日本語例文バンク（Jreibun）プロジェクト中間報告

― 作成された全例文を使用語彙・文パターンの観点から見る ―

鈴木智美・清水由貴子・渋谷博子・伊達宏子

キーワード：「Jreibun」（日本語例文バンク）、項目の選定、例文作成、使用語彙、Ngram、音声録音

1. 本稿の目的

本稿の目的は、「日本語例文バンク」（Jreibun）プロジェクト（日本学術振興会科学研究補助金基盤研究（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」）[1] の進捗状況についての中間報告を行い、2021 ～ 2022 年度にかけてプロジェクトにおいて作成されてきた約 9,000 の例文について、その使用語彙および観察される文パターンを通じて検討を行うことである。同時に、鈴木・中村（2023）で述べられている例文作成の注意点をふまえた上で、プロジェクトメンバーの視点からあらためて例文作成の実際についてその一例を簡潔に報告する。また、作成例文には一部、読み上げ音声も付すこととしており、2022 年度に行った録音・編集作業についても報告を行う。

2.「Jreibun」プロジェクトの目的および実施体制

「Jreibun」プロジェクトとは、日本語学習者を対象とする各種アプリ、ウェブサイト等のツール開発において、質の良い日本語の例文が利用可能となるよう例文データベース（英訳付き）を作成し、オープンデータとして公開することを目的としたプロジェクトである。鈴木他（2022）で、その概要および着想に至った経緯等が報告されており、鈴木（2023）では 4 年間のプロジェクトの前半（2021 ～ 2022 年度）の中間報告がなされ、鈴木・中村（2023）では、例文作成の際に注意すべき点、および英訳を行うにあたっての留意点が詳細に報告されている。

「Jreibun」プロジェクトで作成した例文は、プロジェクト代表者の鈴木（本稿の筆頭執筆者、当該科研の研究代表者）が全てに目を通し、完成することとしている。鈴木・中村（2023）にも述べられている通り、研究代表者が直接修正を加えるか、あるいは必要に応じて作成者に修正を依頼し、再チェックするという手順を踏み、全ての例文を完成している。また、プロジェクトメンバーのうち、本科研の研究分担者で文法論を専門とする清水由貴子（本稿の第二執筆者、聖心女子大学）、および意味論を専門とする加藤恵梨（愛知教育大学）の 2 名も、作成途中段階において他のメンバーの作成例文のチェック作業を分担して行っている。

---

[1] 令和 3（2021）年度～令和 6（2024）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、望月源、令和 3（2021）年度に藤村知子
プロジェクトのホームページは下記の通り。
https://www.tufs.ac.jp/ts/ personal/SUZUKI_Tomomi/jreibun/index-jreibun.html
科学研究費補助金の当該研究課題のページは下記の通り。
https://kaken.nii.ac.jp/ja/grant/KAKENHI-PROJECT-21H00535/

例文作成については、2021 年度は上記 3 名、および本科研の研究分担者の中村彰（東京外国語大学）、伊達宏子（本稿の共同執筆者、東京外国語大学）の 2 名、さらに本科研の研究協力者 10 名[2]をあわせ、全 15 名で分担して行った。2022 年度は研究分担者の中村は英訳の監修作業に専従することとし、研究協力者 10 名[3]のほか、上に挙げた研究分担者（清水、加藤、伊達）3 名、および研究代表者（鈴木）をあわせ、全 14 名で作成を行った。2023 年度は研究協力者が 12 名[4]となり、同じく全 16 名で作成を行っている。例文作成の方針、手順および作成にあたって注意すべき点は、鈴木他（2022）、鈴木・中村（2023）、および鈴木（2023）で述べられている通りであるが、本稿第 6 節では、研究プロジェクト開始初年度より継続して参加している研究協力者のうち 1 名（渋谷博子）が、実際の例文作成の手順と留意している点についてあらためて報告を行う。

また、「Jreibun」には、例文データのほか、それぞれの例文の英訳データも含まれることとなっており、完成した例文については、順次プロジェクト外部の英語翻訳専門家にその英語訳を依頼している[5]。それらの英語訳については、さらに日本語教育・英語教育の観点から、研究分担者の中村が目を通し、必要に応じて修正等を行うこととしている[6]。また、例文には一部音声データを付すこととし、2022 年度に一部の例文の読み上げ音声の録音・編集作業を行った[7]。その概要は、研究分担者の伊達が本稿第 7 節で報告を行う。このほか、例文データベースにはふりがなデータも付すこととしており、ふりがな入力作業についても、別途研究補助者に行ってもらっている。

なお、本稿第 4 節および第 5 節では、「Jreibun」プロジェクトにおいてこれまで作成された約 9,000 の例文を、その使用語彙、および観察される文パターンの観点から、それぞれ研究代表者の鈴木、および研究分担者の清水が分析する。その際の基礎データについては、本科研の研究分担者の望月源（東京外国語大学）が全例文の形態素解析を行い、それぞれ必要な処理を行ったものである。また、本稿第 3 節では、「Jreibun」において例文作成のきっかけとしている約 10,000 項目[8]について、研究代表者の鈴木がその選出手順の報告を行うが、これに関わる具体的な作業は、2020 年度から 2021 年度（本科研の初年度）にかけて研究補助を務めた韓金柱氏に行ってもらった。

また、本プロジェクトではこれまで 2 回の公開研究会[9]と、計 8 回のプロジェクト内勉強会[10]を

---

[2] 饗場淳子、浅野涼子、内海陽子、大山祐李、岡葉子、加藤理恵、渋谷博子、田丸望、中村亜美、三谷閑子の 10 名（敬称略、五十音順）である。

[3] 饗場淳子、浅野涼子、内海陽子、岡葉子、加藤理恵、渋谷博子、田丸望、中村亜美、野田大志、三谷閑子の 10 名（敬称略、五十音順）である。なお、この年度の後半に藤村春菜氏にいくつかの項目について例文の試作を行ってもらっており、2023 年度から新しく協力者として参加してもらうこととなった。

[4] 饗場淳子、浅野涼子、家田章子、内海陽子、岡葉子、加藤理恵、渋谷博子、田丸望、中村亜美、西島絵里子、藤村春菜、三谷閑子の 12 名（敬称略、五十音順）である。

[5] 英語論文校閲等を専門とする「リンクサイエンス」代表の川上輪子氏に依頼している。

[6] 英訳において注意すべきいくつかの点については鈴木・中村（2023）において報告されている。

[7] 音声読み上げは、フリーアナウンサー・日本語教師の梅田エリカ氏に依頼している。

[8] 「例文作成のきっかけとする項目」とは、わかりやすく言えば、通常の辞書におけるいわゆる「見出し語・見出し項目」にあたるものである。ただし、「Jreibun」は、日本語学習アプリやウェブサイトにおいて活用されることを念頭に作成しているため、作成された例文は、作成のきっかけとなった 1 つの項目のみに 1 対 1 で対応するわけではなく、その例文中に使用されている様々な語・表現からアクセス可能なものとなるため、このような呼び方をしている。

[9] 2021 年 7 月 18 日に第 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」を、また 2023 年 3 月 4 日に第 2 回公開研究会「日本語例文バンク Jreibun プロジェクト中間報告」を、いずれもオンライン形式で行っている。

[10] 勉強会はいずれもオンライン形式で行っている。

行っている。公開研究会では、プロジェクトの目的、概要、例文作成の方針・方法、進捗状況などを広く共有することとし、またプロジェクト内の勉強会においては、具体的な例文作成の手順や修正のポイントなどについて、メンバー間で疑問点などを出し合い、検討し、解決する機会としている。第2回公開研究会では、本プロジェクトの外部協力者として、日英対訳辞書サイト「Jisho.org」の開発・管理者のキム・アールストロム氏にも、当該サイトへの「Jreibun」作成例文の取り込み作業について報告してもらった。

3. 作成のきっかけとする項目の選定について

「Jreibun」では、例文作成のきっかけとする項目約10,000のリストを作成し、それぞれの項目を中心として例文を作成することとしている。項目選出の方法は、鈴木他（2022）においても簡潔に述べられているが、その手順の詳細を下記に述べる[11]。まず、項目選出のための基礎資料としたのは、以下（1）a～fの6種である。

（1）「Jreibun」において例文作成のきっかけとする項目の選出に使用した資料（6種）

a. 「日本語能力試験出題基準」語彙リスト
（旧4級～1級までの語、約8,100項目）

b. 『日本語コロケーション辞典』見出し項目
（動詞、形容詞、形容動詞を合わせて、約2,200項目）

c. 『日本語口語表現辞典』見出し項目
（会話やスピーチでよく使用される口語表現[12]、約3,300項目）

d. 「日本語学術共通語彙リスト」[13]見出し項目
（日本語学術テキストにおいて高いテキストカバー率を示す語、約2,500項目）

e. 「留学生のための基礎的専門語」[14]見出し項目
（化学、物理、経済を合わせて、約1,500項目）

f. 『初級日本語』『中級日本語』『上級日本語』[15]文型・語句
（初級～上級レベルの文法等の項目、あわせて約680項目）

上記6種の資料のうち、e.「留学生のための基礎的専門語」（約1,500項目）は、かなり専門的な用語も含まれているため[16]、準備段階として、より一般性の高いと思われるもののみを抽出した

---

[11] 項目選出作業は本プロジェクトが科研費を得て本格的に始動する前の2020年度から既に着手しており、本科研の初年度（2021年度）において最終の調整を行った。

[12] 「後はよろしく」「切れがある」「スープの冷めない距離」「見切りをつける」等、いわゆる「語」よりも長い単位で項目が立てられている。

[13] 松下達彦氏作成。日本語能力試験の旧1級語彙までには含まれていないが、アカデミックな文脈でよく使用される重要語がカバーされている。

[14] 小宮千鶴子氏作成。化学、物理、経済、数学のうち数学を除く3分野を資料とした。

[15] 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の日本語教科書である。

[16] 例えば、化学の専門語として「アゾ化合物」「アルキド樹脂」「オストワルト法」、物理の専門語として、「角周波数」「偶力」「コンプトン効果」など。経済の専門語は比較的一般語彙と重なるものが多いが、「ケインズ」「アダム・スミス」「マルクス」などの人名や「プラザ合意」「ブレトン・ウッズ協定」などの特定の事項の名称も含まれている。

