第6回 国語分科会言語資源小委員会（Web開催）議事録

令和7年2月17日
10時30分～11時35分
文部科学省16階・16F4会議室

〔出席者〕

（委員）相澤主査、石川副主査、植木、神永、齊藤（美）、武田、
前川各委員（計7名）
（文部科学省・文化庁）村瀬国語課長、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

〔配布資料〕

- 1 第5回国語分科会言語資源小委員会議事録（案）（委員限り）
- 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）（素案）

〔参考資料〕

- 1 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）
- 2 言語資源の在り方に関するこれまでの議論の整理
- 3 ジャパンサーチの概要
- 4 国文学研究資料館「データ駆動による課題解決型人文学の創成プロジェクト」

〔経過概要〕

- 1 事務局から定足数の確認が行われた。
- 2 事務局から配布資料の確認が行われた。
- 3 前回の議事録（案）が確認された。
- 4 今期最後の言語資源小委員会の開催に当たり、村瀬国語課長から挨拶があった。
- 5 事務局から配布資料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）（素案）」、参考資料1「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）」、参考資料2「言語資源の在り方に関するこれまで

の議論の整理」、参考資料3「ジャパンサーチの概要」、参考資料4「国文学研究資料館「データ駆動による課題解決型人文学の創成プロジェクト」」について説明があり、質疑応答及び意見交換が行われた。意見を踏まえての配布資料2の修正及び国語分科会における報告については、相澤主査に一任することが了承された。

- 6 今期最後の言語資源小委員会に当たり、各委員から感想等が述べられた。
- 7 今期最後の言語資源小委員会閉会に当たり、村瀬国語課長から挨拶があった。
- 8 今期最後の国語分科会は、ローマ字小委員会における議論の状況を踏まえて、令和7年3月中に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。また、今回の言語資源小委員会の議事録については、委員による2回の確認を経て公開することが確認された。
- 9 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。

○相澤主査

定刻になりましたので、ただ今から、第6回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。

本日も、文部科学省の会議室で御出席の皆様と、オンラインで御参加くださっている方々とのハイブリッド開催となります。また、傍聴者の方々もオンラインでこの会議を御覧になっていることを御承知おきください。

事前に資料を御覧いただいておりますとおり、本日はこれまでの審議内容を踏まえて、今期の議論を取りまとめたと考えております。この言語資源小委員会で、次回の国語分科会にお示しする審議のまとめ(案)のための素案の検討となります。それぞれのお立場から、幅広い御発言をお願いいたします。

では、本日の言語資源小委員会の審議に当たりまして、村瀬国語課長から御挨拶をお願いしたいと存じます。

○村瀬国語課長

本日は、年度末のお忙しいところ、皆様、御出席を賜りまして、誠にありがとうございます。本日、取りまとめに向けた御議論ということになるわけでございます。

ので、どうぞ精力的に御審議をお願い申し上げまして、御指導、御鞭撻^{べんたつ}のほどよろしくお願い申し上げます。

○相澤主査

ありがとうございます。本日の議事は、(1)言語資源の在り方、そして、その他となっております。

まず、言語資源の在り方につきまして、この言語資源小委員会の今期の審議まとめ(案)を取りまとめるという趣旨でございます。まずは審議の対象となる配布資料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)(素案)」と、参考資料1、2、3、4について、事務局から御説明をよろしく願います。

○鈴木国語調査官

配布資料2及び参考資料1～4について御説明申し上げます。

言語資源小委員会の審議内容については、参考資料1において、その論点を初めにお示しいたしました。そして、参考資料2という形で、審議で出された御意見をまとめさせていただいたところでございます。

この参考資料3と参考資料4は、新しい情報ではありますけれども、この内容も盛り込んだ形で報告の素案としようと考え、先取りした形で配布資料2の最後のところで触れております。

では、配布資料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)(素案)」を御覧ください。こちら、「素案」と書きましたのは、この報告というのは国語分科会から文化審議会に報告する形でございますので、国語分科会において審議する段階が「(報告)(案)」となります。ですから、その前のこの小委員会の段階では「案の案」ということで、「素案」とさせていただいた次第でございます。

この素案をまとめるに当たりまして、核となる用語について整理いたしました。その点の確認をさせていただければと思います。その核となる用語というのは、「言語資源」、「デジタル言語資源」、そして「言語コーパス」です。この三つの用語がこの「(報告)(素案)」の中には出てまいります。この相互の関係についての確認

を、初めにしておければと思っております。

まず、「言語資源」でございますが、これはアナログのもの、例えば書籍ですとか、古文書のようなものも含めてでございますが、言語が使われている資料、それらの総称として「言語資源」という用語を用いております。ですから、最も広い概念です。そして、このアナログのもの等を含めた「言語資源」、それをデジタル化したもの、これが「デジタル言語資源」でございます。ですから、「デジタル言語資源」は、「言語資源」に内包されるとお考えください。次に、この「デジタル言語資源」のうち、形態素分析などの情報が付されたテキストデータこちらを「言語コーパス」と呼んでおります。ですから、「言語資源」が最も広い概念、その中に「デジタル言語資源」がございまして、その「デジタル言語資源」の内側に「言語コーパス」があるという関係であるということで、この用語を用いているということを初めに確認させていただきます。

