

国立国語研究所学術情報リポジトリ

日本語学習者コーパス開発の歴史

メタデータ	言語: Japanese 出版者: 研究社 公開日: 2026-07-17 キーワード: 中間言語コーパス, 技能, 母語, 習熟度, W-CoLeJa 作成者: 石黒, 圭, 呉, 楚琦 メールアドレス: 所属: 国立国語研究所, 一橋大学大学院博士後期課程, 国立国語研究所プロジェクト非常勤研究員
URL	https://doi.org/10.15084/0002000626

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.



第2章 | 日本語学習者コーパス開発の歴史

学習者コーパスは、学習者が話したり書いたりした目標言語の使用を収集してデータベース化し、研究教育に活用するもので、産出された記号のみならず、産出主体である学習者という人間を重視する点に特徴がある。本章では、日本語教育の学習者コーパス開発の歴史を、タスクなどの技能、学習者の母語、目標言語の習熟度という三つの観点から振り返り、学習者コーパス開発史における『日本語学習者縦断作文コーパス (W-CoLeJa)』の位置づけについて検討した。

KEYWORD

中間言語コーパス | 技能 | 母語 | 習熟度 | W-CoLeJa

□1 学習者コーパスの特徴

コーパスとは、書き言葉（文章）や話し言葉（談話）のデータを体系的に収集してデータベース化した研究のための言語資料を指すが、コーパスのなかでも学習者コーパスは異彩を放っている。学習者コーパス (learner corpus) は、第二言語として目標言語を学んでいる人が、目標言語で話したり書いたりした**中間言語コーパス** (interlanguage corpus) だからである。

母語話者が話したり書いたりした一般的なコーパスは、自然な日本語であることが前提であり、誰が話したり書いたりしたかということとはあまり問題にならない。話し言葉と書き言葉の違いに代表される**レジスター** (register: 言語使用域) の異なり、現代語と古典語の違いに代表される時代的な異なり、共通語と方言の違いに代表される地理的な異なりといった観点から区分され、区分された領域で用いられている言語の使用実態を把握するのに使われる。もちろん、『日本語日常会話コーパス (CEJC)』の場合のように、年代差や性差という話し手の属性によって区分することも可能ではあるが、一般的なコーパスでは、記号としての言語の持つしくみや偏りを分析することに重きが置かれていると理解するほうが自然であろう。

一方、学習者コーパスの場合、もちろん、直接の研究対象は言語という記号であるが、そこで重視されるのは、言語の背後にいる「学習者」という言語使用者、すなわち人間である。なぜ学習者コーパスにおいて人間を重視するのかと言うと、学習者コーパスは単に言語の使用実態を問題にしているのではなく、言語の使用実態の把握をとおして学習者という人間の言語使用面での成長過程を研究し、その研究成果を教育に応用することを目指しているからである。人間が学習によって第二言語を身に着けていく発達過程を研究する分野は第二言語習得研究 (Second Language Acquisition: SLA) と呼ばれるが、学習者コーパスはこの第二言語習得研究に貴重な分析の材料を提供するものである。その意味で、学習者コーパスは、第二言語習得研究に基礎資料を提供することで習得過程を明らかにすると同時に、その研究成果を教授法、教材作成、試験開発、カリキュラム構築といった学習環境の設計に役立てることを目指すものである。

以上の検討からわかるように、学習者コーパスはコーパスの一種でありながら、記号としての言語のしくみや偏りを分析するだけではなく、人間の記号使用の背後に透けて見える習得上の成長や困難点を明らかにしようとする点で特異なコーパスであると言える。こうした学習者コーパスの特異性を踏まえたうえで、どのような学習者コーパスが開発され、どのように発展してきたのか、その歴史を振り返りたい (石黒 2023も参照)。ここで取り上げる学習者コーパスは、読者への便宜を図るため、現在でもWebサイト等で入手が比較的容易なものを中心に紹介する。

□2 学習者コーパスの展開

2.1 学習者コーパスが出現するまでの研究状況

コーパスは、書き言葉や話し言葉を体系的に収集してデータベース化した研究のための言語資料であり、コンピュータで処理されることを前提にデジタル化されている。しかし、昨今のようにコーパスが整備されるはるか以前から、日本語学習者の言語使用の研究は行われていた。学習者コーパスが出現する以前の研究はどんな方法で行われてきたのだろうか。