抜粋版リストを作成した。一部、関連すると思われる語を追加した上で[17]、経済・物理・化学の用語を約 420 語に絞った。その上で、a ～ e まで（e. 専門語については上記抜粋版を使用）の統合版を作成し、相互に重複している項目を整理した結果、約 13,000 項目を得た。これに、f の初級～上級の日本語教科書における「文型・語句」680 項目を合わせると、約 13,680 項目が得られることとなった。以上をプロジェクト準備段階として 2020 年度に行った。

2021 年度より科研費を得て「Jreibun」プロジェクトを本格的にスタートすることとなり、この段階で、人員・予算の観点から、現実的な作業量を考え、例文作成の対象とする項目（作成のきっかけとする項目）を、約 10,000 項目まで絞り込むことにした。

まず、f の初級～上級の日本語教科書掲載の「文型・語句」には、接続詞、副詞、疑問詞、種々の複合動詞等、いわゆる「語」にあたるものも含まれているため、上記 a ～ e の語彙統合リストに既に収録されているものについては重複を削除することとし、また述語の連用中止法[18]やテ形接続[19]、「～という。」「～じゃないか。」「～と見てよい。」「～う／ようとは思っていません。」「～という心配／危険が出てくる。」などの文末表現、あるいは「～うちはよいが、…」などの接続表現など、「項目」として扱いにくいと考えられるものを除くことにした。以上の観点から「文型・語句」のリストについては 116 項目にまで絞り込んだ。なお、この 116 項目については、「Jreibun」で例文作成の対象として扱われている「文法項目」として、本稿の末尾にリストとして示す。

次に、準備段階で得られた約 13,000 項目の語彙関係のリストの中から、感嘆詞[20]、口語表現[21]、現代語としては使用頻度が低いと思われる語[22]を中心に優先順位の低い項目を確認しながら取捨選択を行い、語彙の統合リストを 10,432 項目に絞り込んだ。

以上のような手順を踏み、上記の語彙の統合リスト 10,432 項目（(1) a ～ e の相互の重複を除いた 13,000 項目に、取捨選択を加えたもの）、および文法項目 116 項目（(1) f に基づき絞り込みを行ったもの）をあわせ、例文作成のきっかけとする項目を計 10,548 項目にまとめた。プロジェクトでは、これらの項目を担当者ごとに配分し、例文作成を進めていくことにした。作成にあたっては、特に品詞等で分担を分けることはせず、五十音順に並んだ項目リストを、順に割り振っていくことにした。なお、作成のきっかけとしているこれらの項目のリストについては、プロジェクト後半には、Jreibun ホームページに掲載する予定で進めている[23]。

---

[17] 例えば、「乳酸菌」「気象予報士」「ビブラート」「贈収賄」「コンプライアンス」などの項目を追加した。この段階では、他のリストとの重複は考えていない。

[18] 中級文型として扱われているもので、「先生は学生に本を<u>読ませ</u>、漢字の練習をさせる。」「技術が<u>進歩し</u>、便利な機械がいろいろ作られて、生活が向上した。」などの下線部分の表現方法である。（例文はいずれも『中級日本語』より）

[19] 述語のテ形接続には様々な用法があるが、例えば初級文型として扱われているもので、「ナイフとフォークを<u>使って</u>肉を切ってください。」「何度も<u>くり返して</u>テキストの CD を聞いた方がいいです。」などは前件が後件の手段を表す用法になっている。（例文はいずれも『初級日本語』より）

[20] 「日本語能力試験出題基準」語彙リスト記載の「ああ」「あっ」「えっ」など。

[21] 例えば『日本語口語表現辞典』の「足元が悪い」「足元にも及ばない」「足元を見る」などは「足元」に、「いい気味」「いい薬」「いい根性してる」「いい線いってる」などは「いい（良い・好い）」に集約させ、「あばよ」「あんよ」「楽ちん」「レア物」などは優先度が低いと考え、今回は削除するなど、調整を行った。

[22] 「日本語能力試験出題基準」語彙リスト記載の「洋品店」「お母様」「ストロボ」など。

[23] 実際に例文の作成を始めてから、例えば他動詞の「ほぐす」は項目にあるが自動詞の「ほぐれる」がない、動詞「歪む（ひずむ）」があるが名詞「歪み（ひずみ）」の例文も加えたいなどの場合には、それらを項目として追加することとしている。「Jreibun」は鈴木他（2022）でも述べられているが、例文の改良・差し替え・追加等を可能とする「成長するデータベース」を基本方針としており、例文作成のきっかけ

また、2022年度に行った第2回公開研究会（2023年3月）では、(1) dに挙げた「日本語学術共通語彙リスト」作成者の松下達彦氏にコメンテーターを依頼したが、そのコメントの中で「日本語能力試験出題基準」を基にすると、外来語が不足する懸念があるのではないかというものがあったが、上記の手順を見るとわかるように、「Jreibun」が作成のきっかけとする項目は「日本語能力試験出題基準」のみに基づくものではなく、(1) のa～fに挙げた6種の資料に基づき、相互に重複を排除した上で得られた項目をさらに絞り込んだもの（一部追加したものも含まれる）となっている。具体的に、松下氏より「旧日本語能力試験の初級レベル（旧4級、3級）にないが、初級語彙として追加すべき外来語」のリスト102語[24]が提示されたが、研究会終了後に確認したところ、うち78語は「Jreibun」が作成のきっかけとする10,548項目に既に含まれており、基本的な外来語が大きく抜けているということはないことが確認された。

4. 例文の検討：使用されている語彙の観点から

これまで、プロジェクト初年度の2021年度には、2,266項目について3,956例文、2年目の2022年度には、2,812項目について5,021例文、2年間で計5,078項目をきっかけとして全8,977の例文が作成され、それぞれに英訳が付されている。ここでは、それらの例文に使用されている語彙に着目し、作成例文にどのような特徴や傾向が見られるかを検討する。

4.1　例文に使用されている語の総数

解析の下処理においては[25]、まず、作成されたすべての例文が1文ごとに切り分けられる。ここで「1文」とは、句点「。」で終わる文のことであるが、引用符「 」内に見られる「。」については区切らずに処理を行っている[26]。その結果、上記全8,977の例文は、計10,223の文により構成されていることがわかった。即ち、この数を見れば、「Jreibun」の例文には、1つの例文が複数の文により構成されているものがあることが見てとれる[27]。

次に、上記のすべての文を形態素解析器「MeCab」を用いて解析した結果を見ていく[28]。上記全

とする項目についても、プロジェクト開始時の10,548項目に固定するものではなく、項目の追加は随時可能である。

[24] ただし、102語の中には単独では語として使用されにくいと思われる「マン」「ワン」「ツー」「デー」「ルーム」のようなものも含まれている。

[25] 第2節で述べたように、本節で使用する基礎データについては、本科研の研究分担者の望月氏により提供されたものである。

[26] 例えば「野党議員は、「これまで再三申し上げてきましたように、今は格差解消が最優先するべき課題ではないのでしょうか。これについて総理のお考えをはっきりとお聞かせ願いたい。」と述べた。」(「再三」を作成のきっかけ項目とする例文）については、「」内の「。」では区切らず、この全体で「1文」となる。

[27] 例えば、「「日豪交際交流プロジェクト」の一環で、オーストラリアの高校生がこの夏1週間うちでホームステイをすることになった。そこで、春から英会話の勉強を始めることにした。」（「そこで」を作成のきっかけ項目とする例文）のように、作成のきっかけとする項目が接続詞等である場合、また、「野生の生物が置かれている状況は厳しく、絶滅の恐れがある種も少なくない。生物の多様性を守るための取り組みが急務となっている。」（「急務」を作成のきっかけ項目とする例文）のように、対象とする語・表現の意味するところを文脈において明快に示すために、文を複数に分けて述べているものなど、1つの例文が複数の文から構成されるものがある。

[28] 語彙については「表層形」と「基本形」でカウントしたデータがあるが、ここでは「基本形」のデータを見る。また、品詞の特定にあたり、「MeCab」標準の「IPA辞書」のほかに、「NEologd」（「LINE」提供の「MeCab」用のシステム辞書で、新語や固有名詞に強いとされる）、および「UniDic」（国立国語研究所開発の解析用辞書）を使用した解析も行われているが、ここでは主として「IPA辞書」を用いた解

8,977 の例文については、まず単純に解析結果を見ると、異なり語数 16,171 語、延べ語数 275,026 語で構成されている[29]。そのうち、句読点や引用符などの記号が 17 種、計 27,832 見られるため、それらを除くと、例文に使用されている語の異なり語数は 16,155 語、延べ語数は 247,194 語となる。例えば、旧日本語能力試験において、1 級レベルの出題語彙の範囲の目安が 10,000 語とされていることを考えると、計約 5,000 項目をきっかけとして作例された全例文において、異なり語数約 16,000 語が含まれているというのは、語数から見て決定的な不足はないと思われる。以下、主要な品詞（動詞、形容詞、形容動詞、副詞、名詞）ごとに使用されている語を見る。各品詞において使用数の多い上位 30 語については、本稿末尾に表として資料を付している。なお、本稿では使用語彙の傾向性を形態素解析の結果を通じて大きくつかむことを目的としており、以下に各品詞ごとの異なり語数等を述べていくが、誤解析の可能性のある箇所をすべて解消した上で、一の位まで厳密にそれぞれの語数を導き出そうとすることを意図したものではない。

4.2　主要な品詞別に見た例文によく使われている語

動詞は、異なり語数 2,437 語、延べ語数 40,881 語という結果となった。ただし、この中には受身・使役形の接辞「れる」「られる」「せる」「させる」や、接辞「がる」、縮約形「ちゃう」「てる」[30]、誤解析のもの[31]が含まれており、これらを除くと、異なり語数は 2,425 語となる[32]。使用されている動詞の上位 4 語（いずれも使用数 1,000 以上）は、「する」「いる」「なる」「ある」である。次に使用数 500 以上のものに「しまう」[33]「できる」「言う」[34]「くる」[35]があり、以下、200 以上のものに「見る」「行う」「行く」[36]「思う」「出る」「使う」「いく」「くれる」が見られる。いずれも基本的な動詞が使用数の上位を占めていることがわかる。

