

中間言語対照分析における話者背景情報の組み込み
—I-JAS for CIA Integrated Face Sheet および
I-JAS for CIA V2.0 の開発—

石川 慎一郎(神戸大学)[†]

Integration of Learner Background Data into Contrastive
Interlanguage Analysis
—Development of the I-JAS for CIA Integrated Face Sheet
and I-JAS for CIA V2.0—

Shin'ichiro ISHIKAWA (Kobe University)

要旨

「多言語母語の日本語学習者横断コーパス」(I-JAS) (迫田, 2020) は中間言語対照分析 (CIA) に適した言語資源であるが、「中納言」上でも、ダウンロード版データを使っても、様々な関連要因を統制して必要なデータを取り出すことはむづかしい。報告者は、2024 年に、I-JAS の産出データを 1 枚のエクセルシート上に展開し、習熟度・母語・タスクの各観点をキーとするデータ抽出を可能にする I-JAS for CIA V1.0 (石川, 2024) をリリースした。今般、I-JAS の 3 種の「フェイスシート」データを統合した I-JAS for CIA Integrated Face Sheet と、主要な属性情報を独自コード化して産出データに組み込んだ I-JAS for CIA V2.0 を開発した。これにより、従前の 3 観点に加え、性別・年齢といった基本属性、学習動機づけ関連要因、学習履歴関連要因、学習行動関連要因の各々をキーとして該当する産出データを取り出せることとなり、従前以上に精緻な CIA の実践が可能になった。

Abstract

The I-JAS (Sakoda, 2020) is a language resource well-suited to Interlanguage Contrastive Analysis (CIA); however, even when using the downloadable data or querying on the Chunagon corpus query platform, it is highly difficult to extract the necessary data whilst controlling for various confounding factors. In 2024, the author released 'I-JAS for CIA V1.0' (Ishikawa, 2024), which organises the I-JAS output data onto a single Excel spreadsheet, enabling data extraction based on the key categories of proficiency level, native language, and task. The author, therefore, has now developed 'I-JAS for CIA Integrated Face Sheet', which integrates the three types of I-JAS 'Face Sheet' data, and 'I-JAS for CIA V2.0', which incorporates key perspectives into the output data using a proprietary coding system. Consequently, in addition to the three previous categories, it is now possible to extract relevant output data using basic attributes such as gender and age, as well as factors related to learning motivation, learning history and learning behaviour as keys, thereby enabling the implementation of CIA with greater

[†] iskvwshin@gmail.com

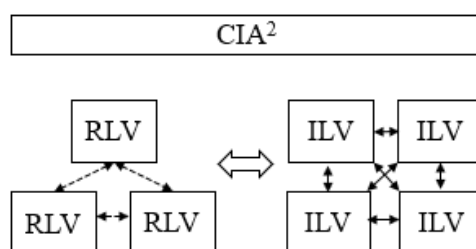
precision than before.

1. はじめに

1.1 中間言語対照分析

学習者コーパス研究では、中間言語対照分析(contrastive interlanguage analysis: CIA)という分析手法が一般的に使用される。初期の CIA モデル(Granger, 1996)は、母語話者・学習者比較と、母語を異にする学習者比較という 2 つのモジュールで構成されていたが、改訂 CIA モデル(Granger, 2015)では、「母語話者」という概念に含まれる問題点をふまえ、母語話者データを参照データに置き換えた上で、(1)多様な参照データ(reference language varieties: RLV)間での比較、(2)多様な学習者中間言語データ(interlanguage varieties: ILV)間での比較、(3)前述の RLV と ILV 間の比較、の 3 種を同時並行的に展開することの重要性が示されている(図 1)。

図 1 Granger (2015)の改訂 CIA モデル(石川, 2026b, p. 11 の図を転載)



Granger は上図の ILV の「変容」の例として task variables と learner variables の 2 点に言及しているが、多様な learner variable のうち、母語と習熟度以外の属性要因の影響の検討は十分進んでいないのが実情である。

1.2 中間言語対照分析のための I-JAS for CIA

学習者の属性データを使って信頼できる CIA を実施するには、議論しようとする観点以外のパラメタをできる限り統一することが望ましい。トピック差を見るのであればトピック以外を、習熟度を見るのであれば習熟度以外を、性差を見るのであれば性差以外を厳密に統一することで、当該観点の影響の度合いを科学的に測定することが可能になる。

世界の 1,000 名の日本語学習者に同一のインタビューを受けさせ、統制的なデザインで、多様な話し言葉・書き言葉の産出を収集した「多言語母語の日本語学習者横断コーパス」(I-JAS) (迫田, 2020) は CIA に適した言語資源である。I-JAS には非母語話者参加者の習熟度データも含まれるため、習熟度を統制して母語を比較したり、母語を比較して習熟度を比較したりすることもできる。

もともと、「中納言」サイトからダウンロードできる I-JAS のオリジナルデータは、各種の情報が紐づけされておらず、必ずしも CIA 研究にとって使いやすいものではなかった。そこで筆者は、ダウンロード版の I-JAS のデータを全面的に再構成し、I-JAS for CIA V1.0 (石川, 2024) をリリースした。I-JAS for CIA は、I-JAS に含まれる膨大な産出データを 1 枚のエクセルシート上に展開するとともに、習熟度(I-JAS、SPOT のスコアに加え、両者を組み合わせて独自に算出した JLPT 段階化レベルが使用可能)・母語・タスクを指定して合致するデータを抽出し、相互比較ができるようになっている。

1.3 「個人の中心化」研究と I-JAS

I-JAS for CIA V1.0 のリリースにより、タスク・習熟度・母語の 3 観点を組み合わせた CIA の実行のハードルは下がったが、母語や習熟度以外の参加者背景属性を使った分析は依然として困難である。I-JAS で調査された膨大な参加者の属性情報はフェイスシートという独立ファイルに記録されており、産出データと直接に紐づけられていないためである。

応用言語学や SLA(第二言語習得)の変遷を考える場合、I-JAS のフェイスシートデータの重要性は非常に高まっていると思われる。初期の応用学研究では、もっぱら、言語習得に伴う学習者の普遍的な心的反応や認知機制的の解明が目指されていた。その後、1990 年代ごろから、いわゆる“social turn”が起こり、社会文化要因を組み込んだインタラクション研究の重要性が認識されるようになった。さらに近年では“individual turn”とも呼ぶべき転換が進んでおり、学習者を安易に概括化・集約化せず、個々の学習者の背景を丁寧に見とる重要性が強調されるようになってきている。この点に関して、Benson(2019)は、過去の応用言語学における学習者の位置づけが 3 段階で変化してきたとする。第 1 段階は invisible learner(学習者の不可視化)の時代で、普遍的理論の構築を優先する構造言語学や小規模 SLA 調査が主流で、学習者は入力-出力モデルの中の情報処理装置とみなされ、学習者個々のアイデンティや背景は不可視化されていた。第 2 段階は learner-centredness(学習者の中心化)の時代で、コミュニケーション教育論や社会言語学の影響のもと、学習者は言語教育研究の中核に据えられるようになり、学習方略や学習スタイルの差異の影響が広く調査されるようになった。ただし、議論の目的は属性集団の特性の解明による教授の効率化で、個人の内部に踏み込んでいく視点はほとんど存在しなかった。そして第 3 段階が person-centredness(個人の中心化)の時代で、最近の SLA に見られるように、学習者個々を独立した人格、かつ社会的存在としてとらえ、インタビュー、ナラティブ調査、ケーススタディによって、当該人物の全体像を記述することに関心が向けられるようになっている。こうした新しい学習者観に立つ学習者コーパス研究の先駆的事例として、砂川(2025)は、I-JAS と類似の方針で縦断データを集めた B-JAS に含まれる 2 名の学習者に注目し、フェイスシートにある「日本語母語話者の友人がいるか」「日本語でどのような活動をしているか」「学習のきっかけは何か」といった情報を独自に組み合わせ、両名がそれぞれ「音声・画像・文字の情報から一方的に学ぶ」スタイルと、「音声・画像・文字の情報からインターアクションを通じて学ぶ」というスタイルを持つとして、それぞれのスタイルの違いが接続表現の獲得に及ぼした影響を丁寧に検証している。