日本国内の戦後の日本語研究は、日本語学習者の誤用や質問を手がかりに発展してきたという特徴がある。誤用は、学習者が論理的に考えた結果生じたものであり、そこに他の言語とは異なる日本語の特徴が隠れていることが多い。また、学習者の質問は、母語話者が当たり前のものとして見すごしてきたことを学習者の視点から鋭く切りだすものである。そうした誤用や質問への的確な

対処を探索するなかで、日本語の記述文法は発展してきた。そうした研究を当時主導したのが寺村秀夫氏であり、その出発点となったのが、『寺村誤用例集データベース』として知られる科研費の報告書(寺村 1990)である。そのデータの総数は4601文、そのうち誤用を含む文は3131文である。寺村氏の急逝後、その志が大阪大学の仁田義雄氏・宮島達夫氏に受け継がれ、2003年から2010年にかけてまとめられた日本語記述文法研究会『現代日本語文法(全7巻)』(くろしお出版)が当時の記述文法の学術的成果の集大成であり、その後も記述文法の研究は連綿と続いている。

しかし、細かな言語現象の記述が進めば進むほど、日本語学としての精緻化には貢献するものの、学習者の役には立たないものとなり、日本語学と日本語教育学の乖離が進行する。21世紀に入り、学習者の習得に資する**教育文法**(*pedagogical grammar*)を別に構想する必要があること(小林 2002; 野田編 2005など)、寺村(1990)のような母語や習熟度の統制が取れていない雑多なデータベースでは、学習者の第二言語習得過程が明らかにできないことといった理由から、学習者の日本語を研究するための新たなデータベースの作成が模索されるようになった。同時に、ICT技術の向上により、各種母語話者コーパスが開発されるようになると、日本語教育の世界でも、第二言語の習得過程の解明につながる学習者コーパス開発の機運が高まってきた。

学習者コーパスは、学習者が第二言語について何をどこまで習得しているか、その習得過程においてどのような発達段階を経るのかを明らかにする第二言語習得研究の基礎資料となるデータベースである。第二言語習得研究に十分に貢献しうる学習者コーパスを開発する場合、次の三つの観点からコーパスの設計を考えることが重要になる。第一は技能、第二は母語、第三は習熟度である。以下では、この三つの観点から学習者コーパスの展開について整理する。

2.2 技能から見た学習者コーパスの展開

コーパス開発のさいの第一の重要な観点は、**技能**、いわゆるタスクである。技能は、聞く・話す・読む・書くの四技能に代表される、学習者の言語活動に基づくものであり、レジスター(言語使用域)と深い関わりを持つ。学習者コーパスを開発する場合、漫然と学習者の言語使用のデータを収集するのではなく、学習者が第二言語を使用する具体的な場面を想定し、そこで用いられる特定の技能に絞ることが必要である。CEFR(ヨーロッパ言語共通参照枠)をはじめ、多くの言語参照枠で技能別の観点が取り入れられていることからわかるように、

第二言語能力を正確に測ろうとした場合、言語活動をベースにした技能別に言語能力を考えるのが現実的であり、その言語を使って何がどこまでできるかというCan-do記述 (can-do statements) が妥当であると考えられる (塩澤 2010; 文化庁 2023など)。

一般的なコーパスが話し言葉のコーパスと書き言葉のコーパスに大別されるように、学習者コーパスにおいても話し言葉のコーパスと書き言葉のコーパスに大別されることが多い。さらに細かく分けると、話し言葉では、対話 (dialogue) のコーパスと独話 (monologue) のコーパスの二つに分かれ、前者は雑談、話し合い、ロールプレイなど、後者は発表、スピーチ、ストーリーテリングなどが試みられている。

公開されているコーパスのうち、独話のデータは少ないが、対話のものは比較的多い。従来よく使われてきたコーパスに、上村隆一氏が開発し、10を超える言語の母語話者の会話を集めた横断コーパス『上村コーパス』、鎌田修・山内博之の両氏がOPI (Oral Proficiency Interview : ACTFLによって開発された言語運用能力を測る会話テスト) のインタビュー手法を生かして開発した、中国語・英語・韓国語話者のデータを収録した『KYコーパス』の二つがある (上村 1997; 鎌田 2006 を参照)。ただし、『上村コーパス』は現在では入手不可能なので詳細は割愛し、『KYコーパス』については2.4節で紹介する。