形容詞については異なり語数 361 語、延べ語数 4,323 語である。ただし、ここには「やすい／にくい／がたい／づらい」のように、複合語の後項として用いられ、物事の難易や傾向性を表すものも含まれている。また接辞「ぽい／っぽい」や、誤解析のもの[37]なども含まれており、これ

析データを見る。

[29] ここで語数を述べる際の「語」とは、「MeCab」およびその標準の「IPA 辞書」によって解析された単位を示すものとなる。

[30] 「ている」の縮約形で、「売れるなんて豪語してるけど」「合格はできてると思うけど」の下線部分のようなものである。

[31] 動詞として「く」「ぶす」などを抽出する誤解析が見られた。例えば、「日和ったままで」が「日和（名詞）」「く（動詞）」「たま（名詞）」「まで（助詞）」、また「ぽっちゃりとして健康そうな」が「ぽい（形容詞）」「く（動詞）」「ちゃり（名詞）」「として（助詞）」と、また「ぶしつけな」が「ぶす（動詞）」「つける（動詞）」「だ（助動詞）」のように解析されている。そのほか、「おせち料理」「まどろっこしい」「ふいにしてしまった」から「おせる」「どる」「ふう」という「動詞」を切り出すという誤解析が見られた。

[32] 複合動詞後項の「きる」「きれる」「かねる」なども動詞 1 語としてカウントされている。

[33] 単独の動詞としてではなく、補助動詞としての用法が多いと考えられる。

[34] 「言う」と「いう」は別語としてカウントされている。ひらがな表記の「いう」には、文末の「～という。」のほか、「～からといって」「何かというと～」「～というほどの」「～という意味で」等の例が含まれる。これとは別に引用形式の「という」というかたまりが「～というのは…」「～というNは…」「～ということだ。」などの例において、助詞としてカウントされている。

[35] 「くる」と「来る」は別語としてカウントされている。ひらがな表記の「くる」については、補助動詞として用いられたものが多いと考えられる。

[36] 「行く」と「いく」も別語としてカウントされている。ひらがな表記の「いく」については、補助動詞として用いられたものが多いと考えられる。

[37] 形容詞として「くい」「うい」などを抽出する誤解析が見られた。「行くことを決めた」が「行く（動詞）」

らの接辞と誤解析のものを除くと、異なり語数は357語、延べ語数4,311語となる。使用されている形容詞で最も多いのは（使用数500以上）「ない・無い」[38]である。次に使用数200以上のものとして「いい・よい」[39]「多い」がある。使用数100以上のものには「高い」「新しい」「良い」[40]「悪い」「やすい」[41]が見られ、50以上のものに「長い」「難しい」「大きい」「ほしい・欲しい」[42]「強い」「美しい」「少ない」「うまい・上手い・旨い・巧い」[43]「早い」「にくい」[44]「若い」が続く。動詞と同様、基本的な形容詞が上位を占めている。

形容動詞については、「MeCab」の標準の解析辞書では独立した品詞としては区別されておらず、名詞の範疇に含まれている[45]。したがって、ここでは「名詞」に分類された語の中から形容動詞の用法があると思われるものを目視にて抜き出したところ、計561見つかった[46]。また、名詞に分類されているものの中には「積極」「消極」「本格」「相対」「民主」「共和」「文部」「当事」など、専ら複合語の語基として用いられるタイプのものも含まれているが、このうち「積極」は28例のうち27例が「積極的」として用いられており[47]、「本格」も8例のうち6例が「本格的」[48]、「相対」は使用されている5例のうち4例が「相対的」[49]、「消極」は使用されている2例がいずれも「消極的」となっている。これらの「積極的」「本格的」「相対的」「消極的」も形容動詞の1つであるため、これらを含めると形容動詞の異なり語数は計565となる[50]。なお、抜き出した形容動詞には、名詞用法と形容動詞用法の双方を持つもの（「健康」「心配」「元気」「危険」など）と、形容動詞としてのみ用いられるもの（「きれい」「丁寧」「豊か」「静か」など）がいずれも含まれている。したがって、以下で述べる使用数が多く見られた形容動詞というのは、名詞あるいは形容動詞の

「くい（形容詞）」「こと（名詞）」、また「恩師を失い」が「失（動詞）」「くい（形容詞）」に、「手洗い、うがい」が「うい（形容詞）」「がい（名詞）」などと誤解析されている。

[38] 否定の接辞「ない」は別個に「助動詞」としてカウントされており、形容詞としてカウントされた「ない」は「元気がない」「可能性がない」「面影がない」「予定が何もない」「異常はない」「年齢や性別に関係ない」など、“有無”の“無”を意味するものである。ただし、「うらやましくてしかたがない」などもここに含まれている。ほとんどがひらがな表記で、漢字表記の「無い」は2例のみである。

[39] 「いい」と「よい」は別個にカウントされているが、ここでは合わせた数で見た。これには、「～たほうがいい／よい」「～といい／よい」や、「よく考えて」の「よい」などが含まれる。

[40] 「良い」については、「品質が良い」「町を良くする」「良い教育を受けさせる」「行儀が良くない」「仲の良い」「血液の循環を良くする」「良いお年を」など、単独の形容詞として「物事の優劣・善悪・適否」に関わる意味で用いられている。

[41] 「やすい」は「歩きやすい」「～になりやすい」など、容易であること、あるいは傾向性を示すものとして用いられている。価格・値段が低いものであることを示す「安い」は別にカウントされている。

[42] 「ほしい」と「欲しい」は別にカウントされているが、ここでは合わせた数で見た。「ほしい」は「貸してほしい」「来てほしい」など、ほかの人に対する希望を表す用法のものである。

[43] これらの表記の違いは別個にカウントされているが、ここでは合わせた数で見た。また「巧い」1例は、「巧（名詞）」「い（名詞）」と誤解析されたものである。

[44] 「にくい」は、「わかりにくい」「聞こえにくい」などである。単独の形容詞としての「憎い」については別にカウントされており、2例見られる。

[45] 注28に述べた解析辞書のうち「UniDic」のみでは「形状詞」という範疇が立てられ、形容動詞が区別されている。「UniDic」の解析データに基づくと、「形状詞」は異なり語数で448語見つかる。ただし、この「形状詞」には「よう」「そう」「どんな」「そんな」「こんな」なども含まれている。

[46] 「惨憺」「断固」「忸怩」など「タリ活用」のものも含まれる。

[47] 残り1例は「積極性」である。

[48] 残り2例は「本格化」「本格始動」である。

[49] 残り1例は「相対評価」である。

[50] 「民主」については「民主主義」「民主政権」で用いられており、「民主的」の使用例は見られなかった。

いずれかとして、その語が例文に多く使用されているということになる[51]。

使用数が100以上見られる形容動詞は「必要」「様々・さまざま」[52]「好き」の3語である。「～が必要である」あるいは「様々な～」などは、例文の文脈を適切にまとめていくために使用頻度の高い表現であると思われる。また、何らかの物事について「好きである」か、あるいは「好きではない」かということが文脈の中で触れられることが多いことも読み取れる[53]。次に、使用数50以上のものは「大切」「健康」[54]「可能」「自然」[55]「重要」である。これらも、「～が大切である」「～が可能である」「～が重要である」として例文をまとめるのによく使われる表現であると思われる。また、例文において、「健康」「自然」に関連する話題が取り上げられることも比較的多いのではないかということが見てとれる。そのほか、使用数20以上のものには「きれい・綺麗」[56]「心配」[57]「簡単」「有名」「自由」[58]「危険」[59]「丁寧」「無事」「不安」[60]「大変」[61]「豊か」「嫌・いや・イヤ」[62]「元気」[63]「大事」「新た」「適切」「幸せ」[64]と続き、主要な形容動詞が用いられていることがわかる。

また、様々な外来語形容動詞も用いられている。「スムーズ」「シンプル」「グローバル」「アナログ」「マイペース」「カジュアル」「リアル」「ゴージャス」「フォーマル」「ユニーク」「クリエイティブ」「マクロ」「クリーン」「メジャー」「エレガント」「ネガティブ」「ナンセンス」「ハード」「ロマンチック」「フルーティー」「ダイレクト」「シャープ」「セクシー」「シャイ」「アクティブ」「リーズナブル」「オリジナル」「ダイナミック」「ドライ」「シリアス」「コミカル」「クリア」「ナチュラル」「インタラクティブ」「ヘルシー」「ローカル」「ラフ」「ポジティブ」「スピーディー」などが例文に使われている。

副詞については、異なり語数612、延べ語数3,888となった。程度を表す「非常に」および「大

---

[51] ここではすべての例について名詞用法か形容動詞用法かを遡って区別することはしていない。

[52] 「様々」と「さまざま」は別個にカウントされているが、ここでは合わせた数で見た。

[53] 100以上の使用数があるため、全ての例を挙げることはできないが、例えば「私が好きなアニメの主人公は」「相手の話を聞くのは好きだが」「好きな歌手が出る番組だけは」「海の好きな夫は」「なぞなぞ好きの息子は」「甘いものが好きか聞かれたので」「冬が一番好きだ」「冬という季節は好きになれない」「チーズがお好きなんですか」「柔軟剤の香りは好きな人がいる一方で」「好きな音楽を2、3曲聞くと」「それぞれ好きなものを買ってきて」「引っかかるような言い方をするので、好きになれない」「君が好きなように決めてよい」など、「好きである」あるいは「好きではない」ことに触れる様々な文脈において用いられている。

[54] 「健康」については、「健康のため」「健康を維持する」「健康を悪化させる」「健康診断」「健康食品」「健康被害」「健康保険」等、名詞の用法も多く見られる。

[55] 「自然」についても、「自然の風景」「自然の美しさ」「自然の営み」「自然の中で遊ぶ」「自然をないがしろにする」「豊かな大自然」「自然豊かなところ」「自然災害」「自然遺産」「自然乾燥」「自然消滅」などの名詞の用法が多く見られる。

[56] ほとんどがひらがな表記で、「綺麗」が1例のみ見られる。

[57] 「心配」については、名詞の用法もある。

[58] 「自由」についても、名詞と形容動詞の双方の用法がある。

[59] 「危険」も名詞の用法がある。

[60] 「不安」も名詞の用法がある。

[61] 「大変」については程度副詞として用いられているものをここでは除いている。

[62] 「嫌」「いや」「イヤ」は別個にカウントされているが、合わせた数で見た。なお、ひらがな表記の「いや」の中には「いやが上にも」(「弥が上にも」で「嫌」とは異なる)、また「いやに」という副詞用法が含まれていたが、これを除いてカウントした。カタカナ表記の「イヤ」は子どもの「イヤイヤ期」という例で「イヤ」が2回カウントされている。

