

民話データベースの課題と可能性

0. はじめに

この報告で「民話」という言葉を使用することをお許してください。私は、フランス民話の研究から出発しました。フランスでは、口承文芸 (littérature orale) とほぼ同じ意味で民話 (conte populaire) という言葉を使用します。また、民間伝承 (tradition populaire)、口頭伝承 (tradition orale) などという表現も同様の意味で使用します。

これから、報告する「日本民話データベース」「沖縄民話データベース」という2つのデータベースにも、「昔話」「伝説」「神話」「世間話」などのいわゆる「話」のほかに、伝承の事情や俗信、神唄や祈り、神の祀り方や民謡などさまざまなジャンルのデータが含まれています。

このことは、日本の各地で民話あるいは昔話の聞き取りをなさった方なら、どなたでもご理解いただけると思います。私たちは、語り手から話を聞こうとするとき、語りにまつわる全てを聞き出そうと努力します。語りは、暮らしの中に生きているのですから、語りを支える暮らしの全てを記録しようと思います。

私が、ここで「民話」という言葉を用いて表現しようとするのは、こうした語りを取りまく全ての情報です。私たちの「日本民話データベース」と「沖縄伝承話データベース」は、いわゆる「話」のほかに、「話」を支える全ての情報が含まれています。

1. 日本の民話記録とデータベース化作業

1-1. 日本の民話記録

2010年は、柳田國男が『遠野物語』を著してから100年にあたるのだそうです。日本では、『遠野物語』以前にも、民話の記録は、さまざまな形で残されていますが、「聞き手が特定の語り手から語りを聞き取り、意識的に記録した」という意味では、やはり『遠野物語』は一つの時代を画し、日本の口承文芸学研究の第一歩を記したと言ってよいでしょう。

これ以降、次第に緻密かされていく民俗学・口承文芸学の枠組みの中で、多くの研究者が、優れた語りの記録を残してきました。

こうした事情は、ヨーロッパでも同じで、1812年の『グリム童話集』の刊行を一つの画期とすれば、日本よりも100年ほど早く記録が始まり、19世紀の末には極めて質の高い資料集が各地で生まれていたと言ってよいでしょう。

しかし、こうした記録の作業は1960年代に入ると一変したと言ってよいと思います。テープレコーダの普及とともに各地で音声ともなう忠実な記録活動が始まったのです。とくに日本ではオープンリールに続いてコンパクトなカセットレコーダが普及し、誰もが簡単な操作で音声記録を残せる環境が整いました。

私たちは、2001年から、それらの記録のデータベース化に取り組み始めたのですが、当初、1960年代から今日に至る半世紀の間に、少なくとも20万件から30万件の記録が残されたに違いないと考えました。これは膨大な数字ですが、後に、沖縄で遠藤庄治さんが当

時の学生とともに記録した話数が約 75000 件であるという事実を知り、それが決して大きな間違いではなかったと確信しました。

1-2. 民話記録の保存とデジタル化の必要性

私たちは、2001 年 4 月に「日本民話データベース作成委員会」をスタートさせたのですが、当時はデジタル化によるデータの保存ということが大きな課題で、データベース化の大切さということは十分に理解していませんでした。

民話の記録に携わるものなら誰もが気がついているように、20 世紀の終焉とともに、伝統的な語りの方に大きな変化が訪れました。もちろん現在でも、村々を訪れて語りを記録している方々は少なくありませんが、その一方で 1980 年代までは気楽に聞くことのできた話が、つぎつぎと消えていったことも事実です。その一方で、1960 年代から記録を続けてきた話の記録者の高齢化も進行しました。

語りの記録者の記録方法には、テープやカードの整理などに一定の共通ルールがあることは事実ですが、やはり手づくりの部分が多く、語りの記録の大半は、記録者が作業を放棄すると、ただの粗大ごみになってしまいます。幸運にも地域の図書館や大学の研究室に残された場合でも、管理者が変わると素性のよく分からない「使えない」資料になって埃をかぶってしまうのが実情です。

私たちは、資料のデジタル化によって、まず時の流れとともに劣化するアナログ・テープ資料を借り受けてデジタルにし、集中管理するすることを目指しました。デジタルの資料は、操作を誤ると一瞬のうちに消失してしまうという恐ろしい危険もありますが、複数の担当者が、複数の手段で保存すれば、恒久的に保存することが可能です。

私たちが作業をはじめた 2001 年には、コンピュータの能力も低く、音声の処理もままありませんでしたが、現在ではごく簡単な作業になりました。デジタル化した資料を保存する手段も、飛躍的に向上し、当初は 2 ギガ程度だったハードディスクの容量も 200 ギガがあたりまえになり、外付けのディスクを利用すればテラバイト単位のデータ保存も難しくありません。

さらに音声データの圧縮技術も進歩し、MP3 などの普及によって、たとえば iPod のような小型ツールのなかに 2 万話以上のデータを収納できるようになったのです。

1-3. データベース化のメリット

こうしたデジタル技術の進化と信頼性の向上が進む一方で、次第に明らかになってきたのが、データベース化のメリットです。これから「日本民話データベース」を実際に使用しながら、具体的にお話ししますが、民話のアナログ記録データには大きな問題がありました。「沖縄伝承話データベース」に関しては、比嘉久さんが詳しく説明なさる予定ですから、その内容には深く立ち入りませんが、その基礎となった沖縄国際大学の遠藤庄治さんと学生達の調査記録は、これまで日本でなされた民話記録のうちで最も体系的なものであり、方法的にも一貫したものです。この記録はカセットテープ約 4500 本に収められ、ほとんど全ての資料がカードに整理されています。この資料は、現在データベース化作業の最中ですが、作業が完了するごとに、テープは沖縄県立博物館の収蔵庫に保管されて行きま

す。日本の民話記録のうちで、このような手順で保存された記録は質量ともに他にないと思います。

しかし、誰かがこの資料を利用して、たとえば「アカマタ婿入り」と「浜おり」について研究しようと思いついたとしましょう。その研究者は、何日もおもろ町の沖縄県立博物館に通いつめて、テープカードを調べて沖縄各地の「アカマタ婿入り」と「浜おり」についての記録を確認し、当該テープを見つけ出し、テープレコーダにかけ、早送りとまき戻しを繰り返して「アカマタ婿入り」と「浜おり」の語りを探し出さなければならないのです。この作業には、膨大なコストがかかります。よほどの覚悟がないかぎり不可能な仕事です。

