

シンポジウム報告

「情報処理語学文学研究会」のテキスト・アーカイブス

内田 保廣[†]

共立女子大学の内田と申します。

私は国文学をやっております、江戸時代の文学を中心に研究しております国文学者で、ですから、長瀬先生のように世界的な広がりも持っていませんし、安永先生のように機械に対する知識もほとんどありません。そういう状態ですが、機械は使い始めてからかれこれもう20年近くはたっているんですけども、確かに非常に幾つかの大きな問題がありまして、そういう問題にぶつかりながら、同じような境遇の研究者と集まりまして情報処理語学文学研究会というのを始めました。お配りしてある資料の中に、前回のニュースレター16号を入れてございます。その中に古相正美さんが私どもの会の紹介を書いてくださっております。

よく外国の研究者に接触した時に言われることなんですけれども、「どうして日本はこんなにコンピュータがいろいろあるのに、日本の古典文学のアーカイブスはないのか」というようなことを言われたりすると、英語がわからないふりしてニコニコ笑って、「サンキュー、サンキュー」って言って握手してごまかすという、そのぐらいの手しかないという状況です。

今、資料館のほうで『古典大系』入れてくださって、この『古典大系』が外に出るようになれば、やはり相当な意味があるだろうと思うんですが、残念ながら、まだ今のところは、一般的には売ることみなさっておられないわけで、非常にこれは難しい、残念な問題です。

さきほど長瀬先生のお話の中にも出てきていたんですが、日本人といいますか、日本の文化の中で、著作権とか所有者、知的所有権ですね、この概念が多少希薄なんじゃないかというような言い方、これはよく聞くんですけども。これはやはり、ずいぶんこれまでの日本の著作物についての考え方と申し

ますか、そういうものの中にこれは内蔵されていたものではないかなと私などは思っております。

たまたまきょうここに今持っているんですが、これは私の恩師といいますか、慶応大学の西村亭先生がお書きになった『歌と民族学』という本なんです。この中で、「歌というもの、和歌というものにいったい作者というものがあつたんだろうか」ということが非常に問題になるくらい、われわれに問い掛けられている問題なんです。

文学作品にいったい作者というものがあつたのか。これは、今われわれには、これは誰その歌だとか、これは赤人の歌だとかなんだとかいってますけれども、これほんとにそうなのか。あるいは、今われわれが持っている作家、作者の概念をそのままあてはめていっていいのかという問題がずっとありました。

江戸時代になりまして作品というものが商品化されてきた過程でも、はたして、作家と作品の対応関係といいますか、そういうものを現代と同じようにとらえていいのかということはずいぶんまだ問題があります。という具合に、国文学者というのはどうしてもその辺のことをそのように考えてしまうものですから、著作権とか知的所有権とかいうことに対してはやや意識が希薄であろうと、これは仕方のないことだろうと思っています。

私どもの会が、出来上がりました時の基本的な考え方は、こちらに古相さんが書いてくださっているんですけども、基本的には3つありまして、まず、研究者が自分の研究に必要なものを――本文です、特に入力していると。これは自分で索引を作りたい場合もあるし、それを翻刻という、さきほどの安永先生のお話の中で出てきておりましたけれども、活字本にするためにワードプロセッサの代わりに機械を使うということもあるわけで、翻刻しているというようなことをします。そういうことをしていった、終わったあとですね、これをそのま

[†]共立女子大学

まにしておくのはもったいないじゃないかということが、まずあります。これはほかの人も使えるんだからお互いに提供しようということが、まずありました。

次に、もう1つは、たとえば、『源氏物語』の本文というのは電子化されていれば検索が非常に楽です。もちろん、総索引がございすけれども、索引なしでも検索が出来るというところで、非常に楽になります。だからといって、索引が要らなくなるということではないんですけれども、かなり楽に使えます。