近年では、会話、ロールプレイ、ストーリーテリング、絵描写といった多様なタスクを含むコーパスである『多言語母語の日本語学習者横断コーパス (I-JAS)』を用いた研究が多いが、こちらは2.3節であらためて紹介する。また、俵山編 (2023) には『課題別会話コーパス』が用いられ、伝言、訂正、弁明、説明、アドバイス、話し合い、コメントといった多様な課題を含む談話データが用いられており、会話の多様な機能別分析が可能になっている。

一方、書き言葉では、物語文 (narrative text) と説明文 (expository text) に分けられ、前者は生活文や体験文など、後者は意見文や学術レポートなどのコーパスがそれぞれ存在する。書き言葉は、「作文コーパス」という名称のもとで収集されることが多く、収集が比較的容易であるため、博士論文執筆のさいに個人が収集しているものが相当数存在すると見込まれる。なかには小規模ではあるが、貴重と見られるデータも散見されるが、そのほとんどが公開されておらず、埋もれてしまっているという残念な現状となっている。今やオープンサイエンスの時代であり、第三者による論証過程の適切な再検証ができるよう、調査協力者の書面による同意を得、個人情報の匿名化などの処理を適切に行った

うえでオープンデータとして公開されることが望まれる。

こうしたなか、タスク別の書き言葉コーパスとして著名なのが、当時横浜国立大学に在籍していた金澤裕之氏が中心となって収集した『YNU書き言葉コーパス』(金澤編2014)である。メール文がもっとも多いが、メール文以外にもレポート、新聞、広報誌などの媒体を想定し、説明、紹介、意見、依頼などを行う12種類のタスクが、日本語話者、韓国語話者、中国語話者各30名によって書かれており、計1080本の文章が収録されている。メール文ならば、『中国語・韓国語を母語とする日本語学習者縦断的メール文コーパス』や、後述のI-JASにも一部含まれている。

また、400字から800字前後の短い作文が多い現状を踏まえ、説明文・意見文・歴史文という3ジャンルにおいて、JLPT のN1相当以上の力を持つ学習者に2000字以上の分量の作文を書いてもらい、文章の結束性や文章構成を見た金井勇人氏らの『JCK作文コーパス』もある。『JCK作文コーパス』を用いた研究成果については石黒圭編(2017)を参照されたい。そのほか、『台湾人日本語学習者コーパス(CTLJ)』や『華東政法大学作文コーパス』も比較的入手が容易な中国語母語話者の作文コーパスである。

2.3 母語から見た学習者コーパスの展開

コーパス開発のさいの第二の重要な観点は、**母語(native language)**である。学習者が第二言語を習得する場合、母語の知識を前提とすることが多く、母語によって習得のプロセスが異なることが少なくない。日本語が学習者一般にとって難しい言語かどうかという質問はあまり意味がなく、学習者の母語との言語間距離が大きな意味を持つ。文法が日本語と近い韓国語話者の場合、韓国語の文法知識を生かせるので、学びやすい言語となるだろうし、漢字という共通の表記を持つ中国語話者の場合、中国語の漢字知識を生かせ、やはり学びやすい言語となろう。しかし、そうでない言語を母語とする学習者の場合、日本語の学習は相対的に困難になるにちがいない。学習者は母語から出発して、徐々に目標言語(target language)に近づけていくという**中間言語(inter-language)**の考え方が浸透したことで、母語の重要性がより強く認知されるようになった。もちろん、母語によらず、多くの言語話者に共通して難しい日本語の特徴は存在し、母語だけで難しさが決まるわけではないが、目標言語の習得は母語による影響をかなりの程度受けるため、母語を考慮したコーパスの設計が重要になる。事実、現在開発されているもので母語への配慮がない学習者

コーパスは皆無であり、それだけ母語の重要性が浸透してきたと言える。

多様な母語を背景とする代表的なコーパスが、第1章執筆担当の迫田久美子氏を中心となって国立国語研究所在職中に開発したI-JASである。世界各地で日本語を学習している、12の言語（中国語・韓国語・英語・フランス語・ドイツ語・ロシア語・ハンガリー語・インドネシア語・スペイン語・トルコ語・ベトナム語・タイ語）を母語とする1,000名もの学習者横断コーパスである。この巨大なコーパスの登場により、日本語学習者の言語使用に見られる誤用や偏りが、母語に由来するものなのか、学習者に共通するものなのかといった判断の精度が飛躍的に向上し、第二言語習得研究に大きな影響を与えている（迫田・石川・李編 2020）。

