

第12回 国語分科会言語資源小委員会(Web開催)議事録

令和 8 年 2 月 26 日
10 時 00 分 ~ 11 時 00 分
文部科学省3階 3F2特別会議室

〔出席者〕

(委員)相澤主査、植木、神永、齊藤(美)、武田、前川、森各委員(計7名)
(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、武田委員及び事務局は、3F2特別会議室にて参加。

〔配布資料〕

- 1 第11回国語分科会言語資源小委員会議事録(案)(委員限り)
- 2 今期の言語資源小委員会における審議経過報告(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)

〔経過概要〕

- 1 事務局から定足数の確認があった。
- 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 事務局から配布資料2「今期の言語資源小委員会における審議経過報告(案)」について説明があり、質疑応答、について意見交換が行われ、審議経過報告(案)の修正について主査一任とすることが了承された。
- 5 今期最後の言語資源小委員会に当たり、国語課長から挨拶があった。
- 6 今期最後の言語資源小委員会の議事録の扱いについて、2回の委員による確認を経て公開することが確認された。また、国語分科会については、3月16日(月)15時

から予定していることが確認された。

7 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12回、今期6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。

本日も、オンサイト、文部科学省の会議室で御参加の方々と、オンラインで御参加の方々と合わせてのハイブリッド開催となります。また、傍聴者の方々におかれましても、オンラインでこの会議を御覧になられていることを御承知おきください。

本日は、お手元の議事次第のとおり、「言語資源の整備と活用」について、「その他」という、二つの内容で協議を行いたいと考えております。本日は、今期の言語資源小委員会の審議経過のまとめを行う場でございます。

それでは、議事に入ります。議事(1)としては、今期の審議のまとめということで、今期の審議内容について事務局の方で整理していただいた資料がございますので、配布資料2に沿いまして、まず、事務局から説明をお願いいたします。

○鈴木国語調査官

それでは、配布資料2を説明いたします。

配布資料2は、参考資料1にあります、今期の主な審議事項に基づいた構成といたしました。「はじめに」に全体的な、この報告が何をまとめたものかということを書いた上で、主な審議事項として挙げたものの順番に、それぞれヒアリングを行った中身と、それから、審議事項に基づいた意見交換で出てきた意見を並べているという形でまとめたものでございます。審議事項に挙げた四つ以外の事柄についても話題になりましたので、それらは別の柱として整理したという形でございます。これが全体的な構成でございます。

まず、「はじめに」でございますが、最初の3行は、実は、昨年度まとめていただいた報告で使っている表現をそのまま持ってきております。言語資源、デジタル言語資源、言語コーパス、こちらについての基本的な考え方を示しているのです、それをそのまま用いております。そして、令和6年度にはこの報告というものを取りまとめ、令和7年度のこの言語

資源小委員会におきましては、報告で示された方向性を踏まえ、補完的な事項について審議を行ってきたということで、四つの課題をここで示しているという形です。そして、この四つの課題と+αの事柄について、どのような審議を行ったのかということ、以下、課題ごとに整理したという形で、全体の流れを説明しております。

1点目としまして、「言語コーパスなどデジタル言語資源整備の取組の進捗状況」。これは、国立国語研究所のBCCWJ(現代日本語書き言葉均衡コーパス)拡充の取組の進捗状況を把握するということを中心に据えた課題でございましたので、前川委員から御報告を頂きました。その中で、この拡充作業自体は比較的順調に進行しているということがまずございまして、2025年度末に最初の部分的なデータの公開を予定している—ただ、それは、BCCWJ1—既にあるBCCWJ—とは別のコーパスという形で公開するけれども、検索に関しては、串刺し検索が可能な形で公開する予定であるというお話でした。そして、来年度からは成果発表会の開催を予定していること、それから、2028年度で構築を完了する予定で進行しているというお話がございました。

どのようなデータを追加するのかということで示されたことは、2025年まで20年間にわたる出版書籍、これがサンプリングされ、語数としては1億語となるということ。このところを主に言われていたところでございますが、実は、そのほかにも、ここ3回の学習指導要領改訂があり、その度に教科書が新たに出ておりますので、その教科書—小学校から高等学校までの全ての教科の教科書で、最も採択数の多い教科書について1種類ずつ、全文をデータとすること。大体2,000万語ぐらいになるのではないかとこのお話がございました。それと、もう一つ話題になったのが、SNSからも、2024年から2026年のもので70万語ぐらいのものをデータとして取った形で追加する計画になっているという御紹介がありまして、特にSNSのデータというのは、これまでのコーパスに含まれていない、対話となっている書き言葉という形ですので、新しい価値、新しい使い道の可能性があるということになります。そして、もう一つ、著作権との絡みがある点で、検索においてこれまで、前後、かなり長いものを示すことができたのですが、思想又は感情の享受を目的としない、軽微利用の範囲に収まるように、ここの表示できる語数というところは少し絞った形になるだろうというお話がございました。

