

国立国語研究所学術情報リポジトリ

研究情報・資料の蓄積と利用のためのシステム：
国立国語研究所研究情報資料データベース

メタデータ	言語: Japanese 出版者: 公開日: 2021-06-11 キーワード (Ja): キーワード (En): 作成者: 熊谷, 康雄, 磯部, よし子 メールアドレス: 所属:
URL	https://doi.org/10.15084/00003315

研究情報・資料の蓄積と利用のためのシステム

国立国語研究所 研究情報資料データベース

情報資料研究部 第2研究室 熊谷康雄

(E-mail: ykumagai@kokken.go.jp)

情報資料研究部 第2研究室 磯部よし子

(E-mail: yoshiko@kokken.go.jp)

要旨:

国立国語研究所は1948年の設立以来、日本語および日本人の言語生活に関する科学的な各種の調査研究を行って来た。それらの資料・研究成果や現在進行中の研究・事業による資料・研究成果の蓄積と利用を組織的に行うための仕組みを構築し、運用することが目的である。ここでは、(1) 物理的な資料の保存と管理、(2) 電子媒体による資料の保存と管理、(3) 資料に関する情報の作成と管理、(4) 資料の利用と利用管理の各側面が問題になる。これらを過去から現在、そして将来に向けて整えていくことが課題である。「国立国語研究所 研究情報資料データベース」(仮称)の構築に向けて現在進めている仕事を通して、現状と問題点および目指しているところの概略を本稿と当日のデモンストレーションを通して示す。

キーワード: 研究情報 文献情報 研究資料 資料保存 ネットワーク 電子メディア

1. 目的

国立国語研究所は1948年の設立以来、日本語および日本人の言語生活に関する科学的な調査研究を行って来た。その成果は、漢字、語彙、文法、敬語、方言、音声、言語発達、言語と社会、地域の言語、言語の対照研究、日本語教育等の広い分野にわたり、これらの成果は研究報告、資料集、データ集等の刊行物の形で、広く言語研究、教育、国語施策、情報処理、その他各方面で参照され利用されてきている。

これらの研究成果は、日本語に関する基礎的資料として、また、多くの先駆的な研究として、日本語研究の中で重要な位置を占めて来ている。また、長期的な歴史のながれの中に置いてみれば、多くの研究資料は日本語の貴重な歴史的な記録でもある。

この50年間におよぶ多くの研究の蓄積の中にある情報を埋もれさせることなく、新しい研究の中で有効に活用していくと共に、日本語の基礎的研究資料として後世に伝えていくことが課題である。そのためには、研究成果の組織的な蓄積・保存・管理体制の整備と、その蓄積全体に関する情報検索・アクセス手段の整備が必要であり、そのシステムの構築が目的である。「研究情報資料データベース」の構築に向けて、多くの仕事がまだまだ途上にあるが、本稿では現在進めている仕事を通して、現状と問題点および目指しているところの概略を示す。なお、より具体的な内容・情報はデモンストレーションにより示す。

2. アプローチ

資料の保管・管理に対して、(1) 実体としての資料と (2) その資料に関する情報のふたつの側面からのアプローチが必要である。前者は資料庫、後者はデータベースとすることもできる。仕事を進めていく上での観点としては、(1) 既存の研究成果の蓄積の組織化、

(2) 知識・経験の組織化、(3) 成果・資料の保存継承、(4) 蓄積の利用や公開などがある。
ここで、研究資料・成果といっても、その内容は多岐に渡る。きちんとした形になった報告、論文というものから、中間的な資料やメモ的なもの、調査の原資料等さまざまな内容、形態のものがある。

(A) はっきりとした形になったもの：

研究者（研究の実行主体）から離れて、研究が外在化されて独立して存在し、
他者が利用できる知的産物（知識、技術、経験）
形態：報告書、論文、資料、データ、プログラム
媒体：紙、電子媒体（装置、機器）

(B) はっきりとした形にはなっていないもの：

研究の実行主体自身と一体化されていて、外在化の程度が低い
中間的な各種の資料、思いつき、メモ、経験、ノウハウ、等々。

研究活動の節目は (B) を (A) に作り上げていくことにあり、(A) は (B) から生まれてくるものであるが ((A) と (B) を行ったり来たりという見方もできるが)、資料保存の立場から見ると、研究活動や事業によって生まれ、蓄積すべき研究資料・成果は (A) から (B) までの間で、いろいろな形で存在する。そこで、アプローチの仕方として、いくつかのアプローチを組み合わせることを考えている。そのひとつとして、形になったものから入るというアプローチを考え、成果物として世に問うているものを入り口として整理するという行きかたをひとつの柱とする。すなわち、(1) 成果・資料としてはっきりしているものから整理する。その上で、(2) 上のような枠とそれ以外の枠を組み合わせることを考える。また、(B) を積極的に外在化する仕事として、あるいは (A) が生まれた背景としての情報を資料化する仕事を行う。例えば、聞き取り調査（録音・録画）などが考えられ、試行している。

