

Digital Humanities の実践例

—国文学研究の裾野を広げるために—

福田 智子（同志社大学・准教授）

注：職名は報告当時。現職は教授。

1、はじめに

同志社大学・文化情報学部の福田智子でございます。お声掛けいただいた当初から考えておりましたことをここでご報告させていただきたいと思っております。

「初中等学校における古典教育」研究会ということですが、これからお話いたします内容は、私が同志社大学の文化情報学部に勤め始めましてから実際に大学生、大学院生に対して指導してまいりました内容です。初中等学校において実践するとなると多少難しい面があるかもしれませんが、指導の仕方によってはこれらの生徒の皆さんにも取り組んでもらえる教材になり得るかもしれません。いずれにせよ、この試みは今回のお話の副題にも付けました国文学研究の裾野を広げるための実験とも言えるものではないかと考えております。まだまだ問題点は多くございます。ご意見・ご感想を頂戴できれば幸いです。

先ほど申しましたように、私は同志社大学の文化情報学部という学部には所属していません。大学の学部の名前は短いほうが良い。たとえば、文学部、理学部、医学部というように考えておりましたので、文化情報とはなんだろう、何か得体の知れない学問・研究をやっているような気持ちがしたものでした。今から11年前、2005年のことです。それをたとえば、学部ホームページでは現在、「興味を学びに。好奇心を学びに。データサイエンスで文化を読み解く」と紹介しています。何をするかは具体的に分からないけれども、何か良さそうな学部の雰囲気は伝わってきそうです。うまく書いたものです。この学部コピーのキーワードは既にお気付きのとおりデータサイエンスです。データを科学的に分析する。文化情報学部は文化を科学するデータ分析の学部です。これを文学研究に当てはめるならば、文学作品を対象に具体的な用例を集めて分析する実証的な研究ということになるでしょう。

このような学部にくる学生は、文学部、理学部、医学部に入学してくる学生に比べ、大学に入学して何を学びたいのか目的意識を明確に持っていない傾向があります。良く言えばいろいろな分野の物事を広く学ぶ姿勢があるとも言えそうですが、こと国文学、日本古典文学について言うと、大変分が悪いのが現状です。文化情報学部では、高等学校までの古文の授業で、日本古典文学に対する苦手意識を植え付けられている学生が少なくありません。そもそも古文が好きな学生は文学部国文学科入学を目指します。文化情報学部には文学少年・文学少女はまずいないのです。これに加えて文化情報学部は、設置当初から文理融合を合言葉にしていました。この言葉もどこかうさんくさい感じがしなくもないわけ

ですが、ここで申し上げたいことは、文化情報学部には高等学校で文系理系に分かれて受験勉強をしてきた学生たちが、入学以後は文理に関わらず同じカリキュラムを受講する、——もちろん選択科目はあるのですけれども——基本的には文系の学生と理系の学生が机を並べることになるわけです。先ほど文化情報学部には文学少年・文学少女はまずいないと申しましたが、現実はそのどころか理系の学生まで交じっているというのが、この学部の特徴です。融合というからには、学生がもともと文系であっても理系であっても、その枠にとらわれることなく学んでほしいところです。教員の側も情報科学をはじめとする理系の教員と、いわゆる文系の教員が一つの学部に所属して、授業によっては文系の教員と理系の教員が組んで一つの授業を担当するという場合もあります。

ですから、たとえば演習系の授業で文学作品を取り上げるときは、できれば具体的な用例収集の段階で、ある程度専門的な情報科学の技術を習得し、分析手法として利用しながら実証的な文学研究に結び付けてほしいと考えています。ですが、それも理想的な授業運びにするのは難しいところです。学生に文学と情報科学、両方学びなさいと言っているわけですから、計算機プログラミングは得意だけれども、古文はもとより本なんかまともに読んだことがないという理系の学生には、拷問のような授業になりかねません。そこで、文学部国文学科以外の学生に向けてどのような古典文学の授業を行ってきたか、文化情報学部、および文化情報学研究科において約 10 年にわたって行ってきた授業の中から、その特徴を示す実践例をご紹介します。

2、デジタル・ヒューマニティーズ (Digital Humanities)

その前に本日のお話の題目にも挙げておりますデジタル・ヒューマニティーズについて少し触れておきましょう。次のスライドは、東京大学大学院横断型教育プログラム、デジタル・ヒューマニティーズのトップページの一部です。そこにはデジタル・ヒューマニティーズとは、ということで次のように説明されています。

「デジタル技術は、人類の知的資源の保存、研究、発信の方法を大きく変えて、情報社会の新しい知識基盤を形成しています。この変化に対応すべく、デジタル媒体による学術資料のアーカイブ構築、文化コンテンツの分析、学術成果の公開や展示の方法などを、文系・理系の枠組みを横断して研究するデジタル・ヒューマニティーズの動きが世界的に広がっています。本プログラムは、デジタル技術を用いたヒューマニティーズの新しい方法とともに、情報学を文化資源の構築に活かしていく方法を広く学びます。」 (<http://dh.iii.u-tokyo.ac.jp/> 閲覧日 2016/8/30)

ここで述べられているデジタル媒体による学術資料のアーカイブ構築、文化コンテンツの分析、それから文系・理系の枠組み横断といった点からも、今までご紹介してまいりました文化情報学部の学問教育の一端がまさにこれであることがお分かりいただけると思います。

3、文化情報学部および文化情報学研究科における授業の実践例

それでは、具体的な授業実践例に話を移しましょう。今回は四つの実践例を準備してまいりました。お時間の許す限り、順にご紹介していきたいと思っております。いずれの実践例も文学作品に対する姿勢にある共通点を持っています。それは何かというと、文学作品を単なる文字列データとして捉える、言い換えれば、授業の初めには学生に意味を考えさせない、目の前の文字列が文学作品であることはいったん置いて、まるで遺伝子の塩基配列のように文字列を見せるということです。これはある意味では大変乱暴な見方、文学作品に対する冒涇とも受け取られるかもしれません。しかし、これまで述べてまいりましたように、これらの授業は全て文化情報学部という、まるで文学に興味がない、さらには古文なんて大嫌いという学生が多くを占める学部での授業です。思い切って作品を文字列と言い切ることで、たとえば理系の計算機プログラミングに慣れた学生ならば、まず文学作品の文字列の統計処理を難なくこなしてしまうこともあるでしょう。そうすると、その後が肝心なのですが、やはり、作品そのものに何が書いてあるのか意味が知りたくなる。そうなったときにあらためて知識を与えることで、その学生が、国文学研究者になることはほばないだろうけれども、日本古典文学をある程度知っている、あるいは、日本古典文学の奥深さ、魅力を味わうことの意義を理解することになるのではないかと考えました。

こうして考案した授業の概略を示すのが、このスライドです。

2, Digital Humanitiesの実践例—目次—

実践例（1）和歌の文字列統計

授業名：文化情報学演習 1

対象：大学1年生（必修）

担当：教員12名が12クラスを分担

*教員の専門分野によりコンテンツは異なる

期間：1クラス7回

実践例（2）和歌の本文異同

授業名：日本古典文学情報特論

対象：大学院博士前期課程（選択）

担当：教員1名（文系）

期間：半期

実践例（3）物語の本文異同

授業名：ジョイント・リサーチ

対象：大学3年生（必修）

担当：教員2名（文系・理系）

期間：半期

実践例（4）字母データベース構築

授業名：ジョイント・リサーチ

対象：大学3年生（必修）

担当：教員2名（文系・理系）

期間：半期

文学作品を単なる文字列データとして捉える
(最初は意味を考えない=まずカタチから入る)

