まえがき

本書は，迫田久美子とそのチームが中心となって構築した「多言語母語の日本語学習者横断コーパス」(International Corpus of Japanese as a Second Language: I-JAS) の構築理念・構築過程および収集したデータの概要を紹介し，日本語習得研究・日本語教育への I-JAS の応用可能性について解説を行った書物である。

I-JAS には，学習者 1,000 名，日本語母語話者 50 名，合計 1,050 名の参加者を対象とする対面調査で収集した発話と作文，一部の参加者から収集した任意作文，および，全参加者に関する詳細な背景情報のデータが含まれる。

I-JAS は，対面調査だけで 807.6 万語（うち参加者の産出は 461.6 万語）の産出データを収集しており，名実ともに世界最大の日本語学習者コーパスであるが，その真の価値は，第二言語習得研究の枠組みにおいて，多様な対照研究に利用できるよう綿密な設計がなされている点にある。

まず，対面調査には性質を異にする 5 種のタスク（ストーリーテリング，対話，ロールプレイ，絵描写，ストーリーライティング）が含まれ，タスク間の比較を行うことができる。さらに，任意作文との比較も可能である。

次に，1,000 名の学習者は，海外学習者 850 名，国内教室環境学習者 100 名，国内自然環境学習者 50 名にわけられており，海外学習環境と国内学習環境，教室学習環境と自然学習環境の比較が可能である。

さらに，海外学習者の母語は 12 種に及び，母語別の比較や，言語系統別の比較もできる。これらの学習者と日本語母語話者の比較も行える。

加えて，すべての学習者は 2 種の習熟度テストを受けており，習熟度別の比較も可能である。このほか，背景調査のデータを併用することで，性別・年齢・日本語学習歴・日本語学習スタイルなど，幅広い観点で比較研究を行うことが可能である。

これだけの幅広いデータを体系的に集めるのは決して容易なことではなかった。I-JAS プロジェクトでは，研修を受けた 17 名の調査者が，海外 17 ヵ国 20 ヵ所，日本国内 10 ヵ所に赴き，共通のプロトコルで対面調査を

実施し，統制的にデータを収集した。

なお，ここで特筆しておきたいことは，このプロジェクトが研究プロジェクトであるとともに，日本語の学びの輪を世界に広げる国際的な教育プロジェクトでもあったということである。調査者は，タスクという枠組みの中であっても，参加者の話に真摯に耳を傾け，心を通わせ，必要な助言や励ましを行った。また，調査者は，日本の風物が印刷された絵葉書カードを持参し，個々の参加者への感謝と学習助言をカードにしたため，参加者全員に手渡しした。参加者の中には，今回のプロジェクトを通して，生まれて初めて，日本語母語話者と長時間話す経験を持った者も少なくなかった。プロジェクトへの参加は，彼ら・彼女らの日本語の学びにとって大きな刺激となった。

I-JAS プロジェクトの完成には，このような調査（前の段落の内容を受けて）に参加してくれた多くの日本語学習者，また同時に多くの機関・個人の助力を得た。巻末資料に関係者および協力者の一覧を掲載し，深く感謝の意を示したい。以下，関係するプロジェクトおよび科学研究費助成金事業を挙げ，本書がその成果の一部であることを記す。

- 国立国語研究所共同研究プロジェクト「多文化共生社会における日本語教育研究」（期間：2009–2015 年度／代表者：迫田久美子）
- 国立国語研究所共同研究プロジェクト「日本語学習者のコミュニケーションの多角的解明」（期間：2016–2021 年度／代表者：石黒圭）
- 科学研究費助成事業（基盤研究 A）「海外連携による日本語学習者コーパスの構築―研究と構築の有機的な繋がりに基づいて―」（期間：平成 24–27 年度／代表者：迫田久美子／課題番号：24251010 ／共同研究者：岩立志津夫・野田尚史・田中真理・松見法男・金田智子・大関浩美・奥野由紀子・峯布由紀・李在鎬）
- 科学研究費助成事業（基盤研究 A）「海外連携による日本語学習者コーパスの構築および言語習得と教育への応用研究」（期間：平成 28–31 年度／代表者：迫田久美子／課題番号：16H01934 ／共同研究者：野田尚史・田中真理・李在鎬・砂川有里子・松見法男・野山広・奥野由紀子・望月圭子・宇佐美まゆみ・小柳かおる・石川慎一郎）

