

第7回 国語分科会言語資源小委員会(Web開催)議事録

令和 7 年 5 月 29 日
10 時 00 分 ～ 11 時 35 分
文部科学省3階 3F1特別会議室

〔出席者〕

(委員)相澤主査、石川副主査、神永、齊藤(美)、武田、前川、森各委員(計7名)

(ゲスト)渡部泰明国文学研究資料館館長ほか随行者

(文部科学省・文化庁)山田国語課長、中田課長補佐、武田主任国語調査官、

鈴木国語調査官、町田国語調査官ほか関係官

※ 相澤主査、石川副主査、神永、武田各委員、渡部館長、随行者及び
事務局は3F1特別会議室にて参加。

〔配布資料〕

- 1 文化審議会国語分科会言語資源小委員会委員名簿
- 2 文化審議会国語分科会言語資源小委員会の会議の公開について(案)
- 3 言語資源小委員会における主な審議事項及び当面の予定(案)
- 4 データ駆動による課題解決型人文学の創成～データ基盤の構築・活用による
次世代型人文学研究の開拓～(渡部泰明国文学研究資料館館長御提出資料)

〔参考資料〕

- 1 文化審議会国語分科会運営規則
- 2 今後における日本語のデジタル言語資源の整備・活用の在り方(報告)
(文化審議会国語分科会 令和7年3月17日)

〔経過概要〕

- 1 事務局から定足数の確認があった。
- 2 今期1回目の言語資源小委員会の開催に当たり、山田国語課長から挨拶があつ

た。

- 3 事務局から配布資料の確認が行われた。
- 4 参考資料1「文化審議会国語分科会運営規則」に基づき、委員の互選により、相澤委員が言語資源小委員会主査に選出され、主査就任に当たっての挨拶があった。
- 5 参考資料1「文化審議会国語分科会運営規則」に基づき、相澤主査により、石川委員が副主査に指名され、副主査就任に当たっての挨拶があった。
- 6 事務局から配布資料2「文化審議会国語分科会言語資源小委員会の会議の公開について(案)」の説明があり、言語資源小委員会の会議を公開することが了承され、以降の議事が公開された。
- 7 事務局から配布資料3「言語資源小委員会における主な審議事項及び当面の予定(案)」について説明があり、説明に対する質問及び意見を尋ねたが、特に発言はなかった。
- 8 渡部泰明国文学研究資料館館長から配布資料4「データ駆動による課題解決型人文学の創成～データ基盤の構築・活用による次世代型人文学研究の開拓～」について説明があり、説明に対する質疑応答の後、意見交換が行われた。
- 9 次回の言語資源小委員会について、令和7年8月4日(月)10時から12時まで、対面及びオンラインで開催する予定であることが確認された。
- 10 質疑応答及び意見交換における各委員の発言等は次のとおりである。

○鈴木国語調査官

定刻になりましたので、ただ今から、第7回、今期1回目の文化審議会国語分科会言語資源小委員会を開催いたします。

本日は、今期最初の小委員会ですので、後ほど主査を御選出いただくまでの間、事務局で進行をいたしますので、御了承願います。

なお、会議は実際にお集まりいただく形でも、オンラインで参加していただく形でも可能な形式で、今後も継続する予定でございます。また、後ほど会議が公開となりましたら、傍聴の方にもオンラインでこの会議を御覧になることを御承知おきくださいませ。

定足数の確認を最初に行います。言語資源小委員会は8名で構成されており、本日は

7名の御出席を頂いております。文化審議会国語分科会運営規則の第4条に定められた、会議開催に必要な過半数出席という条件を満たしていることを御報告いたします。

次に、事務局の異動について御報告いたします。令和7年4月1日付けで国語課長として山田亜紀子が、課長補佐として中田尚樹が着任いたしました。

初めに、国語課長の山田から御挨拶申し上げます。

○山田国語課長

国語課長の山田でございます。本日、第7回の言語資源小委員会、今期1回目ということで、開催に当たりまして、一言、私の方から御挨拶申し上げます。

言語資源につきましては、これまで、令和5年9月の国語分科会におきまして、「日本語のデジタル言語資源の整備に関する国語分科会の見解」と、さらにその後、令和7年3月ー本年3月の国語分科会におきまして、「今後における日本語のデジタル言語資源の整備・活用の在り方」という報告をおまとめいただいております。

この間、文化庁の方の具体的な事業としましては、昨年度から、現代日本語の保存・活用のためのデジタル基盤整備事業という新しい事業を開始いたしまして、国立国語研究所の「現代日本語書き言葉均衡コーパス」の拡充を進めていただいております。

近年、生成AI、それから、その基となる大規模言語モデルの開発ということが急速に進められているということでございまして、国の政策としまして、ちょうど昨日、AI推進法が成立したということで、政府の方の体制、取組の強化というのも求められているところでございます。

このAI推進法の法案審議の中におきましても、日本語のデジタル言語資源の充実というものに関するやり取りがあったということもございまして、一層、この重要性が高まっているという認識をしております。そうした中で、前期に引き続きまして、この言語資源小委員会が設置されたということでございます。

今期につきましては、前期の報告のフォローアップということで、国内の様々なデジタル言語資源に関する取組というものを取り上げていただくとともに、その他の社会的な課題につきましても御議論いただくということを期待しているところでございます。

委員の皆様におかれましては、大変御多忙の中、恐縮でございますけれども、言語資源の整備と活用ということにつきまして、それぞれの専門的なお立場から活発な御議論を頂ければと思っておりますので、どうぞよろしくお願いいたします。

【委員の互選により相澤委員を主査に選出】

○鈴木国語調査官

それでは、相澤主査、御挨拶をお願いいたします。

○相澤主査

どうもありがとうございます。情報分野からの参加ということで、大変僭越ではございますけれども、皆様の見識あるお立場からの御意見や、御講演いただく先生方の御知見などについて、素朴な質問等で議論を活性化できればと存じますので、どうぞよろしくお願い申し上げます。

また、私、生成AIー特に大規模言語モデルを構築する立場におりまして、去年、この小委員会が始まった当初は、大規模言語モデルが生成するテキストというのが、人間のテキストと不可分な形で世の中に広まるという現象がリアルに想像できていませんでしたが、正にそういったことも問題になってくる時代に突入していると存じますので、そういった観点からも、議論を楽しみにしております。

どうぞよろしくお願い申し上げます。

【相澤主査の指名により石川委員を副主査に】

○鈴木国語調査官

それでは、石川副主査から御挨拶をお願いいたします。

○石川副主査

神戸大学の石川でございます。本日の課長の御挨拶にもございましたように、言語資源ーデジタル言語資源を国を挙げて整備をしていく、また、その活用の可能性を考えていくということは、非常に重要な問題であろうと存じております。力不足ではありますが、このミッションの実現に全力を尽くしたいと存じます。どうぞよろしくお願い申し上げます。

○鈴木国語調査官

ここからの議事につきましては、相澤主査にお願い申し上げます。

○相澤主査

それでは、配布資料2「文化審議会国語分科会言語資源小委員会の会議の公開について(案)」を御覧ください。こちらにつきましては、5月20日の国語分科会で御了解いただいた、「文化審議会国語分科会の会議の公開について」に倣った案で、言語資源小委員会の会議について公開の在り方を定めようとするものとなっております。

では、事務局から御説明をよろしく願いいたします。

○鈴木国語調査官

配布資料2について御説明申し上げます。これは、今、相澤主査からお話があったとおり、国語分科会の公開に倣いまして、この言語小委員会も公開するという形で整理されたもので、基本的に会議は公開をするという形でございます。