実践例(1)は、大学1年生の必修授業で、12名の教員がクラス別に担当しておりまして、演習の内容は担当教員に委ねられています。半期をさらに半分に分け、1クラスの授業回

数は7回です。私のクラスでは、和歌の文字列統計を取るため九州大学大学院システム情報科学研究院の竹田正幸氏との共同研究で開発した文字列解析器を使っています。和歌の表現に関する知見を入れるのが目標です。また、実践例(2)は大学院の選択授業なのですが、和歌の本文異同を把握するため特定の歌集の複数の原本を和歌用のデジタル校本作成ツールを用いて分析するというものです。このツールも先ほどの竹田氏との共同研究で独自に開発したものです。そして、実践例(3)(4)では、文系の教員と理系の教員が一つの教室に出向いて行う大学3年生対象の演習授業です。これが文化情報学部の特色ある授業の代表格とも言えるもので、実践例(3)の物語の本文異同のほうは文化情報学部1期生の学生が作成し、残っていてくれた校本作成支援ツールに既存のソフトを組み合わせて使用しています。また、実践例4。これは現時点で進行している授業なのですが、影印などの画像に翻字や字母などの注釈を付けるシステムを独自に開発し、授業中に学生とともにデータベースの作成を始めました。

では、これらの実践例を一つずつ見ていきたいと思います。

4、和歌の文字列統計

まず1例目は和歌の文字列統計です(スライド「実践例(1)」)。

2、Digital Humanitiesの実践例(1) 和歌の文字列統計

竹田正幸氏(九州大学大学院システム情報科学研究院)作成
文字列解析ツールe-CSA “efficient character string analyzer, イークサ”

- ◎ データ(作品)を単なる文字の連鎖として捉える。
=形態素解析(品詞分解)をせずに、
- (1) あるテキストデータから、部分文字列を抽出し、
どのような文字列が何回出現するか(出現頻度)を提示する。
- (2) 複数のテキスト間で共通文字列の出現頻度を比較する。

南里一郎・竹田正幸・福田智子
文字列解析ツールe-CSA ver. 1.00—国語学・国文学研究者用マニュアル
<http://catalog.lib.kyushu-u.ac.jp/handle/2324/8971/pa001.pdf>
(『文献探究』第42号、2004年3月)

2016.9.5

Doshisha University

五一七一五一七一七の三十一文字の仮名文字テキストをもとに竹田正幸氏が作成した文字列解析ツール e-CSA を使って、文字列の統計を取るという授業です。e-CSA はスライドに挙げておりますように、文学作品でも何でも、とにかく文字列になっているデータであ

れば、どんなデータでも解析することができます。すなわち、文字列を意味は無視して言葉、日本語を扱う場合は形態素解析、いわゆる品詞分解をしないで単なる文字の連鎖として捉え、あるテキストデータからありとあらゆる部分文字列を抽出し、どのような文字列が何回出現するか、その出現頻度を提示したり、複数のテキストの間で共通する文字列の出現頻度を比較したりすることができます。授業ではこのツールと次にご覧に入れます和歌のデータとを学生に与え、自由に文字列統計を取ってみるという演習を行っています。では、画面を和歌データに切り替えます。前のスクリーンをご覧ください。

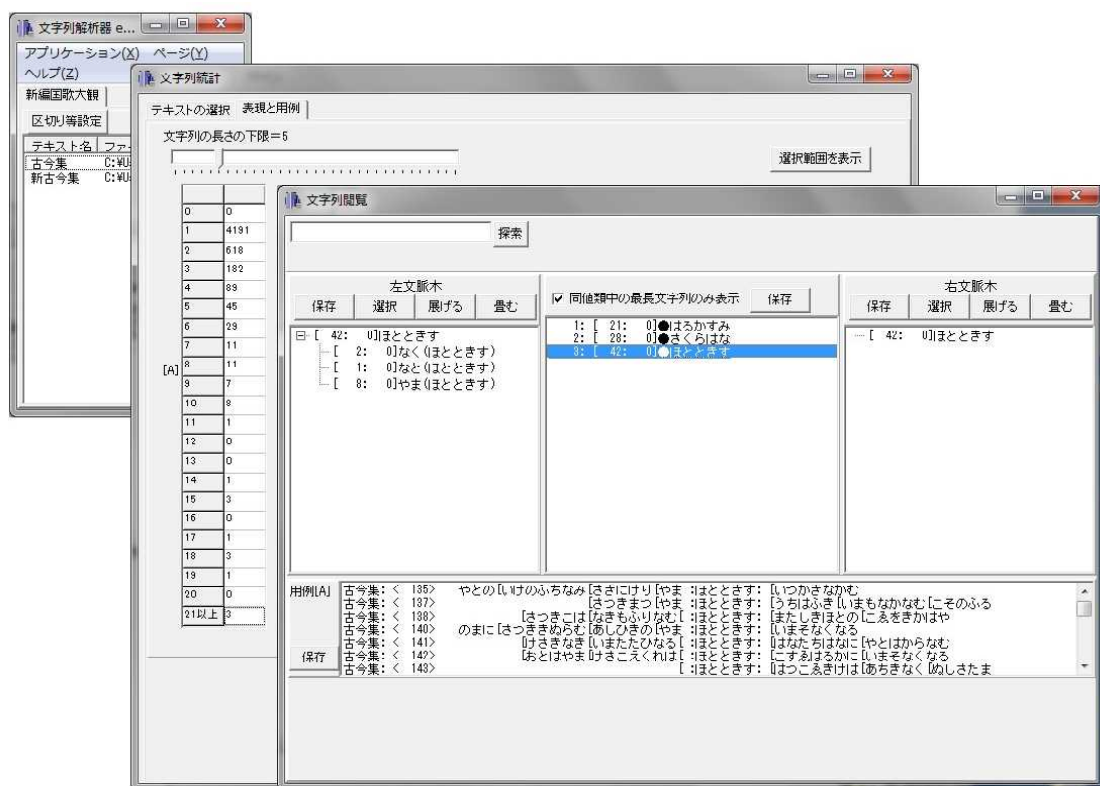
```
<D>
<R>
<A>一卷 1古今 </A>
<N> 1</N>
<P>[としのうちに][春はきにけり][ひととせを][こぞとやいはむ][ことしとやいはむ]</P>
<K>[としのうちに][はるはきにけり][ひととせを][こそとやいはむ][ことしとやいはむ]</K>
</R>
<R>
<A>一卷 1古今 </A>
<N> 2</N>
<P>[袖ひちて][むすびし水の][こほれるを][春立つけふの][風やとくらむ]</P>
<K>[そてひちて][むすひしみつの][こほれるを][はるたつけふの][かせやとくらむ]</K>
</R>
<R>
<A>一卷 1古今 </A>
<N> 3</N>
<P>[春霞][たてるやいつこ][みよしのの][よしのの山に][雪はふりつつ]</P>
<K>[はるかすみ][たてるやいつこ][みよしのの][よしののやまに][ゆきはふりつつ]</K>
</R>
.....
</D>
```