3. 文献情報からのアプローチ

このように考えたとき、文献情報は過去から現在に至る研究活動の蓄積へ接近するための重要な手がかりとして捉えることができる。このような考えの上に立って、文献情報の整備を資料の保存・蓄積の立場から積極的に位置付け直して捉えており、この立場から、研究所における文献情報の整備に積極的に関わっている。国語年鑑のデータベース出版による編集出版システムの機械化や図書館システムの構築なども担当者との協力の下に進めてきており、これらを含む全体を文献のシステムとして捉え、整備を進めている。

資料情報のうち、成果がはっきりした形になっているものに対する情報として、文献情報は基本である。研究成果へのインデックスとしての機能や研究成果の蓄積に関する情報を組織化する機能を果たすものとして文献情報を捉え、研究情報システム中への組み込みの試みを行っている。そのために、書誌情報の他に、現在までに、目次のデータベース化、索引のデータベース化、図表目次のデータベース化、英文概要（英文による研究報告の要約を主要部分とする）のデータベース化を行なっている。また、これまでに刊行して来た国立国語研究所の報告書本文の全文のデータの計画がある。研究所の刊行物の文献情報は、研究報告という書物としての研究成果へのインデックスであると同時に、研究資料のアーカイブにおける役割という視点も与えることができる。調査資料やデータなど研究資料のデータベース化の仕事の中で、資料データベースとリンクしたものとして、文献情報を捉え、その整備を進めている。

研究活動の過程において生み出される資料は、内容、形式、その意味など、多様なもの

である。この研究活動の成果を記録蓄積していくという活動、研究によって生産される資料を保存、蓄積すると同時に、新たな研究の中でも活用できるような仕組みを作っていくことが必要となっている。この資料の保存蓄積という観点から見たとき、研究報告などの文献は、記録として安定したものであり、研究者である著者によって組織化されたものであるという特徴があり、資料データベースとのリンクを図ることによって、研究活動の成果を記録蓄積していくという活動の中の重要な柱になるものと捉えている。

必要な研究情報へアクセスする仕組みをどのように作ればいいのか、埋もれている情報を表に出すにはどのような仕組みを作ればいいのか、ということに関して、様々な検討をしてきているが、そのような広い枠組みの中の重要な柱のひとつが、この文献情報の整備だと考えて、情報の整備を進めて来ている。その内容はデモンストレーションの中で示すが、以下に関連する文献情報データも含めリストを示す。

研究所の刊行物に関して：

1. 書誌データ（研究所の刊行物について下記の目次データと一体のものとして作成）
2. 目次データ（約 9,900 レコード。書誌データと共にホームページで公開）
3. 索引データ（索引を持つ報告書についてはほぼ完了、約 34,000 レコード）
4. 図表目次（社会言語学関係の報告書を中心に作成）
5. 英文要旨（1988年発行の第3版のHTML版をホームページで公開中。
第4版は1998年12月に公開予定）
6. 本文データ（計画中）

研究所に限定しない日本語研究に関する文献情報として以下のものがある。

1. 日本語研究文献目録
（国語学会との共同事業による。「国語学文献総索引データ」としてホームページ上に学会員向けに試験公開中。約10万件）
2. 海外日本語研究文献目録
（国語学会との共同事業による。ホームページ上で公開。総数約13500件）
3. データベース出版による国語年鑑
4. 国語学研究文献目録データベース
（1と3をつなぐデータとしてデータ整備を進行中）
5. 国立国語研究所図書館蔵書目録データベース（遡及入力作業中、現在、約4万件）