これに対して、「沖縄伝承話データベース」が完成すれば、小さなノートコンピュータに全ての資料が収納されているわけですから、東京にしようと北海道にしようと、たとえば「キーワード検索」で「アカマタ」「浜おり」と入力することで、必要なデータのほとんどを瞬時に入手することができるのです。

同じような作業が、「統合版・日本民話データベース」を利用すれば、沖縄だけではなく日本全国の現時点で4万話あまりの話について可能になります。

こうしたデータベース化のメリットは、2006年に岩倉千春さんの手で完成した「日本民話データベース」の試作版ができあがって、はじめて経験されたことでした。

2. 日本民話データベースと沖縄伝承話データベースの現状と課題

2-1. データベース化作業の現状

こうして「日本民話データベース作成委員会」は、2001年4月に作業を開始し、以後5年間にわたって日本学術振興会から2900万円程度の助成金を受けて、21000件の民話に関するデータを処理し、データベース化しました。この間、ホームページを立ち上げて、日本語のほかに、韓国語、中国語、英語、フランス語、スペイン語のサイトを用意し、アップロード可能なサンプル・データを発信してきました。(ホームページは、私たちの委員会の活動の一部を紹介する役割を果たしています。)

日本の場合は、学術振興会による助成は大体5年が上限ですから、私たちは2006年度は一度活動を休止し、その間にデータのプログラム・ミスなどをチェックしました。

(2005年度の研究助成終了時点での日本民話データベース作成委員会のメンバーは、岩倉千春、小野和子、剣持弘子、高津美保子、武田正、竹原威滋、立石憲利、田畑千秋、常光徹、花部英雄、牧ヶ野靖子、丸山顕徳、三原幸久、山下欣一、米屋陽一です。)

そして2007年4月から、今度は沖縄伝承話資料センターと協力して「沖縄伝承話データベース作成委員会」というもう一つの組織を立ち上げ、地域を沖縄に限って民話関連データの処理を行うこととし、同じく日本学術振興会から2007年度からの4年間の間に2340万円の補助を受けて20357件のデータを処理しました。2011年度の作業は始まったばかりですが、830万円の補助金によって最終的に27202件を処理する予定です。

「日本民話データベース」と「沖縄伝承話データベース」は、それぞれ独立していますが、まったく同じプラットフォームの上で作業が進められているので、現在は「日本民話

データベース」「沖縄伝承話データベース」「統合版・日本民話データベース」の3つが機能し、2012年4月現在には「統合版・日本民話データベース」を使用すれば40000件程度の日本民話関連資料を検索できます。

2-2. データベース化の技術的な問題点とその克服

この間、私たちが日本民話データベースに着手した時に直面した大きな問題は、2つありました。

2-2-1. データベースの公共性と個人の業績の保証

一つは、日本民話記録のデータベース化ということに対する理解の不足です。すでに述べたとおり、日本民話の音声記録は1960年代に始まり、以後40年の間に、少なくとも約30万件程度のデータが記録されたと推測されますが、それらのほとんど全ては研究者個人の努力によるもので、記録は個人によって所有されています。研究者個人は、自分の調査に個人的な愛着がありますから、日本民話データベースのような公共的性格の強い資料体に、情報を提供することに消極的です。

ですから、最初は記録テープの借り出しに時間がかかりました。日本民話データベース作成委員会は、当然のことながら、記録テープをデータ化する際に「当該テープの所有者が誰か」「調査した組織は何か」「それぞれの話を聴いた調査者は誰か」を明らかにし、データベース化された資料から、テープ所有者、調査組織、調査者等の名前が検索可能なものとし、それぞれの研究者の個人的な業績を保障しました。その上で、アナログのテープを借り出して、データベース化の作業を行い、作業が終わると、CD-ROMにデータを記録し、資料提供者にテープとともに返還しました。研究者の中には、コンピュータを利用しない人もいたので、オーディオCDも用意して、一般のステレオでも簡単に記録を確認できるようなサービスを行いました。

資料を所有する研究者たちは「今後、自分の調査した資料が、どう使われるか」に対して不安を抱いていましたから、次第に資料提供に対して肯定的になり、最終的には限られた予算では処理しきれないほど大量のデータが集まりました。

2-2-2. データベース入力作業者の養成とその業績の評価

二番目に直面した大きな問題は、作業者の確保です。当初は、学生のアルバイトなどを期待していましたが、最近の学生は、データ入力のような細かい作業を好みません。そこで、役に立ったのが民話に関心を持ち、民話の調査の経験を持つ主婦です。アナログテープをデジタルデータにするのは、簡単な機械的作業ではありません。極めて熟練のいる作業です。

まずアナログ録音のノイズを取り除きながらデジタルに変換する作業があります。これは比較的簡単な機械的な作業で、誰でもできます。

次に、このデータを1件ごとに切り分ける作業があります。これは、調査者によって作業手順が違うので、個別の対応が必要で、深い経験を必要とする、複雑な作業です。

この場合、一番簡単なのは、調査者自身が「これから何処で、誰の、何という話を聴きます。それでは始めます。」とって調査を始め、「これで記録を終わります」とってテ

テープを止める調査です。この場合は、調査者自身が一件毎にデータを整理して録音しているので、作業は簡単です。テープが貴重であった初期の調査では、この方法が多く採用されました。

しかし、次第に録音技術も進歩し、テープが安く大量に入手できるようになると、1時間半あるいは2時間の間テープが途切れなく録音されることが多くなります。民話の記録調査を経験した人なら誰でも理解できるように、一つの話が語られる合間の雑談の時間に、語り手が当該の話に関して貴重な情報を提供することが少なくないからです。調査の時には、いつ貴重な民俗情報が提供されるか予測できないわけですから、テープは調査の間中、途切れなく記録しておくのがよいのです。

こういう調査全体を記録したテープを整理するのは大変な作業です。調査者自身には、どこに重要な情報があるのかが分かっているからよいのですが、第三者が情報を整理するためには、まず全体を聴き、テープの何処に重要な情報があるかを判断し、テープの情報を1件ずつ切り分けなければなりません。場合によっては、話者が、テープのA面で語られた情報を、B面で補っていることもあります。そうするとAとBの情報を組み合わせて1件の情報を合成しなければならなくなります。こういう作業が日常茶飯の、当たり前の事になるのです。