9

これが古今集のテキストデータの構成内容です。和歌の本文だけを取り上げて実習に使っています。今、画面に映し出しておりますのが使っているデータで古今集の和歌データなのですけれども、ご覧になっていただくとすぐ分かりますように、和歌1首分のデータが<K>のタグでくくられています。<K>のタグでくくられているのが仮名開き本文で、他にも<P>のタグというのがあるのですが、これは漢字仮名交じり本文です。ですから、分析の目的に応じて e-CSA で分析する対象文字列を変更することができるのですけれども、今回は<K>のタグ、全部仮名に開いたものを使います。本来だったらこのテキストをつくること自体が学生にはかなり勉強になるのではないかと思います。漢字仮名交じりのものを仮名に直す、歴史的仮名遣いに直すということなので勉強になるかと思うのですけれども、この授業の受講生が1回生ということもあり、今回はこのテキストを学生に与えるところから始めています。

さて、このようなデータを用意して実際に文字列統計を取ってみます。ちょっと e-CSA というツールを立ち上げてみます。これです。

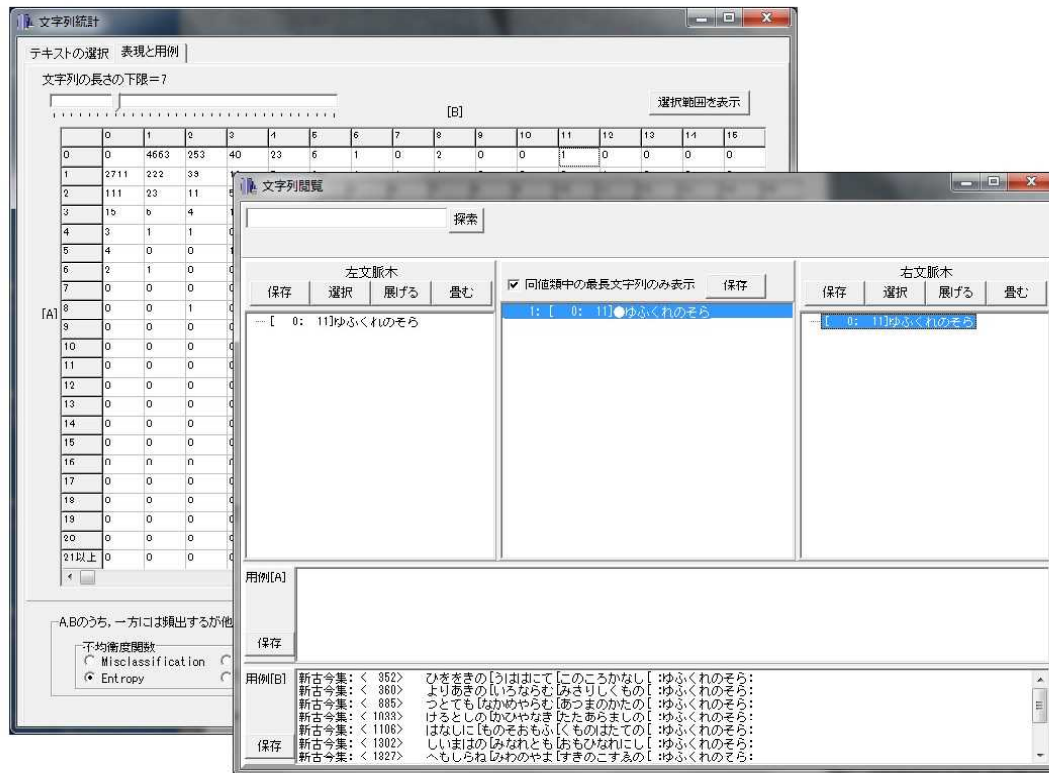
たとえばですが、さっきの『古今集』のテキストをこういうふう読み込ませます。それで、このまま実行を押しますと、こちらのスライダーバーで文字列の長さの下限というのを変えることができます。たとえば、ここ加減を 5 にいたしますと、5 文字以上の文字列で……。これ全部、仮名なのですけれども、5 文字以上の文字列で『古今集』の中で何回出てくる文字列が何種類あるのか、ということを経統計処理でまとめて提示します。たとえば、ここに 21 回以上出てくる 5 文字以上の文字列が、これ「3」とありますので 3 種類出てくるということです。ここをカチカチとクリックしますと、「はるかすみ」「さくらはな」「ほととぎす」という文字列が出てきます。「ほととぎす」が一番多くて 42 回出てくるということが分かるわけです。さらに、こうやってここをクリックすると、何番の歌に出てきたかということも分かります。



このようなことが一つできて、あとは、ふたつの和歌の集合を比べるということで、今回、小山さんのために『新古今集』を出してみようかなと思います。文字列解析器の画面は同じようなつくりになっていますが、これで AB で出現頻度を見てみます。

そうすると、実行を押して、もう少し文字列の長さの下を上げて 7 文字にしちゃうと、これ 7 文字以上の文字列で、ということで、大変縛りがきつくなっていくのですが、何回出てくるかということが分かります。先ほどの A というのが『古今集』で、B というのが『新古今集』を読み込ませましたから、たとえばこのクロス表の「6 : 0」というところ、これ「2」と書いてあるのですけれども、これは『古今集』に 6 回出るけれども『新古今集』には出てきていない、そういう文字列が 2 種類あるということです。これをクリックすると、たとえば『古今集』には「うつろひにけれ」が 6 回出てくるけど、『新古今集』

にはない。「ものにそありける」もやはり、同様の分布になっているということが分かります。もう一つ見てみると、たとえば A の『古今集』が「0」で、B の『新古今集』が「11」のところ、この「1」種類ある文字列はなんだろう。「ゆふくれのそら」だそうでございます。研究者の坎から言っても当たりかな。後で聞いてみたいと思いますが、ざっとこういうふうな感じです。



こういうふうにしていくと、どういう文化情報学部の学生にどういふふうな古典文学への入り口になっていくかということなのですが、統計処理ということで、結局、今のでもそうなのですが、古今集に頻出する文字列は何か、また新古今集ではどうかということがすぐに把握できる。また、古今集に頻出するのに新古今集には出ないとか、あるいはその逆とかいった文字列があることが把握しやすくなってきます。うちの学部の1回生、恐らく『古今集』や『新古今集』を読んだことも、本を手にとったこともない、ひょっとしたら名前も知らないという、そういう学生が交じっている中では、こういった文字列の出現頻度の差をまず認識してもらおうという、そういう試みの授業です。

もっとも先ほどの統計結果、たとえばこのクロス表なのですが、このクロス表を目の前にしてどうしていいかわからないという学生も、文系理系関わりなく少なくありません。そういうときには、「どこでもいいから、文字列解析器は爆発しやしないので、どれでもいいからクロス表の中のセルをクリックしてご覧なさい」というふうに指導しています。そうすると大体、このクロス表の端っこのあたりがすごく興味が出てくるはずなのですが、それも教えないで「とにかくクリックをしてご覧なさい」と。そうして、何かを発見してくれるといいなと。このスライダーバーも動かせば、文字列の下限が調節でき、クロス表の数値が動きますので、最も何かを発見しやすい、文字列の出現の偏りを最も発