例外規定としては、ただ今あった主査の選任等、人事に関すること、それから、その内容が公開するのにはふさわしくないという主査の御判断のあった場合、そちらにつきましては、公開をしないという場合がございます。

そして、この会議を傍聴していただく方は、オンラインでの傍聴という形になりますが、事前登録を必要とするという形を取っております。

それから、配布資料については、基本的には会議の前日に公開するという形を取っております。なお、議事録につきましては、前期と同様、確定をしてから公開いたしますので、会議資料として提出する場合も「委員限り」という扱いで進めさせていただくことにいたします。議事録の公開については、最終的にはウェブサイトでの公開となりますが、確定の後ということでございます。確定までの間、修正が必要なところにつきましては、事務局の方に御連絡いただければ、それに合わせた形で修正させていただきます。

以上でございます。

○相澤主査

ありがとうございます。では、ただ今の御説明について御質問等ございましたら、よろしくをお願いします。（挙手なし。）

よろしいでしょうか。では、こちらはお認めいただけるということでよろしいでしょうか。

(「異議なし」の声あり。)

では、こちらの文化審議会国語分科会言語資源小委員会の会議の公開について、「(案)」を取りまして、合意いただいたものと決定いたします。これに従いまして、ここからの本日の議事については、オンライン上で公開となります。どうぞ御準備をよろしくお願いいたします。

【YouTube配信開始】

○相澤主査

審議事項に入っております。前期で報告をまとめたこの言語資源小委員会における今期の主な審議事項について、まず確認をしてから、具体的な審議に入っております。

お手元の配布資料3「言語資源小委員会における主な審議事項及び当面の予定(案)」についての御説明、山田課長からよろしくお願いいたします。

○山田国語課長

国語課長の山田でございます。それでは、配布資料3に基づきまして御説明いたします。

まず、これまでの審議状況ですけれども、前期につきましては、昨年度から国立国語研究所の「現代日本語書き言葉均衡コーパス」—こちらの拡充を文化庁の予算で行うということがございましたので、これに合わせて、日本語のデジタル言語資源の整備・活用の在り方について幅広く御議論を頂いたということでございまして、その結果については、本年3月にこちらに記載のとおり、報告書という形でまとめていただいたところでございます。

これを受けまして、今期の審議事項ですけれども、この3月の報告におきまして、言わば理論的な部分について、新たな考え方というのをお示しいただきましたので、それに基づきまして、今期はより具体的な議論をしていくということで考えております。

具体的には、国立国語研究所のコーパスの進捗状況などに関する情報共有、それから、本日、事例発表を頂くことも含めまして、様々な言語資源を活用した活用事例、社会課題解決のための取組、そして、地域の文化的基盤としての地域の言語資源の在り方ということで、例えば方言のようなものに関する考え方の整理、さらには外来語をめぐる課題の検討といったようなことなど、より具体的な観点からの御議論を深めていただければと思っていますところでございます。

日程につきましては、「3」に記載のとおりでございます。

どうぞ皆様、御指導のほどよろしくお願いいたします。

○相澤主査

ありがとうございます。それでは、ただ今の御説明に対して、何か御質問、御意見等ありましたら、御自由に御発言をお願いいたします。オンライン御参加の方におかれましては、挙手いただければと思います。よろしいでしょうか。（挙手なし。）

では、よろしいようでしたら、こちらの予定されている審議事項に留意しつつ、今後の予定ということで、現在のところ、5月、8月、9月と、事例紹介を含めて小委員会の開催が予定されておりますので、それに従って進めてまいりたいと存じます。

それでは、本日のメインとなりますが、前期の報告書では、言語コーパスをはじめとしたデジタル言語資源の継続的整備の必要性とともに、このデジタル言語資源を活用した社会課題の解決ということの重要性について、特に強調しようということで、「コーパスアプローチ」というキーワードで検討をいたしました。

その最後の小委員会でも、リーフレットの形で、この課題解決というキーワードで御紹介がありましたが、本日は国文学研究資料館のプロジェクト、「データ駆動による課題解決型人文学の創成 データ基盤の構築・活用による次世代型人文学研究の開拓」について、渡部館長から御説明いただくこととなっております。

渡部館長におかれましては、御多忙の中、どうもありがとうございます。それでは、よろしくお願いいたします。

○渡部発表者

おはようございます。国文学研究資料館の渡部です。どうぞよろしくお願いいたしま。本日は、当館の事業について御説明の時間を取ってくださいます、本当に感謝申し上げます。

お手元の配布資料及び投影資料に基づきまして御説明申し上げていきたいと思ひます。

1ページの表紙の部分を見ていただきますと、大規模学術フロンティア促進事業とございます。文部科学省の事業で、略してフロンティア事業などと私たちは申しておりますが、十数件の事業がこれに採用されまして、現在進んでおります。

その中で、文系としては唯一の事業でございます。そのほかの理系の事業はスーパーカミオカンデであるとか、国を挙げての事業ですけれども、文系として私たちがこれを進め

ております。それが、「データ駆動による課題解決型人文学の創成」、副題として、「データ基盤の構築・活用による次世代型人文学研究の開拓」という長い題になっております。

余りにも長いので、略称で「国文研DDHプロジェクト」—これでもそんなに短くないんですが—と申しております。DDHというのは、データ・ドリブン・ヒューマニティーズということで、データ駆動の人文学だということなんですけれども、これについてかいつまんで、そして、皆様の御関心に即してお話をしていきたいと思います。

まず、2ページをお開けください。このプロジェクトの内容に入る前に、当館、国文学研究資料館、略して国文研の概要をそこに記しております。日本各地に存在する日本の古典籍、歴史資料など—その周辺資料も含めますけれども—そうした資料、日本はこれが生の形で残っている、本当にこれだけ天災の多い国でよくぞこれだけ残っていたものだと思います。それは、「残った」じゃなくて、「残してきた」わけなんですけれども、その資料を収集して、そして公開していくということを、一番大きな使命として53年の間進めてまいりました。

その進めてまいりました事業の一つの総決算としまして、2014年から、このページの一番下のやや右側のほうに、一番下に年表がありますけど、2014年4月からフロンティア事業を開始とあります。これは、現在のフロンティア事業ではなくて、その前のフロンティア事業です。2014年から2023年までの10年間進めてまいりました。

「歴史的典籍ネットワーク事業」などと呼んでおりますけれども、ここでやったことというのは、要するに古典籍をデジタル画像にして、そして、それをサイト、データベースとして公開するという事業です。30万点の古典籍、「源氏物語」でも1点ですから、コマ数でいうと3,200万コマを公開いたしました。その成果が評価されまして、これを継続してというか、更に続けて次の事業を行うようにということを許されまして、そして始めたのが、去年から、2024年4月から開始いたしましたフロンティア事業、このDDHプロジェクトという形になります。

ですから、30万点、3,200万コマのデジタル画像—これは古典籍を高精細の画像で写しているわけなんですけれども—これをどのように利活用していくかということが、大きな柱になっていくわけです。

そこで3ページを御覧ください。国文研DDHプロジェクトの概要として、二つの柱をお示しいたしました。まずは、研究基盤データベースのデータ拡充・高度化として取り上げてあります。画像データの範囲を、今までは江戸時代までだったんですけれども、それを近

代の初頭まで拡張いたします。そして、さらにそれを機械可読のテキストデータに変換する。機械可読ということは、人間も、機械も読み取れるように。画像データというのは、当然これは版本や写本ですからくずし字で書かれているわけで、一つのハードルがあったーそういうくずし字が読める人にしか読めないということになっていた。それを、人間も読める、機械も読める、そういう形にしようではないか。いつでも、どこでも、どなたでもと一右にありますけれどもーそういうことを更に目指して、国書データベースー我々がやっておりますこのデータベースを拡充して、高度化していく。ここに「ただで」という言葉も付け加えてあります。

