

知識の ' 柔らかさ ' 計量の 試みとその意義

藤原 鎮 男

- 1 定 義 : 相関の度合いで見る ' 知識の柔らかさ '
- 2 資 料 : 化学文献語彙、国文学連歌語彙
- 3 相 関 の 構 造 : 類推、自然界の類縁事例、本稿の例
- 4 結 語

概要

知識は人間の知的活動の成果であり、文化や科学の本体である。本報告は、この知識の特性を定量的に捉える試みである。試みは、<人文科学、自然科学を通じてその「知識」は「普遍と個別」の二面から成る>という認識と、<この二面は、知識の表現媒体である語彙の出現の様相で分かる>という考えに立っている。

実際に科学および国文学資料の語彙の出現頻度を解析した結果は、出現頻度上位の語彙と下位のものとのそれぞれこの二面に対応することを具体的に示した。とくに最上位から数えて数番までの語彙群と、最下位から上に向けて数番の語彙群は、順位に対して厳密な指数関係に立つことが明かになった。指数関係であるので、上位の数語は全語彙の出現頻度の過半を占めることになる。すなわち、それらは全体に相関することになるので、これらの語彙が優越するドキュメントは柔らかいと定義し得ると考える。すなわち出現頻度から普遍的ないし個別的語彙の別を知り、その相対比率で知識の ' 柔らかさ ' が計量し得ると考えるのである。

以上の普遍と個別の語彙の分類は、それらが均一分散であるとして見た結果である。こうして分類された普遍性ないし個別性の語彙群について、さらにそれらの間の相関を調べたところ、化学語彙ではその専門分科の特性に、国文学資料では情報表現の特性に対応する二次の相関の存在が見出された。これは、知識の高次構造の指示である。すなわち、<語彙の解析が、知識の専門分科の形成ないし、知識の表現特性の把握>に及ぶことを示唆するのである。

あらためて全体を見直すと、最初に見出された相関は成分間の相関が弱い段階であり、各成分は個別、独立に近いとされる。ただし、相関はあるので外力の影響は全体に吸収分散され、その意味でこの系は<柔らかい>ことになる。近似的には系の全成分は一様、均一である。これに対して第二段階では成分間の相関はある限度を超えた状態であり、「知識」にクラスター（群、ないし専門と言ってもよい）構造が生まれることになる。

この「知識における構造の認識」は、広く見ると、自然界の基本現象として他にも見出される。かくて、ここで述べた相関による「知識の柔らかさ」の計量は、人文科学と自然科学を通ずる「自然原理」へのアプローチになると考えられる。これをより一般的な解析演算路に乗せることを今後の情報知識学の課題として努力したい。

1 定 義

「柔らかい知と硬い知」 : 語彙の解析結果による定義

演者は先年来、語彙の研究に関わって来た(1)(2)。その経過においてとくに語彙の相関に注目し、その出現頻度を解析した結果(3)(4)、概要に記したように、柔らかい語彙、硬い語彙の分類を考えるに至り、さらに、語彙でつ

くられる文献資料、したがってまた、知識もその柔らかさを同様に考え得ると思
うに至ったのである。

2 資料

語彙

演者は先に、わが国の文化の宗本として連歌を採りその語彙(3)と、科学で
は文献資料の組織化が進んでいる化学について同じくその語彙(4)の出現頻度
を調べた。経過を述べると、(4)は大量情報のデータベース利用が可能になっ
た直後、120万篇の化学論文の約1千万のキーワードを集めて解析した結果で
あり、(3)は個人の研究室レベルで情報処理が可能になった時点で、室町時代
の連歌十種千句の表八句の語彙3600語を解析したものである。連歌を採った
理由は、連歌が古代からの日本文化を継承し、それが俳諧、俳句を通じて現代文
化につながるものであり、すなわち、日本民族の意識の表示と考えられたからで
ある。室町時代の連歌を採ったのは、連歌がこの時代に満開し、その代表たり得
るからである。