I-JAS がこうした最新の研究トレンドにも対応できる言語資源としてより広く活用されるためには、研究チームが収集したフェイスシートのデータをわかりやすく再構成し、産出データと直接紐づけることが必要になると思われる。そこで、筆者はまず、公開されている 3 種のフェイスシートのデータを単一シート上に統合し、項目と配列を再整理した I-JAS for CIA Integrated Face Sheet を開発した。次に、フェイスシート情報の中から基本属性・動機づけ・学習履歴・学習行動の 4 観点の情報を独自に整理し、産出データと紐づけた I-JAS for CIA V2.0 を開発した。

本稿では、以下、第 2 節において、I-JAS フェイスシートについて概説する。その後、第 3 節と第 4 節で開発した 2 種のデータシートについて解説を行い、第 5 節で I-JAS for CIA V2.0 を使用した研究例を示す。

2. I-JAS フェイスシートの概要と課題

I-JAS の特徴の一つは、教室環境学習者、国内自然環境学習者、日本語母語話者、という異なるタイプの参加者から、産出データと多様な属性データを集めていることである。

I-JAS では 3 種の学習環境を区別してデータを収集したわけであるが、同じ学習環境の学習者であったとしても、男性か女性か、日本語にどの程度接触しているか、どのような指導を受けているのか、どの程度の習熟度段階にあるか、といった点で、さらに細かい違いが存在する。I-JAS ではこうした背景情報が習得研究に不可欠であると考え、学習者の属性、言語環境、日本語学習や日本語との関わり、習熟度などについての情報を細かく収集することとした。(迫田・石川・李, 2020, p. xx)

I-JAS チームは、上記の方針に従い、事前にオンラインでの回答を依頼し、不足や不備があればインタビュー当日に再確認するという慎重なデータ収集計画を立案して実行に移した。収録したデータを記録したフェイスシートは一般公開されており、プロジェクトサイトよりダウンロードすることができる。

質問内容は、参加者種別により、異なりがある。データの多数を占める海外教室環境学習者の場合、質問項目は下記の 20 種である(表 1)。

表 1 海外教室環境学習者用の質問項目

質問タイプ	質問項目
言語以外の属性(7 種)	1. 氏名, 2. 性別, 3. 所属, 4. 職業経験(アルバイトは除く), 5. 出身国, 6. 出生年, 7. 現在住んでいる国とその年数
言語環境(7 種・9 項)	8. 母語(中国語の場合は方言も確認), 9. 家族の母語, 10-1. 住んでいるところで日常的に日本語が話されているか, 10-2. 「はい」の場合, だれが話しているか, 11. 親しい友人に日本語母語話者はいるか, 12. 日本語の授業以外でどのような時に日本語を使うか, 13-1. 母語以外の日常的に使える言語があるか, 13-2. 「はい」の場合, その言語は何か, 14. 外国語の授業以外で先生が使用する言語は何か(大学入学前／大学)
日本語学習(6 種・8 項)	15. 日本語学習のきっかけは何か, 16. 現在, 日本語をどのように学んでいるか, 17. 日本語でする活動は何か, 18-1. 教育機関で日本語学習の経験があるか, 18-2. 「はい」の場合, 機関と時期, 19. これまでに勉強した日本語の教科書は何か, 20-1. 《日本以外》来日経験の有無, 《日本在住者》以前の来日経験の有無, 20-2. 「はい」の場合, 期間・目的

上記をベースとして、参加者種別によって質問が追加された(表 2)。

表 2 参加者種別による追加質問項目

参加者種別	質問項目
国内教室環境学習者	4-2. 職業経験は国内が海外か, 7-2. 今住んでいる県はどこか, 9-2. 現在一緒に住んでいる人の情報(家族/友人/日本人の有無), 14-2. 「14」の解答は母国のことか日本のことか, 18-3. 「18-2」は母国のことか日本のことか
自然環境学習者	1. 来日年, 滞日年数, 2. これまでの移住地, 3. 自国の授業で使用された言語, 4. 職業経験(海外/日本), 日本での職業内容・期間・職場での日本語使用の有無と使用相手, 5. 家庭・仕事先以外で日本語を誰と話すか, 6. 現在の日本語学習の有無と時間, 7. これまでの日本語の学習経験の有無と時間(※連番は本稿筆者が仮に付与)

また、日本語母語話者には以下の 12 種の内容を尋ねた(表 3)。

表 3 日本語母語話者用の質問項目

質問タイプ	質問項目
言語以外 (7 種)	1. 氏名, 2. 性別, 3. 出生年, 4. 学生か非学生か, 5. 職業経験の有無, 6. 小学校卒業時(12 歳)までの居住地(都道府県)とその年数, 7. 現在住んでいる都道府県とその年数
言語環境 (5 種 7 項)	8-1. 海外生活の経験の有無, 8-2. 「はい」の場合その場所と時期, 9-1. 日本語以外で日常的に使える言語の有無, 9-2. 「はい」の場合, その言語は何か, 10. 日本語教師の経験の有無, 11-1. 学校の授業を日本語以外の言語で受けたことがあるか, 11-2. 「はい」の場合, その言語は何か

上述のように、多岐にわたるデータが収集されていることは I-JAS のすぐれた特性であるが、公開されているフェイスシートを CIA 研究に活用しようとする、以下のような問題があることに気が付く。

- (1) 海外・国内教室環境学習者、国内自然環境学習者、日本語母語話者の 3 種の独立したフェイスシートが作られており、フェイスシートをまたいだ検索ができない。
- (2) 3 種のフェイスシートを統合しようとしても、それぞれで問われている項目の数と内容にずれが大きく、また、同種の設問が異なるヘッダで別の場所に位置している場合も多く、機械的な統合自体ができない。
- (3) 「住んでいるところで日常的に日本語が話されているか」「『はい』の場合、だれが話しているか」のように、本来は単一にまとめられる質問内容が 2 項目の分岐形式で調査されている場合も多く、質問項目が必要以上に増加している。
- (4) (海外生活経験の)「場所と時期」を同時に答えさせるなど、1 つの質問項目に複数の情報を回答させるものが多く、それぞれを区別したデータ抽出ができない。
- (5) 多肢選択型の設問には複数選択を認め、かつ該当項目を記載させる形式になっているため、個々の項目の選択有無に基づくデータの抽出ができない。
- (6) 自由記述については、回答内容の整理やコード化がなされておらず、それらを用いたデータの抽出ができない。
- (7) 自由記述には日本語以外の言語(英語、フランス語、インドネシア語など)も用いられており、翻訳が示されていないものもあり、そのままでは活用が困難である。