12月の国語分科会で御報告いただきましたのが、参考資料2でございます。この段階では、柱立てに基づいて、委員の皆様のお発言、それから資料に示された内容を箇条書にして並べておりました。今回、この「(報告) (素案)」では、それらを文章化いたしました。そして、構成の方も「はじめに」を付けまして、「基本的な考え方」、「言語コーパスの現在」、「言語コーパスの今後」、「将来社会を見据えた考慮事項」と来て、「おわりに」—そのような構成にいたしました。

「参考資料」として委員名簿、審議経過、それから、前期国語分科会がまとめられたデジタル言語資源に関する見解、それから、石川副主査の御報告内容と、そこで行われた意見交換の内容をまとめた言語コーパスの社会実装の実際の例—そうした参考資料を添えた構成にしております。

まず、1ページの「はじめに」でございます。ここは、言語資源小委員会での審議を行うに至った経緯を紹介している部分でございます。前期の国語分科会でまとめられました「見解」の中から、深く関係する部分を引用しまして、最後に今後の更なる検討を求めているというところを受けて、この小委員会での審議が始められたんだというまとめ方をしております。

1ページの後半には、「基本的な考え方」を立てております。ここでは、言語の働きや形態、言語資源がいかなるものであるかという認識を示した上で、検討の観点を提示しているという、そういう部分でございます。言語の働きや形態に関しては、

言語というものが人間の思考、それから他者との意思疎通において不可欠な存在であること、そして、アイデンティティと深く関わること、文化芸術の基盤を成すこと、それから、生きる力の土台となること、歴史と伝統の中で培われてきたものであること—そのような認識をまず掲げまして、書き言葉から話し言葉に至る多様な形態が認められるものであるとまとめております。

そして、このような言語についての認識の下で、言語資源というものがデジタル化されることで、様々な研究、政策、教育、そして生成AIの基盤となる大規模言語モデルの開発などにとって有用なものであるということ、言語資源そのものが、言語が使われた途端、過去のものとなっていくという性格を持っておりますので、それを記録・保存しているものであるということ、つまり、それは潜在的な文化財としての価値がある、言わば「未来へ伝える文化遺産」と位置付けられるという提案をした上で、この多様な言語コーパスの構築、そして、それを活用していく土壌を整えていく必要があるんだという形でまとめているところでございます。

その審議結果ということで、「言語コーパスの意義・現況」、それから、「言語コーパスの活用事例」、「生成AI等社会の変化、進展がもたらす影響」、「今後に向けた言語コーパスの整備・運用の方向性」等の柱の下に、記述が進んでまいります。

2ページに行きまして、「言語コーパスの現在」という見出しの下、「言語コーパスの意義・現況」と「言語コーパスの活用事例」という二つの小見出しの下にまとめております。

「言語コーパスの意義・現況」は、言語コーパスの定義と意義、それから国立国語研究所の言語コーパスなどの整備状況をお示ししたところでございます。「言語コーパスの活用事例」については、実際どのようなことで活用されているのか、また、活用可能なのかということ日本語、それから英語、フランス語、ドイツ語について紹介をしております。

言語コーパスとはということで、実際の言語体系を体系的に記録し、電子的に検索可能な状態にして公開した言語資料—これは前川委員の御発表の中で示された定義を使わせていただいております—そして、その要件として七つ挙げられておりましたので、それを紹介した形といたしました。

国立国語研究所においては、四半世紀にわたり、日本語の言語コーパスの整備に取り組んでいらっしゃいました。現在では、おおむね奈良時代から現代に至る日本

語の歴史的な変化をたどることができ、書き言葉、話し言葉の多様な言語使用域も把握できるものとなっているという現状を紹介した上で、書き言葉における代表的なコーパスとして、「現代日本語書き言葉均衡コーパス（BCCWJ）」について触れております。これには2006年以降のデータが収録されていないという課題を抱えているということから、現在の動きとして、2006年から2025年までのデータの追加が順次行われていく、そのような事業が進行中であるというところまで紹介しております。

コーパスの意義につきましては、多様な文脈や経年による違いのある言語を研究するためのデータ提供、計算機で言語を処理するための学習データとして活用できるだけでなく、ある時点の言語を記録した潜在的文化財としての価値があると述べた上で、言語コーパスを作る上においては、著作権の問題もあるので、そのことについては、ここで触れているという形です。

3 ページの最後のところに進みます。言語コーパスの活用事例を、日本語の場合は、国語施策における例、それから辞書編集における六つの活用が考えられるという御紹介がありましたので、それを紹介しました。そして、教育において、また、学術研究における活用というところを挙げております。

