平成 29 年度　東京外国語大学オープンアカデミー
東京外国語大学語学研究所 企画
『コーパスから見えることば・文化・社会』
2017 年 10 月 17 日（火）　第 2 回

「コーパスから学ぶフランス語」

東京外国語大学准教授
秋廣 尚恵

皆様はじめまして。私は本学東京外国語大学でフランス語を教えております秋廣尚恵と申します。本日は皆様に、コーパスから見えるフランス語ということでお話をさせていただきます。どうぞよろしくお願い致します。

録音機械の発明、その後の目覚ましい情報処理技術の向上により、大規模な言語のデータを蓄積して自動処理することが出来るようになりました。色々な言語において、億単位の語数を持つ大規模コーパスが次々に作成されています。コーパス、すなわち、実際に話された言葉の資料体の作成と応用は、現代の情報化の時代において、新しい段階を迎えているといえるでしょう。

また、コーパスは、人文科学、そして社会科学にとって経験的研究の観点から非常に重要なリソースを提供してくれています。経験的と言うことは、実際のデータに基づいて、実証的に出来うる限り客観的に分析していくということです。

この授業では、フランス語コーパスの現状を紹介したいと思います。そして、コーパスにはこんなものだと解説を聞いても、聴講者の皆さま自身の学習とあまり関係ないなあという感想で終わってしまっては、もったいないので、コーパスを使って、フランス語を勉強する方法にはどんな方法があるか、あるいはコーパスを使って、フランスの言葉だけではなくて、フランスの社会や、フランスの文化などを調べていくにはどうしたらよいか、といった身近な利用法を紹介したいと思います。そして、コーパスの提供するデータを単なる情報ではなくて、皆さま自身の知識へと変え、フランス語やフランス文化への理解を深め、さらには一般に人間科学におけるデータの役割とは何かといった問題の一端を考えて頂く機会になれば幸いです。

さて、ここで、簡潔に、現代的な意味でのコーパスとは何か、ということからお話ししましょう。

コーパスというのは先ほども申し上げましたように言語データの蓄積なのですが、ただ数だけ集めればいいというわけではありません。それなりの研究目的―例えば、バルザックの作品を集めて、バルザックの文体を調べたいといった目的など―に基づいてある程度の量のデータを集めたものです。このように、何等かの研究目的をもって、その目的をかなえる代表的なデータとなるものを集めたものをコーパスと言います。

コーパスは量、質ともに、何らかのサンプルとして十分に信頼にたるデータでなくてはなりません。例えばバルザックの勉強をするのに、バルザックの作品の一つだけを見ていては不十分です。ある程度の量とタイプの作品をデータ化して集積したものを参照する必要があります。

同様に、フランス語の「俗語」――つまり、砕けた言葉―を勉強したいと思ったら、様々な状況、話者によって発せられた言葉をある程度の量、集めなければ、サンプルコーパスとしては使えません。ある集合体の代表となるような質と量を備えていなければサンプルとはなり得ないからです。

さて、先程も言いましたように、現代的な意味でのコーパスというのは電子化された大規模なデータのことです。電子化されていますから、PC で閲覧したり聞いたり、サーチエンジンによって検索したり、計算ソフトなどを用いて統計処理を行ったり、音声処理ソフトを使って音声分析をしたり、統語解析、形態素解析のソフトを使ってタグ付けをすることができるデータのことです。

「タグ」というのは、コーパスのデータを分析して、様々な情報を書き加えたラベルのようなものです。言語の単位に様々な分析レベルにおけるタグを付与することが可能です。Tu aimes le chocolat?　という発話に、どのようなタグが付けられるか考えてみましょう。

音韻レベルでは、Tu aimes le chocolat? という発話の末尾のイントネーションが上昇している場合、その「上昇型イントネーション」というタグを末尾に付与できます。形態論レベルでは、Tu は「2 人称、単数、無強勢代名詞、主格」、aimes は「動詞 aimer の直説法、現在、2 人称、単数」、le は「定冠詞、男性、単数」、chocolat は「男性名詞、単数」といった品詞や形態論的情報が、タグとして各要素に付けられます。また、統語論レベルでは、aimes という動詞に依存する統語的要素として、Tu は「主語」、le chocolat は「目的語」、というように統語的機能やそれぞれの要素の持つ意味的役割（ Tu 「経験者」、le chocolat「対象」）を加えることも可能です。

このように、コーパスに現れた要素（単語や句や文など）に音声的特徴や、文法的カテゴリー、統語的機能、意味的特徴といった様々な言語学的な特徴をタグとして書き込んでいく作業をタグ付けといいます。

タグ付けにあたっては、多くの場合、Tree tagger などのタグを自動的につけるソフトを使用して、まず、機械による自動タグ付けを行った後、それを人間の目で確認して、修正を加えつつ最終的なタグを決定していきます。そのようにしてタグをつけることによって、ある特定の特徴を持つ要素を一斉に検索することが可能になるので、統計処理などを行う際には大変便利です。AI の開発も伴って、コーパスを用いた言語分析や自動言語処理は現在、大きな産業になっています。

