

国語学・国語科教育におけるコンピュータ利用

国文学科・准教授 中俣 尚己

1. 国語学におけるコンピュータ利用

筆者は隔年開講である「国語学特講 A」という授業で、IPC を利用している。文系科目、特に国語科とコンピュータの利用にどのような関係があるのか、と思われる方がいるかもしれない。

筆者の専門領域は国語学（日本語学）で、いわば日本語という言語を研究対象としている。これまでの言語の研究では例文を自分で作成し、内省に基づいてその例文を判断するというスタイルで研究を行ってきた。例えば、「*夜に甘いものを食べないべき」とは言えないので、否定表現は「べき」の前に来ることはない」と記述する、といった具合である。

しかし、近年ではコーパスと呼ばれる大規模な言語データが整備されるようになり、言語の研究も、実際にそのような言われ方がされるのかどうかという観察を元に論じることが多くなってきている。言うまでもなく、これは自然科学としては当然の態度である。実際、先程の例について言えば、web から大量の文書を集めた日本語最大のコーパス「国語研日本語ウェブコーパス」(100 億語)を検索すると、「べきでない」は 15,250 件見つかるのに対し、「ないべき」も 2,420 件見つかる。これは例外として切り捨てるにはやや多い数字である。実際のところは、現在、「べき」の否定形には揺れが生じていると見るべきであろう。さらにデータを観察すると、「ないべき」は「食べるべきか、食べないべきか」のように対句的に用いる場合に多く用いられるということもわかる。

2. 「国語学特講 A」の授業内容

まず、コーパスというものを簡単に定義する。石川(2012:13)はコーパスの条件として、(1)書き言葉や話し言葉などの現実の言語を、(2)大規模に、(3)基準に沿って網羅的・代表的に収集し、(4)コンピュータ上で処理できるデータとして保存し、(5)言語研究に使用するものとしている。

このコーパスを元に研究を進める方法としては2つのタイプがある。1つは、手元にテキストを用意して、自分で目的に合わせたコーパスを作り、分析を行う手法である。これは児童の書いた作文の特徴を調べるとか、教科書に収録されている作品の重要語句を抽出するといった目的に適した研究法である。実際、筆者が担当する大学院の授業「国語学特論」ではこのタイプの研究手法を扱っている。しかし、これは大量のテキストデータを用意する必要があること、正規表現が使えるテキストエディタや「KH Coder」(樋口 2014)などのテキストマイニングソフトウェアを使う必要があることから、学部授業では行っていない。

もう1つのタイプは、他の研究者が構築した大規模コーパスを利用する方法である。コーパスの種類にもよるが、例えば日本語の書き言葉をバランスよく集めたコーパスからは日本語の全体像を窺い知ることができる。先にあげた「べき」の否定形の現状の調査などはこのタイプの研究である。「国語学特講 A」でもこのタイプの研究手法を身につけることを目的としている。

「国語学特講 A」では国立国語研究所が開発した『現代日本語書き言葉均衡コーパス』(Balanced Corpus of Contemporary Written Japanese:BCCWJ)とコーパス検索アプリケーション『中納言』を利用している。BCCWJは1億語の規模を誇る世界でも有数の均衡コーパスである。均衡コー

パスとは書籍、雑誌、新聞、白書、教科書、広報紙、ベストセラー、Yahoo!知恵袋、Yahoo!ブログ、韻文、法律、国会会議録と 13 種類の文書からバランス良くデータを集めたという意味で、日本語の全体像に迫ることを可能にしている。「中納言」は BCCWJ を始めとする種々のコーパスを検索するための web アプリである。最大の特徴は、語の様々な情報を元に検索できることであり、例えば、普通の検索システムでは「行く」と検索すれば、この形のみが検索され、「行きます」「行って」のような活用形は検索できないが、中納言は複数の表記や活用をまとめて検索できる。さらには、副詞「いよいよ」から後方 5 語以内に出現する動詞を全て検索する、といった複雑な条件の検索も行える。このような検索をプログラミングの知識なしに出来るので、文系の学生にとっては非常に使いやすいツールである。