見しやすい文字列の下限はどのあたりか、ということも学生にも考えさせるというようなことです。で、何か特徴的なことを見つけたら、これはもう常識的なことになるかと思うのですけれども、でも何でも何か特徴的なことを見つけたら、「自分で見つけ出せて良かったね、さすがだね、やったね」というふうに褒めてあげると。で、自分でも見つけ出してうれしい、何かここに発見する種が埋まっているかもしれないというふうに思ってもらおう。できるだけそういうふうな意欲というものを持ってもらいたいなというふうに思っています。

で、こうして今まで全くなじみのなかった和歌に触れた学生の中には、特に促さなくても言葉の意味を知りたいという子がぼつりぼつりと、やはり、出てきます。それまでは、「単なる文字列だから意味は知らなくていい」って言っていたわけなのですが、じゃあこの言葉の意味はなんだろうというのを当然知りたくなってくるわけです。そのタイミングでしかるべき文献やウェブサイトなどを紹介することで、これまで学生が分析対象としていた単なる文字列が、意味を持つ言葉、文学作品として再認識されることになる。これが理想です。

ここからやっと、文学部で行う講義や演習と同様の作品そのものの読解に入っていくことになるわけですが、実際の授業ではその前段階で時間切れというふうになってしまいます。半期の授業をさらに半分に分けた全7回の授業回数ですので、その先は2年生以上が受ける講義や演習につながるのが精いっぱいといったところです。

5、和歌の本文異同

2、Digital Humanitiesの実践例（2）和歌の本文異同

同じ歌番号の歌（自動生成した書き入れ本文も含む）の本文を列挙し、異同箇所を句ごとに表示

同じ歌番号を持つ歌を比較						
桂宮	530	さよふけて	いつらん月を	たかやまの	みねによそへて	なかもつる哉
	0	さよふけて	いつらむつきを	たかやまの	みねによそへて	なかもつるかな
	1	さよふけて	いつらむつきを	たかやまの	みね(のしろくも	かくしてしかな)
	2	さよふけて	いつらむつきを	たかやまの	みね(のしろくも	かくしてむかも))
永香	530	さよふけて	いつらん月を	たかやまの	みねによそへて	なかもつるかな
	0	さよふけて	いつらむつきを	たかやまの	みねによそへて	なかもつるかな
	1	さよふけて	いつらむつきを	たかやまの	みね(のしろくも	かくしてしかな)
	2	さよふけて	いつらむつきを	たかやまの	みね(のしろくも	かくしてむかも))
寛文	530	さ夜更て	いつらん月を	たかやまの	みねによそへて	なかもつる哉
	0	さよふけて	いつらむつきを	たかやまの	みねによそへて	なかもつるかな

科学研究費補助金基盤研究（C）

「文字列データ解析システムの構築と平安朝文学の伝本と表現に関する総合的研究」
（課題番号22500236、平成22～24年度）研究成果報告書

<https://kaken.nii.ac.jp/file/KAKENHI-PROJECT-22500236/22500236seika.pdf>

次に 2 例目に入りたいと思います。2 例目は和歌の本文異同ということで、いわばデジタル校本の試みです（スライド「実践例（2）」）。

このツールも先ほどの竹田氏との共同研究で作成しました。つい昨晚、このツールに名前が付きまして「e-CHA（イーチャ）」といます。余談になって恐縮なのですが、情報科学の方で自分でツールをつくと魅力的な名前を付けたいみたいなのですね。さっきの「e-CSA（イクサ）」というのも博多弁で、すごいいいよという意味なのですが、この「イーチャ」もやっぱり博多弁です。これで名前が「イーチャ」というふうに付きました。このツールの特色は、ありとあらゆる墨付きをデータ化して、本行の傍らに書き入れられた文字なども、その書き入れが反映された本文を自動生成して本文異同を示すというところなんです。それを実現するためには、書き入れの文字にタグを付ける必要があります。

それでまた画面を切り替えます。まずどんなテキストなのかということをお見せいたします。

```
4
*
みふのたゝみね
春たつと/いふはかりにや/みよしのゝ/山もかすみて/今朝はみゆらん
はるたつと/いふはかりにや/みよしのの/やまもかすみて/けさはみゆらむ
*
s{拾一春}
5
*
*
やまかせに/とくるこほりの/ひまことに/うちいつるなみや/春のはつ花
やまかせに/とくるこほりの/ひまことに/うちいつるなみや/はるのはつはな
*
s{古一春上},1[1-2]r{谷イ|たに}
6
む月
大伴坂上郎女
うちのほる/さほのかはへの/あをやきの/もえいつるはるに/なりにけるかな
うちのほる/さほのかはへの/あをやきの/もえいつるはるに/なりにけるかな
*
```

これが先ほどと大変よく似たデータなのですけれども、これでたとえば、この 5 番のところを見ていただいたら、この 5 番というのは歌番号なのですけれども、やはり、上段と下段で漢字仮名交じりと仮名開きのテキストをつくっているのですが、その書き入れをどういうふうにデジタル化するかというところで、タグを付けるというのはこんな感じで。たとえば、この<R>というの仮名開き本文で、初句「やまかせに」の 1 文字目から 2 文字目にかけて、「たに イ」という異文表記があるということです。だから、これ「やまかせ」というのが「たにかぜ」という異文を持っているということなのですけれども、それをこういう形でデジタル処理をしやすいタグを付けることによって、あとはもう読み込ませる、ソフト「イーチャ」に読み込ませるというようなことでデータを分析していくということです。ですから、後ほど申しますけれども、学生がこれをつくることができればデータ処理ができるという状態であります。

それでやってみます。実は、このツールを開発した先生であります竹田先生と、それか

らデータ処理にはうるさい南里一郎（立命館大学非常勤講師・共同研究者）が、一生懸命デモに付き合ってくれました。ここで失敗すると、後から怒られるということなのですが、ともかく立ち上げてみたいと思います。ここに 10 本ぐらいデータ、同じ歌集の伝本をデータとして入れているのですが、たとえば桂宮本と寛文版本というのをクリックいたしまして、文字が小さくて恐縮なのですが「表記の異同」「読みの異同」というのがあります。これ「表記の異同」だったら桂宮本と寛文版本と一致しないものが多過ぎるのですが、読み、つまり意味の関わる異同ではどうかってなったときには、結構一致するものも多いのですが違うものもある。じゃあ違うものはなんだろうということで見ると、こういうふう

に列挙されます。そして、たとえばこういうふう

に右側のほうに——今、桂宮本をもとにしていますので本文異同があるところが、句ごとにこういう形で赤で示されるというようなことです。こうやってくると、たとえばこんな感じでどこが違うかというのが見えてきたり、書き入れの本文がある箇所なんかもここに出てきます。横に歌番号があるのは漢字仮名交じりの本文を出して、ここからこの下がいわゆる仮名開き本文の分析結果です。このデフォルト「0」というのが、これ本行の本文なのですが、あと 1、2、3、4 というふう

に異文があればあるほど、ここに書き入れがたくさんあればあるほど、ここにだあっと書き入れを加味した本文のバリエーションが並ぶというようなことになります。で、こういうふう