それから、『源氏』だけではありません、いろいろな作品があれば、われわれは言葉を調べるのがとても楽になります。そういう時に、じゃあそれを作ろうと思った人がいて、『源氏物語』を一生懸命入力を始める。ただ、これを入力している時に、ほかでもやっているかもしれないわけです。こういう人たちが並行して同じ作品を何か所かで入れるのは、これはばかばかしいわけで、それならば、何人かで手分けをして、さきほどの話でも出てまいりましたが、『源氏物語』には異本があり、同じ『源氏』入れるなら、5人なら5つの異本を入れればいいんで、1つの本を5人が集まったらこれはばかばかしいじゃないかと、こういうことの調整をしようじゃないかということですね。つまり、同じ作品を方々で入れることはないだろうと、お互いにうまくそれを調整しあっていこうじゃないかということなんです。

国語研究資料館のような所で非常にスタンダードな、いわゆる校訂された本文や、あるいは基本的なものが入れられていくというところで、われわれのほうは、どうしても自分の研究に必要なものになりますから、非常に個別的なものになってきます。

この、古相さんのものの中で右側のほうの、下にちょこっとありますが、そこから右にかけてこう出ておりますものがありますが、これなどご覧になっていただくと、ずいぶん特殊なものです。たとえば、そうですね、高木元さんのお作りになった『旧幕引継文書『類集撰要』巻46』の翻刻テキストなんていうのは、一般的にはどう考えても使うもんじゃないんですが、近世文学の研究者の中ではこれは非常に使い途が、もちろん一部ですけれども、一部といってもこれは数10人だろうと思いますが、使い途があるもんです。一応こういうものまで入っているといいですか、こういうものをお互い同士流通す

るということが1つの目的ですから、当然入っているわけなんです。こういう作り方がされていくわけで、つまり、『源氏物語』だって岩波の『古典大系』や、あるいは小学館の本文ではなくて、異なった本文のものが欲しいと思う人は結構多いわけです。

全体の日本人の読書人口の中からいえばすごく少なくなるかもしれませんが、『源氏物語』の研究者の中の比率でいえば、非常に高くなってくるわけで、これは当然ですけれども、そうになると、そこいら辺のものも並行して入れていってほしいし、あるいは入れるべきだという考えが出てくるわけで、それだったら手分けをして入れようよと、これは当然の発想だろうと思うんです。

それから次に、長い本文、『源氏物語』の場合もそうですけれども、非常に長い本文が幾つかある場合、こういうものは1人で全部入れるんじゃないくて、何人かで手分けをして入れようというような形になってきます。そうすると、まあ共同作業のようなことが出来上がってきます。

国文学というのは、今まであまり共同のプロジェクトで何かをするということは馴染まない分野で、さきほどの安永先生のお話ですか、書斎の中で1人でやっているというのが非常に多いというお話ですが、ほんとに共同研究というのはあんまりやってないところなんですけれども、まあ、ここいら辺で、本文を作るなんてことになると、共同研究の形が出てくるということがあるんじゃないかと思います。

そういつて出来上がってきたデータがあります。これは私どもグレップなどを使って検索するだけでも十分に意味があるんですが、それ以外に、これを使ったらこんなことが出来たというような情報をお互いに持ち寄ろうということになりました。この3つが柱になりまして、こういうこと、同じようなことをやっている人間が集まって研究会を持とうではないかというので、1986年に一応有志が集まって、この会の発足といいますか準備会が始まったという状態です。

なぜこんなに、さきほどらい出ておりますように、外国のデータベースに比べて遅れたかという、最大の理由は文字の問題で、つまり、漢字が簡単に使われるシステムがそれまでなかったということですね。漢字セットを使って仕事をするために数千万円かかってしまうという環境では、今までそういった大きな予算を取ったことのない文学部の国文

科の研究室では、そんなものは取りようがなかったわけです。

また仮に、そんなものの願いを出したところを通るわけがなかったわけですが、1つのセットが、当時のおカネで40万ぐらいでしょうかね、そのぐらいでそろうということになって、初めてこれは手に入れることが出来たわけです。ですから、その脈絡を引いておりますから、今でも国文学者が機械を使う時は自前でやっておりますので、高いものは使えないというのが原則になっているわけです。