[63] 「元気」も名詞の用法がある。

[64] 「幸せ」も名詞の用法がある。

変」は名詞「非常」「大変」として分類されており、これを副詞として数えると[65]、異なり語数は614となる。100以上見られるものは「よく」「いつも」であり、50以上のものには「すぐ」「なかなか」「まだ」「少し」「とても」「どう」「初めて」「まず」「あまり」が見られる。以下、使用数が多くないものまで順に確認していくと、物事のありよう[66]や程度[67]、また時間的な展開[68]、評価[69]に関して用いられるもの、あるいは文末の表現と呼応するタイプのもの[70]まで、様々な副詞が用いられていることがわかる。

名詞に関しては、約12,000の異なり語数、約90,000の延べ語数が見られる。ただし、「MeCab」の標準解析辞書で特定しきれなかった名詞約3,000が1つにまとめて分類されており[71]、これらを1つ1つ特定すると、異なり語数はもう少し増えるように思われる[72]。形式的な意味を表す「こと」「の」「よう」「ため」、あるいは接辞「的」「たち」などを除くと[73]、上位に見られるのは「人」「私」「者」「日本」「子ども」「家」「会社」「仕事」「自分」「時間」「大学」「友人」「世界」「問題」「時代」などで、これらの語が例文によく使われている語であることがわかる。家、子ども、友人な

[65] 副詞用法の「非常に」は27例見られ、「非常に重要だ」「非常に寒い」「非常に強い」「非常に義理固い」「非常に興味深い」「非常に巧い」「非常に喜ばしい」「非常にスケールが大きい」「非常に残念だ」「非常に困難だ」「非常に優秀だ」「非常に厳格だった」「非常に仲の良かった」「非常に充実している」「非常に緊張していた」「非常に便利なツール」「非常にストレスを感じてしまう」「非常に苦痛を感じる」「非常に落胆させられた」などである。形容動詞としての用法は「非常な勢いで」が1例、それ以外の名詞用法は「非常事態」「非常ボタン」「非常の際」などである。「大変」は40例見られるが、形容動詞として用いられているものが31例で、「覚えるのが大変だ」「～は大変な仕事だ」「掃除が大変になる」「手入れ／世話が大変」「混雑して大変だ」「思ったより大変だった」などで、副詞として用いられているものは「大変面白い」「大変厳しい」「大変美しかった」「大変人気だった」「大変困った」「大変お世話になった」「大変恐縮した」「大変申し訳ないのですが」など、9例見られる。

[66] 例えば「はっきり」「ゆっくり」「すっかり」「一気に」「黙々と」「のんびり」「こつこつ」「どんどん」「ちゃんと」「こっそり」「ぴったり」「うっすら」「ふんだんに」「いちいち」など。

[67] 例えば「かなり」「全く」「ずっと」「もっと」「大変」「さらに」「できるだけ」「なるべく」「極めて」「結構」「一層」「大いに」「ある程度」「比較的」「そこそこ」など。

[68] 例えば「時々」「たいてい」「やっと」「いよいよ」「そろそろ」「だんだん」「ついに」「いったん」「いきなり」「早速」「あっという間に」「既に」「今や」「今にも」「あらかじめ」「たまたま」「しょっちゅう」「急遽」「とりあえず」「もとより」「もはや」など。

[69] 例えば「当然」「もちろん」「あいにく」「やはり」「やっぱり」「せっかく」「さすが」「やむなく」「かえって」「惜しくも」「どうせ」「あえて」など。

[70] 例えば「決して」「必ずしも」「一向に」「それほど」「ぜひ」「たとえ」「もし／もしも」「たぶん」「何とか／なんとか」「どうも」「どうやら」「どうしても」「いざ」「さぞかし」「まさか」など。

[71] 年号のほか、「ブッカー（賞）」「FIFA」などの固有名詞、「カッパドキア」などの地名、「ワンオペ」「サブスクリプション」「アンインストール」「ジェネリック」「ディフューザー」「デジタル・デトックス」「RPG（ロール・プレイング・ゲーム）」などの比較的新しく使われるようになったと思われる外来語（和製英語を含む）やその縮約形などが特定しきれていない。なお、上記に挙げた例は「NEologd」および「UniDic」辞書ではいずれも解析・特定されている。「アンインストール」「デジタル・デトックス」「ロール・プレイング・ゲーム」については「NEologd」では1語とカウントし、「UniDic」では「アン」「インストール」、「デジタル」「デトックス」の各2語に、「ロール」「プレーイング」「ゲーム」の3語に解析している。ただし「UniDic」では「un」を「アン -Anne」と人名としてカウントするという誤解析は生じている。

[72] 一方、誤解析されていると思われるものも数十の単位で見つかるため、名詞の数はここでは厳密には特定しない。

[73] ほかに接辞「がち」「ら」「だらけ」「げ」「ちゃん」や、助数詞「枚」「個」「件」「兆」「か所」、また序数詞「号」「位」「条」「号室」「度目」なども名詞としてカウントされている。そのほか、複合語の一部を構成する以下のような漢字形態素もいずれも名詞としてカウントされている。例えば、「制」「派」「校」「園」「状」「剤」「割」「料」「省」「業」「医」「高」「大」「観」「視」「誌」「権」「史」「画」「圏」「科」「官」「集」「界」「審」「展」「路」「長」「製」「令」「価」「群」「活」「婚」「帽」「公」「帳」「重」「廃」「庁」「治」「犯」「報」「記」「半」「教」「失」「録」「簿」「植」「給」「歴」「敗」「近」「乗」「毎」などである。

どの身近な話題、仕事や会社、大学に関することなどが話題に含まれ、また世界という範囲、時代という概念[74]、何らかの「問題」に言及することも多い[75]ということがこれらの使用語彙から見てとれる。

以上、「Jreibun」において、作成のきっかけとしている項目を含め、全ての例文中に使用されている語を見ると、現時点において異なり語数は約 16,000 語、延べ語数は約 247,000 語、主な品詞別には、異なり語数で動詞 2,425、形容詞 357、形容動詞 565、副詞 614、名詞約 12,000（ただし、いずれも誤解析をすべて完全に排除した数ではない）という結果となった。語と語の結びつきに着目し、多くの用例を示している『日本語コロケーション辞典』を例にとると、その収録項目は、動詞 1,370 語、形容詞 361 語、形容動詞 540 語とあり[76]、「Jreibun」の例文中に使用されている語も、これらの主要な品詞においては特に偏りなくバランスの範囲内にあると思われる。

5. 例文の検討：観察される文パターンの観点から

次に、2021 ～ 2022 年度に作成した 8,977 の例文（全 10,223 文）[77] に、どのような文のパターンが使用されているかを明らかにするために、Ngram を作成して文のパターンを抽出し、分析を行う。

5.1　Ngram による文パターンの抽出

Ngram とは「N 個の形態素の連鎖」という意味で、ここでは、3gram ～ 5gram の連鎖を機械的に作成し、その中から意味のある連鎖を目視で抽出して分析していく。手順は以下の①～③の通りである[78]。

① 8,977 の例文（全 10,223 の文）を MeCab（IPA 辞書使用）で形態素に切る。
② 3gram、4gram、5gram の連鎖を機械的に作成する。
③ 3gram、4gram、5gram の連鎖の各パターンの個数を集計する。

例えば、「私は毎朝、近所を走ることにしている。」という文を形態素に切ると、「私 - は - 毎朝

---

[74] 「学生時代」「大学時代」「高校時代」「子ども時代」「小学校時代」「縄文時代」「弥生時代」「飛鳥時代」「平安時代」「戦国時代」「江戸時代」「明治時代」「絶対王政時代」「冷戦の時代」「ポスト石油時代」「今の時代」「～が栄えていた時代」「活気にあふれていた時代」「不況の時代」「コンピュータの時代」「大学全入時代」「なんでもネットで検索できる時代」「いつの時代でも／にも」「時代／時代の流れとともに」「時代を先見する力」「時代が変わる」「今の時代にも通じる」「～の喜びを味わえる時代になった」等の例が見られる。

[75] 「問題」については、例えば「～に問題がある」「～は問題ない／問題がない」「～が問題となっている」「～は切実な問題だ」「～は問題外だ」「問題がなければ～」「問題を抱えながらも」「多くの問題を内包する」「注目されている問題の一つだ」「～では根本的な問題解決にならない」「問題点を洗い出す」「～の行動を問題視する」「環境問題」「人口問題」「社会問題」「国際政治問題」「入学試験問題」「問題発言」「基地をめぐる問題」「憲法をめぐる問題」「農作物の輸出入の問題」等、様々な例が見られる。

[76] 「前書き」に記載されている。

[77] 第 4 節で述べた通り、例文には複数の文からなるものもある。

[78] 第 4 節で使用した資料と同様に、①～③に示す作業は、本科研の研究分担者の望月氏に依頼し、データの提供を受けている。

-、- 近所 - を - 走る - こと - に - し - て - いる -。」（以下、形態素の区切り目を「-」で示す）となる。「、」「。」も 1 つの形態素としてカウントされる。これをもとに、3gram、4gram、5gram を作ると以下の（2）のようになる。当該の文からは、3gram で 11、4gram で 10、5gram では 9 つのパターンがそれぞれ抽出されることがわかる。

（2）「私は毎朝、近所を走ることにしている。」をもとに作成した 3gram、4gram、5gram

3gram	4gram	5gram
私 - は - 毎朝 は - 毎朝 -、 毎朝 -、- 近所 、- 近所 - を 近所 - を - 走る を - 走る - こと 走る - こと - に こと - に - し に - し - て し - て - いる て - いる -。	私 - は - 毎朝 -、 は - 毎朝 -、- 近所 毎朝 -、- 近所 - を 、- 近所 - を - 走る 近所 - を - 走る - こと を - 走る - こと - に 走る - こと - に - し こと - に - し - て に - し - て - いる し - て - いる -。	私 - は - 毎朝 -、- 近所 は - 毎朝 -、- 近所 - を 毎朝 -、- 近所 - を - 走る 、- 近所 - を - 走る - こと 近所 - を - 走る - こと - に を - 走る - こと - に - し 走る - こと - に - し - て こと - に - し - て - いる に - し - て - いる -。