なお，本書の関連資料の提供の場として，専用のサイト（www.9640.jp/ijas/）を用意した。サイトでは，誤植・誤記の修正情報のほか，書籍に載せられなかったデータや資料を公開する予定である。あわせて活用していただきたい。

最後に，本書の意義を認め，出版を快諾くださったくろしお出版，とくに，本書の編集を担当された池上達昭氏に感謝を申し上げたい。

今後，本書を手ほどきとして，I-JAS が第二言語としての日本語の習得研究や日本語教育の分野で幅広く使用され，世界の日本語研究および日本語教育の一層の発展に資することを願っている。

2019 年 12 月
編著者

目　次

第Ⅰ部　I-JAS の設計と構築

I-JAS は，従来にないユニークな特徴を持った学習者コーパスである。I-JAS はどのような背景で構想され，どのようにして作られたのであろうか？

I-JAS の開発にあたっては，(1) 参加者の決定，(2) 調査の設計と実施，(3) 音声データのテキスト化，(4) データの公開，という 4 つのステップで作業を進めた。

第 1 部では，まず，第 1 章において，I-JAS プロジェクト誕生の経緯や，既存のコーパスとの関係性，また，「I-JAS の 10 大特徴」について簡単にまとめた後，第 2 ～ 5 章で上記の 4 つのステップについて順に説明する。

第 2 章では，I-JAS の調査に参加した学習者のタイプやその背景を示す。

第 3 章では，対面調査を中心とする参加調査の全体をどのように設計・実施したかを示す。

第 4 章では，対面調査で集めた音声データをどのように書き起こし，言語分析用に形態素解析したかを示す。

最後に，第 5 章では，公開されるデータの全体像を示し，ファイルのコードやデータの読み方について説明する。

第 1 章

I-JAS 誕生の経緯

1.1　はじめに

本章では，なぜ，I-JAS という多様な母語の日本語学習者のコーパスが誕生したのかについて，これまでの研究を振り返りながら，I-JAS の背景および経緯を述べる。

第二言語（外国語を含む）学習者は，目標言語を習得する過程で様々な学習者特有の言語表現を産出する。正用も誤用も含んだ第二言語学習者特有の言語は「中間言語」と呼ばれ，1970 年代から研究が盛んに進められるようになった。

日本語を第二言語とする学習者にも，（1）～（3）のような様々な誤用が観察されている。（→）で正用を示し，国名は学習者の国籍を示す。

（1）　田中さん**が**（→は）どこですか　（韓国　市川, 2010, p.59）

（2）　**さむくて**（→寒いから），ヒーターをつけよう　（インドネシア　同上, p.389）

（3）　私たちは三年前にけっこん**していますが**（→しましたが），こどもがまだありません。　（アメリカ　同上, p.414）

国内における日本語教育が盛んになり始めた 1970 年代は，上記のような誤用は学習者の母語と日本語の違いが起因して起きると考える研究者が多く，日本語と学習者の母語との対照研究が盛んに行われた。しかし，日本語教育の現場にいる日本語教師は，母語の異なる学習者から同種の誤用が産出される実態を見て，学習者の母語の影響，つまり母語の言語転移だけでは片

第 2 章

参加者の決定

2.1　はじめに

学習者コーパスを作るのは膨大な時間と労力を要する作業である。I-JAS のプロジェクトでは，大まかに 4 段階の検討および作業のステップを経て最終的なコーパスの完成にこぎつけた。2 〜 5 章ではそれぞれのステップについて実際の作業の内容を紹介していく。

［1］　参加者の決定（どのような学習者を対象とするか？）
［2］　調査の設計と実施（どのような調査を行うか？）
［3］　音声データのテキスト化（音声データをどう処理するか？）
［4］　データの公開（集めたデータをどういう形で公開するか？）

まず，2 章では「参加者の決定」に関して述べる。コーパスを作る上で重要なことは，誰を対象としてデータを集めるのかを明確に決めることである。これによって集められるデータの特徴が異なってくる。本章では，I-JAS のデータがどのような参加者から取られたものなのかを具体的に紹介する。

2.2　調査対象とする参加者

日本語学習者コーパスを作ろうとする場合，その対象が日本語学習者であることは自明である。しかし，一口に日本語学習者と言っても，日本語を習い始めたばかりで，日本語をまったく書いたり話したりできないとすれば，データの収集は困難になるだろう。そこで，I-JAS では，対象とする日本語学習者について以下の 3 つの基礎要件を定めた。

