

CA1821

辞書の向こう側： 生きた用例と辞書を往き来する

たかはし
高橋さきの*

「電子化されたテキスト」というかたちで蓄積されてきた先人の文章を「電子化された辞書」と行き来しつつ手軽に利用できる環境が実現しつつある。

1. 記憶をたどって青空文庫を読む

青空文庫という存在の大きさについては、いまさら多言を要すまい。青空文庫を支えてこられたみなさんのおかげで、著作権保護期間を満了した作品を、書籍の香りをしっかりと残したかたちで作品として読むことができる。しかしそれだけではない。『トロッコ』で「トロッコはどんな音をたてて動き出したんだっけ」という疑問も、テキストデータである青空文庫なら、検索という手段でたちどころに解決する。

そう、トロッコは3人の子どもの力が揃うと「突然ごろりと車輪をまわし」、この音にひやりとしつつも「良平」たちが押し続けると、「ごろり、ごろり、——トロッコはそう云う音と共に、三人の手に押されながら、そろそろ線路を登って行った」のである。

しかし大抵の場合、前後も読んでみたくなるはずで、『トロッコ』のこのくだりは、冒頭の物語りが動き出すシーンなので、視線は自ずと先へと向かうことになる。

かくして検索者は、「良平」がトロッコに飛び乗ると、トロッコが最初は「徐おもむろに」、それから「見る見る勢いきおいよく」動き出し、さらには風景が「忽たちまち両側へ分かれるように、ずんずん目の前へ展開して来」る現場に立ち会うことになる。

「ごろり」を探してたどりついたはずの『トロッコ』冒頭という現場で、「ごろり」という状況に「突然」遭遇し、「ごろり、ごろり」に耳を澄まし、「そろそろ」、「徐ろに」、「見る見る勢よく」動くトロッコを観察し、さらには「良平」と一緒に「ずんずん」を体感したということだ。

文脈という存在のもとで作品の「現場」を体験するというのは、たとえばこういうことなのだと思う。「用例は、まるごとのすがたで法則をおしえてくれる」⁽¹⁾。

青空文庫は、こうしたかたちでも利用されてきた。

2. コーパスとして青空文庫を使う

記憶の奥深くに沈潜していた特定作品の特定場面に

検索を介してたどりつくのではなく、「ごろり」の使用場面を多数確認したいという場合には、『トロッコ』1作品のみでなく、青空文庫全体を検索対象とすることになる。これは、青空文庫を文章表現のコーパスとして利用するということである。

「検索」や「コーパス」というと、コンピュータ以降の世界のことと思われがちだが、そんなことはない。文学作品は、電子化される前から、文法研究などに利用されてきた。高橋の「文法資料としての文学作品—文法形式のゆたかさの面から」から一節を引用する⁽²⁾。「文学作品は、もの見かたの独自性とむすびついて、表現のしかたがゆたかである。これは、新聞や評論の文章とくらべてタイプ化がよわいことであるが、このことが、ゆたかな形式を得られるという、文法研究にとって有効な結果をもたらすことになる。具体的にいえば、文学作品を資料にすれば、基本的なことが、だいたいいえるということである。」

ここには書かれていないが、文学作品であれば文庫本が使えた。コピーがまだまだ高かった時代、比較的廉価で、カードに貼り込みやすい文庫本の存在は貴重だった。当時は、文庫本のページを張り込んで必要箇所を赤鉛筆で囲むなどしてこしらえたカードを特注の引出しにあいうえお順にしまうなどして、検索可能な環境を整備していたのである。こうしたカードは辞書の語釈執筆の際にも使われた⁽³⁾。

文学作品を中心とする青空文庫には、パターン化を免れた表現の多様性や、よく推敲された安定感が確かにある。さらに、文学作品からははみだす幅広いジャンルの作品、たとえば科学随筆や評論（例：寺田寅彦、戸坂潤）も収載されているし、事態が順調に推移すれば、2021年には1970年没の作家の作品までが収載されることになる。