I-JAS 開発チームは、フェイスシートを、個人産出データの解釈で疑問が生じた際に当該学習者の背景を確認するための一種の「カルテ」と位置付けており、現行の体裁はその目的に叶ったものであったと思われる。ただ、CIA 研究に応用するには上記の点が障壁となり、フェイスシートの存在を知りつつも、その情報をうまく活用できないまま研究を終えてしまう例も見られる。こうした現状において、属性データがより広く活用されるためには、3 種のフェイスシートの統合・整理と、産出データとの紐づけが不可欠であると思われる。

3. I-JAS for CIA Integrated Face Sheet の開発

すでに述べたように、このプロジェクトでは、既存のフェイスシートの内容を整理・統合した I-JAS

for CIA Integrated Face Sheet と、そのデータの一部を加工して学習者産出データと関連づけた I-JAS for CIA V2.0 の開発を行う。はじめに、I-JAS for CIA Integrated Face Sheet の作成手順を示す。

作業準備として、まず、既存の 3 種のデータシートのヘッダ項目のみを 1 枚のエクセルシート上に転記する(図 2)。

図 2 I-JAS フェイスシートのヘッダ項目の一覧

	A	B	C
1	NS参加者	教室	国内自然
2	ID	ID	ID
3	データセット	データセット	データセット
4	1.性別	1.調査地	1.性別
5	2.年齢	2.性別	2.身分
	3.所属		
6	学生/学生ではない	3. 身分	3.出身国
7	4.職業経験の有無	4.職業経験	4.年齢
	5.小学校卒業(12歳)までに住んでいた所とその年数を教えてください。	5.出身国	5.最初の来日年 (西暦年)
9	6. 現在住んでいるところとその年数	6.年齢	6(1).日本に何年住んでいますか。
10	7(1).海外で生活をしたことがあります	7.現在住んでいる国	6(2).現在住んでいる場所とその年数
11	7(2).「はい」 の場合、居住地(年数)	7.住んでいる年数	6(3).今まで、どこに住んでいましたか。
12	8(1).日本語以外で日常的に使える言語は	8.母語 (一番強い言語)	7.母語 (一番強い言語)
13	8(2).「はい」 の場合、言語	9.家族の母語	8.家族の母語
	9(1).学校の授業を、日本語以外の言語で受けたことがありますか。	10.住んでいるところで日常的に日本語が話されているか。	9.家族で誰と何語で話しますか。
14			
15	9(2).「はい」 の場合、言語	10.「はい」 の場合は、だれが話していますか。	10(1).日本語とあなたの国の言葉の他に、よく使う言葉や話せる言葉はありますか

上記にあるように、ヘッダの見出し項目は、しばしば、実際の質問文のままになっている。ゆえに、整理のため、これらを短いラベルに変更する。また、以後の整理のために連番コード(Seq)を新たに付与する(図 3)。

図 3 ヘッダ項目名称の変更

	A	B	C	D
1	Type	母語話者	教室	国内自然
2	Seq_001	ID	ID	ID
3	Seq_002	データセット	データセット	データセット
4	Seq_003	1.性別	1.調査地	1.性別
5	Seq_004	2.年齢	2.性別	2.身分
6	Seq_005	3.所属（学生/非学生）	3. 身分	3.出身国
7	Seq_006	4.職業経験の有無	4.職業経験	4.年齢
8	Seq_007	5.過去（小学校）居所	5.出身国	5.初来日年
9	Seq_008	6. 現在居所	6.年齢	6(1).在日年数
10	Seq_009	7(1).海外生活経験の有無	7.現在居住国	6(2).現在居所
11	Seq_010	7(2).当該経験の詳細	7.居住年数	6(3).過去居所
12	Seq_011	8(1).日常使用可能外国語の有無	8.母語	7.母語
13	Seq_012	8(2).当該外国語	9.家族母語	8.家族母語
14	Seq_013	9(1). 外国語での授業経験の有無	10.居住地の日本語使用の有無	9.家庭内使用言語
15	Seq_014	9(2).当該外国語	10.該当する場合の話者	10(1).日本語以外の日常使用外国語

次に、個々の回答内容を見ながら、同じ内容を聞いていると判断できるものは同じ行に移動して整理していく。たとえば、母語話者向けの質問の 3 番の「所属」は、内容的には、教室環境学習者向け質問の 3 番の「身分」、国内自然学習者向けの質問の 2 場の「身分」と同等であることから、これらは同じ行に整理する(図 4)。

図 4 同一内容項目の集約

F	G	H
母語話者	教室	国内自然
ID	ID	ID
データセット	データセット	データセット
	調査対象者データ	
A1. 作文データの有無	A1. 作文データの有無	A1. 作文（任意）データ
A2. 作文プロンプト	A2. 作文プロンプト	A2. 作文プロンプト
1.性別	2.性別	1.性別
2.年齢	6.年齢	4.年齢
3.所属（学生/非学生）	3. 身分	2.身分
NA	5.出身国	3.出身国
NA	1.調査地	NA
5.過去（小学校）居所	NA	6(3).過去居所
6. 現在居所	7.現在居住国	6(2).現在居所
NA	7.居住年数	6(1).在日年数
NA	NA	5.初来日年

上記において、NA とは、当該内容の設問が存在しないことを示す。また、作文関係の 2 つの設問には元ファイルで番号が付与されていないため、便宜的に A1、A2 の整理コードを付与している。

こうして、フェイスシート調査で尋ねた内容の全体を一覧の形で書き出した後、整理のために新たな「観点」を立て、全 10 観点、50 項目に再整理する(表 4)。

表 4 ヘッダ項目の整理(10 観点 50 項目)

観点	下位項目	観点	下位項目
作文	作文データの有無	日本語接触	居住地日本語使用の有無
	作文プロンプト		居住地日本語話者
基本属性	性別		NS 友人の有無
	年齢		授業以外日本語使用場面
	身分		日本勤務日本語使用
居所情報	出身国		勤務日本語使用詳細
	調査地		使用相手(家族仕事以外)
	過去居所		(従前)渡日経験
	現在居所		渡日経験詳細
	居住年数	学習行動	学習動機
	初来日年		学習方法
母語情報	母語		学習活動
	家族母語		使用教科書
	家庭内使用言語	機関学習	機関学習歴
	通常授業教師使用言語		機関学習経験詳細
	当該使用言語詳細		過去学習歴有無
外国語能力	日常使用可能外国語有無		過去学習形態
	当該外国語		過去機関学習詳細
	海外生活経験有無	習熟度	J-CAT 聴解
	当該経験詳細		J-CAT 語彙
	外国語授業経験有無		J-CAT 文法
	当該外国語		J-CAT 読解
職業	職業経験有無		J-CAT 合計
	自国勤務形態		SPOT 得点
	日本勤務形態		
	日本勤務領域		

