

日本語例文バンク科研 (Jreibun) 第1回公開研究会報告

— 日本語学習ツールに使用可能な良質な例文をオープンデータで公開する —

鈴木智美・清水由貴子・中村彰・加藤恵梨

キーワード：「Jreibun」(日本語例文バンク)、例文データベース、学習ツール、辞書アプリ、オープンデータ

1. 本稿の目的および研究の背景

本稿では、日本学術振興会科学研究補助金の助成による「辞書サイト・アプリ開発に資する質の高い日本語例文バンクの構築とその応用研究」(基盤研究(B), 令和3年度～6年度 課題番号: 21H00535, 研究代表者: 鈴木智美, 研究分担者: 清水由貴子, 中村彰, 加藤恵梨, 伊達宏子, 藤村知子, 望月源) について、その概要を記すとともに、2021年7月に行われた第1回公開研究会について報告する。以下、本稿では、同研究プロジェクトを「日本語例文バンク」(Jreibun) プロジェクトと呼ぶこととする。

この研究プロジェクトは、日本語教育の専門家と、情報工学・教育工学の専門家とがチームを組み、日本国内外において広く日本語学習のために開発される辞書サイト・アプリ等、学習ツールの内容的質の向上に資することのできる、教育的観点から見た質の高い日本語例文データベース(英訳付き、以下「例文バンク」)を構築することを目的とする。作成した例文バンクには、「クリエイティブ・コモンズ・ライセンス」によるライセンス「CC BY-NC」(著作権者を表示した上で非営利目的で利用可能)を付与し、日本国内外におけるアプリ・サイト等のツール開発者が活用することのできるオープンデータとして公開する。さらに、この例文バンクによる例文を、実際に世界の日本語学習者に広く使用されている辞書サイトの1つ¹に、その開発・管理者の協力を得て掲載し、同サイトを通じユーザーからのフィードバックを得ることとしている。

2021年度は、この研究プロジェクトの開始年度であり、第1回の研究会を公開形式で行い、研究の目的と意義とを、プロジェクトメンバー(本研究の代表者、分担者、協力者)のみならず、関心のある人たちと広く共有することを目指し、今後の活動への理解や協力等をあわせて呼びかけることにした。

2. 公開研究会の概要

公開研究会の概要、および当日のプログラムは以下の表1、表2の通りである。2020年より続いているCOVID-19(新型コロナウイルス感染症)拡大の状況が落ち着けば、対面形式にて研究会を行うことも考えたが、東京都の感染状況が先行きの読めない状況であったため、6月半ばには、本研究会はオンライン形式で開催することを決定し、準備を進めることとした²。

¹ 「jisho」(<https://jisho.org>)

² 研究会のお知らせ(広報ポスター)を、本稿末尾に資料として掲示しておく。

表1 日本語例文バンク科研 第1回公開研究会の概要

テーマ	「例文バンク」プロジェクトの目指すもの —良質な例文をオープンデータで—
日時	2021年7月18日(日) 13:00～15:15
方式	ミーティングツール (Zoom) を利用したオンライン形式
対象者	一般公開
申し込み・参加人数	当日参加者人数 / 87名 (事前申し込み人数 / 115名)

約1か月前の参加申し込み開始から申し込み締め切りまでの約3週間の間に、計115名からの参加登録があり、研究会前日には、申込者全員宛に、資料(スライド資料PDF)のダウンロード先、および当日のミーティング参加URLを通知した。当日の参加者は計87名となり、日本語教育関係者(大学、日本語学校等の教師)、ウェブサイト等のツール開発者、研究者、辞書編集者(出版社)、大学院生、その他一般の方々と、日本国内外からの参加があった。対面形式で研究会を行うよりも、海外を含めた様々な地域から多くの方々に参加してもらうことができ、結果的にオンライン形式による開催の利点が活かされた形となった。

表2 日本語例文バンク科研 第1回公開研究会プログラム

【第一部】日本語学習ツール(アプリ・サイト)に求められるもの	13:00～13:45
<開会挨拶・進行の説明>	
1. 世界の日本語学習者のツール使用状況	鈴木 智美 (東京外国語大学)
2. 日本語辞書アプリ「imiwa?」の利便性と落とし穴	中村 彰 (東京外国語大学)
【第二部】よりよい例文作成を目指して: よい例文を作るために必要なこと	13:55～14:55
1. 「例文バンク」プロジェクトで目指すもの: 教科書・辞書作成の実体験をふまえて	鈴木 智美 (東京外国語大学)
2. ツールを使った例文作成 (1): 例文作成を合理的に考える	清水 由貴子 (聖心女子大学)
3. ツールを使った例文作成 (2): 評価を表す形容詞を用いた例文作りを考える	加藤 恵梨 (大手前大学)
【コメンテーター】砂川有里子 (筑波大学名誉教授)	
<閉会挨拶>	～15:15

第一部では、本研究の発想に至った経緯、および現状の日本語学習ツールに見られる問題点等を示し、第二部では、本研究の目指すところ、および具体的な例文作成の手法を解説し、参加者との意識の共有を目指した。1つの発表は、質疑応答を含めて20分に設定した。質問がある場合、参加者はチャット機能を使って質問事項を書き込み、発表者が適宜取り上げ、回答する方式とした。最後に、研究会全体へのコメントを、学習者用の辞書開発等の研究に長く携わる砂川有里子氏(筑波大学名誉教授)に依頼した。また、参加者の方には、記名式にて事後アンケートへの回答協力を呼びかけた³。

³ 事後アンケートを記名式としたのは、アンケートを通じて、研究協力者としてプロジェクトに参加可能である旨を表明してくれる方があった場合に、連絡をとることを可能な形とするためであった。

以下、第3節～第4節にて、研究会の各セッションの内容を順に報告する。

3. 第一部：日本語学習ツール（アプリ・サイト）に求められるもの

研究会第一部では、本研究の着想を得るに至った経緯を述べ、学習ツールの中で、特に辞書アプリ等に使われている例文に見られる問題点を具体的に示し、参加者と問題意識の共有を図ることにした。