このような作業を 10,223 の文すべてに対して機械的に行う。3gram、4gram、5gram の各パターンの個数について集計した結果を上位 20 位まで示すと、次ページに示した表 1 の通りとなる[79]。3gram において、最も多く観察されるパターンは「ている。」で、個数は 1,502 例、即ち全 10,223 文のうち 14.7% においてこのパターンが観察されるということがわかる。3gram、4gram、5gram のリストを順に見ていくと、連鎖が長くなるほどこのパターンはさらに「している。」「にしている。」「をしている。」等と細分化され、各パターンの発現個数がそれに応じて減っていき、より詳細で具体的な連鎖が見えてくることがわかる。（以下の図 1 を参照）

3gram	4gram	5gram
「ている。」（1,502）	「している。」（389）	「にしている。」（102）
		「をしている。」（46）
	「れている。」（294）	「されている。」（94）
		「言われている。」（87）
	「なっている。」（106）	「になっている。」（67）
		「となっている。」（36）

図 1 「ている。」を含む 3gram、4gram、5gram の連鎖（（　）内は全例文における使用数）

[79] なお、全 10,223 の文を区切った結果、取り出された全パターンは、3gram で約 18 万、4gram で約 21 万、5gram で約 22 万である。

表 1　「Jreibun」の作成例文（2021 ～ 2022 年度）における 3gram、4gram、5gram の各パターンの数
（上位 20 位）

順位	3gram	個数	順位	4gram	個数	順位	5gram	個数
1	ている。	1,502	1	している。	389	1	ことにした。	133
2	している	717	2	れている。	294	2	ていたが、	112
3	ていた	694	3	てしまった。	272	3	にしている。	102
4	した。	648	4	ていた。	239	4	されている。	94
5	たが、	585	5	になった。	179	5	と言われている	88
6	である。	405	6	てきた。	158	6	言われている。	87
7	れている	402	7	していた	151	7	になっている。	67
8	には、	359	8	されている	143	8	ようになった。	64
9	では、	351	9	にした。	142	9	なければならない。	63
10	てしまった	339	10	ことにした	140	10	ことになった。	57
11	になった	304	11	いたが、	126	11	なってしまった。	55
12	てきた	300	12	された。	120	12	ようにしている	53
13	をして	297	13	ていたが	116	13	たほうがいい。	49
13	がある。	297	13	ているが、	116	14	をしている。	46
15	になって	293	15	にしている	115	15	になってしまった	44
16	しまった。	281	16	をしている	113	15	することにした	44
17	なった。	280	17	と言われて	107	17	となっている。	36
18	された	277	18	なっている。	106	17	ことができた。	36
19	されて	260	19	言われている	100	19	をしていた	35
20	れた。	259	20	になっている	92	19	していた。	35
			20	ことがある。	92			

このほか、上記の表 1 からは次のようなことも読み取れる。3gram では「には、」「では、」といった主題化の表現が見られる。4gram では、「てしまった。」「てきた。」「ことがある。」などの文末の表現や「ているが、」という節末の表現が見られる。5gram では、「ことにした。」「ようになった。」「なければならない。」「ことになった。」「たほうがいい。」「ことができた。」などの文末の表現や、「ていたが、」という節末の表現も見られるということなどである。

5. 2　例文に観察される文パターン

以下では、特に数の多かった 7 つのパターン（以下 a ～ g）を、具体的な例文を挙げながら見ていく。（該当箇所に下線を付し、例文末尾の【　】に、例文作成のきっかけとなった語･表現を示す。）

a.　意志的な決定を表す「ことにする」、非意志的な決定を表す「ことになる」

意志的な決定を表す「ことにした。」というパターンは 5gram において使用数 133 回で、最も

よく見られる文末のパターンである。一方、非意志的な決定を表す「ことになった。」というパターンも 57 回ある。以下の（3）は「ことにした。」、（4）は「ことになった。」の例である。

（3）カスタマーセンターに電話をかけたが、なかなかつながらなかったため、しばらく時間をおいてから、再度かけ直すことにした。【しばらく】
（4）商品の販売と輸送を効率よく結びつけるため、我が社では新しい在庫管理システムを構築することになった。【構築】

「ことにする」と「ことになる」は、話者の意志性の有無の違いはあるものの、「決定」を表す表現という点で共通している。加えて、両者ともある決定をするに至った原因や理由が例文中に記されていることが多く、例えば（3）では「（電話が）なかなかつながらなかったため」、（4）では「商品の販売と輸送を効率よく結びつけるため」がそれに当たる。このことをふまえると、「…ため、～ことにした。」や「…ため、～ことになった。」のように、文のパターンをより広くとらえることもできる。

b. 前置きの「～が」

逆接の「が」を伴う「ていたが、」という節末のパターンは 112 回あり、後に続く主節の内容の前置きや状況説明の後に現れる。本節の分析では 5gram までを資料としているため表 1 には分析結果は現れていないが、「ていたが、」の直前の動詞を確認してみると、最も多いのが「していたが、」（34 回）、次いで「思っていたが、」（11 回）であった。

（5）音楽を聞きながら宿題をしていたが、電話がかかってきたので、音楽のボリュームを下げ、電話に出た。【ボリューム】
（6）色が落ち、穴があいたジーンズは、古くてくたびれたものかと思っていたが、ファッションとして意図的に加工したものもあると知って驚いた。【くたびれる】

c. 時々起こることを表す「ことがある」

「ことがある。」という文末のパターンは 92 回見られる。このうち、「時々あることが起こる」の意味で用いられているものは 88 回あり、以下の（7）のような個人的な経験や、（8）のような一般的に知られていることを説明するのに使われている。

（7）疲れると、漬物などしょっぱいものを食べたくなることがある。体が塩分を必要としているのかもしれない。【しょっぱい】
（8）2000 年代に生まれ、幼い頃からスマートフォンや SNS に親しむ人たちを Z 世代と呼ぶことがある。【年代】

なお、「ことがある。」のうち、「たことがある。」という「経験」の意味で使われているものは 4 回のみで、用法によって例文における使用数には偏りがあるようである。

d. 非意志的な変化を表す「ようになる」

変化を表す「ようになった。」は 64 回見られる。(9) は能力の変化、(10) は状況の変化、(11) は習慣の変化を表している。いずれも自然と起こるさまざまな変化を表す文である。

(9) 高校時代のテニス部の顧問の先生はとても厳しい指導をしていたが、それは私たちへの期待の裏返しだったということが、大人になった今、ようやくわかるようになった。【裏返し】

(10) ペーパータオルは使い捨てにすることができ、タオルやふきんと比べて衛生的で便利なのでよく使われるようになった。【衛生】

(11) 私の趣味は雲の写真を撮ることだ。きっかけは、あるときふと空を見上げたら、犬のような形の雲が見えて、面白いと思って写真を撮ったことだ。それから、面白い形の雲や、太陽に染まった雲など、いろいろな雲を撮るようになった。【見上げる】

e. 意識して行っている習慣を表す「ようにしている」

意識して行っている習慣を表す「ようにしている」は 53 回見られる。個々の例文には、(12) ～ (14) のように話者の個人的な習慣が書かれているが、どういう状況・理由で、その動作を意識して行っているのかを説明する内容になっている。

(12) 私は毎晩寝る前に温めた牛乳を飲むようにしている。そうするとなんだかよく眠れるような気がするのだ。【毎晩】

(13) 新しい家電を買うときはネットで口コミ評価を調べてから、買うようにしている。【調べる】

(14) 仕事が忙しく、これまでは外食が多かったが、最近は、栄養が偏らないように自炊をするようにしている。【栄養】

f. 一般論であることを表す「と言われている」「されている」

「れている。」(294 回)、「された。」(120 回)、「されている。」(94 回)、「言われている。」(87 回) など、受身の形で終わる文は少なくない。直前にはさまざまな動詞が使われるが、「する」を除けば、最も多いのは「言う」で、以下の (15) のように「(と) 言われている。」というパターンで一般論を述べることが多い。

(15) 虫歯の主な原因菌はミュータンス菌だと言われている。虫歯予防には、食べたら歯をみがくことを心がけることが大切だ。【菌】

「されている。」は、以下の (16) のよう単純に受身を表す場合もあるが、(17) のように一般論を述べることの方が多い。

(16) 市の中央図書館のロビーに、フォトコンテストの入賞作品が展示されている。【入賞】

(17) 学校においても家庭においても、様々な体験を通して子どもを教育することの重要性が指摘されている。【教育】

g. 話者の判断を表す文末表現

義務を表す「なければならない。」という文末のパターンは 63 回用いられている。また、提案・勧めを表す「たほうがいい。」という文末のパターンは 49 回（「だほうがいい。」を加えると 51 回）あった。

(18) 先月アルバイトが一人辞めたので、新しい人を雇わなければならない。【雇う】
(19) 100 円ショップで売っているものが、すべて 100 円とは限らない。300 円の商品や 500 円の商品もあるので、買うときは値段を確認したほうがいい。【ショップ】

そのほか、上位 20 位には入らないものの、36 位「ことが大切だ。」(23 回)、および 37 位「注意が必要だ。」(22 回）という文末のパターンも特徴的かと考えられる。これらも「なければならない。」「たほうがいい。」と同様、話者の判断を表す表現である。

(20) 健康な毎日を過ごすためには、規則正しい生活を送ることが大切だ。【毎日】
(21) 地震の後は、津波による被害が生ずる恐れがあるので、沿岸部では注意が必要だ。【生ずる】

以上のように、2021 ～ 2022 年度に作成された 8,977 の例文（全 10,223 の文）のデータから Ngram を作成して文のパターンを観察した結果、「決定：ことにする／ことになる」「前置き：～が」「時々起こること：ことがある」「自然な変化：ようになる」「意識して行っている習慣：ことにしている」「一般論：言われている／されている」「話し手の判断を表す文末表現：なければならない／たほうがい／ことが大切だ／注意が必要だ」といった表現が多用されているという傾向をつかむことができた。

