

第4回 国語分科会言語資源小委員会（Web開催）議事録

令和6年11月28日
10時00分～11時55分
文部科学省5階・5F2会議室

〔出席者〕

（委員）相澤主査、石川副主査、植木、神永、齊藤（美）、前川各委員（計6名）
（ゲスト）杉山香織西南学院大学外国語学部教授
（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3回国語分科会言語資源小委員会議事録（案）（委員限り）
- 2 フランス語コーパスの現状と活用事例（杉山香織氏御提出資料）
- 3 言語資源の社会実装―「集めた後」を見据えて―（石川副主査御提出資料）

〔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 杉山香織氏から配布資料2「フランス語コーパスの現状と活用事例」に基づいて説明があり、質疑応答が行われた。
- 5 石川副主査から配布資料3「言語資源の社会実装―「集めた後」を見据えて―」に基づいて説明があり、質疑応答が行われた。
- 6 二つの説明を踏まえ、意見交換が行われた。意見交換の後、国語分科会におけ

る言語資源小委員会の審議状況の報告については、相澤主査に一任することが確認された。

7 次回の言語資源小委員会について、令和6年12月26日（木）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。また、国語分科会については、令和6年12月10日（火）15時から17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。

8 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

それでは、定刻になりましたので、ただ今から第4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に実際にお越しくくださった方々とオンラインで御参加の方々とがいらっしゃいます。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ということで、御承知おきください。

さて、本日は議事次第のとおり、「言語資源の在り方について」、「有識者による事例等紹介」、「その他」という内容で協議を行いたいと思います。事前に資料を御覧になっていただいているとおり、これまでの審議内容を踏まえまして具体的な情報ということで、海外のコーパスの状況、活用事例について、西南学院大学の杉山香織先生からフランス語について、また、石川副主査に英語コーパスを社会課題解決に活用している事例ということでお話いただくこととなっております。意見交換では、このような事例に倣って日本におけるコーパスの可能性なども含めまして、皆様から広く御意見いただければと思っていますのでよろしくお願いいたします。

では、議事次第では2となっている事例等紹介、1番目の御紹介ということで、西南学院大学の杉山先生から「フランス語コーパスの現状と活用事例」と題して状況を御紹介いただくこととなっております。

杉山先生、よろしくお願いいたします。

○杉山発表者

おはようございます。本日は、どうぞよろしくお願いいたします。西南学院大学

の杉山香織です。本日は、フランス語コーパスプロジェクトの事例を紹介してまいります。どうぞよろしくお願いいたします。

発表の流れは、このようになっております。初めに、フランス語の主要なコーパスを説明いたします。次に、フランス語コーパスプロジェクトの独自性について述べます。最後に、コーパスを活用した成果について幾つか事例を御紹介いたします。

まず、主なコーパスについて、書き言葉コーパスと話し言葉コーパスに分けて御紹介いたします。今回は書き言葉コーパスとしてFRANTEXT、話し言葉コーパスとしてESLOコーパスとPFCコーパスを取り上げます。書き言葉コーパスであるFRANTEXTは、フランス語のコーパスとしては最大規模となります。もちろん、インターネット上のフランス語を収集したフランス語のT e n t eですと2023年時点で238億語と規模が桁違いとなっておりますが、出典や著者が明らかであり、体系付けられて構築されたコーパスですと、このFRANTEXTが最大だと言えます。総語数は2億7,200万語で、9世紀から21世紀のテキストが収録されています。こちらの右下のグラフは、世紀ごとのテキストの収録数を示しています。ここからお分かりのとおり、19世紀から20世紀のテキストが主となっています。テキストの内訳は、文学作品が約80%、非文学テキストが約20%となっております。非文学テキストには新聞や雑誌などが含まれております。こちらがURLです。

FRANTEXTは、ATILFというフランス語情報処理分析センターが運営しています。ATILFは2001年1月に設立されて、CNRS（フランス国立科学研究センター）とロレーヌ大学が共同で運営しています。フランス語の言語学研究に特化した研究機関であり、辞書編纂や言語データ^{さん}のデジタル化を行っています。日本の国立国語研究所に当たると言えます。日本学術振興会の中に研究ユニットがあって、その研究ユニットとある大学が共同して国立国語研究所を運営しているという図をイメージしていただくと分かりやすいと思います。

FRANTEXTのプラットフォームは、御覧のとおりです。誰でもアクセスできるこのプラットフォームでは、単語の頻度を検索したり、単語がどのように文脈の中で使用されているかを表示したり、また、単語と共に使用される語を検索したりすることもできます。今、皆さんに御覧いただいているのは、二つの単語がどのような文脈で使用されているのかを調べるために使用するページです。検索する単語はそのままでもよいですし、正規表現で条件を細かく指定することもできます。

時代区分も指定できます。また、辞書系か活用系かなども指定できますし、品詞を指定することも可能です。

こちらがM a i s—英語で言うB u tです—、これとO u i—英語で言うY e s—という二つの単語がどのような文脈で使用されているのかを検索した結果の提示例です。作家や出典、作家がこちらで、出典、また、出版年などが分かります。また、この作品のどの部分でこれらの2語が出現するのかを見ることができます。ページ数が記載されています。どちらかというと、このコーパスは、これまで文学者などによって多く使用されてきました。

また、言語学者でも用例検索で用いられることも多かったです。ただ、このコーパスを用いて定量分析するには、提供されているプラットフォーム上では不十分であると言えます。そのため、もう少し踏み込んだ統計分析を行うのであれば、コーパスをダウンロードして分析目的によって異なる分析ソフトを使用する必要があります。FRANTEXTは研究センター、図書館、研究者、教員、学生は教育・研究目的で 사용할 ことができます。ただし、全部のデータにアクセスするには有料となっており、機関での契約と個人での契約が可能になっています。また、有料版であってもFRANTEXTに収録されたテキストの中には、ダウンロード可能なものとそうでないものがあります。

次に二つの話し言葉コーパスを紹介します。一つ目のESLOコーパスは社会言語学の調査を目的としたコーパスです。そのため、年齢、性別、社会的背景の異なる話者を含みます。また、このコーパスの特徴は、二つの時期に分けて調査が行われているということです。1回目が1968年から1974年の間、2回目が2014年以降からデータを収集しています。そのため、通時的研究が可能になります。また、かなり古くから話し言葉コーパス構築に着手しているという点も注目に値します。ファイル数は1,289で、750.5時間の会話が録音されており、総語数は約1,500万語です。様々なトピックを扱っており、大学の講義、講演といったモノログでフォーマルなものから、食事中の会話といった複数人の会話でインフォーマルなものまであります。こちらがURLです。

ESLOコーパスは、ロレーヌ言語研究所が運営しています。この研究所はフランス国立科学研究センター、オルレアン大学、トゥール大学、フランス国立図書館の共同ユニットになっています。ESLOコーパスのインターフェースは、このよ

うになっています。インターフェース上では、ファイル選択、単語検索、単語リストを検索することができます。コーパスの中身を検索するためには、真ん中のところをクリックします。すると、このような画面になります。

このページでできることは、以下のことです。まず、1960年代から1970年代のデータ、E S L O 1 か、それ以降の、2014年以降の現在のデータか、あるいは両方のデータを検索することができます。次にトピックを選択します。例えば面接、講演、映画などから選択できます。また、話者の年齢、性別、学歴なども指定できます。そして、調べたい単語がどのような文脈で使用されているのかをチェックするために調べたい単語を入力します。表出系だけではなく、単語の一部を指定したりすることもできます。

今回は、現在のコーパスから食事中の会話コーパスを選択して、全ての話者を選択しました。そして、「M a i s」という単語を調べました。指定した単語がどのように文脈内で使用されているのかを見たり、その箇所を再生したりすることができます。また、結果をC S Vファイルに保存することもできます。しかし、先ほどのF R A N T E X Tと同様に、このインターフェースだけでは細かい統計分析は、することができませんので、用例検索以上の分析を行う際には、コーパスをダウンロードする必要性が生じます。E S L Oコーパスのダウンロードについては、後ほど説明いたします。

二つ目にP F Cコーパスを紹介いたします。こちらは現代フランス語音韻論と名前が付いているとおり、現代の会話が収録されている音韻論の研究を目的に作成されたコーパスです。1999年からこのプロジェクトが始まりました。49地点で調査を行っています。調査地点にはフランス本土だけではなく、レユニオン島などの海外県並びに海外地域圏やカナダやルイジアナ州などの北アメリカ、セネガルなどのアフリカも含まれます。年齢、性別、社会的背景の異なる400人以上のデータが収録されており、被験者は単語リストの音読、テキストの音読、インタビュー、自然会話などのタスクを行っています。

E S L Oコーパスと異なる点は、自由会話やインタビューのコーパスだけではなく、単語リストの音読やテキストの音読が含まれているという点です。これらは地域フランス語に特有の音素を分析したり、フランス語特有の音韻現象を分析したりするのに役立ちます。

P F C コーパスは、中心となる研究所はあるものの、世界各国の研究者が連携して運営している国際プロジェクトです。元々フランスの研究者2名とノルウェーの研究者1名でこのプロジェクトが始動しました。現在はノルウェー、フランス、スイス、日本、韓国、ドイツの研究者がこの運営者として、このプロジェクトを管理しています。早稲田のシルヴァン・ドゥテ先生が運営者の一人を務められております。

P F C のインターフェースは、このようになっております。録音地点、リエゾンやシュワ（曖昧母音）といったフランス語の音韻体系に特徴的な現象の項目、そして転写という項目が設けられています。例えば転写から「M a i s」という単語を検索すると、このような画面になります。話者のID、調査地点、検索語を含む文脈を見ることができます。音声、音韻を研究する目的で構築されたコーパスという特性があることから、検索語を含む一部について音声を聞くことができます。また、結果はCSVファイルに保存可能です。

