

# 民話データベースの現状と国際協力

(これは、2010年10月29日に韓国の韓国学中央研究院で行われた「韓日民話データベース標準化問題シンポジウム」での報告に加筆したものです。)

## 1. 日本における民話データベース作業の現状

私たち「日本民話データベース作成委員会」は、2001年4月から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、データベース化してきました。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。

そして2007年4月から、今度は「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て、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3年間の間に1710万円の補助を受けて14687件のデータを処理しました。2010年度の作業は進行中ですが、630万円の補助金によって5400件を処理する予定です。学術振興会の助成は、これまで順調に推移しているので、2011年度の助成も期待され、5年間の総計で25000件程度のデータをデータベース化することが期待されます。

「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つが機能していて、2010年10月現在では「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. 日本民話データベースと沖縄伝承話データベースの技術的な問題点とその克服

2010年に、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。

### 2-1. 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。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も、実は「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. 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、日本の学生は、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。」と調べて調査を始め、「これで記録を終わります」として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。こういうことが日常茶飯の、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。また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も起こります。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

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、私に分か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。今回、私が持参したマキントッシュのノートには、実は 1 年以上前のデータしか入っていないので、2つのデータベース合わせて 29000 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 Ipod 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、図1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。

### 3-3. データベースのアナログデータに対する優位性

こうした検索機能を備えたデータベースは、アナログのデータに比べて圧倒的に便利で、研究者をはじめ、地域の図書館、博物館、公民館、学校などに研究や教育の素材を容易に提供する可能性を示しています。

その最も分かりやすい例が、従来の沖縄伝承話に関するアナログ資料と「沖縄伝承話データベース」の違いです。沖縄の場合、1960年代後半から、沖縄国際大学に赴任した遠藤庄治さんと学生達の努力下、遠藤さんが他界された2006年の時点でテープにして約4500本、話数約70000件の沖縄全域にわたる民話のアナログ記録がありました。これは、日本で行われた最も完璧な調査で、沖縄県のほぼ全域を網羅し、テープごとにカード整理され、保管され、沖縄各地の資料集（文字化資料）の作成にも利用されてきました。

しかし、この資料を現在、有効に活用することは、ほとんど不可能です。

たとえば、沖縄全域に広く分布する「天人女房」(AT400、Seki118、KT205)について考えてみましょう。アナログテープを利用してこの話型の研究を試みる場合には、膨大な調査カードを検索し、4500本のテープとの照合を行い、さらにテープの保管されている資料室まで出向いて、ケースに納められた複数のテープを最初から一つ一つ聴きなおさなければなりません。

それに対して、統合版・日本民話データベースを使用すれば、沖縄だけでなくデータベースに納められた全てのデータを、瞬時に、話型を含めた複数の条件で検索し、それぞれのデータを検出し、さらにその音声資料を聞くことが出来るのです。

しかも、このデータベースは公共のものですから、地域の図書館が所蔵すれば、日本全

国どこからでも利用できます。アナログテープを保管する沖縄まで出張し、伝承話資料センターや県立博物館まで行き、使用許可を申請する必要はないのです。

データベースは、小さなノートパソコンに収納できますから、一定の運用規則にさえ従えば、自宅に帯出することも可能です。

さらに、インターネットに公開された資料であれば、全世界から瞬時にアクセスできます。

## 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。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 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、さまざまな問題が発生します。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、自動車工場で現場の作業員達が、

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しますので、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、以下の33項目の情報を入力します。

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 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

	A	B	C
1	レコード番号	04002021	04002022
2	CD番号	04002019	04002019
3	決定題名	金の卵	赤次おひなまつ
4	話者がつけた題名	金の卵の話	赤次おひなまつ
5	話者名	水浦誠吾	水浦誠吾
6	話者名かな	ながのせいき	ながのせいき
7	生年月日	19090417	19090417
8	性別	男	男
9	調査時話者住所	宮城県登米郡南方町香島屋敷72	宮城県登米郡南方町香島屋敷72
10	出身地	宮城県登米郡南方町香島屋敷72	宮城県登米郡南方町香島屋敷72
11	記録日	2000110	2000110
12	記録場所	宮城県登米郡南方町ハの森40-1 南方町公民館	宮城県登米郡南方町ハの森40-1 南方町公民館
13	探訪者名	小野和子・山田和郎・山田裕子・山田伸司	小野和子・山田和郎・山田裕子・山田伸司
14	記録者の所属組織	みやぎ民謡の会	みやぎ民謡の会
15	元テープ番号	67A	67A
16	元テープ管理者	小野和子	小野和子
17	分類	12	12
18	大成名	金の卵	民謡 龍吹き盤
19	大成番号	226	211 119
20	AT番号	AT729	AT510
21	(通観番号)	52金の卵	1011655 223月台篇
22	発句	おひな	おひな
23	結句	舞	舞
24	あいづち	うん んで	うん んで
25	伝承事情	指母よふさんかみ、面伊裏橋かまじの焚き口などに座って袖と袖に聞く。	指母よふさんかみ、面伊裏橋かまじの焚き口などに座って
26	文字化資料	水浦誠吾語り みやぎ民謡の会編みやぎ民謡の会調査集 香島屋敷名物伝説 中巻みやぎ民謡の会 2001p.66-67	水浦誠吾語り みやぎ民謡の会編みやぎ民謡の会調査集 香島屋敷名物伝説 中巻みやぎ民謡の会 2001p.66-73
27	備考	話者水浦誠吾さんは2002年9月7日逝去。誠吾さんは昔話を定型句で語り始めるのを好むが、ここでは変形の成れで結句を欠いている。	話者水浦誠吾さんは2002年9月7日逝去。誠吾さんは昔話を定型句で語り始めるのを好むが、ここでは変形の成れで結句を欠いている。
28	写真	有	有
29	キーワード	水浦の節 湯釜 終の魂と金の卵 コイ 膝の節伝説 水浦せい	長崎屋敷門 長崎屋敷の節 打節の節 湯釜 湯釜

上図の場合は、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。

4-6. p p

### 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。

幸い、韓国の場合には韓国学中央研究所の行った組織的な調査がありますから問題は少ないと思いますが、そこにもやはり限界はあるでしょう。かつての民俗の分布は、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. 作業者による入力作業の偏り

これも既に述べたことですが、複数の作業者が「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、作業者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瞑る」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、現在のウィンドウズ7のレベルでも十分に可能です。

韓国で民話のデータベースを作成する場合にも、基本設計に、音声のほかに文字、画像、動画などの情報を付け加えることは難しくありません。

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. 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

最後に、「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」についての考えを述べたいと思います。それは、難しいことではありません。

### 7-1. 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「西陽雜俎」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。

### 7-2. 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・インデッ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ったので、『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・インデックスの後に、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・インデッ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・インデッ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・インデックスを参照しています。

### 7-3. 「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ィス・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。

スティス・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・インデックスに付された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。

### 7-4. 東アジアの民話データベースをどう構築するか1 ; 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。

#### 7-5. 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、さまざまな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 レポート番号、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 採訪者名、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よいと思います。)

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。

## 8. 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。