

国立国語研究所学術情報リポジトリ

A Study on the Stability of L1 Production Seen in I-JAS : Japanese Learner Corpora and L2 Teaching

メタデータ	言語: jpn 出版者: 公開日: 2018-03-20 キーワード (Ja): キーワード (En): 作成者: 石川, 慎一郎, ISHIKAWA, Shin'ichiro メールアドレス: 所属:
URL	https://doi.org/10.15084/00001474

日本語学習者コーパスの教育応用における留意点 —『多言語母語の日本語学習者横断コーパス』に見る 母語話者 L1 産出データの安定性検証を中心に—

石川 慎一郎（神戸大学）

A Study on the Stability of L1 Production Seen in I-JAS: Japanese Learner Corpora and L2 Teaching

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）を初めとする大型の日本語学習者コーパスの整備が進んだことで、母語話者と学習者の言語運用を比較し、学習者の逸脱性を客観的に明らかにした上で L2 教育の質的改善を図る可能性が拓かれつつある。しかし、こうした研究を実践する際には、母語話者データおよび学習者データの性質を十分に理解し、得られた結果を慎重に解釈する必要がある。本研究では、日本語学習者コーパスの教育応用を考える際に留意すべき問題点を概観した後、とくに母語話者による L1 産出データの安定性の問題を取り上げ、I-JAS を使った検証を行う。検証の結果、母語話者の L1 産出であっても、その正確性や言語特性については想像以上の多様性が存在することが示された。

1. はじめに

学習者コーパス研究においては、中間言語対照分析 (contrastive interlanguage analysis : CIA) という分析手法が標準的に使用される (Granger, 1996 ; Granger, 1998 ; Granger et al., 2002)。中間言語対照分析では、多くの場合、母語話者による第 1 言語 (L1) 産出と学習者による第 2 言語 (L2) 産出が比較され、これにより、学習者特有の過剰使用 (overuse)、過小使用 (underuse)、誤用 (misuse) などが特定される。L2 運用における学習者の逸脱の詳細が明らかになれば、それらをふまえて L2 指導の内容を改善することができる (石川, 2012)。

英語においてはこうした学習者コーパスの教育応用がすでに広く試みられているが、日本語の場合、使用できるデータに制約があり、従来、同様の研究は必ずしも一般的ではなかった。しかし、近年、世界の日本語学習者と日本語母語話者による話し言葉および書き言葉の産出を大規模に収集する『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 : 以下、I-JAS）（迫田他, 2016 ; 迫田 2016a）が開発され、今後、日本語においても、学習者コーパスを用いた中間言語対照分析と、その結果をふまえた教育内容の改善が大いに期待される場所である。

もっとも、学習者コーパスの教育応用に関しては、比較の規準となる母語話者による L1 産出の安定性、比較に使用する産出データの統制性、比較結果の教育応用の是非など、検討されるべき問題点も残されている。以下、2 節において 3 つの論点を概観した後、3 節において、とくに L1 産出の安定性の問題を取り上げ、I-JAS の日本語母語話者による L1 産出データを用いた検証を行う。なお、本研究は、2016 年 5 月 9 日にリリースされた I-JAS の第 1 次公開データ（学習者 210 名・母語話者 15 名）に基づく。

2. 日本語学習者コーパスの教育応用に関わる 3 つの問題

以下、3 つの問題について順に見ていきたい。

2.1 L1 産出の安定性

1 点目は、比較の規準となる母語話者による L1 産出が、正確性や言語特性の点で、真に安定しているかどうかということである。一般に、学習者コーパス研究では、学習者については多様性（国籍、L1、L2 習熟度、年齢、動機づけなど）を前提とした議論がなされる一方、母語話者については一枚岩的なとらえ方をすることが多い。

しかし、Leech (1998) は次のように述べて、この点に警鐘を鳴らしている。

The conventional prescriptive view has been that the goal of foreign language learning is to approximate closer and closer to the performance of native speakers. Yet which native speakers? American, Australian, British or Caribbean? Highly educated or less so? Old or young? Such questions as these cause difficulties, although in practice teachers probably have covert answers to them. The problem becomes more noticeable when we compare learner corpora with a native-speaker 'reference corpus'.... Native-speaking students do not necessarily provide models that everyone would want to imitate. And, when we come to examine a reference corpus of native-speaker speech, the less admirable features of the native speaker's performance can show up especially clearly... (Leech, 1998, p. xix)

上記でも示唆されるように、母語話者の側にも、地域・教育・年齢といった点で多様なばらつきが想定しうる。また、母語話者が常に模倣すべき「手本」となるような無謬の文を産出するとも言いきれない。

I-JAS は、母語話者について、20～50 代の年齢層にまたがる 50 名のデータを公開予定であるが（第 1 次公開データでは 15 名分のみ）、これらについても、文法的正確性や基本的言語特性における安定性を実証的に検証することが重要であると思われる。