図1 『中納言』の検索インターフェース

「国語学特講 A」の授業計画は講義（実習）と学生の発表からなる。講義は 3 段階に分かれ、第 1 段階では、まずは中納言を使って検索を行うやりかたを学ぶ。中納言にはブラウザでアクセスする。図 1 の通り、シンプルなインターフェースではあるが、自分が調べたい内容を自在に調べるにはある程度の訓練が必要である。また、「いよいよ」を漏れなく検索するためには、「語彙素 が 愈」というように、見慣れない漢字を打ち込む必要がある。もちろん、どのように打ち込めば良いかは中納言を使って調べればわかるのであるが、その調べ方を身につけなければならない。参考までにこの段階で学生に課された課題をいくつか紹介する。

- (1) 「かもしれない」が中納言でどのような語彙素・品詞になっているか調べよ。
- (2) 動詞「見る」の後 3 語以内に「ない」が出現する例を全て検索せよ。

第 2 段階は検索した結果を加工、集計する段階である。検索結果は図 2 のように表示される。いわゆる KWIC(Key Word in Context)と呼ばれる表示形式であるが、これを眺めても、例えば「いよいよ」の後にどんな語が何回出現するのか、その個数まではわからない。中納言には検索結果を CSV 形式でダウンロードする機能がある。これを利用して CSV ファイルをエクセルで開き、ピボットテーブルやフィルタを利用して、集計したり、自分がほしい例を抽出するといった

作業を行う。この段階は完全にエクセルの授業と言ってよいであろう。この段階で行う課題には以下のようなものがある。

- (3) 「てもらう」の用例から「てもらいたい」という形だけを抽出せよ。
- (4) 「名詞+を+とる」というという形で検索、ダウンロードし、共起する名詞の頻度表を作れ。

2,037 件の検索結果が見つかりました。そのうち 500 件を表示しています。
検索対象語数: 124,100,964 記号・補助記号・空白を除いた検索対象語数: 104,911,460

サンプル ID	開始位置	連番	前文脈	キー	後文脈	語彙素読み	語彙素	語彙素細分類	品詞	活用型	活用形	レジスター
OB6X_00081	3630	2440	の略称)の活動も、軌道に乗っていき たし、(いよいよ)世界を舞台に	し	て、本格的な活動を開始する時 が来たといっはいいだろう」#	スル	為る		動詞- 非自立 可能	サ行変 格	連用形 一般	特定目 的・ベス トセラー
PM41_00395	44520	27490	裏打ちが終ると、木杵に戻され、 絵具のはかれた部分に、(いよいよ) 色ぬり	加え	られる。#わたしも驚いたため、この 補彰で、は、オリジナルと同じ種類の の絵具で	クフェ ル	加える		動詞- 一般	下一段 -ア行	未然形 一般	出版・雑 誌
LBh3_00029	6930	4290	こうでなければ(寛慢だと言わんば) りです。# それゆえ(いよいよ)アメリカ をたとうと	し	て(出国)手続きをしていいたとき、そ の係員と話しているうちに、「	スル	為る		動詞- 非自立 可能	サ行変 格	連用形 一般	図書館・ 書籍
LBm9_00023	106900	71510	参詣者も倍増し、以前は長谷寺の 配下にあったのが今度(いよいよ)独 立	し	て(真言宗の密生派の本山という こと)に昇格したため、それゆえに	スル	為る		動詞- 非自立 可能	サ行変 格	連用形 一般	図書館・ 書籍
LBk3_00001	82790	56350	の面々が刀を抜いてたちどころに 二、三人を斬り落とす。#先(いよいよ) 激昂	し	、(驚も)火事(装束で駆け付け、足輕 小者を打ち倒し、ついには中納言 の乗物へ	スル	為る		動詞- 非自立 可能	サ行変 格	連用形 一般	図書館・ 書籍
PB59_00130	31190	20880	して(名前が同じ)か(容貌が)そっくり と、か(設定では、(いよいよ)馬鹿馬鹿 しいと	受け取 ら	れる)ばかりなのに、(ま)くろを(転生) の(証拠)とした。#これなら(二十世紀) の	ウケ ル	受け取 る		動詞- 一般	五段- ア行	未然形 一般	出版・書 籍

図2 『中納言』の検索結果

第3段階は得られた集計結果に対し、統計的な操作を行う段階である。統計といっても極めて初歩的なもので、母数が異なるサブコーパスの出現数を比較するために調整頻度という指標を計算したり、あるいは出現数の差が有意であるかどうかを、 χ^2 検定を用いて確かめるといった内容である。