そして、2番目の柱として、その下にあります、異分野融合によるデータ駆動型研究の推進と。先ほどの1番目の柱が事業ベースであったのに対して、この2番目の柱は研究開発ベースと申したらいいでしょうか、研究は当然、我々、研究機関ですから、中心となるものでもあります。さらに国文学に限らず、異分野融合を大いに進めていこうではないかということで、四つの研究計画、4種類のー4類のと言った方がよいかもしれません。

データインフラストラクチャーの構築、人文系データ分析技術の開発、コンテンツ解析、マテリアル分析・解析の4類の研究計画を立てております。このうち、トピックとして取り扱う、今回御紹介したいというものに限って御紹介をしていきたいと思っております。

4ページをお開けください。先ほど申し上げました2本柱のうちの最初の一つです。研究基盤データベースのデータ拡充・高度化。これは、先ほども申しましたが、30万点の歴史的典籍ー古典籍なんですけれども、歴史資料もちょっと含まれておりますので、歴史的典籍と言っていますがーこれをOCRによって、機械で読み取れる形に読み取っていく。そして、テキストデータを作成していくということです。

左側にその一つのフローを書いておりますけれども、2020年度実績とあります。OCRによってテキストデータ化していくと。先週、聞いたところでは、今、OCRでくずし字が92%読めるようになっていると。もちろん、特殊なものは除きます。浄瑠璃本のようなのは、我々でも読むのに苦労する。どこからどこまでが字なのか、ちょっと分かりにくいような非常に特殊なものもありますので、それを除くとして、普通の写本であれば92%。このAIを使ったOCRの技術は高まったのだと実感します。しかし、92%ということは、8%間違っているわけで、400字詰め原稿用紙で1枚につき32字誤字があるということになるので、それをそのまま研究に使うというわけには、なかなか難しいところがある。

もっとも、その状態でも調べられることはあるので、それもまた大いに使っていこうでは

ないかとー我々の中では。なかなか公表するというわけには、難しいところもありますけれどもーそれを用いて進めていこうとはしております。

今申し上げましたように、8%なりの読めていないものがあるわけで、それは人間の手で正していかなければならない。その正したものを校正テキストデータと私たちは呼んでおります。人間の手ですけれども、くずし字が読めさえすれば、古典に対するそれなりの理解さえあれば、多くの人にやっていただけるし、また、大変手間の掛かることですから、たくさんの方に参加していただきたい。そこで、ワークフローを作って、さらに業務運用管理ツールを開発することによって、学生、院生などでもどんどん参加してもらおう。そうすることによって、大量のテキストデータを作っていけるだろう。それは、また、してもらおう院生、学生の立場というか、院生を中心に考えれば、それは育成の一環でもあるわけです。是非こういう世界に目を開いてもらって、我々自身の研究分野の活性化にもつなげていきたいし、次世代を育成していきたいと思います。

その一例として、右側に校正テキストデータの例として「紙漉重寶記」と、あと「江戸大地震出火場明細記」というのを挙げておきました。これは、要するに絵と言葉が同じ画面の中に共存しているような例です。こういうようなものも、今かなり読めるようになってきています。

それから、その下の方に、共同研究として「AI技術を用いた大規模古典籍画像に対する新たな検索手法の研究」ということで、昨年度開始したものですけれども、これCLIPと申すんでしょうか、画像と言語の情報を同時に処理するモデルということですが、これなどを使うことによって、画像を文章の形で検索できるようになる。結構長文にわたる一長文と言っても、もちろん限界はありますが、それなりの長さを持つテキストから画像を検索することができるようになるだろうと思います。

日本の古典籍の大きな特徴として、絵と文が補い合って一つの作品を形作っている。そういう特色がありますけれども、そういう日本古典籍を幅広く検索、分析することを可能にする研究と思っております。

次のページへ参ります。異分野融合によるデータ駆動型研究の推進ということで、これは、最初に申し上げた大きな柱の2番目ということになります。研究ベース、研究開発ベースの研究計画です。今年度から、重点課題研究、テキストデータ関連の重点課題研究ですけれども、これを開始いたしました。

その中で、文献観光資源学というものを推進していこうというのが、一つ大きな柱にな

っております。それは、要するに今まで、従来、私たち、53年前から始めていた事業というのは、地方にある資料を撮影して、マイクロフィルムなり、紙焼き写真にして集めて、そして当館に来てもらって、それを見てもらうというものでした。これだけでも随分省力化になるんですが、しかし、結局これは地方のものを中央に集めたー吸収したということになって、地域の方々はそれを自由にすぐに見るということはなかなか難しかった。

それを、地域密着型ー中央へ吸収するだけじゃなくて、地域の研究者と一緒に調査研究して、その研究成果を地域社会に還元して、そして、地域の市民や学生と共に学んで、さらに地域から発信していくーこういうようなことを行って、地域振興と言うとなかなか大げさかもしれませんが、あえて使わせていただければ、地域振興に役立てていきたい、観光資源としていきたい。

例えば、そういう中の例として、右側に、岡山県の正宗文庫^{まさむねぶんこ}というものがあります。これは正宗敦夫^{まさむねあつお}という優れた国文学者が、本当にこつこつと集めた和書、そして、彼自身が学者として様々な基盤データをそれぞれ作った作品の類い、これを独力で作っていったというなかなか偉い学者なんです。しかし、この方、岡山県にずっと在住していて、決して東京へ出てこなかった。

お兄さんの正宗白鳥^{まさむねはくちょう}はとっとと正宗家を捨てて東京へ出て行ったんですけれども、正宗

敦夫はずっと残って、地方文化の発展に尽くしたんです。その資料を公開してもらうことになったのがもう既に十数年前ですが、それを国文学研究資料館が地域の研究者と共に調査してまいりました。その調査の結果、展示を行ったりーこれは岡山県立博物館で展示を行い、また、「正宗文庫セミナー」と称して、その研究成果を公表し、皆さんに聞いていただくということを、岡山県で行ったりいたしました。それは、本当に正宗敦夫の精神を受け継ぐというんですか、そういう事業だと自負しております。

同じようなのが一番下にあります。大分県の廣瀬資料館、これは大分県の日田というところにあります。行くと、すごく不便なところですが、実は江戸時代にはここが一つ文化の拠点であったということが分かっております。現在、「進撃の巨人」で有名になったところですから。ここに廣瀬淡窓という大変優れた儒学者がおりました。

この廣瀬淡窓^{ひろせたんそう}が咸宜園^{かんぎえん}という、要は私塾を開きまして、ここにも和本の類いが各地域から集められて、大変優れた文庫を形成しています。そして、また、これは私塾ですから、教育も大変ユニークで、「三奪の法」と称していたようです。要するに、身分の上下にかかわ

らず、全員同じスタートで始めて、そこから非常に厳しい試験を通った者だけが進級していく。これは本当に身分に関係なく行うという、非常に当時としては画期的な教育を行っていたところですよ。大村益次郎であるとか、高野長英であるとか、そういった人たちを生み出したところですよ。

こうしたところで同じようにセミナーを開き、展示なども行うということを協力してやっております。言わば、廣瀬淡窓の精神をそのまま受け継いでいくということになりましょう。