一方、特定の母語話者に集中したコーパスも存在する。なかでも多いのが中国語話者のコーパスであり、たとえば、書き言葉では、『YUKタグ付き中国語母語話者日本語学習者作文コーパス』（Ver.14）は56校の大学において中国語母語話者の日本語作文のみを収集したコーパスであり、5,118本という非常に多くの作文が収録されている。話し言葉では、『魯東大学会話コーパス』のほか、後述する縦断学習者コーパス『LARP@SCU語料庫』『湖南大学学習者中間言語コーパス』『北京日本語学習者縦断コーパス（B-JAS）』はいずれも中国語話者のコーパスである。これにたいし、『住みやすい国コーパス』はヨーロッパで収集されたコーパスで、話し言葉編ではドイツ語話者のものが、書き言葉編ではドイツ語・セルビア語・クロアチア語・スロベニア語・ボスニア語・日本語話者のものが公開されている。

また、学習者が日本語で書いた内容を母語でどう表現するのかを合わせて示した、宇佐美洋氏を中心となって国立国語研究所在職中に開発した『作文対訳データベース』も長く用いられてきた。1,754本の作文が収録され、執筆した学習者の国籍は21ヶ国に及び、作文執筆者による母語訳（またはもっとも楽に書ける言語への翻訳）を含むものである。このような母語対訳という発想も、目標言語にたいする母語の影響を考える意味で有用だと思われる。

なお、今後のコーパス開発を考える場合、母語という考え方が現実と整合しなくなっている点を指摘しておきたい。その一つは、世界中の非英語圏の学習者は第二言語（second language: L2）として英語を学習することが多く、その影響が看過できないという点である。このため、目標言語の習得を第三言語（third language: L3）習得として捉える研究が増えてきている。もう一つは、世界のグローバル化にともない、複言語話者（multilingual）が急増している点である。単一母語を想定することが難しいバイリンガル、トリリンガルなどの複

言語話者が少なくなく、母語の影響自体が複雑になっている学習者も少なくない。今後、こうした現実を前提に設計された学習者コーパスの出現が強く望まれる。

2.4 習熟度から見た学習者コーパスの展開

コーパス開発のさいの第三の重要な観点は、目標言語の**習熟度 (proficiency level)**、いわゆるレベルである。学習者コーパスを開発する場合、異なる習得レベルの学習者の比較ができるよう、学習者の目標言語での習熟度を考慮する必要がある。学習歴や試験の成績などによって学習者の第二言語能力を初級、中級、上級などと分け、データを収集してコーパスを構築するのが一般的である。

話し言葉のコーパスで、学習者の習熟度について深く考慮しているのは、すでに見た『KYコーパス』である。『KYコーパス』は、OPIの手法に基づいたコーパスであり、熟達度による能力レベルが明示されている点に特徴がある。90名分の会話データを文字化した言語資料であり、母語別に見ると、中国語、英語、韓国語がそれぞれ30名ずつ、OPIの判定結果別の内訳は、それぞれ初級5名、中級10名、上級10名、超級5名となっている。インタビューの構成がしっかりとおり、データ同士の比較が容易である点、発話単位の認定が比較的容易であり、数量化・定量化を行いやすい点で優れているとされる。『KYコーパス』と同様に、OPIの手法を生かして収集されたコーパスとして、国立国語研究所の野山広氏らが開発した390名対象のインタビューデータ『日本語学習者会話データベース』があり、大学生、日本語学校生などのいわゆる学生だけでなく、会社員や公務員、研修生などの社会人が含まれている点に特徴がある。

一方、書き言葉のコーパスで、習熟度の整理が行き届いているのは、東京外国語大学留学生日本語教育センターが構築した『JLPTUFS作文コーパス』であろう。異なるレベルの留学生を豊富に抱える外国語大学ならではの強みを生かし、55の国・地域出身の、入門～初級、初級後半～初中級、初中級、中級前半、中級後半、上級前半、上級後半、超級の8レベルにわたる学習者の1,515本の作文を収録している。これによって、どのような母語のどのような日本語能力の学習者が、どのような作文を書くのか詳しくわかるようになっている。