この報告を踏まえた意見交換におきましては、特に国語辞典の編集において非常に活用ができると同時に、国語辞典の内容も大幅に変わることにつながり得るというお話ですとか、あとは、BCCWJのデータというのは、生成AIの影響をぎりぎり受けていない時代

の日本語のデータ集合になると考えられるので、大規模言語モデルの評価などで使用するという、そうした価値もあるのではないかと、そのような御意見が出ておりました。

2点目としましては、「言語コーパスなどデジタル言語資源を活用した社会課題解決のための取組」ということで、こちらは、整備ではなくて、活用の側面についての実例と、意見交換でございます。

まず、最初の小委員会で、国文学研究資料館の渡部館長に、「データ駆動による課題解決型人文学の創生～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～」というテーマで、国文学研究資料館で行っている取組について、お話しいただきました。古典籍のデジタル化、そしてデータベースとしての公開というところで、データの範囲を少し広げるというお話と同時に、このデータベースを利用して異分野融合という形でデータ駆動型の研究を推進するという動き出しているというお話で、その例として、文献観光資源学ですとか、典籍防災学といったものが挙げられておりました。

そして、これに関しての意見交換では、現代語訳ですとか、校訂本文、英訳などもあるといいのではないかというお話。それから、テキストの精度が100%でない段階であってもテキストデータというのを使わせてもらえると有り難いということで、公開を早めてもらう方がいいだろうという御意見。それから、このような基盤を整備するという取組に関しては国からの長期的な支援が必要であることが確認されたということでございます。さらに、国立国語研究所で形態素分析の技術というものをもちますので、これを組み合わせることで、さらに課題解決につながる、いろいろな研究が進むのではないかと。それと同時に、そもそもデジタル化する基になるオリジナル資料が保管されていることが実は重要なことであるという御指摘もございました。

もう一つ、より具体的なお話として、国語辞典におけるコーパス利用について、神永委員からお話しいただきました。『日本国大辞典』の改訂ということで、『日本国語大辞典』は用例主義を採っておりますので、用例というところでこのデジタル化された資料というものが非常に活用できるということで、どういう点で活用できるかということをお話しいただきました。そして、特に『日本国語大辞典』の場合は、この辞書を入り口として多種多様な知的情報への動線をデザインするという非常に新しい発想での辞書の在り方というものをお話しを計画しているというお話がございました。

これに対する意見交換ということで、一つは、紙の辞書だと改訂されたところが分かるのだけでも、データ更新だけになってしまうと変わった部分が分からなくなるので、その

点について何か検討する必要があるのではないか。それから、もう一つは、今、語釈や語義の分類でAIを活用することができるようになりつつあるので、そういったものの活用はどうなんだろうか。このようなことが示されておりました。

3点目としましては、「地域の文化的基盤としての地域の言語資源の在り方」。これは主に方言の話題でございます。岩手大学名誉教授の大野眞男先生から「消滅危機方言の記録作成及び啓発事業の意義－日本語の多様性のためにできること－」ということでお話しいただき、危機言語・方言に係るユネスコや文化庁の動きというのが、各地の方言の価値を見直し、活性化させる取組の後押しになっているという評価ですとか、そもそも言語には、意思疎通の道具という側面とアイデンティティーの表出という側面があって、方言のことを考える場合、後者、アイデンティティーの表出の機能というところに注目して取り組むことが必要であるということ。そして、消滅危機方言の記録作成のための新たな調査ということももちろん必要ではありますが、方言に関しては埋もれている既存のデータがかなりあるということで、それを掘り起こして言語資源の整備というところにつなげていく必要があるであろうということ。さらに、方言のデータというものは社会還元されるべきものであり、医療・介護の分野ですとか教育現場において特にそういうことが可能であり、必要であること。それから、方言の価値の見直し、方言を通じた地域への愛着の深化というのは、地域文化の活性化に寄与するものである。このようなお話でございました。