先ほど少し触れましたリエゾンといった音韻現象についても検索できます。今回は簡単に説明すると、リエゾンとは単独で発音されない音が後続する母音とともに出現するという現象です。フランス語では、してもしなくてもいいリエゾンというものが存在します。選択的リエゾンと呼ばれていますが、選択的リエゾン使用の割合が話者の学歴や話す状況によって異なるということは、先行研究から指摘されていることです。ここではZがリエゾンとして出現するか、しないかを調べた研究の一部を示しています。このように音声・音韻研究と社会言語学的な研究が行われています。こちらのプラットフォームも先述のFRANTEXTやESLOコーパスと同様、定量的分析には適しません。コーパスのダウンロードが必要になります。

次に、フランス語に関するコーパスプロジェクトの独自性について触れていきます。フランス語コーパスに関しましては、今まで紹介してきた代表的なコーパスも含めて、大規模コーパスと呼べるものは少なく、小規模コーパスや中規模コーパスが多く作成されてきました。研究所や研究ユニットごとに作成されたコーパスというのが一因にあります。それぞれ研究の興味が異なるからか、複数の研究機関が協力して作成するコーパスというものも少ないです。そこで、フランスでは2012年に個々のコーパスを一元化するORTOLANGというネットワークインフラを構築しています。こちらがURLです。

ORTOLANGの目的は、言語データ、つまり、コーパス、辞書、専門用語などのレポジトリとその分析処理のためのツールを含む一元化されたネットワークインフラを構築することでした。このプロジェクトは、首相府投資総務庁というところが主導する未来投資プログラムの一環です。この未来投資プログラムは、フランスの成長と雇用の可能性を高めるため、革新的で有望なプロジェクトに資金提供をしています。特に戦略的分野において革新的なプロジェクトをターゲットにしています。その中に高等教育・研究も含まれています。期待される成果としては、一つ目が、質が保証されたリソースをプールすることにより、言語の分析・モデリング・自動処理に関する研究を国際的な最高レベルまで促進すること。二つ目が、公的研究機関内に設置されたリソースやツールの利用を促進し、産業パートナーと共有すること。その中でも特に言語処理のためのリソースやツールを開発できないことが多い中小企業との共有を促進すること。三つ目が、公的研究機関が獲得した専門知識を共有することにより、フランス語とフランスの地方言語を促進することを挙げています。

こちらは、ORTOLANG内の個々のコーパスの提示例です。コーパス名、コーパス構築に携わった研究所、協力団体名がまず記載されています。そして、中心のこの青い部分は引用形式です。ORTOLANGは、研究チームでなく、個人でもコーパスをアップロードすることができます。ただし、品質保証のため、コーパスがきちんとしたガイドラインに従って構築されているのかなどについての審査があります。知的所有権が尊重されるため、資源の出典情報についても明記されています。こちら、コーパスを使用する際には、この引用が義務付けられています。

こちらに、下にコーパスの説明があり、キーワードが挙げられています。そのプロジェクトのリンク先も示されています。右側を見ていただくと、開発者へのメールを送る、フォームへのリンクやダウンロードページへのリンクが示されています。SNSへの共有も可能です。また、ORTOLANGには検証があり、管理者、構築者、使用者に関わる条約が存在します。フランス語だけでなく、英語版も存在します。こちらは英語のページになっています。

現在、ORTOLANGは書き言葉コーパスが78、話し言葉コーパスが193、マルチモーダルコーパスが53リソース収録されています。マルチモーダルコーパスには主に動画が収録されたものです。先ほど紹介したESLOコーパスやPFCコーパ

スもORTOLANGからダウンロードすることができます。ただし、プロジェクトで構築されたコーパスの全てではないようです。例えばPFCコーパスですと、匿名化されたデータのみ公開されています。また、ダウンロードできるデータについては、各プロジェクトで形式が異なっており、HTML形式のもの、テキスト形式のもの、テキストグリッド形式、オーディオ形式のものなど様ではありません。

フランス語だけではなく、英語やドイツ語などのフランス語以外の言語のコーパスも収録されています。今回、この発表に向けて調べていく上で、ORTOLANGの総語数等を調べてみましたが、総語数は明記されていませんでした。コーパス以外にも辞書、専門用語、そしてコーパスツールも収録されていると先ほど申し上げました。それぞれ簡単に例を御紹介します。

まず、辞書には特定の地域フランス語辞典や類義語辞典などが収録されています。CSV形式やXML形式で公開されています。専門用語に関するセクションには、39ファイルありますが、そのうち36ファイルがINISTという科学技術情報研究所が提供するデータになっています。気候変動や生物多様性などの項目が多く収録されています。例えば気候変動については、科学論文のコーパスと国際気候変動委員会の報告書に基づいた用語を抽出したものがあります。このようにドキュメントがXML形式で保存されているものが多いです。コーパスツールとしては、音声認識や依存構造分析に関するものがあります。今御覧いただいているのは、音声自動アノテーションツールのページです。自動転写だけではなく、音声と転写部分をひも付けたりすることができるツールです。

ここまでフランス語の主要なコーパス三つとORTOLANGというネットワークインフラについて紹介してまいりました。最後にフランス語のコーパスがどのように社会や人々に還元されているのか、そして活用されているのかについて幾つか御紹介したいと思います。辞書作成において、コーパスが重要な役割を果たしているというのは、どの言語のコーパスでも同じであると思います。フランス語もTrésor de la langue françaiseという辞書があります。19世紀から20世紀までのフランス語の単語、表現、用法、派生語、語源が収録されています。Trésor de la langue françaiseという辞書は紙媒体として1971年から1994年に掛けて出版されたものです。フランス語の膨大な資料を集めて辞書を編纂^{さん}することを目的に、1957年にプロジェクトが始動し

ました。この資料収集がFRANTEXTの先駆けとなったものです。

辞書には、各単語の歴史的変遷や文学作品内での使用例が記載されています。このように文学作品はもちろんのこと、哲学や科学のテキストから豊富な引用例がありますが、こちらはFRANTEXTから取ってきたものです。1994年に全16巻が出版され、紙媒体でのプロジェクトが完了しましたが、辞書の内容をデジタル化し、インターネット上で公開されています。こちらがTLFiというものです。研究者だけでなく、誰でもアクセス可能になっており、*Trésor de la langue française*、TLFとその電子版であるTLFiはフランス語の歴史的及び現代的な用法を理解するための重要な資料として幅広く使用されています。こちらがTLFiのURLになっています。

この辞書はFRANTEXTと同様にフランス語情報処理分析センターが管理しています。ただし、紙媒体の辞書については、1994年にプロジェクトが終了しており、当時、携わっていた研究者たちは職を離れてしまっています。また、電子化された辞書の方も紙媒体の辞書をそのまま電子化したものであり、内容は紙媒体のものと同一です。そのため、辞書の改訂は30年間行われておらず、時代に即さない内容が含まれている可能性もあります。

コーパスが教材や評価に使用されている例は、日本語コーパスでも、英語コーパスでも珍しくないと思いますが、二つ目のコーパス活用例としてフランス語の事例を御紹介します。まず、FLELexというツールです。こちらはベルギーのルーヴァン・カトリック大学のCENTALという自動言語処理センターが開発したものです。このツールの開発に向けて、まずはCEFR準拠の教科書をコーパス化することから始めました。そのコーパスに見られる出現頻度等を基に調整を重ねて、単語の難易度をCEFR基準で提示するツールを開発しています。例えば御覧いただいている例は、「星の王子様」の冒頭部分になります。薄い緑色がA1、濃い緑色がA2のように単語にCEFRレベルが付与されています。インターフェースでこのように示すことも可能ですし、単語の辞書系とそのCEFRレベルなどの情報が記述されているCSVファイルをダウンロードして、独自の研究に応用することも可能です。こちらは教材の作成や学習者の語彙習得状況の評価に役立っています。こちらがURLです。

同じくCENTALが開発したものにCEFR-FR Ratorというものが

あります。これはフランス語学習者のライティングの自動評価システムです。このシステムを作成する上で、TCF—フランス語学力テストというテストの中にあるライティングセクションの学習者の作文をコーパス化しています。そのコーパスには学習者の最終的なレベル評価、学習者の母語といったメタデータも含まれています。作文コーパスを語彙的、形態統合的観点から変数分析を行い、さらに機械学習を行うことで学習者の作文にCEFRレベルを自動付与するAIモデルを作成しました。CEFRはリーディングセクションとリスニングセクションは多肢選択式のため、採点が容易なのですが、オプションとして設定しているスピーキングとライティングセクションは自由回答のために、採点にかなりの時間と労力を費やしてきました。このシステムは、まだ実用化されてはいないのですが、実用化に向けて準備段階です。導入されればコストカットが期待できます。

このプロジェクトは、フランス教育国際機関の協力の下、フランスの公共事業変革基金やベルギーの若手研究者向けの科学リーダーシッププログラムの助成を受けています。フランス教育国際機関は、公的機関であり、国民教育・青年省、日本で言う文部科学省の監督下にあります。先ほど申し上げたTCFという試験の運営団体でもあります。このように国を超えて公的機関が支えているプロジェクトであると言えます。