英語に関しては、代表的な言語コーパスを紹介した上で、非常に幅広いユーザーが想定されたインターフェースが準備されていることですか、コーパスアプローチという形で、社会的な課題の解決に導くという社会実装の例ということで、ヘルスコミュニケーションですか、企業コミュニケーション、ヘイトスピーチの分析などを挙げているところでございます。

フランス語の場合も、代表的な言語コーパスを紹介いたしまして、その上で小規模・中規模コーパスの一元化と、その一元化をする目的というところが紹介された上で、活用事例としては、辞書、学習教材、それから、学習の評価システムの開発、それから地方言語の研究など、御発表で紹介された事例を挙げております。

ドイツ語の場合も、代表的な言語コーパスを紹介した上で、特にドイツ語の正書法の改正に深く関わっているということ、それから方言研究で活用されているという辺りの紹介をいたしました。

続いて、6 ページでございます。「言語コーパスの今後」。こちらでは、生成AIに代表される社会の変化・進展がもたらす影響、それから、今後に向けた言語コー

パスの整備・運用の方向性を述べている部分でございます。ここでは、日本語が変化していく要因としてのサイバー化、高齢化、多文化・多言語化、それから生成AIの影響などが考えられるということで、日本語の変化する要因というものをまず挙げた上で、その中でも、特に生成AIの影響について触れているところでございます。

大規模言語モデルがどのようなものかという整理をした上で、この大規模言語モデルというのは言語データの中に埋め込まれた知識等を獲得しているという方向と、生成されたテキストが内包するバイアスが人間社会に影響を及ぼすという方向—双方方向の関係にあるということ、それから、学習用のテキストデータの不足ということが指摘されているという問題—AIによって生成されたテキストが学習データとして活用されていくことが予想されるということを紹介しております。つまり、ここでは学習データのテキストの品質というものの保持という課題が大きな問題になるというところで、言語コーパスというものの構築というのがそこに深く関わっていくというところまで言及しているところでございます。

そして、今後の言語コーパスの在り方については、取組の継続、普及、そして社会実装という、この3点が重要になるということです。取組の継続については、言語変化を追うためには、同じサンプリング基準で継続的にデータを収集する必要がある、その実現のためには、言語コーパス開発に取り組んでいる機関への継続的な支援が必要であるというような整理をしているところでございます。

普及に関しては、言語研究者や辞書編集者のためだけでなく、一般ユーザーの使用を視野に入れて、言語コーパスの価値を実感できる工夫というものが重要だとしています。その中では、例えば辞書における言語コーパスの活用を進めて、その意義を発信していくことで、言語コーパスというものを多くの人に知ってもらい、そこに触れてもらうというきっかけを作るということですか、使いやすいインターフェイスの開発などを求めているところです。

社会実装としては、意思疎通の手段である言語が人間関係の構築とか、信頼感の形成に関わりますので、円滑なコミュニケーション基盤の構築を必要とする教育、医療、社会福祉、子育て、通訳、言語運用などの現場で活用できるように社会実装に結び付けていくことを考えていく必要があるという整理をしているところでございます。

8 ページに参ります。「将来社会を見据えた考慮事項」—ここでは、言語コーパスというのが、言語研究や言語政策にとって有用なだけではなく、社会課題の解決に資する役割も期待されているということを述べている部分でございます。そのために、今述べた継続的支援ですとか、普及策の工夫、そして、コーパスアプローチの実践というところまで触れた上で、さらにコーパスの開発ですとか、コーパスのメンテナンス自体の意義というものをしっかりと認識していく必要があるということ、それから言語データの円滑な収集、多様な言語の蓄積、人材育成が求められる、そのような留意点を整理したところでございます。加えて、日本語母語話者でない人々の話す日本語の言語コーパスの整備ですとか、生成AIによる日本語への影響についての評価の在り方、そうしたところが今後の研究対象となっていくであろうと最後触れています。

最後、「おわりに」です。こちらは言語コーパスというよりは、むしろ、幅広いデジタル言語資源というところで、デジタルヒューマニティーズという分野がございまして、そちらでの動きを補足したということでございます。それが本日の参考資料3と参考資料4に当たります。参考資料3—これは国の取組—実際に運営しているのは国立国会図書館ですが—デジタルアーカイブの統一プラットフォームというものを作成しております、これは言語のテキストだけではなく、非常に幅広い対象—絵巻物ですとか様々なものがあります。それらを横断的に検索ができるという「ジャパンサーチ」というものがありますので、それを一つ、紹介をいたしました。参考資料4は、国文学研究資料館が今年度からスタートさせたプロジェクトのリーフレットでございます。このプロジェクトのスタートする前に、国文学研究資料館の持っている古典籍のデジタル化というものを集中的に行っていたんですが、次、それをどうやって活用するかということで始まったプロジェクトです。例えば、古典籍の資料は崩し字ですので、それをAIでテキストデータにするということはもう既に実際に行われております。そして、それを使ってどうするかという例としては、古典籍の中にある災害の記録—津波ですとか地震—そうした災害の記録を分析して、防災に役立てるという研究が既に始まっているということも書かれているかと思えます。このような動きがあるということで、「おわりに」に挙げさせていただきました。