もう一人、東洋美術印刷株式会社の畑中悠樹様から、「方言を活かしたブランディングの取り組みについて」という報告を頂きました。こちらは経済的な側面から方言というものを位置付けるという内容でございましたが、現在、方言というものは非常に肯定的なイメージを持ってもらえるものになっていて、かつ、多くの場合は共通語でやり取りがされている中で方言が出てくるとインパクトが強く、特に話題になりやすい。そういう意味で、ブランディングで使う効果があるということではあるのですが、ただ、下手な使い方をすると不自然で、敬遠される原因になってしまうという面があり、慎重さが必要であるということが述べられました。加えて、方言の使用というのは、一見すると県外の観光客向けという見方をされがちですけれども、それだけではなく、地元の人にとってふるさとへの帰属意識というものを刺激する、そういう存在でもあるということが示されております。特に、最後の地元の人にとってというのは、大野先生がおっしゃっていたアイデンティティーの表出と重なる要素として示されておりました。

こうした報告を踏まえて意見交換がございました。一つは、方言をデータ化をするとき、

方言は多様であるからこそ、ある程度の画一性を持たせる必要のあるデータ化という部分とは、なかなか相性が悪い面があるのではないかという御意見。そもそもデータベース化して将来のために残すというときに、どういうふうにするかということにそれは重なってまいります。それと同時に、各地で方言資料を作成していて、それをいかに取り込んでいくかということを考えないと、各地の記録というもの、特に磁気テープで残されているものは消失してしまうおそれがあるので、これは実は緊急性が高いという御指摘。それから、方言のブランディングという点においても、このデータベースがあると、確実な方言の使い方等が確認できるという意味で活用されるであろうということ。そして、どうしても短期的な目標に傾きがちな中で、方言の継承ということの動機付けというのはなかなか工夫が必要であろうということ。そのような御意見を頂いたかと思えます。

4点目としまして、「外来語をめぐる課題と言語コーパスの活用」。これは立命館大学教授の小椋秀樹先生と日本大学准教授の金愛蘭先生に御発表いただいたものです。

小椋先生は「コーパスを活用した外来語表記のゆれに関する調査」ということで、国立国語研究所のBCCWJをはじめとした様々なコーパスを利用して、外来語の表記をめぐる状況を調査されました。その結果の報告でございますけども、それは同時に、こうしたコーパスというものの充実—継続的にしっかり拡充されていくということが、外来語をはじめとした国語施策を考える上において、日本語の実態把握という点で非常に重要なことであるという、そのようなまとめを述べておられたということでございます。

金先生は、新聞の語彙の中で抽象的な意味を表す外来語が基本語化しているという発見をお話しになりましたけども、その中で、20世紀後半の大規模な新聞経年コーパスを御自身で作られて、それに基づいて調査を行ったということで、こうしたコーパスというものが充実することが言葉の変化等を把握する上において重要なことであるというお話でございました。

これらの報告を踏まえての意見交換では、一つは、同じような意味、類義語についての使い分けというのは、同じ語が続くのを避けるといった面からも行われるという御指摘。あと、科学的に議論していくには、経年でたどれるコーパスが必須であるということ。一方で新聞社の方は、自社の記事データについての扱いが、外の人間が使う場合に難しい側面が出てきているので、そうしたものの共通プラットフォームができることが望ましいのではないかというお話。それから、辞書の中での外来語の扱いというものを少し見直す必要が、特に用法の変化というものを考えたときにあるだろうという御指摘。一方で、表記の統

一、外来語の表記の統一というところに関して、どこまでそれが必要なんだろうかということも改めて考えてみてもいいのではないかという御指摘もございました。

石川副主査の御発表の中でも「言語資源と外来語」という項でお話がありまして、外来語の使用実態を考えると、定点観測調査と結果の報告というものが継続的に行われることが必要であり、そのことが国民の意識や公的機関での使用というところに響いていくであろうということ。新しい概念を際立たせるために外来語の使用というものはありなんだけれども、そこでは受け手に正しく伝わるための工夫というものは必須であるという、このことは落としてはいけないだろうということ。そして、コーパスの継続的な整備の必要性というものを述べられていました。

ここまでが四つの柱でございまして、それ以外で、「生成AI時代の言語資源」ということで、生成AIと大規模言語モデルとの絡みで相澤主査に「大規模言語モデルと言語資源」ということでお話いただいたということがございますので、これについても意見交換がございました。大規模言語モデルの構築に用いるコーパスの品質は知識問題を解かせた正答率で調べているということですか、良質のコーパスを作成するには様々なクリーニングを行う必要があるというお話がありました。大規模言語モデルの世界で言う「コーパス」という用語と、この言語資源小委員会で用いている「コーパス」という用語にはちょっと違いがあるということを含めての御発表でございました。この言語資源小委員会では、個々に言語情報を与えた総体としてのデータセットを「コーパス」と呼んでおり、大規模言語モデルの中で言う「コーパス」とは、若干、指すものの幅が違っているということがございます。その点を確認できるところでございました。