に、句ごとの異同なのですけれども、一応示すことができます。

同じ歌番号をもつ歌を比較									
桂宮	530	さよふけて	いつらん月を	たかやまの	みねによそへて	なかつる哉	530	*	*
	0	さよふけて	いつらむつきを	たかやまの	みねによそへて	なかつるかな			
	1	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
	2	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
永青	530	さよふけて	いつらん月を	たかやまの	みねによそへて	なかつるかな	530	*	*
	0	さよふけて	いつらむつきを	たかやまの	みねによそへて	なかつるかな			
	1	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
	2	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
御所	530	さよふけて	いつらん月を	たか山の	みねによそへて	なかつるかな	530	*	*
	0	さよふけて	いつらむつきを	たかやまの	みねによそへて	なかつるかな			
	1	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	なかつるかな			
	2	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	なかつるかな			
寛文	530	夜更て	出らん月を	たか山の	麓によそへて	なかつる哉	530	*	*
	0	さよふけて	いつらむつきを	たかやまの	みねによそへて	なかつるかな			
	1	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
	2	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
黒川	530	小夜ふけて	いつらん月を	高山の	麓によそへて	なかつるかな	530	*	*
	0	さよふけて	いつらむつきを	たかやまの	みねによそへて	なかつるかな			
	1	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
	2	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
和学	530	夜更けは	出らん月を	高山の	麓によそへて	なかつるかな	530	*	*
	0	さよふけて	いつらむつきを	たかやまの	みねによそへて	なかつるかな			
	1	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
	2	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
江華	530	夜更て	出らん月を	高山の	麓によそへて	なかつる哉	530	*	*
	0	さよふけて	いつらむつきを	たかやまの	みねによそへて	なかつるかな			
	1	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
	2	さよふけて	いつらむつきを	たかやまの	みねのしらくもが	きあかころのけ			
富崎	530	夜更けは	いつらん月を	高山の	麓によそへて	なかつる哉	530	*	*

そして、たとえば今、実はこのツールに読み込ませているのは、『古今和歌六帖』の伝本 10 本なのですけれども、たとえばこれ全部使ってどんなふうになるか。これまだ工夫の余地がすごくあるのですけれども、全て選択を押して先ほどの同じ歌を歌番号で検索す

ると、こういう形で出てきます。一番上が漢字仮名交じり、ここから下というのが、「0」というのが本行の本文。それから、1 以下は書き入れ等々を組み合わせで生成された異文も込みの本文です。これで面白いのは、今、底本を桂宮本にしているからこういうふうな朱の出方、赤の出方をしていますけど、これを寛文版本に変えると、当然のことながら赤のハイライトの位置が変わります。こうやってデータがある中で、結局どの本文について自分が研究したいだろうなんていうことを、初心者でもテキストさえあれば、目算をつけやすい、割と使えるツールではなかろうかなというふうには思います。

他にもいろいろ本文比較の機能があるのですが、今回はこの程度でお話を進めさせていただこうと思います。それで、このツールは卒業研究や大学院博士前期課程の授業で使いました。たとえば『新編私家集大成 CD-ROM 版』（エムワイ企画、2008 年 12 月）に収められている私家集の複数の伝本のリストを使って本文異同を見ていくというようなことをやった学生もいました。また、『古今集』の伝本として定家本と元永本と伝藤原公任筆本を比較したと。とはいっても特定の部立だけちょこちょこっとテキスト作ってやってみただけではあったのですけれども。でも、そういう学生も出てきています。さすがに書き入れまでを扱った学生というのは、まだいないのですけれども、自分でテキストをつくって登録することができることもあって、具体的な問題意識を持ってそれなりに頑張った学生もいたようです。

6、理系教員とともに——（1）物語の本文異同——

それで、次の 3 例目、4 例目が先ほど申しました 3 年生の演習系の必修授業です。いずれもこれ、文系教員である私とともに理系教員、特に情報科学の研究者が常に手を携えて——というか、こちらからすり寄っていつているわけなのですけれども——情報科学研究者が同じ教室で指導をするという、そういう形態を取っています。

この 3 件目の物語の本文異同なのでも（スライド「実践例（3）」）、たとえば『源氏物語』でも CD-ROM『角川古典大観』（角川書店、1999 年 10 月）には翻刻本文が 4 種類も収録されています。また、それらのテキストデータをダウンロードすることもできます。そこでこの授業では、それらのテキストに別の伝本の翻刻テキストを加えて——これは授業で作るのですけれども——、本文異同の出現状況を分かりやすく図示した上で、それぞれの伝本の性格を具体的に把握することを試みました。

2, Digital Humanitiesの実践例 (3) 物語の本文異同

坂田桂一氏 (同志社大学文化情報学部1期生) 作成
散文用校本作成支援ツール



エクセルのマクロで異同箇所を分割・分類
(深川大路氏 (同志社大学文化情報学部) 指導)



SplitsTreeを用いた分析
(矢野環氏 (同志社大学文化情報学部) 指導)

科学研究費補助金基盤研究 (C)
「文字列データ解析システムの構築と平安朝文学の伝本と表現に関する総合的研究」
(課題番号22500236、平成22~24年度) 研究成果報告書
<https://kaken.nii.ac.jp/file/KAKENHI-PROJECT-22500236/22500236seika.pdf>

2016.9.5

Doshisha University
16

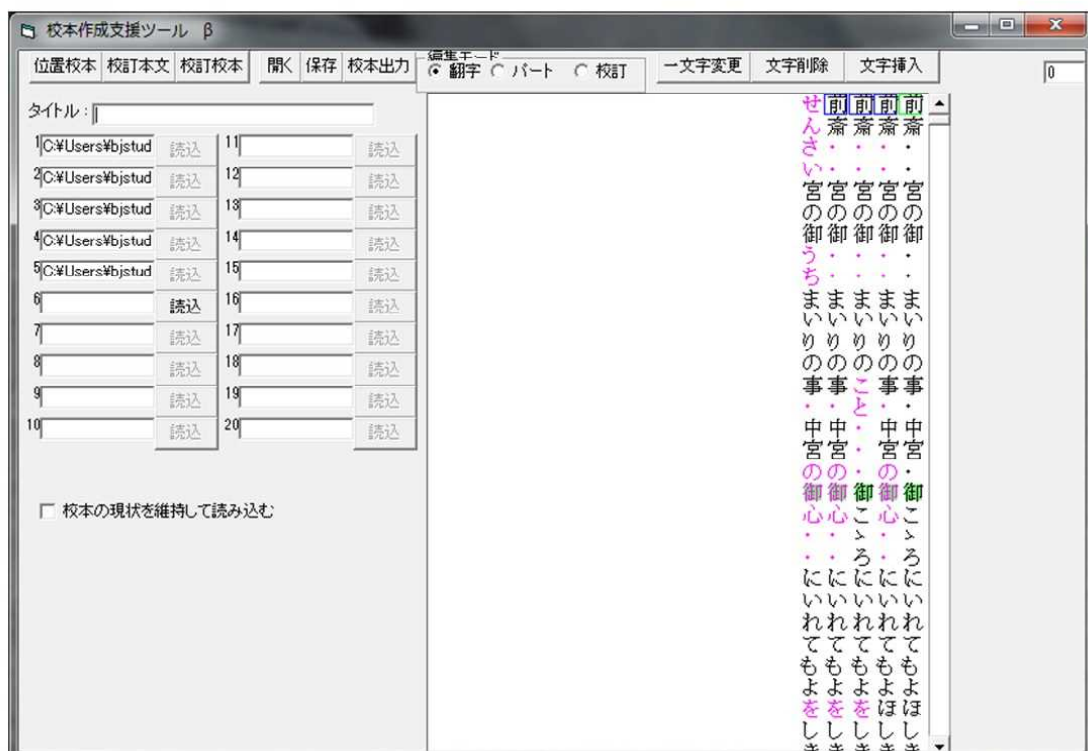
で、演習の手順はスライドに挙げておりますように、大きく3段階に分かれます。まず、散文用の校本作成支援ツールですが、これに先ほどの翻刻テキストを読み込んで校本を作成します。このツールは坂田桂一君という文化情報学部1期生が作成して残しておいてくれたものです。これやっぱり専門家が見ると見方も違って、情報科学研究者に言わせると、学生レベルを超えるプログラムではないというふうなことだったのですけれども、校本の作成段階から底本を決める必要がなかったり、いったんできた校本に新たな本を足すことができるというようなことを、この子は考えてつくってくれましたので、そういう意味ではうちの学部や研究科では結構重宝して使っているものです。