そのほか、ここに取り上げなかったんですけども、私たちがこれから目玉にしていきたいというものがあります。口頭になりますけれども、福島県の大熊町、富岡町、浪江町という、いわゆる複合災害のあった、現在でも帰還困難地域が多く残っている、そういう市町村がございます。ここで、最初は資料レスキューというところから始めまして、災害に遭った資料を守っていく。そして、これからさらにそれを保全していく、そして、それを公開し、その地域の方々と共に学んでいくというような、そういう活動をするための協定を、今申し上げた大熊町、富岡町、浪江町と結んでおります。これから、大いにその活動を進めていきたいと思っております。

続きまして、6ページへ参ります。6ページ、分野融合のデータ駆動型研究のまた一つであります典籍防災学、これもネットワーク事業―前の事業から続けております研究分野です。例えば、左側に宇宙物理学―地球物理学でもありますけれども、片岡龍峰さんという、極地研究所の研究者がいらっしゃいますが、その片岡さんと協力して、日本の古文文献に残っている「赤気」と書きますけれども、これはオーロラのことらしいということを特定いたしました。要するに、「中緯度オーロラ」です。オーロラだったら、本来は極地で見られるものですけれども、日本でも見られることがある。それは、巨大な磁気嵐が起こったときにそういう現象が見られるので、何百年かに1回、そういう記録が残っている。それを見ますと、そういう巨大磁気嵐が起こったときが分かるわけで―ということは、それに伴う天災などもそこから分析することができる。そうすると、それは天災への対応に資するのではないかとということで、その論文―本来片岡さんは理系の雑誌に載せていますけれども―当館の雑誌にもその論文を載せていただいたということです。これには、もちろん当館の研究者が協力しております。

そういったような視点から、古気候の解析、災害への適応事例などを判断して、例えば茨城大学などとも協力して大規模自然災害、気候変動というものを古典籍などから分析していくという研究を更に進めております。これは右側にあります。

7ページの方へ参ります。データ駆動型研究の一つとして、TEIに準拠した構造化テキストの作成というものが挙げられます。TEI=Text Encoding Initiative、テキストデータを構造化していく、その仕方などの指針を私たち自身が示していった、これもたくさんの方々の協力が必要になりますけれども、このTEIで構造化していくことを進めていきたいと思っております。

既に進めているものの例として、右側に廣瀬本の可視化を示しました。廣瀬本萬葉集をこのTEIによって様々な面から検索することが可能になっています。現在我々が見る萬葉集というのは、鎌倉時代以降の訓読が基になっているんですけれども、鎌倉時代以前の万葉集の訓読を復元できる廣瀬本萬葉集という写本が三十数年前、発見されたわけなんです。

それには、古い萬葉集に見せ消ちの形で、様々な時代の訓読が積み重なっているわけです。ちょうど地層の断面図みたいなものです。一番下からずっと積み上がっている。だから、一見するとごちゃごちゃして、何だかよく分からないんですけれども、その一つ一つを剥いでいくように、これは第何地層の何とかという訓読であるといったことを比定していきながら、TEIで構造化していきますと、何時代の注の部分抜き出せとか、あるいは、この言葉を持っている注の部分、見せ消ちの部分抜き出すとか、そういうようなことができるようになる。

これは、研究者にとっても本当に福音と言っているような大きな対応で、これはまだ完成していないんですけれども、でも、試行として一試しとして部分的に公開をしております。

8ページへ参ります。このようなTEIに準拠した構造化テキストの作成というのも大変手間が掛かり、多くの人手が掛かります。これに対して、画像データがあり、翻刻して、さらにそれをマークアップしていく=構造化していくという、その作業、これも、学生や院生なども動員しながら、次世代育成も兼ねながら、こういうマークアップしていくような方針を作成しました。

そして、学会などにも呼び掛けて、最近、例えば去年、和歌文学会でこのTEIマークアップについてのシンポジウムが行われました。大変関心が高いところでもあります。

さらに、これは研究者にとっては、もちろん、これまでにない検索、分析が可能になるわけなんですけれども、それだけではなくて、学校教育にも大いに利用できるんじゃないか。このTEIによってマークアップしたものを、例えば教科書掲載、教科書に載るような古典をそうすることによって、自由自在に古典の言葉を引き出して——自由自在と言うと大げさか

もしれませんが、様々な観点から引き出してこられる。

そうすると、古典探求などの科目－なかなか古典で探求学習って難しいんですけども－実際に学生が自分の力で古典にアプローチしていける、そのきっかけになるだろうと私たちは考えております。もちろん、学校現場で使われるということは、一般市民の方にも大いに使っていただけますし、それから、海外でも大いに使っていただけるのではないかと考えております。

10ページへ参ります。これはTEIに限りません。国書データベース、画像データ、テキストデータ、全部含めまして、これを教育の現場で、国語の授業でどのように使っていかというのはこれからの大きな課題になろうと思います。左側の真ん中辺りに、大井将生さんという方－これは人間文化研究機構に所属しておられたんですけども、この4月から同志社大学に転出なさいました－こういう国書データベースを、教育現場でどのように使うかというような研究をしておられ、同志社大学に転出されましたけれども、続けて共同研究をしてもらっています。特にこの重点研究として、進めておられます。

私も教育に関心があるものですから、その研究協力者になっておりますけれども、ますますそういう要望は高くなっていくだろうと考えております。

中学校、高校と協働した取組ということで、右側に。これはもう既にやったことですが、紙すき体験－紙をすく体験、それから当館がやっております、今回、御説明は割愛いたしましたけれども、顕微鏡で本を見ることによって、いろいろな情報が分かってまいります。そこに髪の毛がすき込んであるとか。そうすると、髪の毛が分かると、髪の毛から当時の人の栄養状態であるとか、様々なことも分かってまいります。

それが、さらにこのDDHプロジェクトでは、微小部X線分析装置という設備を買うことができて、それによって、顔料であるとか、紙質であるとかを更に細かく分析することができて、例えば成立年代であるとか、そうしたものも科学的に解析していけると私たちは考えております。

こういったような多面的なものになって、あっちもこっちもという欲張りな計画でもあるんですけども、でも、その中から、とにかくトライアルを重ねていくということも、我々にとって非常に重要なことで、異分野融合ってそう簡単にはできません。理系等なりの他分野の方とウィン・ウィンでなければならないわけですから、そういったマッチングが大変大事でもありますし、成果が出るかどうかというのは、トライアルを重ねた上で出てくるものと考えております。

始まったばかりの当館の計画ですけれども、皆様の御支援を頂きながら、御理解を頂きながら進めてまいりたいと思います。

御清聴どうもありがとうございました。

○相澤主査

どうもありがとうございました。それでは、質問の時間に移らせていただきます。ただ今の渡部館長の御発表に関しまして、御質問等ありましたら、どうぞよろしくお願いいたします。どなたからでも結構でございます。

○石川副主査

よろしいでしょうか。どうもありがとうございました。大変勉強になりました。

3点、質問を差し上げたいのですが、まず1点目です。前回のプロジェクトでは江戸の後期まで、今回のプロジェクトで明治に範囲を広げて作業されていらっしゃるわけですが、記録すべき、あるいは確保すべきと考えておられる作品の数を100%といたしますと、現時点で画像化まで進んでいるもの、さらに、その先のテキスト化まで進んでいるものは、どのぐらいの比率になるのでしょうか。

○渡部発表者

それ、本当に大変痛いところを突かれたという感じなんですけれども、要するに、じゃあ、古典籍がどのくらいあるのかというのは、これもまだよく分からないわけです。まだ発見されているものもありますし、お蔵にあることは分かっている、それがどのくらいあるのか調査が及んでいないところもありますので、本当に分からないところもありますが。