そのほか、大規模言語モデルの学習には3段階あること、非常に話題になったのは、モデルごとに出カテキストに「口癖」に当たるようなものがあって、違いが区別できたりするというお話。それから、生成された言語が人間の言語に影響を与えているというのは、既に現状として存在しているというお話。さらに、生成する言語を解析するには、モデル自体の透明性が求められているが、その透明性をたどれるようなオープンなモデルというものは数えるほどしか存在しないという御指摘があったかと思います。

それを受けて、学生が生成AIで作ってきたものが、内容的にも、言葉としても、だんだん不自然さがなくなってきたりすると、教育の現場でそれをどのようにしていったらいいかということは大きな課題になってくるというお話ですとか、先ほど申し上げましたけれども、既に作られたBCCWJ1、それから、現在拡充しているBCCWJ2というのは、生成AIが関

与していない状態の日本語であるという点を重視して、評価等で利用できるという、その可能性の御発言もありました。

石川副主査からも、「生成AI時代の言語資源研究」ということでお話があり、基礎データのブラックボックス化というところの課題というものを御指摘の上で、研究の方向が、当初、生成AIを使つての研究というところだと思われていたものとだんだん違う方向性が出てきているのではないかということ、むしろ、人間が適切な用例を抽出して、それを生成AIに解釈させるという、このような方向というものが出てきつつあるというところで、人間が行う部分というのがコーパスの開発の部分と適切な用例の抽出の部分になっていくのではないかと、そういう可能性を指摘されていたということがございました。

最後のところは、「これからのデジタル言語資源の整備・活用における課題」として、いろいろな課題の指摘がありましたが、石川副主査の御発表の中でまとめてお示しになられたところがございます。多種多様なコーパスが作成されているけれども、その多くは公開されていないこと。そもそも、コーパス作成が目的ではなくて、手段であるためにこういうことが起こっているということがございます。作成されたコーパスが他の研究者の他の関心事にとって有益である可能性があるのも、これはできるだけ公開されるべきものなのではないか。そのための仕掛けというところでいろいろ考えていく必要があるのではないかと、この御発表でございました。

これに対して、公開を前提とまでしてしまうとさすがに酷ではないかというお話とともに、公開を前提に作成されるものを「コーパス」と呼んで、コーパスの定義をより狭い形できちんと示す。そして、そうでないものはデータとかデータベースと言って区別して呼んでいくということも一方で考えられるという御意見が出されておりました。

以上が、今期、ヒアリングを中心にして委員の皆様の意見交換の中で出てきたところを、本来は皆様にもっと御発言いただいていたところを、エッセンスだけに絞って整理した形でお示したものでございます。

そして、「おわりに」ということで、「コーパスアプローチ」という言葉を残しつつ、デジタル言語資源の整備と活用というところについては、小規模、中規模のコーパスの存在の指摘がございましたので、そうした実態を捉えていく問題というところに触れ、さらに、著作権の関係のお話も話題の中に出てまいりましたので、そういう関連領域も視野に入れて課題を考えていくということの必要性を触れさせていただいて、まとめとしております。

以上でございます。

○相澤主査

ありがとうございました。

それでは、これから先の議論の進め方ですが、まず、今日のメインは、本資料につきまして皆様の御意見や御感想をお伺いして、修正すべきところがあれば併せて御提案いただくということです。十分に時間がありますので、もしかしたら少し早めに終了という形になるかもしれません。

その前に、ただ今、御説明いただいた内容について、御質問がありましたら、お願いいたします。その後、意見交換の時間とさせていただきます。（→ 挙手なし。）

よろしいでしょうか。

では、御意見、御感想の交換の場に入らせていただきます。気になる点や、この辺りは重要、この辺りをこうしてはどうかということ、本当にどの切り口でも御自由に御発言いただければと思います。よろしくお願ひいたします。

○齊藤(美)委員

オンラインの齊藤(美)です。よろしいでしょうか。

○相澤主査

よろしくお願ひいたします。

○齊藤(美)委員

まず、大変きれいにおまとめいただいて、ありがとうございました。私の発言内容をまとめて記載いただいたところだと思うんですけど、ちょっと私の言い方が曖昧だったなと思ひまして、もし可能であれば、訂正させていただいて、修正を入れていただければと思うところがありますので、それを申し上げます。