そして、青空文庫であれば、検索した表現を、上であげた『トロッコ』のケースのような前後数段落にとどまらず、それこそ作品全体という現場で検討することもたやすい。

さて、ここまでは、十数年前から言語関連分野の研究者の間で日常的に行われてきた作業である。しかし、文章執筆をなりわいとする者にとって、文章執筆の傍ら青空文庫をコーパスとして使うというのはハードルが高かった。つまり、辞書を引きながら、その語彙の複数の用例を青空文庫で検索して文脈を確認し、そうした作業を通じてその語彙をきっちり身につけ、自信をもって執筆中の文章で使うというのは、大変手間のかかる作業だったということだ。

特に、日本語を母語とする自分のような翻訳者の場合、訳出対象文書に書かれている内容そのものの理解や確認だけでなく、相対的に不得意な英語側表現の検

* 翻訳者・お茶の水女子大学非常勤講師

討作業にも時間をとられる分、言語力の源たるはずの母語側表現の検討に割ける時間が限られてしまい、宝の山がそこに横たわっているとわかってはいても、青空文庫の参照を励行するのは正直難しかった⁽⁴⁾。

そうした状況が一変したのが2013年だった。

3. 青空 WING がやってきた

少々時間を遡る。

紙の辞書全盛時代に翻訳、それも基本的に理系の翻訳の仕事をしていたころは、机の前の書棚のみならず、手の届く限りが辞書の海だった。英和辞典、和英辞典、国語辞典、英英辞典などのいわゆる語学辞書だけでなく、理化学辞典、生物学辞典をはじめとする英日双方の専門辞書もたくさん使うから、辞書引きの作業はちょっとした筋肉労働だった。

90年代になって、CD-ROM形式で辞書が販売されるようになると、多数の辞書をパソコンのハードディスク・ドライブに保存しておいて、辞書ブラウザ(辞書閲覧用のソフトウェア)で、串刺し検索(複数の辞書を一度に検索して、その検索結果を行き来しながら読むような検索)を行うのが普通になった⁽⁵⁾。フリーウェア、シェアウェアの作者さんたちが優れた辞書ブラウザを作ってくださったおかげである⁽⁶⁾。

当時CD-ROMのかたちで販売された電子辞書の多くが採用していたのが、EPWINGという電子辞書に特化したデータ形式だった。言い換えると、他の形式の辞書データや、青空文庫のような辞書用データではないものも、EPWING形式に変換・加工することさえできれば、EPWING用の辞書ブラウザを使った串刺し検索が可能になる。ただし、これはあくまでも理屈の上での話で、そう簡単ではない。

そして2013年、「青空 WING」によって、青空文庫が辞書ブラウザで検索できるようになった。もう少しきちんと書くと、青空文庫のテキストデータをEPWING形式に加工した「青空 WING」を作ってくださった作者さんがおられ、辞書ブラウザで「青空 WING」を引くことができるようになったということだ。青空文庫のテキストデータは、辞書データではないわけで、それを、どのように実際に使いやすかたちに加工するかというのは、一律に決まるような機械的作業ではないし、「青空 WING」は、現在もどんどん進化中である⁽⁷⁾。

かくして、辞書を引きながら、青空文庫で多くの用例を確認し、それからまた辞書に戻るといった作業が楽に行えるようになった。

4. 辞書の向こう側

通常の国語辞典には、ごくごく大ざっぱに分けて、3種類くらいあるように思う⁽⁸⁾。

一つは、収録語彙数が多く、新語や百科事典の記載も数多く含まれるような大型辞典、もう一つは、すでに知っている語彙の表記を確認するのに適していたり、辞書執筆者の顔が見えたりするような辞典で、こちらは基本的に小型である。

