

Archive)があり、既に25カ国920冊の機械可読テキストを世界各国の研究者に提供している。しかしこれまでは日本語のテキストは供託されておらず、「源氏物語」が最初のものとなる。

いずれにせよ世界中の研究者が「源氏物語」のテキスト・データベースの恩恵に預かることができるようになる訳で、遅滞きながらこの分野での国際的な学术交流へ多少とも寄与できるようになった。

今回の発表では、この程2年がかりで完成したパイロット・プロジェクトの作成経過並びに試用結果を中心に、テキストの選定、入力、作業経過、コピーライト、サービス、出来上ったテキスト・データベースの利用、問題点などを解説し、今後のテキスト・データベースの展望と実用性を検討した。

文章の伸縮とデータベース

大阪府立大学 総合科学部 樺島忠夫

文章伸縮の必要

今、ワープロやパソコンでレイアウトを行って、きれいな印刷用の版下を作ることができるようになっている。このとき、文章の最後の数行あるいは数文字が次頁送りになったために、その頁の大部分を空白のままに残さなければならなくなることもある。こういう場合、文章を縮めて、頁の中に収まるようにする必要が出てくる。また、図とその説明の文章との位置がうまく一致しないために、文章を伸縮して調節する必要が生じる場合もある。

こんなとき、数行または数文字を削除または追加するためにあれやこれやと頭を悩まさなければならぬが、これを自動的に行うことができるならば便利である。

文章伸縮の一方法

前回、統計数理研究所で行われたデータベース研究会で、パソコンによる手紙文の推敲システムを紹介した。目的に応じて、手紙の文例を画面に出し、必要に応じてその文章の部分と交換可能な表現を示し、そのどれかを指定することによって、要求に近い手紙文を作成するというものである。この方法は、手紙文に限らず、制限はあるが他の文章にも及ぼすことができる。

この推敲システムにおいて、文章の長さ(文字数)と文章を構成する各表現の長さを測り、文章の長さが指定した長さになる(または近づく)ように調節すると、文章伸縮の目的を達することができる。

推敲システムの作成

この推敲システムを作成するために、次のことを考えている。

(1) 文章作成支援アウトライン・システム作成

作成したい文章の種類(手紙、説明の文章、報告の文章など)と、必要な条件を指定すると、その文章のアウトラインが提供される。そのアウトラインは、例えば、論文ならば

問題提起、観察事実の報告、そこから立てられる仮説、調査方法、調査内容、考察、結論
依頼の手紙文ならば、

頭語、時候のあいさつ、感謝の言葉、経緯、依頼内容、結びの言葉、結語

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