したがって、テープ整理は「民話に対して深い理解と愛着」のある人でないと出来ない仕事になります。民話をよく知っていないと、何が大切な情報であるかを理解できないので、機械的に大切な情報を切り捨ててしまうことが起こります。テープの整理には、1件ごとに30項目ほどのデータを入力するという細かい作業が不可欠ですが、こうした細かい入力作業よりも、話を1件ごとの情報単位として切り分ける作業が重要です。

作業者は、最初はコンピュータの使用法から学び始め、テープの整理をするうちに少しずつ「大切な情報が何であるか、何処にあるか」を学んでいきます。「日本民話データベース」の場合は、だいたい日本民話の会に所属する主婦が作業を行い、「沖縄伝承話データベース」の場合には、かつて沖縄国際大学で学び、民話の調査を経験した主婦が、主として自分の調査したテープをデータベース化しています。いずれの場合も、コンピュータの作業手順を理解した後は、非常に質の高い仕事をしています。

こうした作業には、当然対価が支払われますが、日本民話データベースと沖縄伝承話データベースは、1件ごとの入力作業者の個人名も記号を利用して記録し、誰がこのデータベースの作成に関わったかもデータ化しています。そうすることによって作業者の責任意識とモチベーションが高まります。

3. データベースの紹介

3-1. 作成ソフト:ファイルメーカー

それでは次に「実際のデータベースで何ができるか」を具体的に紹介したいと思います。

データベースは、日本民話データベースも沖縄伝承話データベースもファイルメーカーというソフトを利用して作成されています。作成の担当者は、両データベースの作成委員会委員の岩倉千春さんで、コンピュータに弱い私には、このソフトが如何なるものかを技術的に、正確に説明することができません。本来であれば、私の代わりに岩倉さんにお話

を聞いた方がいいのですが、今回は、私が見られる範囲で説明させていただきます。

ファイルメーカーは、ウインドウズとマキントッシュの二つの OS でプログラムを作成できるソフトですが、作成者の岩倉さんがマキントッシュのユーザーなので、完全なプログラムはマキントッシュで組まれています。データベースは作成途中ですから、今回、私が持参したマキントッシュのノートには、2つのデータベース合わせて3700件ほどの検索しか出来ません。しかし、今回、みなさんに御覧いただくには十分かと存じます。ほかに、ウインドウズ上で起動する日本民話データベースの CD を持参しました。これは、サンプルとして口承文芸学会に差し上げますが、音声に連動していないので、話のデータを音として聴くことができないのが残念です。

音声データは、基本的には WAVE 形式で保存されていますが、ノートパソコンや Ipad を利用する場合には、MP3 の形式で保存されたデータが利用されています。

3-2. 検索項目 33

さて、このデータベースの検索項目は、次の 33 項目です。

- 1 レコード番号、2 CD 番号、3 決定題名、4 話者がつけた題名、5 話者名、6 話者名かな、7 生年月日、8 性別、9 調査時話者住所、10 出身地、11 記録日、12 記録場所、13 採訪者名、14 記録者の所属組織、15 元テープ番号、16 元テープ管理者、17 分類、18 大成名、19 大成番号、20 AT 番号、21 (通観番号)、22 発句、23 結句、24 あいづち、25 伝承事情、26 文字化資料、27 備考、28 写真、29 キーワード、30 梗概、31 話の容量、32 公開の了解、33 公開の了解者

このデータベースを開くと上図のようなページが開かれ、番号は記されていませんが、「語り手の名前」や「話の題名」など、項目別に記入事項が書かれています。たとえば、上の図の場合は、語り手は山形県米沢市の「平田幸一」さんで、話の題名は「桃太郎」、記録したのは「水野道子」さんで、1984年8月のことだったなどということが分かります。

これを新しく検索しなおして、たとえば「岡山」県の「桃太郎」について知りたければ、左図の決定題名に「桃太郎」、記録場所に「岡山」と書き込んで「実行」のボタンをクリックすればよいのです。

検索は、いくらでも複雑に行うことができます。たとえば>とい記号と1970年とを使用すれば、1970年代以前の調査で記録された「岡山」

の「桃太郎」などというデータも抽出できます。

キーワードに狐や虎を入れれば、話型に関わらず、虎や狐が登場する話のデータも抽出できます。梗概に虎や狐を入れれば、キーワードとして採用されなかった虎や狐の登場する話のデータをクリッカー一つで入手できるのです。

4. 入力作業の実際

次に、データベース作成の作業手順についてお話しします。

4-1. アナログデータのデジタル化

作業者は、まずアナログテープを Wave 形式でデジタル化します。変換ソフトは、なにを利用してかまいませんが、テープによってはノイズが大きいので、ノイズを除去する作業が必要になります。調査中にテレビの音や電話の音が録音されてしまうこともありますが、これを取り除くには専門的な知識が必要ですし、現在除去することができない場合も、将来、技術が進歩すれば簡単に処理することが出来るかもしれませんから、あまり神経質にノイズを除去する必要はないと思います。

これに対して、現時点では音声を Wave 形式で保存することは、絶対に必要です。Wave 形式で保存しておけば、これを MP3 あるいは MP4 に変換することは簡単ですし、オーディオ・ファイルに変換して、コンピュータ以外の一般のオーディオ・プレーヤーで聴くこともできます。まず、この作業を行い、オリジナルのテープを Wave に変換して、ハード・ディスクに保存すると同時に、CD-ROM にもコピーして保管します。

デジタル化されたデータをハード・ディスクと CD-ROM に二重に保管して管理するのは、デジタルデータが極めて破損しやすく、紛失の危険が大きいからです。幸い CD-ROM もハード・ディスクも安くなりましたし、とくにハード・ディスクには安全性の高いものが増えました。しかし、データは、かならず複数の媒体で保存する方がよいと思います。

4-2. データの切り分け

作業者は、次にオリジナルのテープに収められていた 90 分あるいは 120 分の情報を、話ごとに、1 件ずつ切り分けます。この作業は、たいへん熟練を必要とします。未熟な作業者は、何が大切な情報かを判断できず、機械的に作業を進め、重要な情報を削除してしまう可能性があるからです。