6. 質の良い例文作成を目指して

「Jreibun」プロジェクトにおける例文作成の基本方針、および基本的な作成手順の例については、鈴木他（2022)、鈴木（2023）で述べられており、また例文作成の際に注意すべき点については、鈴木・中村（2023）で、文構造が明快であること、語義を的確にとらえること、文意が明瞭となるよう文脈を整えること、例文の内容に誤りや偏りがないよう配慮することという各点が示されている。本節では、本プロジェクトに初年度より参加する研究協力者のうち 1 名が、例文作成の現場の視点から、あらためて例文作成の手順とその際に注意している点について報告を行う。

例文は、具体的な場面が設定されており、学習者にとって一般的な辞書の例文よりもその語句や表現を使う文脈が理解しやすく、また広く日本に関する情報が含まれていたり、知識として学ぶ意義のある内容であるものを作成するようにしている。

日本語を教える際、教師は誰でも、提示する場面や使われる表現がわかりやすく、学習者の記憶に残るような例文を提示したいと考えるだろう。しかし、授業準備のために新出の語句・表現等を辞書などで調べても、辞書に示されている例文がそのまま学習者に提示できることは少ない。例えば「またとない」という語句について調べてみると、「これはまたとないチャンスだ」のように具体的な場面設定のない例文しか掲載されておらず、例文中の「これ」とは一体何かを教師が

補って提示しなければならないといったことが起こる。

「Jreibun」プロジェクトにより提示された例文作成の方針に基づき、本節の執筆者は、実際に以下のような手順で例文作成を進めている。まず、例文作成の対象となる語句・表現について、自身の直感でどのような語と共起し、どのような例文が考えられるかを頭におき、国語辞書で意味を確認する。時に、自身が見落としていた意味が辞書に記載されていることもあり、それらは、自身の知識の偏りか、辞書に記載はあっても実際の使用例が少ないものであるのか、その後の作業で注意するようメモしておく。次に、「NINJAL-LWP for BCCWJ」（以下 NLB）を用いて表記の揺れや使用されるレジスターの特徴、用例が多いコロケーションを確認する。その後、当該の語について、「現代日本語書き言葉均衡コーパス」（以下 BCCWJ）をアプリケーション「中納言」を用いて検索する。検索結果を表形式でダウンロードすれば、当該の語の前後にどのような表現が共起するか、「反転前文脈」「後文脈」による並べ替えもでき、その語の使用文脈について、より詳細な特徴をつかむことができる。ただし、「BCCWJ」に収録されているデータは 1976 年から 2005 年のものとなっているため、一般の検索エンジンを用いてウェブ検索を行うことで、より新しい使用例を探る場合もある。また、「BCCWJ」は書き言葉がデータとなっていることから、単純に用例数の多いものから拾っていくのではなく、レジスターを確認し、使用場面との兼ね合いを考慮に入れる。例えば法律、医療、コンピュータ等の関連分野で突出して多く用いられている用法があっても、使用場面が限られていると考えられる場合には、例文に取り入れるかについては検討の余地があると考える。

多義語の場合は、用例数が多い意味から作例し、NLB 等の検索で多く挙がったコロケーションを生かす。例文が唐突に「これは」で始まるというように、文脈が不明なものとならないよう、具体的な場面を設定する。一方で、詳細な状況説明を意図するあまり、例文が必要以上に長くなってしまう場合もあるため、説明不足にならず、かつ冗長にならないように留意する。取り上げる内容についての配慮は、以下に具体例を挙げて述べる。

(22) 19 世紀末に、通訳として活躍したジョン万次郎（1827-1898 年）は、元は漁師だったが、漁に出た際に遭難し、漂流しているところをアメリカの捕鯨船に救助された。これがその後のジョン万次郎の人生を大きく変えることとなった。【「捕鯨」】

「捕鯨」を「BCCWJ」で検索すると、日本の捕鯨政策を批判している文脈が多く見られるが、政治的・文化的に意見の分かれる内容は避け、日本と海外との懸け橋の先駆けともいえる人物「ジョン万次郎」を素材として取り上げ、作例を行った。

また、次の（23）は実際に日本語学習者が混乱するということを素材として取り上げたものである。

(23) 日本ではテストを採点するとき、正しいものに○をつけるのが一般的だが、文化によっては間違っているものに○をつけることもある[80]。【「丸」】

[80] 当初作成した例文は「日本ではテストなどの解答では、正答に○をつけるが、ほかの国では誤答に○をつけるところもある。」であったが、解答者が選択した解答に「○をつける」という意味か、それとも採点する者が、採点のために、正答あるいは誤答に「○をつける」という意味か、判然としなかったため、

納得できる例文を完成するためには、一度に完璧な例文を目指すのではなく、新しい項目の例文作成を順次進めていきながら、少しずつ元に戻って作成済みの例文の内容や表現を再検討し、修正を加えるという手順を踏んでいる。

以上、本プロジェクトの研究協力者により、例文作成の手順と作成の際に留意している点について実際の経験に基づき報告した。本プロジェクトでは随時15名前後の作成者が分担協力して例文作成を行っている。例文作成は時間と労力を要するものであるため、学習者にとって意義のある例文を作成し、それを学習者のもとへ届けたいという強い意欲に支えられていなければ、継続的に作業にあたっていくことは難しい。一方、本プロジェクトの強みは、異なる作成者の目および発想により、それぞれに異なる種々の場面や話題が例文に設定されるというバラエティの豊かさにあると考えられる。ごく少数の作成者によって短期間に良質の例文を多量に生み出すことは現実的に考えて困難であり、メンバーが限定的になれば、場面や話題の広がりも生まれにくくなる可能性がある。本プロジェクトで例文作成にあたっているのは全員現職の日本語教師であり、例文作成の一定の質が確保されることに加え、それぞれに異なる環境で職務にあたっており、その発想の広がりについてはある程度確保されているものと考えられる[81]。

7. 例文の音声収録

「Jreibun」では、例文の一部に読み上げ音声も付すこととしており、本節では2022年度に試みとして行った録音・編集作業について報告を行う。

7.1　音声収録例文数と録音機材

2021年度作成の例文のごく一部のものであるが、2022年度に音声収録を行った。収録した例文数は174文である。例文の読み上げは第2節注7に記した通り、フリーアナウンサー・日本語教師の梅田エリカ氏（女性1名）に依頼し、共通語アクセントで読み上げてもらい、音声を収録した。収録は東京外国語大学の録音編集室で行った。当日の録音作業の時間は約3時間半であった。録音・編集については、録音編集室に備え付けのマイクやミキサーを使って行うことも可能であるが、試行の結果、収録した音声にどうしてもノイズが乗ってしまい、無音区間にノイズの入らないきれいな波形の音声を収録することが難しいことがわかった。そのため、音声収録には別途ノートPCとヘッドセットマイクロフォンを持ち込んで行うこととし、録音室は防音のためにのみ使用することとなった。また、録音にはノートPCにインストールされた音声分析ソフト「Praat」を使用し、サンプリング周波数48kHz、量子化ビット数16bitで収録した[82]。

7.2　読み上げ原稿と音声収録の手順

読み上げ原稿は、今回の例文音声収録作業のために選定された230の例文[83]が記載されたPDF

複数回の修正を経て（23）のような形となった。

[81] 例文作成の発想にどの程度の影響があるかは明確には言えないが、性別で言うと男性の例文作成協力者は2022年度に1名のみと少ない。

[82] 音声収録における一般的なフォーマットを用いた。

[83] 2022年度には、この230の例文のうち174例文の音声を収録した。引き続き残りの例文および他の例文についても、順次2023年度に音声録音を継続して行っている。

ファイル（全 34 ページ）である。図 2 に示すように、ファイルの 1 列目には、左から右に、読み上げ例文の「通し番号」（1 ～ 230）、例文作成のきっかけとする項目の「項目番号」、項目中の「例文番号」、例文作成のきっかけとなる「項目」、項目の「読み方」、作成された「例文」、例文の「ふりがな」の情報が記載され、2 列目から読み上げ対象となる例文が記載されている。本ファイルを事前に梅田氏に送付し、アクセントやイントネーションの確認、発話の速度調整等を行ってもらい[84]、収録当日を迎えた。防音室には、梅田氏の他、研究代表者の鈴木、研究分担者の伊達が入り、梅田氏の読み上げに合わせて、PC 上での録音操作、ファイルの保存などの作業を行った。また、読み上げにおいて、例文や語句の読み飛ばしがないかチェックし、また、アクセントやイントネーションの適切性を適宜協議し、読み直しが必要な場合は再度当該例文の録音を行った。

通し番号	項目番号	例文番号	項目	読み方	例文	ふりがな
1	3	01	あいかわらず	あいかわらず	高校の同窓会で友だちと久しぶりに会ったが、相変わらず元気そうだった。	高校の(同窓会;どうそうかい)で(友だち;ともだち)と(久しぶり;ひさしぶり)に会ったが、(相変わらず;あいかわらず)元気そうだった。
2	11	01	愛する	あいする	父は娘に本当に愛する人と結ばれて幸せになってほしいと願った。	父は(娘;むすめ)に本当に愛する人と結ばれて幸せになってほしいと願った。
3	608	01	愛おしい	いとおしい	踏まれてもたくましく生きる雑草に愛おしさを感じる。	(踏まれても;ふまれても)たくましく生きる(雑草;ざっそう)に(愛おしさ;いとおしさ)を感じる。
4	610	01	営み	いとなみ	厳しい自然の営みから学ぶことは多い。	(厳しい;きびしい)自然の(営み;いとなみ)から学ぶことは多い。
5	611	01	営む	いとなむ	私の両親は、田舎で農業を営んでいる。	私の両親は、(田舎;いなか)で農業を(営んで;いとなんで)いる。

図２　例文の読み上げ原稿の一部（1 ページ目冒頭）

7.3　音声ファイルの保存と切り出し作業および微調整

読み上げ音声は、7.2 節で述べた読み上げ原稿 PDF ファイルの各ページ（1 ページに 6 ～ 9 例文程度の記載）を読み終わるごとに、「01.wav」などの名前（ページ番号 .wav）を付けて保存した。読み直しを行った場合は、「01 1101.wav」などの名前（ページ番号 項目番号・例文番号 .wav）で別ファイルを保存した。

