

連歌語彙に見る情報の通性 ：普遍と個別、文章性・画像性・ファクト 性の三態

Fundamentals of Information as seen by Analyses of Terms of Renga

○藤原鎮男、立川美彦

Fujiwara Shizuo & Tachikawa Yoshihiko

Abstract

As the followup of the former analyses of the frequencies of appearance of keywords of the chemical literatures which has revealed the dual characters of specificity and generality, the terms of renga has been analysed. Same results as above have been obtained. The sample terms of Renga have been classified, furthermore, into three codes of literature-, graphic- and factual natures according to the general schemes of presentation of information. The terms which belong to each code has shown the dual characters as cited above. Hence, it is concluded that the dual characters of generality and specificity and the presentation of information in the three modes form the bases of information and those bases can be studied by their frequencies of appearance.

1 はしがき

普遍性を旨とする学術論文のキーワードが普遍性のものと個別性のものの二群に分かれることを先に明らかにした(1)。よって我々は、個別性に勝つと推察される国文学の語彙を、この視点で解析することを試みる。

さらに、情報の表現は、文字による文章と、画像と、数値・事象に関わるファクトの三態によってなされている。すなわち、この三態は情報の基本の要素なのである。しかしながら、「情報」は今日まで、量的には文章が主体である。

我々は本研究で、文章の構成要素である語彙に、この三態の特性が付与されていて、語彙をこれらの三態に分類することが出来ることを明らかにする。その情報学上の意義と今後の研究課題についても触れたい。

2 考えの基本

(情報の三態) 上記の情報の三態は、それぞれが異なる特長をもっている。この特長を文章性、画像性、ファクト性とし、L、G、Fで表すとしよう。

Lは一次元的な文字列を媒介として、この文字列を読者が読み進む間に、読者は著者の主張を読みとるのであり、いわば、その過程で著者主導により、著者と読者は認識形成の共同作業をするのである。それ故、Lは著者(主)主導の主体性を特長とする。

これに対して、画像は二次元、三次元的の世界の提示である。その構築には、かかった時間の経過があるが、それは表には出ず、情報の伝達は一挙になされ、受け手は瞬時に全情報を受け取る。画像情報は、すでに存在している客体情報であり、所与である。すなわ

ち、画像情報の本質は客体性にある。

ファクトや事象は、L、Gと異なっている。それは、主客と離れたところで決められた規約、約束、条件があって、それが衆知されていることを前提として、提示された情報である。数・信号・度量衡・時制などに関係する。ここでは規約性が大事である。

かくて、上に述べた普遍性と個別性、さらには、L、G、Fの三態は情報の基本的特性と言い得る。

(語彙の三態)

ところで文章はL性としたが、実際にはそう簡単ではない。「文章の表現」という事柄の実体は、文章に含まれる漢字、かな、カナや、それが構築する文字列構造が読者に誘起するものまで含んでいるのであって、それは、L、G、Fにわたるのである。我々は、そういう文章の持つ機能が、語彙の誕生と生育に間に語彙に染みこんでいいると考え、語彙を、その機能の考慮の下に、三態に振り分けた。

(連歌の語彙)

連歌は、万葉集から古今、新古今の和歌の伝統を継いで室町時代に隆昌し、俳諧、俳句の現代に結ばれるわが国固有の文芸である。すなわち、宮廷文化と今日の庶民文化の接点に立って日本民族の性情を窺わせる素材である。我々はこう考えて、これまで行って来た文章情報の用語・語彙の頻度解析を試みる本研究を企画した。

3 方法と対象

(一般性と個別性) 先に藤原は、化学の論文に使われる用語の解析をした(1)。すなわち、化学論文の要旨のデータベースの機械処理が始まった直後、約100万の論文の用語(キーワード)延べ約1000万について、それを25の、化学内の分野に相当するサブセクションに分け、各サブセクションごとに、出現頻度の解析をした。その結果、頻度と順位は対数関係にあり、一般語と、専門語に分かれること、

100位までの高出現頻度のキーワードについて、サブセクション間の相関を調べると、部門間に著しい差が見られることが分かった。有機化学、薬学系のサブセクションは相関が高く、物理化学、分析化学は相関がはなはだ低い。前者では専門語と一般語の数の比が高く、後者は反対であることなどの分野の特性が計量的に分かる。

(情報の三態) さらに埼玉県約100の市町村の広報紙の最初の数ページの紙面について、文字文章と画像と告知や数表などのファクトのそれぞれが占める面積の相対比、すなわち、L、G、Fの比を測ったところ、広報紙の通性や市町村の特長を計量的に窺い知ることが出来た。

これらの結果は、語彙の頻度解析が有用な情報学の研究方法の一つであることを示唆する。よって我々はこれを連歌に適用した。

研究対象：連歌千句の表八句

取り上げた資料は、ほぼ西暦1500年を中心とする前後50年間に編まれた10種の千句*の表八句である。これは、100の百韻の各巻頭の八句であるから都合800句となる。この800句を対象として、名詞、代名詞、動詞、形容詞、副詞など単独で思想内容を表現する言葉(詞、いわゆる自立語)を語彙として切り出した。助動詞、助詞など言

語主体の態度を表現する言葉（辞、いわゆる付属語）はこれを無視した。切り出された語彙は3600語であった。以下、「表八句の語彙」とはこの3600語を指す。

この内、語彙の種の数、すなわち異なり語の数は1143である。

（*美濃千句、因幡千句、表佐（おさ）千句、熊野千句、河越千句、三嶋千句、葉守千句、名所千句、東山千句、伊庭千句）

取り上げた千句は、宗祇らの連歌最盛期の作品である。連歌は百韻を一座とする。千句はこの百韻十巻を一座とし、かつ各百韻に独立性を持たせる。それ故、千句集は形式が整った大量の連歌のサンプルとして最も適した素材である。