3.1 世界の日本語学習者のツール使用状況

ここでは、まず、本研究会の実施メンバーが中心となりこれまで行ってきた研究の成果に触れながら、本研究プロジェクトの発想へとつながった背景事情について述べることにした。本研究へとつながる直近の研究成果は、「日本語学習者の学習ツール使用状況の解明と教師の教育支援リテラシーを結ぶ総合的研究」（2017～2019年度、日本学術振興会学術研究助成金 基盤研究（C）、課題番号：17K02842、研究代表者：鈴木智美、研究分担者：藤村知子、中村彰、清水由貴子、渋谷博子）によるものである⁴。以下、当セッションにおける報告内容は、その研究成果（鈴木他2018, 2019, 2020）をふまえたものである。

今、日本語学習者をとりまく学習環境は、高度な情報化社会の発展の中で、特に2010年代以降に大きく変化が見られるようになってきている。その変化はごくおおまかに言えば、いわゆるアナログからデジタルへの変化ということになるだろう。2000年代初め頃、教育現場で使用される主たる教科書・教材は印刷された紙媒体のものであり、学習者は、言葉の意味や漢字の読み書き等を調べたりする際には、書籍タイプの辞書に代わって、当時主流となりつつあった電子辞書（辞書専用器）を使うというスタイルが見られた。しかし、2010年前後頃からスマートフォンの使用が急激に拡大するにともない、スマートフォン上の様々なアプリが日本語学習上においても急速かつ活発に使われるようになり、同時に、ウェブ上でも、各種のオンライン教材や、自律学習をサポートするサイトなどが次々に生まれてきた。

このようなツールの変化にともない、学習者の学習スタイルのあり方も、また教師側の教育活動の設計のあり方も、現在は、教育全体が大きく変化している時代とも言えると思われる。そのような中で、学習者は、日本語を学ぶ際に具体的にどのようなツールを日常的に使っているのか、その現状を把握し、それを教師たちの間で共有し、時代の変化をより明確に認識する必要があると考えたことが、研究の始まりである。

このような問題意識のもと、上記研究では、2017年から2018年にかけて、学習者を対象に学習ツール使用に関するオンラインアンケート調査を行い（一部2016年末に実施済みを含む）、日本の大学（東京外国語大学）で日本語を学ぶ留学生222名、海外6つの国・地域（セルビア、タイ、中国（香港）、オランダ、エジプト、英国）の大学で日本語を学ぶ学習者387名、計609名からの回答を得た。回答者の国籍は、全70の国・地域にわたった。

各種ツールについて、どのくらいよく使っているかを5段階評価で問うた結果、スマートフォン等のアプリは、「非常によく使う」との回答が全体の56%と半数以上を占めるという顕著な結果が出た。「よく使う」（19%）と合わせると回答者の75%がスマートフォンの各種アプリをよく

⁴ 同研究の報告書は、鈴木智美（編）（2020）。

使っている。ウェブサイトは、「非常によく使う」から「あまり使わない」まで、使用頻度にはややばらつきが見られたが、「非常によく使う」と「よく使う」の回答をあわせると、やはり約半数が日本語学習の際にウェブサイトをよく使っているとの結果となった。よく使うアプリやウェブサイトの名前を挙げてもらったところ、計124種のアプリ（延べ742回答）、および計98種のサイト名（延べ667回答）が挙げられた。2000年代初めに使用者数の多く見られた電子辞書（辞書専用器）については、この2017～2018年の調査では、「所有していない」との回答が7割以上を占め、辞書専用器としての電子辞書のユーザーは、やはり減少傾向にあることがわかった。

さらに、鈴木他（2020）では、回答数の多かった辞書アプリ4種⁵を実際に試用した結果が述べられている。辞書アプリは、手軽に言葉が検索でき、知りたい語の意味や、漢字表記、また動詞の各活用形の情報なども手早く得られる点で、非常に使いやすく、便利なツールである⁶。一方で、誤訳や、不自然な訳が見られることもしばしば指摘されている。

しかし、実際に試用してみると、誤訳以上に改善の余地があるのではないかと考えられたのは、アプリに表示される例文である。語学学習用に特に形式や内容が整えられているとは考えにくい、いわば生のままに近い文が、前後の文脈のないまま示されたり、あるいは、類似した単純な文ばかりがいくつも表示されたりすることがあることがわかった。アプリ等における例文は、その辞書アプリが依拠している何らかのデータベースに基づいて表示されることとなっている。そのため、例文の文脈を整えたり、形式・意味の両面において語学学習に役立つ情報を盛り込んだりしたいと考える場合には、元の例文データベースを改良していくことが必要となる。しかし、現行の辞書アプリやサイト開発において使用可能な例文データベースはごく限定的なものとなっており、それは特に日本語教育の観点からの精査を経た例文のデータというわけではないようである。

このような現状を見る中で、研究グループでは、辞書の例文として機能するよりよい例文を日本語教育の観点から作成し、これをデータベースとして整えた上で、アプリやウェブサイト開発に利用可能なようにオープンデータとして公開するという考えを、次の研究課題として考えることとなった。

3.2 日本語辞書アプリ「imiwa?」の利便性と落とし穴

3.1節で述べられたように、鈴木他（2020）の調査によると、日本語学習者の半数以上が、スマートフォン等のアプリを非常によく使うということである。ここでは、その中で使用者が多かったアプリの1つの「imiwa?」⁷を取り上げ、具体的にその利便性と落とし穴について見る。なお、以下の内容は、本セクションの担当者が鈴木他（2020）において実際にアプリを試用し、得られた結果をもとに報告するものである。

「imiwa?」とはiPhone・iPad対応の無料の辞書アプリである。Jim Breen氏による日本語辞書プロジェクト「EDICT」⁸に基づき、オーストラリア・モナシュ大学のElectronic Dictionary Research

⁵ 海外にて回答数上位であった辞書アプリの中から、調査を行った6つの国・地域のうち5つ以上の国・地域で名前の挙がっていたアプリ4種を取り上げている。

⁶ カメラ機能により対象テキストを読み込み、形態素解析を経て即時逐語翻訳が表示される点など、回答者からは種々の機能が便利な点として指摘されている。

⁷ 「imiwa?」の基本情報については以下のサイトを参照。<http://www.imiwaapp.com>

⁸ EDICTの基本情報については以下のサイトを参照。<http://www.edr.org/jmdict/edict.html>

and Development Group⁹が開発した日本語 - 複数言語対応辞書「JMDict」プロジェクト¹⁰のデータベースファイルを利用している。2021年7月21日の時点で、利用者の評価は5点満点で4.7と非常に高い評価である¹¹。その時点でのバージョンは4.2.1で、更新は定期的に行われているようである。