このツールで作成した校本を次にエクセルのマクロで異同が生じている箇所を分割して異同箇所1カ所について、いったい何種類の異同が生じているのかというのを数値化しますが、分類してしまうわけですね。これは簡単なプログラムで実行できるということなのですが、私自身はプログラミングなんかはととてもとてもできるものじゃございませんので、こういうときは情報科学研究者をお願いすると。今回は深川大路氏に書いていただきました。そうすると、異同箇所について番号が打たれます。「1・1・1・1・2」とか。

で、そういうデータが出来上がりますと、これはもうご存じの方も多いと思うのですが、Splits Treeという、遺伝子を分析して生物の進化、系統図をみるとか、そういうのでよく使われているソフトらしいのですが、本文異同のデータって本当に遺伝子の解析とほぼ同様の分析の方法になっています。そういった「0」「1」になってしまった異同データを Splits Tree という——これフリーソフトなのでどなたでもご利用になることができます——そのソフトを使って、異同の出現状況を目に見える形で図に表します。このソフトは現在、早稲田大学にいらっしゃる新美哲彦氏も以前お使いになっていたものです。新美

氏が呉工業高等専門学校にお勤めの頃にこの Splits Tree をご紹介になったのが、現在、文化情報学部にいらっしゃる矢野環氏で、当時、いろいろなところでこのソフトを宣伝なさっていたようですが、国文学研究者で興味を持たれたのは新美氏だけだったそうです。その矢野氏と授業をご一緒させていただいて、授業の中で Splits Tree とはどのようなソフトか、その使い方についても指導していただいて、学生とともに私も学びました。

では、一連の手順を実際にやってみたいと思います。だんだんと手順が多くなってきて私も自信がないのですが頑張ります。こっちが坂田桂一君が作ってくれた校本作成支援ツール。これを読み込んでいくのですが、テキストをご準備してまいりました。飯島本のテキスト。『源氏物語』CD-ROM から同じ箇所のテキストをダウンロードします。飯島本はとにかく翻字をする。そして、それらのテキストを、改行を解消して1本の文字データにします。これをツールを読み込んでいくわけですが、順番にこのように読み込んでまいります。すると、1行ずつ行が増えていることがお分かりいただけるかと思います。これで横並びに文字を見ると、少しずれたりとかしているのですが、それは手作業での修正が必要になります。この支援ツールでは、ぴったりと校本を作ることはできません、やはり。形態素解析などは全くしておりませんので。ただ、さまざま調整をいたしまして、なるべく後から手を入れないで出力できる塩梅はどこかということを実験して詰めた感じがこれです。だから、こういうふうになんか横並びの文字がずれたりとかしているところがあるわけです。



そこで取りあえずいろいろごによごによ修正をいたしまして、修正したものをたとえば、こういう形で出力をいたしますと、どういうものが出てくるかということ、こういうものが

出来上がります。こんな感じです。

前斎・宮の御・まいりの事・中宮・御こゝろにいれてもよほしきこえたまふごまかなる御とふらひまてとりたてまつる。

 〃の御心。給。た。

 〃こと。を。た。

 〃の御心。給。た。

 せんさい うち。の御心。を。た。

 □↓

 前斎・宮の御・まいりの事・中宮の御心・にいれてもよほしきこえ給。ごまかなる御とふらひまてとりたてまつる。

 〃御こゝろ。たまふ。まつ。

 〃こと。御こゝろ。を。たまふ。

 〃。を。ふ。

 せんさい うち。を。たまふ。

 □↓

 前斎・宮の御・まいりのこと。御こゝろにいれてもよほしきこえたまふごまかなる御とふらひまてとりたてまつる。

 〃事・中宮。ほ。まつ。

 〃事・中宮の御心。ほ。給。

 〃事・中宮の御心。給。

 せんさい うち。事・中宮の御心。

 □↓

 前斎・宮の御・まいりの事・中宮の御心・にいれてもよほしきこえ給。ふごまかなる御とふらひまてとりたてまつる。

 〃御こゝろ。ほ。たま。まつ。

 〃。ほ。

 〃こと。御こゝろ。たま。

 せんさい うち。たま。

 □↓

 せんさい宮の御うちまいりの事・中宮の御心・にいれてもよほしきこえたまふごまかなる御とふらひまてとりたてまつる。

 前斎。御こゝろ。ほ。まつ。

 前斎。ほ。給。

 前斎。こと。御こゝろ。

 前斎。給。

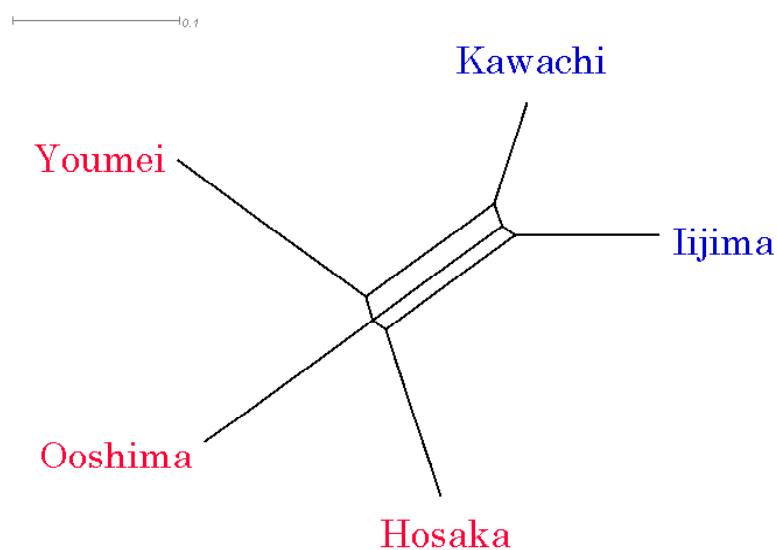
これ、伝本を5本読み込んで5種類の校本が出てきているというのは、もはや説明を要しないだろうと思います。5本すべての伝本を、順繰りに底本にして、5種類の校本を一度に出力するわけです。一番上の行にあるのが底本です。だから、同じデータですけど、校本によって見え方の印象が変わります。だから、やっぱりこれを底本にしたら面白そうだということがここで分かれば、その校本で研究を進めることができます。これがあれば、後はいろいろと気づくことがあるはずなのですが、ただそれでもうちの学部の子は、もう一越え、何か導いてあげなければならないと。