このタグ付けの作業に、言語学的な分析そのものが大きく関わることは想像がつくと思います。言語単位をどのように区切っていくか、分析レベルをどのように設定するか、統語的カテゴリーをどうやって分類していくか、イントネーションをどのように特徴づけていくか、といった基本的な概念に於いて、言語学者の見解が一致していないことが多いのです。タグ付けをする言語学者の主観に流されないために、コーパスを扱う研究グループごとに複数の作業者の間で話し合いながら作業を進めていく必要があります。したがって、膨大な量のデータのタグ付けには気の遠くなるような大変な時間を要します。またそのための技術者や分析者といった人材の確保も難しいと言われています。

さて、コーパスには書かれたデータと話された音声データがあります。書き言葉のコーパスというのは、書かれたテキストを電子化したものになります。一方、話し言葉は、音声データに加え、それを書き起こし、電子化したデータが伴います。ところがこの文字起こしに関しても、先ほどのタグ付け同様、転写者により、転写の仕方に大きな違いが現れるため、複数で校正作業を進めなくてはなりません。例えば、フランスで最大の話し言葉を有するオルレアン大学では、転写を少なくとも 3 人の言語学者が見直すという体制を取っているそうです。そして、最終段階の 3 人目のチェックの段階でも一つの録音会話につき 1000 以上のミスを発見することもあるのだそうです。客観的で公正なるデータを得ることの難しさをこのエピソードは語っています。

つまり、我々は、言語分析を精密なものとするために大規模なデータを必要としていますが、実はそのデータを処理する段階で、既にある程度大まかな言語分析をしていかなくてはならないという、大きな矛盾に直面しているのです。

さて、ここで、フランスにおけるコーパス言語学の現状についてお話しましょ

う。フランスというのはドイツや他の国とはちょっとことなった特殊な事情にあります。これまで、フランスでは個々の研究者、あるいは研究チームが別々にコーパスを集めていました。先ほども申し上げた通り、コーパスのデータ処理それ自体が言語学的分析を要するため、あまり統一的な見解が存在しないわけです。よく言えば個性を重んじる風潮なのでしょうが、悪く言えばバラバラなわけです。せっかく集積したコーパスも、あちらこちらの研究室に分散されたままで、国レベルで大規模なデータとしての集積が出来ていなかったというのがフランスのこれまでの特殊な事情でした。

しかしながら、言語情報学の分野やコミュニケーション分野からの要請もあり、フランス語学者たちも、国レベルで、あるいは国際レベルで、データをお互い提供しあうシステムを作り出さなくてはならないということに気が付きました。もちろん、分析の方法論の部分でお互いの見方を共有することはできないにせよ、データをいろいろな人が使用できるように一か所にまとめよう、という機運が生まれました。それがこの ORTOLANG という機関です。ORTOLANG というのは「Outils et Recherche pour un Traitement Optimisé de LANGue（言語の最適化された処理のためのツールと研究)」という研究の名称の略語です。この ORTOLANG のウェブページに行ってみますと、様々なコーパスが提供されていることが分かります。そこに、登録をして、それぞれの各個人が持っているコーパスをアップロードしたり、あるいはダウンロードしたりできるようになっているのです。

データをアップロードする際には、ORTOLANG の研究者がサポートをしてくれます。データはアップロードされるために、ある程度の基準を満たしてないといけないので、最低限の基準を満たしているかどうかを調べてくれます。そしてその許可を得られるならば、アップロードが出来ます。

さらに、ORTOLANG にアップロードされたコーパスの一部については、こちらの CNRTL (Centre National de Resources Textuelles Lexicales　テキストと語彙のリソースの国立研究所)という研究機関のサイトにもアップロードされています。こちらの方は、オープンリソースですので、登録しなくても、ご自宅から個々のサイトに行くと見ることができます。ORTOLANG の方がもちろんデータとしては大きいのですが、こちらも十分大きいです。そしてこちらの方はアクセスがより簡単です。もしフランス語のコーパスをちょっと覗いてみたいという人は、むしろこちらの方が使いやすいかもしれません。

さて、今まで話したのはむしろ話し言葉の現状でした。話し言葉コーパスの収集とデータ処理については、まだまだやることが残されていますし、完成されたものとは言えないということが分かって頂けたでのではないでしょうか。

一方、フランス語の書き言葉のコーパスは非常に良くまとまっています。フランスには、滔々とした研究の蓄積がフランス文学の分野にはありますので、テキストの電子化はかなり進んでおり、とりわけ近現代の時代のものについては著作権さえ同意が認められるならば、コーパスとして公開するということが比較的容易にできています。