第 3 章

調査の設計と実施

3.1 はじめに

2.1 節で述べたように，学習者コーパスの構築は，(1) 参加者の決定，(2) 調査の設計と実施，(3) 音声データのテキスト化，(4) データの公開という 4 つの段階で進められる。3 章では，このうち，「調査の設計と実施」に関して述べる。

I-JAS のデータの中核部は，調査者（インタビュワー役を担当した教員）と参加者（学習者および日本語母語話者）の間で行われた対面調査で収集された。多様な母語背景を持つ学習者が調査に参加していること，そうした学習者が性質を異にする多様なタスクに取り組んでいること，また，すべての参加者に同条件で調査を実施したことは，I-JAS の特筆すべき特徴であるが，I-JAS の対面調査とはどのようなものであったのだろうか。

本章では，I-JAS 構築のために行った調査の全体像を紹介する。I-JAS を正しく利用するには，I-JAS の調査がそもそもどのようなもので，「何をどう調査したのか」を正確に理解しておくことが不可欠であると言えよう。

3.2 データ収集の基本方針

学習者の日本語の習得状況を観察するには，日本語学習者の話し言葉や書き言葉のデータが不可欠である。では，具体的にどのようなタスクを与え，どのようなデータを集めればよいのであろうか。

学習者コーパスの開発に際して，収集すべきデータは一義的に決まるものではない。というのも，研究目的に応じて必要なデータは異なるからである。逆に言えば，データ収集に先立ち，「何を目的としてデータを集めるの

第 4 章

音声データのテキスト化

4.1 はじめに

2 章および 3 章で示したように，I-JAS プロジェクトでは，初めに対象とする調査参加者を決定し，次いで，対面調査を中核とする I-JAS 調査の全体的な枠組みを決めて調査を実施してきた。

世界各国で行った調査によって膨大な量の音声データが得られたわけであるが，収集した音声データをテキストデータとして整備するためには，書き起こし（transcription）と形態素解析（morphological analysis）という 2 つのデータ加工のプロセスが必要となる。4 章ではこれらの詳細について述べる。

書き起こしとは，音声を聞き，それを文字として書き出していく作業のことで，「文字化」や「文字起こし」と呼ばれることもある。また，形態素解析とは，書き起こしたテキストを形態素（テキストを構成する最小単位。いわゆる「語」に準じる）に分割し，個々の語について，読みや品詞，また，活用のタイプや終止形などの情報を加える作業のことである（小木曽，2014）。

そもそも，書き起こしや形態素解析はなぜ必要なのであろうか。ここで，以下の例を見てみることにしよう。(1) は録音された音声を，(2) は書き起こしされたテキストを，(3) は形態素解析されたテキストを示す（縦線は形態素の切れ目を示す）。なお，この例は，中国語母語話者（CCS40）の対話タスクにおける実際の発話から取ったものである。

(1) ハハハアレワケッコーマエノニジュウネングライマエノエイガーデスケド…

(2) {笑} あれは結構前の二十年ぐらい前の映画ーですけど…

第 5 章

データの公開

5.1 はじめに

2 ～ 4 章で述べたように，I-JAS プロジェクトでは，(1) 参加者の決定，(2) 調査の設計と実施，(3) 音声データのテキスト化という作業をそれぞれ綿密な計画のもとに実施してきた。コーパス構築のための作業の最後に来るのは (4) データの公開である。5 章ではこの点について述べたい。

コーパスの公開にあたって，かつては，コーパス全体をダウンロードさせる方式が一般的であった。利用者は，ダウンロードしたデータを各自のコンピュータ上で処理して必要な検索や分析を行った。こうした公開方法は，コーパスの処理技術に精通した利用者には有益なものであるが，幅広い利用者の便宜に叶うものとは言いがたい。

このため，近年では，ダウンロード版よりもオンライン版でのコーパス公開が一般的になりつつある。オンライン版では，通例，検索システムが同時に提供されるため，データ処理の技術を持たない幅広い利用者がコーパスを活用し，自身の教育や研究に役立てることができる。こうした動きをふまえ，I-JAS でも，発話書き起こしや作文といった産出テキストについては，オンライン版とダウンロード版の 2 つの形でデータを提供することとした。