さきほどの、長瀬先生のお話のほうでいろいろ考えてしまうことがあったんですが、たとえば、諸外国の、これだけ出来ていて、それなりの価格で売られていると。これを諸外国の研究者は買うことが出来るわけで、これは所得の問題とか、そういうこともあるかもしれませんが、そういった文系の研究者の研究補助の体制というものが、日本と外国とでは全然違うわけです。日本では、こういうものを買うのはまだまだ難しい状態にありますから、これはパーソナルに買っていくしかない状態なんですね。そうしますと、これをパーソナルに買っていくとなると、この今の価格でもかなり苦しくなってくるんじゃないかと思います。

同時に、これは研究者の生活状態にかかわってきます。大学の教師をやっております者よりはずっと収入の少ない大学院生とか、あるいは学部学生というのを抱えておりまして、この人たちの利用というものはこの価格になったらとても無理だろうと思われまう。大学でサイトユーザーになって、それを一般の学生たちに使わせるという環境はまだまだ整っておりませんから、この辺も非常に、こちらとあちらの事情の違いというものがあって、このような形にいきなりは当分取れないのではないだろうかというように私などは思います。

話はちょっと横道に逸れてしまいましたけれども、ともかく、会としてはそういう方向でデータを集めてまいりました。ここに出ておりますのも最近のものなのですが、現状ではもう少しこれに加わっているという状態です。

そして、ここに集まってまいりましたデータの性格なんですけれども、これは基本的には原本から起こしていくものがいけば多いのではないかと思います。もちろん、中には活字本からのものもありますが、この活字本の場合にも、多くの場合、50年以

上前の活字本ということが多いようです。最近のものから起こしているというのは、そうですね、もちろんないわけではありません。しかし、最近の活字から電子化するということにはやはりちょっと多少の抵抗があります。

これは、たとえば校訂者のオーケーを得ていても出版社のほうからクレームがつくというところを経験することがありまして、そのために、最近の活字本からの翻刻というのはだんだんみんなが見合わせるようになっていきます。これは、実はよろしくないこととか、つまらないことなどではないかと思えます。私もど仲間うちでは、「これは5社協定みたいなもんだ」と言っているんです。5社協定というのは、テレビが出来た時に映画会社が協定を結んで、有名な俳優をテレビに出さないようにしてしまった協定がありましたね。あれと同じようなもので、実際、さきほどの長瀬さんのお話の中にありましたように、電子化テキストが出てくると本文をみんなが欲しがるといふ、そういう傾向を持っておりますから、実際には、出版社が自分で自分の首を締めていることになっているんじゃないかなと私も思います。しかし、それは言わずに、そういうものは作っていても、少なくともこのリストには載せないという形を取っています。

つまり、活字本として出ているもので生きているものといえますかね、著作権がまだ生きているものについては、われわれはなるべくこれを抑制する方向です。それに対して死んだものと申しますのは、もう50年以上たっていてその出版社も滅んでしまったようなもの、たとえば、「帝国文庫」なんていう立派な叢書がありまして、これは、博文館という有名な本屋さんが明治時代に作った本です。これは博文館ございませんから、今これを活字本から直しても、誰からも文句は来ようがないんで、これは翻刻しても大丈夫です。それでは翻刻させていただいて配ると。

ここに載っております私のやりました『近世説美少年録』は、これは博文館ではありません有朋堂文庫という書物から起こしました。それを原本に対象して、原本のほうの情報を盛り込んでといいますか、それで校正をして出しました。これなどは、仮に有朋堂文庫のページ数を入れてもたぶん問題の起用はないんじゃないかと思っています。私どものほうとしては、これは大丈夫だろうと思ってやっているこ

となんです、それ以外で、やはり今生きているものだと問題が出た時に困ってしまうなというところはあります。

原本から直接やるもの、これは全く問題がないだろうと思います。著作権者はもちろん死んでおりますし、死んで何百年もたっています。その出した版元も、江戸のニシムライチロベイなどという名前で、京都にあります、まずこれが主張してくることはなかろうと思われています。そちらのほうは安心だという状態で作業は進んでいます。