三つ目が、収録語彙数は第一のタイプの辞書ほど多くないものの、丁寧な語義分類を特徴とし、典型的用例や類義語の記載も充実したタイプの辞書で、その筆頭は『学研国語大辞典』だろう。用例や類義語の記載は別として、『大辞泉』(小学館)、『三省堂国語辞典』などもこのタイプの辞書だといえる。コーパスとしての青空文庫と組み合わせて日常的に使用するのに適しているのは、このタイプの辞書だと思う。

『学研国語大辞典』の場合、用例の多くが出典明示のうえで近代文学から採られている。一例を挙げれば、「満面」を引くと出てくる用例二つのうち片方は、『蜜柑』(芥川龍之介)の「この煙を満面に浴びせられたおかげで、殆息もつけない程咳(せ)きこまなければならなかった」というものである(もう一つは「満面に笑みをたたえて」のような表情に関する用例)。これでも、この語彙の含意の大枠はわかる。

しかし、青空 WING で、『蜜柑』に実際にあたってみれば(図1)、これが「元来咽喉のどを害していた私」の乗った汽車がトンネルに入った際の描写であることがわかるし、「火の粉が降るやうに満面に吹き付けて」(田山花袋『重右衛門の最後』)といった他の用例も簡



図1 青空 WING で「満面」をひいたところ(辞書ブラウザはEBWin 4)。片方の(ここでは左)ペインに検索語の含まれる作品の段落のリストが並び、そこで選んだ段落が前後とともにもう片方の(ここでは右)ペインに表示される。

単に拾え、この表現が、実際にはからだ全体が晒されているが、そのことを顔で感じたような場面で使用されていることがわかる。さらにいえば表情に関する用例にくらべて圧倒的に使用頻度が低いことや、「帆を満面にはらませた船」、「爽やかな春風を満面にはらんだ椎の樹の梢」といった顔以外の場面でも使用可能なこともわかる。

「満面」の例は、辞書に出典明示の用例が載っていた例であるが⁽⁹⁾、辞書の用例を経由せずに直接検索する場合も基本は同じで、複数の用例を見わたした後、先人の知の集積としての辞書の記載を再度確認しておくことになるはずだ。なお、自分の場合は翻訳者でもあるので、日本語の用例に関しては、国語辞典だけでなく和英辞典も参照して言語間の交通をつけているが、研究社の『新和英大辞典』の記載は、歴代の国語辞典に影響を与え続けてきただけあって、青空文庫の用例の広がりによく符合する。青空文庫の用例を通じて、国語辞典と和英辞典が結ばれているような具合だ。

なお、辞書ブラウザを使わずに、ウェブ上の青空文庫を検索しても、ある程度似たような作業を行うことはできる⁽¹⁰⁾。

《辞書+青空文庫》の組み合わせということで以上に述べてきたような使い方は、語義を示す辞書の奥行きを広げるような使い方かもしれない。

5. コロケーション辞書への広がり

青空文庫の辞書利用には、少しちがったかたちのものもある。その一つが、格関係などの文法事象を利用したコロケーション辞書としての利用だろう。

コロケーション辞典を過去に遡れば、勝俣銓吉郎の『英和活用大辞典』（研究社、1939）に行き着く。『新英和活用大辞典』（研究社、1958）の序文には、「語義を示すのではなくて、語が他の語と慣習的に結合して一つの表現単位をなすその姿を広く採集し、これを文法的に排列（ママ）したもので、その狙いは英語活動態（English in action）を展示しようとするにある」と書かれている。

日本語のコロケーション辞典としては、書籍版の小内一の『てにをは辞典』（三省堂、2010）の利用者が多い。この辞書は、手作業も介して大変丁寧に作られており、それぞれの語彙の「～を」などのかたちが実際にどのように使われているかを、共起語を手がかりとして知ることができる。この辞書は、電子辞書化の要望が群を抜いて高い辞書でもある。

ウェブ上のコロケーション辞書としては、国立国語研究所が構築した『現代日本語書き言葉均衡コーパス』（Balanced Corpus of Contemporary Written Japanese: BCCWJ）を利用して、Lago 言語研究所が

