

# 国立国語研究所所蔵資料アクセス環境改善への取り組み

山口亮・関川雅彦（国立国語研究所プロジェクト非常勤研究員）

国立国語研究所（以下国語研）は、1948（昭和23）年12月の開所以来70年近くに渡って方言や語彙、音声、日本語教育、言語コーパスといった多様な研究プロジェクトを積み重ねてきた。これら研究調査の中で使われた紙・カード資料をはじめ、視聴覚データ・電子データも多数蓄積されている（紙資料は資料群ごとに、メディアは媒体ごとに）。研究情報発信センター研究資料室では、これら資料の整備及び公開に向けての作業を続けている。本稿は、2016年度の取り組みについて述べていくものである。

## An improvement of access to research materials in the National Institute for Japanese Language and Linguistics(NINJAL)

Ryo Yamaguchi / Masahiko Sekikawa (Temporary Researcher, NINJAL)

Since 1948, the National Institute for Japanese Language and Linguistics(NINJAL) has been conducting various research projects related to Japanese language, linguistic and corpus. Further, NINJAL has been collecting and maintaining research materials and data of these projects. The Center for Research Resources, NINJAL, is getting ready to make them public, but recently, media for linguistics (e.g., phonetic data recordings on cassettes and digital audio tapes) have been increasing and difficult to organize. This paper reports how it can be done in 2016.

### 1. まえがき

国語研研究情報発信センター研究資料室では、2003年度以降、EAD（Encoded Archival Description）を採用した資料記述・検索システムの構築に着手し、2005年9月までに一定程度の完成をみた（詳細は森本論文参照）[1]。2013年10月、新たな資料管理者に交替し、利用者が研究資料にアクセスしやすい環境の整備を行なった。具体的には、資料庫（紙資料が中心）総点検を実施し資料の所在自体を確認するとともに、EAD記述の簡略化や移管票の改良（後者は未実施）を試みた（詳細は寺島論文参照）[2]。2016年度よりさらに新たな陣容・体制となり、収蔵資料の公開に向けて、不十分であった点の改善や新移管票・資料群概要の策定、未実施であったメディア資料の点検などの抜本的な資料管理・移管システムの改良を行なっている。本稿では、これまでに行なってきた事業の概要とその問題点、そしてこれから取り組んでいく作業の改良点と今後の展望について述べていく。

### 2. 国語研研究資料室における資料収集の流れ

#### 2.1. 既存の国語研コーパス・データベースとの関係

国語研では既に多数のコーパスやデータベースを公開しているが、研究資料室に移管される資料群は、前者が研究成果及び結果を示すデータベースとすれば、その結果を生み出す過程で生成された資料であるという点に特徴がある。言語を扱う研究機関であるために、資料の媒体が多様であり、また言語の研究には欠かせない録音・録画資料が多数存在する。特に方言などの音声資料や個票は今後二度と収集できない貴重なものも多く、これら資料を再活用する環境を整える必要がある。また、移管資料そのものが研究の過程を示す痕跡でもあり、プロジェクト研究をどのような道筋で行なうべきかを示唆するパイロットデータともなり得る。

#### 2.2. 研究資料移管の流れ

研究資料移管の流れを簡単に表すと、次のとお

りである。

- ①資料移管者(主に研究プロジェクトの代表者)より研究資料室に移管を申し出  
↓
- ②資料移管者が「移管票」を作成・提出  
↓
- ③所内委員会審議, 受け入れ可否決定  
↓
- ④可の場合, 保存箱を資料保管室へ仮運び込み  
↓
- ⑤研究資料室担当者が移管票をもとに「資料群概要」作成, 紙資料・電子データを目録(リスト)に登録  
↓
- ⑥保存箱の正式配架(中央資料庫, 中央メディア保管庫)  
①では, まず資料移管者(以下移管者)が研究情報発信センターの研究資料室担当者(以下担当者)に申し出を行なう。  
②に移行する前に, 担当者は移管者に移管準備作業の説明を行ない, 実際の移管資料を確認し, 資料を入れる移管箱と移管箱の区別に使う貼付用ラベル, 移管票の各シートを渡す。  
②では, 移管者は渡された移管票に必要事項を記入し, 同時に移管箱への資料箱詰め作業を行なう。記入が完了したら移管票を担当者に渡す。  
③までに担当者は移管票に記入漏れがないかなどを確認し, 確認後, 所内委員会に移管票を提出し, 審議にかける。  
④委員会が移管予定資料を研究・教育に寄与するものと判断した場合, 移管が許可される。移管者は, 担当者立ち会いのもと, 移管箱を資料保管室に運び込む(仮置き)。  
⑤では, 担当者が研究資料室にて登録・配架の準備を行なう。具体的には, 移管資料のフォンド番号(fo○○○○)を確定し, 移管票をもとに資料群概要を作成するとともに, 紙資料は配架リスト上での登録用箱番号を確定し, メディア資料はメディア媒体ごとのID番号を確定して各メディアリストに情報を登録する。  
⑥では, 資料群概要, 各リストの登録完了後に移管資料をそれぞれ中央資料庫・メディア資料庫