なお、一部の質問は特定参加者種別にのみなされていることから、質問ごとに、外(海外・国内教室環境学習者)、内(国内自然環境学習者)、母(母語話者)のコードを新規に付与する。たとえ

ば、「外・内」であれば、教室環境学習者と国内自然環境学習者のみを対象とした質問で、母語話者には該当しないことがわかる。

以下は、完成した I-JAS for CIA Integrated Face Sheet のメイン画面の一部(図 5)と、既存フェイスシートの質問項目との対応表(図 6)である。

図 5 統合フェイスシートの一部

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	基本事項_1	基本事項_2	基本事項_3	作文_1	作文_2	属性_1	属性_2	属性_3	居所_1	居所_2	居所_3	居所_4	居所_5
2	外・内・母	外・内・母	外	外・内・母	外・内・母	外・内・母	外・内・母	外・内・母	外・内	外	内・母	外・内・母	外・内
3	ID	収集回数	調査対象者データ	作文データの有無	作文プロンプト	性別	年齢	身分	出身国	調査地	過去居所	現在居所	居住年数
4	CCH02	第四次データ		無		女性	22	学生	中華人民共和国海外(中国2)			中華人民共和国22年	
5	CCH03	第四次データ		無		男性	21	学生	中華人民共和国海外(中国2)			中華人民共和国21年	
6	CCH06	第四次データ		無		女性	21	学生	中華人民共和国海外(中国2)			中華人民共和国21年	
7	CCH07	第四次データ		無		女性	20	学生	中華人民共和国海外(中国2)			中華人民共和国21年	
8	CCH08	第四次データ		有	海外教習	女性	21	学生	中華人民共和国海外(中国2)			中華人民共和国21年	
9	CCH09	第四次データ		有	海外教習	女性	21	学生	中華人民共和国海外(中国2)			中華人民共和国21年	
10	CCH10	第四次データ		無		女性	22	学生	中華人民共和国海外(中国2)			中華人民共和国22年	
11	CCH11	第四次データ		有	海外教習	女性	19	学生	中華人民共和国海外(中国2)			中華人民共和国19年	
12	CCH12	第四次データ		有	海外教習	男性	19	学生	中華人民共和国海外(中国2)			中華人民共和国20年	
13	CCH13	第四次データ		有	海外教習	女性	22	学生	ポルトガル共和国海外(中国2)			中華人民共和国22年	
14	CCH15	第四次データ		無		女性	21	学生	中華人民共和国海外(中国2)			中華人民共和国21年	

図 5 に示すように、統合フェイスシートの 1 行目には新設した観点、2 行目に当該質問がなされた参加者種別、3 行目に実際の調査項目が記載されている。3 行目にはあらかじめデータ抽出用のフィルタが設定されているので、条件を指定して、該当する学習者コードを調べることができる。たとえば、「日本語接触」観点の 3 つ目にある「日本語母語話者友人の有無」は教室環境学習者 950 名を対象とした設問であるが、フィルタ機能を使うことで、「はい」が 351 名(CCH09、CCH29、CCH46 など)、「いいえ」が 579 名(CCH02、CCH03、CCH06 など)、「未回答」が 20 名(CCM46、CCT15、CCT31)であることが即時に確認できる。母語種別とのクロス集計を行えば、母語別に、日本語母語話者友人のいる率を求めるといったことも簡単に行えるようになる。

図 6 統合フェイスシート用項目リスト

Type	Seq_1	Seq_2	コード	対象	母語話者	教室	国内自然
コード	Seq_001	基本事項_1	ID	外・内・母	ID	ID	ID
	Seq_002	基本事項_2	収集回数	外・内・母	データセット	データセット	データセット
	Seq_002x	基本事項_3	調査対象者データ	外		調査対象者データ	
作文	Seq_003	作文_1	作文データの有無	外・内・母	A1. 作文データの有無	A1. 作文データの有無	A1. 作文（任意）データ
	Seq_004	作文_2	作文プロンプト	外・内・母	A2. 作文プロンプト	A2. 作文プロンプト	A2. 作文プロンプト
基本属性	Seq_005	属性_1	性別	外・内・母	1.性別	2.性別	1.性別
	Seq_006	属性_2	年齢	外・内・母	2.年齢	6.年齢	4.年齢
	Seq_007	属性_3	身分	外・内・母	3.所属（学生/非学	3. 身分	2.身分
居所情報	Seq_008	居所_1	出身国	外・内	NA	5.出身国	3.出身国
	Seq_009	居所_2	調査地	外	NA	1.調査地	NA
	Seq_010	居所_3	過去居所	内・母	5.過去（小学校）居所	NA	6(3).過去居所
	Seq_011	居所_4	現在居所	外・内・母	6. 現在居所	7.現在居住国	6(2).現在居所
	Seq_012	居所_5	居住年数	外・内	NA	7.居住年数	6(1).在日年数
	Seq_013	居所_6	初来日年	内	NA	NA	5.初来日年
母語情報	Seq_014	母語_1	母語	外・内	NA	8.母語	7.母語
	Seq_015	母語_2	家族母語	外・内	NA	9.家族母語	8.家族母語
	Seq_016	母語_3	家庭内使用言語	内	NA	NA	9.家庭内使用言語
	Seq_017	母語_4	通常授業教師使用言語	外・内	NA	14.通常授業の教師使用言語	11.通常授業の教師使用言語
	Seq_018	母語_5	当該使用言語詳細	内	NA		11.備考(使用言語詳細)
	Seq_019	外国語_1	日常使用可能外国語の有無	外・内・母	8(1).日常使用可能外国語の有無	13.日常使用可能外国語の有無	10(1).日本語以外の日常使用外国語
外国語能力	Seq_020	外国語_2	当該外国語	外・内・母	8(2).当該外国語	13.当該外国語	10(2).当該外国語
	Seq_021	外国語_3	海外生活経験有無	母	7(1).海外生活経験の有無		
	Seq_022	外国語_4	当該経験詳細	母	7(2).当該経験の詳細		
	Seq_023	外国語_5	外国語授業経験有無	母	9(1).外国語での授業経験の有無		
	Seq_024	外国語_6	当該外国語	母	9(2).当該外国語		
	Seq_025	職業_1	職業経験有無	外・母	4.職業経験の有無	4.職業経験	NA
職業	Seq_026	職業_2	自国勤務形態	内	NA	NA	12.自国での勤務形態
	Seq_027	職業_3	日本勤務形態	内	NA	NA	13(1).日本での勤務形態
	Seq_028	職業_4	日本勤務領域	内	NA	NA	13(2).日本での勤務領域
	Seq_029	日本語接触_1	居住地日本語使用の有無	外	NA	10.居住地の日本語使用の有無	NA
日本語接触	Seq_030	日本語接触_2	居住地日本語話者	外	NA	10.該当する場合の話者	NA
	Seq_031	日本語接触_3	NS友人の有無	外	NA	11.NS友人の有無	NA
	Seq_032	日本語接触_4	授業以外日本語使用場面	外	NA	12.授業以外日本語使用場面	NA
	Seq_033	日本語接触_5	日本勤務日本語使用	内	NA	NA	13(3).日本での勤務における日本語使用有無
	Seq_034	日本語接触_6	勤務日本語使用詳細	内	NA	NA	13(4)勤務での日本語使用の詳細
	Seq_035	日本語接触_7	日本語使用相手（家族仕事以外）	内	NA	NA	14.日本語使用相手（家族・仕事先以外）
日本語習熟度	Seq_036	日本語接触_8	（従前）渡日経験	外	NA	20.（従前）渡日経験	NA
	Seq_037	日本語接触_9	渡日経験詳細	外	NA	20.渡日経験詳細	NA
	Seq_038	学習行動_1	学習動機	外・内	NA	15.学習動機	16.学習動機
学習行動	Seq_039	学習行動_2	学習方法	外	NA	16.学習方法	NA
	Seq_040	学習行動_3	学習活動	外・内	NA	17.学習活動	15.学習活動
	Seq_041	学習行動_4	使用教科書	外・内	NA	19.使用教科書	19.使用教科書
	Seq_042	機関学習_1	機関学習歴	外・内	NA	18.機関学習歴	17（1）.現在学習形態
機関学習	Seq_043	機関学習_2	機関学習経験詳細	外・内	NA	18.当該経験の詳細	17（2）.現在の機関学習の
	Seq_044	機関学習_3	過去学習歴有無	内	NA	NA	18(1).過去学習歴の有無
	Seq_045	機関学習_4	過去学習形態	内	NA	NA	18(2).過去学習形態
	Seq_046	機関学習_5	過去機関学習詳細	内	NA	NA	18(3).1. 過去の機関学習の
	Seq_047	習熟度_1	J-CAT聴解	外・内	NA	A4. J-CAT聴解/100	A3. J-CAT聴解/100
日本語習熟度	Seq_048	習熟度_2	J-CAT語彙	外・内	NA	A5. J-CAT語彙/100	A4. J-CAT語彙/100
	Seq_049	習熟度_3	J-CAT文法	外・内	NA	A6. J-CAT文法/100	A5. J-CAT文法/100
	Seq_050	習熟度_4	J-CAT読解	外・内	NA	A7. J-CAT読解/100	A6. J-CAT読解/100
	Seq_051	習熟度_5	J-CAT合計	外・内	NA	A8. J-CAT合計/400	A7. J-CAT合計/400
	Seq_052	習熟度_6	SPOT得点	外・内	NA	A9. SPOT得点/90	A8. SPOT得点/90