ここに挙げた FRANTEXT というサイトに行くと、そうした書き言葉のデータを検索することができます。残念ながら、無料で公開されているのはデモンストレーション版の僅か一部しかありません。全部を見るためには、毎年、お金がかかります。多分、個人会員だと一万円位だと思います。ちょっと高いですね。

東京外国語大学では、この大学内のネットワーク内でサイトにアクセスして頂くと FRANTEXT は自由に検索できます。それは学生、それから先生方のために、大学が一年ごとのお金を払っているからです。もしみなさん機会がありましたら、例えば外語大の図書館のパソコンなどから、FRANTEXT を試しに使用してみてください。

この FRANTEXT というのは、16 世紀から現代にいたるまでの文学作品を中心とするものなので、大体 8 割が文学作品、残り 2 割がその他の書き言葉です。例えば、先程言ったバルザックの作品だけ調べたければ、バルザックだけ取り出してみることが出来ます。そうではなくて、時代だけ、例えば 2010 年に出版されたコーパスだけ見たいとか、そういうことも可能です。あるいは 16 世紀のコーパスと 19 世紀のコーパスを比べてみたい。そういうのも可能です。データの幅も非常に広いですし、検索も割と可能に、またきめ細やかに作ってありますので、ぜひこれは試してみてください。

例えば、本学の学生でフランス語における「色」の概念に関心のある学生がいます。ある作家の作品に出てくる「赤」という色のイメージを詳しく調べたいと思うならば、「赤 rouge」という単語を検索します。そうすると、その言葉の現れるコンテクストがコンコーダンス形式で表れますので、その作家がどんなコンテクストでその言葉を使ったのかを調べることができるわけです。こんな風に、言語学だけではなくて、文学研究でも使用することができます。

FRANTEXT はほぼ 80 パーセントが文学作品であるので、その他のジャンルの

書き言葉に関しては別なコーパスを参照しなくてはなりません。

例えば、新聞のコーパスもあります。これも有料なので高くついてしまうのですが、「Le Monde（ルモンド)」コーパスが有名です。「Le Monde（ルモンド)」の新聞記事はインターネットでダイジェスト版をただで見られると思いますが、このコーパスは、その全文をテキストファイルに保存したものです。「European Language Resources Association（ヨーロッパ言語リソース協会)」という機関から売り出されています。一年分買うと 5~6 万円してしまうので、ちょっと高いかなという感じですが、これも研究室で購入しています。私の研究室配架なので、気になる人はどうぞ研究室に来てください。

こうした新聞コーパスなどを使って行われた研究の例としては、政治家の使うフランス語の研究があります。政治家の演説による社会的効果や文体の関係については、かなり多くの研究が残されています。

さて、ここで、再び話し言葉コーパスの問題に話を戻しましょう。先にも申し上げましたが、フランスでは話し言葉コーパスは研究室ごとに集積された個別なものが多いです。いくつかの個別コーパスを見ていきましょう。一番目に挙げるのは TCOF (Traitememt de Corpus Oraux en Français　フランス語の口語コーパス処理)という、これは ATILF――日本で言う国語研究所みたいなものに相当します――が出しているコーパスです。これもこのサイトに行っていただくと実際にコーパスをすぐ無料で検索できます。話者のタイプ―例えば、年齢、性別、学歴に分けた話者のカテゴリー、あるいは、発話状況―例えば、家の中で友達と親しく話したとか、家族と親しく話している様子なのか、あるいはビジネスの場で仕事の場で使っているのか、会議なのか、あるいは講演なのか、といったタイプごとに、データは整理されています。ですから自分が調べたいと思うタイプのデータにしぼって検索することが可能です。例えば、このデータを使用することで、例えば、フランス人の小さい子供が、いつ頃から、どんな風にフランス語の le la les という冠詞を身につけているのかを調べるということができます。このコーパスの弱点は、その規模が小さく、40 万から 50 万しかないという点です。どうして、こんなに小さいのかというと、このコーパスを作っている研究グループ GARS は統語分析に力を入れ、非常に一つ一つの単位の統語分析に時間を割いていることにあります。その分、データの質は非常に高いと思います。分析とか転写において非常にクオリティの高い信頼に足るコーパスです。

次の ESLO というのは、オルレアン大学で作っている非常に有名なコーパスで

す。こちらのコーパスはデータ収録の時期が二つに分かれていて、1968 年から 1974 年の初期と、2008 年から現在の後期があります。初期のコーパスは、外国人にフランス語を学習させるためにデータとして必要だという理由で、イギリス人が収録したものです。その後、フランス人がそのあとを引き継いでコーパスを整備して ESLO1 として公開しました。そして、後期のコーパスは今なお、ESLO2 として命名され、現在も拡充されている状態です。最近はビデオコーパスなどの収録も始めたようです。オルレアンのチームは、社会言語学的な問題にこだわっています。年令が違う人、性別が違う人、そういった話者のカテゴリーが違う場合、それらの人たちの話すフランス語には、どんな違いが出るのか、新しい言葉というのはどんな風に生まれて、どんなふうに広がっていくのか、――そういったことを社会学的なコミュニケーションの役割の観点からむしろ見ていきたい、というチームです。フランス随一の大きさのコーパスで、サイトにコーパスがオープンリソースとして公開されています。無料で、確実に自宅からもアクセスすることが出来ますので、みなさんもここのサイトに行ってみてください。転写をみながら、音も聞けるようになっています。