に配架する。

移管資料は, 大まかに以上のような流れで準備し, 登録・配架される。

### 2.3. 平成 28 年度及び今後の公開作業計画

まず, 収蔵資料の資料群概要の一般公開準備作業は, 4月から12月までに全約230件の概要記述の整備を行ない, 翌1月から3月にかけて公開準備を行なう。

収蔵資料の公開と並行して収蔵音源の所内配信システムサービスも実施する予定である。後述のように, カセットテープやオープンリールの音声資料をMP3化し, 所内限定で研究利用のため試聴できる仕組みを構築中である。音声資料の大部分はインフォーマントのインタビュー記録であり, 個人情報を含むため, 所内限定利用とする。

また, 煩雑であった中央資料庫収蔵資料の配架替えも行なう。4月から9月を準備期間とし, 10月から11月に配架替えを行ない, 12月から3月に配架リストの修正作業に移る。

年度を通じて収蔵資料目録(研究資料室データベース)の拡張整備を行ない, また, 音源資料のメディア変換も実施する(約1800件)。合わせて劣化資料対策も施す。

後述する資料移管処理の再検討を行ない, 12月以降, 新処理手順による移管業務を行なう。

## 3. 移管票の改良

### 3.1. これまでの移管票と問題点

前章で説明した通り, 資料を研究資料室に移管する際には, 移管者は移管票を提出することになっているが, これまで移管者(一部分担当者)が記入すべき移管票は2種類あり, 記入必要事項は以下の通りであった。

#### A. 移管資料の概要

- ①移管年月日
- ②移管量(保存箱(紙)・メディア箱・カードケース箱ごとに数量を記載)
- ③移管者名
- ④受取者名
- ⑤調査・研究名称及び期間
- ⑥担当部署・担当者名
- ⑦研究・調査の概要(成果刊行物・関係調査等

があれば記載)

⑧保存・利用にあたっての留意事項(資料の媒体・個人情報の扱い等)

⑨調査ID(=fondID. 担当者が記載)

**B. 資料移管リスト(カード資料を除く)**

①保存箱(メディア箱)番号

②内容(メディアの場合, その種別)

③調査名

④備考(メディアの場合のみ)

上記のようなフォーマットをもとに, 担当者はEADや目録を作成してきたのであるが, 運用にあたっていくつかの問題が生じていた.

まず, 移管票Aにおいて, 記入すべき事項の説明が大雑把であった点である. 例えば⑦・⑧のように, 特に記載してほしい点をカッコ書きにしまったがために, 移管票によって内容・分量に差が生じ, 記入不足の際には担当者の概要記述作成に不便が出ていた.

また, 移管票Bにおいてはメディアの項目が不十分であり, 例えば, メディア種別と数量のみが移管票に記入される一方, 1点1点のメディアタイトルや内容を記入する項目がなかったために, 移管されてから担当者がメディアのタイトルを目視してリストに記載しなければならない場合が多々あった. これでは担当者がメディアリストの作成により多くの時間を費やしてしまうばかりでなく, 移管当事者でないためにメディアの内容が不明で, 適切でないタイトルをリストに記載してしまう可能性もあり, この点はそのまま放置できない問題であった.

**3.2. 新移管票の作成と目的**

前項のような問題を踏まえて, 資料移管に関わる手続きの大幅な改定を行なった.

新しい移管票の記入項目は以下のようにした.

