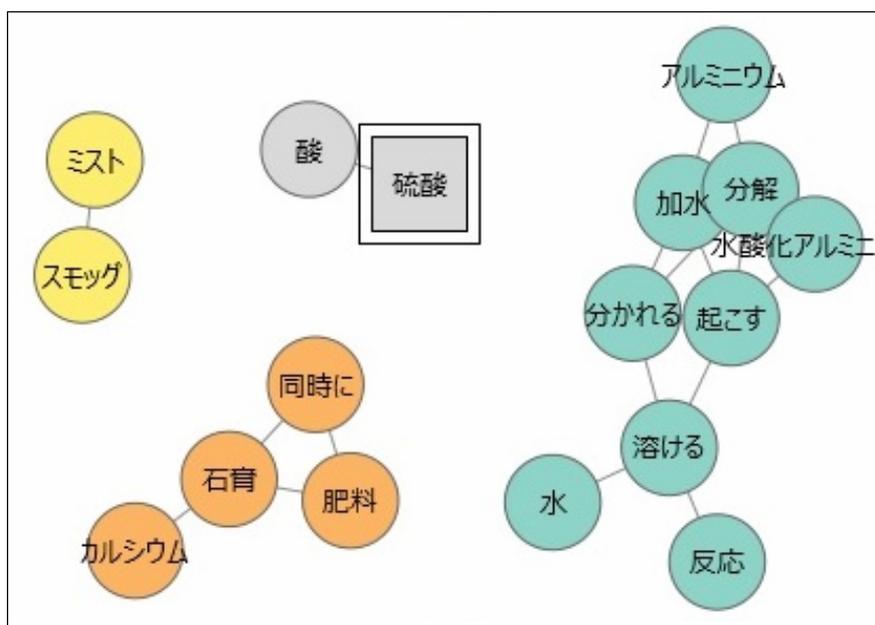


図2 硫酸及び関連語の共起ネットワーク（一部）

文科系読み物「俳句のたのしさ」に頻出する科学用語は、多い順に、雪（理科）60回、月（地学）59回、以上（数学）38回、逆（数学）24回、光（物理）24回、波（物理、情報）21回、焦点（物理）20回である。理科に関するものは科学用語というよりも自然を表す言葉として使用されおり、数学に関するものは日常的によく使われている言葉である。表4に「月」の関連語分析の結果を示す。「月」と共起する語として、「句」、「夜」、「感じ」という語が抽出された。これらの語の共起ネットワークを図3に示す。





の順に多いことが明らかになった。また、各コンテンツに出現する語の共起ネットワークを描くことによって、科学用語がどのような語と関連しているのかが明らかになった。今後の課題は、様々な日本語コンテンツに同様な分析を試みて、科学用語の実態と特徴を明らかにしたい。

#### [謝辞]

本研究の一部は、JSPS 科研費・基盤(C)17K01009（代表者：坂谷内勝）の助成を受けて行った。

#### 参考文献

- (1) 加納寛子、他：「文部科学省検定教科書高等学校「情報」の用語分析」、『日本科学教育学会年会論文集 37』, pp. 152-155, 2013.
- (2) 公益社団法人日本化学会：「高等学校化学で用いる用語に関する提案（1）」, <http://www.chemistry.or.jp/news/press/1-1.html> (2015年3月17日).
- (3) 樋口耕一：『社会調査のための計量テキスト分析 —内容分析の継承と発展を目指して—』, ナカニシヤ出版, 2014.
- (4) 米山正信：『化学とんち問答』, 講談社, 1991.
- (5) 鷹羽狩行：『俳句のたのしさ』, 講談社, 1976.
- (6) 山田洋次・朝間義隆：『男はつらいよ・ぼくの伯父さん』, 松竹株式会社, 1989.