さらに、私たち東京外国語大学でもささやかながらコーパスを作っています。実際にフランスのパリ、ボルドー、エックスに行きまして、学生のインフォーマルな自由会話を録って、それを転写して、一部はタグ付けをしているので、品詞で検索ができます。字列だけではなく、動詞だけとか、そういう形で検索もできるようになっています。規模 40 万です。量、質ともにまだまだ向上していかなくてはなりません。私たちがどうしてインフォーマルな会話ばかりを録音しに行ったかというと、多分に教育的価値も考えていたからです。日本でフランス語を学んだ我々は、教科書で習う規範的なフランス語に慣れ親しんでいますが、巷で実際に話されている自然なフランス語のことはよく分かりません。だからこそ、それをデータとして録って調べてみよう――と言うのが一番の大きな動機です。

その次に、これは比較的新しくて 2000 年くらいからですけれど、パリ第三大学の、パリの話し言葉コーパスが出ています。これはパリに住むインフォーマントの会話を集めたコーパスです。

インフォーマントというのはデータを提供する協力者です。つまり、実際録音に協力してくれる人です。パリに住む人たちも、色々出身が違いますので、必ずしも、パリのフランス語を体現しているわけではないだろうと思いますが、このコーパスに関しても、非常に細かく様々な基準を指定して検索することが可能で

す。これも無料で公開されていますので、ご覧になる機会がありましたら是非見て下さい。

それから最後に、フランス語の話し言葉コーパスで忘れてはいけないのは、PFC と言われるもので、「Phonologie du Français Contemporain、現代のフランス語の音韻論研究」というものです。これは音にこだわって作成されたコーパスです。フランス語の音の地域による違いやその多様性を見るのに適しています。皆さん、カナダのフランス語と、フランスのフランス語というのは同じだと思いますか？違いますね。語彙も全然違うのでしょうけれども、まず音で聞いて、これはちょっと同じフランス語に聞こえないというくらい違うのです。そういった違いをこのコーパスで聞くことができます。最初は、この PFC というプロジェクトは国内のフランス語について始めたものですが、最近は、国外のフランス語圏に行って、そういうフランス語を録ることを始めています。ですからコーパスはどんどんどんどん大きくなっています。フランス語圏というと、アフリカもあるし、アメリカにもあるし、色々ある。北アメリカにもあるし、オセアニアにもある。今後は、規範的なフランスで話されるフランス語だけではなく、地域ごとにその多様性というのを認めるべきだとフランス人の言語学者が動き始めたところなのです。このプロジェクトに関わっている人の一人に、早稲田大学のシルヴァン＝ドゥテさんも述べていることですが、フランス語というのは、非常に豊かなレパートリーを持った言語です。地域、時代、文体に応じて様々な多様性が存在する。規範に合致しない用法を全て消してしまうことは、言葉の豊かさを失うことなんじゃないでしょうか。そこで、PCF のようなプロジェクトを立てて、フランス語の地域版―例えば、オック語ではなくて、オック語の影響を受けた地域版のフランス語―というものを聞くことが出来るように、サイトを立ち上げています。

さて、ここから先は、少しコーパスを使った研究成果について見ていきましょう。コーパスをやっている中で一番大切な成果というのは語彙の研究です。TLF (Trésor de la langue française　フランス語大辞典)というのは本当に素晴らしい辞書です。フランス語をちょっとでも読める人がいたら、ぜひともこの TLF をよく見て読んでいただきたいなと思います。これは、先程紹介した FRANTEXT コーパスを基盤として作成されたフランス語の大辞典です。1994 年に紙版が出来上がり――紙版は本学の図書館にもあります―現在では、素晴らしいことに、インターネット上で無料公開されています。

ただし、この辞書というのは FRANTEXT を基盤にしているので、書き言葉の

データでわかる範囲でしか記述がされていません。ですから、話し言葉の例というのはあまり考慮されていません。したがって話し言葉特有の用法は抜け落ちてしまっていることが多いです。

例えば、après という前置詞があります。フランス語で「あと」という意味の前置詞です。この après というのは　après le repas「食事のあと」のように、主に、前置詞としても使えます。また、節を導いて、après que...「〜したあと」というようにも使う。そして例えば、je viens après「あとで来ます」のように副詞としても使う。おおよそ、TLF で記述されるのはこうした用法にとどまります。