Iijima	前斎	宮の御	まいりの	事・中宮	御	こゝろ	にいれてもよ	ほ	しきこえ	たま	ふ
Hosaka	〃	〃	〃	〃	の	御	心	〃	〃	給	〃
Kawachi	〃	〃	〃	こと	〃	〃	〃	〃	を	〃	〃
Ooshima	〃	〃	〃	〃	の	御	心	〃	〃	給	〃
Youmei	せんさい	うち	〃	〃	の	御	心	〃	〃	〃	〃
Iijima	1	1	1	1	1	1	1	1	1	1	1
Hosaka	1	1	1	1	1	2	1	2	1	1	2
Kawachi	1	1	1	1	2	1	1	1	1	2	1
Ooshima	1	1	1	1	1	2	1	2	1	2	1
Youmei	2	1	2	1	1	2	1	2	1	2	1

たとえば、左から順番に見ていったときに、もうここで既に異文が、「1・1・1・1・2」というふうに出てきています。こういうふうな異文が出てきているところのいわゆる文節というわけではないのですが、異文ごとに種類に分けて、「1・1・1・1・2」は何か所ある、というふうに整理するためにもやっぱり、ずっと表を読んでいくのは面倒くさいというふ

うに言われます。先ほどのこれですね。これがいわゆるエクセルのマクロで、左端から順番に文字列が違ってくるにきたらセルを切って、ということで、異文の種類別を行います。こういう感じです。そうすると、あつという間に「1・1・1・1・2」ってなっています。これは一番上の……。今回は飯島本ですが、飯島本がデフォルトで「1」です。でも、陽明文庫本は、これ「せんさい」と仮名になっていますので、「2」というふうになっています。これをもし、表記ではなくて語句の本文異同にしたいと思われる場合は、ここのテキストを調整しなければなりません。いわゆる表記も統一するということが必要になっていきます。こうやって見てくると、「1・1・1・1・2」とか、同じパターンで出てくるかと思えば、「1・2・1・2・2」とか、そういうふうに出てくるのがあって。こうなってくると、同じ数字の組み合わせのところを拾っていきたくはなりますが、それも一応、これで見るとは可能です。

ただ、これを先ほどの矢野氏や新美氏がお使いの Splits Tree で見ると、これが Splits Tree の実際に読み込むテキストファイルなのですが、ここに伝本の名称、そして、先ほどの「1・1・1・1・2」をここに張り付ける。全く同じ文字列だったら全部1ですから、Splits Tree の図は、中央の一点に集まるはずですが、当然、点はばらけてきますよね。そのばらけ具合というのを分かりやすく視認性高くということで、先ほどのファイルを読み込んでいきますと、どこかでご覧になったことがあるかなというようなものが出来上がる。



これ、私も正直言って正確に説明しろと言われても勉強が必要なのですから、ただすぐ分かることは、距離と方向で分類が視覚的に分かりやすいように示されるようです。ですから、この場合、私、テキストをつくっていても感じていたのですけれども、ここの飯島本は河内本に近いなと思っていたのです。そうしたらちょうど、こういうふうに二つに大きく分けるところで切れるということがわかります。けれども、もうご推察のとおり、こちらでもわずかながら切れる箇所がある。要するに、陽明文庫本と河内本が何か特徴的なものをごく少数、共通して持っているものがちょっとだけあると。だけど、大まかに言うと、先ほどのように分かれる。

ここまでやると、少しうちの学生もやる気を出して、「じゃあ私、陽明文庫本を中心に

やります」とかいうことになってきて、ここでようやく、いわゆる文学部でなさっているような授業になっていくわけです。こういうふうなことで結局、その Splits Tree で図を描いてみますと、複数の伝本間における本文異同の出現状況の概要が一目で取りあえずは把握できる。だけど、もちろん図を描くことが目的ではありませんから、ここからどの伝本に注目していくのか、本文異同の質はどんなものか、具体例に戻って考察することになりますが。でもやっぱり、この授業も半期 15 回なので、肝心のこの考察部分が手薄になりがちなのが問題です。それでもここで興味を持った学生の中には、卒業研究で『源氏物語』の伝本研究に取り組んだ子も少なからずいます。

7、理系教員とともに——（2）画像注釈システムの構築——

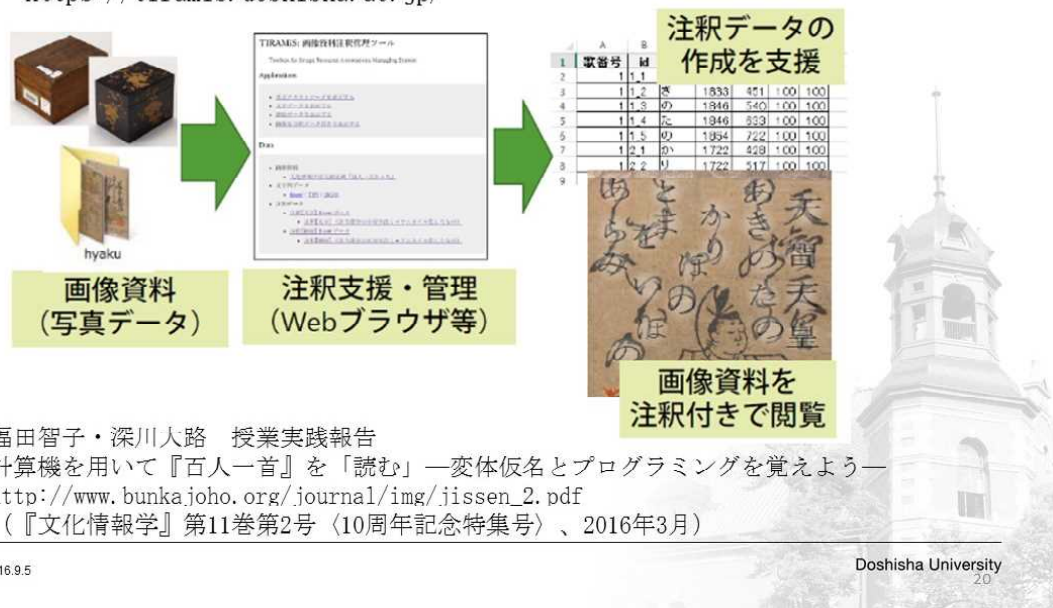
さて、最後の話題にまいります（スライド「実践例（4）」）。

2、Digital Humanitiesの実践例（4）字母データベース構築

深川大路氏（同志社大学文化情報学部）作成

TIRAMiS “Toolbox for Image Resource Annotation ManagIng System (ティラミス)”