5.2 2 種類のデータ

I-JAS のデータは，参加者による発話の書き起こしや作文といった産出テキストデータと，付属データ（音声＋フェイスシート＋関連情報）に大別することができる。このうち，前者についてはオンライン版とダウンロード版という 2 種類の使用形態がある。

第 2 部　I-JAS の量的外観

第 1 部では，(1) 参加者の決定，(2) 調査の設計と実施，(3) 音声データのテキスト化，(4) データの公開，という 4 つのステップで I-JAS の構築が進められたことを示した。この過程で，1,050 名の発話者から，800 万語を超える大量のデータが収集された。

第 2 部では，こうして集められたデータを計量的観点から概観する。

まず，第 6 章では，I-JAS の参加者の人数や基本的な属性情報を整理して示す。

第 7 章では，参加者のうち，学習者の日本語習熟度の概況について詳しく述べる。参加者は，すべて 2 種類の日本語習熟度テストを受験しており，テストのスコア情報を活用することで，学習者を習熟度別に分けて分析することが可能になる。

第 8 章では，参加者の産出語数に注目し，タスク別・母語別の語数について紹介する。

第 6 章

参加者の属性

6.1　はじめに

第 1 部では，I-JAS の構築過程と集められたデータの概要について説明した。すでに述べたように，I-JAS の最大の特徴は 1,000 名の学習者と 50 名の日本語母語話者から，800 万語を超える大量の産出データを集めたことである。

I-JAS を用いて研究を行う際には，参加者の産出データ（作文・発話）を調べるだけでなく，その属性情報や習熟度情報を加味して分析を行うことが重要である。また，異なるタスク間で頻度を比較する際には，タスクごとに産出の総語数を調べ，得られた頻度を標準化しておく必要がある。この目的に沿い，以下，第 2 部の 6 〜 8 章では，参加者の属性・習熟度・産出語数の 3 点について計量的な概観を行うこととしたい。なお，6 〜 8 章の分析のもとになっているのは，I-JAS の関連情報の 1 つである「フェイスシート調査（背景調査）結果」と「語数表」の情報である。

さて，6 章においては，参加者の一般的な属性情報に注目する。フェイスシートからわかる情報は多岐にわたるが，以下では，(1) 母語別・学習環境別の参加者数，(2) 職業・性別・年齢，(3) 言語環境（複言語使用状況），(4) 日本語学習環境（教育機関・学習のきっかけ・日本語使用活動）の 4 観点を取り上げ，状況を概観したい。

第 7 章

参加者の習熟度

7.1 はじめに

I-JAS は大規模な日本語学習者コーパスであるが，特筆すべきは，学習者の日本語習熟度データとして，3.6.1 節と 3.6.2 節で述べた「J-CAT」と「SPOT」のスコアが付属していることである。この情報を活用することで，学習者の産出分析を精緻化し，考察をさらに深めることができる。

本章では，J-CAT と SPOT の得点分布を中心に報告する。得点分布についての理解は，I-JAS の性質を捉える上でもっとも基礎的かつ重要な情報である。なお，以下の節では，得点の度数分布表をもとにしたヒストグラムと母語別の得点情報を示す。

7.2 J-CAT と SPOT の全般的得点分布

I-JAS に収録されている全データをもとに，J-CAT の合計得点（400 点満点）と SPOT の得点（90 点満点）を調査したところ，以下のような得点分布が確認された。なお，J-CAT の平均点は 201.95，標準偏差は 58.89，SPOT の平均点は 66.23，標準偏差は 12.06 である。

ヒストグラムは，横軸がテストの得点，縦軸が度数（人数）である。ヒストグラムを通して得点の全体分布を確認することができる。図を見る際のポイントとして，中央（中程度の能力の集団）が盛り上がって，両端に伸びる分布，すなわち正規分布になっているかどうかを確認する。

図 1 の J-CAT であれば，200 点台がもっとも盛り上がっていて，左端の 100 点台（能力が低い集団）と右端の 300 点台（能力が高い集団）が少なくなっていることが確認できる。図 2 に関しても分布のパターンとしては概ね図 1

第 8 章

参加者の産出語数

8.1 はじめに

コーパス分析では，個別の頻度を議論するにあたり，全体のサイズ情報を参照して正規化を行う。たとえば，20 万語中の 20 回と 30 万語中の 9 回であれば，それぞれの総語数で割って 10 万語あたりの調整頻度に換算し，10 回と 3 回として議論を行う。