また、地域フランス語の研究にもコーパスは用いられています。この現状のタイトルを日本語に訳すと、「フランス語圏で話されるフランス語の多様性、教育リソース」となります。タイトルから分かるように、教育でフランス語の多様性を扱うことができるように、学生や教師向けに書かれた本です。日本語訳も出版されており、私も編訳を担当しています。PFCコーパスで紹介したように、フランス本土だけではなく、スイス、アフリカ、海外県と海外領域圏、北アメリカで収録された自由会話を基に、その地域特有の表現や音韻的特徴を説明しています。また、音声データもDVDの中に収録されています。こちらが、この本のURLになっています。

これまでフランス本土の話を中心に行ってまいりましたが、最後のコーパス活用例として、ケベックにおける言語政策を取り上げようと思います。ケベック州は御存じのとおり、カナダのフランス語圏になります。カナダでは1969年に公用語法が制定され、英語とフランス語を公用語とすることが定められています。しかし、各

州は自らの言語政策を制定する権限を持っています。ケベック州では、フランス語が唯一の公用語として法的に定められています。その一方で、ケベック州内では英語話者が経済的な権力を持つことが多く、フランス系住民は其中で劣位に置かれることがあります。このため、フランス語を強化することで、自らの経済的地位を向上させる狙い、動きがあります。

また、フランス語は単なるコミュニケーションの手段ではなく、その文化的アイデンティティの核心です。フランス語を守ることはケベック州民としてのアイデンティティを維持するためにも重視されています。このため、州政府はフランス語の保護や使用促進に力を入れています。最近では2020年にケベック州法第95号が制定されました。これはケベックにおけるフランス語の地位をより強化するものです。また、ケベック州には、フランス語の地位を保護、促進するためのケベック・フランス語局という公的機関も置かれています。

このようにケベック州では、フランス語の地位の保護に対してかなりしっかりと政策が敷かれています。コーパスに関する事例として、*Grand dictionnaire terminologique*という大用語辞典を挙げるができます。これもケベック・フランス語局によって作成されたものです。オンラインで利用可能な用語のデータベースです。法律、経営、情報技術、健康などの様々な分野における専門用語の定義や翻訳を提供することを目的としています。各用語のページには、定義、フランス語と英語の用語、同義語、イラスト、略語が含まれています。また、英語の用語に対するフランス語の同等語を提供するだけでなく、避けるべき用語や適切な代替案を提案することもあります。この辞書の作成や新語の追加を検討するに当たって、ケベック州における言語使用状況を把握するためにも、コーパスが使用されています。このようにフランス語の地位保護に向けてもコーパスが一役買っていることが分かります。

以上、大変駆け足となりましたが、フランス語のコーパスの現状と活用事例を御紹介いたしました。御清聴、どうもありがとうございました。

○相澤主査

杉山先生、どうもありがとうございました。

それでは、ただ今の御説明について御質問がありましたら、よろしくお願いいたします

します。意見交換は、この後の石川副主査のお話の後にまとめて行う予定でありますので、御質問ということで、よろしくお願いします。何かございますでしょうか。

○石川副主査

よろしいでしょうか。

○相澤主査

はい。お願いいたします。

○石川副主査

杉山先生、大変啓発的なお話をありがとうございました。1点質問なのですが、FRANTEXTには文学作品が多く入っているというお話であったわけですが、これは作品中の一部をカットして入れているのか、あるいは作品を全部入れているのかという辺り、いかがでしょうか。

○杉山発表者

全体を見渡したわけではないのですが、著作権が切れているものに関しては全体が入っていることは見えていますが、全部見ていないので、もしかしたら一部の収録のものもあるかもしれません。

○石川副主査

分かりました。ということは、どの作品からも同じ量だけ抜いているというより、取れるものは全部入れておこうという、そのような方針で作られているということでしょうか。

○杉山委員

はい。そうです。

○石川副主査

分かりました。ありがとうございます。

○相澤主査

ほかは、いかがですか。

○齊藤（美）委員

よろしいでしょうか。

○相澤主査

お願いいたします。

○齊藤（美）委員

齊藤と申します。御発表、ありがとうございました。大変興味深く伺いました。ORTOLANGのことで少し伺いたいのですけれども、期待される成果というところで、具体的にいろいろ挙げられているのをなるほどと思って伺っていたのですが、この二つ目にあった産業パートナーと、中小企業と杉山先生はおっしゃっていたかと思うのですけれども、共有することというのは何か具体的にこのような分野の企業とかあったり、事例があったりするのでしょうか。もし御存じのことがあれば伺えたら幸いです。

○杉山発表者

一応、ORTOLANGのホームページ上では、そのようなことが書いてありますけれども、実際の事例としては、まだ挙げられていなくて、もしかしたらあるかもしれないですけれども、そこまでまだ調べられておりませんでした。一応、目標として掲げられてはいます。

○齊藤（美）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、いかがでしょうか。

では、私からも1点、ORTOLANGについてですけれども、多様なコーパス作成者がアップロード、申請をしてクオリティーのチェックを受けた後に共有できるという仕組みが非常にいいと思ってお伺いしていたのですけれども、そのときにツールを共有する仕組みもあるということでありましたので、何か共通のフォーマットのような指定があったりするものなのでしょうか。

○杉山発表者

見てみると、各自、もともとORTOLANGというプラットフォームができる前から作成したコーパスとかも多く、それぞれ作ってきたので、中身を見ると結構、形は全然統一されていないという感じです。

○相澤主査

なるほど。分かりました。

○杉山発表者

ただ、何かツールに関しては、開発者が恐らく同じ研究所とか、元々同じところにいた人とか、そのようなこともあって、何となく似たような一例えばツールで多かったのは音声系の分析ツールなので、そのようなところに所属していた人たちとか、今でも研究している人たちが上げていて、プラートをもう少し専門的な分析ができるようにちょっと改良したというか、そのような系のツールが結構多かったです。コーパスに関しては、フォーマットはまだ全然一様ではないという感じです。

○相澤主査

分かりました。ありがとうございます。では、ツールは音声認識など比較的に汎用的な処理のものが中心で、場合によってはコーパス、あるコーパスを扱うためのツールなども含まれるかもしれないけれども、それほど一般的ではないという。

○杉山発表者

はい。

○相澤主査

ありがとうございます。

ほかは、いかがでしょうか。

○前川委員

よろしいですか。

○相澤主査

はい。お願いいたします。

○前川委員

国語研の前川です。二つ伺いたいのですが、一つは全く趣味的な話なんですけれども、FRANTEXTの方が9世紀からとなっていましたよね。

○杉山発表者

はい。

○前川委員

あれ、何でしたっけ、最初のフランス語と言われる *Serments de Strasbourg*（ストラスブールの宣誓）でしたっけ、あれから入っているということなんですか。

○杉山発表者

どの文脈か、どこの作品かというのはまだ、調べれば分かりますが、パッと答えられません。

○前川委員

多分、9世紀というとかさうかなという気がするのですがけれども…。それは趣味的なことなんですけれども、もう一つ、ORTOLANGの方で、いろいろツ

ルとか、それから、音声的なコーパスとかをディストリビュートしているわけですが、たまたまフランスの、同じフランスに拠点を置いている—これは民間の企業、有限会社みたいな企業になると思いますけれども、エルダというのがあるんですね。E l d a と書いてエルダという。これは言語資源全体の配布をヨーロッパで担っているかなり存在感があるというか、そのようなところなんですけれども、そことのすみ分けみたいなものについては何か御存じだったら教えていただきたいんですけども…。

○杉山発表者

エルダについては、全く存じ上げませんでした。

○前川委員

そうですね。はい。分かりました。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、いかがでしょうか。（→ 挙手なし。）

では、また後ほど意見交換もしますので、質問は以上といたします。

続きまして、石川副主査に社会課題解決の一助としてコーパスを活用している事例、配布資料3をお使いいただきまして、御発表をいただきます。よろしくお願いいたします。

○石川副主査

それでは、始めさせていただきます。本日は「言語資源の社会実装—「集めた後」を見据えて—」というテーマでございます。御覧いただいております七つのトピックについて順次お話を進めてまいります。

1. 言語資源開発に必要な3つの観点

まず、一つ目、言語資源開発に必要な三つの観点ということです。私は前回の報告におきまして、データを集めた終わった段階においては、(1)継続的なデータ収集、

つまり、新しいデータを入れてそれをアップデートしていくこと、(2)多くの人に使ってもらえるような普及のための工夫、そして、(3)収集した言語資源を社会とつなげていくこと、以上3点が重要であるということを申し上げました。本日は、この3点目に絞ってお話をいたします。

2. 言語資源の社会実装を考える必要性

続いて、配布資料3にございます「言語資源の社会実装を考える必要」ということでありますが、「コーパスに価値がある」、「コーパスが非常に重要なものである」ということは、ここにいらっしゃる委員の皆様を含めまして、専門家には自明でございます。しかしながら、余りに自明であるがゆえに、専門家集団の外に向けて、そのことを強くPRするという努力は、私自身を含めて不足していたところがございます。言語学以外の分野でも、あるいは言語学の分野内でも多くの研究ニーズがあり、公的な支援を必要としているところは非常にたくさんあるわけでございます。その中で、あえて言語資源の整備・拡充に支援を求めようとするのであれば、言語学者にとってのメリットを示すのみならず、それを超えて、広く国民に対する有益性や社会的課題への解決の可能性を示していくこともまた重要であろうと考えるところです。

3. 言語資源を社会に開く—応用言語学の方向性—

このように考えていきますと、言語研究を社会に開くということになってまいるわけでございますが、この点に関しましては、アメリカやヨーロッパにおける応用言語学の歴史を考えることも有益かと思えます。御承知のように、応用言語学は、主として理論系の言語学の中から枝分かれするような形で出てまいりました。応用言語学は、社会の問題が先にありきでして、主として言語学的なアプローチでその問題の解決に迫っていきます。