<https://tiramis.doshisha.ac.jp/>



福田智子・深川大路 授業実践報告

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—

http://www.bunkajoho.org/journal/img/jissen_2.pdf

（『文化情報学』第11巻第2号〈10周年記念特集号〉、2016年3月）

2016.9.5

Doshisha University
20

最後なのですが、これは今、継続中の授業で、近年、ここ国文学研究資料館の大型プロジェクト（「歴史的典籍 NW 事業」）でも力を入れて取り組んでいらっしゃるように、古典籍の画像をウェブ上に公開するということが盛んに行われています。そこで、このような画像データを学生や国文学初心者の方にもさらに身近に感じられるように、先ほどもお名前を出しました深川大路氏とともに、画像に注釈を付けるツール、これティラミス（Tiramis）といいます。やっぱり研究者によってネーミングの系統がすごい違うなと思うのですが、この方、ティラミスというふうにお付けになりました。ティラミスというのを開発しております。同様のツールは、東京大学の史料編纂所や奈良文化財研究所でも積極的に開発・利用されていますが、この授業では文化情報学部ということもありまして自





分で何か作ってみようかな、ツール開発から入って、国文学を学生に身近に感じてもらうかなというようなことがあります、独自のツールを使って学生とともに注釈を付けています。

注釈の内容はどのようなものでも構いませんが、変体仮名やくずし字で記された文字は現代人にとってなかなか読みづらいものですので、現在使われている表記文字に置き換えたデータを作成して画像にリンクさせる、あるいはもう少し頑張って変体仮名の字母データをつくってやはり、画像にリンクさせる、その作業を学生とともに行うわけです。

作成中のデータなのですが……。URL を先に出さないといけません。これ実は、ネットに上がっておりますのでご覧になれます。ただ、作成中のデータをこの URL にアップしております。完成前ということで、まだリンク切れの部分が多いため、このユーザー名とパスワードを付けて一部の方のみにご覧に入れるという段階です。この点をご理解いただいた上で、少しだけ作成中のデータを見てみたいと思います。

TIRAMIS: 画像資料注釈管理ツール > 本文テキストデータを表示する

1. あまのたの

number	phrase	pos	text	jibo	yomi	thumbnail
1	0	1	天	天	てん	
1	0	2	智	智	ち	
1	0	3	天	天	てん	
1	0	4	皇	皇	わう	
						

さて、それで本文のテキストデータを表示するとこんな感じです。ちょっと「天智天皇」とかこんなふうを書いてあって。読み仮名も付けようというので。こんなふう画像を切り取って、それで仮名と字母と切り取って自分たちで入力するというようなことです。これをやっています。基本、学生たちは変体仮名を読めません。ですので、このデータベースをつくるためにといって学生には変体仮名を読みましようと言っています。こういうふうなのをつくります。それと文字データを表示する。ここで少しだんだんと研究に学生を引きずり込むという段階になってくるわけですが、そうやって画像を切り取って、これが「か」とか「あ」とかやっていたら、それもすなわちデータベースを作っているわけですね。そうすると、集計は簡単なので、たとえばこれ同志社のかるたなのですけども、「か」の字母が、「加」と「可」と「閑」ですが、どれが多いかなんていうこともここですぐ分かるわけで、こうやって見られるということですね。こんなふうに見られるということ。だから、同じ「か」でも実はちょっと字体が違っている、みたいなことも、これで割と分かりやすく見ることができるので、かるた以外にも、もしこういうのをやってくれ

るんだったら基本的なことですけど、学生の卒研などにいかがでしょうと私はいつも宣伝をしております。

TIRAMiS: 画像資料注釈管理ツール > 文字データを表示する

文字 (文字列) か (304B): 161件 字母 ALL

number	phrase	pos	text		i	thumbnail
<u>1</u>	2	1	か	可	か	
<u>1</u>	4	2	か	可	か	
<u>2</u>	5	4	か	可	か	
<u>3</u>	4	2	か	加	か	
<u>3</u>	5	4	か	加	か	

それから、これ応用編というか、ちょっと面白みをつけようと思って。これかるたなのですけれども。小山さん、お使い立てして申し訳ないのですが、この下句の中央の部分のこれをこういうふうに、「る」はここか、とかやってみてくださる？（「形が重なるものを選んでいくということですね」と操作される）これ今、影印なのですが、当然、影印の裏側に字母や今の仮名に直したものがあるので。ちょっと間違ってみたりしてくださる？形をはめるといことです。



これ、本当に遊びなのですからけれども、でも学生ってゲームをかなり時間をかけてやっていますよね。それだったらこういうので遊んでほしい、という感じで。これは別にゲームが目的ではなくて、こういうものがつくれるのであれば、もっと小さな子たちでも文字の形だったら分かりますので、ぴゅっぴゅっぴゅって、まず文字を文字として認識しなくても、「これなんか形が同じや」というふうにやってみる。

というわけで、一つゲームを作りましょうという提案で引っ張って、かるたから入って、気が付いたら字母データベースを作っていたというような状態なのですが、それでも学生は前よりは、選択科目の中のこのクラスに集まってきているように思います。前までは、たとえば仮名の字母の分布を見ましようとか、それから字余りの法則を自ら見いだしましようとかやっていたのですが、本当にかすかすの人数しか集まらなかった。どうしたらいいのかなというので、これで多少持ち直したということがありました。これはまさにネット上に出しておりますので、ぜひ見ていただきたいと思いますし、あとは私自身もいろいろな意味で改良の余地が多分にありますので、それは学生とともに鋭意やっていきたいと思っています。

8、おわりに

今回のお話は以上ですが、最後に文化情報学部の学生と日頃接していて一つ気になっていることをお話して結びに代えたいと思います。1年生と3年生の学生が4年生の卒業研究の発表を聞くという授業があります。ある程度専門的な内容を聞くことになるわけですが、下級生はまだ自分のゼミを決めるには至っておらず、まさに何でもありの発表を聞くことになります。聞く側の学生にとっては発表の中に興味のある内容もそうでないものもあるでしょう。それはよく分かります。しかし、発表に対する質問の前置きに、「僕は〇〇には詳しくないので」とか、「私は△△をよく知らないのだ」とかいう傾向があるのです。たとえば、私のゼミの学生の発表に対する質問であれば、「僕は古典に詳しくないので」とか、「私は和歌をよく知らないのだ」とかいった具合です。この「詳しくない」とか「知らない」とかいう言葉は恐らく、これからも学ぶ気はないというニュアンスを多分に含んでいる発言ではないかと思っています。「どんな学問分野でも学ぶ意思があるからこそこの学部に来たのでしょうか」というふうに思わず質問したくもなりますが、通常あり得ないこんな言い訳から質問に入ってしまうのは学ぶ対象が明確ではない文化情報学部の学生ならではの逃げ道であり、それは教える側からの悩みの種とも言えます。こういうとき国文学に限らず、学ぶとはどういうことか、学問研究に対する姿勢を身に付けさせるにはどうしたらいいのかと、そうしたことも考えながら教壇に立たなければならないと切に思います。

以上でご報告を終わらせていただきます。今回は、文学部国文学科以外の学生に向けて国文学研究の裾野を広げるための実験的な授業実践例をご説明いたしましたが、分析ツールに関しましては作成者にご相談の上、ご興味のある方にお使いいただけるように、今後手配していきたいと思っていますので、ぜひおっしゃっていただきたいと思っています。以上でご報告を終わります。ご清聴ありがとうございました。