そして、今後の方向として、言語コーパスの整備ということが前期の「見解」で

は述べられていたんですけども、今回の報告では、言語コーパスの活用というところまで視野に入れる必要があることを訴えていくということで、最後まとめております。

飛ばし飛ばしで申し訳ありませんが、以上が、この配布資料2、それから、参考資料1～4の説明でございます。

○相澤主査

ありがとうございます。では、ただ今の説明について御質問がありましたら、この場でお願いいたします。意見交換は後ほど行います。御質問ということで、何かございますでしょうか。よろしいでしょうか。 (→ 挙手なし。)

では、意見交換の方に進めていきたいと思えます。まず、秋までの審議について整理した資料として、参考資料2がございませぬ。また、その後の審議内容を加えて、文章化したものが、ただ今御紹介いただきました配布資料2の「(報告) (素案)」となっておりますので、こちらについて審議を進めたいと思えます。

この配布資料2、「(報告) (素案)」につきましては、事前にお手元に送付をさせていただきますにございませぬ、既に御提出いただいた意見については、反映されたものという位置付けでございます。

本日の説明を踏まえまして、配布資料2について改めて御意見を頂戴いたします。既にいろいろ議論は尽くされたところですので、「追加の意見なし」ということも含めまして、皆様、全員から御発言をいただくという形にいたしたいと思えますので、よろしくお願ひいたします。

まず、御意見のある方がいらっしゃいましたら、先にお願ひできますでしょうか。

○石川副主査

石川でございます。ありがとうございます。まず、この度のこの(報告) (素案)がまとまってきたこと、委員の一人として大変感謝しております。委員の皆様の御協力、また、文化庁の皆様の御協力で、現代における言語資源の重要性や、その基盤整備の重要性が明示的に示されたということは非常に重要であります。また、言語資源が言語学の研究にとどまらず、様々な社会課題の解決に通じるものであるという見通しが示されたことも大変意義のあることと考えております。

その上で、この報告書の趣旨をより明確にするため、付け加えることができるかもしれないと思うことを3点申し上げたいと思います。1点目は用語の定義に関する記述、2点目は危機言語などのコーパスに関する記述、3点目は国語教育との関連についての記述でございます。

まず、1点目の用語の定義についてですが、1ページの「はじめに」の冒頭部分などで、先ほど鈴木国語調査官が口頭で加えてくださったような説明、つまり、本報告書における「言語資源」、「デジタル言語資源」、「言語コーパス」といった用語の定義を示しておいてはどうかと思います。また、これに伴いまして、2ページの見出しの「2（1）」の2行目にある「電子的に検索可能な状態にして公開した言語資源」という部分は、「デジタル言語資源」としておいた方が整合性が高まるかと感じました。

続いて2点目、危機言語関連の記述ということですが、言語資源に期待される役割の一つとして危機言語の記述ということがございます。この点に関して、3ページには、これまでにいろいろな言語資源が作られてきたことが述べられていますが、危機言語コーパスについての記述が不足しているようです。例えば、第3段落「大学等の研究機関で様々な…」という段落内のBCCWJの記述の後に、これまでに、文化庁の支援の下、各種の方言コーパスなどが作成されていることを記載しておいてはどうかと感じました。

最後に、3点目の国語教育との関連についてでございますが、4ページに辞書編集に関する記述があり、それを受けて「教育面でも、言語コーパスは…」という内容が始まりますが、現在は外国人のための日本語教育のことだけが前面に出ているような印象がありますので、この箇所に、例えば、「国語教育や国語教材のアップデートの基盤資料となる」といった内容を加えることで、言語資源の教育的活用について、よりバランスの取れた内容になるのではと感じました。

以上、少し長くなりましたが、定義に関する記述、危機言語に関する記述を、国語教育に関する記述を、可能な範囲で盛り込むことも御検討いただけたらと思いついて、発言をいたしました。

○相澤主査

石川副主査、素晴らしい御提案をありがとうございます。今の御意見に対して何

かもし補足等、よろしいでしょうか。ありがとうございます。

では、ほかにはいかがでしょうか。

○前川委員

よろしいですか。

○相澤主査

前川委員、お願いします。

○前川委員

私も実は、石川副主査が前半におっしゃった定義のことはあった方がいいなと思いました。口頭での御説明は非常にはっきりしたもので、それでいいわけですが、**「言語資源」**という言葉に関して、人によっていろいろな理解があると思いますので、おっしゃった、ただ**「言語資源」**と言うと、アナログーデジタル化する前のものを含むということ、それから、デジタル化したものは、デジタルという言葉をつけて表現する、そして、コーパスというものは、さらにもうちょっと限定した形でというような形ですね。