Widdowson (2000) も、「言語学の応用」と「応用言語学」は違うのだということを申しております。「言語学の応用」というのは、飽くまでも視点は言語学の枠内にあり、ウェイトは言語の方でございます。一方、「応用言語学」というのは、言語学を超えた部分、社会の方の課題を出発点としており、ウェイトはむしろ社会の側に置かれているわけです。このように考えますと、広義の言語研究に含まれる諸分

野は、言語ないし社会へのウェイトの置き方を軸に整理することができるかもしれませんが。お示ししている図では、左側が言語により焦点を当てたもの、右側はより社会に焦点を当てたものとなっています。コーパス言語学や辞書学というのは、一般に、用例重視の実証系の言語研究とみなせます。文脈を持ったテキスト、つまり、社会とのつながりを持った資料を使って研究を行うという点である程度の社会とのつながりはあるわけですが、どちらかと言えばウェイトは言語の中にあり、社会との結び付きの強さは必ずしも強くありません。しかし、コーパス研究のデータや分析の枠組みをうまく生かすことで、言語資源を用いた研究をより社会の方に近づけていくことは十分に可能であろうと考えます。

4. CASSの取り組みに学ぶ

ここで英国のCASSの事例を御紹介したいと思います。イギリスには政府が管轄するイギリス研究技術革新機構—UKRI というのがございます。UKRIの下には、分野別の評議会がございます。人文も、自然科学もあるのですが、その中に、経済学・社会学を統括するESRC—英国経済社会研究評議会というところがございます。このESRCが一定の資金を拠出しまして、ランカスター大学の中にESRC Centre for Corpus Approaches to Social Science—略してCASSというセンターが設置されているわけです。

CASSのミッションについてでありますけれども、まず1点目はコーパスアプローチの社会的活用ということです。コーパスそのもの、コーパス言語学そのものではありませんで、コーパスの分析技術やコーパス的な言語データの見方を中核とするコーパスアプローチを、これまでそうしたものを使ったことがない社会科学分野に新規に応用していくということですね。2点目は、コーパス言語学に専念するのではなく、言語分析の最新技術を社会科学の様々な問題に応用するという事です。3点目として、これは5年間の集中投資のプロジェクトですが、これにより、今後15年間、人材を育成していく体制を作るということですね。4点目は、これらによって社会科学におけるコーパス技術の普及を促進するという事です。

全体を総括して、「世界の中で、英国において非常に大きな成功を収めたコーパス言語学をベースとして、同じ英国を舞台として、今度は、「コーパス言語学を応

用した社会科学」という新しい成功例を示し、世界に示していく」という構想が示されています。

CASSのプロジェクトでは、もちろん文学ですとか言語教育などの内容もありますが、それ以上に、経済学や社会学との結び付きの強いプロジェクトが精力的に展開されています。ここでは五つの点に分けて御紹介します。

まず、1点目、公衆衛生系でありますけれども、例えばヘルスコミュニケーションの総合的研究では、不安症、精神症、肥満症等がメディアの中でどう表象されており、そこにどのようなスティグマやステレオタイプがあるのかということ明らかにしていきます。患者フィードバック研究では、NHS—イギリスにおきまところの国民保険のようなものでありますが—これに対する患者さんからのフィードバックテキストを効果的に分析し、NHSのサービス改善を支援することが目指されます。ワクチンやマスクに関する言説研究では、先般のCOVID-19のパンデミックにおいても、ワクチンやマスクに関する様々な言説が流布されたところですが、こうした言説を分析する手法の開発が目指されます。最後に、終末期医療の関連言説研究では、終末期医療や延命治療というものが言語表象される場合にどのようなステレオタイプやイメージが結び付けられているのかを明らかにしていきます。こうした問題については、無知や偏見、差別意識等によって適正な対応が取れていないという側面があり、大量のテキスト分析を通して問題の真相を特定し、解決につなげる方向を探ることが期待されています。

次に、2点目、経済・ビジネス系です。ここには、金融テキストの言語処理研究や、企業コミュニケーションの言語研究などが含まれます。後者に関しては、例えば、企業が株主に出します文書などを言語学的に分析し、分析結果を企業にフィードバックすることで、コーポレートガバナンスの改善に資することが狙いとされています。このほか、企業が発行する財務情報の分析や、欧州における海上安全保障関係資料の言語分析などもなされています。

3点目の社会心理・社会思想系としては、例えば、DOOM分析やヘイトスピーチ分析などがなされています。DOOMというのは、Discourse of Online Misogyny のことで、オンライン上に散見される女性嫌悪の発言や女性に対する脅迫的な発話などが分析対象となります。DOOMやヘイトスピーチなど、社会的に問題のあるテキストをあえて集めて解析することで、そうし

た言説がなぜ起こるのか、今後オンライン上でそうしたことが起こらないようにするにはどのような仕組みが必要なのか、といったことが研究されているわけです。ここでは、法学や犯罪学と連動した分析も試みられています。さらに、時系列データを使い、「貧困」や「リスク」に対する意識の変遷を探る研究、国内の地域と世論がどのようにリンクしているかを解明する研究、動物に関するテキスト分析を通じて動物に対するフレーミングを解明する研究なども行われています。社会意識や社会思想というのは、つかみどころがないように思えるわけですが、コーパスアプローチの活用について、これらについても一定のエビデンスに基づく分析が可能になってまいります。

4点目は、英国の国内政治や国際問題などに関する研究です。例えばイスラムに対する表象や人々の意識がどのように変化してきたのか、今後、イギリスに置いてイスラム報道をどう行っていくべきなのか、建設的なジャーナリズムはどうあるべきなのかといった議論がなされています。また、ブラジルとの共同研究として、ブラジル都市部の暴力問題に関して、公式統計、メディアの表現、市民の見解など、テキストデータと統計データを組み合わせるアプローチで新しい解決策を探るといった研究もなされています。同様の枠組みで、気候変動の言説分析や、移民へのインタビューコーパスに基づく移民の宗教意識研究、各種の批判的談話分析研究なども多角的に進められております。

最後に、5点目は、文学系・言語系・教育系の研究です。もちろん、言語資源を使うということで、これらについても新機軸が強調されています。例えば、GIS—地理情報データと文学作品をリンクした空間人文科学研究、大規模な言語試験コーパスを用いた外国語学習者の研究、心理学や医学の知見を踏まえ、事象関連電位等のデータを組み込んだ言語処理の研究、障害のある児童の言語特徴研究など、また、BNC 2014の構築・分析研究もここに含められています。

以上、駆け足でCASSの研究を御紹介してきました。

ここで一つ指摘しておきたいことがあります。言語資源研究と言うと、大規模均衡コーパスの開発がすぐ頭に思い浮かぶわけですが、英国のBNCや日本のBCCWJといった大規模汎用コーパスだけでは、実のところ、具体的な社会問題に応用するのは意外に難しいということです。やはり社会の特定の問題に即効性を有するのは、その問題に特化して構築されたコーパスとなります。この意味で、コーパス

言語学者も、これまでのように、大規模均衡コーパスを開発し、それを分析するというだけでは十分ではないかもしれません。むしろ、積極的に他の分野に出掛けていって、そこでのニーズを聞き取り、各分野で必要とされる目的特化型の小規模コーパスの構築を手助けし、さらには、その分析方法のガイダンスを行う、といった新しい仕事が今後ますます重要になってくるのではないかと思います。CASSがやっているのも、正にこうしたことです。そして、このようにしていきますと、学問分野を超えて、様々な社会的課題に特化した小さなコーパスが無数に作られていくこととなります。ここで重要なのは、そうした言語資源をばらばらの孤立的存在とせず、緩やかなネットワークでつなぎ、そのネットワークの中核として、つまりは、ネットワーク内にある各種の小規模コーパスにとっての絶対的な参照基準として、BNCやBCCWJのような大規模均衡コーパスの意義を再定義するということです。なお、様々な人の巻き込み方については、BCCWJの構築時に行われた「コーパス日本語学の創生」というプロジェクトが参考になるでしょう。このプロジェクトでは、言語学の枠内に限られたものではありませんでしたが、辞書学・音声学・文法学・自然言語処理など、それまでつながりの薄かった様々な分野の研究者が出会い、一緒になって新しい研究を開発する基盤が提供されました。

5. 様々な実践例

さて、ここで、CASSの話をいったん終えまして、次の話題、つまり言語資源の社会応用を目指した様々な実践例の紹介に入りたいと思います。

まず、1点目として、医療系では、ASD（自閉スペクトラム症）の研究が多く行われています。軽度と重度のASD児の発話を集めて違いを調べたり、ASD者と非ASD者の発話コーパスを分析したりする研究です。ある研究では、ASD者は「……だわ（よ）」とか、「……だね」といった文末終助詞の使い方に特徴的な傾向があるということです。また、抑鬱症患者の発話の言語学的・音声学的特徴を明らかにしようとした研究などもあります。こうした研究が進めば、発話データを見ることで、そうした症状の有無や程度を早期に診断することが可能になるかもしれません。

2点目は、教育系研究です。一例として、アメリカでは、いわゆる母語話者と、そうでない方が共に看護師として働いておられるわけですが、様々な看護師の発話

を集めてテキスト分析をすることで、看護師向けの新しい発話法の指導プログラムを開発するという研究がなされています。テキストデータを根拠とすることで、どういう言葉遣いをすれば患者さんとのコミュニケーションが深まるかが、ある程度、定式化でき、それを指導プログラムに反映していくわけです。このほか、英語教育や日本語教育の例として、日本の大学英語教育学会（JACET）は、各種のコーパスで語彙頻度を調べ、日本の大学生が学ぶべき英語語彙の枠組みを確定するという研究を長年にわたって進めています。また、発表者による試行的な研究ですが、I-JASという学習者コーパスに含まれるストーリーライティング（同じ連続イラストを見て内容を作文する）のデータを用い、学習者の習熟度レベルと、使用した語彙の頻度を統計的に分析することで、学習者の日本語レベルを簡易測定する手法の開発などが試みられています。