こうした頻度の調整は母語間の比較を行う際に多く行われるが，タスク間の産出の比較を行う際にも同様の調整が必要となる。とくに I-JAS の場合，「対面調査」と言っても，そこには性質を異にする多くのタスクが含まれており，これらを比較する研究も必要であろう。そこで，本章では，I-JAS の対面調査のデータを用い，参加者がそれぞれのタスクにおいて産出した語数を概観することとしたい。

なお，インタビューデータの総語数については，参加者発話（K）だけを指す場合と，参加者発話と調査者発話の両方（K + C）を指す場合があるので注意が必要である。

8.2 タスク別産出語数

I-JAS には 800 万語のデータが収録されているが，その内訳はどのようになっているのであろうか。まず，タスク別および全体の総語数について確認する。なお，本章で言う語数とは形態素の数のことである。

第 3 部　I-JAS の使用法

本書第 1 部では，I-JAS の構築過程を示し，第 2 部では収集されたデータを計量的に概観してきた。いよいよ第 3 部ではコーパスを「実際に使ってみる」こととしたい。

まず，第 9 章と第 10 章では，国立国語研究所コーパス検索アプリケーション「中納言」上で I-JAS を検索するやり方を詳しく解説する。「中納言」では，文字列検索と，形態素解析を行ったデータをもとに行う短単位での検索の 2 種類方法があるが，9 章では前者を，10 章では後者を扱う。

第 11 章では，コーパスから得られた頻度の処理に関して，計量的な学習者コーパス研究の進め方や心構えについて解説を行う。

第 3 部の 3 つの章は，コーパス研究になじみのない入門者を主たる対象者として，具体的な手順について詳しく説明しているので，説明に沿って，ぜひ実際に I-JAS を検索していただきたい。

第9章

「中納言」の基本検索

9.1 はじめに

5.2節で述べたように，I-JASにはダウンロード版とオンライン版がある。前者はI-JASに含まれるテキストデータをダウンロードし，利用者が各自のコンピュータ上で自由に分析する方法である。使用するコンコーダンサ（コーパスから用例を検索するソフトウェア）は利用者の側で用意する。一方，後者は国立国語研究所コーパス検索アプリケーション「中納言」（以下「中納言」）上で検索を行う方法である。この場合，インターネットに繋がる環境さえあれば，利用者の側で他に用意すべきものはない。

両者を比較すると，ダウンロード版のほうが検索の自由度が高いが，コーパスやコンコーダンサに関する知識や経験がないと，I-JASのような大型コーパスからほしい用例を自力で取り出すのは難しい。一方，「中納言」を使えば，初心者であっても，ほしい用例を簡単に取り出すことができる。

そこで，第9章と第10章では「中納言」でのI-JAS検索方法について解説する。「中納言」には「文字列検索」と「短単位検索」という2つの検索方法があるので，第9章では前者を中心に基礎的な検索手法を解説し，10章では後者を中心に応用的な検索手法を解説する。ただし，「中納言」上でI-JASを使用するには登録手続きが必要なので，次節では，まず，その方法から紹介したい。

9.2 事前登録

「中納言」でI-JASを利用するには，(1)「中納言」のユーザー登録と，(2)I-JASの利用申請という2つの手続きが必要になるが，これらは同時に完了

第 10 章

「中納言」の応用検索

10.1 はじめに

短単位検索では，形態素解析で付与された形態論情報を利用することができる。これにより，文字列検索ではできなかった，以下のような応用検索が可能となる。

［1］ 活用形一括検索
［2］ 品詞指定検索
［3］ 共起検索

［1］は，動詞の活用形（例：「行かない」「行きます」「行く」「行けない」「行けば」「行こう」「行って」）や語の異表記（例：「やはり」「やっぱり」「やっぱ」など）を一括で検索する方法である。短単位検索では，短単位ごとにその基本形の情報が付与されたデータを使うことになるので，基本形（各種の活用形の祖型という意味で「語彙素」と呼ぶ）を指定するだけで，上記のような例を一度に検索対象にすることができる。

［2］は，特定の語について品詞を指定して検索したり，あるいは，語を指定せず，任意の品詞の語をすべて検索したりする方法である。たとえば，「感動詞（フィラー）」を一括検索することもできる。