それを、先ほど石川副主査がおっしゃったような形でもいいし、最初に、以下ではこういう意味で用いますというような定義を最初に導入してもいいし、それは是非、そうなされた方が誤解を防ぐのでいいと思います。全体としては、大変よくまとまっていて、特に申し上げることは、私からはございません。

○相澤主査

ありがとうございます。では、ほかにはいかがでしょうか。

○齊藤（美）委員

齊藤です。よろしいでしょうか。

○相澤主査

はい、お願いいたします。

○齊藤（美）委員

詳しい御説明、ありがとうございます。話題になっております定義について、私も先ほどの鈴木国語調査官の口頭での御説明が最初にあって、大変分かりやすいなと思いましたので、加えていただけるようでしたら、私も有り難いなと思います。

一つ、資料を拝見して、もう少し御説明が詳しいといいかもしれないなと思ったことがあるので、申し上げます。「コーパスアプローチ」という言葉です。今の定義の話と、もしかしたら関係してくるかもしれないんですけども、確かこれは11月に石川副主査が御発表くださった中でも御使用になって、詳しい御説明を頂いたと思うんですけども、批判的談話分析などとも関わってくるというような話もあったかもしれないけど、そこまでここに詳しく書くのは難しいかもしれないんですけども、これは非常に重要なアイデア、概念で、最後の「おわりに」のところにも、再度、3回目が、使われていますので、これを強く打ち出せると、社会実装のところでも強みになるようなキーワードのようにもなるのではないかなと思いましたので、もし、うまく入れ込めるようでしたら、もうちょっとこの「コーパスアプローチ」というのが言語学的な分析を行うということは、もちろん書いてくださっているんですけども、さらに、もう少しあってもいいのかもしれないと思いました。

あと、先ほど石川副主査が危機言語のことをおっしゃっているのを聞いて、ああ、なるほどと思いました。そうであればというか、日本の中にある日本語以外の言語という意味で、これまで手話言語について話題になったことがあったらどうかと思いました。日本語のコーパスということで作っていこうという話ですので、手話言語はまた別の言語としてあるわけなので、必ずしも含める必要はないと思うんですけども、例えば、将来的にそれも視野に入れていますよといったことを、ちょっと触れておくのもいいのかもしれないと思いました。これは思い付きですが、申し上げます。

○相澤主査

ありがとうございます。手話言語については、確かにあまり話題に出てきませんでした。これまでの取組状況など、いかがでしょうか。

○村瀬国語課長

では、私の方から。手話言語につきましては、障害者施策の中で、正に障害者関連法の中で、言語としての手話言語の重要性といったものがうたわれてきているということが、まず一義的にあるところでございます。このため、国会におきましても、手話施策推進法という形で、手話を使った、いわゆる教育であったり、福祉であったり、そのためのいわゆる通訳をする方々であったり、そういったものを進めていこうという動きが御議論として進んでいるところでございます。

広い意味での「言語」におきましては、手話言語もここで御議論されていた事柄、それから、各委員の皆様が御研究されている事柄と同じような形での論点ということになるかと存じますが、一義的には、先ほど申し上げました手話政策という中で、まずもって取組があり、そして、また、いわゆるアカデミックな部分として、この手話言語をどう保存していくのかといったような位置関係に、今、行政上はあるかと受け止めております。

したがいまして、今、齊藤（美）委員からもお話があった点も含めまして、今後、我々も研究課題として受け止めていきたいというふうに、まず印象として受け止める次第でございます。

○相澤主査

ありがとうございます。言語とは多様であるというメッセージの形で、どこかに入れることは可能かと思えます。

ほかはいかがでしょうか。（→ 挙手なし。）

そうしましたら、「あいうえお」順でまいりたいと思いますが、植木委員、よろしくをお願いします。

○植木委員

ありがとうございます。本当に長い議論を、このように整理してまとめてくださいまして、ありがとうございます。特に私から付け加えることはないのですが、こういう報告のときには、前川委員がよくおっしゃっているように、予算を取ってきて、作っていくというようなところでは、いかに社会実装—社会に役立つのかということを強調するべきだと思います。それはもちろん、そうなのですけれ

ども、冒頭のところで、「文化遺産」と位置付けるというのは、両輪の一つとして重要だと個人的には感じております。

古語など、なくなっていくってしまうものが記録されているということ自体、活用するというだけでなく、ただ保存されているということにも、非常に貴重な意味があると思っています。ちょうど石川副主査が危機言語のことをおっしゃったので、正に、古語と言わなくても、今使われているものでも、消えていってしまうものを保存することにも重要な意味があるということを一直してほしいということではないんですけれども一そこも十分強調したいところだなと思った次第です。

○相澤主査

ありがとうございます。では、神永委員、よろしく申し上げます。

○神永委員

意見と言いますか、ちょっと感想的な感じになってしまうかもしれないんですけれども、今回この小委員会で各国のコーパスの状況をお聞かせいただきまして、大変勉強になりました。ありがとうございました。

