

「源氏物語」ハイパー・テキストの開発の試み

長瀬真理

城西国際大学

パーソナル・コンピュータの発展と共に、文科系の教育や研究にもコンピュータが利用されるようになった。中でも文字や画像を同時に組み込んだハイパー・テキストへの関心が深まっている。又海外における日本語・日本文化研究も盛んになっており、基礎的な資料である日本文学や古典のデータベースの需要も高まっている。そこで既にオックスフォード大学と東京大学大型電算機センターで公開されている日英対照「源氏物語」テキスト・データベースを土台に、画像データも取り込んだハイパー・テキストの開発を試みた。

An Experiment for Developing a Hypertext of the "Tale of Genji"

Mari NAGASE

Josai International University

There has been a considerable growth in the development of computers which can be applied to various researches in the field of humanities.

The demand for hypertexts of literary texts incorporated with graphics are increasing. Good electronic texts are also required in the Japanese studies which is getting popular in the foreign countries. This paper describes the experiment of developing a hypertext build on the bases of the English and Japanese versions of the machine-readable texts of the "Tale of Genji".

1 はじめに

10年ほど前から欧米では文科系の教育や研究にコンピュータが積極的に導入され、テキスト・データベースやハイパーテキストが作成され多面的な授業や学際的な研究に広く利用されている。

同時に海外での日本語・日本文化研究も盛んになっており、基礎的な資料としての日本文学や古典のテキスト・データベースへの需要も高まっている。しかし残念ながら現在国際的に公開されている日本語のテキスト・データベースは非常に少ない。

このような現状に鑑み、日本の代表的古典である「源氏物語」を題材に、海外での公開も含め、他の研究者にも利用可能なハイパー・テキストの開発を試みた。なお今回の研究は既にオックスフォード大学と東京大学大型電算機センターで公開されている日英対照「源氏物語」テキスト・データベースを基礎にし、新たに現代語訳、注や画像をデータとして加えた。

アルネ・シフェールのフランス語訳「源氏物語」の翻訳も発行されたので、実験的にテキスト・データベース化しシステムに組込んだ。フランスでも日本研究が盛んであり需要も大きいことが予想されたからである。

1 システム開発

1.1 機種を選定

古典テキストのハイパー化は商業用も含め近年幾つか作成されているが、理工系研究者主導によるものが多い。使用されているマシンもワークステーションが中心で、近年安価になったとはいえ、まだまだ人文系の研究者には手が届かない。

しかしパーソナル・コンピュータ及び周辺機器の方も、ここ1-2年の内に劇的に安価になった。又マルチ・メディアの台頭により画像入力用のスキャナー等も安くても高性能な物が出回っており、ハイパー・テキストの作成や利用に関して、必ずしも高額なマシンが優位を保つ状況ではなくなった。

そこでプログラムの対応機種は、比較的安価に購入が可能で現在急速に普及が進んでいるWindowsマシンとした。

1.2 プログラムの開発

ハイパー・テキスト開発には、既存のオーサリング・システムやパッケージ・ソフトを使用する方法がある。しかし代表的な製品である”Tool.kit”などは既に販売を中止している他、ほとんどの製品が会社の紹介や各種商品の紹介を主眼としたスライドショー的なものが多く、今回のプロジェクトに使える物はなかった。そのため外注の形で自主開発した。

大枠の古典、現代語訳、英語訳、仏語訳を対象に複数のテストを同時に相互参照しながら読み進むためのシステム開発には、Windows.3.0のMDI(Multiple Document Interface)機能を使いC言語で作成した。又画像との連結にもMDIを使った。そのほか注の参照にはメッセージ・ボックスを利用した。

主な機能は次の通りである。(Max. として示した数値はプログラムを変更することにより Windows の仕様範囲内で変更可能である)