仮にどうなんでしょう、100万点と考えれば、まだまだ道は遠いところがございます。そしてまた、あることは分かっても、なかなかその画像をまず許していただくまでにも随分長い道のりになりますし、さらに、それをテキストデータとして公開していいかとなると、またさらにハードルがあるわけです。総数として45万点までは行こうと思っています。それが半分ぐらいに当たるのか、なかなか明確に割合を申し上げることがしにくいんですけれども…。

ただ、御理解は急速に進んでいると言いましょか、今まで本当に壁が高かった、厚かったんですけれども、こういう事業を進めると本当に感じるのは、「ああ、あれですね」というふうに分かっていたいて、逆にあちらさんの方から、例えば静嘉堂文庫－三菱の

古書や古物などを集めた静嘉堂文庫などは、やってくれませんかという、あちらの方からお声掛けを頂くということも生まれていますので、とにかく一步一步進めていかざるを得ないなと感じているところです。

割合としては、その程度で申し訳ありません。

○石川副主査

分かりました。ありがとうございます。

質問の2点目です。このプロジェクトでは、いわゆるOCRでくずし字の自動解析をなさるということで、AIについての御言及もあったところですが、これは自主開発で作っていらっしゃるのでしょうか。そうであればどのような仕組みになっておりますでしょうか。

○渡部発表者

私たちがいろいろ開発しているところはありますけれども、そういう技術的な問題のところは、竹内さん、答えていただけますか。

○竹内随行者

私、OCRについて、すごく詳しいというわけではないんですけれども、私が知っているものとしては、国立国語研究所がいろいろジャンルであるとか、時代などで辞書の開発をしているので、そういうものを学習させて、OCRを掛けた結果、精度を見るとか校正に役立てるとか、そういうことをしているという話は聞いています。

○石川副主査

なるほど、ありがとうございます。

最後の質問ですが、これまで読めなかったものがテキスト化されて検索できるようになるというのは非常に意義深いことですが、その先として、より多くの人に読んでもらおうと思うと、やはり現代語訳が欲しいという気もいたします。こうした点に関しまして、現時点での御方針があれば、お聞かせください。

○渡部発表者

ありがとうございます。まさしく、そこが後継に当たるところと、私は考えております。現

在のところでは、直接的にその計画の中には組み入れておりませんが、しかし、その先には必ずそれが求められますし、考えていかなきゃいけないということで、この計画とはまた別筋で、私自身、進めています。現代語訳を付して、そして校訂本文―誰もが扱えるようなしっかりとした本文に現代語訳を付けて、そして、それをオープンデータとして公開していく。今、試みに「新古今和歌集」でそれをやっているんですけども、私の自分の専門ですが、ほぼそれができそうだという見込みが立ちました。これを公開することによって、こういう形で、「源氏物語」であるとか、「平家物語」であるとか、そうしたものにも広げていきたいと思っています。

更に言えば、先週、海外にこれから勤める人―この人は日本人ですけども―一番大事なのはやはり英訳だと。誰もが使えて、すぐ見られる。デジタルな形での英訳が必要であると。もちろん、それは仏訳や独訳もあればいいわけですけども、そうしたものもこれから作るということを考えなきゃいけないんじゃないかと言われて、そうだと。現代語訳さえあれば、AIで翻訳してくれるかなと思ったんですけど、それだけじゃなくて、確かな訳が必要ですね。

そこまでも視野に含めながら進めていきたいと私自身は考えております。

○石川副主査

分かりました。どうもありがとうございました。

○渡部発表者

それから、大変申し訳ありません。私、一番大事なところを飛ばしてしまったんですが、9ページのところです。構造化テキストをコーパスにお使いいただくよう提供するという、9ページのところに示したもの―ここのところ。これは、国立国語研究所と連携させていただいて、デジタル画像、OCR処理、そして校正テキストデータとしたもの、これを提供して、国語研の「OpenCHJ」というか、中納言コーパスの一つとしてこれを利用させていただくという連携を図っておりますので、このことを申し上げるのを忘れて、大変申し訳ございません。

○相澤主査

ほかはいかがでございますか。

○神永委員

すみません、2点、よろしいでしょうか。ちょっと細かなことではあるんですけども、5ページ目、廣瀬資料館－廣瀬淡窓の資料について御説明がありましたけれども、これは廣瀬淡窓の書いたものと、あと実際に廣瀬淡窓が持っていたものと、両方ということになるわけですか。

○渡部発表者

そうですね。

○神永委員

これは、ある程度テキスト化とか、そういうことは考えていらっしゃるのでしょうか。

○渡部発表者

もちろん、考えております。

○神永委員

既に進んでいらっしゃるかと。

○渡部発表者

進んでおります。

○神永委員

そうですか、分かりました。ありがとうございます。

○渡部発表者

それから、淡窓だけではなくて、この廣瀬家がずっと代々、咸宜園というのを守ってまいりました。

○神永委員

咸宜園、そうですね。

○渡部発表者

その資料も、もちろんございます。それもテキスト化を進めております。

○神永委員

ありがとうございます。あともう1点なんですけれども、今、最後にお話が追加でありました、9ページ目の国語研との連携というお話なんですけれども、テキストデータ化を行っている、それで提供、連携を図っているとおっしゃっていましたが、提供できるテキストデータというのは、かなりなものがもう既に蓄積されているのでしょうか。

○渡部発表者

その校正作業ですね。

○神永委員

校正作業中ですか。

○渡部発表者

そうです。

○神永委員

じゃあ、今後ということになるわけですね。

○渡部発表者

そうですね。

○神永委員

分かりました。ありがとうございます。

○相澤主査

ありがとうございます。ほかはいかがでございましょうか。

では、私から1点だけ、その発想が効率重視かもしれないんですけども、校正作業というのは非常に人手が掛かって、時間も要するというお話をお伺いしまして、その中で何を先に校正するべきかという重要度の見積りが重要で、そこにこのAI技術を使うということなのかなというふうにも感じたんです。

その辺りの戦略というか、重要度、優先度付与等についてはどのような形で。

○渡部発表者

まずは、コミュニティーの意見ですね。これは重要だろうと。もちろん、私たち自身も判断いたしますけれども、これは是非してもらいたいというような意見も広く聞かなきゃいけないと思いますし。ただ、また、それを私自身がそれを判断していかないと。

確かにその重要度はなかなか大事な問題だろうというふうに思います。やはり、それは研究重視だろうと思いますね。この資料は是非開いていきたいというようなことだと思います。

○相澤主査

なるほど、ありがとうございます。

○齊藤(美)委員

順天堂大学の齊藤と申します。お話、聞かせていただいて、ありがとうございました。大変勉強になりました。今、ちょうど相澤主査から校正のことのお話、御質問があったんですが、私もその点に関してお聞きしたいなと思っておりました。今お話の中で、ものにもよるけれども、92%は読み取れるというふうに伺いまして、残りの8%は人の手で直していくと。利用者側、ユーザーとしては、完璧に仕上げていただいたものが公開されるのはとても有り難い、うれしいことなんですけども、とても時間と労力が掛かるだろうなと思います。

公開するに当たっての精度、クオリティーの判定というか、どこまで求めるというか、あるいはタイミング、早さよりも精度を重視するというお考え、どういうところに基づくのかなと思いました。今日のお話の中から推察できる部分もあるんですけども、直接伺えましたら幸いです。

というのは、例えば国立国会図書館のネクストデジタルライブラリーなどでは、精度が

100%ではないテキストデータも使わせてもらえるようになっています。私も使っていると、自分で直さなきゃいけないな、みたいなどころはあるんですけども、ただ、100%じゃなくても、それが利用できることはとても有り難かったりもするので、どのように考えていらっしゃるのかと思って、お尋ね申し上げます。よろしくお願いいたします。