この、今日お配りいただきました配布資料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）（素案）」についてなんですけれども、特に「言語コーパスの今後」というところで、「取組の継続」というところに「言語コーパスの普及」というところがあります。7ページ目なんですけれども、「辞書における言語コーパスの活用を進め、その意義を発信することで言語コーパスに触れるきっかけを作るとともに、言語コーパスの意義を理解してもらい取組が考えられる」という一文を加えていただいたのは、辞書編集者としては、とても有り難い内容だと思っています。

今までの、私の発言を繰り返すことになりますけれども、現在、日本最大の50万項目収録しております「日本国語大辞典」という辞典の第3版の改訂作業を始めております。やっぱり現在の日本では、コーパス活用の最先端と言いますか、一番恩恵にあずかっているのは、ひょっとすると国語辞典の現場ではないかなというふうに考えております。また、規模の大きな辞典というのは、コーパスを活用するだけでなく、逆にコーパスを入りにできる存在ではないかなというふうにも、私は考

えております。

そのようなこともありまして、「日本国語大辞典」第3版では、新しく編集委員を委嘱した方の中に、国語研のBCCWJとか、「日本語歴史コーパス」の構築に実際に関わった方もいらっしゃいますので、そういった方たちに御協力いただきたいと考えております。

特に、「日本語歴史コーパス」に関して言いますと、例えば現在、平安時代編からジャパンナレッジという、辞書や文学全集などの検索サイトがあるんですけども、そこに収録されています小学館の「新編 日本古典文学全集」へのリンクができるようになっております。これは、検索アプリケーションであります「中納言」の検索結果から、その用例の存在する「新編 日本古典文学全集」の当該ページの本文と、現代語訳と、それから注釈を直接参照できるものなんです。

「新編 日本古典文学全集」と同じ出版社で「日本国語大辞典」というのは出しておりますので、その機能と、それから得られる情報を大いに活用しまして、デジタルに特化されてしまうだろうとは思いますが、国語研と協力しながら、デジタル時代に対応した新しい形の国語辞典に進化させたいということを考えております。

実際に、このような取組に掛かれるのはまだ先ではあるんですけども、今年度、国語分科会で言語資源に関する小委員会を開催したわけですから、今年度限りで終わりにしてしまっただけでは、非常に惜しい気がしております。何年か後に「日本国語大辞典」の第3版は2032年の刊行を目指しておりますけれども、私もその間、元気でいられるかどうか分かりませんが、私以外の人間になるかもしれませんが、どのような形で「日本国語大辞典」でコーパスを活用しているのかを報告できるような機会を設けさせていただけると、非常に有り難いと思っておりますので、その辺のところを御検討いただくと有り難いと思っておりますので、よろしく願いいたします。

○相澤主査

どうもありがとうございます。では、武田委員、よろしく申し上げます。

○武田委員

ここまでの議論を伺いまして、今回の「(報告) (素案)」ですけれども、大變的確にまとまっていると思っております。

また、本日、石川副主査、前川委員より補足もいただきましたけれども、非常に重要な御指摘になるかと思えます。古語教育ですとか、危機言語に関する配慮ですとか、そういったところも非常に重要な観点かと考えております。特に、私の方から御指摘することはこれ以上ないかと思えます。

この素案の表現なんですが、大變^ま瑣末ではございますけれども、8ページに、表現として「裨益^ひする」という言葉が出てきております。若干難しい表現かと思われ、これも以前から一貫して用いられている表現のようですけれども、この表現が妥当かどうか。これでよければ、このままということでもよろしいかとは思いますが、この点だけちょっと気になったところでございます。

○相澤主査

ありがとうございます。では、最後に、私からで、細かい点への配慮がすばらしく、拝読する際に一気に最後まで読むことができ、文章全体が頭にさっと入ってくる書きぶりである点に感銘を受けました。

私たちが議論を始めたときには、それほど強調されていなかった「コーパスアプローチ」という言葉が本小委員会での議論を進めるうちに出てきて、皆様が賛同をされたということで、その点が重要であると感じています。もしも、もう少し強調するとすると、先ほど植木委員もおっしゃっていた2ページ目の、デジタル言語資源を「未来へ伝える文化遺産」として位置付けるというところ、文化遺産だけではなくて、社会課題解決の手立てでもあるということを併記する、例えば「…位置付けるとともに、構築し、社会課題を解決する」といったふうに、頭出しをすることも可能ではないかと思いました。

あと、「コーパスアプローチ」の議論の時に、人材育成の話が話題になったことを記憶していますので、一番最後の10ページで、コーパスアプローチが望まれるという記述のところ、人材育成への言及が少しあってもよいのではないかと思います。無理にということではないんですけれども、そういう手もあるかなというふうに感じました。