6. ICNACLEプロジェクトから考える《継続・普及・実装》

以上、言語資源の社会応用に関する研究例を御紹介しました。最後に、冒頭に申し上げた言語資源活用の三つのポイント、すなわち、データの継続的収集、普及の工夫、社会実装化が、それぞれ具体的にどのように行われるのか、発表者自身が長年関わっているICNACLE（International Corpus Network of Asian Learners of English）プロジェクトを例として御紹介します。

ICNACLEとは、アジア圏の10か国・地域の大学生から、共通条件で、英語による作文、独話、対話などのデータを体系的に集めた国際学習者コーパスです。現時点では、アジア圏学習者の公開コーパスとしては、世界最大級のものになっています。

まず、1点目、継続的収集ということですが、このプロジェクトが始まったのは2007年です。それから20年近くになりますが、今もプロジェクトは継続しています。この間、まず、作文、次いで独話、対話、評価とデータを段階的に拡張してきましたが、それだけでなく、既に集めたデータについても、反復的な検証によって修正や増補を行っています。例えば、作文データの現在のバージョンは2.6となっています。当初のリリース以降、既に10回以上、細かな修正や増補が加えられています。いかに大きなコーパスを作ったとしても、継続的なアップデートやメンテナンスがなければ、研究資源としての価値はやはり下がっていくのではないかと思います。

2点目、普及のための工夫ということですが、集めたデータについては、ダウンロードして各自で高度な分析をしていただけるようになっているだけでなく、I C N A L E O n l i n e という簡便なオンラインでの検索システムを開発して提供しています。これはコーパス言語学の知識がない一般の方でも使えることを重視しておりまして、シンプルな操作で、各国の学習者がある語をどのように使ったか、ある語の共起語にはどのようなものが多いか、日本人学習者vs英語母語話者、日本人学習者vs中国人学習者など、2群を比較した場合に、一方が特徴的に多用する語句にはどのようなものがあるか、学習者のオリジナルの作文と専門家が校閲した版ではどこがどういうふう変わっているかなどを知ることができます。また、様々な出力の工夫もあり、特定の語の学習者群別頻度をグラフで確認したり、学習者の発話の音声を聞いたり、学習者がインタビューに答えている様子を動画で確認したりすることもできます。こうした簡易な検索システムがあれば、コーパス学者だけでなく、英語を勉強しているアジア圏の高校生や大学生の方、学校の先生方、各種の英語教材の開発を考えている方々にも広く集めた資料を活用していただくことが可能になります。

3点目、社会実装ということですが、まずは、研究面で、広く使っていただけるようになりました。I C N A L Eを使った卒論や修論・博論、また、トップジャーナルに掲載される論文などが世界中で数多く出ています。例えば、タイ、インドネシア、パキスタンなど、これまで、英語学習者データが体系的に集められていなかったような地域では、特に盛んに使っているようです。続いて社会実装ということですが、これはまだまだ十分ではないのですが、海外の著名なグラマーチェッカー（英文の文法ミスを自動で指摘するアプリケーション）のトレーニングデータに活用されたり、内外の自然言語処理系学会における「シェアドタスク」のベースデータに使われたりしています。さらには、I C N A L Eの分析から分かった日本人学習者の英単語の使用の癖や問題点を踏まえて、和英辞書用の語法コラムなども書かれています。

言語資源には、文化遺産の側面もあるということが本小委員会でも議論されておりますが、学習者コーパスについても同様のことが言えるかもしれません。先般、タイの学会に参加したのですが、その席で、現地の研究者から、「これまで、世界の英語学習者研究においてタイ人学習者は注目されてこなかったが、タイ人のデータ

を含む ICNALE が広まったことで、タイ人英語学習者への関心が世界的に高まり、タイでも同様の研究が増えてきた」ということを言われました。コーパスにデータが入ることで、その存在が記録され、可視化され、研究対象になるということなのだと考えます。

7. 小括：報告者による提言

以下、本日は、多くのお話をさせていただきましたが、最後に小括として個人的な提言として5点をお示しし、報告を終わります。

1点目は、言語資源は作るだけではなく、完成後もアップデートを行いつつ、かつ広く使っていただける工夫を凝らすことで社会実装への道筋を考えていくべきだということです。2点目は、応用言語学の理念を踏まえ、言語資源研究においても「社会的課題解決のための言語研究」という方向性を積極的に探っていくべきだということです。3点目は、CASSの実践にも学びつつ、コーパスそのものの開発・分析にとどまらず、コーパスの分析技術やテキストに対する見方を広くコーパスアプローチとして捉え、それを接点として、戦略的に他分野を巻き込んでいく方向性も検討すべきだということです。4点目は、様々な社会問題の解決に貢献できるよう、他分野のニーズに基づき、他分野のデータを使って各種の小規模コーパスの作成と解析を支援し、あわせて、それらの絶対参照基準として基盤的な日本語コーパスの維持管理の必要性を再定義していくべきだということです。最後に、5点目は、長期にわたる安定した公的資金の投入によって、大型コーパスを核とする「社会実装志向型の異分野共創言語資源活用研究」という新しい研究領域を展開していくべきだということです。個人的な意見ではありますが、今後の議論のきっかけになれば幸いです。

以上で報告を終わります。本日は、御清聴、ありがとうございました。

○相澤主査

石川副主査、どうもありがとうございました。

それでは、まず、ただ今の御説明についての質問の時間としたいと思います。御質問がありましたらよろしく願いいたします。

○齊藤（美）委員

一つよろしいでしょうか。

○相澤主査

はい。お願いします。

○齊藤（美）委員

石川副主査、ありがとうございました。「コーパスアプローチ」という言葉、最初、どのような意味なのかなと思ったのですが、お話を聞いてよく分かりましたし、その重要性というのもよく分かりました。ありがとうございます。

いろいろ情報量、大変多かったので、もう少しだけ御説明、追加いただければなと思うことがいろいろあるのですが、一つ伺いたいのが、石川副主査の御研究を基に、「ささやかな」と書いていらっしゃるんですが、社会実装のところで、二つ目の海外の工学系学会の多数の Shared task の training データに利用されますが、これは言語教育を前提とした母語判定とか、習熟度判定、それから、フィードバックコメントということになりますでしょうか。

○石川副主査

ありがとうございます。自然言語処理などの分野では、様々な共通課題を設定し、各自が課題に沿ったシステムを開発し、その性能などを競い合うといったことをなさっておいでのようなようです。おっしゃるように、学習者テキストに対する母語判定や、自動フィードバック付与といった課題の際に、ベースデータとして ICNALE を使っているようです。無償で公開されているので、使いやすい、ということもあるのかなと考えています。

○齊藤（美）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、いかがでしょうか。では、私から1点、大変貴重なお話を頂いて、ありがとうございました。本当になるほどなと思うところが多かったのですけれども、社会実装の実現という観点から見たときに、今、大規模言語モデルが何に役立ちますかという議論の中で出てくるものと共通点が多いと感じました。

○石川副主査

はい、共通点は多いと私も思います。

○相澤主査

石川副主査のお話の中で出てきた医療等も大変話題になりましたし、あるいは、たまたま消費者保護の話聞く機会があったのですが、ダークパターンと呼ばれるよこしまな意図で悪用される言語パターンの問題ですとか、本当に切実に社会実装で必要とされている場面が多々あると感じます。それで、そのときに分析の技術が重要で、それを汎用的なものとして、学問として、あるいは技術として体系化していくというのは、やはり訴えていくために重要だと感じていまして、その辺りで何かこれがポイントであるとか、あるいはもしかしたら、今、大規模言語モデルで苦労している点との共通点のようなものがありましたら、是非御意見を頂きたいです。

○石川副主査

ありがとうございます。コーパス分析技術、いわゆるコーパスアプローチというのが何を意味するかというのは、なかなか難しいところなのですが、私見では、大規模データからの統計的なパターン抽出と、質的なテキスト解釈の二つが両軸となるのではと思います。前者については、工学系でも広く実践されており、多変量解析などの手法を組み合わせることで、言語的なパターンを半自動で抽出するといったことができます。後者については、批判的談話分析(Critical discourse analysis : CDA) などとの連携も考えられると思います。CDAは談話分析を拡張したもので、テキスト(談話)の詳細な分析を通して、書き手や話し手のバイアス、特に、政治的バイアスをあぶり出すといったことを課題としています。例えば、「太郎が窓を割った」と言わず「窓が割られた」と言うことで意図的に責任を曖昧にし

ている、といったような分析がなされます。これは非常に面白く、価値の高い研究分野なのですが、一つの制約としては分析者の主観に影響されやすいという問題があります。ここにコーパスアプローチを組み込んだものが、最近注目されている、コーパス準拠型批判的談話分析（corpus-based CDA : CCDA）というもので、コーパスからの言語的エビデンスを基盤とすることで、主観と客観のバランスを取って研究を進めていくことができます。こうした量的手法と質的手法を整理していくことが、コーパスの分析手法の体系化の上で必要ではないかと考えます。

○相澤主査

ありがとうございます。我々もアノテーターの方々へのガイドラインを作成するとか、アノテーターの方々のトレーニングをすることはよく行うのですが、その際に言語学の方から見たらプリミティブな、初歩的なことをやっているのかもしれない、そういったところで何かコーパスアプローチがどんどん重要性を増していくということがあるのかなと思いました。