**A. 資料移管票**

①申請年月日

②移管年月日

③移管量(保存箱・メディア箱・カードケース箱ごとに数量を記載)※

④移管者名・所属

⑤受取者名

⑥調査・研究名称(あれば通称)※

⑦担当部署・実施期間※

⑧担当者名※

⑨資料使用言語

⑩管理歴

⑪調査・研究の概要※

⑫資料群に含まれる資料の年代

⑬利用にあたっての留意点※

⑭複写条件※

⑮関連調査・研究

⑯成果刊行物

⑰備考

**B. 資料移管リスト(紙資料)**

①保存箱番号

②内容

③備考

**C. 資料移管リスト(視聴覚メディア・電子メディア)**

①メディア箱番号

②メディア種別

③内容

④数量

⑤備考

**D. 資料移管リスト(カード・その他)**

①番号

②形態

③内容

④数量・単位

⑤備考

**E. 移管資料詳細リスト(視聴覚メディア・電子メディア)**

①メディア種別

②通し番号

③メディア箱番号

④内容

⑤特記事項

特に改良・新設したのは移管票A, Eである. 移管票A(図1)では旧来のものと比べて記入する項目を増設ないしは細分化し, また※を必須記入項目とした. ※の項目は資料群概要を記述する際に最低限必要であり, こうすることにより, 移管票ごとの記述のムラを軽減するとともに, 移管票を元にした資料群概要の記述をより充実させることが可能となる.



「EADの導入時には多様な資料群に対応すべく記述項目が数多く設けられたが、かえって各階層の項目が過大となる結果も生み、どの項目に、どのような情報を記述するのか不明になりがちであった。この問題のために記述者によって情報量にバラつきが出るようになった。同様の理由で、過大な情報量のEADを反映した概要記述は長大になることがあり、概要を示すにも関わらず概要把握には向かないという問題に陥った」。

以上のような問題が発生したために、XML形式によるEADに代わり、概要情報と紙・メディア目録とがリンクしたデータベースを新たに構築し、これに伴う資料移管体制の見直しが大きな課題となったのである。前年度までは新たな移管体制に向けての基礎作業を行ってきた。

#### 4.2. (新) 資料群概要の内容と改善点

前記の基礎作業の後、本年度より新たな資料群概要の作成に着手した。まずは資料群概要(図3)の項目を以下に列記する。

- ①資料群 ID (fond 番号)
- ②移管年月日
- ③移管者名
- ④記述作成年月日
- ⑤記述作成者
- ⑥記述言語
- ⑦概要
- ⑧配架位置
- ⑨公開年月日
- ⑩表題
- ⑪資料群に含まれる資料の年代
- ⑫数量・形態
- ⑬資料作成者
- ⑭資料作成組織の履歴
- ⑮資料作成年月日
- ⑯管理歴
- ⑰入手情報
- ⑱資料内容
- ⑲評価・廃棄
- ⑳追加受入
- ㉑利用条件
- ㉒複写条件

- ㉓資料使用言語
- ㉔物理的特徴
- ㉕検索手段
- ㉖オリジナル資料の存在
- ㉗複製の存在
- ㉘関連調査・研究
- ㉙成果刊行物 (DB 等含む)
- ㉚備考

資料群概要の各項目は基本的にISAD(G)に準拠している。これに加えて、一般公開した際に検索に資する目的で⑦(研究資料の)概要を付け加えた。この資料群概要は先述した新移管票と項目がリンクしており、これまでEAD記述、資料概要、各リストでバラバラであった記述内容の統一を図っている。なお、これまでのEAD記述や移管票など、旧方式で記述された研究資料についても、新資料群概要への転記を行なっている。

#### 5. 電子データの整理作業

研究資料室には5万点を超える電子データ(数量の多い順にDAT23,000本以上、カセットテープ16,000本以上、フロッピーディスク4,000枚以上ほか)が収蔵されているが、紙資料はほぼリスト(目録)への記載が完了しているものの電子データ資料に関しては大半がリスト未登録ないしは再チェックが必要な状態である。現在、メディア資料の1つ1つにIDをふり、内容(タイトル)、配架場所等を記入したリスト(Excel形式)を作成中である。このメディアリストは、最終的には資料群概要とリンクさせ、検索者が資料群概要から1点1点の電子データにたどり着くようなシステムの構築を目指している。

また、CD、DVD、MOなどについては、それぞれ中に入っている拡張子の情報もリスト化することができた。所蔵CD907枚、DVD866枚、MO80枚(ファイル数CD172,235個、DVD70,279個、MO29,542個)を記録しているが、電子データは個々のタイトルやファイル数を把握しているだけでは意味がない。これらファイルの拡張子を網羅し、かつファイルを閲覧するためのソフトウェアも備えていなければならない。現在判明している拡張子の種類は、CD126種、DVD21種、MO677種である。目下開くことのできるファイ

