

計算機を用いて『百人一首』を「読む」

—変体仮名とプログラミングを覚えよう—

福田 智子・深川 大路

1. はじめに

2015年6月、文部科学相が、全国の国立大学法人に対し、人文社会科学系の学部・大学院の廃止や転換に取り組むことを求める通知を出したことは、あまりにも記憶に新しい。その後の調査によれば、人文社会科学系学部をもつ国立大学法人の約8割が、再編を検討しているという(7月19日NHK調査)。

この国立大学法人の回答は、とくに意外性をもつものではないであろう。それだけ、現今の人文社会科学、とりわけ人文科学という学問分野に対する社会からの風当たりは強い。

ここで、世界に思いを馳せてみよう。たとえばヨーロッパでは、その国の古典を重んじる文化が根付いていると聞く。実際、第一著者が国際会議に出席した時、ただ日本の古典を研究しているというだけで、日本では経験したことのない敬意を向けられたことがある。

ところが、今日の日本では、古典文学研究は風前の灯火、時間や労力をかけて継承していく価値のない学問に成り下がってしまったようである。

このような状況の背景には、長らく景気の低迷を経験してきた現在の日本の余裕のない現実がある。だがそれ以上に、今日の日本の古典文学研究者が、高度に専門化され、個人的で蜻蛉化した研究のあり方を肯定し、存続させ続けてきた上、後進の教育に手薄であったことにあるのではないか¹。

同志社大学文化情報学部は、京都府南部、京田辺の地に位置する。今出川校地を文系・基礎学の学部が存するキャンパスとするならば、京田辺校地は、理系・応用学(文理融合)の学部の集結と

見做されよう。それならば、前述の国立大学法人を対象とする、しかも、人文系の学部に対する文部科学相からの通知は、本学部とは直接的には関係がないようにも思える。

しかしながら、文理融合だからこそ、また、応用学だからこそ、研究・教育において、基礎学の重要性を痛感することが多い。とくに教育に関しては、基礎力がなければ、学生は、自分が学んでいることが応用であることすら理解しがたい。一方、基礎力を身に付けることができた学生の応用力は、自ずから醸成されていく。

以上のような国・大学・学部の現況を鑑み、本稿では、本学部3年生の演習授業「ジョイント・リサーチ」の中から、「教養としての『遊び』—カードゲームと歌謡の歴史—」を採り上げ、2015年度春学期の実践例を紹介する。日本の古典文学を対象に、情報科学の技術を学ばせることにより、古典への理解を深め、プログラミングの技術を身に付けるとともに、論理的思考力を養成することを目的とする。

2. 授業の動機付け

本授業は本学部における3年生の必修授業で、年度が始まる前に、学生に向けてクラス分けプレゼンテーションが行われる。その際に、授業の題目からもわかるように、まず、「カードゲーム」

¹このことは、第一著者(福田)が編集委員として参加した全国大学国語国文学会編『文学・語学』第213号「〈中古文学小特集〉時空を超える平安文学」(2015年8月、113頁)でも指摘したところである。

を扱うと説明した。そして、初回の授業では、子どもの頃にどんな遊びをしたか、どんなカードゲームで遊んだか、そこで何を学んだか、カードゲームを楽しむためにはどんな能力が必要か、といったことを、ともに考えた。

受講生の世代のみならず、現代の子どもたちにとっても、カードゲームは魅力的な遊びのひとつである。そこで時代を遡り、江戸時代にもカードゲームとしてかるたを楽しんでいたこと、そして、そのかるた遊びが、識字率を上げ、和歌の教養を身に付けるのに役立っていたことなどを説明した。その上で、受講生ひとりひとりが実物の江戸期のかるたを手にとって見る機会を与えた。

鎌倉期の古写本ならばともかく、江戸後期のかるたであれば、よほど貴重な品を除けば、直接触れることは、保存の上でもそれほど問題はない。それでも、昨今の大学生は、江戸期の文物を身近で意識的に観察したことがない者が多い。実物を見た受講生の反応は大きく、今まで歴史の教科書でしか想像したことのない江戸時代の人と文化が、急に身近な存在として認識されたようである。

なお、本授業では、子どもたちに文字を教えるための教材（ゲーム）を作ることを最終目的として設定した。だが、それは単にゲーム開発を行おうとしているのではない。子どもに何か教えたい、ゲームを作りたい、という意志のもとに授業に参加する過程で、受講生自身が身に付けるものは、日本語の表記文字についての理解と、変体仮名・くずし字を読む力、そして、それらの文字列や画像を有効に分析するプログラムを組む力である。その上で、論理的な思考でデータを分析する力を養っていく。