原本の作成までのところは今までお話いただいているので、流通の問題について少し詳しくお話をしたいと思います。といってもあまり時間がないので簡単に済ませますが、流通の問題については私どもの会は規約を作っておりまして、その規約、一応全部持ってまいりましたけれども、テーブルの上に置いてあるものですが、まあ、2枚綴られておりまして、その中の、「研究活動細則」というものがまずいちばん大事なもののじゃないかと思っています。

「研究活動細則」の7、8が基本的にこの問題にかかわってきます。7、8、9になりますかね、10ぐらいまでちょっとまとめて見ていただくとありがたいんですが、7には、「会員は製作した機械可読テキスト等を会を通じて会員に無償で提供することができる。ただし、提供を受けた会員は、提供者の合意なしに非会員に複製物を配布することはできない」と書かれております。

つまりこれは、会員は会員の中にももちろん配るわけですね、欲しい人がいたらお渡ししますが、それを会員外には勝手に複製してもらっては困るよということですね。これは、私どものこの会のテキスト・アーカイブは、会員にのみ配布したいか、あるいは会員外に配布してもよろしいかという、そういうことを聞きながら提供しておりますので、「会員内のみ」というのはこういう形で、会員の中だけ、第3者には複製はしないでくださいという形になっているわけです。

それから8には、「機械可読テキストの利用の際は、その作成者に対して利用を通告し、その成果を報告し、さらにその成果を発表する場合には、利用を明記する義務を負う」と書かれております。これは非常に固い文章で言っておるんですけど、簡単に言ってしまうと、使わせてもらうよと言って欲しいということです。さらに、これを使ってこういう

ことをやったよということは教えてくださいということなんです。

つまり、自分のデータをどのように使ったのか、どういう成果が上がったのかというのは、これはデータ作成者としてはまず知りたいだろう、データ作成者も研究者の一員ですからそれは知りたいだろう。それは教えてあげるようにしよう。お互いに教えっこをしようというわけです。そのために会があるわけですから。

それからもう1つ、どこかに発表した場合には、これを利用したということも出してくれと。これは普通の論文の引用と同じであるというわれわれの解釈です。それで、誰々の作ったどういうテキストデータを使ったということを明記してほしいという要望を出したんです。

次が9番目ですけれども、「テキスト作成者は、提供する機械可読テキストの利用に関する条件を付加することが出来る」というようにしました。

これは、たとえば、ある翻刻本文をどこかの出版社から印刷して出したいと思っている製作者がいたとして、この本が出るまではこれについては利用を見合わせてほしいというような要望を加えることが出来るという形をとりました。そうしますと、それが各会員の研究者の所に、使いたい人は使ってもいいんですが、それを使って何か論文を書いたりなんかした時には、その本が出るまで待ってくれということをするのが出来るだろうという形ですね。これは相互の了解の上でそういう形を取ろうじゃないかということです。

また、これの特にいちばん大きい条件は、「会員外には配布しないほしい」というようなことですね。これがいちばん大きい条件になると思います。

それから10番目には、利用者が誤りを発見した場合には、これは運営委員会というものがございまして、そこを通じて連絡をとるということです。これは利用者の発見した誤りというものが常に作成者が誤りと認めるとは限らないわけで、そのために、いったん連絡という形でひとつステップを置くという形にいたしました。

そして、今度は運営委員会の役割になります。公開テキスト、それ1つ1つについて、誰が使っているか、これを使う時はこういう条件がかかっているから気をつけてくれ、それから、これはいつ改訂したかというようなことを全部記録して持っている

いうことです。

つまり、バージョンアップが行なわれることがあるわけで、作った人間が、「これについてはちょっと字が間違ってたから取り替える」なんてことがありますと、運営委員会のほうに、「これは間違っていたから取り替えたい」ということをおっしゃってくださる。それをこちらは取り替えて、そして管理しております。これは、いつ変わったかということはこちらで持ってまして、私のデータはいついつこれを受け取ったんだけど、これはそのあとバージョンアップされているかどうかということは、こちらの運営委員会に問い合わせてくださればすぐわかる。ということが一応こちらの体制になっているというわけです。