また、たとえ作業者が熟練していても、調査時の記録が不十分であったり、ノイズが大きくて、十分に聞き取れなかったりして、当該の記録を保存すべきか否か、判断に迷う場合もあります。日本民話データベースと沖縄伝承話データベースの場合は、不十分と思われるデータも、必ず一件として保存することにしました。

とくに、話者が話を途中で忘れてしまったり、語り間違えたり、訂正したり、データとして不完全な場合は、必ず保存するようにしました。

以上のように、データの切り分け作業には、さまざまの問題が発生します。したがって、多くの作業者が関わる場合には、最初によく話し合い、作業に共通理解を徹底することが必要です。そして、作業者が判断に迷った時には、熟練した作業者に相談できるような連絡体制を作る必要があります。

日本民話データベースの場合は、幸い初期の作業者は少数で、難しい問題に直面するたびに作業者同士が相談しあうことが容易でした。(これは、自動車工場で見場の作業員達が、作業の手順を教えあうトヨタの「カイゼン」のやり方に似ています。プロジェクト全体を指揮する人たちの指令も大切ですが、具体的な問題に直面する作業者同士のチームワークと工夫の努力の方が、さらに大切です。)

2007年3月に沖縄伝承話データベースの作成作業が始まった時には、すでに日本各地に5年間の作業経験をもつ熟練者がいましたので、沖縄の作業者にその Know-How を全て伝えることができました。沖縄でも、作業を理解した少数の作業者が、初心者に作業方法を伝授するという方法で問題を解決し、現在では沖縄だけで20名ほどの熟練作業者がいます。

韓国の場合は、作業にあたるのは学生だと思いますので、まずマニュアルを作り、いくつかの作業グループにわけて、作業責任者を設け、作業グループ内で作業者同士が緊密に連絡をとる体制を作られることをお勧めします。

4-3. エクセル入力

テープの切り分けが出来たら、つぎはエクセルを利用して、6ページで紹介した33項目の情報を入力します。

右図の場合は、エクセルの空欄に、全ての項目が入力されていますが、これは例外です。資料によっては、話者名も調査者名も、調査地や調査日時も不明なものがあります。しかし、資料はどんな資料でも大切です。それに、内容さえ確認できれば分類は可能ですし、使用された方言などから地域を特定できるかもしれません。ですから、エクセルの入力には、空欄があってもかまいません。

4-4. 完成データの提出とデータベースの作成

以上の作業を終えた作業者は、①オリジナルテープ、②オリジナルテープをデジタルに変換した CD-ROM (Wave) 1枚、③データごとに切り分けた「完成 CD-ROM」(Wave) 2枚、④エクセルに整理されたデータ、という4つをセットにして管理責任者に提出します。

これを受け取った管理責任者は、エクセルのデータと「完成 CD-ROM」1枚をプログラム作成責任者に提出し、プログラム作成責任者は、ハード・ディスクにエクセル・データと「完成 CD-ROM」の内容を保存し、ファイルメーカー等のソフトを利用して、データベースを作成します。その際、「完成 CD-ROM」の Wave を MP3 などの容量の軽いデータに変換して、ノートパソコンなどで簡単に利用できる「軽量バージョン」のプログラムも作成します。

以上の手順でデータベースの作成が終わっても、データ管理の問題が残ります。

4-5. データの保管と管理

すでに述べたように、デジタルのデータは、簡単にコピーできますが、たいへん破損しやすい性格をもっています。CD-ROMにも耐用年数がありますし、ハードディスクは数年の寿命しかなく、ある日、突然壊れます。そのために、そのために日本民話データベース作成委員会では、データをCD-ROMとハードディスクという2つのメディアを使って保管し、さらに①作業員、②管理責任者、③プログラム作成者の三者が、それぞれのデータを責任を持って保管します。とくに、作業員が作業データを保管することが大切です。というのは、場合によっては、作業に使用した機材やソフトが故障していて音声が十分にデジタル化されていなかったりして、プログラム作成段階で不具合が起こる可能性があります。そのような場合は、もう一度、作業員が自分の仕事をチェックして、資料を作成しなおす必要があります。

データの管理責任者が、データを保存しなければいけないのは当然のことです。幸いなことに、コンピュータのハードディスクは加速度的に進化し、容量も大きく、管理も容易になってきています。しかし、事故によるデータの喪失は、何時、何処で起こるかわかりませんから、複数の人たちが、複数の場所で、複数のやり方で保管するのがよいと思います。また、データベース化が終了しても、アナログ・テープの保管は必要です。デジタル・データに比べて、アナログ・テープの方が耐用年数が長く、安定しています。

5. 日本民話データベースの問題点

5-1. 地域的な偏り

すでに述べたように、日本における民話調査は、そのほとんど全てが個人の研究者によるものです。したがって、地域による偏りが多いのです。たとえば、日本民話データベース作成委員会のメンバーの立石憲利さんは、きわめて優れた民話研究者で50年にわたって調査を続け、200冊の著書を刊行しました。そのほとんどが民話資料集です。彼のフィールドは岡山県ですから、岡山周辺の資料は実に豊富です。また、宮城県の小野和子さんは、宮城民話の会を組織して、仙台を中心に現在も調査活動を行っています。さらに沖縄の遠藤庄治さんが沖縄全域の調査を調査し、すぐれた資料を残したことは、すでに述べた通りです。しかし、これらは例外で、たとえば近畿地方などは、ほとんど手つかずの状態です。データベースの空白地域が多いのです。

しかも、民俗の分布は、複雑で、山や川を一つ隔てれば、まったく言葉や習慣が違うことが一般です。Aという地域を集中的に調査しても、すぐ隣のBという地域には、まったく違った民俗がある可能性があります。したがって、どれほど調査しても十分ということはないのですが、日本民話データベースの場合には「基本的な資料に地域的な偏りがあるという欠陥がある」という批判は免れえません。

5-2. 作業員による入力作業の偏り

これも既に述べたことですが、複数の作業員が「データの切り分け」と「エクセル入力（情報入力）」に携わります。そこには、当然のことながら個人の能力や嗜好の差が表れます。たとえば、「キーワード」として何を選ぶか、「梗概（要約）」にどんな基準を設けるか、などには、個人差が生まれます。基本的な「話型」の決定に関してすら、作業員の能力や