2.2 産出データの統制性

2 点目は、比較に使用する母語話者による L1 産出や学習者による L2 産出のデータが相互に比較可能な形で統制的に収集されているかどうかということである。従来の学習者コーパス研究では、多様なデータを収集することが重視され、データの内容や産出条件の統制は必ずしも十分に考慮されていなかった。たとえば、Granger et al. (2003/ 2009) によって開発された International Corpus of Learner English (ICLE) の場合、学習者による作文のトピックは 800 種を超え、辞書使用の有無、時間制限の有無といった点でも条件はまちまちである。また、比較対象となる母語話者の作文は学習者とはまったく異なる環境で収集されている。

このように、トピックや、発話条件・執筆条件が統制されていない場合、対照分析で得られた差異を合理的に解釈することは極めて困難になる。たとえば、環境問題をテーマとして、辞書なし・時間制限ありという条件で書かれた L1 作文と、夏休みの思い出をテーマとして、辞書あり・時間制限なしという条件で書かれた L2 作文があった場合、両者の対照によって得られた差を母語話者・学習者間の差とみなすのは危険である (Ishikawa, 2013 ; Ishikawa, 2014)。このことは、母語話者・学習者の比較だけでなく、異なる L1 を持つ学習者間の比較にもあてはまる。

I-JAS は、こうした既存の多くの学習者コーパスの制約をふまえ、母語話者・学習者の双方に共通のタスクを与えることで、相互比較を可能にする統制度の高いデータを収集している。OPI (oral proficiency interview) をアレンジした独自のプロトコルに基づき、発話については、ストーリーテリング (2 タスク)、対話、ロールプレイ (2 タスク)、絵描写課

題の 6 タスクが、作文については、ストーリーライティング (2 タスク) とその他の作文 (4 タスク) の 6 タスクが用意されており、それぞれ、内容や条件が明示的に規定されている (迫田, 2016a; 迫田, 2016b)。もっとも、絵描写課題とその他の作文については全員から収集しているわけではないので、分析の際には注意が必要である。

2.3 比較結果の教育応用の是非

3 点目は、比較によって得られた研究をどのような形で L2 教育に応用していくかということである。学習者コーパス研究は伝統的に L2 教育との親和性が高く、母語話者との比較によって学習者の「逸脱的」な言語使用が明らかになれば、教育的介入によって、その矯正を行うべきであるとする主張が広くなされてきた。

しかしながら、母語話者との比較で検出された差異のすべてを問題ととらえ、教育的に介入することは必ずしも適切ではない。前述のように、言語使用のすべての面において母語話者が常に無謬の手本というわけではないし、加えて、母語話者の L1 産出から乖離していたとしても、コミュニケーションを阻害するものでない限り、それらを矯正する根拠は希薄だからである。

この点に関して、Granger (2009) は図 1 のようなモデルを示し、中間言語対照分析と指導項目選定の間には、質的な取捨選択の過程が存在するべきだと強調している。

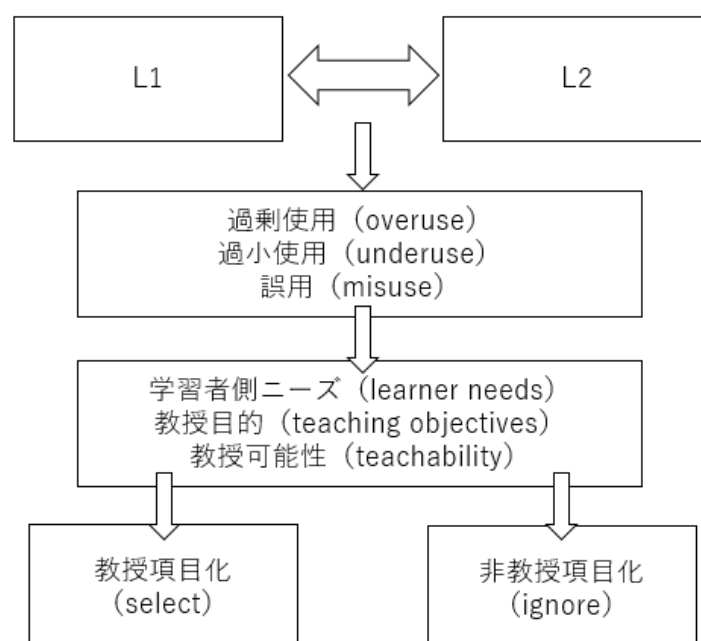


図 1 中間言語対照分析の教育応用過程 (Granger, 2009, Fig.4 を改変)