3. 文化情報学部蔵百人一首かるた

本授業では、かるたの中でも最もオーソドックスな百人一首かるたを教材とした。

幸いなことに、文化情報学部文献室には、江戸期のかるたが複数所蔵されている。そのうち本授業で用いた百人一首かるたは、江戸後期に木版印刷で出回ったものである。通常、百人一首かるたは、札に歌仙絵が描かれることが多いが、文化情報学部所蔵のかるたは、上句札の歌仙絵だけではなく、下句札に歌意絵（和歌の意味内容を描いた絵）が描かれており（〔画像 1〕参照）、比較的珍しいものようである。その歌意絵によって、江戸時代の人々が、百人一首の歌の内容をどのよう



〔画像 1〕 同志社大学文化情報学部蔵百人一首かるた

に理解していたのか、ひいては、文学作品の絵画化、イメージの意匠化がいかになされていったのか、といった点でも研究テーマとなり得る。

4. かるたのデータ化

本授業で用いた百人一首かるたは、上句に作者名と歌仙絵が記される札と、下句に歌意絵が描かれる札が 100 対、200 枚ある。受講生は 20 名。そこで、4 名ずつ 5 班に分かれ、百人一首の歌をひとり 5 首ずつ担当した。そして、かるたに記された文字と挿絵をリンクしたデータベース作成を試みた。その手順について、以下に述べていこう。

4-1. 文字のデータ化

かるたに記されている文字は、いわゆる変体仮名・くずし字で記されている。一般に現代人には馴染みの薄い文字である。けれども、今の大学生の三代前の曾祖父・曾祖母の時代には、当たり前のように手紙に変体仮名を用いていた時代があった。ところが、次世代以降の文化の断絶には、目を覆うものがある。

変体仮名は、字母を覚えたり、字母の崩し方のレベルに対応したり、連綿の文字に慣れたりする必要があるが、受講生には、簡単な字体表を配付し、札に記される文字を、現代の通行の文字に直したテキストの作成を求めた。と同時に、仮名字母のテキスト、すべて歴史的仮名遣いで表記した仮名開きテキストも作成した。

仮名字母のデータは、一作品内で、字母がどのように分布しているか、といった国語学の視点からの考察も期待できる。

4-2. 挿絵のデータ化

挿絵をデータ化するにあたり、まず、受講生に対し、名詞や動詞ひとつで指示できる要素を挿絵から見出し、なるべく多く指摘するように求めた。いわゆる「絵引き」の作成である。

「男」「女」「立つ」「座る」「山」「川」「農夫」「眺める」「屋根」などの要素が寄せられた。たとえば〔画像1〕では、「男」「女」と「彦星」「織女」、両方の要素を包含しながら、なるべく多くのデータを収集することにした。

もっとも、班ごと、あるいは個人ごとに、着眼点が違えば、要素を示す語の選択も異なる。データベースを作成するには、均質なデータを収集することも重要だが、今回は、とにかく思いつくりの要素をなるべく多く提出させることにした。そしてその中から、比較的基礎的で穏当な要素を選び取って、データベースに格納した。

とはいえ、文化事象のデータ化には、試行錯誤が付きものである。挿絵データのみならず、文字データに関しても、データの修正・入れ換えが柔軟に行えるシステム作りが求められる。

5. データベースを用いた発見・創造

百人一首かるたの文字データと、挿絵データが出来上がったところで、本授業では、Rを用いた分析と、第2著者（深川）作成の画像資料注釈管理ツール TIRAMiS による考察を行った。