5ページ目にあります、「4 外来語をめぐる課題と言語コーパスの活用」のところの意見交換の部分の一番下は、多分、私の発言を基にしてくださっていると思うんですが、「明治期の文献を読むと、表記は様々で…」というのですね。これを「外来語の表記方法を統一するかどうかというのは、言葉はどうあるべきかという各使用者の考え方と関わるため、慎重な議論が必要だろう」ぐらいに変えていただけないかなと思ひました。発言の意

図としては、うまく言えてなかったと思うんですが、言葉はどうあるべきかという、その考え方で私が意味しているのは、例えば、多少、表記が異なっても伝わればいいよねと考える人もいれば、言葉というのは、正しい／間違っているというのがあって、表記方法についても正しい唯一のものがあって、使用者はみんなそれに従うことを目指すべきというような考え方などもあるだろうと考えています。なので、日本語の使用者と一くりに言っても、年代も様々ですし、地域ごとの違いもありますし、また、第二言語、第三言語として学習している方々もいますので、言葉はどうあるべきかと考えるかは、一様ではない、いろいろあるだろうと思われましたので、また、私自身も、明治期に書かれたものなども含めて、いろいろな方が書かれたものを読む中で、どうあるべきかと決めるのはちょっと難しいなど思うことがあるため、申し上げておりました。

以上です。御検討、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した。

ほかはいかがでございましょうか。

では、フォローアップということで、私の部分についても、1点、補足でございます。発表したときに言語モデルごとに「口癖」があるということが大変印象に残ったかもしれませんが、その後、日本語についても調べまして、日本語についても非常に高い精度で見分けることができるということを確認いたしましたので、フォローとして御報告をさせていただきます。

以上でございます。

では、御意見等を御自由にお願ひしたいと思います。よろしいでしょうか。

そうしましたら、また例によって順番を決めて御発言をお願いしていく形にしたいと思います。皆様、御準備の時間もあると思いますので、まず、私から簡単な感想を述べさせていただきます。

ただ今の資料にもありましたとおり、私、長い間、ある意味お手軽な、ウェブから抽出したようなテキストも対象としながら、自然言語処理、コンピューターを使った言語処理の研究をしてきました。そのときの思いとしては、やはりウェブから切り取られたテキストというものには文脈がない—余り生きていないと申しますか—誰がどういうシチュエーションで発信した文であるかが分からないままに機械処理をするというのが不自然な印象でおり

ました。

一方で、BCCWJのようなコーパスは、誰がどのような状況で書いたものであるかという出典をたどることができる点が非常に重要で、コーパスの一つの大きな価値であると感じています。

それを前提にしつつ、今の生成AIを見てみますと、実は様々なテキストを組み合わせただけなのですけれども、非常に豊かな文脈を持っているように感じます。その文脈が正しいものであるかどうかは問題なわけですが、仮想でありながら、豊かな文脈がある。例えば、今日、私、開会の挨拶をするのに「こんにちは」という言葉でいいですかと生成AIに聞いてみたところ、「皆様、午後の眠くなる時期に晴れやかな声で「こんにちは」と言うことによって、場がぱっと明るくなるでしょう、しかも自然光中で言うとさらに映えることでしょう」みたいな、どこまでが現実でどこまでが仮想か分からないながら、非常に豊かな文脈を持って答えてくる。それがどこから生まれてくるものかというのも非常に興味深く感じておりました、BCCWJ2は更に分量も増えるということですし、様々な、言語というものに含まれる仮想的な文脈というものと対比できるような、例えば出典が分かっているものの強さを生かして検証的な役割を果たすとか、様々な使い道を今後模索していけるといいなと感じております。以上です。

続きまして、名簿順ですとアイウエオ順ということになりますけれども、皆様に順番にお願いをしていきたいと思っております。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。うまくまとまっていないのですが、感想を申し上げます。

この会は言語資源の整備・活用を考えるということなので、その範囲を超えるというか、ちょっと先のことになってしまうかとは思いますが、整備されたものをどう使っていくかという点—整備したものにどこまで責任を負わなければいけないのかという問題もありますが、日常的にそこに非常に危惧を感じるころではあります。ここでも何度かお話ししたのですが、特に教育現場において、学生は生成AIやデータを使って調査をするのですが、情報リテラシーが十分身に付いておらず、例えば、自分の見たデータにないものは「ない」と言うんですね、平気で。なので、そうではないこと、データに網羅されている範囲の限定性や、大分正確にはなってきたと思っておりますけれども、データのいろいろな偏りなど、そういうことの問題点、課題とともに認識していく必要性を言っていくことも大事ではないかと思