［3］は，複数の短単位からなる連鎖を検索する方法である。前章の例で言えば，「それ」＋「から」と指定して，「それから」を検索したり，さらには，品詞を組み合わせて，「代名詞」＋「助詞」と指定して該当するすべての用例（例：「私は」「彼が」「彼女の」など）を検索したりすることができる。

第 11 章

計量研究の方法

11.1 はじめに

第 9 章および第 10 章では，国立国語研究所コーパス検索アプリケーション「中納言」上で I-JAS のデータを検索する方法を具体的に示した。I-JAS の検索によって，我々は，様々な用例だけでなく，各種の頻度データを得ることができる。学習者コーパス研究では，質的な用例分析と計量的なデータ分析が車の両輪となる。本章では，このうち，後者に注目し，計量的な学習者コーパス研究の進め方について入門的な解説を行う。

計量研究の実践に関して述べるべきことは多いが，ここでは，(1) 必要標本数の確認，(2) 有意性検定，(3) 多変量解析の 3 点に限って，その概要を示したい。また，計量研究をめぐる最近の学界の状況についても言及を行う。

11.2 必要標本数の確認

11.2.1 必要標本数とはなにか

母語話者コーパスと異なり，学習者コーパスのサイズは総じて限定的である。ゆえに，学習者コーパスの研究者は，議論しようとする母集団に対して手元のデータがどの程度の代表性を持っているか確認しておく必要があるだろう。

統計学では母集団に対して必要な標本数を決める公式が存在する。母集団の人数を N，想定する回答の枝分かれ率を p（通例 0.5），許容する誤差を ME（Margin of Error），必要とする信頼水準（Confidence Level：CL）に呼応する定数を k とすると，必要な最低標本数 n は以下の式で定義される（Ishikawa, 2017; 石川, 2019b）。

第 4 部　I-JAS の分析

本書第 1 部では I-JAS の構築過程を示し，第 2 部では収集されたデータを計量的に概観し，第 3 部では I-JAS の使用方法について述べた。第 4 部では，これまでの議論をふまえ，I-JAS データを用いた簡単な研究の一例を示したい。第 4 部には 5 つの章が含まれるが，第 12 章と第 13 章では，作文データをマクロ的に分析し，その全体特徴を議論する。一方，第 14 章と第 15 章では発話の書き起こしデータを用い，個別的な調査対象を設定した上で，ミクロ的に問題を検討する。第 16 章では本書全体の総括と今後の展望を示す。

まず，第 12 章では，jReadability というオンラインシステムを用いてストーリーライティングの作文データを分析し，語数・漢字等使用率・文長等を議論する。

第 13 章では，同じく jReadability を使用してストーリーライティングの作文データを分析し，語彙レベルとリーダビリティを議論する。

第 14 章では，対話タスクのデータを用い，語彙の中でも，とくに副詞に対象を絞って，その使用状況を概観する。

続いて，第 15 章では，同じく対話タスクのデータを用い，語法・文法の中で，とくに，丁寧体否定文の使い分け（ないです／ません）の問題を取り上げて検討する。

最後に，第 16 章では，関連する研究分野（日本語教育・第二言語習得研究，縦断研究，計量言語学研究，世界の学習者コーパス研究）から見た I-JAS への期待について述べる。

第 12 章

作文における産出量と語彙特徴

12.1 はじめに

多くのコーパス分析では，何らかの検索システムに対して，キーワードを入れて，データを得るという方法が用いられている。これは，コーパスの部分に対するアプローチであり，コーパスをミクロ的に分析していることになる。

これとは別にコーパスデータの全体を丸ごと分析するという方法がある。これはコーパスに含まれるテキストデータを何らかの解析システムで解析し，数値データを得る方法であり，コーパスを言わばマクロ的に分析していることになる。

前者の方法を部分検索型と呼ぶなら，後者の方法は全文解析型の研究と呼ぶことができる。

表 1 部分検索型と全文解析型

部分検索型が有効な場合	全文解析型が有効な場合
• 調査のターゲットが明示的に決まっている場合 • 何らかの仮説を持っている場合 • 特定の語句の使い方を細かく調査したい場合	• 調査のターゲットは必ずしも明示的でなく，データを解析しながら事実を発見したい場合 • コーパス全体の傾向を把握したい場合