習熟度を考える場合、似た環境にある学習者を集め、学習歴や試験の成績によって学習者をいくつかの群に分けてデータを収集し、習熟度別に習得状況を見ていくのが一般的である。しかし、その場合、別の人物同士の比較となり、

学習者の成長を分析したくても、見かけの成長しか比較できない。つまり、学習者の発達過程を正確に追いかけるには、同一人物の成長を経年的に追跡する必要がある。そうしたコーパスは**縦断コーパス (longitudinal corpus)**と呼ばれ、数は少ないが収集・公開されているものがある。

縦断コーパスの先駆けは東呉大学の『LARP@SCU語料庫』であり、37名（最終的には26名）の学習者に3年半のあいだ、毎月作文と、作文についてのインタビューを行ったものである。また、名古屋大学の杉村泰氏を中心となって関わった『湖南大学学習者中間言語コーパス』もよく知られており、日本語学科の中国語話者の日本語会話（計7回）と日本語作文（計19回）を94名分、やはり3年半のあいだ収集・構築したものである。国立国語研究所在職当時の迫田久美子氏と筆者が中心となった『北京日本語学習者縦断コーパス (B-JAS)』は、日本語学習経験がゼロないしは初級の段階で日本語学科に入学し、4年間日本語を集中的に学んで卒業した北京師範大学の学部学生17名を対象に、8回のインタビュー調査、および11回の作文調査を継続的にを行い収集したものである。I-JASに準拠し、I-JASと組み合わせて分析できる点に特徴があり、研究成果は林・徐・迫田主編（2024）にまとめられている。なお、『中国語・韓国語母語の日本語学習者縦断発話コーパス (C-JAS)』は、中国語母語話者・韓国語母語話者各3名を約3年追跡した話し言葉コーパスであり、後のI-JAS、B-JASの着想のもととなった、迫田久美子氏が博士論文執筆に用いたデータである。

2.5 そのほかの個性的な学習者コーパスの展開

以上、技能、母語、習熟度の観点から学習者コーパスを概観してきたが、それとは異なる観点からの興味深いコーパスもあるので、簡単にご紹介したい。

一つ目は、**添削タグ**のあるコーパスである。『寺村誤用例集データベース』の時代から、日本語学習者の作文は誤用に注目されてきた。しかし、誤用は判定者によっても自然さの判断は異なるし、学習者の日本語レベルによっても判断は揺れる。また、誤用をどう直すか、どこまで直すかも、判定者によってその判断は揺れることも多い。したがって、誤用のアノテーションを行うのは、困難を伴う決断であるとも言える。そうした研究の草分け的存在である『学習者作文コーパス「なたね」』は、285本の作文数にたいし、3名の日本語教師によって6,775件の誤用タグを付与されており、学習者の誤用を検討するうえで、また、判定者による誤用の判定基準を検討するうえで参考になる。このほか、既述の『作文対訳データベース』にも添削データが付与されており、その添削

データに誤りタイプの情報をXML形式で付与して機械学習による自動分類を可能にした研究も存在し(大山・小町・松本 2016)、誤用の定量的分析への応用が期待される。さらに、『ICJSデータサイト』で提供されている作文データにも誤用情報が付与され、「日本語誤用辞典」としてサイト上での検索が可能のほか、既述の『YUKタグ付き中国語母語話者日本語学習者作文コーパス』にも、5,118の作文数にたいし、122,179の誤用タグが付与されており、豊富な誤用情報が提供されている。

二つ目は、**接触場面**が豊富なコーパスである。宇佐美まゆみ氏が中心となって開発した『BTSJ1000人日本語自然会話コーパス』は、BTSJと呼ばれる、談話分析に適した基本的文字化の原則に従って構築されたコーパスで、母語話者同士の会話のみならず、学習者同士、さらには母語話者と学習者の接触場면을豊富に取り入れている点に特徴がある(宇佐美編 2020a, 2020b など)。また、初対面同士の会話と友人同士の会話の区別もあるため、話し手と聞き手の親疎関係によって、さらには母語話者と学習者の接触場面において、ボライトネスがどのように変化するかを見ることができるコーパスとなっている。また、『日中Skype会話コーパス』は日本国内と中国国内の学部生同士による接触場面のオンライン会話を収録したもので、2012年に収録されたデータである。コロナ禍以降、オンラインコミュニケーションは一般的になっているが、当時としては画期的な、先駆的な会話コーパスと見ることができる(中俣ほか 2013)。