ホームページは <http://www.kokken.go.jp>

資料に関するものとして：

1. 資料情報データ
所在、形態、内容、注記等
（現在までに主に社会言語学関係の調査に関する部分について作成している）

上記の細かい内容にわたらないが所蔵資料全体に関する粗いものとして資料の目録データを作成中。

4. 資料の保管管理体制

情報の整備と平行して、きちんとした物理的な物の保管システムを整備しなければならないが、現状ではなかなか根本的な対応が難しい状況にある。

原資料の蓄積・保管に関しても、資料の保存のための十分な手当が必要であるが、研究室に保管されているもの以外は、現在は完備した資料保管スペースというものが整備され

ておらず、現状では、段ボールに詰めて「資料庫」や空きスペースにしまい込むというような最低限のことを実行しているにすぎないものが多い。また、カード等の保管に関しても、旧式のカードボックスにしまわれたまま、保存のための環境としては好ましくないところに置かれているものも多い。現段階で可能な範囲で、資料の散逸や環境の悪化に対するの注意は行っており、資料の現状把握に努めてはいるが、物理的な資料の体系だった保管の体制は取ることができていない。近い将来に移転も控え、できる限り資料の保管体制を整えておくよう、物理的な保管方法についても、整理保管用具を整備し、情報のみならず、物理的な整理を行う必要がある。資料庫等の物理的なスペースや設備の問題もあり、可能な方策とし、保管管理体制の整備に取り組んでいるところである。目録・所在情報の整備と資料およびそれに関する情報の流れに対する全体的な共通の理解とそのシステムの運用という形を定着させるべく取り組んでいる。

5. 資料の電子化と電子化の問題

資料の保存・利用・管理のシステムを実現するための基盤のひとつの側面として、急速な発達と普及が進行しているコンピュータ、電子媒体、ネットワーク、電子機器というものがある。これらに対する評価と利用を進めることが課題である。

関連した事業として、情報2研では言語資料緊急整備というタイトルで、電子媒体で残されている資料の新しい媒体への変換・複写を中心とした保存作業を行っている。現在までは調査の音声テープが中心であり、そのDAT化を順次進めている。また、その他のメディアについては検討を進めている。

資料の電子化の有効性の一方、電子化資料の持つもろさをシステムの中で考えておく必要がある。例えば、媒体の劣化、記録の劣化、物理的な変形・破損、再生装置の消失、記録フォーマットの時代遅れ（例えば、古いワープロの文書ファイルはそのソフトがないと読めない）、プログラムの実行環境の消失（プログラムは実行できる計算機環境がなくなると使えない）、ソフトウェア資産の継承の問題等である。電子化そのものが、新しい問題を持ち込むことがあるし、すでに電子化されている資産については、このような負の側面があることを忘れてはならない。電子化によって生じる新しいコストという観点でも考える必要がある。

電子化した資料そのものが劣化に強い訳ではない。デジタル化した資料はコピーが可能であり、デジタルからデジタルへコピーしている限りにおいては、複写による劣化はないが、時間の経過と共に媒体は劣化する。また、ある媒体やフォーマットがいつまでも安泰である訳ではない。電子化した資料の資料としての安定性という観点からも種々の問題を抱えている。電子化した資料を利用するためのプラットフォームという点からも気をつけるべきことは多い。また、資料保存という観点からは、電子化した資料と実物資料の双方を保存しておく必要がある。

6. 権利の保護

著作権： 研究報告等のほとんどは国立国語研究所が著作権を持つ（執筆者が著作権を持つものが中にある）ものであり、このようなものの公開に当たって著作権に関わる障害が少ないが、資料の中には調査目的に応じて、それぞれに著作権者との一定の契約のもとに作成した資料が少なからずある。これらについては、その著作権に関する情報を明示的に集中管理しておく必要がある。

被調査者の情報の保護、プライバシー・肖像権等の保護： すでに公刊されている情報

に関しては問題はないが、もとの資料、データの扱いに関しては、被調査者の情報、権利の保護に関しての注意が必要である。調査票、録音資料、録画資料など、様々な資料があり、これらについても情報や権利の保護に関する情報を明示的に集中管理しておく必要がある。

情報公開・利用の促進という観点からも、これらの側面に対する対応の仕組みを十分に検討した上で、進んでいく必要がある。

7. 研究成果の一次情報の提供

今日、世界的にコンピュータネットワークの急速な普及が進んでおり、インターネットによる世界的な規模での情報の共有と利用が可能になってきている。また、情報処理技術の発展によって新しいメディアによる大量の情報の蓄積利用の手段が一般に利用可能になってきている。一方、印刷物としての刊行物は、学術書の多くがそうであるように、品切れ等によって入手できなくなるものが多い。また、海外においては基本的な文献として国語研究所の刊行物を情報検索によって検索することができても、その本体に触れることは、多くの場合、非常に困難である。インターネット上の2次情報の整備により、研究や資料の存在を知ることが容易にできるようになってはいるが、あるいは、できるようになってきているからこそ、研究・資料の本体にも同様の容易さでアクセスできるようにすることの必要性は高い。

この必要に応え、さらに冒頭の目的に向かうためにつぎの計画を始めようとしている。

- すなわち、国立国語研究所が1948年の設立以来行って来た日本語研究の成果を、
- (1) その中心的な存在である研究報告を電子化し、これに対する情報検索手段を整備し、これを研究成果と資料へのアクセス手段として整備する。
 - (2) インターネット等を利用して研究成果の公開を推進する。
 - (3) これらと対応する形で、その研究の元になった原資料やデータの組織的な管理・保管システムに関する研究と整備を行い、総合的かつ組織的な研究の蓄積・保存・管理・利用と公開を行うシステムに関する研究を行ないシステムの構築を行なう。