4. I-JAS for CIA V2.0 の開発

4.1 作業の基本指針

前節で見たように、既存の 3 種のフェイスデータの情報を I-JAS for CIA Integrated Face Sheet に統合したことにより、たとえば、「(従前) 渡日経験」(日本語接触_8)の項目において、「はい」と答えた学習者が誰か調べることができるようになった。

しかし、個々の項目への合致の有無を網羅的に調べたり、自由記述データに含まれる内容をキーとして検索を行ったり、条件に合致する学習者の産出データをダイレクトに取り出したりすることは依然として困難である。

そこで、I-JAS for CIA Integrated Face Sheet に含まれる情報のうち、CIA で特に重要と思われる 4 観点(基本属性・動機づけ・学習履歴・学習行動)について、(1)選択式回答のコード化、(2)自由記述回答のコード化、(3)産出データとの統合、という 3 種の加工を加えることとした。以下、これらの作業について概説する。

4.2 選択式回答のコード化

選択式回答については、すべて 0/1 コードに変換することとした。これにより、質問の聞き方に応じて「はい・いいえ」や「あり・なし」などの文言で記載されている回答はすべて数値化される。

また、1 つの項目に複数の選択肢が文言として入っているものについては、個々の選択肢をすべて独立の項目として立て、0/1 コードで合致の有無を示すこととした。

作業の例として、はじめに、動機づけ関連質問(「日本語学習のきっかけは何か」)で用いた作業用ファイルの一部を示す(図 7)。

図 7 動機づけ多肢選択回答のコード化

ID	学習動機	1_日本・文化興味	2_周囲の動機	3_日本生活	4_仕事将来	5_留学
CCH02						
CCH03	日本や日本の文化に興味がある	1				
CCH06						
CCH07	日本や日本の文化に興味がある	1				
CCH08	日本や日本の文化に興味がある	1				
CCH09	日本や日本の文化に興味がある	1				
CCH10	日本や日本の文化に興味がある 家族・友人の勧め	1	1			
CCH11	喜日本国					
CCH12	仕事・将来のため				1	
CCH13	日本や日本の文化に興味がある	1				
CCH15	日本や日本の文化に興味がある 仕事・将来のため 留学のため	1			1	1

この質問では、5 種の選択肢(「1: 日本や日本の文化に興味がある」「2: 家族・友人の勧め」「3: 日本で生活するため」「4: 仕事・将来のため」「5: 留学のため」)のうち当てはまるものを任意の数だけ選ぶように尋ねていたため、1 つの回答セルに、回答者により、0~5 個の文言が記載されている。そこで、「1_日本・文化興味」~「5_留学」という 5 つの列を新設し、当該内容の記載がある場合には各列に 1 を記載する。たとえば、CCH15 であれば、3 つの選択肢を選んでいるため、右側の 1 番、4 番、5 番に 1 が入っている。

こうした整理により、属性情報の計量的把握が容易になる。たとえば、縦方向で数値を合計すると、動機づけタイプごとに合致する人数がわかる。また、横方向で数値を合計すると、個人ごとに、動機要因数がわかる。

次に、日本語学習行動関連質問(「日本語でする活動は何か」)で用いた作業用ファイルの一部を示す(図 8)。

図 8 学習行動多肢選択回答のコード化

ID	学習活動	1_聴解行動	2_読解行動	3_発信行動
CCH02	テレビ・ドラマ・映画・アニメを見る（インターネット上で見るものを含む） 新聞・雑誌・小説・漫画を読む（インターネット上で読むものを含む）	1	1	
CCH03	テレビ・ドラマ・映画・アニメを見る（インターネット上で見るものを含む） 新聞・雑誌・小説・漫画を読む（インターネット上で読むものを含む）	1	1	
CCH05	テレビ・ドラマ・映画・アニメを見る（インターネット上で見るものを含む） 新聞・雑誌・小説・漫画を読む（インターネット上で読むものを含む）	1	1	
CCH07	テレビ・ドラマ・映画・アニメを見る（インターネット上で見るものを含む） 新聞・雑誌・小説・漫画を読む（インターネット上で読むものを含む）	1	1	
CCH08	テレビ・ドラマ・映画・アニメを見る（インターネット上で見るものを含む） 新聞・雑誌・小説・漫画を読む（インターネット上で読むものを含む） チャット・スカイプ・ゲーム・カラオケをする	1	1	1
CCH09	テレビ・ドラマ・映画・アニメを見る（インターネット上で見るものを含む）	1		

この質問では、3 種の選択肢（「テレビ・ドラマ・映画・アニメを見る」「新聞・雑誌・新聞・漫画を読む」「チャット・スカイプ・ゲーム・カラオケをする」）のうち当てはまるものを任意の数だけ選ぶように尋ねていたため、回答セルには 0～3 個の文言が記載されている。そこで、「1_聴解行動」～「3_発信行動」という 3 つの列を新設し、当該内容の記載がある場合に各列に 1 を記載する。なお、「チャット・スカイプ・ゲーム・カラオケをする」という選択肢については、聴解行動・発信行動・交流行動が入り混じっているが、I-JAS 開発チームへのヒアリング結果もふまえ、便宜的に主体的な「発信行動」とみなしている。