います。特に、私たちは、ここまでいろんな情報化が進む前の時代を知っていて、例えば、手で調べたり、紙の索引を引いたり、そういう作業の経験もあって、そしてデジタルデータが出てきたわけなのですが、デジタルネイティブーその世代の人たちは、用例を手作業で調べたり、データがどうやって作られているのかを理解しないまま、すぐに便利に使うのが当然になっているので、活用する際の課題というものについても、啓発していく必要があるのかなということ強く感じました。以上です。

○相澤主査

ありがとうございます。

では、神永委員、よろしく願いいたします。

○神永委員

2年間になりますけれども、コーパスなどの言語資源の重要性というのを改めて気付かされる機会を設けていただきまして、本当に有り難いなと思っています。辞書編集者としても、大変勉強になりました。

今期、私が御報告したことも、とても分かりやすくまとめていただきまして、有り難かったかなと思っています。ちょっと先走ったと言いますか、先の話なんですけれども、実は、私が今期御報告したことから更に、『日本国語大辞典』としては、国語研のコーパスと連携を取りながら、いろいろと作業を進めております。実際にそれを担当している者もおりまして、国語研のコーパスに限らず、ほかのものも含めて、今、検討しておりますので、できましたら、私よりも直接の担当者が御報告した方がいいかなと思うんです。そういった、国語研のコーパス、あるいは、それ以外のものも含めて、『日本国語大辞典』としてどのような展開を考えているのかということ、具体的な内容を報告するような機会を設けさせていただくとすごく有り難いなと思っています。

もう一つ、それに関してと言いますか、関連するんですが、前回、方言のことも申し上げましたけれども、『日本国語大辞典』には方言欄がありまして、第二版では昭和の戦前までの方言資料が主に基になって方言欄が設けられているんですが、それ以降の方言を全国規模でまとめようということ始めておりますので、こちらも担当者がおりますから、できましたら、その担当者にそういう話をできるような機会を設けさせていただくと有り難いなと思っています。もちろん、方言に関しまして全国規模で行っているものはいろいろと

あるんですけども、『日本語大辞典』も大きな存在の一つだと思いますので、そういった意味で、そういうことができれば有り難いなと思っております。

あと、外来語に関してなんですけれども、申し上げていますように、外来語コーパスで、頻度数とか、そういうを出していただくと国語辞典の見出しの語形に影響するものであります、もちろん、こうじゃなきゃいけないというものを決めていただきたいわけではないのですが、ある程度、よく使われる外来語に関しましては、こういった語形があるよというのを、異形も含めてなんですけれども、示していただけるとすごく有り難いなと思います。例えば、「カルシウム」なんですけれども、辞典によっては異形で「カルシウム」というのを入れています。私、割と最近見付けたんですけども、「カルチウム」というのがあるんですね。戦前に例があるんですけども、これをどうするのか。「カルチウム」なんていうのを使う人はほとんどいないかもしれませんが、ひょっとしたらどこかでまだ生き延びているのかもしれませんが、そういったことが分かると面白いですし、載せられる辞典は限られてくるのかもしれませんが、そういったものが参考になって国語辞典の見出し語が成り立っていくわけですから、できましたら、そういうものをコーパスから抽出してきたということで示していただければ有り難いなあとと思っております。

何か、感想というよりは、これからのお願いも含めまして、話をさせていただきました。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。御提案いただきまして、どうもありがとうございます。

齊藤(美)先生、追加で何かございましたら、よろしく願います。

○齊藤(美)委員

ありがとうございます。今年度、この小委員会を通じて、また、自身の大学ですが、職場での同僚や学生とのやり取りなどを通じて、先ほどの報告の中でも、学生がAIをどう使っていくかとか、教育の中での話もありましたけれども、私も正にそういう大学の現場で、言語資源のデジタルでの保存、あるいは整理を充実させる意義というのを改めて認識しました。

AIは、今日の経過報告でも言及されているとおり、実に様々な場面で活用が進んでいますので、自分が読めない言語で書かれた文書であっても、デジタルデータであれば、テ

キストデータになってなくても、画像データであっても、AIを活用して、ある程度、内容を把握できる可能性が増えているというのを実感しております。

国文学研究資料館での古典籍のデータベース化のように、日本の歴史上の情報、また、現在の情報についても同様ですが、非日本語使用者も対象として国内外に発信するためにも、デジタルの言語資源というのは有意義だなというのを改めて理解できました。私もですし、同僚や学生なんかも活用しています。