「コーパスアプローチ」という概念そのものが、今回のまとめ案で今後へ伝えた

い重要なキーワードだと感じていますので、以上コメントさせていただきます。

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、長い間にわたりまして御審議をいただきまして、ありがとうございます。
このまとめの素案についてのめどもほぼ立ったところだと思いますので、本日頂いた御意見につきましては、配布資料2を修正の上、次回の国語分科会での報告をするということで、資料の修正及び報告については、主査である私に御一任をいただければと思います。よろしいでしょうか。

(→ 「異議なし」の声あり、小委員会として了承)

ありがとうございます。それでは、配布資料2の修正案は、皆様と共有していただくということで、修正された配布資料2をもって国語分科会で報告することで決定をいたします。

もし何か、追加で御発言がありましたら、よろしく願いいたします。よろしいでしょうか。

加えて、もう一言ずつ、皆様に全体を通しての御感想を頂ければと思っております。では、石川副主査からよろしく願いします。

○石川副主査

小委員会の報告で、コーパスを社会課題の解決に生かす「コーパスアプローチ」という概念を紹介させていただいたのですが、恥ずかしながら、自分では余りそうしたことができておりませんでした。

この間の小委員会での審議や、その後の皆様との意見交換を通じて、たとえ小さなことでも、実際に何か動き出してみる必要があるのではないかと感じるようになり、先般、関西地方の教育委員会の方とお話する機会を持ったところです。皆様御承知のように、現在、日本の小中高には、日本語の理解に問題を抱えている外国ルーツのお子さんが多数おられるわけですが、例えば、児童や生徒の皆さんに統一のテーマで作文を書いてもらい、それをコーパス化して言語解析することで、日本語に問題を抱えている可能性のあるお子さんを早期に見付けるといったことができないか相談を始めたところです。

今後、言語学の枠内のみならず、社会の各層で、各種の課題に特化した多様なコーパスが構築され、それらを生かして、社会課題の解決を図る動きが広まってくる

とすれば、すばらしいことだと思います。

○相澤主査

ありがとうございます。では、植木委員、よろしくお願いいたします。

○植木委員

ありがとうございます。私は本当に専門外で、勉強させていただくことばかりで、こちらはほとんど貢献できなかったのですが、学ばせていただきまして、ありがとうございました。

私は、大学の国文学科に所属していて、入ってくる学生たちは言葉に関して、もちろんある程度興味がある人たちなのですが、日本語を母語とする人たちでも、その言葉に対しての自信のなさというか、若い人たちの一へんてこな敬語を使ってしまうとか一そういうようなところで、非常に言葉に対して興味もあり、自分の言葉遣いについて心配もしているようです。

一方では、非常に新しい、略語のようなSNS上での言葉、余り論理的でないような言葉、そういうものもどんどんできています。本当に新しい言葉の中でみんな困惑していると言うか、指針を求めていると言うか、そういう感じを日々受けているので、言葉を集めて、分析していく、その基になるコーパスというのは本当に重要なんだなというのを、日常的にも感じています。それを作っていくにはお金も人手も掛かるわけですが、とても意義のあることだと改めて思っております。

ありがとうございました。

○相澤主査

ありがとうございます。では、神永委員、よろしくお願いいたします。

○神永委員

今回、辞書編集者の立場としていろいろとお話しさせていただいたんですけども、今日、石川副主査のお話で国語教育ということをお聞きしまして、はっとしまして、ああ、それ、抜けていたよねというような思いがありました。

私も、もちろん国語教育に関しては門外漢ではあるんですけども、子供時代の

語彙教育というのはとても大事なことはないかなと考えておりますので、その中で、このコーパスをそういったものに活用できるのであれば、それは非常に素晴らしいことだと思います。

やはり、辞典にも同じようなことが言えるんですけども、日本語は常に変わっておりまして、いろいろと変化している部分があるんですが、その変化がコーパスから読み取れるわけです。子供たちに、じゃあ、正しい日本語って何だろうか、日本語ってどんなものなんだろうか、一体それが正しいものだろうかということをお教える一つの材料にもなるのが、コーパスだと思うんです。

そういった意味で、うまくコーパスを使いこなせると一私が言うのも変なんですけれども一本当に小学校の国語の授業もまた楽しい言葉の指導の時間が作れるんじゃないかなんて、勝手なことを考えたわけなんです。是非、そういった方面での活用を期待したいなと思っております。

今回、どうも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、齊藤（美）委員、よろしくお願いいたします。

○齊藤（美）委員

こちらの小委員会に参加させていただいて、ありがとうございました。私も言語資源、コーパスについては専門ではないので、一体どんなお役に立てるのだろうかと不安で、不安なまま終わった感じもあるんですけども。ただ、私が翻訳について研究する上で一明治時代の翻訳の研究が多いんですけども一実は点と点で利用していた、BCCWJであったり、国文学研究資料館のデジタル化された明治時代の翻訳者が書いた自筆の原稿であったりとか、それから、国立国会図書館のデジタルコレクションであるとか、そういったものが実は「デジタル言語資源」だったり、「言語コーパス」だったりして、自分もその恩恵を受けて研究をしていたんだなということを改めて理解して、有り難く思っておりました。