3. コーパス調査から得られた知見を国語科教育に活かす

講義の後には、学生が2回ずつ、自由にテーマを選んで発表する。この授業の当初の目的は、「日本語の使用実態を調査する技法を身につけること」であり、付随してコンピュータの操作や統計について習熟することを挙げていたが、授業を進める中で、コーパス調査の結果は国語科教育の授業にも応用できる可能性があることがわかってきた。

例えば、ある学生は「美しい」と「きれいな」について、直後に修飾される名詞を調べてきた。この授業の発表ではレジメも用意してもらうが、それだけではよくわからない場合、学生が調査したエクセルの画面を全員で共有しながら考える。「美しい」と「きれいな」の場合、学生は直後に出現する名詞の頻度表までは用意してきたが、その結果をうまくまとめられていなかった。例えば「花」や「女性」は「美しい」の後にも「きれいな」の後も多く出現する。

しかし、他の学生から「景色・光景・自然」は「美しい」が多いということなどから、「きれいな」は単純な性質だが、「美しい」はそれを見る人間の解釈を交えた複雑な性質ではないかという意見が出された。これはなかなか鋭い観察である。さらに、「きれいな水」とは「飲める水・安全

な水」という意味であるが、「美しい水」は実際には自然の美しさについて描写する際に使われているという違いも指摘された。ということは、「美しい」とはモノに内在する性質というよりは、それを見た人間の心の動きを表す語であるということになる。

(5) きれいな水 を～ごくごく飲むのさ～…懐かしいw

(6) 島根県邑南町の自然豊かな風土、澄んだ空気、美しい水、そこで育つ良く作物をエサに大切に育てられた「石見ポーク」ちゃん。

このような違いは、辞書には記述されておらず、コーパスを使って「生きた用例」を多数観察することで初めて気がつくことである。

その後、別の学生の教育実習の実地指導に行く機会があったが、その内容は高校1年の論説文「美しさの発見」(高階秀爾)であった。この論説文の主張がまさしく、「美しさ」とはものに内在する性質ではなく、それを見る人間の心の中に存在するのであるという内容であった。もし、この学生がコーパスを使った「美しい」と「きれいな」の分析を聞いていたら、より深く筆者の授業を理解できたかもしれない。また、生徒達にコーパスから得られた情報を提示することで、筆者の意見に対して実感を持って納得することができたかもしれない。偶然のタイミングではあったが、コーパスを用いた言語研究と国語科教育の実践を繋げる可能性を感じた。

また、大学院の授業ではあるが、中学2年の文学教材「故郷」(魯迅)において、情景描写として使われている「鉛色の空」と「紺碧の空」に着目し、それぞれの色が他にどのように使われているのかを調査した学生もいた。結果、「鉛色の」は曇り空の描写に用いられる事が最も多い。他方、「紺碧の」は海、それも「地中海」とか「太平洋」「アドリア海」といったリゾート地の海の描写の例が多いことがわかった。文学の読解において肝要なのは、その色が実際にどのような色なのか知るのではなく、「鉛色の」「紺碧の」といった語が喚起するイメージを掴むことではないだろうか。コーパスを使って大量の用例を集める方法は、そのために最も適した教材研究の方法と言える。

4. 今後の課題

筆者のバックグラウンドは非母語話者のための日本語教育であり、コーパスの利用も外国語学習としての日本語学習のための利用という観点から考えることが多かった(中俣 2014)。しかし、学生の発表を通して、国語科教育における読解においてもコーパスから得られる情報は有益であるということがわかったのは大きな収穫であり、今後、さらに有効な活用法を検討していきたい。

また、児童・生徒の語彙力を増やすための教材のアイディアも得られた。これは今後実地での試用を経た後に、商品化していくつもりである。

最後に、国語学特講 A では他者が構築したコーパスを使って日本語の全体像に迫る手法を扱ったが、教材分析という観点からは学習教材そのものを分析する手法も有益である。今後はそのような研究も学部で授業で扱っていきたい。

参考文献

石川慎一郎(2012)『ベーシックコーパス言語学』ひつじ書房

中俣尚己(2014)『日本語教育のための文法コロケーションハンドブック』くろしお出版

樋口耕一(2014)『社会調査のための計量テキスト分析—内容分析の継承と発展を目指して』ナカニシヤ出版