

# 1 U-4 イメージをインタフェースとする異本の比較考証を支援するツール — マルチレイヤーのデータモデルとテキスト処理 —

北村 啓子 (国文学研究資料館 研究情報部)

## 1 はじめに

インタフェースとしてのイメージデータ、計算機処理のための内部表現としてのテキストデータとその対応関係から成るマルチメディアデータモデルに取り組んでいる [2][4]。このモデルは、イメージ-テキスト対応関係を保持し、イメージの持つ字形情報 ↔ テキストの持つ文字情報を互いに補完しながら使うことができ、ユーザから見て字形情報と文字情報をあたかも一つのデータのように扱える。このモデルに基づいて異本の比較考証を支援することを和歌を題材に述べる。

## 2 国文学のイメージインタフェース

国文学の対象は草書体の手書きで書かれた古書であり、それに文字認識、翻刻、考証、解釈などを施して活字化したものは研究成果物である。従って計算機上で研究をするには古書のイメージを直接見ながら計算機処理をするのが当然という単純な発想がある。

現状では計算機処理するためには、古書に書かれている文字情報を抜き出して電子コード化を行なうのだが、この際限られた計算機文字内での代替を強いられる。手書き文字 → 活字文字 → 計算機文字において、文字集合は小さくなっていくので、多くの重要な情報が落ちてしまう。

逆に言うと一つの電子コード化されたテキストがあった時にその元の古書上の表記は多くの可能性があり、潜在的に不確定性を多く含んでいると言える。またその古書の表記自身、最初の手書き文字認識の段階で変体仮名文字 / くずし文字 / 草書体など認定の困難さがあり、絶対的な正解はなく曖昧性を含んでいる。ここが出発点であることが重要な点であり、他の活字文字や計算機文字から作られたテキストとの大きな違いである。さらに、翻刻時の研究上の意図的な修正も多いため [1] テキストから古書へ正確にもどるのは不可能に近い。

以上述べたような古書を電子コード化したテキストの性格から、直接イメージデータを見ることが不可欠な要素となる。

Image Interface for study variant text of Japanese Literature

Keiko Kitamura  
National Institute of Japanese Literature

## 3 和歌の同定について

写本を繰り返されてきた古書は現存する本を集めるとオリジナル（著者の書いた元本）は無く、似て非なる本が何十何百と存在する。それらを比較しオリジナルはどう書かれていたのかを推定するのが重要な研究テーマの一つである。

和歌集には特に和歌を同定する上での問題がある。異本による差は散文と同様に、写本時の書き間違い、写し間違い、伝承時の記憶違い、創作的に修正するなどが原因で発生する。歌人名や歌の並び順の入替りによって違う歌人の和歌になったり選ばれた和歌集合が異なるなどは、「XX 勅選集に選ばれたのは誰のどの和歌か」に関わる重大な問題となる。また和歌の独特な類似として本歌取りでは五つの句の中のまとまった二つの句、三つの句が用いられるという大きく違っているものもある。

この和歌を同定するプロセスを計算機で支援する方法として、類似している和歌を捜し、同じ歌である候補の集合として提示することを考える。「類似」をどう定義するかについて述べる。類似の指針として次の3次元の要素を考える。

### I. 類似の種類（類似の度合を示すものではない）

- ・ 平均的に似ている  
各句で異なる文字数が1文字以下  
全体で異なる文字数がn文字以下
- ・ 部分的に異なる  
初句のみ異なる、一つの句/二つの句だけ異なる
- ・ 順番が異なる  
n句とm句が入替っている
- ・ 特徴が似ている  
上句/下句が体言止めで終る