このアプリが日本語学習者にとって便利な点として、以下の各点が挙げられる。

- ・無料である
- ・英和・和英辞典であるだけでなく、一部はフランス語、ドイツ語、スペイン語、ロシア語、イタリア語、ポルトガル語にも対応している
- ・オフラインで使用可能である
- ・例文とその多言語訳がある
- ・Analyserという機能があり、日本語の文をコピーすると、自動的に形態素解析を行い、漢字にふりがなが振られ、それぞれの単語の他言語訳が表示される
- ・漢字辞典の機能もあり、書き順、日本語能力試験の級別対応の漢字リスト、学年別学習漢字の漢字リストがある
- ・書籍形態の辞書、電子辞書と比較すると、新語、時事用語への対応が早い

「imiwa?」には上記のような利点がある反面、辞書の“要”でもある例文で、種々の誤りが多く見られるという大きな欠点もある。「imiwa?」の例文は、「Tanaka Corpus」¹²及び「Tatoeba:多言語例文プロジェクト」¹³のものが使われている。「Tanaka Corpus」というのは、大学生に英文を日本語に翻訳させたものを集めたもので、15万弱の文が英語と日本語で書かれている。したがって、その日本語は日本語母語話者が自然に話したり書いたりしたものではなく、なかには誤訳もある¹⁴。「Tatoeba:多言語例文プロジェクト」は外国語学習者向けの例文を掲載しているオンライン協同データベースである。

寺嶋・板井(2019)は「imiwa?」と「Google Translate」を比較し、「imiwa?」の方が「Google Translate」よりも学習者に勧められるが、例文には以下のような7つの問題点があることを指摘している。

- ・文脈がわかりにくい
- ・文が長すぎる
- ・文が専門的すぎる

⁹ Electronic Dictionary Research and Development Groupの基本情報については以下のサイトを参照。<http://www.edrdg.org>

¹⁰ JMDictプロジェクトの基本情報については以下のサイトを参照。https://www.edrdg.org/jmdict/j_jmdict.html

¹¹ Apple App Storeのユーザーレビューによる。

¹² 「Tanaka Corpus」については以下のサイトを参照。このサイト内のリンクから、「Tanaka Corpus」をダウンロードすることが可能。http://www.edrdg.org/wiki/index.php/Tanaka_Corpus

¹³ 「Tatoeba:多言語例文プロジェクト」については以下のサイトを参照。<https://tatoeba.org/ja>

¹⁴ 「Tanaka Corpus」の問題点については以下のサイトで簡潔にまとめられている。<https://www.manythings.org/corpus/warning.html>

- ・不自然なコロケーションとなっている
- ・文法的な誤りがある
- ・漢字の読み方の誤りで意味が成立しない
- ・見出し語と例文が合わない

実際に「imiwa?」を試用してみると、上記の問題点の中で、漢字のふりがなの誤りは特に多く見られる。漢字が弱い学習者にとっては大きな問題である。また、寺嶋・板井（2019）が指摘した上記の問題点に加え、次のような問題点もある。

- ・例文に偏りがある（同じような例文が続く）
- ・例文は「Certified Example」（質が保証された例文）とその他の例文に分かれるが、前者にも誤りが見られる
- ・例文の訳に誤りがある

「imiwa?」は確かに利便性は高いアプリだが、以上のように、深刻な問題点も多いことがわかる。「Tanaka Corpus」や「Tatoeba：多言語例文プロジェクト」のようなコーパスやデータベースは、利用に関して著作権等の問題がなく、アプリ開発者にとっては、利用しやすいものであるが、そこから得られた日本語の文をそのまま学習者向け辞書に使うと問題が出てくる。したがって、吟味された、質のよい例文を効率的に提示する必要性は高い。

4. 第二部：よりよい例文作成を目指して：よい例文を作るために必要なこと

第一部では、現在、日本語学習者の間では、各種アプリの使用が活発に行われているが、日本語教育の観点から見ると、現行の辞書アプリに提示される例文には問題が見られることがあることを指摘し、そのため、本研究では、質のよい日本語例文のデータベースを作成し、各種ツール開発者が利用可能なオープンデータの形で公開することを目指すこととなったという経緯を述べた。

第二部では、「日本語例文バンク」(Jreibun) プロジェクトの目指す内容について、より詳しく説明し、具体的な例文作成の手法をいくつか紹介することで、参加者との意識の共有をはかることを目指した。

4.1 「日本語例文バンク」(Jreibun) プロジェクトで目指すもの：教科書・辞書作成の実体験をふまえて

4.1.1 「日本語例文バンク」の基本設計方針

第一部で見たように、現在、日本語学習者がよく使用する辞書アプリについては、その例文が、時に文が複雑でわかりにくい、使用文脈が個別的、特殊過ぎる場合がある、辞書としては、その項目の形式的な情報（前後に接する要素や、その形式自体の形の変化）、あるいは共起する典型的表現などについて、情報が十分に得られるようにしたほうがいいのではないかと、あるいは例文は数多く表示されるが、単純で似たような文ばかりが並ぶのはあまり意味がないのではないかなど、いくつかの問題点が指摘できることがわかってきた。確かに、生に近い例文というのは、そこで取り上げられている話題などは興味深く、言語のいわば生きた姿を伝えるものとは考えられ

るが、しかし、辞書に示す例文として考えると、その対象となる表現の持つ形式的あるいは意味的な特徴および多様性（バリエーション）が、明瞭に、かつ十分な前後文脈を伴って表示されることも必要なのではないかと思われる。

そこで、この「日本語例文バンク」(Jreibun) プロジェクトでは、アプリやウェブサイト開発に使用可能な1つの選択肢として、日本語教育の観点から見て、内容的な質を整えた日英対訳の例文データベースを作成し、それを、オープンデータとして公開することとした。日本語教育の専門家と、情報工学・教育工学の専門家がチームを組み、例文のデータベースは、ツール開発に応用可能なファイル形式に整え、公開する。各種ツールの開発者は、著作権者を表示した上で、非営利目的に限り、このデータベースを使用することが可能となる。4年間の開発計画の中では、数回に分けてデータを公開することとし、学習者の間で広く使用されている辞書サイト「jisho」の開発・運営責任者からの助言も受けつつ、実際に本例文バンクの例文を同サイトに掲載することとしている。

