

については、既にオンラインによる公開サービスを行っている大型計算機上の検索システムと比較しながら説明した。最後に、現在までに館内で出てきた改良すべきいくつかの点にも触れた。

自動品詞分解を通しての古語のデータベース化

大阪樟蔭女子大学 学芸学部 西 端 幸 雄

わが国の古典文学作品の語彙（古語）データベースを構築しようとする場合の作業の流れを振り返ると、

単位語句句切 → 語句の諸情報付加 → 手書きでカード化 → キーボードで入力

というように、すべての作業過程で人間が介在する。それに、その作業に携わる人間は、対象とする作品の大きさにもよるが、一般的には複数人間が必要となる。このように、複数人間によって作業を進めようとする、いくらしっかりした作業マニュアルを作っておいたとしても、結果的には、不統一な面が残る可能性が高くなってしまふ。

こうした人間が介在することによって生じる不統一を、パソコンを用いることにより極力少なくし、さらに作業の流れを自動化するというのが今回の試みである。なお、今回の作業では平安時代の和歌集を資料とした。

ところで、この作業のねらいとするところが、データベース作成に関わる作業過程での人間の介在を少なくする点にあるので、事前に単位語句句切と品詞分解を行なう際の基準となる『古語辞典』（約4万語）を作成した。この『古語辞典』には自立語辞書・助詞辞書・助動詞辞書の3種類がある。

まず、入力された本文データから、14文字を最長として、漸次1文字ずつ文字長を減らし、『古語辞典』との一致を検索させ、一致すれば、その一致語句の末尾に句切符号を付けるようにした。現在のところ、この単位語句句切の正答率は約90%で、誤った句切についてはキーボードより修正できるようにしている。

次に、すべての本文を単位語句に句切ることができた後、その句切符号付きの本文データに対して、品詞分解を行なわせる。この場合も、『古語辞典』を用いて、本文データの後部から前部に向けて順次取り出した句切符号までの語句との一致を検索し、一致すれば、『古語辞典』内に登録してある文法情報や接続情報を基に、当該語句の文法的性格を決定させるようにした。この場合の正答率も約90%である。

以上の作業が完了すれば、次のようなフォーマットで、データベースとして登録するのである。

A	B	C	D	E	F	G	H	I	J	K
A	歌番号			4	バイト	(半角4桁)				
B	行数			3	バイト	(半角3桁)				
C	本文語句			60	バイト	(全角30文字)				
D	見出し用語句			60	バイト	(全角30文字)				
E	漢字			60	バイト	(全角30文字)				
F	参照語句			60	バイト	(全角30文字)				

G 品 詞	2 バイト (半角 2 桁)
H 掛 詞	1 バイト (半角の「*」)
I 所 在	2 バイト (半角 2 桁) <1 行中の位置>
J 索引種	1 バイト (半角 1 桁) <W・K・N で表現>
K 作 品	2 バイト (半角 2 桁) <01~21・二十一代集>

このように、データベース構築を自動化することにより、これまでの人手に頼っていたときよりも、その作業にかかる時間を大幅に短縮でき、データベースとして登録するデータ内容も、統一化できるようになった。

「おくのほそ道」データベースの試作 ——研究者のための個人用データベースの一例——

図書館情報大学 図書館情報学部 松 本 浩 一

1. 人文系研究と個人用データベース

人文系の研究においては、研究資料の収集、蓄積、処理が個人的な環境の中で進められることが多いが、最近のパーソナルコンピュータとそのDBソフトの普及によって、そのような研究にたずさわる人たちのコンピュータ利用の機会は、一気に拡大されたように思われる。そしてその取り扱い資料が様々な形態を備えていることを反映して、パソコンの利用方法も様々なものが考えられる。たとえばある古典の作品のテキストに関するものであれば、以前から見られたようにその全文を入力して索引作成に利用したり、カードにかわるものとして、資料として有用な部分を抜出して入力しておき、論文作成の際にワープロソフトと組合わせて利用したり、あるいはある一部の記事や文書に簡単な要約やキーワードをつけておいて、資料索引として利用したりすることが行なわれている。またこの他に関連の資料や研究論文を扱った書誌のデータベース、民俗学の調査報告書など様々な調査のデータを扱ったデータベースなどが考えられる。しかしここで注目すべきことは、個人で人文系の研究を進めていく場合、これら様々な情報の間の関係が特に重要になってくることである。たとえば必要な資料の書誌データの場合であれば、解題、版本、研究論文あるいはその中の記事に対するコメント等の情報が、それに関連して必要となってくる。またテキストデータであれば、校勘、注釈、研究論文あるいは研究者自身のコメント、そして関連の資料等が関わってくる。しかし、いままでのところ、研究の過程で関わってくる様々な形態の情報を、パソコン上にいかに有効に組織化していくかについてあまり議論がなされていない。ここでは古典のテキストとそれに関係するいくつかの情報をと、パソコンのDBを利用して組織化を試みた一例として、吉田貴子氏の昨年度の卒業研究「おくのほそ道」データベースの試作を紹介した。

2. 「おくのほそ道」データベース

周知のように「おくのほそ道」の本文は、文章の部分と発句とからなる。ここではテキストは『校本芭蕉全集』にあるものを底本とし、関連情報としては、校異、注釈(菅菰抄等の古注、明治以後の新注)、関連資料(芭蕉の書簡、曾良旅日記、俳諧書留等)、研究論文等を取りあげ、PC98上にDBソフト「桐」を使用してデータベース化した。なおテキストや注釈の採用、また検索上の要求等については、専門の研究者にアドバイスを仰いだ。そして関連情報のうち、曾