ところがもう一つ大事な用法が、話し言葉では沢山現れています、それがいわゆる単独で接続表現として働く例なんですね。例えば、Je vais aller au cinema ce soir, tu viens avec moi?「今日映画に行くんだけど一緒に行かない？」Mais après, si tu veux pas.「でも、君が嫌だっていうんだったらね〜」というふうに、何か言った後に、前に述べた発話に対して、留保するような内容の発話を付け足すような感じで、après というのが良く使われます。このような après は、話し言葉でしか使われません。

日本語の「あと」も同じですね。学生の話で「飲み会は武蔵境がいいな。あと駅に近いところね」――例えばこんな風に言うのを日本語の話し言葉コーパスの中で聞いたことがあります。また、うちの息子が母の日の手紙を書いたところ、先生に直された箇所がありました。息子は「いつも仕事してくれてありがとう。あとお弁当作ってくれてありがとう」と書いたのですが、先生は「あと」ではなくて、「そして」にしなさいと直します。話し言葉でよく見られる発話の継続性を表す「あと」の用法ですが、規範的な話し言葉や書き言葉では使わない、あるいは誤りとされています。

みなさんに、ここで理解していただきたいのは、データをどうとるかで言語分析が大きく変わってくるということです。たとえ大きな辞書であっても、それが書き言葉をベースにして記述されたものであれば、話し言葉でよく使う用例が抜け落ちてしまうということはよくあり得るのです。

言語学をちょっとかじった方はご存じかもしれませんが、20 世紀の智の偉人と呼ばれるアメリカの言語学者であるチョムスキーという人がいます。チョムスキーという人はコーパスには大反対だった人です。どうしてだか分かりますか？「コーパスに出てこない例だっていくらでも人間は作れる」からです。コーパスには収録できるデータに限界があり、確かにコーパスで全てが分かるということ

はありません。その意味で確かにチョムスキーは正しい。コーパスを補完するために直観テストというものを組み合わせる必要があります。この文は合っているか、合っていないかということを直観のあるネイティブに判断させて、テストしていくのが直観テストです。

ところがこのテストにも、実は大きな落とし穴があります。私が最近よく若者が使っている「ほぼほぼ」という表現を間違いだと言ったら、それは正しいことでしょうか？よく考えてみてください。人の直観、言葉に対する直観というのは、その人の美的感覚といいますか、非常に主観的な判断や評価になってしまうものなのです。ですから、もし、直観テストをするのであれば、たった一人のネイティブに頼らず、できうる限り多くのネイティブの直観テストを用いるべきです。また、「ほぼほぼ」があっているかどうか、という点については、コンテクストをはっきりとさせて、テストをする必要があります。例えば、私は「ほぼほぼ」をこのような講演の中では使いませんし、また学生のレポートの中に書いてあったら赤でしっかり訂正します。しかし、学生が雑談の中で「ほぼほぼ」としゃべっていても、それは理解できますし、直しもしません。

つまり、直観テストをする場合には、それを判断する人たちに、どのような状況で用いられる例であるか、をきちんと説明した上で行わないといけないわけです。調査者のコントロールがしっかりできていれば、おそらく、直観テストとコーパスを組み合わせることにより、より精密な調査が可能になることでしょう。

コーパスを利用した研究の例をさらに挙げましょう。フランス人というのは文章を書く際にどのように論理を組み立てて論拠を挙げて説得力ある文章を書くのかを研究する。それには、文の内部の構成ではなく、文と文の繋がり、談話自体の内部構成を見る必要があります。作文コーパスなるものを作成し、データを集めて検証することで、こうした問題にも取り組むことができるでしょう。

また、ある程度の時間を通してみた言葉の変化なども通時的なコーパスを元に研究することが可能です。とりわけ借用語や新語などについて、新しい言葉がどうやって入って変化して、また廃れていくかというようなことを研究していくことで、言葉のダイナミックな側面を明らかにすることが可能でしょう。

さらに、方言研究については、そもそも、コーパスに頼ってやっていたところがあります。フランスでは、方言のコーパスなども地図とともにかなり電子化されています。膨大なコーパスの研究というのは、実は方言研究では昔から当たり前に存在しています。

統語の分野でもコーパスは重要です。例えば、受け身の使い方をテキストごとに調べるということができるでしょう。フランス語は受身の形が結構たくさんあります。「être+過去分詞」というのもあるし、「代名動詞」の形もあるし、「使役受身」みたいのもあるし、on という主語をちょっと受身に代用することもある。すごくバラエティに富んでいます。テキストのジャンルの違いやコンテクストの違いに応じて、これらの形式がどう使い分けられるのかということは、コーパスを調査することで明らかになる部分があります。

このように、コーパスを使って、研究することのメリットは、規範文法に書かれていることに縛られずに、皆さんが自分の頭で考えることが出来ることだと思います。そしてその考えるというのはとても大事な活動であろうと私は思います。様々な言語学者や規範文法で書かれたり言われたりしていることを知識として覚えるだけではなく、そうしたことを実際に自分でもデータを見て検討していく、自分の頭で考え直してみる必要はいつでもありますし、そうしたことから、将来、学問的にも役立つ発見が生まれていくと思います。