このプロジェクトは、何らかのオリジナルな辞書を新たに作成しようとするものではなく、独自の学習サイトや辞書サイトなどを新規に立ち上げるものでもない。本プロジェクトは、あくまで、そのようなツールの開発において使用可能な、例文のデータベースを作成することを目的としている。これまで既に、日本語教育関連の教材や学習サイトは多く作成されてきているが、それらは基本的に、その教材あるいはサイトの中で完結するものとなっており、そこで使われている日本語の例文が、他のアプリやウェブサイトの開発で使用可能な素材として公開されたものとなっているわけではない。本研究プロジェクトは、ツール開発側のニーズに立ってこのような点を根本的に解決し、コンテンツとなる例文はオープンデータとして提供する。

また、通常日本語教科書等とは異なり、仮に、何らかの文法項目をきっかけとして例文を作成する場合でも、日本語教育において言われるその文法項目の学習レベルと、同じく語彙レベルとを必ずしも結びつける必要はない。アプリやウェブサイトでは、データベースから取り込まれる例文データは、基本的に形態素解析を経た上で使用される。書籍形態の辞書のように、1つの例文がある1つの項目のみに対応しているということではなく、ツール開発側の設定次第で、1つの例文は、その文に含まれる様々な語・表現からアクセス可能な素材として活用され得る。したがって、例えば、初級の文型として導入されるような項目でも、現実的なその文型の使用状況を考えた場合、必ずしもすべて初級の語彙を用いて例文を作成しなければならないということではなく、むしろ、多様な語彙と組み合わせることを考えている。

また、この例文バンクは、まずは4年間の研究プロジェクトの中で作成していくが、その後も、例文の改良・差し替え・追加等の作業および管理を継続的に行うことで、成長し続けていくデータベースとする。データベースには、作成の意図、助成研究費、助成期間のほか、作成に関わったメンバーを明記し、開発の趣意に賛同が得られ、現在の本研究グループの外から、著作権の問題がなく公開可能な例文データの提供が得られたならば、その作成者・協力者を明記することで、データを組み込んでいくことを可能とする。

実際に例文を作成する際のきっかけとなる項目については、「日本語能力試験出題基準」旧4級～1級までの語彙¹⁵、『日本語コロケーション辞典』見出し項目¹⁶、「日本学術共通語彙リスト」

¹⁵ 約8,000項目

¹⁶ 動詞、形容詞、形容動詞を合わせて約2,200項目掲載されている。

見出し項目¹⁷、初級～上級レベルの文法・語句等の項目（『初級日本語』『中級日本語』『上級日本語』いずれも東京外国語大学留学生日本語教育センター編著）¹⁸等を参考として、相互の重複を削除し、約10,000項目を抽出しており、これに基づいて例文を作成していくこととしている。

4.1.2 例文作成の観点

実際にどのような点に注意しながら例文を作成していくかということについては、まず、対象となる項目について、その形式的な多様性を示す情報が必要である。例えば、動詞項目を中心に例文を作成する場合ならば、とり得る形の多様性を、使用頻度等の情報を参考にしながら例文として示していくことになるだろう。いわゆる「て形」や、否定の「ない形」、意志形、受身形、使役形、使役受身形などの各活用形のほか、「～ざるを得ない」「～たまま」「～しかない」など、特徴的に用いられる何らかの形があれば、そのような形も適宜例文として示していく。

また、語は多義的であるため、基本的意味だけではなく、頻度の高い派生的意味についても例文を示す必要がある。述語となる要素には、それがどのような補語をとるかなど、文中で共起する語の情報も重要である。例えば、「判断する」という語であれば、「何を」、「どのように」、「どう」判断するのかという情報は、これと共起する語の特徴として、例えば「{状況／可能性}を判断する」、「{慎重に／冷静に／瞬時に／自分で／…た上で}判断する」、「{必要（である）と／～ものと}判断する」などの形として示されることになるだろう。

また、例文においては、その的確な使用文脈を考える必要がある。例えば、「協力する」「協議する」「維持する」「対応する」などの動詞を例に考えてみると、これらは実際に、どんな状況・場面で、どんなことに対して、誰が、どのような行動をとる場合に使われるのか、文脈によってそれをなるべくわかりやすく示す必要があるだろう。例えば、「関係各国で協議し、貿易に関する新たな協定を結んだ。」「最近は、よく運動しているので、体調がいい。このまま、現状を維持したい。」「電車で傘を忘れて、駅員に相談したら、親切に対応してくれたのですぐに見つかった。」などの例文¹⁹で、それを示すことが考えられる。

ほかにも、会話表現や接続表現等においては、複数の文を用いて例を示したり、文化的な情報は積極的に取り入れ、かつ社会・倫理面において適切な内容であることに配慮するなどを、基本的な作成方針としている。

4.2 ツールを使った例文作成 (1)：例文作成を合理的に考える

ここでは、データベースとしてコーパスを使い、多くの実例から様々なヒントを得ながら、例文を作成する方法を紹介する。

中俣（編）（2017）では、中上級のアカデミックライティング等で必要となる項目を取り上げ、日本語教師のための例文作りのヒントを提示している。そこでは「典型的な例を、根拠を持って

¹⁷ 東京大学松下達彦氏作成の語彙リストで、日本語の学術テキストにおいて一般テキストよりも高いテキストカバー率を示す語とされるもの。日本語能力試験の旧1級語彙までに含まれていない、しかし、アカデミックな文脈で使用される語が約2,500項目示されている。

¹⁸ あわせて約680項目である、

¹⁹ 実際に東京外国語大学留学生日本語教育センター編著（2017）『中級日本語漢字練習帳Ⅰ』第2版の改訂作業において、作成した例文である。

示すことが必要」という立場を取り、そのためにはコーパスを利用することが有効であり、かつ、不可欠であると述べている。

確かに、コーパスを利用すれば、多くの用例を取り出すことができるが、大量の用例から、何を優先して、どう吟味して用例を選ばばいいかという判断は、容易ではない。ここでは、以下の3つの情報が必要だと考える。