個人的な判断が影響します。

したがって最終的に完成したデータベースにも、多くのエラーが発生します。データベースのプログラムを作成し、作業者の作成したエクセルと音声のデータを次々と加えていく「プログラム作成責任者」は、当然、その作成作業で個々のデータの不備や不足をチェックして、自ら修正したり、作業者にフィードバックして作業をやり直させたりするのですが、それにも限界があります。

データベース作成の作業は、膨大な労力を必要とする作業ですから、ある場合には「ある程度の不足には目をつむる」ことが必要な場合もあります。

6. データベースの問題点を、どう解決するか

つぎに「以上二つの問題をどう解決するか」についてお話しします。いずれの場合も、決定的な解決策はないのですが、デジタル・データベースは、①技術的にさらに進歩し、その可能性は、ほぼ無限であること、②データベースの仕様は、いかようにも変更可能であること、という2つの特性をもつので、基本の設計を誤らなければ、かならず「よりよい解決」が見つかると考えられます。

6-1. 地域的な偏りを、どう補うか

上記のように、日本民話データベースの地域的なデータの偏りは致命的なものです。今後、学術振興会の援助をどこまで得られるか明らかではありませんが、まず為すべきことは、資料空白地帯において行われた過去の調査資料（アナログ・テープ）を借り受けて、データベース化を推進することです。日本民話データベースがまだ処理していないアナログ資料は、少なくとも100000件はあると考えられます。沖縄だけでも40000件の処理が残されています。

しかし、次にまず着手すべきことは、音声資料だけではなく過去の文字資料をPDF等のファイルにして、統合版日本民話データベースに付け加えることだと思われます。文字化された資料は、音声資料よりも長期にわたって保存が可能ですから、時間をかけて作業を進めることができます。とくに昨今のように電子図書が流通し、活字のデジタル化が進めば、データベース化は容易になるでしょう。

統合版日本民話データベースは、ファジーな構造を持っていますから、音声情報と文字情報、さらには映像や動画の情報を付け加えることが、簡単にできます。問題は、コンピュータのCPUの計算速度とハードディスクの容量ですが、これは、例えば、現在のウインドウズ7のレベルでも十分に可能です。

日本民話データベース作成委員会は、できるかぎり地域的な偏りを排して、全国にわたる音声情報を入力したいと考えますが、予算や時間の不足によって、今すぐこれが達成できない場合にも、最終的には十分な結果を得られると考えています。

6-2. 作業による入力作業の偏りを、どうするか

作業による入力作業の偏りの問題に関しても、あまり大きな問題はないと考えています。それは、やはり日本民話データベースのファジーな構造によります。たとえば、キー

ワードや話型の入力ミスは深刻な問題です。しかし、それによってプログラム全体が動作を止めてしまうことはないのです。

こうした誤りは、いつでも修正できますし、不足は追加できます。デジタル・データベースの強みは、常に進化しているということです。最初から「完全なデータベース」を作る必要はないのです。たくさんの人が、一定のルールに従ってデータを自由に使用し、一定のルールに従って、誤りを指摘し、修正してゆけばよいのです。

誤りの訂正は、それほど難しくはありません。大切なのは、まずデータベースを正しく設計すること、つぎは使用上のルールを正しく決めること、最後は決められたルールに従ってデータベースを正しく管理することです。

7. 民話データベースの国際協力

7-1. 韓国の民話記録

7-1-1. 1945年の光復までの民話記録

韓国における民話の記録は1920年代にはじまり、1926年に沈宜麟の『朝鮮童話大集』、1927年に鄭寅燮の『温突夜話』、1930年に孫晋泰の『朝鮮民譚集』が刊行されます。とくに孫晋泰の『朝鮮民譚集』は20年代初頭からの記録で、語り手や記録日時が正確に記された極めて質の高い記録です。その後、秋葉隆らによって、朝鮮半島全域にわたる文化人類学的な調査が開始されると、彼等と行動をともにした任哲宰の調査記録が開始されます。任哲宰の記録は、1987年から93年にかけて『任哲宰全集・韓国口傳説話』全12巻として刊行され半島全域にわたる貴重な民話記録となりました。(任哲宰は、音声録音の普及した1960年代以降は、韓国全土の民謡調査を行い、その記録はこれから紹介する韓国学中央研究院に寄贈され、データベース化されています。)

7-1-2. 1945年の光復以降の民話記録

日本の植民地支配が終わった1945年以降は、韓国独自の民俗調査が行われますが、その一方で南北分断という状況下で、私たちの目にすることのできる記録の大半は、朝鮮半島(韓半島)南部のみのもとなります。

1945年以降の韓国の民俗記録の大きな特徴は、大規模な調査が国家事業として行われたことです。その第一次調査は、1970年代初頭に文化広報部が韓国文化人類学会に委嘱しておこなったもので、韓国の全道を対象に、各年度ごとに一道の調査を集中的に行うという組織的なものでした。この調査の結果は竹田旦先生と任東権先生のご努力で翻訳され『韓国の民俗大系』という形で、大部分の成果を日本語で読むことができます。そこに「口碑伝承」という形で納められた記録は大きく民謡・説話・謎と諺・方言などに分類され、民話は「説話」として分類され、その扱いは各道ごとにばらつきがありますが、貴重な記録であるといえます。

韓国政府主導で行われた第二次の調査は、1978年から1984年まで韓国精神文化研究院(現・韓国学中央研究所)が、全国60の市町村(市・郡・面)で行った「全国口碑文学調

査研究」で、その成果は『韓国口碑文学大系』全 82 巻に収められています。この調査は、やはり説話と民謡という 2 つのジャンルに分類されていますが、当初から「口碑文学調査方法」という厳密な調査マニュアルが用意され、その成果である 15107 話の説話 (= 民話) は、この調査を主導した趙東一によって話型分類されています。(説話と同時に調査された民謡 6,187、巫歌 376、その他資料 21 も『民謡巫歌類型分類集』によって分類整理されています。)