今年度も参加させていただいて、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。

武田委員、よろしく申し上げます。

○武田委員

本日の審議経過報告(案)につきましては、大変適切に今年度の検討をおまとめいただいていると思われておりまして、特に付け加える点はございません。

改めまして、昨年来、様々な観点で勉強させていただくことになりましたけれども、話者人口の多い日本語として、諸言語の状況を見るまでもなく、言語資源の構築と整備、そして利用が必要であるということは、全く賛成でございます。その点で、BCCWJ2の構築も進行しているということで、御苦労もあるかと思えますけれども、楽しみにさせていただきたいと思えます。

議論の中では、BCCWJ2は、生成AIの影響のない、人間だけが作っていた言葉の時代の最後の日本語のデータ集合になるかもしれないという観点も示されておりまして、大変興味深く受け止められるところでした。

また、BCCWJ2におきましては著作権についても配慮がなされているということで、この点も期待したいと思えます。

その他、様々な先行研究にも改めて敬意を表したいと考えております。

それでもなお、今後の課題もあるかと思われそうですが、外来語の表記の問題ですとか、地域の言語一方言の問題ですとか、そういったところは課題が多々残されていると思えますので、引き続き、言語資源を構築して、有効活用につなげられるようになればと考えております。

有効活用という点では、辞書での利用、『日本国語大辞典』をはじめとして、その他の辞書でも様々な活用の場があると思いますので、幅広く日本語の言語文化を豊かにしていくという観点でも有効かと思われまし、ブランディングというお話もありましたけれども、実際に地域なり日本の社会に影響を及ぼすような活用の仕方も十分考えられると思いますので、継続的に取り組んでいく必要があると思われております。

その点においては、財政的な支援が必要だろうと思われまし、是非、国としても御検討いただきたいと思いま。

簡単ですけれども、以上でございます。

○相澤主査

ありがとうございました。

では、前川委員、よろしくお願いたします。

○前川委員

今年度の取りまとめ、どうもありがとうございます。大変手際よくまとまっていたと思いま。

私はどちらかという作る立場に近いところにおいま人間なので、その意味で、この委員会ではどういふうに使われるかというい観点からのいいろいろな御意見を聞くことができたのは、大変有り難かったと思いま。作る側、利用する側、どちらの観点からしても、私、一つ気に掛かるとい言いまるか、この小委員会で一度押さえていく必要があるのかなあと思いまのは、知財関係のことなんですね。私はほんの数日前に人間文化研究機構のDH研究会が主催した知財関係の講習会みたいのをちょっと聞いたんですけども、そこで聞いた話によると、EUなんかでは大分、そういう言語資源に関係するい形でのいいろいろなテキストの利用に関するガイドラインとい言いまるか、そういうものが、EUの規約とい言いまるか、そういう形で決まってきたとい。実は私、そういうことを知らなかったんですけども、言われてみると、ヨーロッパ、特にドイツ語圏なんかでのコーパス整備が最近ちょっと変わってきたといですね。非常に進度が早くなってきたといんですけども、そういうことの内にはそういう法整備の問題があるのかなあと感じた次第です。私自身もそこら辺は知識が全然ないので、できたら、今後、この小委員会なんかで、知財関係の、特に海外の事情をよく御存じの方をお招きして、そこら辺を説明していただくとい言いまるか、作る

側としても、利用する側としても、大変有益ではないかなと、最近感じた次第です。以上です。

○相澤主査

ありがとうございます。

では、森委員、よろしくお願いいたします。

○森委員

今年度から委員に加えていただきまして、ありがとうございました。1年のまとめも拝見しまして、こんなにもいろんなゲストスピーカーの皆様、専門家の皆様からお話が聞けたんだということを、改めて有り難く思いました。学びました。

社内外で文化庁の委員というのは何をやっているんだと聞かれたときに、コーパスという言葉を知らない人とか、存在を知らない人が思った以上に多いということに気がしました。研究者とか専門家のためだけではなくて、広く利用されて親しまれるものになっていくといいなあというのが、1年の間に感じたことです。文化庁が発表されている言葉の調査も、これが正しい言葉で、最近、誤用が広まっていますという言い方ではないはずなんです。言葉には、○か×かだけではない、グラデーションが必ずあるはずで、時代の変化とか、地域性とか、そういうものも見えてくるのがコーパスだと思うので、もっと気軽に楽しく使えるような入り口もあっていいんだろうなと思いました。それが広められるような役割も、マスコミとしては担っていきたいなとも考えております。ありがとうございました。