- ① 直前には、どんな語が接続するのか（前接語）
- ② 前／後文脈には、どんな語が共起するのか（共起語）
- ③ どんなレジスターで使用されるのか（使用域）

これらの情報を集計して、高頻度で使われているものに着目する。また、高頻度ではないが、意味や形に特徴のある前接語・共起語も拾っていく。このようにして得られた情報を例文作りの素材・材料とし、もっとも典型的な例文を作っていく。

<使用コーパスおよび検索ツール>

コーパスは取り上げる項目によって適宜選択するが、今回は『現代日本語書き言葉均衡コーパス (BCCWJ)』を使用する。

検索ツールは、『NINJAL-LWP for BCCWJ (NLB)』と『中納言』を使用する。NLBは、語と語の結びつき（コロケーション）を見るのに便利なツールである。品詞レベルで共起パターンも示す。NLBを使って、対象の語が高頻度で現れる (A) パターン を調べる。もう一方の中納言は、対象とする語の1つ前の語を指定して調べたり、前後の文脈を取り出すことができる。また、レジスターの情報も付いている。これを用いて、対象の語の直前に接続する (B) 前接語、(C) 前の文脈に多い語、(D) 後ろの文脈に多い語、そして、その語が高頻度で現れる (E) レジスター を調べる。

以上の2種類のツールを使い、上記 (A) ~ (E) の5つの情報を取り出す。

<事例1「皮肉」>

まず、「皮肉」という語がどう使われているか、コーパスを検索し、情報を取り出す。

- (A) 「皮肉+助動詞」「皮肉+助詞」「皮肉な+名詞」といったパターン
- (B) 前接語は、「は／が／の」などの助詞、また「句点」や「読点」といった記号類
- (C) 前文脈に多い語は、「ていたのは／というのは／のだから／たが」といった複合辞、「何とも」や「痛烈な」という副詞
- (D) 後文脈は「にも／なことに／である。／な（もの／話）だ。／を（込めて）言う／が込められている／といえば皮肉だ」といった語。前文脈と後文脈を合わせると、「～ていたのは、皮肉である。／～というのは、皮肉なものだ。／～したが、皮肉にも／皮肉なことに／何とも皮肉な話だ／痛烈な皮肉が込められている／皮肉といえば皮肉である。」となる
- (E) レジスターは、書籍・雑誌・新聞

これらの語が含まれているコーパスの用例の一例を紹介する。

- ・さんざん待たされたあげく、九月に入ってからエアコンは設置されたのだが、皮肉なことに その日から気温は下がり、十月下旬の気温ですと連日天気予報が告げている。(PM22_00055)

角田光代『群像』講談社 2002)

- ・一方、この間、日本が模範としてきた欧米諸国においては、病気と食生活の関連性に対する関心の高まりから、日本型食生活を普及する動きが高まっていることは皮肉である。(PB26_00094 片野學・石川達也『食・農・からだの社会学』新曜社 2002)
- ・「ずいぶん羽振りがいいね、ダグ。イーストエンドの労働者連中が見たら、ただじゃすまないぞ」ヴァージニアは、そこに皮肉がこめられているのを感じたが、モーティマーは頓着しない。(PM41_00643 白石征・逢坂剛『IN POCKET』講談社 2004)

上記(A)～(E)の情報とコーパスの用例を参考にし、「皮肉」の例文を作成すると以下のようになる。なお、「皮肉」には大きく2つの意味があるため、意味ごとに例文を示す。

1) 想定外のことが起こる

- ① 新しいソファを買ったが、皮肉なことに玄関のドアを通らなかった。
- ② 一般家庭に育った彼女は憧れだった女優になり、すべてを手に入れた。しかし今は芸能生活に疲れ、元の暮らしに戻ることを望んでいる。なんとも、皮肉な話である。

2) 嫌味

- ③ 「かわいいね」と、皮肉を込めて言った。
- ④ 「ありがとう」という言葉には、皮肉が込められていた。

1つめは「想定外のことが起こる」という意味で、「ある出来事があって、次はこうなるだろうと想定していたことが起こらない」という文脈で使われる例である。コーパスの用例でも見られるように、良い出来事の後に悪い出来事が生じるという順を採用した。①は「皮肉なことに」という挿入句を付け、後に続く悪い出来事を予告させるような形にした。②は想定外の事態が発生したことに対して、「皮肉な話だ」「何とも皮肉である」というように文末で判断を述べる形である。

2つめは「嫌味を言う」という意味の「皮肉」の例文である。「皮肉を込めて言う」「皮肉が込められている」のように「込める」を使用し、引用符の中には、言われて嬉しい言葉(かわいいね、ありがとう等)を入れた(③、④)。

<事例2「構築」>

次に、「構築」という語がどう使われているか、コーパスを検索し、情報を取り出す。

- (A) 「構築+助詞」「名詞(の)構築」「接頭辞+構築」といったパターン
- (B) 前接語は「を/が/の」といった助詞、また接頭辞「再」も多い
- (C) 前文脈は、「については/としては/においては/ためには」といった複合辞。名詞では「システム/ネットワーク/体制/制度/平和/関係/社会/政策」といった語。辞書では、「城を構築する」という具体物「城」が例として挙がっていたが、コーパスの用例では「組織/情報網/政策/制度」など抽象的なものが上位を占めた
- (D) 後文脈は、「(構築)する/している/しなければならない」で「構築する」という動詞として使用。また、「(構築)に取り組んでいる/努めている」「(構築)を図る/目指す/進める」といった動詞とも共起。その他、「必要がある/必要となっている」というような「必

要」とも相性がいい

(E) レジスターは、白書・書籍・新聞・雑誌・国会会議録

これらの語が含まれているコーパスの用例の一例を紹介する。

- ・初動時の情報収集をより迅速にする世界最先端の災害緊急情報伝達・収集ネットワークの構築が必要不可欠である。(OW6X_00080 『消防白書』2005)
- ・核開発、処理能力を持つ五カ国の専門家を動員、核保有国が廃棄検証に関与した南アフリカの事例を念頭に徹底的な廃棄システムの構築を目指す。(PN3k_00013 『神戸新聞』2003)
- ・第一次石油危機から二十年を経た今日、エネルギー政策全般の再構築を図り、地球環境・問題の顕在化に対応しつつ、エネルギーの安定的かつ効率的な供給の確保に努めてまいります。(OM41_00005 国会会議録 1993)