- ・単一の画面内に複数のテキスト窓(Max. 5-Files)を開く。
- ・各テキストはページ単位(Max. 12KByte * 300Page)に分割され、窓内に入りきらないテキストについては窓右のバーによりページ内で上下スクロールを行う。
- ・ページ送り(前、後、ページ指定)はメニュー項目をクリックして行う。
- ・各テキスト間の制御の移行は、テキスト窓内の何れかの個所をマウスでクリックすることにより行う。制御移行後の窓が画面の最前面に移動する。
- ・あるテキスト窓から他のテキスト窓へ制御を移行する際、移行前のページ位置に対応するページ位置(ページ情報が一致する位置)のテキストが表示される。
- ・ページは更に複数のセクション(Max. 300)に細分割可能であり、セクション情報が一致する場合にはその位置のテキストを表示する。(対応するセクション情報がない場合はページ先頭)
- ・各テキスト窓は画面内の自由な位置に配置でき、大きさも自由に設定が可能である。
- ・テキスト内の記述に関して注釈がある場合、別色表示とし、その参照個所をマウスでクリックすることにより注釈(詳細情報 Max. 2KByte)を画面中央のメッセージ・ボックス内に呼び出す。(注釈数 Max. 100/Page)
- ・テキストに関連した挿絵の表示を行う。(メニュー項目)

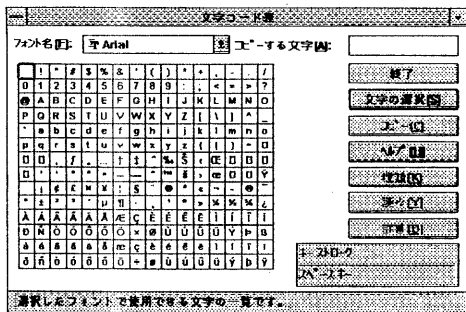
なお今回、対象としたテキストは源氏物語であるが、開発に当たっては他のテキストにも対応可能なよう、汎用性について十分に留意した。詳細は環境設定の箇所で解説する。

2. データ作成と加工

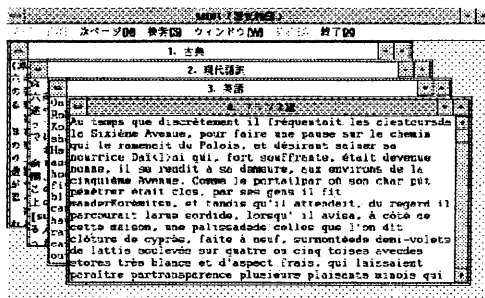
2.1 データ作成

古典テキストと英語訳については前回で作成されたデータを利用したが、新たに加えられた、現代語訳、仏訳、現代語訳の注についてはOCR(光学読みとり装置を使った。なお今回は試験的な開発ということで仏語訳と注の入力は4章の「夕顔」に限定した。英語の注は手入力した。

フランス語の入力は前回のプロジェクトでも協力してもらったサカタインクス(株)に依頼した。UNIXマシンで入力したためアクセント記号やセディーユなど英語にないキャラクターは文字化けを起こしていた。そこでWindows上では内蔵されている文字コード表の中のArialコード(図1参照)を使ってフランス語表示に切り替えた。



(図1) フランス語コード表



(図2) メニュー画面

2.2 テキストへの加工 (情報埋め込み)

今回、システムの対象としたテキストは源氏物語の古典、現代語訳、英語訳、仏語訳の四カ国語であるが、各々のテキストについて、相互参照や注釈の参照などをコンピュータで行うためにはテキスト自体に位置や注釈番号などの情報を埋め込む必要がある。

文章テキストに情報を埋め込む方法としてはText、SGML、RTFなど各種のMark-Up Languageが考案されワープロやDTP、印刷などに既に広く利用されている。又TEIのように、内容分析まで踏み込んだ付加情報を埋め込む試みも始まっている。今回はテキストに最も基本的な情報を、簡便に埋め込める方法として普及しているCOCOAフォーマットを採用することにした。なお加工作業は自動化ができないためすべて手入力で行った。

COCOAフォーマットは情報を特定の文字(テキスト内で他に使用されていない文字、一般的には<と>)で囲みテキスト内にタグとして埋め込んでいく方法である。

< に続く1文字が情報の種類を表すカテゴリーとして取り扱われ、カテゴリーの後に続く1文字分の空白を置いて実際の情報の内容が定義される。情報の終わりは>で示される。