なお、すでに述べたように、0/1 コード化によって個人ごとの合計値を出すことが可能になるが、こうして得られた数値はある種の総合指標の尺度として使用可能である。そこで、動機づけ関連項目に関しては、「2:家族・友人の勧め」を除く 4 項目の合致数を「主体的動機付けスコア」として、学習行動関連項目に関しては 3 つの選択肢の合計値を「学習行動量スコア」としてそれぞれ新しい列を作って記録した。

4.3 自由記述回答のコード化

自由記述回答については、そのままでは 0/1 コード化できないので、回答を目視で確認した後、複数の回答に共通する要素を抜き出し、新規のコードに立てることとした。

作業の例として、はじめに、動機づけ関連質問で用いた作業用ファイルの一部を示す(図 9)。

図 9 動機づけ自由記述の新規コード化

ID_4	自由記述	A_言語関心	B_サブカル関心	C_反動機	D_特殊
CCH11	喜日本画				日本画
CCH17	言語に興味があるから	1			
CCH19	自分の意思ではないんです。			1	
CCH22	専門として				
CCH29	日本語の専門ですから				
CCH33	専門の調整			1	
CCH38	専門は大学に決められた。			1	
CCH39	上大学分数不、被到日			1	
CCH42	自分の意思じゃない、学校に配られたからです。			1	
CCH54	Because I was adjusted to the Japanese major			1	
CCH56	日本語は志望学科ではない			1	
CCM05	語学のことに興味があ	1			
CCM08	大学の専門は日本語です				
CCM31	大学の専門を勝手に選ぶ			1	
CCM35	偶然大学の日本語科に受かりました				
CCM36	大学入学				
CCM52	語学が好きだから。	1			
CCS01	曾經上過日語課，覺得很有趣，便繼續學習 高校で第二外国語として学ぶ。				
CCS14	大學分發成績 大学入学前の、レベル分けのテストの成績。			1	

動機づけに関連する自由記述の多くは、言語的関心と、アニメ・ゲーム等への関心に言及したものであった。また、これらに加え、成績不振で日本語専修を命じられた、など、日本語学習が自身の意思に反するものだったことへの言及も一定数見られた。そこで、「A_言語関心」～「3_反動機」という3つの列を新設し、当該内容の記載がある場合には各列に1を記載する。さらに、D列には、上記以外の特記すべき内容を短い用語にして記載することとした(表5)。

表 5 特殊動機づけ自由記述回答(32人)

日本画、アイドル、旅行、友人、メディア、英語教師として日本で就職希望、バイオリン鈴木メソッド、音楽、幼少期 immersion、文学、旅行・滞在、歌・映画・本、友人、音楽、書道、歌、教師翻訳者志望、友人、友人、自動車、結婚、家族、家族、家族、家族、恋人、親戚が日本に、声優志望、友人、旅行、在日ギタリスト、jpop

なお、自由記述の中には「日本語専攻だったので」や「日本語学科に入ったので」といったものも一定数存在したが、これらは事実の記載であって特定の動機づけとは言い難いため、D欄には記載していない。

次に、日本語学習行動関連質問で用いた作業用ファイルの一部を示す(図10)。

図 10 学習活動自由記述の新規コード化

ID_3	自由記述	A_話す	B_聞く	C_書く	D_交流	E_翻訳
CCH33	授業					
CCT15	接待、青年活動				1	
EAU11	音楽聞く		1			
EAU15	友達と飲みに行	1	1		1	
EAU36	メール・sms送信			1		
EGB04	日本人の友達を	1	1		1	
EGB17	就職用日本語作文、大学テスト対策			1		
ENZ32	音楽を聴く		1			
ENZ35	TVなどで断片的に語を聞き取る程度					
EUS01	文通			1		
EUS02	教会で日本語の通訳					1
EUS03	留学生と会話	1		1	1	
EUS15	友人と話す	1			1	
EUS33	大学で初級日本語を教授					

学習活動に関連する自由記述の多くは、友人と話すといった話す活動、ニュースを聞くといった聞く活動、メールを書くといった書く活動、友人と遊びにいくといった交流活動、母語を日本語に、あるいは日本語を母語に翻訳するなどの活動に言及したものであった。そこで、「A_話す」～「E_翻訳」という 5 つの列を新設し、当該内容の記載がある場合には各列に 1 を記載する。なお、1 つの回答に複数要素が含まれると考えられる場合には、該当する項目にそれぞれコードを記載した。たとえば、「友達と飲みに行く」という行動には、「話す」「聞く」「交流」の 3 要素が含まれているため、A、B、D の列にそれぞれ 1 が記載されている。

最後に、使用した教材関連質問(「これまでに勉強した日本語の教科書は何か」)で用いた作業用ファイルの一部を示す(図 11)。

図 11 使用教科書自由記述の新規コード化

ID	使用教科書	みんな	げんき
CCH55	総合日本語 みんなの日本語 高級日本語, 北京大学出版社 上海外語教育出版社	1	
CCS10	みんなの日本語, 大新	1	
CCS42	みんなの日本語, 大新進学日本語, 大新中級から学ぶ日本語, 大新	1	
CCS51	みんなの日本語, 大新日語語法, 文筆書局	1	
CCS53	みんなの日本語 初級Ⅰ, 大新書局楽しく聞こうⅠ, 大新書局楽し	1	
CCS60	学ぼう! にほんご, 尚昂文化みんなの日本語, 大新書局日本語上級	1	
ENZ02	げんき 1, The Japan Timesげんき 2, The Japan Times中級を学ぼう, スリー		1
ENZ12	げんき 2中級を学ぼう, 3A Corporation		1
ENZ32	げんき2中級を学ぼう		1

自由記述には様々な教科書名が記載されていたが、複数の国の学習者に共通して使用されていたものは「みんなの日本語」と「げんき」であった。そこで、「みんな」と「げんき」という 2 つの列を新設し、当該内容の記載がある場合には各列に 1 を記載する。実際の回答では、「みんなの日本語」を「皆の日本語」と漢字記載したり、英語で「Minnanonihongo」などと記載したりしている事例も多く混在していたが、筆者が目視で該当するものにはすべてコードを与えた。なお、2 種の教材のレベル(巻号数など)については区別していない。

学習者の使用教材の情報は、ユニークで価値の高いものだが、全員が正確に回答しているとは

限らない。同じ学校の同じ学年であれば全員が同じ教科書を使用している可能性が高いと思われるが、I-JAS の回答を精査すると、教材名の記載の有無は個人によりまちまちであった。ゆえに、「みんなの日本語」と「げんき」に関しても、それらを使用した学生のすべてに 1 が入っているわけではないことに留意されたい。

4.4 産出データとの紐づけ

最後に、こうして整理された 4 観点(属性・動機づけ・学習履歴・学習行動)、全 30 項目の情報を I-JAS for CIA V1.0 に統合し、産出データとの紐づけを行った。