上記(A)～(E)の情報とコーパスの用例を参考にし、「構築」の例文を作成すると以下のようになる。

[組織・情報網・政策・制度などを] 組み立てて、築く

- ① 通販サイトなどのシステムを構築する仕事をしている。
- ② 地域住民の生活を支えるネットワークの構築を目指す。
- ③ 今後は、地震などの自然災害の発生も視野に入れ、リスクが最小限になるような物流体制を再構築する必要がある。

コーパスの用例から、「○○を構築する」の目的語の部分には「組織・情報網・政策・制度」がよく使われていることがわかったため、これらを使った(①、②、③)。さらに後文脈では、「構築」の直後に「する」が付いた「構築する」という形(①)や、「目指す、必要がある」という語を使った(②、③)。

以上、コーパスから例文を作るのに必要な情報を取り出し、それらを適切に組み合わせることで典型的な例文が作れるということを、「皮肉」と「構築」を例に説明した。

4.3 ツールを使った例文作成(2)：評価を表す形容詞を用いた例文作りを考える

ここでは、評価を表す形容詞「だらしない」を取りあげ、「だらしない」の例文を作成する際にどのような点に気をつけたら良いかについて述べ、最後に「だらしない」の例文案を提示した。

4.3.1 例文に必要な情報

鈴木(2016:97)は、中級レベル以上の日本語学習者が抽象的な概念を表す日本語表現を的確に理解し、身につけ、使用するために、その表現を導入する際に必要となる情報として、「語の品詞や活用」「文法的・統語的な情報」「意味」「典型的なコロケーション」、さらにその表現が用いられ得る的確かつ具体的な「状況・場面」を挙げている。以下では、抽象的な概念を表す「だらしない」という表現を取りあげ、「だらしない」が用いられ得る的確かつ具体的な状況・場面を提示することを意識し、「だらしない」の例文を作成する。

4.3.2 「だらしない」の例文作り

例文作りは次の手順で行う。まず、辞書の意味および例文を検討する（手順1）。やはり辞書の記述は優れており、意味も例文も熟慮されて挙げられていると考えるからである。次に辞書の記述をもとに、コーパスで「だらしない」を検索する（手順2）。検索結果をもとに、的確かつ具体的な状況・場面を提示するために、どのような例文が適切であるかを考える（手順3）。最後に、例文を作成する（手順4）。

手順1：辞書の意味記述および例文の検討

辞書では「だらしない」の意味および例文を次のように記述している。

『新明解国語辞典 第八版』（p.969）

- ①〔生活態度・物事のやり方に〕望ましい秩序が認められない様子だ。
「金（時間）にだらしない／だらしない服装／だらしく散らかったままの部屋」
- ②期待される精神力（や肉体的な力）の強さが発揮されず、むなしく挫折したととらえられる様子だ。
「一度や二度の失敗であきらめるとはだらしない／これくらいの仕事で音をあげるなんてだらしないぞ」

『大辞林 第四版』（p.1708）

- ①（外面的に）きちんとしていない。整っていない。
「だらしく口を開けて寝ている／だらしない服装」
- ②（内面的に）節度がない。毅然としていない。しっかりしていない。
「金銭にだらしない／政府の顔色をうかがうマスコミのだらしない姿勢が問題だ」

辞書の記述から、「だらしない」には、ある人の外面について言い表す場合と、ある人の内面について言い表す場合とがあると考えられる。

手順2：コーパス検索

はじめに、「まとめて検索『KOTONOHA』（試験公開版）」によって、「だらしない」がどのコーパスで多く使われているのかを見る。その結果、用例数が一番多かったのは『現代日本語書き言葉均衡コーパス』（以下、BCCWJ）であり、用例数は354件であった。よって、BCCWJを用い、調査を行うこととする。

次に、BCCWJにおいて「だらしない」がどのような形（活用形）で用いられているかを調査する。その結果、最も多く使われていたのは連体形（142件）であり、二番目が連用形（125件）、三番目が終止形（65件）であった。それぞれについて詳細に見ていき、的確かつ具体的な状況・場面を提示するために、どのような例文が適切であるかを考える。

手順3：検索結果の考察および的確かつ具体的な状況・場面の検討

まず、「だらしない」が連体形である場合、後続する名詞で最も多いのは「の」（11件）であり、二番目が「人」（10件）、三番目が「格好」（8件）であった。評価を表す「だらしない」が用

いられる的確かつ具体的な状況・場面を提示するには、「だらしない格好」という視覚的に捉えられる表現が適していると考えられることから、「だらしない格好」を用いた例文作りを考える。BCCWJの「だらしない格好」の例には、次の(1)と(2)のようなものが見られた。

- (1) 何故、高齢者がお洒落なのか、

だらしない格好をしない、例えば人前でジャケットを脱がない、シャツやパンツにはプレスがされており、靴もしっかり磨いてある (OC09_08481 Yahoo!知恵袋 2005)

- (2) (前略) いつもの一分の隙もないでたちとは裏腹に髪には櫛を入れた様子も無く、薄く無精ひげを生やし、上着はだらしなく着崩れている。(中略) 女というのは妙な生き物で、亭主のだらしない格好は断固許さなくせに他の男の退廃ぶりには滅法弱い、というのが、短い結婚生活の中で俺が得た教訓の一つだった。

(PB49_00249 鶴田楡『ダンス・ウィズ・キャット』新風舎 2004)

(1)と(2)から「だらしない格好」は、人前でジャケットを脱ぐ、シャツやパンツにはプレスがされていない、靴もしっかり磨かれていないさまや、上着を着崩しているさまであると考えられる。「だらしない格好」はある人の外面について言い表す場合の例とする。

続いて、「だらしない」が連用形である場合、どのような動詞が後続するのかを調査する。最も多いのは「なる」(20件)、二番目が「見える」(11件)、三番目が「笑う／着崩す／開く」(3件)であった。「だらしない」が用いられる的確かつ具体的な状況・場面を提示するには、他者が一目で分かる例文が好ましいと考えられるため、「だらしなく見える」を用いた例文作りを考える。BCCWJの「だらしなく見える」の例には、次の(3)のようなものが見られた。