同研究所とともに開発したNLB (NINJAL-LWP for BCCWJ) があり、これも大変使いやすく、数十字～百字程度の用例を確認できる。この辞書は、通常の全文検索型インタフェースである『少納言』と組みあわせて利用してもよい⁽¹¹⁾。

青空文庫に関しても、辞書ブラウザと組みあわせてのコロケーション型辞典として利用するためのデータが公開されており、現在も着々と改良が重ねられている⁽¹²⁾。

6. さらなる表現検索の可能性

青空文庫の辞書ブラウザ経由での利用には、さらに別の展開もありうる。

たとえば、辞書ブラウザを使うと、辞典や事典の記載も検索対象として使える。辞典や事典の記載というのは、よく推敲された安定感のある文章で、『岩波理化学辞典』、『岩波生物学辞典』のような専門辞書のみならず、『平凡社大百科事典』や『小学館百科全書』などの記載を青空文庫と組みあわせて利用できることのメリットは大きい。

辞書ブラウザによっては、句読点も検索できるので、それらを利用した文頭検索、文末検索、文中切れ目検索（例：「。すなわち」、「のである。」、「すなわち」）も可能である⁽¹³⁾。青空 WING の場合は、改行検索が可能な段落開始・終了記号（↑↓）付きのバージョンも用意されているので、段落冒頭検索や段落末検索（例：「↑すなわち」、「のである↓」）もでき、語彙を検索する際にこうした検索を組みあわせることもできる。

かくして、接続表現や文末表現などに狙いをしぼった表現検索を、青空文庫だけでなく、辞書や辞典の記載も含めて行うことが可能になっている。執筆の傍らで、こうした幅広い表現検索を手軽に行えることの意味は大きい。

文章の執筆という作業は、執筆内容だけではなく表現においても「巨人の肩」上の作業だといえる。先人による整理の結果たる「辞書」と、先人の文章という「現場」を自在に行き来しつつ文章を執筆できる環境が整いつつあることは意義深い。先人の文章を電子テキストとして蓄積してきてくださった皆さん、そうした資産を辞書と行き来しつつ利用できるようにしてくださった皆さんに心から感謝したい。

- (1) 高橋太郎．言語の記述にとって用例とは何か．国文学と鑑賞．1989, (1), p. 10-15.
- (2) 高橋太郎．文法資料としての文学作品：文法形式のゆたかさの面から．国文学と鑑賞．1988, (7), p. 6-18.
- (3) 筆者の記憶による。
- (4) ウェブ空間全体をコーパスとして表現検索を行うことも可能ではあるが、検索対象の信頼性にばらつきがあり、検索ツールである Google が完全一致から曖昧検索に軸足を移し