○渡部発表者

これも利用の仕方によるだろうと思うんですけども、精度の問題です。確かに本当にそこが問題で、恐らくAI技術が更に進んでいったりすれば、このパーセンテージはかなり自動的に上がっていく部分もあるんでしょうけど、絶対100%にはならないだろうと言われています。

研究者によって、あるいは研究課題によって、本当に最高度の精度でないと困るという研究と、それから、本当に今の現状の92%でも、これは何とかなるんだとー例えば大きな、私たちは話形と呼んでいますけれども、例えば物語には一つのパターンがあるわけですよー貴種流離譚^{たん}のような。そのパターンを抽出していきたいんだというときには、92%でも、ある種、92%でも十分対応可能なわけです。

ただ、また和歌の研究のように、一字一句が非常に微妙な問題を提起するような、そういう場合には、100%近くないといけない。その辺りが判断のしどころで、この辺りも研究者コミュニティとのやり取りの中で了解を得ながら、これは、こういう精度であるということとを納得してもらって、でも、可能な限り早く、100%のものではないけれどもということで、出していければとは思っております。現在、何%になったらという形では、ちょっと申し上げにくいところがございます。

○齊藤(美)委員

ありがとうございます。

○渡部発表者

どうもありがとうございます。

○相澤主査

ありがとうございます。ほかはいかがでございましょうか。

○石川副主査

もう1点よろしいでしょうか。この手の事業というのは非常に時間が掛かりますので、私どもの過去の審議におきまして、短期でなく長期の支援、特に、中で仕事をされる若い方がキャリアパスを見据えられるような、十分な長期にわたる継続支援が必要ではなかろうかという話が出ておりました。国文学研究資料館におかれては、10年のプロジェクトを2回連続で取られているということで、通常に比べると長めになっているかと存じますが、逆に言えば、10年たつと支援が打ち切られてしまう可能性もあるということかと思います。一般論として結構ですが、今回のように、広く国民に資する基盤的言語資源を作るような仕事に対する国の支援の在り方、特に、その時間的なスパンの在り方について、館長の御意見があれば、お聞かせください。

○渡部発表者

ありがとうございます、こちらの意を酌んでくださって。

○石川副主査

是非伺っておきたいところでございます。

○渡部発表者

全くそのとおりだと思いますね。本当に長期的支援、私たちの仕事というのは、本当に50年、100年を単位で考えておりまして、それを10年で成果を上げて、だけど、もう10年たったら出さないよというのは困るわけで、本当に私たちも大変な困難に直面するわけです。

だから、3期目と申しましょうか、さらにこの後継計画も考えていって、その中で、私たちがやっていることを位置付けていただかなきゃいけないだろうなどは考えております。現実がそうなっている以上、そうせざるを得ないですが、確かにもっと長期的に支援していただきたいというのは、強く思います。

その長期的な次の目標として、先ほど言っていた現代語訳、英訳、外国語訳、こうしたものは次の課題として必ず入ってくるものだと、私たちは考えております。

○石川副主査

ありがとうございました。

○相澤主査

ありがとうございます。ほかはかがでしょうか。

じゃあ、1点追加で私からも。非常に勉強になるお話で、いろいろ思いは広がる場所なんですけれども、異分野融合という枠組みで、いろいろなところ、学問分野と連携していくのと同時に、関連分野融合と言いますか、非常に近い分野とどうやって力を合わせて、一つの日本文化というカテゴリーでやっていくかというのも重要かと思っています。

卑近な例として石碑のことを思い浮かべての質問ですが、そういったところで、何か取組とか、思いとか、お考えとかがありましたら、お聞かせいただけますか。

○渡部発表者

ありがとうございます。本当に売り文句にやや傾いて、異分野融合という、この言葉を看板に掲げましたけれども、関連する分野の、例えば我々で言えば美術史、思想史、もちろん歴史は当然ですね、そうしたものの関連というのは元から深いですし、さらにそれはより強固にしていかなきゃいけないだろうなと思っております。そのためにも、隣接分野の方々がずっと国文研に入ってこられるように、そのための基盤作りというのが一つ、大きいだろうなとは考えております。

実際に、具体的に今ぱっと出てくるといいんですが、地理情報、これは理系ももちろん関わるわけなんですけれども、歴史、人文地理学、それから自然地理学も関わります。それから、今一つやっているのは食です、食べ物。こうしますと、これは栄養学やらが、関わってまいりますし、それから、もちろん人文系の問題とも、様々、文化の問題とも深く関わるだろうというふうに思っています。

これは結構大きな科研で、私たちが主導しているわけではないですけれども、でも、そういう試みもあったりいたします。そのほか、もちろん歴史や美術史との関わりも。そうですね、美術史との関わりでいうと、先ほど微小部X線分析装置のお話をいたしましたけど、例えばその顔料から成立年代を分析していくとか、そういう方をこの4月、東京藝大にいらっしゃった方をお迎えして、そういう研究分野も中心になって進めているところです。

○相澤主査

ありがとうございました。

○渡部発表者

どうもありがとうございました。

○齊藤(美)委員

すみません、齋藤からもう1点、よろしいでしょうか。再びお願いいたします。

今のお話もありましたけれども、マテリアル部分も含めてデータ化できる、研究できるというところも大変興味を持ったんですけども、この小委員会－言語資源小委員会としては、コーパスをはじめとしたデジタルの部分が特に注目されてきたと思っているんですけども、そういった髪の毛であるとかが付いているといったような、もしかしたら、私も今日のこのお話を聞くまでは、デジタル化は難しいのかなというような部分さえもできるということであれば、元々の紙であるとか、元のデータというのは、もしかしたら、一回、全てデジタルデータ化できたら、要らなくなるというような考え方もあり得るのかなと－私は思わないんですけども、それだけデジタルの方が、今世の中で注目されてきている中で、恐らく国文学研究資料館の皆様は、それでも生のデータも保存をきちんとなさっていると思うんです。その生データを保存し続ける意義というか、それを一般社会の方に伝えるというのは、どのようにしたらよろしいものか、お考えを伺えたら幸いです。よろしくをお願いいたします。

○渡部発表者

ありがとうございます。本当にそれもまた大きな問題だろうと思います。私たちは、少なくとも私は、生の資料－生のデータと言うのはおかしいですね－生のもの、それは不必要になることは絶対ないだろうと思いますね。データに限界があるというのは、それはどんな情報でも、情報だけを取れば限界があるので、生のものや、あるいは人間というもののとの関わりの中で、必ずそのデータというのは生きてくるんだろうと思います。

むしろ、データがこれだけ積み重なってまいりまして、利用ができるようになってくると、ますます、じゃあ、人間ができること、人間の直観、例えば私たち、和本－本当にその道の大家は、本を持った瞬間に、これは何時代のものだって分かる－その重さから、既にいろいろな情報を酌み取る。これは、さすがにデータ化できないわけですよ。何グラムという

ことは書けるかもしれませんが、その質感というんですか、この質感も、更にデータ化していける方向があるなんていうことは、最近言われているようですけども。でも、やっぱりそれは人間との応答関係というか、人間がそれに何か問い掛け、そして、向こうからの情報で、またそれを往復運動で考えを深めていくということがあって、生きてくるんだろうなと思っていますので、生なものの意味というのは決してなくなる。むしろ、大切さが深まるのではないかと考えています。ちょっと甘いかもしれませんが…。

○齊藤(美)委員

いえ、よく分かりました。どうもありがとうございます。

○渡部発表者

こちらこそ、ありがとうございます。

○相澤主査

ありがとうございます。ほかはいかがでしょう。よろしいでしょうか。

(挙手なし。)