ル、もしくは開くことのできないファイルがどの程度あるかの確認に努めている（なお、上記の数量は2016年9月7日現在のものである）。

カセットテープやオープンリール等の視聴覚データについても、順次電子データ化を進めているところである。

これと並行して、各メディアに記録された音声資料をMP3化し、公開の際には、所内限定で音声資料を聴くことができるような体制づくりにも取り組んでいる。オープンリールテープなどの劣化の進む視聴覚データはできるだけ早く電子データ化して保存する必要があり、また音声データをMP3化することで、より手軽な形で所内公開が可能となる。

## 6. 結論

国語研究資料室では、収蔵資料の一般公開に向けてEAD方式でのデータベース構築から始まり、膨大な量の紙資料・メディア資料の登録作業を続けつつ、移管票・資料群概要の改良に取り組んできた。これまでの方式の問題点及び改善した点は本文で述べたとおりである。要するに、これまで蓄積されてきたデータを活用しつつ、EAD方式から資料群概要一紙資料・メディア資料リストをリンクさせての検索方式に移行すべく作業を行なっているのである。これまで資料リスト・EAD記述・旧概要記述の各々で記入基準が不統一であったものを、新しい移管票・資料群概要を採用することにより一本化し、移管の流れもより円滑に、確実にこなえるよう改良を進めている。資料群概要は2017年中の公開を目指す。

他方、電子データにあっては、不明な拡張子の扱いを如何にするか等の問題が残っている。電子データと対応するソフトウェアと一緒に保存しないと無意味であるので、それへの対応が閲覧環境の整備に向けての今後の課題である。

資料群ID (fona番号)	fo0027
移管年月日	
移管者名	
記述作成年月日	2006/7/20
記述作成者	小高京子
記述言語	日本語
概要	幼児・低学年児童は、「どれだけの語彙量を持つか」「どんな語彙体系を持つか」を明らかにすることを目的として、1976年度から1978年度にかけて実施した幼児の言語調査
記述位置	
公開年月日	
表題	幼児・低学年児童の語彙調査
資料群に含まれる資料の年代	1976-1978
数量・形態	紙資料2箱、(カセットテープ47本、オープンリールビデオ6本)
資料作成者	言語教育研究部第1研究室
資料作成組織の履歴	<p>一研究概要一 本研究は、昭和51(1976)年度から昭和53(1978)年度にかけて、科学研究費補助金(一般研究A)の交付を受け、幼児・低学年児童における語彙の形態・特色を現時点で明らかにし、今後の語彙教育に、多少なりと資することを期して行なわれた研究である。また、1974年から1984年まで国立国語研究所言語教育研究部第1研究室で進めていた「幼児・児童の認知発達と語の意味の習得に関する調査研究」との連携により行なわれた。</p> <p>一研究経過一 昭和51(1976)年度は準備調査として方法論の検討。 昭和52(1977)年度は本調査として、東京の幼児・児童を対象に「語彙使用量調査」「24時間語彙調査」「語彙連想調査」「言語生活アンケート調査」を実施。 昭和53(1978)年度は東京と異なる言語地域(秋田)に住む幼児の言語調査を実施し、2女児の語彙表を作成した。詳しくは、「国立国語研究所年報」28-30(昭和51-昭和53年度)を参照のこと。</p> <p>一研究組織一 芦沢節(S51・52代表)・村石昭三(S53代表)・大久保愛・岩田純一・斎藤秀紀</p>
資料作成年月日	1976/4-1978/3
管理履歴	第1資料庫(西が丘庁舎)に保管されていたものを、移転を機に整理し、2003年移管
入手情報	
資料内容	文字化資料、文脈つき用語索引(KWIC)、品詞情報・使用数、五十音語彙表(手書き)、録音・録画資料(カセットテープ・オープンリールビデオ)
評価・産業 追加受入	
利用条件	個人情報を含む資料は、原則として非公開
複写条件	個人情報を含む資料は原則として複写不可
資料使用言語	日本語
物理的特徴	
検索手段	
オリジナル資料の存在	
複製の存在	オープンリールビデオはDV-CAMの複製あり
関連調査・研究	
成果刊行物(DB等含む)	『研究報告集30「就学前幼児の語彙—4児による日常生活の形態—」(国語研究所報告71 1982.3)
備考	

図3 資料群概要の一例

Figure 3 An example of an outline of a group of materials

## 参考文献

- 1) 森本祥子: EADを用いた資料記述システムの開発について-国立国語研究所の事例、アーカイブズ学研究, No.4, pp.92-102 (2006) .
- 2) 寺島宏貴: 日本語研究資料の整備と公開 : 国立国語研究所研究資料室の取り組み, 国立国語研究所論集, No.10, pp.245-263 (2016) .