### II. サーチする対象

- ・ 底本に忠実な表記
- ・ 翻刻した表記
- ・ 読み/異なる読み
- ・ 異表記

下の2つは一般的に使われる検索ヒット率を上げるための手段であるが、古文の世界で網羅度を上げるのは非常に困難である。

### III. 類似の範囲（あいまいさ）を調節

- 完全一致、一文字違い、二文字違い、...
- （置替り、抜け、挿入）

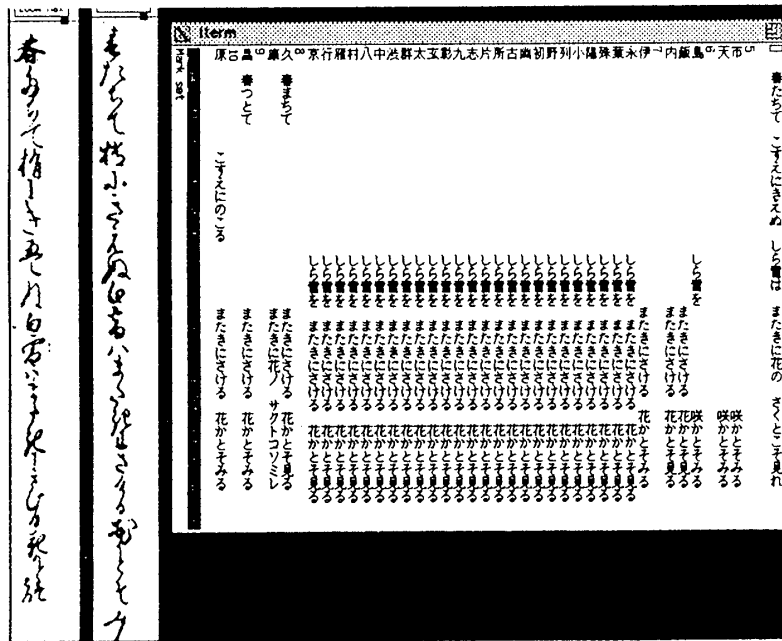


図 1: 和歌に異本比較考証を支援する例

この3次元の各要素を決めると一意に集合が決まり、その要素の組合せにより様々な類似集合を作ることができる。研究者の目的と興味に合わせて和集合を作ったり違った観点での集合を見比べることも有意義である。

#### 4 異本の比較考証を支援するツール

個々の異本をマルチレイヤーモデルで表現する。イメージレイヤーと複数のテキストレイヤー（忠実な表記、読み、異なる読み、異表記）ならびに、イメージテキスト間の対応関係からなる。テキスト処理から得られた比較結果（類似集合と類似の度合）をナビゲーターとして、その結果をイメージで提示しながら比較考証を助ける。

ここで使用するツールは次の4種類である。

- a. テキスト処理ツール類：jperl で開発
  - ・マッチングテーブルの作成  
(句単位にマッチした文字数のテーブル)
  - ・類似の種類別リスティング
  - ・類似度別リスティング
  - ・校訂本文形式フォーマッタ
- b. テキスト縦表示ツール：tterm[5]
- c. イメージ表示ツール：viewer[2]
- d. イメージテキストリンク：
  - nemacs(GNU Lisp)-viewer(C) プロセス間通信 [2]

これらを使って類似集合を表示している例を図1に示す。最初の行にパターンマッチングの基になる和歌を表示し、平均的な似ているものを異りの文字数（数字も表示）の小さい順に並べている。翻刻した表記だけを対象にすべての異りを類似の範囲とした。表示は異りの文字

を含む句とした。

この例では、今回開発した a. を既存の b.c.d. と連動しているが、ユーザビリティの向上と拡張性を考慮し、b.c.d. は tel を使った書替えに取り組んでいる。

#### 5 おわりに

古書を表現する文字集合の変遷という観点でイメージインタフェースの必要性を述べた。和歌の同定を支援するために、3次元の要素による類似集合の考え方を述べた。この類似集合を作り異本の比較考証を支援するツールを試作し報告した。テキスト縦表示、イメージ表示、イメージテキストリンク機能を統合し、ユーザビリティを向上させて国文学研究者の評価を仰ぎたい。

#### 参考文献

- [1] 北村啓子: 古文書を分析するためのHCI, 情処HI研究会, 94-HI-53, pp.63-70,(1994)
- [2] 北村啓子: 古文書を表現するためのマルチメディアデータモデル, Advanced Database Syatem Symposium'93,pp.235-244,(1993)
- [3] 北村啓子: 古い著作物を分析するための計算機環境 - 和歌文学を題材に - 情処全大47回,2C-1,vol.4,pp.95-96,(1993)
- [4] 北村啓子: 古文書を表現するためのマルチメディアデータモデルの構想, 情処全大45回,1R-1,vol.4,pp.99-100,(1992)
- [5] 北村啓子: 縦書きテキスト編集機能の検討とX Window上での試作, 情処全大43回, 7L-1,vol.4,pp.81-82,(1991)