韓国の政府主導で行われた第三次の調査は、2008 年から 10 年の計画ではじまり、現在も進行中のプロジェクトです。この調査は前回の韓国精神文化研究院の調査を引き継ぐ形で韓国学中央研究所が主導しているもので、前回調査の欠落部(未調査地域)を補足して、全国すべての地域の記録を完成し、同時にその成果を全てデータベース化するという計画です。

こうした国家的な調査のほかにも、もちろん個人の調査や研究も存在します。

そのもっとも顕著な業績は、崔仁鶴の仕事であろうと思われます。崔仁鶴は、1960 年代に日本に留学し、関敬吾の『日本昔話集成・大成』に匹敵するタイプインデクスを刊行することによって、韓国昔話の国際比較研究を可能にしました。

崔仁鶴は、自身も韓国各地で聞き取り調査を行い、その成果を日本語でも公表しているので、日本の読者にも親しみ深い存在です。また最近では、李秀子、黄仁徳のように 400 話、500 話クラスの話者の記録を刊行する研究者も出現しました。

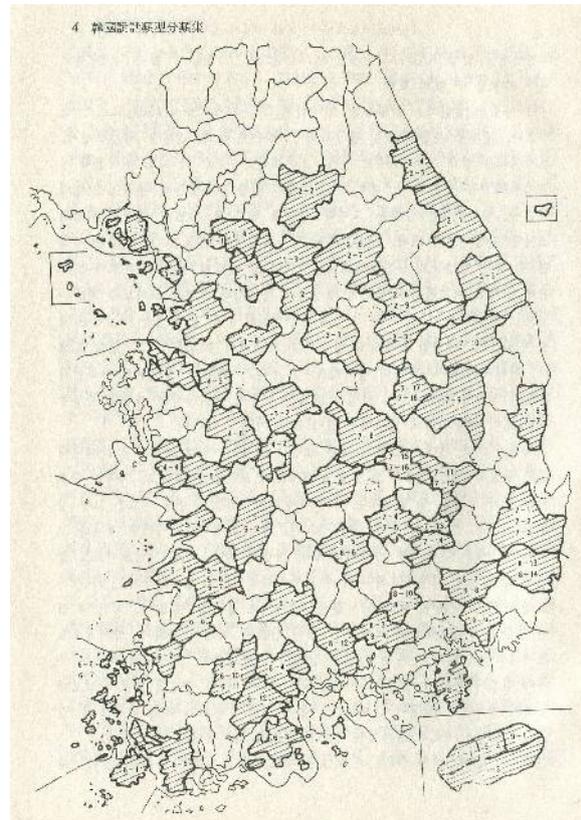
7-2. 韓国のデータベース化の現状

7-2-1. 韓国と日本の民話記録の性格の違い

1960 年代以降、音声で記録された民話記録の資料をデータベース化する上で、日本と韓国が世界で有数の可能性を保持していると考えられます。この 2 つの国が保有する民話記録の質と量は、現時点では、世界のどこの国にも比較できない水準にあると思われます。

もちろん、文化人類学などのフィールドワークを通じて、自らが調査対処とする人々の語りの記録を量的にも質的にも高い水準で保有している研究者の数は、少なくないと思われます。しかし日本や韓国のようなタイプの近代国家のレベルで、これだけの質量を確保することは、今後も難しいのではないのでしょうか。

そして、きわめて興味深いのは、この 2 つの国の保持するデータの性格がきわめて異なるということです。すでに述べたことから容易に想像がつくことですが、韓国のデータは、



国家的なプロジェクトによって得られたものがほとんどであるのに対して、日本にはこのような国家プロジェクトが皆無に近く、ほとんど全ての記録が個人に帰属しているということです。

こうしたデータの性格の違いは、両国のデータベースの性格の違いに反映されます。

韓国のデータベースは、短期集中のローラー作戦によって生み出されますから、今回の第三次プロジェクトが終了すれば、全国の津々浦々を覆い尽くすことになり、地図上の空白地帯は消滅します。日本の場合には、大東島以外のすべてを調査し尽くした沖縄国際大学の調査をのぞけば、このようなタイプの調査は存在しません。後で述べるように、日本民話データベースの地域的な偏りと空白は、目を覆うばかりです。

しかし、その一方で日本の民話調査は、個人が時間をかけて、同じ調査地を何度も訪れて、丁寧に話を聞きます。その結果、100話を越す優れた語り手の数は、韓国に比べて圧倒的に多いと思われます。誤解のないように申し上げますが、韓国のデータベースに納められた話を聞いてみると、一人ひとりの語りの水準はとても高いし、話を聞く調査者の姿勢も実にしっかりしていると思われます。しかし、ある限られた時間内に、ある地域をくまなく調査する使命を与えられた調査者には、限界があるのではないのでしょうか。

日本の場合は、沖縄国際大学の場合も、じつに丁寧な調査を行い、同一地域の同一の語り手を数次にわたって訪れることによって、多くのすぐれた語り手を発見していることを申し添えます。

7-2-2. 韓国におけるデータベース化とその公開状況

つぎに韓国におけるデータベース化の進行状況についてお話しします。

韓国における民話データベース化の作業は、2000年の半ばから始められた「韓国歴史情報統合システム（한국역사정보통합시스템）」の一環として着手されました。これは、韓国学中央研究所に付属する「蔵書閣」という図書館の所蔵する古図書、古文書、写真・



絵画資料、音声資料などをオンラインで公開する事業で、<http://yoksa.aks.ac.kr/> というURLを入力すれば誰でも資料にアクセスできます。このページの音声資料（음성 자료）という中央のタグをクリックすると、誰でも①韓国口碑文学大系、②韓国方言資料集、③韓国民謡大観という3つの音声資料を利用できるのです。