コーパスを使ってできることは、何も研究には限られません。外国語を学習するためにもコーパスが使われることがあります。

外国語を学習したときに経験なさったと思いますが、文法書で規則だけを学んだだけでは発話を組み立てることができません。その規則に語彙がどうやってマッチングしているか、あるいは、どのようなコンテクストで使えるのか、ということを学ばなくては規則を使いこなせないのです。

語や規則の結びつきやコンテクストを知るためにはコーパスがとても役立ちます。受け身の「être＋過去分詞」という形は、砕けた話し言葉の中では限られたタイプの組み合わせ以外はあまり頻繁には用いられません。ちょっと硬い文章であるとか、客観的に物事を表現するコンテクストで使われることがむしろ多いように思います。日常会話では、「物」よりもむしろ「人」を主語にして、代名動詞の使役受身みたいな形で代用させてしまうことが結構好まれるわけです。人を中心として事態を組み立て、その人に対してどのような事態であったか（迷惑や被害、利益などのニュアンスを含め）表現することの方が頻度的に多いわけです。そうしたこともコーパスをじっと見ていると分かることの一つです。

コンテクストに置かれた構文のパターンを知り、それを教授法に取り入れることで、日本人のフランス語力は飛躍的に伸びる可能性があります。さらに、学習者のフランス語コーパスを作成して、その誤用や非用を記述するということも可

能です。

ツールとして、フランス語を勉強されている方には役に立つと思いますが、カナダのサイトで、lextutor というのがあります。これを用いると、コンコーダンス形式でいろいろな語彙を検索できます。実際に語彙の使用されているコンテクストを見ることができます。それからもう一つは、fleuron です。これは、先程話した ATILF――フランス語の国研みたいなところです――が留学生向けに出しているサイトで、フランスの日常生活に関わるいろいろな場面での会話をビデオで見ることができます。場面ごとに、例えばパン屋で買い物するにはどんな手順で話をして買い物をしたらいいか、ピザを注文する時にはどうしたらいいか、大学の登録の際、どのようにして事務の受付の人と話したらいいか、フランスでうまく生活していくにはどうしたらいいかといった会話をモデルとして観察することができます。音声と転写が両方あるので、大変便利です。

先程も話しましたが、フランス語というのは、19 世紀以来学校教育が義務化されて、中央集権的なやり方を認める中で、とりわけ戦後、フランス語の多様性を切り捨ててきてしまった歴史があります。そのせいか、現在フランス語というのは、非常にダイグロシア的な状況にあります。どういうことかというと、規範的なフランス語がリンガフランカとして高層的な言語のレベルをなしています。ところが、インフォーマルなコミュニティ内言語としてのフランス語というのは、きちんと存在しているにもかかわらず、教育の場でも、学問の場でも、あまり光が当たっていないわけです。高層の言語を推進することは悪いことではありません。人々によりよい教育を受けさせ、就職の機会を与えるために、全ての子供たちに、平等に高層レベルのフランス語教育を与えるというのは大事なことでしょう。

しかし、それ以外のことばを切り捨ててしまって良いかというと、そうではないだろうと思います。そこを切り捨てていくと、高層文化からはみ出てしまった人たちが価値を認めてもらえないまま、フラストレーションを抱え、歪んだ社会になっていってしまうのではないでしょうか？インフォーマルなフランス語に、コミュニティ内部の言語としての機能を認め、それに光を当てていくこともまた大切な言語学者の仕事であると思います。

今、カタロニアの話が今盛んになされていますね。フランスの国境のスペイン側とフランス側では同じ方言を使っています。しかしながら、フランスとスペインというのは、国境一本で全然状況が変わっていて、フランス側のバスクやカタ

ロニア語の学校教育は 2 言語併用ではで行われていません。つまり、フランス側では、フランス語だけでしか行われていないのです。ところが反対側のスペイン側では、二言語で学校教育するのが当たりまえなのです。――地方のことばが社会的にもちゃんと市民権を得ている。それどころか、その文化的な多様性を認めるだけにとどまらず、政治的にも独立したがっているわけです。もっともカタロニア地方がスペインでも非常に豊かな地域であるということもあるとは思います。フランスではカタロニア語にしろ、バスク語にしろ、文化的にも独立を認められていないという現状にあります。大きな違いを見ることができます。PCF のコーパスの中で、フランスのバスク地方の青年が話していたことですが、「フランスではまず、僕が大切に思うのは文化的な多様性を認めてくれることなんだ。独立――政治的な独立というわけじゃなくて、文化的にまず認めてほしい」。フランスにおいて少数言語の文化的独立は少なくても認めるべきだという意見は大きいと思います。しかし、文化的独立を認めるということは、多少は政治を介さなくてはなりませんし、それは、実際に、選挙で投票して法律を整備し、予算を割きといった政治的手段を講じる必要があるだろうと思われます。言語学者としては、そうしたこととは別に、人々のメンタリティの中に多様性を重んじることの重要性を学問的に裏付けていくことではないかと思われます。