5-1. Rによる文字列処理

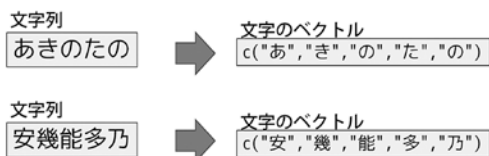
本授業では、かるたに記されている文字を、現行の漢字仮名交じりに翻字したテキストとともに、字母テキスト（ただし漢字も交じっている）を作成している。

```
> hyaku[1:5,]$漢字仮名交じり本文
[1] "あきのたのかりほのいほのときをあらみわころもてはつゆにぬれつゝ"
[2] "春すきて夏きにけらし白たえの衣ほすてふあまのかく山"
[3] "あしひきの山とりのおのしたりおのなかくしよをひとりかもねん"
[4] "たこのうらにうちいてゝ見ればしるたえのふしのたかねにゆきはふりつゝ"
[5] "おく山にもみちふみわけなくしかのこゑきとときそあきはかなしき"
> hyaku[1:5,]$字母本文
[1] "安幾能多乃利本乃以本乃止万遍阿良美和可己呂毛天八川由尔奴連川"
[2] "春春起天夏幾仁連良之白多衣能衣本春天婦安万能可久山"
[3] "安之比幾乃山止利能於乃志多利於乃奈加ノ之与遠比止利加毛祿无"
[4] "太己能字良仁字知似天ノ見連八志呂多衣乃不之乃堂可年耳由幾八不利川ノ"
[5] "於久山尔毛美知不三王遍奈久之可乃己惠幾久止支楚阿幾八可祭之幾"
```

そこで、まず、漢字仮名交じりテキストの仮名を字母テキストと対応させたデータを、次のような手順で作成した。以下、授業中に使用したスライドの一部を引用しながら説明する。

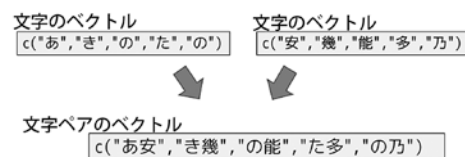
■ 方針 (Step1)

■ 文字列をベクトル化する



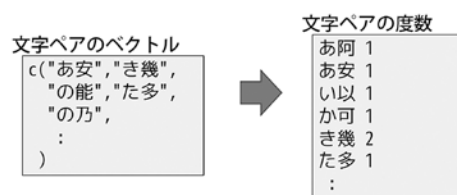
■ 方針 (Step2)

■ 作成した2つのベクトルから文字ペアのベクトルを作る



■ 方針 (Step3)

■ 相異なる文字ペアの出現度数を集計する



上記の Step1 ~ 3 は、仮名 5 文字に対して字母 5 文字が対応する場合であり、処理は最も単純である。

しかし、データを眺めていると、次のような例があることにすぐに気づく。

■ 同じ語に対する異なる表記の例

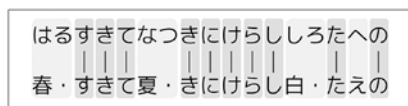
はるすきて	春春起天
はるののにいてて	者留能ノ尔出天
はるのひに	者留能日耳
はるのよの	者留能与乃

このかるたでは、「春」を仮名で表記する際に「者」「留」という字母を用いる傾向があることが見て取れるが、それはともかく、文字列処理の側から言えば、前述の Step1 ~ 3 のようなベクトル化では、仮名に漢字が交じることによって、字母との対応にずれが生じることになるのが問題である。

そこで、演習は次の段階に移る。

■ 解決策

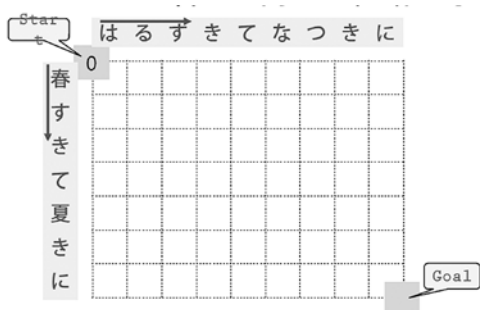
- 編集操作（文字の追加や削除）を考慮した文字列のアラインメントを作成する
- アラインメントを元にペアを作成する



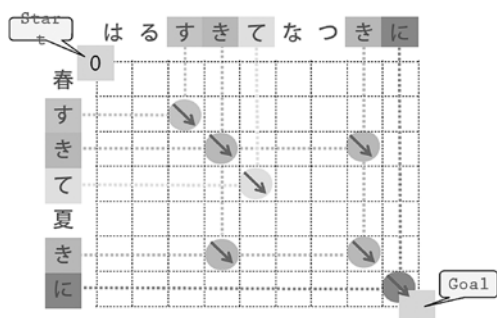
ここで作成する「文字列のアラインメント」は、2本の文字列に対して行うアラインメントで、ペアワイズ・アラインメント (pairwise alignment) という。文字列の長さを m、n とすると、計算時間は $O(mn)$ である。

その計算は、動的計画法 (dynamic programming; DP) と呼ばれ、表を適切な順に埋めていくことで、最適なコストが計算でき、さらに、表を溯って辿ると、最適な経路 (アラインメント) が発見できる。