- (3) それに横ずわりだと、背すじが曲がって、やはりだらしなく見えるし、食事の動作もくずれたものになり、粗相も起こしやすいもの。

(OB2X_00291 千玄室『なんて美しい女性だろう!』主婦と生活社 1981)

(3)は美しい女性のふるまいについて書かれた文章である。横ずわりが「だらしなく見える」のは、背筋が曲がり、食事の動作がくずれたものになるからであると述べている。「だらしなく見える」もある人の外面について言い表す場合の例とする。

さらに、「だらしない」が終止形である場合を調査する。「だらしない」の前接語で最も多いのは「に」(10件)であり、二番目は「が」(8件)、三番目は「は」(7件)であった。また、最も多い「に」は、6件が「金にだらしない」という表現で用いられていた。BCCWJの「金にだらしない」と「[名詞]がだらしない」の例には、次の(4)と(5)のようなものが見られた。

- (4) あなたの生活にかかるお金を、他者(家族も含む)に依存しておきながら、自分の好き勝手に「浪費」しているとしたら「金にだらしない」といわれても仕方が無いでしょう。

(ID OC09_02148 Yahoo!知恵袋 2005)

- (5) 「そもそもは警察がだらしないから、こんな連中がのさばるんじゃないか。もっと警察はしっかりとこいつらを取り締めよ」 (PB59_0020 森詠『砂の時刻』光文社 2005)

(4) から「金にだらしない」は、自分の好き勝手に浪費しているさま、またそのことによって他者に迷惑をかけているさまを表し、(5) から「警察がだらしない」は警察がしっかりと仕事をしていないさまを表していると考えられる。「金にだらしない」と「[名詞]がだらしない」はある人の内面について言い表す場合の例とする。

手順4:「だらしない」の例文案

最後に、前節までの考察をもとに、的確かつ具体的な状況・場面を提示することを意識し作成した「だらしない」の例文を提示する。

1. ある人の外面について言い表す場合

- ・あの人はいつもだらしない格好をしている。人前でジャケットを着ることはめったにないし、たまに着ていても着崩れている。また、シャツやパンツにアイロンがかけられていないし、靴も磨かれておらず、汚れている。
- ・食事のとき、テーブルにひじをついたり、背筋がまがって猫背になっていたりすると、だらしく見えます。背筋をのばして良い姿勢で食べましょう。そうすると、食べこぼすこともありません。

2. ある人の内面について言い表す場合

- ・あの人は金にだらしない。自分の金が足りなくなると、友人から借り、欲しいものを買う。しかし、借りた金を返そうとしない。そのため、金を貸した友人は困っている。
- ・私は仕事が長続きしない。仕事を任されても途中で嫌になり、最後までやり遂げることができない。根性がないのだ。うちの親は「子供がだらしない」と嘆いているようだ。

5. まとめ

以上、研究会の第一部では本研究の発想に至った経緯を、これまでに行ってきたツール使用状況の調査結果、および具体的なアプリの問題点の検討を通じて示し、第二部では本研究の目指すところと、例文作成の基本的考え方、および例文作成の具体的な方法について解説し、研究会参加者との意識の共有をはかった。

研究会の最後には、筑波大学名誉教授の砂川有里子氏から総評を得た。砂川氏自身も学習者辞書の開発等、様々な研究に携わってきた経験から、今回のプロジェクトの趣旨についての共感が示され、その意義を高く評価するとのコメントが得られた。特に、使用可能な例文データベースをオープンデータ形式で公開するという点は非常に現代的で意味のあるものであり、今後、このような形で種々のオープンデータが整えられていくことが、日本語教育の発展に資するのではないかということ、そして、今後、様々な形での協同研究が行われていくことを期待したいとのコメントを得た。

また、研究会後の記名式の事後アンケートには10名の参加者が回答を寄せ、全員がこの例文バンクプロジェクトについて「とても期待している」との回答であった。研究会に参加した感想のほか、例文に関する具体的な質問などもあり、今後、実際に研究成果の例文を示す中で、順次、それらの問いに答えていきたいと考えている。

本研究が、今後の日本語教育の発展に寄与できる点としては、以下の4点が挙げられる。まず、

日本語教育と情報工学との分野を超えた協同形態で実践的な研究を進めていくものであること、次に、研究成果である例文データベースをオープンデータ形式で公開し、ツール開発者等の利用に供するという新しいコンセプトを具現化するものであること、そして、閉じたグループによる閉じた研究として研究成果を終わらせるのではなく、今後のデータベースの成長と発展を可能とするような設計としていることである。そして、何よりも日本語教育の専門家の手による質の高い例文データが新たなリソースとしてアクセス可能となり、それが実際にアプリ・ウェブサイトに活用されることとなれば、ツールを使用する学習者に対して、その学習の質の向上に貢献することができるということが期待される。

(執筆分担：1, 2, 3.1, 4.1, 5 鈴木、4.2 清水、3.2 中村、4.3 加藤)

■本研究は以下の助成を受けて行われている。

日本学術振興会科学研究補助金令和3年度～6年度基盤研究(B)「辞書サイト・アプリ開発に資する質の高い日本語例文バンクの構築とその応用研究」(課題番号：21H00535, 研究代表者：鈴木智美、研究分担者：清水由貴子、中村彰、加藤恵梨、伊達宏子、藤村知子、望月源)

■研究会開催にあたって、広報ポスターのデザインを東京外国語大学大学院生のゾーイ・アガサ・セラノさんに、当日の記録および出席者確認等の作業補助を同上田優花さん、パク・サンゴンさんをお願いした。なお、当日の提示スライドは、鈴木智美がデザインの上、テンプレートの作成をゾーイ・アガサ・セラノさんをお願いしたものである。