それから最後に、インフォーマルなフランス語の研究という観点で、ラップについて言及しておきましょう。セルジ＝ポントワーズ大学の Pruvost 教授が mouv' というラジオ番組に、毎回 Doc Dico という名で登場して行っている番組でラップに出てくる言葉を一つずつ紹介しています。ラップに出てくる言葉なんて、どうしようもない言葉ばかりじゃないのか、と馬鹿にして聞いていたら、実は中世の、昔の古い表現が民衆語の中に残っていたものであったり、何世紀も前にヴァイキングたちを通して、海を渡って、外来語として入ってきた言葉であったり、とその起源はそれなりに由緒正しいものであって、それが様々な変化を経て、フランス語に定着したものばかりで大変面白いのです。例えば、avoir la dalle という表現は、若者ことばで「腹が減る」という俗語なのですが、この表現がそもそもどこから来たかというと、14 世紀に、スカンディナヴィア語から入ってきた dæla とかいう言葉らしいんですね。そして、その言葉が当時は、なんて言ったらいいか、水を流す――「流し」、あるいは「流しの穴の開いた部分」を指していたそうなのです。時代が変わるにつれて、その dalle というのは、現在フランスでは水場の床に敷くタイルみたいな意味に変化してしまったのですが、14 世紀く

らいに入ってきたときには、「排水溝の部分」を表していたそうです。実はその意味が、ラップで使われている la dalle に残っているのです。つまり、「喉」という意味でのメタファーになっているわけです。そこから排水溝に水が流れる、というので、rincer la dalle というと「喉を流す」という意味で、まあ「酒を飲む」という俗語になり、avoir la dalle というと、「食道」をメタファーとして「お腹が空く」という意味になるのだそうです。

このように、民衆語の中にびっくりするくらい昔の起源のことばの記憶が蓄積されているのです。ラップで用いられる俗語というのは、必ずしも若者が使う新奇な言葉ではなく、実は標準語からは姿を消しつつも、ひっそり生き残った言葉であることが多いのです。そういうのを毎週一回くらい、ラジオ番組で流しているわけです。若者は自分で使っているくせして、そうした言葉の歴史については何も知らない。でもそうした歴史を聞くと、わーっと喜ぶわけです。若者が自分の使う言葉に対する関心や自分の使う言葉に何らかのポジティブなイメージを持つことは大切なことなのではないでしょうか。そうしたことにも一役買っている言語学者がいるということはいいことだと思います。

コーパスに基づく研究が言語学に豊かな実りをもたらしたことは確かな事実です。しかしながら、データは果たして本当に客観的なのか、コーパスは万能なのかという点については、今一度、批判的に考えておかなくてはいけません。

コーパスというのは、転写をし、タグを付けていく段階で、既に主観的な分析を通して作成されているものなのです。ですから、厳密に言えば、実はデータというのは、何一つとして客観的なものは存在しえない。コーパスの選び方、利用の仕方次第で、そこから得られるものは大きく変わってきてしまいます。

今回は、みなさま、コーパスがテーマの一連の講義をお聞きになっていると思いますが、「コーパス至上主義」に陥らないよう、やはり注意をしなければいけないと私は思っています。コーパスは便利なツールですが、どんなコーパスを用いているのか、自分の見ているデータはどのようにして作成され、どのような特性を持つものなのか、をよく吟味したうえで言語分析に役立てるべきであるということを最後に繰り返しておきたいと思います。

恐らくデータと向き合う際、データから見える真実が何であるのかを突き詰めて考えていくならば、分野の違いは問わず、どのような学問であれ、多かれ少なかれ、同じような問題に直面するのだろうと思います。

というわけで、私の話は以上になります。ご清聴ありがとうございました。(完)

コラム † マルセイユのお勧め遠足コース

サンシャルル駅から見たマルセイユの街並み

…・* ―――――――――――――――― *・…

マルセイユ は南フランス、ブッシュ・デュ・ローヌ県の県庁所在地であり、パリ、リヨンに次ぐ大都市だ。地中海に面した港町として歴史のある街で、多くの移民の住むコスモポリタンの街である。なぜか治安の悪いイメージが強くちょっと近寄りがたいイメージを持っている人もいるのではないだろうか。私はマルセイユ近郊のエックス・アン・プロヴァンス市に 8 年半ほど住んでいたこともあり、マルセイユには何かと出かけることが多かった。当時よく行ったお気に入りの日帰り遠足コースの一つを紹介しよう。

マルセイユの鉄道駅 Gare St Charles（サンシャルル駅）は小高い丘の上に立つ近代的で美しい建物だ。パリから TGV で 3 時間と便もよい。上の写真は駅前広場からマルセイユの街を見渡した眺めである。駅の周辺から Noailly