下記は、分類観点、対応項目の一覧である。N は 0/1 コードから得られた各項目の該当者総数である(表 6)。

表 6 産出データと紐づけられた 38 項目

分類	N	項目	分類	N	項目
基本	1000	ID	学習履歴	931	機関
	1000	母語 1		659	大学主専攻
	1000	段階化レベル		841	大学学習歴
	1000	発話者		158	小中高学習歴
	1000	タスク		163	語学教室
	1000	コード		43	家庭教師・ボラ教室
	1000	テキスト 1		191	『みんな』
	1000	テキスト 2		69	『げんき』
属性	996	年齢		500	渡日歴
	1000	性(女=1)	学習行動	892	1_聴解行動
動機づけ	773	1_日本・文化興味		551	2_読解行動
	142	2_周囲の勧奨		481	3_発信行動
	221	3_日本生活		1000	行動スコア
	465	4_仕事将来		18	A_話す
	303	5_留学		10	B_聞く
	1000	動機スコア		14	C_書く
	41	A_言語関心		27	D_交流
	18	B_サブカル関心		4	E_翻訳
	9	C_反動機			
	32	D_特殊			

以下は完成した I-JAS for V2.0 の learner only シートの一部である(図 12)。

図 12 I-JAS for V2.0 の属性情報紐づけデータシートの一部

基本										属性		加味づけ				
1050		1050	1000	1050	1050	1050	1050	1050	996	1000	773	142	221	465		
ID	母語1	段階化レ	発話者	タスク	コード	テキスト1	テキスト2	年齢	性(女=1)	1_日本	2_周囲	3_日本	4_仕事	5_留学		
CCH02	中国語	N2	K	INT	CCH02-IN	CCH04で	CCH04です	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	あぜ	あぜ	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	えCCH02	えCCH02で	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	あー四年生に	あー四年生にな	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	はい	はい	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	つ、あるって	つ、あるって、	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	家はここ、こ	家はここ、ここ	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	家からここま	家からここま	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	はい	はい	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	十分(じゅっ)	十分	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	はい	はい	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	どちらに？	どちらに？	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	あ、どこ	あ、どこ	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	はい自分の、	はい自分の、自	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	あー山東省の	あー山東省の工	22	1	0	0	0	0	0		
CCH02	中国語	N2	K	INT	CCH02-IN	山東省	山東省	22	1	0	0	0	0	0		

V1.0 と同じく、各項目にフィルタボタンがついており、該当する条件を指定するだけで合致する話者による産出だけを容易に取り出すことができる。産出データはテキスト1(編集情報つき)とテキスト2(編集情報クリーニング版)のカラムに記載されている。

5. 属性情報を加味した CIA の試行

I-JAS for CIA V2.0 を使えば、興味を持った学習者の属性を観点別に調べたり、逆に、特定の属性条件を満たす学習者を網羅的に抽出したりすることもできる。「個人の中心化」研究を志向する場合は、個人学習者に密着して議論していくことになるが、I-JAS for CIA V2.0 を用い、典型的な CIA 研究、つまり、特定の条件を満たす学習者群の特性を議論する「学習者の中心化」型研究を行うことも可能である。

ここでは、属性情報を用いた学習者の群属性研究の試行として、動機づけの強い群(自身の日本語学習のきっかけとして、1_日本・文化興味、2_周囲の勧奨、3_日本生活、4_仕事将来、5_留学という5項目のうち、主体的動機づけに関連すると思われる1,3,4,5のすべてを選んでいた者)と、逆に、日本語学習が自身の意に反していた者(「希望でなかった」や「大学に決められた」など、反動機的な背景に自ら言及していた者)を比較し、各群の特徴語抽出を行う。

すでに述べたように、CIA の信頼性を高めるには関心対象以外の要因をできる限りそろえることが必要になるため、母語は全言語、タスクはストーリーテリング 1(ケンとマリはサンドイッチを用意してピクニックに出かけるが、サンドイッチを入れたバスケットに飼い犬の子犬がこっそりもぐりこみ、サンドイッチを食べてしまった)という流れの4コマイラストを見て、その内容を口頭で報告する)、習熟度は N2 で統一する。

I-JAS for CIA V2.0 で、テキスト列の左側にある基本属性タスクに関して、「母語1」はそのまま、「段階化レベル」は N2 のみ、「発話者」はそのまま(K:学習者のみ)、「タスク」は ST1 のみとし、テキスト列の右側にあるその他の属性カラムの中で、「動機スコア」を4のみに指定し、抽出された18人(CCM24, CCS38, CCS56, EAU20, EAU34, ENZ29, GAT21, GAT41, IID46, JJC20, JJC38, RRS11, RRS12, RRS40, TTH13, TTH34, TTH52, TTR43)の口頭産出(事後編集などを除去し、元の話者産出だけを残したテキスト2)をテキストファイルとして保存する(図13)。

図 13 I-JAS for CIA V2.0 からのデータ抽出

ID	母語1	段階化	発話者	タスク	コード	テキスト1	テキスト2	動機スコア (1, 3-5)
CCM24	中国語	N2	K	ST1	CCM24-S'	はい、		
CCM24	中国語	N2	K	ST1	CCM24-S'	その		
CCM24	中国語	N2	K	ST1	CCM24-S'	あー、		
CCM24	中国語	N2	K	ST1	CCM24-S'	えーと		
CCM24	中国語	N2	K	ST1	CCM24-S'	えーと		
CCM24	中国語	N2	K	ST1	CCM24-S'	んーケ		
CCS38	中国語	N2	K	ST1	CCS38-ST	C C S		
CCS38	中国語	N2	K	ST1	CCS38-ST	マリと		
CCS38	中国語	N2	K	ST1	CCS38-ST	えーっ		
CCS38	中国語	N2	K	ST1	CCS38-ST	そして		
CCS38	中国語	N2	K	ST1	CCS38-ST	はい、		
CCS56	中国語	N2	K	ST1	CCS56-ST	はい、		
CCS56	中国語	N2	K	ST1	CCS56-ST	はい、		
CCS56	中国語	N2	K	ST1	CCS56-ST	それで		
CCS56	中国語	N2	K	ST1	CCS56-ST	で、マ		

続いて、動機スコアのフィルタを「すべて選択」にして指定解除した後、「反動機」のカラムのフィルタで 1(該当あり)を指定し、抽出された 9 名 (CCH19, CCH33, CCH38, CCH39, CCH42, CCH54, CCH56, CCM31, CCS14)の産出を別のテキストファイルに保存する。

その後、2 つのテキストファイルをオンライン特徴語抽出ツールの EJTKAN(石川, 2026a)で処理する。EJTKAN は入力されたターゲットテキストと参照テキストを形態素解析し、すべての語の頻度を 2 群間で比較し、ターゲット側とレファレンス側で有意に多用される語を自動的に抽出する。統計値は対数尤度比統計量(LLR)とし、有意水準 5%(多重比較補正なし)に相当する 3.84 を超える語を特徴語とする。