○石川副主査

ありがとうございます。

○相澤主査

ほかは、いかがでしょうか。

○石川副主査

先ほどの相澤主査の御指摘に関して、コーパスアプローチの体系化と言語工学の連携は非常に重要だと思います。アノテーション、例えば、品詞や意味のタグを付けていくといったことに関しては、言語学者は、どちらかと言うと手作業ベースでやる場合が多かったわけですが、大規模言語モデルを使うと、もっと大量に、もっと高速に、しかも、揺れのない形で一気に作業ができるようになると聞いております。ですから、そうした面は機械やAIに任せつつ、残った部分、つまり人間の目で丁寧にテキストデータを見ていくという方面で、コーパスの分析技術を更に磨いていくことが大事なのではと思います。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。

よろしいでしょうか。（→ 挙手なし。）

では、御質問の時間としては以上といたしまして、続きまして意見交換の時間に入りたいと思います。本日、お話のありました海外のコーパスの状況、あるいはコーパスを活用した社会課題の実現などを踏まえまして、皆様の意見を頂くという時間となります。本日の御発表、参考資料1に検討の観点というものもございます、1ページ目の最後に観点が書いてございますけれども、本日の御発表は二つ目と三つ目、コーパスの整備状況と活用の在り方、今後における望ましい言語資源の在り方ということに関係してくるかと考えています。今後の小委員会からの報告にも結び付くものでございますので、そういったことも踏まえながら皆様の御専門とも関連付けながら御意見を頂ければと思います。恒例に従いまして、皆様に御発言いただくということで、よろしく願いいたします。杉山先生もよろしく願いいたします。

○齊藤（美）委員

まだ余り頭の中が整理できていないのですが、とりあえずトップバッターでお話を申し上げます。

○相澤主査

ありがとうございます。お願いいたします。

○齊藤（美）委員

今日の杉山先生と石川副主査のお話を伺っていて、また、最後の質疑応答も伺っていて、異分野の連携がとても大事なのだなということを改めて思いました。杉山先生にも、少し質問させていただきましたが、研究者だけでなく一般企業の人にも役立つようにというような連携のお話もありましたし、石川副主査の方でももちろんいろいろな、研究分野だけでもいろいろ出てきましたし、これが私、リベラルアーツの学部で勤務しておりますので、いろいろな分野の専門家が教員としているの

ですけれども、なかなか、それぞれの学問分野をただ学べますというだけではなくて、やっぱり結び付けながら学んでいくことに意義があるわけなので、どうにかいようにいろいろな授業をやろうとしているのですけれども、なかなか簡単なことではなくて、大学教員という中だけでも、そんなに簡単ではない中で、この大学からとは別の、研究者ではない世界の人たちとも一緒にとなると、本当に大変だろうなと思います。私は異文化コミュニケーション学も研究していたりしますので、何かそのコミュニケーションの取り方というのを併せて考えていく必要があるのかなと思いました。今、こうしたらいいというのが申し上げられるわけではないのですけれども、そういったところ、もし皆さん、何か御経験とかありましたら、御教示いただけたらうれしいです。

以上です。

○相澤主査

ありがとうございます。石川副主査、いかがでございましょうか。

○石川副主査

例えばCASSの場合は、そもそも自分たちのお金は、経済とか社会をやるところからもらっている。そのような人たちと連携するというためにお金が下りてきて、それでプロジェクトが走っているという、外から、ある意味、枠を課せられているというところがありますよね。ですから、最初はやはりそのようなものが必要なのかもしれません。

放っておいて、皆さん、言語学者と仕事をしませんかといって、向こうからもなかなか来てくださらないですし、こちらからもなかなか行くチャンネルもありませんので、最初はそうしたある種の枠は課して、少し幾つかはプッシュしてやっていると。そのような中で広がっていく。場合によってはビジネスの方とか、経済の方とか、医療の方が、言語学者と知り合って、いろいろコーパスというものを初めて一緒に触ってみて、あ、これだったら、こんなカルテの情報も分析できるのではないとか、これだったら患者さんとの会話も、もっと使って診療に活用できるのではないかというふうに広がっていくのではないかという意味で、最初の一步みたいなものが重要なのではないかなと思います。

イギリスも私が知る範囲ではありますけれども、このようなプロジェクトが立ち上がる前に盛んにそのようなことが行われていたという印象は、余り強くはありません。ですから、何かこのような新分野開発的なプロジェクト資金なり、プロジェクトみたいなものがキックアップとしては重要なのかもしれないと感じました。

○齊藤（美）委員

なるほど。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。

では、杉山先生。

○杉山発表者

私も知っているプロジェクトがそんなに多くないんですけれども、先ほど少し御紹介したTCFというフランス語の能力を自動判定するプロジェクトですと、やっぱりTCFというこのフランス語のテストを作っている機関が労力を減らしたいから、どうしたらいいのかということで資金とか、そのデータの提供とか、そのようなのを言語学者というか、自動言語処理の専門家に渡してという感じで、何か問題解決すべき何かが、問題がまずあって、そこからやっぱり研究者に下りてくるといような、そのようなイメージです。

あと、何かフランス語系ですと、この前、これもコーパスに関係すると思うのですけれども、英語だったらInternational Conference on Statistical Analysis of Textual Dataというような学会があるのですけれども、何かそのようなところで発表を聞いてみると、医療系とかのデータを使って発表されている研究とかも結構あるんですけれども、元々多分、言語学者ではなくて医療畑の人なのかなと思われたりとかするので、何か結構、広がってきているようですが、何か連携はまだ個々でやっている一問題があって、資金が提供されてやっているか、個々で何か自発的にやっているというような状態なのではないかなという個人的な印象を持っています。

○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、こちらも言語学の皆さんへの御質問という形になるのですけれども、何かお話をお伺いしていると、キャリアパスの問題もある、関係してくるのかなという気もいたしまして、今、言語学は比較的アカデミア志向のような気がするのですけれども、社会実装という出口を見据えていくと、例えばキャリアパスとして企業の中に積極的に入って行って、その言語的な分析の設計を本当に実用的なものとして役割を担ってプロダクト化するとか、そういったことも、実はあってもいいんじゃないかなと感じる次第です。その辺りの分野として、あるいは学生さんを送り出す先生の立場としてという辺り、いかがなものでしょうか。

○石川副主査

確かにそういう面は必要でしょうね。

○前川委員

よろしいですか。

○相澤主査

はい。お願いします。

○前川委員

我々、国語研は一昨年、総研大——総合研究大学院大学に参加させていただいて、博士後期課程のコースを作ったんですね。日本語言語科学コースというんですけれども、その当初からのポリシーの一つとして、キャリアの一つとして、やはりそのような産業界への貢献というものを挙げているんです。それで、まだ卒業生がいまないので一学位を取ったのがまだ2年目でいままので、そのコースがこれからどうなるかというのは分からないのですけれども、私の主催してきた過去のプロジェクトの中で雇用したポスドクですとか、そのような人で産業界に入っている人はいます。最近、ちょっとした大規模言語モデル関係のプロンプトの書き方ですけれ

ども、そういったことで本を書いて、ちょっとしたベストセラーになっているGoogleの研究者とかいますけれども、元々は言語学の出身という人ですね。

細かく見ていくと、昔から音声認識とか、合成とかの関係で、言語学出身の方が例えばN T Tの研究所とか、K D D Iとか、そのようなところの研究所に入っている例はあるんですよね、数十年前から。それはそれなりの貢献はなさっていると思うのですけれども、ずっと見てみると、最終的にはやっぱり、ずっとその研究所にいるのではなくて、大学の先生に転出するとかいうようなケースが多いように思います。ただ、情報処理の方法が変わってきましたから、これからは必ずしもそうではなくて、そのような研究所の中で独自のキャリアパスが作れるような時代になっていくかなと思います。今ちょうどその端境期みたいな、いい方向に変わる入り口ぐらいにいるのではないかなと私は感じております。

○相澤主査

ありがとうございます。

是非、社会人学生としても積極的に言語学の分野に来ていただけるといいなと感じながら伺っていました。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

ただ今の相澤主査の御指摘、これは非常に重要なところで、キャリアパスの問題を考えますと、大学や大学院における言語学の教育課程の組み方や、どんな内容をどのように教えていくかというところにまで話は及んでくるのだらうと思います。言語学の教育においても、社会で必要とされている言語の知識であるとか、言語に関わる社会の問題といったものも丁寧に取り上げ、社会課題解決のための言語学、という側面についても適切に指導していく必要があるな、とお話を伺って強く感じた次第です。ありがとうございます。

○相澤主査

ありがとうございます。

○石川副主査

では、私からも一つ質問を。

○相澤主査

はい。お願いいたします。

○石川副主査

発表でも述べさせていただいたのですが、前川委員が「コーパス日本語学の創生」というプロジェクトを立ち上げられた時には、言語系のいろいろな分野の研究者に広く声を掛けられ、実際に、いろいろな人が入ってきて、結果として、「コーパス日本語学」という新しい分野が勃興して盛り上がってきました。私自身もその末端でいろいろ関わらせていただいたのですが、あのような仕掛けというのは、どのような経緯で実現したのでしょうか。戦略的に裾野を広げていくという方針が当初からおありだったのでしょうか。