カテゴリーは、テキスト中に同一カテゴリーについての新しい情報定義が現れる迄、直前の情報定義で記述された内容を保持することになる。カテゴリーは、情報の性質、特長等を示す、一文字の変数と考えることができる。カテゴリーとしては半角のアルファベット1文字が使用され、大文字、小文字は別のカテゴリーとして区別される。

実際にCOCOAフォーマットによる情報定義を埋め込んだ例を古典テキストを使って解説する。

```
<title 夕顔>
<Title Yugao>
{Evening Faces}
<C 1><G 1>{源氏、乳母を見舞い、女から扇を贈られる}
<R 1>六条わたり<R>の御<R 2>忍び<R>歩きのころ、内裏より
まかでたまふ<R 3>中宿<R>に、大弐の<R 4>乳母<R>のいたく
```

わづらひて、<R 5>尼<R>になりける<R 6>とぶらはむ<R>とて、五条なる家たづねておはしたり。

ここでは以下の5つのカテゴリが使用されている。

- t 日本語の題名 (title)
- T 英語の題名 (Title)
- C チャプター番号
- G 挿絵番号
- R 注釈部定義

この例では t, T のカテゴリはテキスト全体について同一の情報を保持している。

C, G のカテゴリは 3行目の{Evening Faces}の部分についての情報は持たないが、それ以降のテキストについて同一の情報を保持する。

Rのカテゴリは注釈のある部分についてのみ注釈番号の情報を保持する。

2.3 画像入力

画像は貴重本刊行会の発行による「絵本源氏物語」（財団法人日本古典文学界監修）に使われている画像を利用した。これはサイデンステッカー英語訳本の挿し絵にも使われている。全体で200枚ほどあるが、「夕顔」の巻きでは7枚が使われた。

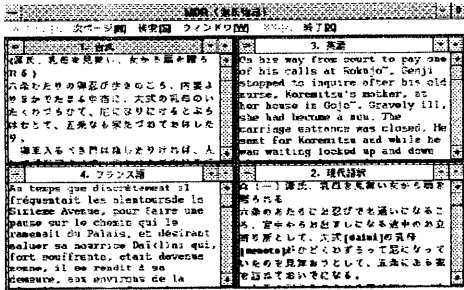
3. 画面操作

3.1 メニュー画面

M D I (Multi Document Interface)はユーザーが Windowsのアプリケーションで複数のドキュメントを同時にオープンして処理するためのインターフェイスである。アプリケーションの要素のひとつであるM D Iフレームウィンドウを使って最初のメニュー画面を作成した。(図1参照)ここで「源氏物語」の複数のテキストをオーバラップさせた形で表示する。ユーザーは4種類の子ウィンドウのタイトルバーをマウスをクリックすることによって必要なテキストを呼び出すことができる。

3.2 複数のテキストの同一画面での比較

ユーザーは個々のテキストのみならず複数のテキストを読み込んで比較研究することができる。ここでは4つのテキスト全てを画面上に読み込んだ場合を紹介する。(図3参照)



(図3) 複数テキストの参照



(図4) 画像の読み込み

3.3 画像や注の呼び出し

画像や注の参照もテキスト上の水色の部分をクリックすることにより可能である。(図4 参照)

3.4 和歌の参照

今回は話者マークを付け加えたのみで、内容分析に深くかかわる複雑な情報は付加価値として挿入していない。しかし共同研究者のオックスフォード大学のP. T. ハリーズの要望で、和歌の情報だけは後からタグとして挿入した。そのため和歌の箇所をクリックすることにより古典語から現代語訳、仏訳、英訳に簡単にアクセスすることができる。

4. 環境設定

今回の開発に当たっては、以下のようなシステム環境設定ファイルを設け、源氏物語に限らず他のテキストにも対応可能なよう考慮した。

環境ファイルはプログラム起動時に読み込まれ、プログラムの動作環境を指定する。テキストファイル名等を書き換えることにより、容易に他のテキストにも対応可能となっている。

```

;*****
;*      源氏物語環境設定ファイル      *
;*      File Name : MDR.ENV            *
;*****
;      ( ; で始まる行はコメント行 )
;
;----- プログラム表題 -----
(源氏物語)
;
;----- テキストファイル数 (Max. 5) -----