解析の結果は、(3)(4)を通じ、<出現頻度が上位の語彙は、その順位と
出現数が指数関係に立つ>ことを明らかにした。自然語文献の語彙の出現頻度と
その順位の間数関係は既に述べられているが、顕著な指数関係の指摘は本研究が
初めてであろう。とくに、自然科学と国文学歴史資料を通じての知見としては新
知見である。科学と人文に亘るといふことは、それが自然の基本則であることを
示唆する。

指数関係であるので、例えば(3)で得た上位の約10語、月、秋、空、春、
花、雪、風、遠き、声、露の出現数は全語彙の出現総数の過半を占める。連歌の
表八句の語彙であるという制限はあるが、我々はこれを、理由をもって国文学の
代表として取り上げたことであるので、演者らは、これらの語彙を日本民族の情
感の基調を表現するものとした。

最下位の語彙は個別性を代表する語彙で出現頻度は1、2、3回である。これ
らの語種数と下からの順位を調べたところ、ここにも指数関係が見出された。こ
れは語彙の発生が指数性であることを示唆し、高次の分類構造がここに存在する
ことを窺わせる。このことについては次節3でもう一度触れる。

3 相関の構造

類推

演者はここであらためて、「宇宙における全元素の総存在量が原子番号と指数
関係にある」ことを想起した。原子番号が1からわずかに数番までの水素、ヘリ
ウム、炭素、窒素、酸素などの数元素の存在量が全元素の総量の過半を占めるの
である。地球も同じである(理科年表 丸善)。これは全く語彙の状況と同じで
ある。

この元素が自然界では群をつくる。そこで類推を敢えてし、語彙の性質、すな
わち、「知識の柔らかさ」にも同様な構造があると考えられるのである。

自然界の類縁事例

相関の代表例は宇宙の天体である。宇宙を構成する元素の、個々の存在量は原
子番号に対して指数関係にある。しかしながら、他方、これらの元素は現実には
太陽系などの群をつくる。原子も同じである。原子は原子核と電子で出来、その
電子はグループをつくり、周期律を形成している。このことは本稿の論旨によれ
ば、次のように言える。すなわち、系を構成する個々の成分を、個別の存在とし

て見れば各自自由であり、その数は原子番号と指数関数関係を示している。この時点では相関は弱く、各自は自由状態の摂動範囲にとどまろう。この範囲の状態を一次相関としよう。しかしながら、それは生まれた直後のことであって、相互の相関がすぐ強まり、相関が高じ、次の段階として成分が群（クラスターと言ってもよい）をつくる段階に入る。自然界において、宇宙の天体や原子構造はこの段階の存在である。このことがある以上、語彙にも同様のことがあるのではないかと思われたわけである。

国文学語彙

以上の視点で国文学の連歌の語彙、化学文献の語彙について検討をした。その結果、演者はそこにクラスター構造の存在を実証することが出来たのである。

連歌では、資料とした語彙を文章性の語彙L、画像性の語彙G、ファクト性の語彙Fに分類した。当然それらは普遍性の頻度上位の語彙でなく、個別性の部類に入る語彙である。それらについて、それぞれの出現頻度を調べたところ、各々について、その頻度と順位との間に指数関係があることが明かになったのである。このことは、〈当初は一応の判断でL、F、Gの分類をしたのであるけれども、得られた分類の各々が、それぞれ独立に指数性を示すのであるから、本質においてそれぞれは一つの独立世界を形成する〉ことの証左である。しかも、この分類は国文学の専門によっても支持されたのであった。すなわち、中世以後の国文学には「体」、「用」などの分類があり、本研究で見出した分類結果は、この伝統的視点の見方に一致したのである。これは、我々の柔らかさの解析が〈時代と所を超えた自然則に則っている〉ものであることを示唆しよう。

化学語彙： 化学の語彙についても同様の結果が得られた。

化学語彙全体はすでに述べた所と同様に出現頻度とその順位と指数関係に立ち、普遍性で柔らかい語彙と、個別性で硬い語彙に大別されることが明かになった。

ところがさらに立ち入ると、硬軟が逆転したクラスター構造が見出されるのである。すなわち、化学語彙の全体を25部門に分け、各部門毎に出現頻度上位の100語彙、すなわち‘柔らかい語彙’のみを集め、それらが所属する部門の間での相関を調べると、部門間の相関に強弱があり、グループ分けが出来ることが明かになった。有機化学、薬学系の部門は相関が強く一群をなしていれば硬い群であり、これに対して物理性、分析化学系の部門は柔らかい。