引用文献

- 国際交流基金・日本国際教育協会編著(2002)「日本語能力試験出題基準」【改訂版】凡人社
- 鈴木智美(2016)「抽象概念語彙を説明するための適切な導入例を考える—現場教師の授業準備に役立つための試案作成に向けて—」『東京外国語大学留学生日本語教育センター論集』第42号 pp.97-110
- 鈴木智美(編)(2020)『日本語学習者の学習ツール使用状況の解明と教師の教育支援リテラシーを結ぶ総合的研究平成29(2017)年度～平成31(2019)年度日本学術振興会学術研究助成金基盤研究(C)研究成果報告書
- 鈴木智美・清水由貴子・渋谷博子・中村彰・藤村知子(2018)「予備教育課程の国費学部留学生の学習ツール使用状況—2016～2017年度実施のアンケート調査の結果から見えるスマートフォンアプリの使用目的の多様化と学習スタイルの変化—」『東京外国語大学留学生日本語教育センター論集』第44号 pp.195-217
- 鈴木智美・清水由貴子・渋谷博子・中村彰・藤村知子(2019)「東京外国語大学全学日本語プログラムで学ぶ留学生の学習ツール使用状況—2016～2017年度実施のアンケート調査の結果と分析」『東京外国語大学留学生日本語教育センター論集』第45号 pp.221-238
- 鈴木智美・清水由貴子・中村彰・渋谷博子(2020)「海外の大学における日本語学習者のツール使用状況の解明—ICT時代における教師の教育設計リテラシーの向上を目指して—」東京外国語大学国際日本研究センター『日本語・日本学研究』第10号 pp.23-48
- 寺嶋弘道・板井芳江(2019)「スマートフォン辞書アプリについての一考察」『日本語教育方法研究会誌』Vol.25 No.2 pp.66-67
- 東京外国語大学留学生日本語教育センター編著(1998)『上級日本語』
- 東京外国語大学留学生日本語教育センター編著(2010)『初級日本語』(上・下)
- 東京外国語大学留学生日本語教育センター編著(2015)『中級日本語』(上・下)
- 東京外国語大学留学生日本語教育センター編著(2017)『中級日本語漢字練習帳Ⅰ』第2版

中俣尚己（編）（2017）『コーパスから始まる例文作り』くろしお出版
姫野昌子監修（2012）『日本語コロケーション辞典』研究社

<ウェブサイト>

「オンライン検索システム NINJAL-LWP for BCCWJ (NLB)」

<https://nlb.ninjal.ac.jp/>

「コーパス検索アプリケーション 中納言」 ver.2.4.5

<https://chunagon.ninjal.ac.jp/>


「日本学術共通語彙リスト」東京大学 松下達彦研究室

<http://www17408ui.sakura.ne.jp/tatsum/list.html>

「jisho」 <https://jisho.org>

<調査資料>

『現代日本語書き言葉均衡コーパス (BCCWJ)』 ver.2021.03 国立国語研究所



日本語例文バンク科研 第1回公開研究会


『「例文バンク」プロジェクトの目指すもの-
良質な例文をオープンデータで』

日本語学習者の使っている辞書アプリを見て、その例文をもっといいものにしたいと思ったことはありませんか。本科研では、アプリ・ウェブサイト開発に使ってもらえるよう、日本語教育の観点から見た質の高い例文バンクを作成し、オープンデータとして公開します。よい例文とは何かをいっしょに考え、作成にご協力いただける方も募集します。

2021年7月18日（日）13:00～15:15

Zoomによるオンライン方式

- **参加申込**：右記よりお申し込みください
- **申込締切**：7月10日（土）まで



*7月16日（金）を目処に、お申し込みされた方のメールアドレス宛に、資料ダウンロード先と当日のZoom情報をお知らせします。

【第一部】日本語学習ツール（アプリ・サイト）に求められるもの

1. 世界の日本語学習者のツール使用状況
鈴木智美（東京外国語大学）
2. 日本語辞書アプリ「imiwa?」の利便性と落とし穴
中村 彰（東京外国語大学）

【第二部】よりよい例文作成を目指して：よい例文を作るために必要なこと

1. 「例文バンク」プロジェクトで目指すもの：
教科書・辞書作成の実体験をふまえて
鈴木智美（東京外国語大学）
2. ツールを使った例文作成（1）：例文作成を合理的に考える
清水由貴子（聖心女子大学）
3. ツールを使った例文作成（2）：
評価を表す形容詞を用いた例文作りを考える
加藤恵梨（大手前大学）

＜コメンテーター＞ 砂川有里子（筑波大学 名誉教授）

日本学術振興会学術研究助成金 基盤研究（B）「辞書サイト・アプリ開発に資する質の高い日本語例文バンクの構築とその応用研究」（課題番号：21H00535, 研究代表者：鈴木智美, 研究分担者：清水由貴子・加藤恵梨・中村 彰・伊達宏子・藤村知子・望月 源）
【お問い合わせ】 jreibun@tufs.ac.jp

（すずき ともみ 東京外国語大学大学院国際日本学研究院 教授）

（しみず ゆきこ 聖心女子大学現代教養学部 講師）

（なかむら あきら 東京外国語大学大学院国際日本学研究院 准教授）

（かとう えり 大手前大学現代社会学部 准教授）

The Report of the First Open Research Meeting of the “Jreibun” (the Bank of Japanese Example Sentences) Grant-in-Aid Project: Making Public High-Quality Example Sentences Usable for Japanese Language Study Tools as Open Data

SUZUKI Tomomi, SHIMIZU Yukiko, NAKAMURA Akira, KATO Eri

KEYWORDS: Jreibun (the Bank of Japanese Example Sentences), the database of example sentences, study tools, dictionary apps, Open Data

In this article, we describe an outline of “Construction and applied research of the database of high-quality Japanese example sentences available for the development of dictionary websites and applications” (JSPS KAKENHI Grant Number 21H00535, Grant-in-Aid for Scientific Research (B), 2021-2024, Principal Investigator SUZUKI Tomomi) and report on the first online research meeting open to interested people held in July, 2021.

The purpose of this research project is to construct a database of high-quality Japanese example sentences with their English translations that can help to improve the quality of the contents of study tools such as dictionary applications and Websites developed for Japanese language learners in and outside of Japan. A Creative Commons license will be assigned to the database of example sentences, which will be made public as Open Data available to developers of tools such as apps and Websites in and outside of Japan.

The first open meeting of this research project took place online on July 18, 2021, with about 90 participants in and outside of Japan. In Part I, we explained what had led to the conception of our research and showed problems found in current Japanese language study tools. In Part II, we described the aim of this research and explained the method of creating concrete example sentences, and sought to share our ideas with the participants.

Our commentator and participants expressed their expectations in regard to collaboration between Japanese language education and information technology, making the result of the research available to users in the form of Open Data, and basic concepts of the database that will enable it to develop further, among other things.