収録後、1 ページごとに作成された音声ファイルは、作業補佐に 1 例文ごとの切り出しを依頼した[85]。作業にあたって、作業補佐には Praat 上での音声の切り出し方、保存方法等を事前に教示した。切り出すファイルは発話の音声波形の開始と終了の前後に 500ms の無音区間を確保して保存するようにした（図 3 を参照）。このようにして、25 ページ分の音声ファイルから 174 の例文ごとの個別ファイルが作成された。ファイルは「3_01_jp_f.wav」などのファイル名（項目番号 _ 例文番号 _ 日本語 _ 女声 .wav）で保存された。

[84] 声の明るさ（トーン）についても、読み上げサンプルを送ってもらった上で、落ち着いたトーンで読み上げてもらえるよう、研究代表者から依頼した。

[85] 今回の録音分について、切り出し作業には約 13 時間を要した。

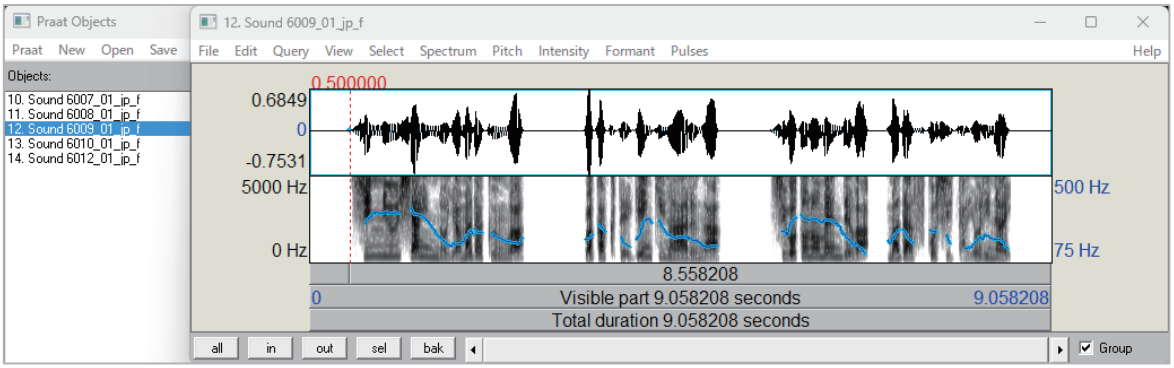

図 3　Praat 上での 1 文ごとの例文音声切り出し作業の様子 [86]

作業補佐から切り出しを終えた WAV ファイルを受領した後、音声編集ソフト「Audacity」用いて音量の調整等（DC オフセットの削除、最大振幅の正規化 [87]）の一括処理を行い（図 4 を参照）、実際のデータ提供の利便性を考え、MP3 形式に圧縮したデータを作成した。なお、例文の音声収録は 2023 年度も引き続き行っていくこととし、進めている。

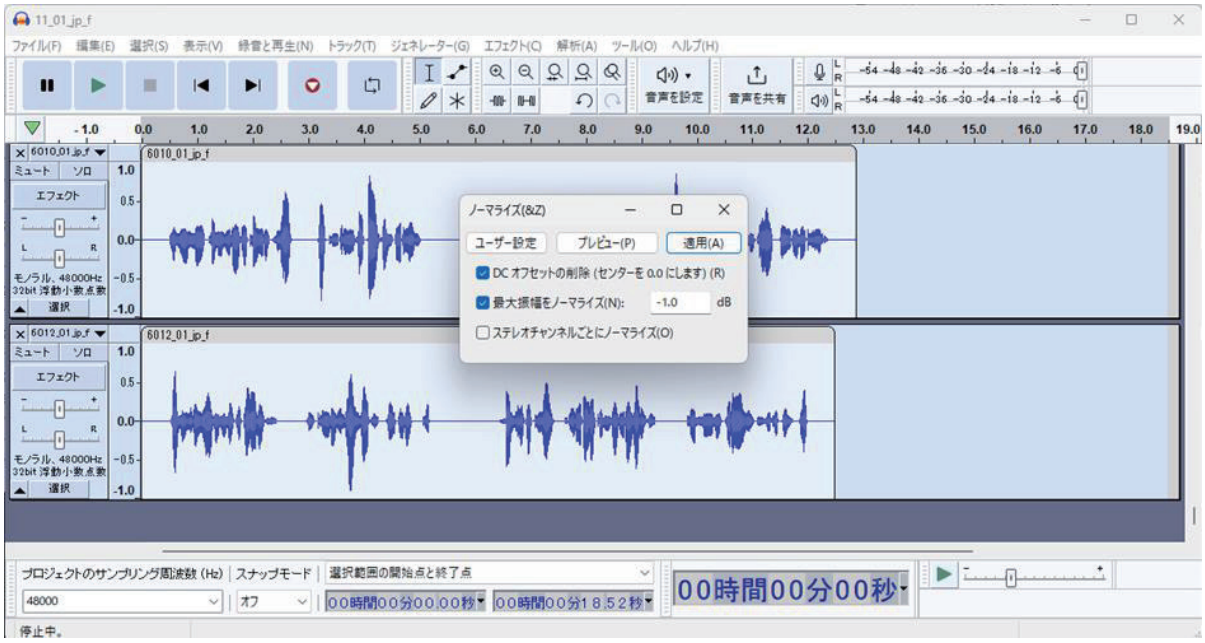

図 4　Audacity 上での録音ファイルに対する DC オフセットの削除と最大振幅の正規化

8. まとめと今後の課題

本稿では、「日本語例文バンク」（Jreibun）プロジェクトの 2021 ～ 2022 年度における進捗状況について中間報告を行った。まず、プロジェクトの実施体制および例文作成のきっかけとする約

[86] 項目「頼りない」、例文「新入社員の山田君は、若くて元気があるが、仕事の面ではまだ少し頼りない」の音声波形（上段・黒）、スペクトログラム（下段・白黒の濃淡）、ピッチパターン（下段・青線）、波形開始前の無音区間 0.5 秒（枠外左上・赤の数字）が示されている。ファイル名は「6009_01_jp_f.wav」である。

[87] DC オフセットとは、マイクなどで PC 内に音を収録する際に、周囲の電気的な影響で直流電流（Direct Current）の信号が入力されてしまうことによって生じる、波形の基準位置からのずれのことである。波形の上下のずれを中心に戻す処理を行ってから、各ファイルの最大音量を揃える調整を行う。

10,000 項目の選出について、その事前準備も含め手順を詳細に記した。

そして、2021 ～ 2022 年度にかけて作成された約 9,000 の例文について、その使用語彙および観察される特徴的な文パターンを通じて検討を行った。作成された全例文には、異なり語数で約 16,000 語、延べ語数で約 247,000 語が用いられており、使用語彙は特に品詞等には偏りなくバランスの範囲内にあることがわかった。また使用数の多い語を見ることで、よく用いられている文のパターンや取り上げられている話題なども見えてくることがわかった。また、作成された全例文を対象に Ngram を作成することにより、例文によく用いられている表現の傾向もつかむことができた。今後、例えば作成された例文を「話題」の観点から分類・分析してみることなども、例文の質を検討する 1 つの切り口となると思われる。これについては、次段階の課題としたい。

例文作成の手順と注意している点についても、プロジェクト初年度より継続して参加している研究協力者により、あらためて報告を行った。さらに作成例文には一部、読み上げ音声も付すこととしており、2022 年度に行った例文読み上げ音声の録音・編集作業についてもその詳細を報告した。予算等の制約により、全例文に音声を付すことは、本プロジェクト終了の 2024 年度までに終えることは難しいが、鈴木他（2022）でも述べられている通り、「Jreibun」は「成長するデータベース」として位置付けられるものあり、本プロジェクトの 4 年間のみですべて結果が固定されるものではない。項目や例文の追加、差し替え、改良等は随時可能であり、読み上げ音声についても、今後、読み上げ者の性別を変えたり、また関西方言等のアクセントによる読み上げ音声を取り入れるなど、各種ツール開発者のフィードバックも受けながら、バリエーションを追加していくことも考えられる。

よく考えられた質の良い例文を人の手により作成し、それをデータベース化していくというのは、労力と時間とを必要とする作業である。しかし、それがこのデータベースの質を確保することにつながる基盤の部分である。本プロジェクトは、例文を作成するメンバーをはじめとして、関係するすべての人の協力のもとに成り立っている。後半の 2 年間においても、引き続き質の良い例文データベースを充実させていくことに邁進していくこととしたい。

（執筆分担：1, 2, 3, 4, 8, および 6 の初めと終わり，全体の調整 鈴木、5 清水、6 渋谷、7 伊達）

■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、望月源）

■本稿の執筆にあたっては、本科研の研究分担者の望月源氏より、作成した全例文の形態素解析および必要な処理を行ったデータの提供を受けている。
■作成例文数の管理等の補助作業は、2021 年度は項目リスト調整作業とあわせ、韓金柱氏に、2022 ～ 2023 年度は東京外国語大学大学院博士後期課程の林燕燕氏に行ってもらっている。

引用文献

鈴木智美（2023）「例文作成の工夫と中間報告」日本語例文バンク Jreibun 第 2 回公開研究会（報告資料）（2023 年 3 月 4 日オンライン形式）

鈴木智美・清水由貴子・中村彰・加藤恵梨（2022）「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－」『東京外国語大学国際日本学研究』第 2 号　pp.191-208

鈴木智美・中村彰（2023）「日本語例文バンク（Jreibun）における例文の質の向上と英訳の工夫」『東京外国語大学国際日本学研究』第 3 号　pp.122-139

〈辞書・教材類〉

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）

姫野昌子監修（2012）『日本語コロケーション辞典』研究社

山根智恵監修（2020）『日本語口語表現辞典』第 2 版 研究社

〈参考資料〉

「日本語学術共通語彙リスト」（松下達彦研究室）http://www17408ui.sakura.ne.jp/tatsum/list.html

『日本語能力試験出題基準【改訂版】』国際交流基金・日本国際教育協会編著 凡人社（2002 年）

「留学生のための基礎的専門語」早稲田大学 小宮千鶴子研究室 http://www.gsjal.jp/komiya/spterms.html

〈ウェブサイト〉

「現代日本語書き言葉均衡コーパス」（BCCWJ）国立国語研究所 https://clrd.ninjal.ac.jp/bccwj（検索アプリケーション「中納言」を使用）