○前川委員

私への質問ということですよ。

○石川副主査

はい。

○前川委員

あのプロジェクト自体は、その前に99年からCSJ関係のプロジェクトがあったんですね。音声認識関係の、東工大の亡くなられた古井貞熙先生がリーダーを務められ、私がお助けしたプロジェクトがあって、それが比較的——比較的と言うか、かなりいい成果が5年間で出たわけです。それでCSJを5年間で公開して、情報処理中心ですけども、非常にたくさん使われ——今でも使われていますけれども、評判がよかった。それで、それが終わって、じゃあ、次はどうしますかということが国語研の中で問題になって、それで、1年間、それを検討する短期のプロジェクトを立てたんですね、所内で。有志のプロジェクトみたいなものですけども、そ

の中でこれから何をやろう、次は何をやろうかということになったときに、やっぱりその均衡コーパスというのは、当時、日本にはありませんでしたから、それを作らなければいけないんじゃないのという話にだんだんまとまってきて、それで短期間の科研を取って1年、当時、そのような大きなプロジェクトを申請するための予備段階で使える科研という非常にいいのがあったんですよね。

それを申請して1年間で100万語のコーパスを作ってみました。そうするとコストが全部分かりますから、それで1億語作るにはこれぐらいコストが掛かるだろうということまで一応押さえて、それで翌年、特定領域研究というのに応募して、幸いすぐ採択していただけたという話なんですね。その今申し上げたぐらいまでのところでは、最初は、国語研の所内での話として始まったんですね。ただ、当然、実際に始めるときは一種のナショナルプロジェクト的なものとしてやらなきゃいけないというのは、やるつもりだということは最初から考えていましたので、そこでは広く声を掛ける。それから、実はその時点で掛けるだけではなくて、あのようなプロジェクトは、もうその前のCSJの時からそうですけれども、情報処理系の力を借りずには絶対できないんですよね。形態解析の技術もそうですし、音声認識はもちろんそうですけれども、いわゆるテキスト的な部分でも絶対必要なんです。

それで、当時から、私、元々音声の仕事をやっていたこともあって、そのような関係の知り合いの先生がたくさんおられたし、それから、所内に奈良先端大の出身者とかもいましたので、そのような関係でNLPの方ともつながっていて、それであのプロジェクトはうまく動き出したと。それで、そのようなそれぞれの領域でしっかりした仕事をなさっている方をハブにして、また声を掛けていただいて、あれは幾つだったかな、六つか七つぐらいの基幹的な研究グループから成っているんですけれども、それぞれの領域に例えばNLPだったら、あの松本裕治さん、奈良先端の松本裕治さんにリーダーになっていただいて、その領域で作っていただくとか、それから、そのようにそれぞれの領域で、この人ならばと思うような人をお願いして、サブリーダーとしてそれぞれの領域を組織していただいた。幸い皆さん非常に優秀で、有能でうまくいったという感じですね。

今日の石川副主査のお話を聞いていて、私、非常に勉強になったんですけれども、実はそのような社会科学なんかへの応用というのは非常に重要な課題だと、今思っているんです。それで、私自身は、実はコーパス言語学というのを専門として名乗

ったことは一度もなく、言語資源とか言語資源学とか名乗っているんですよね、音声学のほかには。それはなぜかという、私が想定する言語資源の価値というか、応用的価値としてはやっぱり、広く諸学への応用、諸学の下支えをするというようなこと、要するにインフラ整備ですね。例の特定領域研究は、日本語研究のインフラ整備ということで取ったんだけど、実はもっと大きく、要するに言語を使ったあらゆる研究、社会科学等も含めたもののインフラになるはずのものなんです。

それを何とかして次の段階で実現したいなとはずっと考えているんですけども、そこはなかなか、先ほど言ったような人の付き合いが、実は私、社会科学系とか、法律とか、そっちの方とはあんまり付き合いがないもので、また、うちの研究所でも、その仕事の上での付き合いがないもので、どうしてもそっちの方に展開するところで、今のところ止まってしまっているという、それが現状です。非常にそのような人のつながり、元々ある人のつながりみたいなものは非常に重要だと思います、このようなプロジェクトを作るときには。

○石川副主査

ありがとうございます。私見ですが、BCCWJプロジェクトの意義は、価値ある言語資源ができたというだけでなく、それを分析できる研究者が同時に育ったという点にあると思います。言語資源とそれを分析する人の両方が同時に世に送り出されたからこそ、今の日本語コーパス研究の活況があると感じています。この意味で、今後、BCCWJ 2ができます時に、当然、良いデータができてくると思うのですが、それと同時に、また新しいタイプの分析者が育成されていくといった仕掛けがあると大変良いのではないかと期待をしながらお話を伺っていたところです。

○前川委員

それはBCCWJ 2とは必ずしも…別個に、一昨年の、令和6年度の概算要求で新しいプロジェクトを通しまして、それで、今年の8月にうちに、国語研に俗にE3P-Linguisticsセンターというものを作ったんですね。日本語で言うと、次世代言語科学研究センターというのですけれども、それは正にそれで、要するに大規模な言語資源が存在するということを前提にして、それをきちんと料理するための、言い切ってしまうと数学的な方法論の開発、ベイズ統計とか、言語

モデリングとか、あるいは種々の実験的な手法、あるいは脳科学的な手法的なものも含めて、そういったものを開発するセンターというのを実際この8月に立ち上げて、それで、まだ重要なメンバーが赴任していないのですけれども、年が明けると赴任してきますので、その赴任してくる人たちは、どっちかというとなlpの出身だったり、あるいは統計数理研究所の教授の人にクロアボで来ていただいたりして、主にそのような言語学とはちょっと外れるんですけれども、そのような統計的なモデリングみたいな研究をそこでやって、そしてそれを言語の研究者に対して普及させようと。毎年ハンズオンセミナーみたいなのを開発—毎年2回かな、開催するというのをその計画の中では立てております。ですから、少しずつそういったことでも貢献できるのではないかなと思っています。

○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した。

では、続きまして植木委員、古典籍なども含めまして、いかがでございましょうか。

○植木委員

ありがとうございます。感想レベルのことになってしまうのですが、今日の石川副主査のお話の、特に社会心理、社会思想系への応用のところに大変私は感銘を受けました。前回の川原先生のお話、そして前々回の石川副主査のお話で、特に医療の関係で、医療従事者の方々がどのような話し方をされていて、それが治療者にとってどのような効果があるかというような、治療に効果的な話し方や言葉の選択を統計的に、客観的に多く示すことができるような可能性がある、そのようなプラス面の話をお聞きしたのですが、今回の、マイナスの事柄、負の意識や問題行動の背後にあるメカニズムを解明して抑止につなげるお話、このような形で、コーパスが役に立つのだというのは、今、本当に社会に対して訴える力があると思いました。

SNS上でも、現在はこれだけたくさんの言葉が飛び交っています。かつて、古典の時代は、その言葉を扱って、言葉を発信する人というのは非常に限られていました。以前から、どのような言葉を取るかという言葉の範囲もこの小委員会でも話題になっておりましたけれども、今は非常にたくさんの人がたくさんの言葉を発している状況の中で、選挙のことを取りましても、負の面が非常に^{ひぼう}出た誹謗中傷や、アンコンシャスバイアス（unconscious bias 無意識の偏見）が掛かった本当にひどい言葉も飛び交っていたと思うのです。それが問題視され、どうしたらよいかという状態になっていて、その解決は一朝一夕にはいかないと思うのですけれども、このようなアプローチの仕方、言葉からのアプローチによって、一つ解決につながる何かがあるかもしれないという、そのようなことは皆が求めていることとつながるのではないかと思います。

先ほど相澤主査が消費者保護のお話をしてくださって、勧誘の言葉についておっしゃってくださったのですが、詐欺のための犯罪の言葉についてもマニュアルが作られているぐらいなようなので、そのような負の言葉の蓄積と解析も、犯罪の抑止や、被害を減らすということでは役立つ側面があるのだというのは、コーパスの必要性を訴えていく上では力があると思いました。

もう1点、古典とのつながりで言うと、石川副主査が「文化遺産」という言葉を使ってくださって、その言語資源の文化遺産としての側面というのが興味深いところですね。もちろん今この小委員会で考えているのは現代語のことですが、それを長期間保存していくということで、今後の人にとっては古典の言葉になるという—そのようなこともあると思うのです。その中で方言—今日、杉山先生がフランスの地方言語を促進するというようなお話をしてくださり—その「地方言語」というのは方言という理解でよろしいのか、ちょっと違うのかもしれないんですが—そのような方言も今どんどん失われつつあります。例えば病気になってしまったり、災害で地域から離れなければいけなくなってしまうたりした人が、故郷の方言を聞くことによって、方言でしゃべって、方言で語り掛けられることによって元氣を取り戻すという、そのようなこともあるとお聞きするので、失われそうになっている言葉を保っていくことも非常に重要な役割になるのかなという気がいたしました。

○相澤主査

ありがとうございます。

杉山先生、方言の話が出ましたけれども、いかがでしょうか。何かコメントがありましたら、お願いします。

○杉山発表者

そこも専門家ではないので、あんまりよく言えないですけども、方言というよりは、やっぱりその地方の一地域のフランス語という感じで、フランス語の中では語られていると思います。例えばフランス語でも、カナダのフランス語って全然違う一本土のフランス語と全然違うように、カナダでも結構、移民をどんどん積極的に受け入れてきましたから、カナダでは結構、フランス語教育、移民に対するフランス語教育というのはかなりしっかり行われていると思います。でも、やっぱり特殊な言い方とか、単語レベルだけではなくて、いろいろあるので、そのようなカナダのケベックで話されているフランス語をちゃんと伝えていこうみたいな動きはあって、そのようなところでやっぱり、その言語データとか、そのようなのは役に立っていると思います。本土ではこう言うけれども、カナダではこう言うという根拠を示すときにも、そのようなデータ、経験だけではなくてデータとかに基づいて、そのような教育、移民に対する教育などが行われているのではないかと考えています。