以上は我々が経験的にすでに持っていた分類、すなわち、国文学の「体」、「用」、あるいは化学の「専門分科」というものが、より根本的な、「自然」の属性に由来することを示唆すると思われる。

4 糸吉 言吾

以上、‘知識’が天体や原子構造と同様に構造を持つこと、したがって、その意味で自然の法則に従うものであることを指摘した。これは「知」における新知見であろう。これは上記〈一次、二次の相関構造を決める‘しきい’値〉の解明や〈専門性の発生と知の展開の関係〉の考察などの課題を提示する。

相関は非線形性の相互作用の結果であり、実社会の事象すべての通性である。この実社会の実データについて、その非線形性、とくに本研究で指摘した「知識の構造」まで考慮した演算法の開発は情報知識学の今後の課題であろう。

付言 近代科学のデータについても本稿の論旨による硬軟の論議が出来る。ここでは述べないが、例えばNMRのデータは、化合物という限定された空間における原子間の相関が、NMR化学シフトと呼ばれる σ （シグマ）値とスピン結合

定数と呼ばれるJ (ジェイ) 値であらわされる。 σ は分子を構築する原子の全体に依存し、いわば分子全体に相関するので柔らかい性格のデータであり、したがって非線形性の量であり、Jは直接結合の原子間の局所的相関であるので硬い相関にあたる。この見方で言えば、化合物そのものが、ここで論じている二次相関の構造を持っていると言える。さらにその内部を見ると三次の構造まで示唆する知見が浮かぶのである。化合物内では σ 値への置換官能基の寄与が加成性で、線形性であり、J値にはそれがないのである。ここではこの事実に言及するに止める。

文献

- (1) 藤原鎮男監修 「科学大辞典」 丸善 1985 東京
- (2) 藤原鎮男、藤原譲編 「総合学術用語集」 紀伊乃国屋書店刊 1987東京
- (3) 藤原鎮男、立川美彦 「連歌の語彙に見る普遍性と個別性」、
国文学研究資料館紀要 22号 213頁 平成8年
- (4) S.Fujiwara, M.Yokoyama, 'Analysis of Key Words in Chemistry'
J. Chemical Information & Computer Sciences, 21, 66 (1981)

A STUDY AND IT'S WORK ON THE QUANTITATIVE EVALUATION OF 'YAWARAKASA (SOFTNESS)' OF KNOWLEDGE

SHIZUO FJIWARA

Analysis is made on the frequencies of the appearance of ten million key words of 1.2 million chemical articles and of 3.6 thousand terms of the typical texts of the Muromachi Renga. The former represents the literature resource of the modern science and the latter that of the classical Japanese culture. The analysis has evidenced that a logarithmic relation holds between the frequencies of the appearance of the top terms and the order of their magnitudes.

By the nature of the logarithmic relation, the top few terms of higher frequencies dominate the total appearance of all other terms, implying that those few terms represent the whole system. The top few terms could be referred to as those of the general nature and the bottom ones as those specific. After we classify the terms into two types of the general and the specific natures, we will be able to evaluate the character of any document by counting the relative ratio of the numbers of the general terms to the specific ones. By definition, any document dominated by the terms of the general nature could be referred to as the 'soft' one.

As further analyses are made with respect to correlations among the terms of the top, and the bottom groups of the appearance, separately, another evidence of correlation has been noticed on both groups. This finding will refer to the fact that the knowledge forms another structure of the higher order. Analogy is referred to for the structures of the universe or those of the atoms.

神奈川大学総合理学研究所、 259-1294 平塚市土屋2948
Research Institute for Information and Knowledge,
Kanagawa University, 2948 Tsuchiya, Hiratsuka 259-1294