さらに表八句は、百韻の導入部として作句に特別の配慮が加えられる定めであって、神祇・釈教・恋・無常・名所・そのほか差し出た言葉を避け、軽く安らかに詠むべきもの、つまり、平常、一般の詠みをする定めである。それ故、ここに用いられる詞は、わが中世文学の美を凝集した選り抜きの語彙群と言えるのである。

3 結果

得られた3600語の語種（異なり語）の数は1143語で、各語種の出現頻度を勘定し、頻度と頻度の順位の間隔を調べたところ、対数関係、すなわち、ガウス分布が成立していることを知った。これは当然、高頻度、少数の上位数語と、多数の低頻度の語彙群に分かれることを言っている。具体的には、全語彙のうち、「月」の頻度が突出して高く、「秋、空、春、花、雪、風、声」がこれに続き、「遠き、寒き」以下ゆるやかに下がってゆく。

ここで得られた語は連歌に普遍的な語彙であり、概念性、一般性の語彙群である。これは「普遍性の語彙群」とも名付け得べく、これこそ連歌の土台を構築する基本語彙である。言い方を変えれば、我々は、ここでわが国民の心情の基本要素を抽出することが出来たと言えるかもしれない。ちなみに、それらは、単語が多く、事物の本質的な特長や総轄的な事象を表し、多数の関連語彙を背景に持つことが特長である。

これに対し、順位の下方面を見ると、末尾第1席は出現1回の語彙群で、所属する語種数は688語である。第2席は出現2回で語種数は168語、第3席は69語。第4席は42語となる。語種は種の出現、発生頻度と見ることが出来る。試みに、末尾からの順位と種の数の間隔を調べたところ、ここにも対数関係を見出すことが出来た。ここに集まる語彙群は、「白雪、淡雪、浜風」「見え隠るる、枯れ立つ、更けわたる」など複合語が多い。それらは、事物の特段の状況、個別的名称、複合した特記すべき作用などを表し、具体的、特殊の指摘の語彙群で、「個別性の語彙群」といえる。

頻度8以上に属する語種は104語、7以下はその10倍の1040語である。その出現の延べ回数は、前者が1786回、後者が1814回で、全語種の3600回をほぼ二分する。すなわち、約9%の高頻度語種が発生語彙全体の半分を担う。

ここに得られた結果は、情報の基本である「普遍性」と「個別性」の相対的重みについて初めて得られた計数値といえよう。

連歌表八句の語彙の三態

3600語、1143語種を下記のように三コードに類別した。

実際の類別は最初藤原と立川が別途、独立におこなった。その結果を見ると、二つの結果はLとGの帰属において全く正反対となった。これは、そうなることが予備的に両目位で議論している間に予測されたことであり、それで、意識して類別を別個に行ったのである。それで、本稿では立川の類別で作業をすることにした。両者の類別が反対になる理由は、個々の「本研究の対象の語彙」に対する両名の認識に違いがあることのようなのである。立川はと検はは国文学の伝統的な見方で見、藤原は科学の経歴によって解析的見方によっているためである。

L類 動作性・形容性・記述的機能の強い語彙（行く、寒き、跡）

G類 対象性・客体性・視覚的要素の強い語彙（花、舟、時雨）

F類 制度的・歴史的に定義された語彙（春・朝・弥生・固有名詞）

この類別には中世以来の国文学の体用（たいゆう）の概念による語彙の分類をも参照した。

こうして類別した語彙群について、さらに出現頻度とその順位を検討した結果、我々はこのにも対数関係が成立すること、すなわち、普遍性の語彙と個別性の種の語彙の二群が分かれることを知った。

我々がここで得た知見は新規なもののように思われるが実はそうではない。すでに世阿弥は花伝書で普遍と個別の本体を論じていて、その論旨はここで述べたところと通ずるように思われるし、また、兼好は徒然草において、わずか1ページの記載になかで、我々が表八句の語彙解析で我々が最も普遍一般性の語彙として得ることが出来た八語彙を、好ましいものとして指摘していて、我々は、そのことに驚かされたのであった。

4 結び

連歌の表八句の頻度解析で得られた一般普遍性の語彙は我々の心情の深層の基本要素を具体的に示してくれる。全文データベースの時代に、ここで得られた語彙をツールとして、一步踏み込んだ、わが文芸の内容分析を行うことが出来れば、文芸の思潮関係の語彙のシソーラス作りを具体的、計量的に進めることが出来よう。

語彙の三態の解析は、それぞれの「態」について一般性の語彙が具体的に何であるかを示してくれた。これもまた、今後、語彙のシソーラス作りの力になろう。

それから、類別コード付けの際に看取された、解析の立場の見方と国文学伝統の見方との違いは興味ある問題を示唆する。

本研究は、語彙の頻度解析が、普遍性と個別性という情報の基本問題の研究法となること、さらに、語彙の、L、G、F性の考察によって、「情報」の本質を研究し得ることを示唆する意義があろう。

文献

- (1) S.Fujiwara, M.Yokoyama, S.Ueda, Analysis of Keywords in Chemistry, J. Chemical Informatin and Computer Sciences 21, 66-70 (1981)

(国文学研究資料館, The National Institute of Japanese Literature)