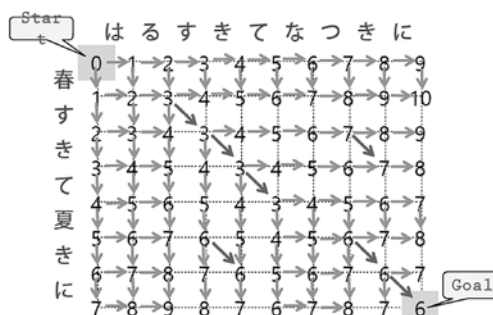
ペアワイズアラインメントの計算は、まず、表の左上から右下に向かう経路を考える。



縦・横の文字が同じである場合 (match) は、斜め矢印に進むことができる。



こうして、可能な経路とそのコストを計算していく。



上記の図の中で、無駄な経路を削除すると、次のようになる。



これで、漢字仮名交じりテキストと漢字を含む字母テキストとの文字対応ができたことになる。

その集計結果をもとに、たとえば次のような表にしてみる。

	A	B	C	
1	仮名	漢字仮名	字母	度数
2	あ	あ	阿	16
3	あ	あ	安	58
4	あぎ	秋	秋	1
5	あぎかせ	秋風	秋風	2
6	い	い	以	47

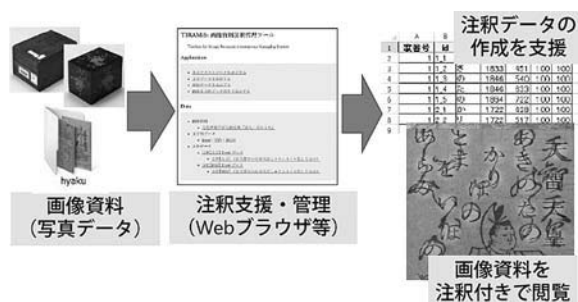
これにより、仮名字母や漢字の出現頻度を、具体的に把握することができる。

5-2. 画像資料注釈管理ツール TIRAMiS

以上の作業を経て、かるたの文字については一定の把握がなされた段階で、次に、テキストと画像を繋いだデータベース作成に入る。

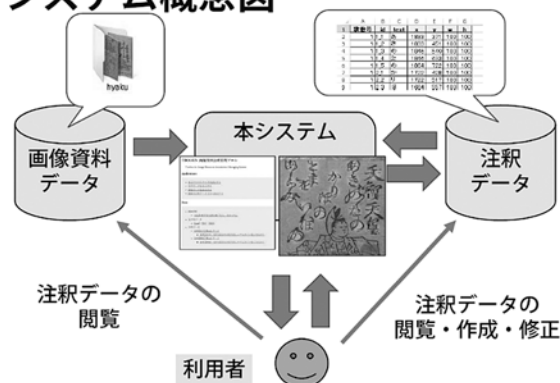
今回、第二著者 (深川) が開発した TIRAMiS “for Image Resource Annotation ManagIng System (ティラミス)” は、画像資料に対する注釈を管理

するためのツールである。



すなわち、画像資料について作成された様々な注釈内容（本授業ではかるたの文字データと挿絵に対する絵引き）を一括管理し、注釈データを画像資料とリンクさせることで、注釈データの作成を支援し、かつ、注釈付き画像を閲覧できる。

システム概念図



このツールを用いて、先に作成した注釈データを画像にリンクさせ、注釈内容の加筆・修正作業を行った。



[画像 2] TIRAMiS による加工画像

作成した文字データを画像とともに閲覧することで、データの誤りに気づきやすく、それをシステム上で修正することも、きわめて容易である。また、方形で囲った1文字1文字は、切り出しが可能である。テキストデータを文字検索し、当該箇所の画像とともに用例を収集することもできる。

この作業を通して、受講生は、かるたの文字と挿絵というコンテンツに対する理解を深めるとともに、いかにゲームとして生かしていくか、いかに楽しく文字や百人一首を子どもたちに覚えてもらうか、といった構想を練った。その内容は、グループ発表というかたちで披露された。

6. レポート指導

こうして進めてきた授業内容について、まとめの学期末レポートを課すが、本授業では、提出締切や分量の他、レポートの構成を詳細に指示した。すなわち、文章構成の「型」を与えるのである。具体的には以下の通り。