「Jisho.org」https://jisho.org

「NINJAL-LWP for BCCWJ」https://nlb.ninjal.ac.jp

〈ツール・アプリ〉

「Audacity」https://www.audacityteam.org/

「Mecab」https://taku910.github.io/mecab/

「NEologd」https://github.com/neologd/mecab-ipadic-neologd

「Praat」https://www.fon.hum.uva.nl/praat/

「UniDic」https://clrd.ninjal.ac.jp/unidic/

（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日本語日本文学科 准教授）

（しぶや　ひろこ　クリエイティブ日本語学校 校長・教務主任）

（だて　ひろこ　東京外国語大学大学院国際日本学研究院 准教授）

【資料 1】「Jreibun」が例文作成のきっかけとする項目のうち、「文法」項目にあたるもの

（116 項目）

文法項目	Jreibun項目（読み方）	補足
～とあいまって	あいまって	
～にあたって	あたって	複合助詞
～あたり	あたり	1日あたり～など
あまりにも	あまりにも	
～如何によって	いかんによって	
～ていく	いく	補助動詞
～以上	いじょう	～する／した以上、…
～ていただけませんか	いただく	「させていただく」は別項目あり
～の上から	うえから	費用の上から考えると／数の上から言えば　等
～上で	うえで	条件　～た／～の上（で）　よく考えた上で返事する　等
～上で・上では	うえで・うえでは	健康を保つ上で必要だ／数の上では～だ
～うちに	うちに	朝の／若い／子どもが寝ているうちに～　等
～うちに	うちに	～ているうちに～（変化）
～うとする	うとする	～しようとしてきた／～しようとしたが／～しようとしている　など
～得る（える・うる）	える・うる	ありえる・えない、起こりうる・えない　など
～において	おいて	複合助詞
～に応じて	おうじて	複合助詞
～うと思う	おもう	～う／ようと思っている、～とは思っていない／思わない
～と思われる	おもわれる	
～が	が	接続助詞　…が、…。
～にもかかわらず	かかわらず	複合助詞
～から～にかけて	かけて	
～かどうか	かどうか	
～かねる	かねる	～することはできない
～かねない	かねない	～にならないとは言い切れない／～になるかもしれない
から	から	起点　「まで」は別項目あり
～からには	からには	～である以上は、当然～
仮に	かりに	仮に～ても／仮に～とすると　等
～きれない	きれない	数え切れない　など
～てくる	くる	補助動詞
～てくれる	くれる	補助動詞
心おきなく	こころおきなく	
こそ	こそ	Nこそ／～てこそ／～のことを思えばこそ　等
こと	こと	修飾語句を伴うもの　病気のこと／あなたのこと／何のこと　等
～ことがある	ことがある	（よく、たまに、時々）～ることがる／～ないことがある
～ことがある／ない	ことがある・ない	経験　～たことがある／ない
～ことができる	ことができる	
～ことにする	ことにする	～る／～ない／～た　ことにする
～さ	さ	高さ、重さなど　頻度の高いものを
さえ	さえ	極端な事柄、同類の事柄添加、十分条件　子供にさえわかる／風だけでなく雨さえ降ってきた／これさえあれば　等々
さえ	さえ	
～（さ）せていただく	させていただく	「～ていただけませんか」別項目あり
～ざる／ず	ざる・ず	否定　知られざる／絶えざる／～ざるを得ない／当たらずとも遠からず／～ずに
～とされる／されている	される・されている	一般的に～
～しか	しか	～しか…ない
～次第だ	しだいだ	努力次第、天気次第
実は	じつは	
～しめる	しめる	なさしめる、生ぜしめる、言わしめる等
主たる	しゅたる	
～に過ぎない	すぎない	それ以上ではない
すら	すら	
～そう	そう	推測　A／Vそうだ・な　恥ずかしそう／心配そう／ありそう／なさそう
～そう	そう	今にも起こりそうだの意　Vそうだ　泣きそう／取れそう　など　「Vそうに／もない」は別項目あり

～そうだ	そうだ	伝聞
～そうに／そうもない	そうにない・そうもない	
その上	そのうえ	接続詞
～たい	たい	希望・願望
たかだか	たかだか	
だからこそ	だからこそ	
～ために	ために	～る／Nのために　目的・利益になるように　原因「ために」は別項目あり
～ため（に）	ため（に）	原因・理由　接続詞「そのため」は別項目あり
誰も	だれも	「誰か」は別項目あり
都度・つど	つど	～のつど／Vつど
～につれて	つれて	複合助詞
～てある	てある	補助動詞
～ておく	ておく	補助動詞
～てはならない	てはならない	許されないことだ
～てみる	てみる	補助動詞
と	と	AとB
という	という	日本語で何と言いますか／～というN
～たところで	ところで	
～として	として	複合助詞
～としても	としても	地震が起きたとしても／間違ったとしても
どのぐらい	どのぐらい	
～とも	とも	二人とも　など
～にともなって	ともなって	複合助詞　～にともなって…（変化）
～を問わず	とわず	複合助詞
～なくして	なくして	Nなくして～ない
なくてはならない	なくてはならない	必須　なくてはならないものだ
～なくてもいい	なくてもいい	
～なければならない	なければならいない	
なり	なり	私なりに努力する　など
何時	なんじ	なんじ
～ばかり	ばかり	Vたばかり
～をはじめ	はじめ	～をはじめ、～をはじめとして
～はずがない	はずがない	可能性がゼロである　「はずだ」は別項目あり
～べきだ・べきではない	べきだ・べきではない	
～ほど	ほど	1時間ほど／山ほど／食べきれないほど／昨日ほど暑くない／すればするほど～／謙虚な人ほど～するものだ　など名詞「程は別項目あり
～まい	まい	打ち消し推量・打ち消し意志
まで	まで	（Aから）Bまで　他　多義多い
までに	までに	期限
も	も	助詞　AもBも　同類のものを並べる
も	も	8時間も／100人も　数量が多いこと
持って行く	もっていく	
持って帰る	もってかえる	
持って来る	もってくる	
もとより	もとより	
～ものだ・ものではない	ものだ・ものではない	傾向性・当為
～ものの	ものの	逆接
～てもらう	もらう	補助動詞
や	や	（並立詞）AやB
～や否や	やいなや	
～うと（も）しない	ようともしない	意志・様子が見られない
～ようだ	ようだ	不確かな断定／「～ような気がする」なども含める　「みたいだ」は別項目あり
～のようだ／な／に	ようだ	比況　まるで～
～のような／に	ような	例示　例えば～
～ように	ように	ご存じのように／以下のように　など
～ように	ように	～る／ないように　目的
～ようになる	ようになる	変化
～によって	よって	Nによって（Vられる）
～によって	よって	Nによって（異なる）
～らしい	らしい	どうも～らしい
～わけだ	わけだ	「わけではない」「わけにはいかない」別項目あり
～わけではない	わけではない	嫌いなわけではない　など
～わけにはいかない	わけにはいかない	
～にわたって	わたって	複合助詞

【資料 2】「Jreibun」の作成例文（2021 ～ 2022 年度）にて使用数の多い語（主要な品詞別）

使用数順位	動詞	形容詞	形容動詞 （名詞としての用法を持つものを含む）	副詞	名詞
1	する	ない・無い	必要	よく	こと
2	いる	いい・よい	様々・さまざま	いつも	の
3	なる	多い	好き	すぐ	よう
4	ある	高い	大切	なかなか	人
5	しまう	新しい	健康	まだ	私
6	できる	良い	可能	少し	ため
7	言う	悪い	自然	とても	時・とき
8	くる	やすい	重要	どう	中
9	見る	長い	きれい・綺麗	初めて	的
10	行う	難しい	心配	まず	もの
11	行く	大きい	簡単	あまり	者
12	思う	ほしい・欲しい	有名	突然	たち
13	出る	強い	自由	もう	日本
14	使う	美しい	危険	特に	子ども
15	いく	少ない	丁寧	しばらく	年
16	くれる	うまい・上手い・旨い・巧い	無事	しっかり	日
17	持つ	早い	不安	全く	家
18	食べる	にくい	大変	必ず	会社
19	いう	若い	豊か	ほとんど	一（いち）
20	受ける	おいしい	嫌・いや・イヤ	ずっと	そう
21	始める	忙しい	元気	ようやく	仕事
22	聞く	厳しい	大事	ちょっと	何
23	考える	寒い	新た	実際	自分
24	入る	広い	適切	つい	時間
25	もらう	小さい	幸せ	非常に	ところ
26	おる	低い	多様	一気に	大学
27	作る	激しい	面倒	ゆっくり	前
28	書く	遅い	平和	常に	友人
29	みる	幼い	いろいろ	最も	今
30	買う	暑い	高級	より まるで	後

Interim Report on "the Bank of Japanese Example Sentences" (Jreibun) Project:

A Comprehensive Analysis of Created Sentences from the Perspective of Vocabulary and Sentence Patterns

SUZUKI Tomomi, SHIMIZU Yukiko, SHIBUYA Hiroko, DATE Hiroko

KEYWORDS: "Jreibun" (the Bank of Japanese Example Sentences), selection of items (words and expressions), creation of example sentences, vocabulary used, Ngram, recording of example sentences

The purpose of this paper is to provide an interim report on the progress of "the Bank of Japanese Example Sentences" (Jreibun) project "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 to examine approximately 9,000 example sentences created during 2021-2022. This examination is conducted through an analysis of the vocabulary used and observed sentence patterns.

First, the paper provides detailed procedures, including the project implementation system and preparation for the selection of approximately 10,000 items that form the basis of example sentence creation.

The analysis reveals that the total sentences created use around 16,000 different vocabulary items and approximately 247,000 words in total. The vocabulary usage is well-balanced across various parts of speech. Examining frequently used words also reveals common sentence patterns and discussed topics. Furthermore, the creation of Ngrams based on all example sentences indicates the frequent use of expressions conveying the meaning of "decision," " preliminary expression," "occasional occurrences," "natural changes," "conscious habitual actions," "general statements," and "sentence-ending expressions indicating the speaker's judgment."

Taking into account previous points made by Suzuki and Nakamura (2023) regarding the considerations in creating example sentences, this paper reports on the actual process of sentence creation from the perspective of one of the project members. Additionally, some example sentences are accompanied by audio recordings, and a report on the recording and editing activities conducted in 2022 is provided.

Creating well-thought-out, high-quality example sentences through human effort, as well as databasing them is not an easy task. However, it forms the foundational aspect that ensures the quality of this database. For the remaining two years, we plan to continue enriching the database with high-quality example sentences.