これは、現在進行中の新たな研究成果の蓄積と公開のためのシステムの基盤ともなり、現在進行中の新しい研究に関する研究情報の蓄積利用の高度化につながるものである。

研究所設立以来、これまでに蓄積されている研究成果はネットワークによる研究成果の共有というようなことが現実的なものとなる以前のものであり、このようなことは考えられていなかった。その後、すでに機械可読となっているデータに関しては、ネットワーク上での公開が検討されているものがあるが、研究所の研究成果物の全体へのアクセスという観点からは部分的なものであり、充実の必要がある。有用な研究資料の電子化という観点からの資料化を資料の保存蓄積という観点からも行うことが必要である。

8. 基盤としてのネットワークの整備

ネットワークによる情報の提供と言っても、ネットワークそのものが整備されなければ現実のものとはならない。国立国語研究所においては研究所全体にわたるネットワークの導入とインターネットの接続を1995年に行った。国立国語研究所におけるネットワークの導入の過程について概略を示す。

- 1993年 パケット交換網により学術情報ネットワークに加入
文部省学術情報センターの NACSIS-CAT (学術情報センター目録所在情報サービス) に加入し、図書館の所蔵図書の遡及入力を開始する。
[これは図書館業務専用のネットワークであり、インターネットへの接続などのためのものではない。ネットワークとしては非常に安定して稼働している。]
- 1994年 IPアドレス (Class C)、ドメイン名 (kokken. go. jp) を取得
1994年に国語研究所では初めてのワークステーション5台とパソコン4台をイーサネットに接続したマルチメディアによる研究用のシステム (音声映像解析システム) の導入を行ない、IPアドレスとドメイン名の取得を行なった。
- 1995年 SINET (学術情報ネットワーク) 経由でインターネット接続を開始
(専用線 128Kbps で、上記パケット交換網とは別の回線によりノードと接続)
研究所のメインフレームを中心として電子計算機システムの更新の機会に、メインフレームの更新だけでなく、ワークステーションと全所的なイーサネットによるネットワークの導入 (基幹ネットは FDDI、各室へは 10BASE-T) を行なった。
IPアドレスは改めて Class C を4つ取得した。ドメイン名は (kokken. go. jp) のまま。
- 1996年 8月 ホームページの外部試験公開運用を開始した。

所内はほぼ全ての部屋 (研究室や事務関係の部屋は全て) にイーサネットのコンセントを取り付け、全ての研究室および事務室からネットワークの利用ができるようになってきている。研究所へのネットワークの導入と同時に各室のパソコンは最低1台はネットワークへの接続をするように全所的に機器の整備が行なわれた。通信ソフト、WWW ブラウザなどのソフトのインストールや設定もある期間で集中的に行い、講習会の開催などを通して、ネットワークの利用を促進するための活動を行った。ネットワーク委員会を構成し、ネットワークの利用についての種々のルールを決めたり、一般のユーザのネットワークの利用を助けたり、ネットワークの管理を行ったりしてきている。このようにして、研究所内において、ネットワークを日常的なコミュニケーションのための道具として根付かせるための努力を続けてきており、研究所内では、ネットワークは日常的に不可欠の道具となっている。また、平行して、ほぼ同時に、社会的な現象とも言えるほどのインターネットの急速な普及が進んでおり、ネットワークを基盤とする研究情報の利用ということが現実的なものとなってきた。もちろん、ネットワークを研究遂行上の道具として使うという観点からは、まだまだ発展させなければならない側面が多いが、そのような展開を現実的なものとするための足場ができてきている段階と捉えている。

9. デモンストレーション

本パネル発表では、ネットワークに接続したパソコンにより、上の情報検索などを実際に示す。また、研究資料・成果の電子化や資料情報とのリンクの例として、映像資料のデジタル化や研究報告の電子化と資料情報、データとのリンクなどを事例を通して示す。

探すためのシステム、使うためのシステムという側面から、二次情報、一次情報、文献情報、資料目録、各種の索引、検索システム、各種のメディアによる資料化、資料保管庫、図書館の蔵書検索、システム構成、ネットワークなどをデモンストレーションを通して説明する。

10. おわりに

電子的なメディアやネットワークなど、現在も急速な発展が進んでいる技術的、社会的な状況・基盤を検討しながら、よりよいシステムを探りながら進んでいく必要がある。