内容：次に挙げる構成に即して、授業で扱った『百人一首かるた』の文字や挿絵の知識、Rを用いた分析の方法やソフト開発の提案をまとめて文章化する。

はじめに

- ・第一～三章の予告を短く書く。

第一章 文化情報学部蔵百人一首かるたについて

- ・配付プリントをもとに、特別な知識をもたない高校生にでもわかるように説明する。

- ・札の枚数、挿絵のある位置、散らし書きの様子など、基本的な情報を記す。

- ・100首全体に目配りしてほしいが、担当した5首について、詳しく説明してもよい。

第二章 文字と挿絵に関する考察

第一節 文字について

- ・仮名文字に関する基礎知識をまとめる。

- ・当該かるたの文字について具体的に考察する。

〔例〕仮名「か」は、平仮名の字母「加」の他に、変体仮名の「可」も用いられる。文字の画像と歌番号を列挙して分布を考察する。

〔例〕当該かるた全体で用いられた仮名字母の数や、出現頻度を考察する。

- ・分析の方法（Rなど）についての説明がほしい。

第二節 挿絵について

- ・当該かるたの挿絵について、構図の種類や特徴など

を考察する。

〔例〕 類型的な構図はあるか、全体像を把握する。

〔例〕 担当した5首の歌の挿絵を、『百人一首像讃抄』の挿絵と比較する。

第三章 教育用ソフト開発の提案

第一節 ソフト開発の目的

- ・子どもたちにどのような力を身に付けてほしいのか、明瞭に記す。
- ・第二章で考察した文字と画像に関する知識を、どのように生かそうとしているのかを必ず書く。

第二節 ソフトの設計

- ・仕様と使い方の説明を丁寧に書く。
- ・札の画像や図などを適宜用いて説明する。

第三節 ソフトの特徴と今後の課題

- ・ソフトの特徴（有効な点）を客観的に書く。
- ・今後、さらに補うべき点、向上させるべき点を具体的に指摘する。

おわりに

- ・なにをしたら何がわかった（実現できそう）か、（第一～三章）を短く（抽象的に）繰り返す。
- ・文字と挿絵、ソフトを関連付けながら、今後の課題をまとめる。

参考文献

- ・百人一首やかるたの他、ソフト開発関連の文献も、参照したものを列挙する。
- ・書誌情報は不足なく挙げる。

以上のように、レポートの構成を具体的に示すことによって、受講生の多くが授業内容を自ら整理し、把握し直す機会になる。

子どもに文字を教えたい、ゲームを作ってみたいという動機付けによって始めた本授業から、変体仮名の字母の分布状況や和歌の絵画化パターンの考察、プログラミング技術を用いたデータ分析の可能性の模索など、文学や情報科学の研究の入り口に、自然と立つことができたのではないだろうか。

7. おわりに

国文学研究資料館を中心として行われている、人文社会科学系初の大規模学術フロンティア促進事業「日本語の歴史的典籍の国際共同ネットワーク

構築計画」(2014～2023年度)の第1回研究集会が、2015年7月31日に開催された。有川節夫氏による基調講演「古典籍共同研究とオープンサイエンス」では、貴重な実験・観測データ等をオープンにし、内外の研究者が相互に利用できるようにする、いわゆる「オープンサイエンス」の重要性が説かれ、さらに、研究者が学術情報基盤を共有するだけではなく、一般市民がデータにアクセスできるようになる「シチズンサイエンス」の構想にまで言及された²。科学の分野では当たり前になりつつあるこういった情報共有のあり方が、文化系の研究分野でも求められ始めている。

本授業の根幹にも、実は、「伝統文化形成に関する総合データベースの構築と平安朝文学の伝承と受容に関する研究」(同志社大学人文科学研究科第18期研究会第17研究、および科学研究費助成事業基盤研究(C)課題番号25330403、いずれも平成25～27年度)の共同研究がある。第二著者(深川)が開発した、画像資料注釈管理ツールTIRAMiSは、その研究成果のひとつであり、文化系コンテンツの「オープンサイエンス」「シチズンサイエンス」の今後に資するところは小さくないと考える。

学期末レポートの中に、「先生が書いたプログラムを利用するだけではいけない。」「もし会社だったら、プログラムを書いてくれる人に頼れないだろう。」といった反省が書かれているものがあつた。確かに今回の授業では、実際にゲームを作る段階には至らず、ゲームの提案に終わってしまったが、このような意識を持たせることこそが、新しい時代の教養を育むのではなかろうか。

²「日本語の歴史的典籍の画像データ化オープンサイエンス」(「日本語の歴史的典籍の国際共同ネットワーク構築計画」ニューズレター『ふみ』第4号、2015年6月)参照。