○相澤主査

ありがとうございます。

では、神永委員、よろしくお願いします。

○神永委員

私も感想みたいなことになってしまうかもしれないんですけども、今日、フランス語と英語のコーパスの実態につきまして、大変興味深いお話をお聞かせいただきまして、ありがとうございました。私は日本語の辞書編集者なので、限られた形でコーパスを使う側であるんですね。だから、そうではあるんですけども、このような海外のお話を伺いますと、その海外での実態を参考にしながら、じゃあ、日

本ではどのようにそのコーパスを活用するのか。私は作る側ではなくて活用する側ですから、どうやって活用したらいいのかということをいろいろと考えさせられました。

具体的にどうしたらいいかということは、まだこれからなんですけれども、日本語独自の意味や用法が当然あるわけで、それはもう英語やフランス語と違うわけですから、そういったものを理解しつつ、私たち、使っていきたいなと思っていましたし、それから、杉山先生のお話にありました方言の件なんですけれども、日本国語大辞典という私に関わってきました辞書なのですが、この中に方言項目がございまして、実は私、方言担当者でもあったんです。

今度、第3版に向けて、じゃあ、方言、どうしようかというようなことも話しておりますと、そうすると、小規模コーパス的なものとしての方言のコーパスというのも一つやっぱり何か考えていかなければいけない。恐らく国語研の方では、もう既にそのような話にもなっているのではないかなと思うのですけれども、日本国語大辞典の中での方言項目と、それから、国語研の中でのやられているものとうまく合わせながら、何か協業でできればいいかなんていうことを考えております。

それから、コーパスですけれども、私たちにとってみれば当たり前のようなものになってしまっているわけなのですが、一般の方に理解してもらうということになると、もう少し具体的なものを示さないと、なかなか難しいのかなというような気もしています。それは石川副主査のお話にもあったわけなんですけれども、そうすると、こんなことができるよということをもっと積極的に可能性を提案していく、発信していくということが大事なのではないかなという、その具体的なものであればあるほど、それがいいわけですし、例えば今回、お話の中でお伺いしました看護の世界のお話なんですけれども、これなんか本当にすぐ日本の中でも活用できるだろうと思いますし、それから、看護だけではなくて介護の方、これから当然、私たちもいろいろとお世話になるような世界なわけですから、そういったものを小規模コーパスを使いながら活用していけるといいのではないかなと思っています。

介護の場合ですと、介護する側とされる側とのコミュニケーションの問題も当然あるわけですし、それから、言葉遣いの問題も当然あるわけですね。それは介護を受ける人間だけではなくて、家族の問題も当然あると思うんです。介護を受けている人を、その施設ないしその介護のところにお願いしている家族の人たちも、や

っぱりその介護をしている方たちとの関係性というか、そのようなのも言葉で非常に重要なことになっているだろうと思いますから、それをお互いにうまく伝えられるような形のものができればいいのではないかなと思います。

それから、当然ですけれども、例えば自分の体の状況を言うときに、なかなかうまく伝えられないという問題もありますから、それもどうやってやったらうまく伝えられるのかということも考えられるのではないかなという気がします。それから、更に言いますと、今、外国人の方たちが介護の世界、非常にいっぱい入ってきていますけれども、そういった人たちの日本語、つまり、日本語を母語としない人たちのためにも介護の現場で対応、ふだん使われている、対応されている適切な言葉を示すことができれば、よりスムーズな活動、形になるのではないかなんていうようなことを考えます。それは本当に飽くまでも私の考えなのですが、そういったものに活用できるのではないかなというような気がしまして、今日のお話を大変興味深く伺いました。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、ほかに何か御発言ございましたら、よろしくお願いいたします。

○村瀬国語課長

事務局からもよろしいでしょうか。

○相澤主査

はい。

○村瀬国語課長

貴重なお話、ありがとうございました。産学連携の経験から申し上げますと、産業界では、アカデミアにおける研究シーズと言いますか、研究成果の活用可能性について、場合によっては、把握しがたい部分があるのではないかと考えられます。この点、例えば、コーパスを切り口としたアプローチ、応用言語学的アプローチ、すなわちコーパスの有用性につきまして、学会など研究者コミュニティのお立場

から、コーパスによる課題解決といったようなものを示していくという方法があり得るかとも考えられますが、このような発信の在り方につきましては、どのようにお考えでしょうか。

○相澤主査

ありがとうございます。

○石川副主査

私は英語と日本語のコーパス系の学会にともに関わっておりますが、学会といえども、基本的には独立した個人研究者の緩やかな集まりですので、学会で意思を統一して、コーパスの社会応用に関して何か外向けにメッセージを出すというようなことは、実際には難しいのではと感じます。先ほど、何か、粹やかせのようなものがないとなかなか人は動けないと申したのですが、研究者の背を押すものが必要だろうと思います。

イギリスの場合は、先ほどお話ししたCASSへの助成がそれに当たります。この意味で、日本においても、コーパスの社会実装やコーパス言語学と異分野の連携を明示的なミッションとする、大型の、しかしながら、間口の広い競争的資金などが用意されると、状況を変えるきっかけになるかもしれません。それを受け取ると、その方向で動く大義と義務が生じます。言語資源研究のネットワークを広げるには、少なくとも最初の段階では、こうした仕掛けが必要でしょう。もちろん、関連研究者のネットワーキングについては、国立国語研究所も既にいろいろなことをやってくださっているので、国語研には引き続きリーダー的な立場でそれを展開していただきつつ、並行して、助成制度の整備を考えつつ、関連の学会や研究者が、それぞれの立場でできることを考えていくという方向になりますでしょうか。

○相澤主査

ありがとうございます。

ほかは、いかがでしょうか。私も今日のお話をお伺いしまして、実は社会実装はあちこちにあったのだなという気付きが非常に強くありました。なので、それがそれぞれの現場で適用したという形になっていて、大きな動きになっていないという

のは確かにもったいない話で、どうすればいいのかというのは本当に重要な問題提起だと感じますし、例えば研究者の研究を評価していく——「評価」と言うとな変な言い方ですけども、どのような研究を求めていくかというときに、やっぱり作ったコーパスは、たとえコミュニケーションのためのある特定の現場に適用するためのコーパスであっても、再利用可能でなければならないし、再利用可能であれば企業の人も、これから使えるということで投資してくださるかもしれないしとか、いろいろな可能性が頭に浮かんで楽しみだなど、むしろ思いました。

ほか、いかがでしょうか。

○齊藤（美）委員

ちょっとだけよろしいでしょうか。

○相澤主査

はい。

○齊藤（美）委員

石川副主査の「かせ」という言葉、大変納得がいきまして、私も日本通訳翻訳学会の会員で、そこでプロジェクトという制度がありまして—それは5名以上の会員で、このようなことをやりたいというのを申請すると、審議されて認められれば、どうぞということで、少額と言うと何なのですが、少しお金を頂けるのですけれども、やはりそのようなのがあると何か一緒にやろうかという感じにはなりますし、その学会は通訳者、翻訳者として働いていることがメインの方、必ずしも大学に常勤で勤めている方ではない方もたくさんいらっしゃるんですけども、そのような方とも一緒にできるかなというのが、できたということがあります。

ただ、その制度があっても、忙しいということで、どうしても参加できませんとか、途中でお辞めになってしまうとかがあるので、何か、その忙しさは、みんな忙しいと思いますので、企業で働いている方も含めて。なので、その忙しさを乗り越えてもできるぐらいの「かせ」が本当にあるといいのだろうなと思いました。

○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す。

では、そろそろ意見交換の時間も終わりに近づいていますが、最後に何かもし御意見や感想等ございましたらお知らせください。よろしいでしょうか。

○石川副主査

最後に一つだけ。今日、杉山先生のお話を伺って、本当にたくさん学ぶことができました。

日本語言語資源の問題を考えようとした場合、もちろん日本のことはよく調べますし、参考として、英語圏の状況ぐらいまでは調べるんですね。しかし、本日のお話で示されたように、日本語・英語に限らず、世界にはいろいろな言語があって、多くの国においていろいろな言語資源が作られ、工夫がなされているわけです。そのような世界の状況を横断的に見ていきながら、日本語の言語資源の拡充と活用の方法を考えていくことが重要であると今日改めて思いました。以上です。

○相澤主査

ありがとうございます。

では、まとめの時間に入らせていただきます。前回は大規模言語モデルや生成AIの問題について御意見を頂きまして、今回はコーパスの整備のされ方、活用について意見交換をいただきました。このコーパスを社会課題解決に使うということについては、今回、お二人の方からの御発表に基づきまして、貴重な意見交換ができたことと思います。この小委員会としても発信をしていくことが必要であると感じました。一度、最初に御案内がありましたとおり、次回、言語資源小委員会の前に国語分科会というのを開催する予定となっております。国語分科会においては、これまでの言語資源小委員会の審議についての報告がございますので、前回の小委員会でお認めいただいた内容につきまして、国語分科会では主査、私が御報告をいたしまして、石川副主査にもサポートいただき、フィードバックがあればまたお伝えするという形にさせていただきます。

また、本小委員会につきましては、引き続き議論の基盤を補強するために事務局とも相談しまして、もう１日、他の外国語コーパスや活用事例などについて有識者からのお話をお伺いすることを考えていきたいと思います。

以上で、本日、もし何かほかにございましたら…、よろしいでしょうか。

では、本日の言語資源小委員会は、こちらで閉会となります。どうもありがとうございました。