```

```

;
;----- COCOAフォーマットの開始・終了記述子 -----
    <      >
;
;----- ページ定義 Category (Max.5) -----
; 各テキストはページ単位 (カテゴリーの持つ情報が変化するまで) で表示され、
; 他のテキストへの移動 (テキスト間の相互参照) は同一のカテゴリー情報を持つ
; ページ位置に移動する。
    Title Chapter
;
;----- セクション定義 Category (Max.8) -----
; 他のテキストへの移動 (テキスト間の相互参照) 時には同一のカテゴリー情報を
; 持つセクション位置に移動する。
    Waka
;
;----- 注釈定義 Category -----
; テキスト内の注釈のある部分、および注釈番号を示す。
; 源氏物語のテキストでは、テキスト内の注釈部は< R  n>で注釈開始位置と
; 注釈番号を示し、< R>で注釈部の終了位置がしめされている。
    Refer
;
;----- テキスト子ウィンドウ表題 -----
; 各テキスト窓の上部に表示される表題をテキスト数分定義する。
1. 古典
2. 現代語訳
3. 英語
4. フランス語
;
;----- テキストファイル名 -----
; 文章テキストが格納されているファイル名をテキスト数分定義する。
c:¥data¥源氏物語¥古典.txt
c:¥data¥源氏物語¥現代語訳.rom
c:¥data¥源氏物語¥英語.txt
c:¥data¥源氏物語¥フランス語.txt
;
;----- 注釈定義ファイル名 -----
; 各テキストに対応する注釈の定義が格納されているファイル名をテキスト数分定義
; する。
; ファイル内の注釈定義は上で指定したカテゴリーを用い、< R  n>のように

```

- ; 注釈番号を付して定義する。
- ; テキスト中に現れる注釈番号は順不同で構わないが、注釈ファイル内の定義は1か
- ; ら順に飛び番なしに定義されていなければならない。
- ; 注釈がないテキストについては空白行を設定してください。

c:\data\源氏物語\古典註.txt

c:\data\源氏物語\英語註.txt

- ;
- ;----- 使用フォント指定 (0:日本語 1:欧米語) -----
- ; 和文は0、欧文（特にアクセント付きの文字を使用する言語）は1を指定し
- ; テキスト数分定義する。
- 0 0 1 1
- ;
- ;----- 挿絵画像定義 Category, 画面数 -----
- ; 挿絵のカテゴリと挿絵の数を定義する。
- G 7
- ;----- 挿絵画像ファイル名 -----
- ; 挿絵データの格納されているファイル名を11で指定した挿絵数分定義する。
- ; テキスト内の挿絵情報は、< G 1 >のように挿絵カテゴリと1から始まる挿絵
- ; 番号とで指定する。

c:\data\源氏物語\graph\t1.bmp

c:\data\源氏物語\graph\t2.bmp

c:\data\源氏物語\graph\t3.bmp

c:\data\源氏物語\graph\t4.bmp

c:\data\源氏物語\graph\t5.bmp

c:\data\源氏物語\graph\t6.bmp

c:\data\源氏物語\graph\t7.bmp

5. 今後の課題

本研究は基礎的な研究であり1995年3月に終了することになっている。しかし今後は「夕顔」のみならず54帖全体へと対象を拡大するつもりである。さらに検索機能を充実させ、インデックスやコンコルダンスを作成し、その結果とテキストの参照を可能にするようなシステムの開発も行いたい。

また開発されたテキストを使った新たな研究にも挑戦したい。現在は日英のテキスト使って「源氏物語」のキーワードである”あはれ”の語彙の自然叙述の利用について比較研究を行っているが、今後は様々な語彙の研究にハイパー・テキストを利用したい。だがそれには現段階では夢のようなことであるが、電子古語辞典や他のテキストとの比較を可能にする大規模テキスト・データベースの開発が前提となる。残念ながら、これには技術や予算の問題もさることながらコピーライトの問題が大きく立ちはだかっている。インフラ整備も含め文科系研究へのコンピュータ導入は多くの課題をかかえている。