植木委員が残すことの重要性と、先ほどもおっしゃっていましたが、こういう明治時代の翻訳の研究で翻訳者が言っていたこととか、訳していたものとかを掘り起こして、私が例えば論文にすることによって、改めてそこに光が当たるという

か、そういう形で残すということもできるのかなとも思いました。それを、もちろん、今の社会、翻訳について考えていく上での役に立てることもできるのかなとも思いましたので、先ほど「コーパスアプローチ」の話もありましたけども、自分が結構利用しているんだなということを改めて意識して、その点をきちんと発信をしていけたらいいなと思っておりました。

以上です。ありがとうございました。

○相澤主査

ありがとうございます。では、武田委員、よろしく願いいたします。

○武田委員

この間の議論に参加させていただきまして、大変勉強になりました。コーパスの必要性、また継続的な開発、維持、そしてまた社会実装、コーパスアプローチというところまで幅広い観点でまとめられたということも、大変よろしかったかと考えております。

出版事業におきましても、まだまだコーパス活用の余地はあると思いますし、今後の活用、そして、社会実装に向けて出版事業としても重要な、今回の報告になったのではないかと考えております。

以上でございます。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、前川委員、よろしく願いいたします。

○前川委員

皆様の御尽力で大変いい報告がまとまったのではないかと思います。特に、私どものように実際そういうことをこれまでもやってきた人間からすると、社会実装の方とか、そういった方に必ずしも目が届かないということがありまして、そういうところをきちんと押さえた答申になったのは、大変良かったなと思っております。これは、本当に皆様の、あと国語課の皆様の御尽力によるものだと思います。

ただ、これは、今回の報告というのは結局「出発点」でありまして、これから先、

実現に向けていろいろとやっていかなきゃいけない。そこにはまた別の問題がいろいろと生じてくる。いろいろな問題、あと著作権の問題なんかも含めて、それについては改めて覚悟を決めておき、それから事前にいろいろと考えておく必要があるなど。

もし、そのようなより具体性を持ったものについて、更に検討するような場が今後設けられれば、それは大変有り難いことだと思っております。

以上です。

○相澤主査

ありがとうございました。では、最後に、もう私は本当にこれ以上申し上げることではなく、ひたすら皆様に感謝をするということで、大変貴重な経験をさせていただき、いろいろなことを知る機会になりましたことも、感謝しておりますし、その結果として、まとまりました素案そのもの、今度、上の国語分科会で報告するのは大変楽しみでございますので、本当にありがとうございました。

元々私は、大規模言語モデルを構築するのに関わっているということで、恐らくお声掛けいただいたと思うんです。このデジタル言語資源の中で、LLM—大規模言語モデルというのはほんの一部でありまして、切り取られたテキストというのは、誰が書いたかも分からない状態では、生きた言語ではないと言えます。言語コーパスという形で、誰が、どのような文脈で、どういう人に伝えようと思って書いたのかということが分かって、初めて言語というのは生きて後世に残せるものになると思っております。

誰かが、正しい言語とは何かというのを常に教育の場でフィードバックをしていかなければ、社会実装の場で生きていくようにならないということを含めて、皆様に教えていただくことが大変多くありました。どうもありがとうございました。

それでは、言語資源小委員会の審議終結に当たりまして、村瀬国語課長から、最後に御挨拶をお願いいたします。

○村瀬国語課長

御挨拶申し上げます。これまで非常に長い間、計6回にわたりまして、委員の皆様には非常に御丁寧に、精力的に御審議を賜りまして、本当にありがとうございました。

した。

本日も「コーパスアプローチ」の話であったり、あるいは、「未来へ伝える文化遺産」ということで、残すことの重要性、そういった中から具体的に古語であったり、辞典であったり、あるいは翻訳であったり、そういったお話が多岐に広がってきたかと存じます。

よくと言いますか、耳伝えに聞いておりますと、国がなくなるときには、まず言葉からなくなっていくというような、そのような話も歴史的に伺っているところでございます。そういった意味で、このたび言語コーパスの意義を整理していただきまして、そして、ただ今もございましたが、社会実装に向けた御提言を頂けたということで、価値創発というところに力点を置いて、今後に向けた新たな見取図をお示しいただいたのではないかと受け止めているところでございます。

重ねまして、これまでの委員の皆様様の御指導、御尽力に御礼申し上げたいと存じます。先ほど前川委員からお話ございましたが、今後もいろいろ課題がある、これが出発点だという叱咤^た激励のお話もございましたので、今後とも、いろいろと委員の皆様と御一緒に、我々もまた取り組んでまいりたいと思います。引き続き御指導のほどよろしくお願いいたします。

私からは以上でございます。

○相澤主査

ありがとうございます。では、これまでの皆様様の御協力に感謝いたしまして、本日の言語資源小委員会はこれにて閉会といたします。どうもありがとうございました。