三つ目は、産出ではなく**理解**に焦点を当てたコーパスである。『日本語非母語話者の読解コーパス』『日本語非母語話者の聴解コーパス』がそれに当たる(野田 2019, 2020, 2024, 2025 など)。これらは、日本語学習者が読んだり聞いたりした内容をどう理解したのか、その理解過程を母語で話してもらった記録であり、必要に応じて質問に答えてもらった記録でもある。読んだり聞いたりといった理解は、書いたり話したりといった産出とは異なり、学習者の脳内で起きる現象であり、通常外部から観察することはできない。しかし、学習者に理解活動をメタ認知してもらい、母語に翻訳して話してもらう発話プロトコル法を用いることで観察可能にした、きわめて斬新な着想に基づく調査である。

ここまで概観してきた主要な学習者コーパスの特徴を、「技能」「母語」「習熟度」という本章の三つの観点から整理すると、次の表1ようになる。

表1 主要な日本語学習者コーパス

コーパス・データベース名	話し言葉	書き言葉	母語の多様性	習熟度縦断性
寺村誤用例集データベース	－	○	○	－
上村コーパス	○	－	○	－
KYコーパス	○	－	○	△
課題別会話コーパス	○	－	○	－
YNU書き言葉コーパス	－	○	△	－
中国語・韓国語を母語とする日本語学習者縦断的メール文コーパス	－	○	△	△
JCK作文コーパス	－	○	△	－
台湾人日本語学習者コーパス (CTLJ)	－	○	－	－
華東政法大学作文コーパス	－	○	－	－
多言語母語の日本語学習者コーパス (I-JAS)	○	○	○	△
YUKタグ付き中国語母語話者日本語学習者作文コーパス	－	○	－	－
魯東大学会話コーパス	○	－	－	－
住みやすい国コーパス	○	○	○	△
作文対訳データベース	－	○	○	－
日本語学習者会話データベース	○	－	○	－
JLPTUFS作文コーパス	－	○	○	△
LARP@SCU語料庫	○	○	－	○
湖南大学学習者中間言語コーパス	○	○	－	○
北京日本語学習者縦断コーパス (B-JAS)	○	○	－	○
中国語・韓国語母語の日本語学習者縦断発話コーパス (C-JAS)	○	－	△	○
学習者作文コーパス「なたね」	－	○	○	－
ICJSデータサイト 日本語学習者コーパス	－	○	○	△
BTSJ 1000人日本語自然会話コーパス	○	－	○	－
日中Skype会話コーパス	○	－	－	△
日本語非母語話者の読解コーパス	－	○	○	－
日本語非母語話者の聴解コーパス	○	－	○	－

□3 総合学習者コーパスCoLeJaの誕生

日本語学習者のコーパス構築の歴史を振り返ってみると、技能、母語、習熟度の三つの観点を踏まえ、多岐にわたる個性的なコーパスがこれまで作られてきたことがわかる。コーパスの開発はきわめて大きな予算と労力がかかる負担の大きい作業であり、先達の熱意と努力には圧倒される。

しかし、上掲の表1から明らかなように、上記の三つの観点を十全に満たすコーパスはいまだ存在せず、その三つの観点を併せ持つ総合コーパスの登場が期待される状況である。そこで、話し言葉も書き言葉も幅広く含む「技能」、世界の多様な話者を含む「母語」、縦断調査によって学習者の成長を扱える「習熟度」の三つの観点を同時に満たす目的で、日本語学習者コーパスCoLeJa (Corpus of Learner Japanese : 通称コレジャ) の開発に着手することにした。日本語学習者コーパスCoLeJaは、石黒がリーダーを務める国立国語研究所の研究プロジェクト「多様な言語資源に基づく日本語非母語話者の言語運用の応用的研究」が開発を担っており、書き言葉である作文を中心に扱うW-CoLeJaと、話し言葉である会話を中心に扱うS-CoLeJaに大きくは分かれる。本書では全体をつうじて、前者W-CoLeJaを取り上げる。まずは、その概要を次の第3章で示し、そのうえで、W-CoLeJaの設計理念と開発過程を第4章から第15章において深く掘り下げ、詳しく解説していくことにしたい。

