

などのような部分を含み、必要に応じて例文を提供することができる。

### (2) 文章推敲データベース作成

各種文章のアウトライン項目ごとに、そこで使われる語句とそれに対する言い換え表現とからなる辞書の集合である文章推敲データベースを作成する。異なる文章であっても、共通に使われるアウトライン項目もあろうと考えられる。

### (3) 推敲システムの作成

アウトライン・システムに従って文章を書いたならば、その文章の部分がどのアウトライン項目に属するかを調べて、置き換え表現を文章推敲データベースから検索し、文章の推敲や伸縮に役立てる。このシステムを作成する。

例として、手紙文(最長 305, 最短 196), スピーチ(最長 349, 最短 222), 説明文(最長 3524, 最短 2422) を実行した結果を示した。

## 品詞の使用率からみた和文体・漢文体の特徴

統計数理研究所 村上 征勝・岸野 洋久\*

鎌倉時代の学術・宗教関係の文章の多くは漢文体で著され、また和文が用いられた文章でも、ほとんどの場合漢文が併用されている。鎌倉時代の宗教家日蓮(1222年~1282年)の遺文の計量分析を進める上で解決しておかなければならない問題の一つに、漢文体・和文体というような異なった文体で書かれた同一人物の文章の間の特徴の異同の問題がある。

そこで、日蓮遺文を中心とした鎌倉期の文献50編を用いて、漢文体(分析には漢文を読み下した漢文訓読体を使用)と和文体(一応漢文混合率10%以下の文章を和文体とした)では、普通名詞、固有名詞、形式名詞、代名詞、数詞、動詞、形容動詞、助詞、接頭語、接尾語、形容詞、感動詞、助動詞、連体詞、副詞、接続詞の使用率に本質的に差があるかどうかを分析し、和文体の文献では動詞、形容詞、助動詞といった活用のある語が多く用いられ、漢文訓読体では普通名詞、代名詞、副詞、接続詞といった活用の無い語が多く用いられているという結果を得たことを報告した。

## コンピュータ使用によるキェルケゴールの文体の解析

大阪教育大学 教育学部 榊 形 公 也

キェルケゴールとコンピュータとの結び付きに関しては、一般には奇異な感じを与えるかもしれない。しかし、彼自身の思想のスタイルからして、コンピュータを利用して彼の文体研究を可能にするような要素がある。その理由をいくつか列挙すれば、次のようなものになるだろう。

キェルケゴールの著作には、本名での著作、仮名での著作、遺稿という三種類のものがあるということ。それに応じて、彼の作品は、宗教的なものと美的なものというように、大きく二つに分けられるということ。彼はまた人間の生存領域として三つの実存領域を提起し、それらの各領域における術語や文体に固有の特質を与え、その伝達の形式に関して固有のものがないければならないとしていること。更に、キリスト者としての語りかけのスタイルの区別として

\* 現 東京大学海洋研究所

は、説教、談話、講話があるということ。

このような、キェルケゴール独特の著作スタイルに注目して、モンリオール、マギル大学の A. マッキノン教授(A. McKinnon)は、今から二十年以上前から、コンピュータを使ってキェルケゴールを研究し始めた。彼はテキスト・データベースとしては、キェルケゴールの全著作を入力し、コンコーダンスとしては、膨大な『キェルケゴール索引』全三巻を出版し、ほぼ 660 のあらゆる出典箇所を提供し、それには原典第二版と、基礎にした第三版、英訳、仏訳、独訳の頁付けも付けてある。

マッキノンのコンピュータは所与の語や語のつながりが出てくる箇所をすべて、full context の中で、原典だけでなく他の版や翻訳の頁付けをもって、ディスプレイすることができ、更に、一つのコンテキストだけでなく、もっと多くの複雑なコンテキストで現れるもののファイルを作成することもできれば、それを全集第三版と他のエディションや翻訳の頁付けを付けて印刷することもできる。

70 年代後半にマッキノンは“aberrant frequency”(abfreq) words について考え始め、D. サンコッフ(David Sankoff)の助けを借りて、そのような words が、ある著作、または一連の著作群の中でどのような役割を果たし、或いは、ある著作、または一連の著作群の中のある概念に対してどう関係するかを identify するための formula を練り上げた。彼は更に、a change point detection program を発展させた。このプログラムは、ある著作、或いはある著作ないし一連の著作の中のある概念に対するある著作家の取扱いにおける主要な変化を identify するための方法としての、“abfreq” words の比例頻度における変化に焦点を当てているものである。この技術ないしテストは、すべての統計的に意味のある“cut(分岐点)”ないし変化の時点 location を identify し、その細部においては、その cut に有意味に貢献したそれぞれの語を示し、同時にその語の貢献の「方向性」と「広がり」とを示している。