○相澤主査

ありがとうございました。

では、一巡しましたところで、幾つか御意見を頂いたことと思います。簡単に俯瞰いたしますと、まず、昨年1年間のまとめを受けて、「コーパスベースアプローチ」ですとか、「社会課題の解決」というキーワードが出てきて、それは今期も引き続き取組として注目をしてきたところかと思えます。本日頂いた御意見の中で「教育におけるAI活用」というキーワードも出てまいりましたので、それについても、もしキーワードとして取り上げることができれば、検討できればいいと感じました。

あと、方言と外来語の重要性ということで、コーパスとともに、双対となる辞書との関係

性についても今後の課題として挙げたということと、次に結び付く課題として、前川委員より、知財関係について、海外の状況を踏まえて、より深く知る機会があるといいという御提案がございました。

まだまだ尽くせないところも多いと思いますので、今まで頂いた御意見に加えまして、さらに配布資料2について反映しておいてはどうかという御意見ありましたら、この場でよろしく願いいたします。（→ 挙手なし。）

よろしいでしょうか。そうしましたら、本日頂きました御意見に基づいて、配布資料2をアップデートする点については、主査に御一任いただければというふうに思いますが、よろしいでしょうか。

（「異議なし」の声あり）

どうもありがとうございます。

では、この場でお認めいただいたということで、配布資料2についての審議は終了したいと思います。

既に御発言も頂いたところですが、来期も言語資源小委員会を置くことが検討されていると伺っております。来期取り上げるべき事項や課題、こんな話が聞きたいということを含めまして、もし追加でありましたら、よろしく願いいたします。方言、外来語、知財等ですね。あと、今期御報告いただいたことの中には、データ駆動型の人文学という話題もございました。それから、教育における言語について、こんなお話を聞けるとうれしいというふうなことがありましたら、この場でも、後ほどでも、お知らせいただければと思います。（→ 挙手なし。）

よろしいでしょうか。そうしましたら、本日、議事は以上でございます。本日の配布資料は、国語分科会に提出をしまして、今期の審議経過として報告する予定でございます。それについてお認めいただいたということにさせていただければと思います。

もし、言い足りないこと、言い残したことがございましたら、よろしく願いします。

（→ 挙手なし。）

よろしいでしょうか。それでは、今期最後の言語資源小委員会、最後になってまいりました。国語課長から御挨拶いただければと思います。よろしく願いいたします。

○山田国語課長

国語課長の山田です。委員の皆様、大変御多忙の中、この1年間、審議に参加していただきまして、ありがとうございました。

この言語資源小委員会ですけれども、国語に関する審議会の中ではかなり新しい部類の委員会かと思えます。言語コーパスを中心としまして、様々な場面での活用の在り方、地方における言語資源、そして、外来語に関するコーパスの活用など、この1年で多数の専門家の方を迎えまして、様々な視点から幅広く御検討いただくことができたと思っております。また、この1年の審議を通じまして、言語資源の整備の重要性、今後の活用の可能性、また、それに伴う様々な課題など、広く議論を深めることができたと思っております、この場をお借りして、改めて皆様にお礼を申し上げます。

来年度につきましては、これから検討ということになりますけれども、本日、皆様から頂いた様々な御意見、それから、文化庁の方でも、来年度、事業をまた展開しますので、BC CWJの更なる整備、方言については新しく事業を拡充する予定となっていることでもありますので、こういった進捗なども踏まえまして、さらに、言語資源に関する施策の充実に向けまして、引き続き、継続される委員の皆様におかれましては、御協力を頂ければと思っております。

生成AIの関係で、教育での利用に関して、今、委員の皆様から少し御意見を頂戴しましたけれども、この委員会は国語に関する施策の委員会でありますので、そもそも、教育におけるAIの功罪といったようなもの、教育の在り方そのものを大きく議論するということはちょっと難しいのかなと思えますが、飽くまでも、国語施策の中でと言うか、国語施策という観点から、取り上げることについては考えさせていただければと思っております。

引き続き、よろしく願いいたします。1年間、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した。

様々な取りまとめ、また、多くの興味深い御講演等の企画運営等を含めまして、大変お世話になりましたことを、私からも御礼申し上げます。

それでは、本日の言語資源小委員会はこれにて閉会といたします。ありがとうございました。