では、これにて質問の時間につきましては終了いたしたいと思います。改めて、御講演いただき、どうもありがとうございました。

○渡部発表者

どうもありがとうございました。

○相澤主査

引き続きまして、意見交換がございますので、是非御意見も頂戴できればと思っております。

それでは、参考資料2の前期取りまとめた内容がお手元にあると思いますが、「今後における日本語のデジタル言語資源の整備・活用の在り方(報告)」ということで、こちらで取りまとめた内容についても踏まえつつ、渡部館長から御紹介のあったデジタル言語資源の活用の事例、テキストの構造化の話などを踏まえまして、御自由に御発言いただく時間とさせていただければと思います。

特にこれについてということではなく、それぞれのお立場からの自由な感想ということで、また恒例に従いまして、皆様、全員から御発言いただく形で進めてまいりたいと思いますので、御準備をよろしくお願いいたします。

また、本日御出席の御陪席の方々も、是非御自由に御発言いただけましたら、幸いです。

そうしましたら、まずは自由な御発言ということで、この報告をまとめましてから二、三か月ということで、その間の変化も踏まえつつ、何かございましたら、是非御発言をよろしくお願いいたします。よろしいでしょうか。

なお、そうしましたら、大変恐縮ながら、まず前川委員に御意見、御発言いただければと存じますが、よろしいでしょうか。

○前川委員

今の国文研の渡部先生のお話についてということによろしいでしょうか。

○相澤主査

それでも結構ですし、さらに、枠を広げて、何でも大丈夫という時間でございます。

○前川委員

はい。今、渡部館長のお話は、プロジェクト自体については、我々、言わば兄弟というか、姉妹の機関ですので、以前から何遍も説明をしていただいているので、比較的よく理解しているつもりでおりますが、今日、改めてお話を伺っていて、私どもがこれまで蓄積してきた、例えば形態素解析の技術のようなものが、国文研のお作りになったテキストと結びつくことによって、今後、それこそいろいろな異分野融合、それから関連分野の融合、そして、最終的にはプロジェクトのタイトルにある課題解決につながるんだろうなと思いました。

つまり、単に読み取っただけではなかなか難しいんだけど、そこに形態素の情報とか、そういったものが付くことによって、読み取りの精度が逆に上がるということもありますし、それから、さらに翻訳とか、そういう可能性が出てきて、それが様々な、例えばその関連、美術史とか、歴史とか、民俗学とか、そういったところに影響を与えていくんだなということを、今日のお話を伺いながら感じておりました。そういう点で、是非お力になりたいと考えております。

それから、あともう一つ、OCRの読み取り精度の話がございました。あれは、例えば私どもの昔の経験をちょっとお話いたしますと、2000年代－2006年とか、2007年とかの頃にBCCWJを作っておりました頃の読み取り精度というのは、もちろん、これは現代語ですけれども、現代語のきちんとした活字の、くずし字でもない、ちゃんとした活字なんですけれども、それでも当時97%ぐらいだったんですよね。

それで、全く同じ問題がありまして、こんなのでは使えないとおっしゃる方もかなりおられた。だけど、私はそうは考えておりませんでした。もちろん、コーパスの一部、全体の5%ぐらいに限っては手作業で精度を上げて、99%ぐらいを目指して、実際それに近い数字まで上げるマニュアル、手作業で上げていったんです。

だけど、全体としては98ぐらいの、それ以外の部分は97とか、98ぐらいの精度で公開をしました。現実はどうだったかというと、いろいろな意味で、それでも非常に活発に使われているわけなんですよ。それから、その後のもちろん、技術的な進歩が素晴らしいものがありますから、今やれば、全体としてもっと高い精度が最初から出るということは確実なわけです。

本当にそういう意味で、今日の渡部館長のお話もそういうことだったかと思いますが、いろいろ言われますけれども、是非公開した方がいいです。広く使ってというのが、税金を使っている立場としては正しいし、科学的にも正しいと私は考えております。

最後にもう1点。これは全く個人的な感想なんですけれども、TEIがありますね、Text Encoding Initiative、私はあれ、非常に使いづらくて、あまり好きじゃないんですけども、今日の廣瀬本の萬葉集ですか、あれのお話、例を伺って、私、不勉強でそういうものがあるということ自体知らなかったんです。廣瀬淡窓文庫のことは聞いておりましたけども、萬葉集というのがそういう注釈が積み重なったものだということは、初めて知りました。そういうのに対しては、TEIは確かに有効だなと思いました。それは大変、認識を新たにいたしました。全く個人の感想ですが、その点は勉強になりました。ありがとうございました。

○相澤主査

ありがとうございました。では、強制的になりますが、武田委員と森委員に御意見を伺ってまいりたいと思います。やや唐突となりますので、私からその前に、1点、御質問を兼ねてということで。

この日本の文化としての発信についてなんですけれども、最近、生成AIも進んで、「鳥

獣戯画」風のコンテンツを自動生成するということではできている状況の中で、やっぱりオリジナルなものと、それっぽいものというのは違うものであるとの考えを持つ人も予想されます。一方でしかしながら、日本文化の発信の手段としてオリジナルが一種のメディアになるという意味では、もしかしたら、日本としては歓迎すべきことなのかもしれないという考えも理解できるため、かなり複雑な問題になっていく気がするのですけれども、その辺りについて、何か御意見がある委員の方がいらっしゃいましたら、あるいは、渡部館長から何か御意見がありましたら、お願いできればと思います。いかがでございましょうか。

○石川副主査

相澤主査の御指摘の点ですけれども、「鳥獣戯画」にしても、オリジナルがきちんとあるので、それを応用して「鳥獣戯画」風の絵を作ることができるわけで、恐らく古典籍についても同じようなことが言えるのではないのでしょうか。オリジナルをしっかり守り、それらをきちんと正しくテキスト化していく。このことが実現して初めて、バリエーションと申しましょうか、古典文化の応用的活用という波及効果も見込めるのではないかという思いを持ちました。

○相澤主査

ありがとうございます。そういう意味でも、しっかりとデータベース化して、広く公開していくというのは価値がある、別の意味でのオリジナリティーを担保するという価値もあるのかなと思ったりして、お伺いしておりました。ありがとうございます。

では、武田委員、よろしくお願いいたします。

○武田委員

本日の渡部館長の御説明について、大変参考になりまして、ありがとうございました。長年にわたる国文学研究資料館の取組について、改めて敬意を表しますとともに、今後について期待したいと考えております。

近代初期までが対象ということで御説明を伺いましたが、それ以降、現代に至るまでの期間についても、様々な機関、研究を進められて、網羅的にそういったコーパスの構築が進むということを期待しております。

簡単ですけれども、以上です。

○相澤主査

ありがとうございます。では、森委員、よろしくお願いします。

○森委員

今年度から委員に加えていただきました、日本テレビの森と申します。よろしくお願いいたしますします。

今日は、勉強になるお話をありがとうございました。

なかなか直接見たり、もちろん触れたりすることができない貴重な資料について、検索が掛けられる状態で、しかも、遠く離れたところからでもアクセスできるということの有り難さを感じました。先ほど前川委員からもお話がありましたが、完璧な状態で出していたくのも確かに大変有り難いんですが、完璧じゃなくても、もう今の段階で、あるんだっから見せてくださいとも感じておりました。