市場くらいまでは治安がよくないので、スリなどに狙われないよう一路旧港を目指してサッサと歩こう。

右写真は、View Port（旧港）である。旧港では朝には魚市が開かれ、新鮮な魚を買うことができる。港付近はカフェやレストランが立ち並ぶ。また港の突端に立てられた地中海美術館では、地中海をテーマとした様々な展示が催されている。写真に写る Notre Dame de la Garde ノートルダム・ドゥ・ラ・ガルド寺院はロマネスク調の美しいカトリック教会だ。航海の無事を願う人々の祈りの場でもある。港から少し西方面へ上ると、Panier という地区がある。最近はすっかり観光地化したが、移民の住民も多い庶民的な界隈だ。年に一度、住民が中心となって近隣祭 Panier 祭りをしている。近隣祭はとても楽しいので、時期的に行けるチャンスがあれば、ぜひともお勧め。

さて、市の中心地からバスにのって海岸にそって移動すること 40 分。終点 Madrague というカランクの入り口の町へと到着（左写真）。地元の人たちのマルセイユ訛を聞きながら、のんびり散歩。海辺のレストランでは、シンプルでおいしい魚料理が食べられる。私のお気に入りは香草で味付けしたマグロのステーキだ。日本人としてみれば、マグロはぜひ刺身にしてもらいたいところだろうが、このステーキはなかなか美味しい。

脚に自信がある人は 6 時間ほどかけてカランクめぐりをしながら、隣町の Cassis カシス（よそ者はカシというが、マルセイユの人たちは末尾の s を発音して、カシスとよぶ。）までハイキングをするといい。危ない道ではないが、石がゴロゴロした海沿いの道を歩かねばならないので靴はしっかりしたものがいい。途中には店も水場もないので水と弁当を持参する。一日歩いた疲れも吹き飛ぶカシスの銘酒 Blanc de Blancs があなたを待っている。カシスからはマルセイユまで船で帰るのもよし。余裕がある人はカシスで一泊して、翌日、Bandol　バンドルまで足を伸ばそう。コートダジュールのリゾートとは一味違った小さく美しい真珠のような港町とカランクの美しさをぜひとも堪能してもらいたい。

フランスを知るための３冊

・*・…‡…・*・*・…‡…・*・…‡…・*・

「フランコフォンの世界 コーパスが明かすフランス語の多様性」La variété du français parlé dans l'espace francophone Ressources pour l'enseignement
Sylvain Detey, Jacques Durand, Bernard Laks, Chantal Lyche
川口裕司・矢頭典枝・秋廣尚恵・杉山香織（日本語訳翻訳）
三省堂　2019 年

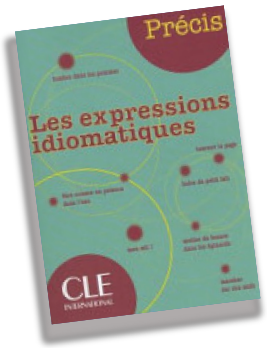
本書は、2010 年 Ophrys 出版社からフランスで発刊された、*La variété du français parlé dans l'espace francophone, Ressources pour l'enseignement*（Sylvain Detey, Jacques Durand, Bernard Laks, Chantal Lyche）の翻案である。翻訳ではなく、内容を一般向けに分かりやすく単純化してリライトしたものである。話し言葉のフランス語の総論に続き、北フランス、南フランス、ベルギー、スイス、アフリカと海外県、海外領域圏、北アメリカの地域で話されるフランス語の特徴を開設したものである。公開講座でも紹介した PCF（現代フランス語の音韻論）プロジェクトの研究成果を反映したものであり、そのコーパスの抜粋とコーパスについての分析が掲載されている。様々な地域のフランス語の多様性を学ぶのによい本である。

…・* ———————————————————— *・…

Atlas du français de nos régions Avanzi Mathieu
Armand Colin 出版社　2017 年

本書は、フランス語圏の様々な地域でどのような言葉の使われ方の違いがあるかについて、具体例を挙げて紹介した語彙地図。チョコレートパン、鉛筆…など日常生活に欠かせない語彙の地域差を分かりやすく示したもの。初中級のフランス語が分かれば十分楽しめる語彙地図。カラフルでとても美しい本でもある。絵本感覚で気軽に開いて、内容に親しんでもらいたい。

…・* ———————————————————— *・…



Les expressions idiomatiques
Isabelle Chollet, Jean-Michel Robert
CLE International 出版社 2008 年

フランス語の慣用表現集。慣用表現は文化的に固定化された表現であり、フランス語学習者にとっては類推がなかなか難しい表現も多いので、頭を悩ませているのではないだろうか。しかし、その反面、フランス文化特有の言い方を学ぶことで、フランス語の面白さを学ぶことができる分野でもある。本書はフランス語の初級中級を学んだ人向けである。慣用表現の意味的なタイプに応じて分類され、例文と共に覚えることができる。

… • * ———————————————————— * • …