- たこと等の問題がある。
- (5) CD-ROM 形式の辞書等については、
翻訳者の薦める辞書・資料. 2014-04-15.
<http://nest.s194.xrea.com/lingua/>, (参照 2014-04-20).
On the Backstage: 翻訳者のための情報源. 2014-03-25.
<http://home.att.ne.jp/blue/onback/>, (参照 2014-04-20).
などが詳しい。
 - (6) DDWin, Jamming, EBWin, Logophile など。
 - (7) 青空 WING. 2014-04-19.
<http://aozorawing.sourceforge.jp/>, (参照 2014-04-20).
 - (8) 小学館『日本国語大辞典』は、現代語のみの辞典ではなく、
語釈掲載順も頻度順でなく時代順、用例も初出重視と性格
が異なるので、ここではとりあげない。
 - (9) プリンストン大学の語釈付きシソーラス WordNet (英語)
を EPWING 化し、ブラウンコーパスの用例も表示可能と
した「Brown Corpus 付き Princeton WordNet 3.1」(青空
WING と同じ作者の方の作品)では、語釈の画面から、コー
パスの元祖とされるブラウンコーパスの当該用例部分を見
ることができる。
WordNet EPWING. 2013-11-24.
<http://wordnetepwing.sourceforge.jp/>, (参照 2014-04-20).
また、京都大学のライフサイエンス辞書プロジェクトによる
「ライフサイエンス辞書オンラインサービス」は、コロ
ケーション辞書機能(「英語共起表現検索」)を備えており、
薬学系ジャーナルのコーパスの検索結果を表示させること
もできるようになっている。
ライフサイエンス辞書プロジェクト.
<http://lsd.pharm.kyoto-u.ac.jp/ja/index.html>, (参照 2014-04-
20).
これらのケースは、辞書とコーパスの一体性が青空 WING
よりさらに高い。
 - (10) 青空文庫.
<http://www.aozora.gr.jp/>, (参照 2014-04-20).
の画面右上の窓、
えあ草紙・青空図書館 | 青空文庫の図書館.
<http://www.satokazzz.com/books/>, (参照 2014-04-20).
での全文検索など。
 - (11) NINJAL-LWP for BCCWJ (NLB).
<http://nlb.ninjal.ac.jp/>, (参照 2014-04-20).
少納言. <http://www.kotonoha.gr.jp/shonagon/>, (参照 2014-
04-20).
 - (12) 上記青空 WING のサイトには、青空文庫のデータをコロケー
ション辞書として使用するための EPWING 形式のデータ
(通称「青空てにをは辞典」)も置かれている。この形態は
進捗が著しく、終止形縮約版と相互情報量順の試作品も公
開されている。
"aozorawing". Sourceforge.jp.
<http://sourceforge.jp/projects/aozorawing/releases/60423>,
(参照 2014-04-20).
筆者は、書籍版『てにをは辞典』と組み合わせて使用して
いる。
 - (13) 句読点を検索する場合に使用する辞書ブラウザは、EBWin
がよい。

[受理: 2014-05-13]

Takahashi Sakino.

Beyond Dictionary: From Dictionary to Corpus and
Back.

CA1822 XXXXXXXXXX 大学間共同利用言語・文化図書館 (BULAC) —欧州随一の言語・文化図書館—

Clotilde Monteiro^{**}
(翻訳 本間渚沙)[†]

1. ドキュメンテーション・多言語教育のための新たな 施設

大学間共同利用言語・文化図書館 (BULAC)⁽¹⁾は
2011年12月12日、建築家イヴ・リオンの設計による
新たな建物内に開館した。ここにはフランス国立東
洋言語文化研究所 (INALCO)⁽²⁾も同居しており、パ
リ13区のフランス国立図書館からほど近くの場所
にある。

BULAC を設立するにあたっては、旧東洋語大学共
同利用図書館 (BIULO) と、4つの大学 (パリ第1大
学、パリ第3大学、パリ第4大学、パリ第7大学)⁽³⁾、
そして5つの高等教育・研究機関 (高等研究実習院、
社会科学高等研究院、INALCO、フランス極東学院、
フランス国立科学研究センター) が連携した。その設
立により、これまでパリ及びその近郊14か所に分散
していたコレクションが一か所に集結することとなっ
た。そのコレクションの一部には今まで一般に公開さ
れていなかったものもあった。BULAC の蔵書が扱う
範囲は、バルカン半島、中央・東欧、マグレブ、中近
東、中央アジア、アフリカ、アジア、オセアニア諸国
といった文化圏から、ネイティブアメリカン文明 (南
北アメリカ、グリーンランド) にまで及んでいる。つ
まり西欧の文明・言語や西欧に由来するものを除いた
全世界を対象としているのである。

構想期間中⁽⁴⁾、司書や教員、研究者らは、その各文
化圏について、資料収集方針の共有や書誌情報の交換



BULAC 外観

© Grégoire Maisonneuve / BULAC

^{**} Bibliothèque universitaire des langues et civilisations
[†] 収集書誌部逐次刊行物・特別資料課