以下は、上記の処理で得られた各群の特徴語の一覧である(表 7)。

表 7 特徴語

動機づけの強い群	動機づけの弱い群
が, から, た, や, まし, しまい, えっと, いい, ハム, き, 思い, 後, 出, それ, 作っ	彼, ら, んー, あん, ば, えー, ある, み, す, みつけ, 大変, 行く, ん, はー, いる, ピース, 話, サンドイッチー, っと, なん, ばす, うん, は, 二人, にー, つい, 決め, 褒める, かれ, まず, あっ, 遠い, 郊外, いー, 楽し, 出す, 芝生, ご飯, 着い, くだ, ずけっ, かー, 物, いま, さがし, 病院, えっ, 行き先, けー

動機づけに関しては、強い(多様な)動機づけを持つ学習者のほうが難度の高い語彙や文法を積極的に使用するのではないかという予測が一般に成り立つ。データを見ると、動機づけの強い群では、動作主を明示する主格の「が」(犬が)や、過去助動詞の「た」(サンドイッチを作りました)、助詞の「から」(それから、箱の中から)、丁寧体の「まし(た)」, 項目列举に用いる副助詞の「や」、完了アスペクトを表す「しまい(ました)」, 複合動詞成分の「き」(>出てきました)など、日本語学習で重要視される指導項目が多用されていた。また、一般に非流暢性のマーカーとされるフィラーは非常に少なく、唯一の例も「えっと」という語彙的フィラーであった。

一方、動機づけの低い群では、2名の登場人物を「彼ら」や「二人」と言い換えたり、不定冠詞の「ある」を名詞に添えたりするなど(ある犬は)、文法的な表現に関連する語もいくつか見られるが、特徴語の多くは非語彙的フィラー(「んー」、「あん」、「えー」)であり、そのほか、言い直しや(「み」、「み」ました、「か」、「かれ」らは)、助詞の引き延ばし(犬「はー」、「サンドイッチー」、箱「にー」など)も多い。また、バスケットの言い間違い(「ば」、「す」、「ばす」、「ずけっ」、「けー」)に起因する語の断片なども多い。

I-JAS 調査で調べたのは学習開始時点の動機づけであって、それ以後、動機づけが様々に変化している可能性はあるものの、学習開始時点において自ら進んで日本語を学ぼうとしていた学習者は教室学習に主体的に取り組み、結果的に、規範的・文法的な日本語を産出する傾向があるのに対し、学習開始時点において日本語に関心を持ていなかった学習者は教室学習への意欲が低く、結果的に、N2 水準に達しても、教科書的な語彙や表現の使用が少なく、むしろフィラーを多用するなどのコミュニケーションストラテジーでタスクをこなす傾向があるのではないかと考えられる。

動機づけの変化の可能性に加え、サンプル数の限られている試行調査の結果を一般化することには慎重でなければならないが、I-JAS for CIA V2.0 を使うことで、こうした探索的研究が平易に実行できるようになったことは意義のあることと言えるだろう。I-JAS for CIA V2.0 は、男性と女性の比較、年齢別の比較、アニメ・ゲームから日本語に興味を持った学習者とそれ以外の学習者の比較、留学意欲ないし渡日経験のある学習者とそうでない学習者の比較、小中高での学習歴のある学習者とそうでない学習者の比較、学習行動において聴解重視型、読解重視型、発信重視型の学習者の比較といった新しいタイプの CIA 研究の広がりや端緒になると期待される。

6. まとめ

以上、本論では、I-JAS のフェイスシートデータの利活用を促進するために開発した I-JAS for CIA Integrated Face Sheet と、I-JAS for CIA V2.0 の内容について紹介を行い、学習者背景を組み込んだ CIA の可能性について簡易な研究実例を示した。

海外の主要な CIA 研究を見ても、比較の観点はずもって母語に置かれ、時に、タスク影響や習熟度影響が調べられるという程度にとどまっているのが現状である。しかしながら、学習者の L2 使用が様々な心理的・社会的・環境的影響を受けるのは自明であり、この点を考えれば、学習者を取り巻く様々な背景属性データについても、むしろ積極的に分析に取り込んでいくことが必要であろう。

もちろん、CIA 研究が志向するのは通例 learner-centredness であって、厳密な意味での person-centredness ではない。そもそも、匿名化された大量のデータを集めることで、集団に潜む傾向の集約と一般化を目指す形で発展してきたコーパス言語学のデータや手法と、個人の背景をその来歴や社会文化環境も含めて詳しく掘り下げようとする SLA の最近のトレンドの完全なすり合わせは容易ではないが、歩み寄りの努力から新しい研究のヒントが生まれてくる可能性もあるだろう。I-JAS for CIA Integrated Face Sheet と I-JAS for CIA V2.0 が、learner-centredness に person-centredness の要素を加味した新しい CIA の萌芽と展開に資することを期待したい。

注

- 1) I-JAS for CIA Integrated Face Sheet と I-JAS for CIA V2.0 は、I-JAS 開発責任者の迫田久美子氏の許諾のもとに作成を行っている。価値ある言語資源を構築された迫田氏をはじめとする開発関係者に敬意を表する。

- 2) I-JAS for CIA V1.0 同様、今回作成した I-JAS for CIA Integrated Face Sheet と I-JAS for CIA V2.0 も、国立国語研究所の「中納言」内リンクからダウンロードできるようになる予定である。
- 3) 本稿は学習者コーパス研究会(2026 年 7 月 19 日)において「I-JAS フェイスシートのさらなる活用のために」という題目で行った報告をもとにしている。
- 4) 本稿は言語資源ワークショップ 2026(LRW2026) (2026 年 8 月 20-21 日)のために提出した草稿版であり、ワークショップ終了後、必要な加筆修正を行い、最終版を別媒体に公開する予定である。別媒体での公表がいわゆる二重投稿に当たらないことについては、言語資源ワークショップ主催者より事前に確認を得ている。

参考文献

- Benson, P. (2019). Ways of seeing: The individual and the social in applied linguistics research methodologies. *Language Teaching*, 52(1), 60 - 70. doi:10.1017/S0261444817000234
- Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), *Languages in contrast: Papers from a symposium on text-based cross-linguistic studies. Lund 4-5 March 1994* (pp. 37-51). Lund University Press.
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), 7-24.
- 石川慎一郎 (2024) 「中間言語対照分析 (CIA) のための I-JAS ダウンロードデータの加工—I-JAS for CIA の整備—」『言語資源ワークショップ 2024 発表論文集』 24-46.
- 石川慎一郎 (2026a) 「英・日テキストの特徴語を自動抽出する「EJTKAN」の開発 : コーパス研究における特徴語分析の意義と可能性」『統計数理研究所共同研究レポート』484, 32-56.
- 石川慎一郎 (2026b) 『ベーシック学習者コーパス研究: 英語・日本語学習者コーパスを用いた研究実践』 ひつじ書房.
- 迫田久美子 (2020) 「I-JAS の設計と構築」迫田久美子・石川慎一郎・李在鎬(編著)『日本語学習者コーパス I-JAS 入門: 研究・教育にどう使うか』(pp. 2-13). くろしお出版.
- 砂川有里子 (2025) 「学習スタイルの違いは接続表現の使用に影響するか: 中国語母語話者の日本語作文縦断コーパスの分析から」 学習者コーパス研究会 (2025 年 2 月 15 日) 発表資料.