これが配布に関する部分で、この配布に関する部分を具体的にしましたのが、その次にあります2枚目のほうになりますが、「複製依頼」と「配布委託願い」です。

最初に作った方、作ってお持ち寄りになる方は、この「配布委託願い」を出していただきます。ここに明記しておりますが、「私が著作権を保有している以下の機械可読テキスト」、つまり、この機械可読テキストについては、本文から起こす場合であっても、すなわち原本から機械可読化する場合であっても、それから何年かたった書物から機械可読化する場合であっても、あるいは、活字本のあるものを機械可読化する場合であっても、私どもに提供してくださる場合にはその著作権はその作成者が持っているということが前提になっているんです。

ですから、ここでの著作権上のトラブルがあることは、会に提供した場合にはありえない。もしあった場合には、それはその提供者本人とそのトラブルを起こした相手との問題になって、会はその中には介在しないという、非常に上手に逃げてしまうといえますか無責任といえますか、そういう形をとらせてもらっています。そこにはあまり入りたくないところですので、「貴研究会には一切ご迷惑をかけない」と。こういうことをつけますとあまりみんな提供してくれないんで困るんですけども、実際にはそれで提供してくださる方はずいぶんいらっしゃいます。あとは、これらの要点、これはもうおわかりいただけると思います。

片一方、「複製依頼」ですね。これは欲しい方が申し出ていただくものですが、これは、研究活動細則

と利用要綱の規定を守ってくださるということで、この書類を出していただくという形でやります。

そういう形でありますので、まだ始まって10年たっていない会ですが、トラブルは今のところ全然起こってはおられません。というよりは、ほとんどの方がこの研究会をやっている間に知り合いになってしまいますから、相互の友人関係の間で行なわれるんで、これは著作権法の関与するところではないという関係で行なわれていることが非常に多いのと、それから、ここに挙がっておりますものをお使いになる方たちが、お使いになって成果を挙げてらっしゃる方もいらっしゃるんですけども、お使いになる方たちがかなりこれらのことをよくご理解といえますか、条文上の理解ではなくて、今までの研究者の間にあるひとつの仁義といえますか、そういうものを明文化しただけのものでございますから、それ以上の問題には発展していないというのが現状です。

ただ、私ども非常に恐れますのは、出版社との関係です。校訂者のほうは実はあんまり怖くないといえますかね、変な言い方になりますけど、あんまり問題ないんです。これは校訂者は大体われわれの仲間、あるものを入力しようと思う時は、その元の本の活字本を作った人というのは大体われわれの仲間か先生か、あるいは、後輩であることもありますけれども、そういうような関係ですから、「お前ちょっと使わせてもらうよ、これをこうさせてもらうよ」と言うことで済んじゃうという、これは狭い世界のひとつの利点でございましょうか、そういう形になります。しかし、本屋さんが絡んでくると話は複雑になるという状態だご理解ください。

最後に、いちばん上のほうに戻りますが、この「入退会細則」というのがございますね。この入会については、まあそれほど問題はない普通のことなんです、退会が少し厄介になっています。

「この会から抜けようと思う人はテキスト利用権を返却してほしい」と。つまり、この会に入っていてテキストを貰っていて、データを貰っていて、それで辞めたいと言った時ですね、ただ、「そうですか、さようなら」というわけにはいかない。それをいったん返す形を取ってほしいということです。つまり、会員にだけしか提供してないテキストについて、データについては使わないでほしいということを含めております。これが退会規則の1つの形になっているということでございます。

今は、大体会員にしか提供していないものについて中心にお話したのですが、非会員に対しての提供形態については、会のほうとしてはなんの書いた規定みたいなものは持っておりません。が、一応大体、GNU に準拠してほしいと思います。GNU というのは、皆さんご存じのような、1 つの PDS ではないですね。あれは著作権を留保して、自分の作った製作物の流通を妨げる行為一切を排除する形で著作権をかけることが出来る。