参考文献

- 石黒圭 (2023)「学習者はどのように日本語を学ぶのか—ポストコロナ時代における日本語運用データの収集法—」『東アジア日本学研究』9: 5-20.
- 石黒圭 (編) (2017)『現場に役立つ日本語教育研究3 わかりやすく書ける作文シラバス』くろしお出版.
- 上村隆一 (1997)「データベースで調べる」『日本語学』16 (12): 60-68, 明治書院.
- 宇佐美まゆみ (編) (2020a)『日本語の自然会話分析—BTSJコーパスから見たコミュニケーションの解明—』くろしお出版.
- 宇佐美まゆみ (編) (2020b)『自然会話分析への語用論的アプローチ—BTSJコーパスを利用して—』ひつじ書房.
- 大山浩美・小町守・松本裕治 (2016)「日本語学習者の作文における誤用タイプの階層的アノテーションに基づく機械学習による自動分類」『自然言語処理』23-2: 195-225.
- 金澤裕之 (編) (2014)『日本語教育のためのタスク別書き言葉コーパス』ひつじ書房.
- 鎌田修 (2006)「KYコーパスと日本語教育研究」『日本語教育』130: 42-51.
- 小林ミナ (2002)「日本語教育における教育文法」『日本語文法』2-1: 153-170.
- 迫田久美子・石川慎一郎・李在鎭 (編) (2020)『日本語学習者コーパスI-JAS入門 研究・教育にどう使うか』くろしお出版.
- 塩澤真季・石司えり・島田徳子 (2010)「言語能力の熟達度を表すCan-do記述の分析—JF Can-do 作成のためのガイドライン策定に向けて—」『国際交流基金日本語教育紀要』6: 23-39.
- 俵山雄司 (編) (2023)『現場に役立つ日本語教育研究4 自由に話せる会話シラバス』くろしお出版.
- 寺村秀夫 (1990)『外国人学習者の日本語誤用例集』平成2年度科学研究費特別推進研究報告書.
- 中俣尚己・漆田彩・小野真依子・北見友香・竹原英里 (2013)「Skypeを活用した日中会話

交流プログラム』『実践國文學』83: 25–48.

日本語記述文法研究会 (2003–2010)『現代日本語文法 (全7巻)』くろしお出版.

野田尚史 (2019)『読んで理解する過程の解明—「読解コーパス」の開発—』野田尚史・迫田久美子 (編)『学習者コーパスと日本語教育研究』23–42, くろしお出版.

野田尚史 (2020)『日本語学習者の読解過程の研究手法と研究課題』野田尚史 (編)『日本語学習者の読解過程』5–24, ココ出版.

野田尚史 (2024)『日本語学習者の聴解困難点と推測技術』『日本語プロフィエンス研究』12: 42–56.

野田尚史 (2025)『日本語学習者による日本語の理解過程—理解困難点と推測技術—』『語用論研究』26: 1–19.

野田尚史 (編) (2005)『コミュニケーションのための日本語教育文法』くろしお出版.

文化庁 (2023)『日本語教育の参照枠』『日本語教育コンテンツ共有システム』 https://www.nihongo-ews.mext.go.jp/information/framework_of_reference

林洪・徐一平・迫田久美子 (主編)、張林・野山広・石黒圭・岩崎拓也 (編集) (2024)『北京日语学习者历时语料库与日语教学 (北京日本語学習者縦断コーパス (B-JAS) と日本語の指導法)』中国国際广播出版社 (中国国際广播出版社).

コーパス情報

ICJS データサイト 日本語学習者コーパス (東京外国語大学の望月圭子氏と海野多枝氏を中心に開発) <https://icjs.jp/llc> (日本語誤用辞典はhttps://icjs.jp/ja_wrong)

上村コーパス (福岡工業大学在職当時の上村隆一氏が開発。現在は非公開)

学習者作文コーパス「なたね」(東京工業大学在職当時の仁科喜久子氏を中心に開発) <https://hinoki-project.org/natane/>

課題別会話コーパス (名古屋大学の俵山雄司氏を中心に開発。くろしお出版の「現場に役立つ日本語教育研究」ダウンロードページからダウンロード可能) <https://www.9640.jp/genba/>