言語コーパスの方もそうなのですが、主に、国語学、国文学の研究者の方々がメインで使っていらっしゃると思うんですけども、例えば研究という名のところに足を踏み入れる大学生よりも前の段階、中学校とか、高校の段階から触れること、「こういうものがあるんだ」、「ここにこんな世界が広がっているんだ」というのを知ってもらうことで、またこの分野に入っていきたいと思う人も増えるでしょうし、便利に使える人たちも増えるでしょうから、マスコミに身を置く者としては、そういうことも、皆さんに知ってもらえる活動をしていきたいなと思いました。

ありがとうございました。

○相澤主査

ありがとうございます。では、今日、実はたくさん時間がございますので、まだ御発言いただけるようでしたらと思いますが。

○神永委員

今日は貴重なお話をありがとうございました。私は、日本国語大辞典という辞典の編集を担当していた者で、特に日本国語大辞典の用例の部分の担当をしておりました。用例を担当するということは、日本の文献をいろいろと調べなければいけないということがご

ざいまして、実は国書データベースは今でもほとんど毎日のように活用させていただいています。

私、日本国語大辞典の第2版を担当していたんですけれども、そのときは、国書総目録だったものですから、これって書籍、書物の文献の読みが何というふうに読むかが分からないと、なかなか引けないという大変な欠点といいますか、当然なんですけれども、あったんですけれども、それが今、国書データベースになったことで、読めなくても引けるというところがありまして、これは非常に有り難いと感じております。

本当に、正にいつでも、どこでも、どなたでもというようなことをおっしゃっていただけども、正にそれなんじゃないかなと思っております。

今、日本国語大辞典－私は会社を退職しておりますOBなんで、現場からは離れていますが、いまだに編集のお手伝いはしております。日本国語大辞典というのは、用例が命というような考え方の辞典なものですから、その用例のための生の資料がいまだに残されているというのはとても魅力で、そういったものの中から一杯用例をすくい上げたいというのが今後の課題だと思っているんです。

今回お話を伺って、日本国語大辞典－2032年に第3版を刊行するようにしているんですけれども、是非国文研と共同作業、連携をさせていただけたら有り難いな、なんていうようなことを思っております。

日本国語大辞典のような辞書は、文献データベースの一つ入り口になるんじゃないかなというようなことを考えておまして、その入り口から、実際のテキスト化されたデータに入って、研究者、さらには一般の方もかなり日本のそういったものを理解できるんじゃないかなということを考えておりますので、是非そういったことが可能になればいいかなんていうことを思っております。

ここで交渉するわけにもいきませんが、是非そういったことをお考えいただけると有り難いと思いますし、それから10ページ目にいろいろなところの連携の話が載っていましたが、この中に日本国語大辞典というの載るようになると有り難いかなと思っておりますので、今後とも是非よろしく願いいたします。

どうも今日はありがとうございました。

○渡部発表者

ありがとうございます。

○相澤主査

ありがとうございます。では、齊藤(美)委員、いかがでしょうか。

○齊藤(美)委員

ありがとうございます。今日、お話を伺っていて、いろいろなことを本当に学ばせていただいたんですけども、そのうちの一つとして異分野融合の件、さっきの、去年のこの小委員会でも度々話題になっていて、それはなかなか難しいよねということになっておりました。

今日、その具体的な方法、事例を複数示していただいて、大変有り難かったです。その中で、特にびっくりしたというか、驚いたのが、観光と結び付けるところです。確かに、私も岡山出身の翻訳者の研究をしていたこともあるんですけども、そうすると、地元に行くと、地元の名士、作家だったり、翻訳者だったり、地元の方が本当に大事にいらして、顕彰をしていて、^た讀えていて、いろいろイベントをなさったりしているので、そういう地元の方々のニーズともマッチするなと思って、大変刺激を受けました。

今後、何か議論や自分の研究の参考にさせていただければと思います。ありがとうございました。

以上です。

○相澤主査

ありがとうございます。では、渡部館長から何かコメントですとか、是非これをというようなことがございましたら。

○渡部発表者

まず、聞いていただいただけでも本当に有り難い上に、今日は本当に勇気付けてもいただきましたし、それから、これからの方向性というものを考える上でも大いに参考になりました。実際に具体的に、本当に連携の道筋をお示しいただいたところもあり、本当にありがとうございます。是非、これからも国文研、そして、このプロジェクトをよろしく願います。

○相澤主査

ありがとうございます。では、石川副主査。

○石川副主査

配布資料の最後の10ページのところで、古典籍と教育の関係の話がございます。くずし字を中等教育で読ませるというのも面白い試みかと存じます。この点に関して、現在、次期指導要領の方向性の議論も進んでいるところですが、古典籍の資料を国語教育のコンテンツとしてより広範囲に取り込んでいたり、あるいは、くずし字の読み方を指導項目の中に入れていたりするような動きというものはあるのでしょうか。

○渡部発表者

そういう動きが確かにあることはあります、学会の中に。

○石川副主査

そうですか。

○渡部発表者

私自身は、これまた勝手なことを言うと、怒られちゃうんだけど、余りくずし字というのをそんなに表立てなくてもいいんじゃないかというのが一つ、基本的にあるんです。それとともに、要するに指導要領の話になりますと、自分もそこに参加していたことがあって、盛り込み過ぎているんじゃないかという反省がずっとありますので、更にコンテンツを増やすのはどうだろうなというのも、一つあります。

ただ、この大井将生さんを取り上げました授業のビデオを見せていただいたんですけど、これは初等教育です。小学校1年生から6年生のずっと授業があって、そこで最初に「奥の細道」の原典を示すわけです。そうすると、我々なんかよりも、はるかに関心度が高いというか、ぱっとは実は入っていけるんです。

それから、音とか、そういう図柄というものに対する感度が、小さい子は我々なんかよりもはるかに高い。我々はどうしても、これは読めないものだとか、縁遠いものだと思ってしまいうんですけれども、むしろ彼らは形から入るんでしょうね。そこをうまく使って、くずし字を読むということ自体よりも、そこを入り口にして関心を高めていくという、そういうやり方は確かにあるんだなということは実感した覚えがございます。

○石川副主査

ありがとうございます。

○相澤主査

ありがとうございました。では、ほかにもしありましたら、このタイミングでお願いしたいと思いますが、いかがでしょうか。

○石川副主

一つ、整理のコメントと言いましようか、感想なのですけれども、今日、渡部館長のお話を伺っていて、私たちが考えないといけない、ある種ターゲットというのが2種あります。

一つは専門研究者、専門研究者の信頼に応える資料を作るとのことと、もう一つは、それを超えて、一般国民と言いましようか、広く国民にその資源の価値を伝えていくという、この二つのある種の仕事の目的と言いましようか、事業の目的というもののバランスを取っていくということが、恐らく古典籍においても重要でありましようし、あるいは、一般的な言語資源ですとか、デジタル言語資源についても同じようなことが言えるんだらうなということ、今日お話を伺って感じていた次第であります。

ありがとうございました。

○相澤主査

ありがとうございました。それでは、本日の議題は以上となります。

本日頂いた御意見につきましては、前期同様、整理しまして、次回以降の御意見も加えまして、まとめていきたいと考えております。

また、今後、この小委員会の委員の方が関わっておられる取組についても、事例紹介としてお話いただくことを考えておりますので、またよろしく願いいたします。

それでは、最後に事務局から御連絡をよろしく願いいたします。

○鈴木国語調査官

それでは、次回の言語資源小委員会について、お知らせいたします。次回、第8回、8月4日、月曜日の10時から12時で予定しております。この会では、委員の方に事例紹介を

お願いするという予定でおります。詳細は改めて御連絡いたします。よろしくお願いいたします。

以上です。

○相澤主査

それでは、本日の言語小委員会、こちらにて閉会となります。どうもありがとうございました。