つまり、著作権をコピーをしろという形で、コピーしていいよという形で、俺の作ったものについて誰かがコピーすることを妨げたら、それは俺の著作権を侵害するんだぞという形でかけているものですね、だから、これに準じてやってほしい。つまり、自由配布をそういう形で保障したいというもので、非会員には配布してほしいというように、文章的にはうたっておりませんがこの点について考えています。

あまり時間がないので簡単にまとめを申し上げますけれども、これはちょっと飛躍した話になってしまいますが、私たちが考えておりますことで、まず、パーソナル・ユースというものを確保、古典の研究の場合ですね、パーソナル・ユースを確保していきたいと。パーソナル・ユースが出来なくなるような状況は作ってほしくないと思っています。

なぜパーソナル・ユースが必要なのか、パーソナル・ユースの利点をどう考えているかといいますと、1 つは、ポータビリティといいますか非常に持ち運びが自由だといいますかね、展開しやすいものですね。いろいろな所に持っていったいろいろな形でそれを役立てることが出来るということが1 つと。それから、自分の研究に併せてそれを作り替えることが出来ます。さきほどの安永先生のお話じゃありませんけれども、ある所から自分用のデータベースをいくらでもつくり出すという機能を持っているわけで、これを殺すような形であらゆる所有権の主張というのは、これは古典研究に対してなされるべきものではないだろうと、このように考えているんです。

最初のところでお話ししましたけれども、古典文学というものは、あるいは古典というものは、もともと著作権などというものはないところから始まっているわけで、どちらかということこっちのほうが長くやっているわけですから、今の社会でこういうものが必要だから、それに合わせた形にしなければ

らないというのはよくわかりますが、それでこの古典の生命といいますか、それを損なうような形でこれはかけられてしまっただろうと思うわけです。

逆に申しますと、今ある知的所有権や著作権などというものは、社会体制なりなんなりの変化で、あるいは経済体制の変化でどうとでもなってしまうものですから、そのどうとでもなってしまうものによってずっと続いてきたものを潰されてしまうのは、これは困ります。もちろん潰すわけではないんですけれども、その自由な発展展開を阻害されてしまうことは非常に残念です。

同時にまた、あまりそういったコンピュータ・サイエンス、その他に古典文学が寄与するところはそれほどないのかもしれませんが、案外ひょっとして大きな展開がこの中に秘められているかもしれないので、その辺の可能性を見ていただいて、なるべくこういう問題については緩やかな見方をしていただきたいと、私たちは心からそう思っているしだいです。

ともかく、こうした国文学者などというものは、法的なことにも機械的なことにもあまり敏感でない人間がそろっているわけですので、なにかと問題点があるかと思います。私どもの会の活動についても何かございましたら、またあとのところで、いろいろお叱りなり、あるいはご意見なり頂戴できれば幸いですと思っています。以上でございます。(拍手)

司会：ありがとうございました。ちょっとお尋ねしますが、会員のメンバーとかちょっと数だけでもお教えていただけますか。

内田：現在会員は、…吉村さん、何人だかわかります？…約 150 人です。

司会：あ、そうですか。流通しているプロダクトの数はどのくらいでしょう。

内田：一応この中に載っていますのが 30 件ぐらいですか。この中に載っていますのがオープンになっているものです。

司会：どうもありがとうございました。

実は、ちょっとここでスケジュールの変更をさせていただきますが、ちょうど時間もよろしいようでございますので、15 分ほどコーヒープレイクにさせていただきます。今、この時計が 15 時 5 分でございますので、15 時 20 分から次に入りたいと思い

ます。じゃあコーヒーの用意もあるようでございますので、どうぞ。

〔休憩〕

〔再開〕

司会：それでは時間になりましたので、後半に入りたいと思います。今まで3人の発表者にご発表願ったわけですが、その内容につきまして、制度面と技術面からそれぞれエキスパートにコメントをいただこうと、こういう趣向でありまして、初めは文化庁著作権課の佐藤さんをお願いいたします。どうぞ。