華東政法大学作文コーパス (名古屋大学の杉村泰氏を中心に開発) <https://www.hum.nagoya-u.ac.jp/about/about-sub2/japanese/sugimura/2-1.html>

KY コーパス (京都外国語大学在職当時の鎌田修氏と実践女子大学の山内博之氏が開発。早稲田大学の李在鎬氏らによって「タグ付き KY コーパス」が公開されていたが、現在は公開されていない模様。ただし、山内博之氏にメールで問い合わせれば現在でも入手可能) <https://www.opi.jp/resources/corpora/>

湖南大学学習者中間言語コーパス (名古屋大学の杉村泰氏を中心に開発) <https://www.hum.nagoya-u.ac.jp/about/about-sub2/japanese/sugimura/2-3.html>

作文対訳データベース (国立国語研究所在職当時の宇佐美洋氏を中心に開発) <https://mmsrv.ninjal.ac.jp/essay/>

JLPTUFS 作文コーパス (東京外国語大学留学生日本語教育センターの鈴木智美氏らを中心に開発) http://www.tufs.ac.jp/ts/personal/SUZUKI_Tomomi/mc-sub/paper/JLPTUFS_Corpus_readme.pdf

JCK作文コーパス (埼玉大学の金井勇人氏を中心に開発) <http://nihongosakubun.sakura.ne.jp/corpus/>

住みやすい国コーパス (ミュンヘン大学の村田裕美子氏を中心に開発) <https://sumiyasui.jpn.org>

台湾人日本語学習者コーパス (CTLJ) (成功大学の黄淑妙氏を中心に開発) <https://j-kumo.>

fild.ncku.edu.tw/ctlj/terms/
 多言語母語の日本語学習者横断コーパス (I-JAS) (広島大学／国立国語研究所在職当時の迫田久美子氏を中心に開発) <https://chunagon.ninjal.ac.jp>
 中国語・韓国語を母語とする日本語学習者縦断的メール文コーパス (帝京大学の東会娟氏が開発) <https://emailtest.sakura.ne.jp/data/>
 中国語・韓国語母語の日本語学習者縦断発話コーパス (C-JAS) (広島大学大学院生時代の迫田久美子氏が開発) <https://chunagon.ninjal.ac.jp>
 寺村誤用例集データベース (筑波大学／大阪大学在職当時の寺村秀夫氏が開発) <https://www2.ninjal.ac.jp/teramuradb/>
 日中 Skype 会話コーパス (実践女子大学に在職当時の中俣尚己氏を中心に開発) <http://nakamata.info/database/>
 日本語学習者会話データベース (国立国語研究所の野山広氏を中心に開発) <http://mmsrv.ninjal.ac.jp/kaiwa/index.html>
 日本語非母語話者の聴解コーパス (国立国語研究所在職当時の野田尚史氏や早田美智子氏を中心に開発) <http://www.nodahisashi.org/jsl-rikai/choukai/>
 日本語非母語話者の読解コーパス (国立国語研究所在職当時の野田尚史氏を中心に開発) <http://www.nodahisashi.org/jsl-rikai/dokkai/>
 BTSJ 1000人日本語自然会話コーパス (東京外国語大学／国立国語研究所在職当時の宇佐美まゆみ氏を中心に開発) <https://mmsrv.ninjal.ac.jp/btsj/>
 北京日本語学習者縦断コーパス (B-JAS) (国立国語研究所在職当時の迫田久美子氏と石黒圭を中心に開発) <https://chunagon.ninjal.ac.jp>
 LARP@SCU語料庫 (東呉大学の陳淑娟氏を中心に開発) https://web-ch.scu.edu.tw/japanese/web_page/3936
 魯東大学会話コーパス (名古屋大学の杉村泰氏を中心に開発) <https://www.hum.nagoya-u.ac.jp/about/about-sub2/japanese/sugimura/2-2.html>
 YNU書き言葉コーパス (横浜国立大学在職当時の金澤裕之氏を中心に開発。書籍を購入することで入手可能) <https://www.hituzi.co.jp/hituzibooks/ISBN978-4-89476-698-3.htm>
 YUKタグ付き中国語母語話者日本語学習者作文コーパス (関西学院大学国際学部の于康氏を中心に開発。非公開だが、「日本語誤用と日本語教育学会」の会員であればデータ請求が可能)

(石黒 圭・呉 楚琦)