部分検索型は，いわゆる仮説検証的研究や課題解決的研究において有効と言える。一方の全文解析型の研究は，データから仮説を発見していくタイプ

第 13 章

作文における語彙レベルとリーダビリティ

13.1 はじめに

第 12 章では，I-JAS のストーリーライティング収集された作文データを分析することで，習熟度の上昇に伴い，語数・タイプ／トークン比率（TTR）・漢字使用率・内容語使用率・文長等に変化が生じることを確認した。本章では，この議論を発展させ，作文における語彙のレベルと作文全体のリーダビリティを概観することとしたい。

この目的に即し，「日本語教育語彙表」の 6 段階の語彙難易度（Sunakawa et al., 2012）と「jReadability」の 6 段階の文章難易度（李, 2017）に基づいて I-JAS のストーリーライティングのデータを分析する。

分析データの作り方は 12.2 節と同様であるため，13.2 節では，「日本語教育語彙表」の概要と「jReadability」を使った研究事例を紹介し，13.3 節で結果を述べる。

13.2 データと方法

13.2.1 「日本語教育語彙表」

語彙の難易度に関する調査研究では，旧日本語能力試験の「出題基準」を用いた研究が多いが，「出題基準」はテスト作成のための資料であること，1980 年代に作成された古い資料であることから，本書では「日本語教育語彙表」を使用する。

「日本語教育語彙表」とは，日本語教育用の語彙データベースであり，17,929 語の見出し語で構成されている。この語彙表には，すべての見出し語に対して，6 段階の難易度「初級前半語彙，初級後半語彙，中級前半語彙，

第 14 章

発話における副詞の使用

14.1 はじめに

すでに述べたように，学習者の L2 産出を分析するには，その全体的な性質を大掴みにすることを目指すマクロ的なアプローチと，特定の語彙や文法項目に絞って調査を行うミクロ的なアプローチが存在する。第 12 章と 13 章では，マクロ的なアプローチによる調査の実践例として，jReadability（李，2017）というオンラインサイトを使用し，I-JAS のストーリーライティングタスクで得られた作文データを「全文解析」し，学習者の習熟度に応じて，語数・漢字漢語使用率・文長・語彙レベル・リーダビリティがどのように変化するかを概観した。

これに対し，第 14 章と第 15 章では，調査対象をより狭く絞ったミクロ的なアプローチに基づき，発話データ分析の実践例を示す。第 14 章では語彙に関して「学習者の副詞使用」を，第 15 章では文法・語法に関して「学習者の丁寧体否定文使用」を論じる。

なお，こうした調査を実施するには，(1) コーパスから必要な用例を抽出し，(2) 用例を整理し，(3) 頻度情報を取り出して分析する，という一連のプロセスが必要になる。本章および次章では，読者が自身で分析過程を追体験できるよう，また，読者自身の関心テーマについて独力で分析を行えるよう，分析の手順についても詳しく解説することとしたい。

14.2 学習者による副詞の使用

日本語の各種品詞の中で，副詞は行為の程度や状態を示す働きを持つ。副詞には様々なものが存在するが，くだけた発話ではどのような副詞が多く使

第 15 章

発話における丁寧体否定文の使用

15.1　はじめに

第 14 章では，調査対象を絞ったミクロ的なアプローチによる研究の実例として，発話における語彙，とくに，学習者の副詞使用の問題を取り上げた。第 15 章では，文法・語法に注目し，同じく発話における学習者の丁寧体否定文使用の問題を考えてみたい。

文法や語法については，従来，正用と誤用を厳格に二分する研究が多かったが，コーパス研究の進展に伴い，両者は連続体に位置付けられるようになり，「言えるか言えないか」よりも，「どのような環境で何がどの程度どのように用いられるのか」という言語の「振る舞い」の解明に研究上の関心が向けられるようになっている。本章においても，こうした立場から，丁寧体否定文使用の実態を探っていく。

15.2　学習者による丁寧体否定文の使用

学習者の問題に入る前に，まずは，母語話者の状況を考えてみよう。そもそも，日本語の文否定は，普通体（常体）では 1 つの文型しか存在しないが，丁寧体（敬体）では 2 つの文型が存在する。

(1)　私はそれを知らない。
(2)　私はそれを知りません。
(3)　私はそれを知らないです。

ここで問題になるのは，動詞否定におけるマセン系とナイデス系の使い分