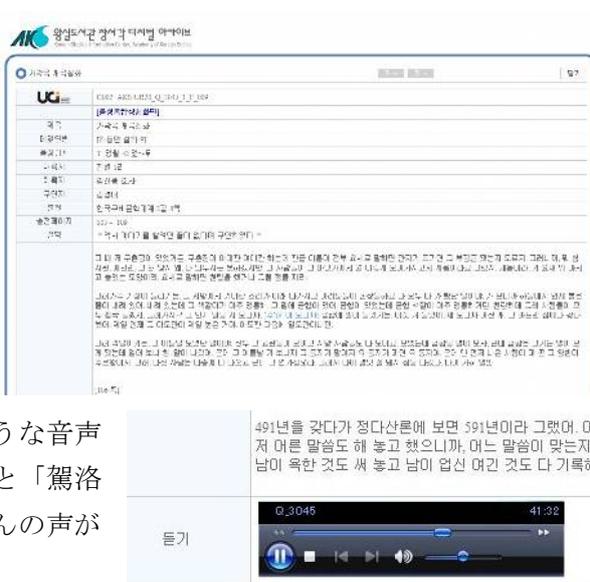
この3つの資料は、それぞれ大変優れたものですが、ここでは「韓国口碑文学大系」について紹介します。

このデータベースには、先に紹介した1978年から1984年までに韓国精神文化研究院（現・韓国学中央研究所）が行った調査データ（説話15107、民謡6,187、巫歌376、その他資料

ここには、いくつかのタグがありますが、一番左の<テキスト・テキスト>というタグをクリックすると、前頁のようなページが表示されます。そこには、話者、記録者、記録地、テープ番号、記録日時などの情報が記されています。

そしてなにより便利なのは、「口碑文学大系」に収録されたテキストが音声とともに収められていることです。

ページを下までスクロールすると、右のような音声用のアイコンがあり、これをスタートさせると「駕洛国（金官伽耶）開国説話」を語る金ヨンデさんの声が流れてきます。



8. 韓国・東アジアとの共同作業の課題: AT というスタンダードの尊重

ここまで、日本と韓国の民話調査とデータベース化の現状について紹介してきましたが、ここで「東アジア、とくに日本と韓国、そして中国の間で、民話の資料を共有し、データベース化を推進するにはどうしたらよいか」という見通しについて、私の考えを述べさせていただきます。それは、難しいことではありません。

8-1. 国際比較の視点の必要

韓国であれ、日本であれ、自国の民話のデータベースを作成しようとする場合に大切なのは、まず自分たちの国の文化遺産である「民話」を記録し、保存し、研究や教育に役立てようということでしょう。とくに1960年代以降に記録されたアナログ・データは、時間とともに消滅する運命にあります。この時代の記録をデジタル化することによって半永久的に保存し、データベース化することによって実用に耐えるものにするのは、現代の民俗学研究者の責務です。

しかし、誰でもすぐに気がつくことですが、民話は世界的な広がりを持っており、簡単に国境を越えてしまいます。たとえば、「シンデレラ」はヨーロッパやアメリカで人気の高い話で、子どもたちなら誰でもディズニーのアニメは知っているし、ディズニー・ランドのシンデレラ城が大好きです。しかしこの話の一番古い記録は中国の唐時代の「酉陽雜俎」にあります。また、韓国には「コンジ・パッジ」(KT450)があり、日本には「米福・粟福(こめふく・あわふく)」(ST205)があります。

民話には、同じような問題が山のようにあります。韓国民話の特性について理解しようとする時には、中国や日本だけでなく、ヨーロッパやアメリカやアフリカや中近東やロシアやオーストラリアのことを考えないわけにはいかないのです。自分たちの国の伝統文化の独自性を理解しようとするならば、東アジアだけでなく全世界に広がる民話・民俗のネットワークと連携する国際比較の視点が必要です。

8-2. 日本のタイプ・インデクス

日本の民俗の研究に最初に目をむけた柳田國男は、当時のヨーロッパ民俗学研究から多くを学んだ人でしたが、こうした国際比較の立場に対して、懐疑的でした。私は、これを大変残念なことだと考えています。

日本の民話・民俗の国際比較に最初に着手したのは、関敬吾です。関は、フィンランドのカールレ・クローン (Kaarle Leopold Krohn 1863-1933) やアンチ・アアルネ (Antti Amatus Aarne 1867-1925) の民話や民俗の研究法に学び、まず『日本昔話集成』(1950-1958)、つづいて『日本昔話大成』(1978-1980) という日本民話のタイプ・インデクスを作成しました。このカタログが優れているのは、日本民話の地域的な特性に配慮しながら、国際比較の視点を忘れなかったことです。関敬吾は、韓国の民話研究者、崔仁鶴氏とも大変親しかつたので、『日本昔話大成』には、とくに韓国民話に対する言及が多いのも大きな特徴です。

日本では、関敬吾のタイプ・インデクスの後に、小沢俊夫と稲田浩二が中心になって『日本昔話通観』(1977-89) という日本民話の地域性に配慮したタイプ・インデクスが作成されました。

日本民話データベースと沖縄伝承話データベースは、『日本昔話大成』のタイプ・インデクスを基本としていますが、『日本昔話通観』も参照しています。そして、国際比較のためには、次に述べるアアルネ・トンプソンのタイプ・インデクスを参照しています。

8-3. 「de facto standard」としてのアアルネ＝トンプソンのカタログ

民話の国際比較という考え方は、おそらくグリム兄弟に始まると考えられます。しかし、それを実行に移したのはイギリスのマリアン・コックス (Marian Roalfe Cox 1860-1916) の *Cinderella: Three Hundred and Forty-Five Variants of Cinderella, Catskin and, Cap O' Rushes* (1897) であろうと思われます。彼女は、ヨーロッパを中心としたシンデレラの類話を 345 話あつめて 5 つのタイプに分類したのです。しかし、さらに世界中の民話を集めて、それらを話型 (Tale-Type) に分類し、カタログ化したのはフィンランドのアンチ・アアルネでした。彼が 1910 年に刊行した『メルヘンの型目録 *Verzeichnis der Märchentypen* (FFC3)』は、わずか 63 ページにすぎませんが、すでに動物昔話 Animal Tales、本格昔話 Ordinary Tales、笑話 Anecdotes and Jokes という基本分類のもとに 2000 の話型が用意されています。このカタログは、後にアメリカのステイス・トンプソン (Stith Thompson) によって 2 度にわたって改定され (*The types of the folktale. a classification and bibliography* FFC 74 & FFC 184)、さらに形式譚 Formula Tales、分類できない話 Unclassified Tales という 500 の話型が追加されましたが、トンプソンのカタログの基本構造は、アアルネと変わっていません。