ここ十年間というもの、マッキノン教授は Benzecri と Greenacre とによって唱道され、correspondence analysis(CA)として知られている“exploratory data analysis” technique を使って、大量の仕事をしている。CA は統計学の中で長い歴史を持っており、1935 年 Hirschfeld, 1940 年 Fisher によって記述され、唱道されたものである。その基本的な方法は八回ほど発見されており、J.-P. Benzecri 版が最新のものであり、Michael J. Greenacre の SIMCA というプログラムによって、成功を収めているものである。

こうして彼の研究は、従来では考えられなかったような方法で、テキストの本質的な、或いは基本的な論議を再構成し、テキストの構造を明らかにする方法を示すのに役立つ。彼の研究方法を学んで、コンピュータによるキェルケゴール研究をしている研究者としては、カナダ、トロント大学の A. カーン教授と、ニューメキシコ大学の A. パージェス教授、それとマッキノンの弟子の S. ホウグ(彼は、国際キェルケゴール文献目録のデータベース the Soeren Kierkegaard International Bibliographic Database を作成している)がいる。以下に彼らの業績を少し挙げておく。

McKinnon, A. (1975). Aberrant frequencies as a basis for clustering the works in a corpus, *Revue CIRPHO Review*, No. 6, 33-52.

McKinnon, A. (1980). Aberrant frequency words: their identification and uses, *Glottometrika*, 2, 108-124.

McKinnon, A. (1982). The du-man polarity in Kierkegaard's works, *Kierkegaard-Studiet*, No. 12, 3-14(「キェルケゴールの作品における「汝」-「ひと」の対極性」(北野裕通 訳), キェルケゴール研究, 第 12 号).

McKinnon, A. (1987). *An Introduction to Correspondence Analysis*, 1-12, Computing Center for the

Humanities, University of Toronto, Ontario.

McKinnon, A. (1989). Mapping the dimensions of a literary corpus, *Literary and Linguistic Computing*, 4, 73-84.

McKinnon, A. and Hogue, S. (1987). Uforanderlige and uforanderlighed: more about their differences, *Literary and Linguistic Computing*, 2, 98-107.

Kahn, A.H. (1982). *Lidenskab in Efterskrift*. in *Kierkegaard, Resources and Results* (edited and with an introduction by Alastair McKinnon), 105-118.

Kahn, A.H. (1985). *Salighed as Happiness? Kierkegaard on the Concept Salighed*, Wilfrid Laurier University Press, Ontario.

## 日蓮遺文の計量分析

### — 思想の変化と文体の変化 —

統計数理研究所 村上 征勝・岸野 洋久\*

立正大学 仏教学部 伊藤 瑞叡

報告者達は、この数年、鎌倉時代の宗教家日蓮の著作、弟子の著作等を用いた日本語文献の計量分析を試みている。今回のテーマは、日蓮の著作をもとに彼の文章に経年変化がみられるか、またみられるとしたら、それは何時頃、どのように変化したのかという問題である。

同一人物の文章の特徴には、一生変化しないものと、時とともに変化するものがある。今回は品詞の使用率を中心に、日蓮の文章において変化のない特徴は何か、変化した特徴は何かを、日蓮遺文 23 編を分析して調べた。

その結果、動詞、副詞は使用率には変化がみられなかったが、他の品詞（普通名詞、固有名詞、形式名詞、代名詞、数詞、形容動詞、助詞、接頭語、接尾語、形容詞、助動詞、連体詞、接続詞）の多くは、1265～1270年までの間に使用率に変化が生じているとの結論を得た。尚、今回の分析にあたっては漢文比率などの諸データを考慮に入れていないため、今後、それらの諸データによる修正を加えた上で、詳細な分析を試みる必要がある。

国際日本文化研究センターの梅原猛氏は「文体は思想の表現である」と述べているが、もし、そうであるならば、日蓮の文体は彼の思想が大きく変化したといわれている佐渡流罪時(1271～1274年)を境に変化がみられるはずである。佐渡流罪以前に多くの品詞の使用率に変化が起きていることは何を意味するのか、品詞の使用率にみられる文体の変化は思想の変化と関係がないのか等、日蓮遺文の計量分析を進める上での問題点を提示した。

## 学術文献情報の関係構造化とそのクラスタ分析

北海道大学 工学部 斉 藤 たつき

学術文献（ここではおもに研究論文をさすものとする）情報間の構造を解明し、研究に対する新しい着想や将来の研究動向を把握する活動は重要である。研究者にとって研究活動を進める上で、文献検索・サーベイ等は不可欠の作業である。学術文献集合を研究対象として、それらから客観的に観測される入力情報や抽出可能な概念情報、あるいは概念情報で構成される学術文献プロフィールの形成は大きな研究課題といえる。

\* 現 東京大学海洋研究所