ステイス・トンプソンは、アメリカ先住民の民話の専門家でしたから、アアルネの作成したカタログでは分類できない話が沢山あることを誰よりもよく理解していたはずですが、しかし彼は、アアルネの分類を忠実に踏襲して、アアルネの分類方

法に従って話型を補い、600 ページほどのカタログを作成したのです。

私は、アアルネのカタログをいわば「de facto standard」として尊重したトンプソンの仕事を尊敬しています。彼は、一方でアメリカというヨーロッパとはまったく違った歴史と民俗をもつ地域の伝承を研究し、民話の地域性を尊重しながら、一方で世界中の民話を比較するための道筋を開いたのです。今日では、アアルネとトンプソンのタイプ・インデクスに付された AT という話型なしには、民話の国際比較を行うことは不可能です。

8-4. 東アジアの民話データベースをどう構築するか;分類の問題

日本に柳田國男、関敬吾、小沢俊夫＝稲田浩二などの、それぞれ特徴ある民話の分類案があるように、韓国には崔仁鶴、曹喜雄、趙東一などの優れた分類案があります。中国にも、丁乃通、金榮華などの分類があります。

私は、民話データベースを作成するにあたっては、それぞれの国の、それぞれの事情に従って、複数の分類案を並行して採用すればいいと考えています。しかし、どうしても忘れて欲しくないのは、国際比較の視点です。さまざまに批判のあることは承知の上ですが、すでに国際比較の上で「de facto standard」として確立しているアアルネ・トンプソンの分類 (AT) を尊重し、必ず検索項目に付け加えることが大切です。

そして、ことに韓国の場合は、国際比較と AT 分類に当初から配慮した崔仁鶴の話型分類 (KT) が存在しますから、この分類だけは必ず検索項目に付け加えて欲しいと考えています。

韓国の人たちが、韓国の民話を韓国固有の仕方で分類するのは、当然のことです。韓国の民話データベースは、まず韓国人の研究者や教育の現場で生かされなければいけませんから、そのために韓国人や韓国文化の専門家だけが利用できる検索項目を、いかに詳細に構築されても構いません。しかし、その一方で、韓国語が十分に理解できなくても、韓国の文化や伝統に関心のある私のような外国人の研究者や一般人が、誰でも利用できることが大切だと思います。

たとえば日本の場合には、すでに多くの韓国民話に関する資料を日本語で利用できる体制が整っていますから、すでに多くの人が韓国民話に親しみを持っています。英語圏にも韓国民話の資料は少なくありません。ですから、ぜひ、民話の国際比較という視点を配慮していただければと考えます。

将来、インターネットを通じて、韓国の民話を世界に発信される場合にも、必ず AT の参照が必要になるでしょう。

8-5. 東アジアの民話データベースをどう構築するか 2;そのほかの検索項目

日本民話データベースと沖縄伝承話データベースの検索項目は、現在 33 あります。しかし、このシステムをさらに進化させる為に、たとえば文字情報とか動画情報とか、さまざまの検索項目を付け加えることは難しくありません。

それは、このデータベースの基本構造が、エクセルによって組み立てられているからです。現在、このデータベースにおいて最も大切なのは音声情報ですが、これに文字情報や

動画情報を追加したり、新しい民話の分類案を追加したりすることは、技術的には難しくありません。データベースの構造は、非常にシンプルなのです。

ここで、もう一度、33の項目を列挙してみましょう。

1レコード番号、2 CD 番号、3 決定題名、4 話者がつけた題名、5 話者名、6 話者名かな、7 生年月日、8 性別、9 調査時話者住所、10 出身地、11 記録日、12 記録場所、13 探訪者名、14 記録者の所属組織、15 元テープ番号、16 元テープ管理者、17 分類、18 大成名、19 大成番号、20 AT 番号、21 (通観番号)、22 発句、23 結句、24 あいづち、25 伝承事情、26 文字化資料、27 備考、28 写真、29 キーワード、30 梗概、31 話の容量、32 公開の了解、33 公開の了解者

このうち、「3 決定題名、4 話者がつけた題名、5 話者名、7 生年月日、8 性別、9 調査時話者住所、10 (話者の) 出身地、11 記録日、12 記録場所、13 探訪者名、14 記録者の所属組織、17 分類、18 大成名 (KT 話型名)、19 大成番号 (KT 番号)、20 AT 番号、22 発句、23 結句、24 あいづち、25 伝承事情、26 文字化資料、27 備考、28 写真、29 キーワード、30 梗概」などは、番号はともかく、韓国でも日本でも大切な情報であろうと思われます。(「17 分類」という項目については、これまで言及しませんでした。日本の記録テープには、昔話のほかに伝説、神話、世間話、民謡、俗信、生活上の経験譚など、さまざまなジャンルの話が登場します。それを全て切り捨てずにデータ化しますから、その分類が必要になるわけです。韓国の場合にも、おそらく調査テープには、さまざまなジャンルの口頭伝承が含まれると思いますから、とりあえずは「17 分類」のような形で整理しておくとうまいと思います。)

もし、今後のデータベース作成上で韓日両国の共同作業が始まれば、双方で、これらの項目の一つ一つに関して検討会を設ければいいと考えます。

9. まとめ

民話データベースを作成する上での問題点を具体的にお話ししたために、たいへん長い話になって恐縮です。

しかし、私の話の内容は極めて短く、次の7つに要約出来ます。

- ① 音声データの保存には、Wave 形式を用いること。
- ② 音声データの1件ごとの切り分けには、熟練が要るので、チームを作り、責任者を決めて、慎重に行うこと。
- ③ 作成チームの責任者は、メンバーの作成したデータを十分にチェックし、メンバーに対する指示を徹底すること、
- ④ データの整理には、エクセルを用い、共通のフォームに従って記入すること、
- ⑤ データの記入項目は、最初に時間をかけて、十分に準備すること。(後で項目を追加することも出来るが、手間がかかるので、最初にしっかり決める方がよい)
- ⑥ 話型分類に、国際比較の視点を必ず取り入れること。
- ⑦ 以上のように、周到な努力を重ねても、「完全無欠なデータベース」は絶対にできない。しかし、データの追加や修正は、いつでも可能だから、失敗を恐れずに、まず作成に着手すること。

以上です。

(本稿は、2011年6月の第35回日本口承文芸学会大会における講演をもとに執筆されたものである。)