母語話者による L1 産出と学習者による L2 産出を対照することで、学習者の過剰使用・過小使用・誤用などの傾向が検出されるが、それらを教授項目として選定するか、問題ないとして無視するかは、学習者側ニーズ、教授目的、教授可能性といった諸点をふまえた現場の教師の判断にゆだねられるべきであるというのが Granger の主張である。つまり、学習者コーパス研究者に一義的に期待される仕事は、何を教え、何を教えないかを直接に指し示すことではなく、現場の教師が教授の内容を吟味する上で必要となる基礎資料を適切な形で提供することであると言える。

I-JAS について言えば、現時点では一部のデータが試行的に公開されているだけで、本格的な母語話者との対照研究や日本語教育への応用はいまだ行われていない。しかし、今後、

データの全体が公開されると、対照研究の結果を日本語教育に応用しようとする研究も増えてくるだろう。その際、安易な教育応用と一線を画する態度が研究者には求められる。

3. I-JAS に見る母語話者 L1 産出データの安定性

3.1 ねらいと RQ

前節では、日本語学習者コーパスの教育応用に関わる 3 つの問題点を概観したが、このうち、とくに重要になるのは、比較の規準となる母語話者による L1 産出の安定性の問題である。本研究は、I-JAS の母語話者 L1 産出データを用い、文法的正確性と、産出の内容に影響されることが少ないと考えられる基本的な言語特性（語数、句読点使用率、高頻度語使用状況）に着目して安定性の検証を行う。リサーチクエスチョン（RQ）は以下の 3 つである。

RQ1 母語話者による L1 産出は、文法的正確性の点でどの程度安定しているか？

RQ2 母語話者による L1 産出は、基本的な言語特性の点でどの程度安定しているか？

RQ3 母語話者による L1 産出は、高頻度語の使用状況に着目した場合、どの程度内部的に一体か？

3.2 データと手法

本研究では、I-JAS の第 1 次公開版に含まれる 15 名の母語話者によるストーリーライティング（SW1）のデータを使用する。各種のタスクの中で、とくにストーリーライティングを選んだのは、話し言葉に比べ、書き言葉のほうが安定的な産出がなされやすいことに加え、I-JAS のストーリーライティングは、すでに行ったストーリーテリングと同一課題で行われているためである。各種のタスクの中で最も高い安定性が予想されるストーリーライティングにおいて、仮に何らかの不安定性が検出されるとすれば、それは、母語話者による L1 産出の安定性の想定に対する強力な反証となる。

SW1 では、ピクニックに関する 5 枚のイラスト（図 2）を見た後、指示された冒頭文（朝、ケンとマリはサンドイッチを作りました）に続けて、イラストに沿ってストーリーを作文する。なお、このイラストは先に行ったストーリーテリングで使用したものと同等であり、被験者はあらかじめ話の内容を十分に理解したうえで作文に臨むことができる。

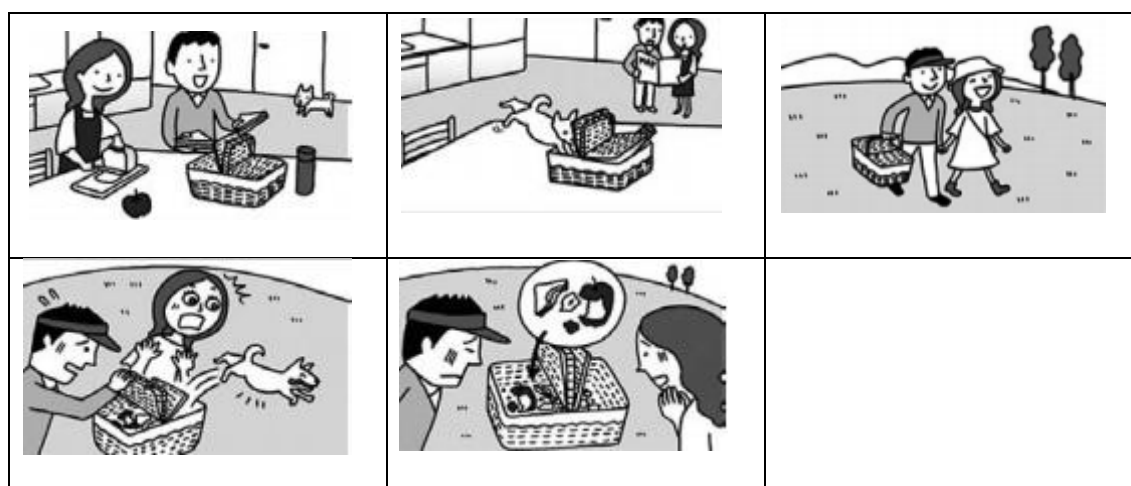


図 2 J-JAS ストーリーライティング第 1 課題（奥野・リスダ，2015 の図 1 を再構成）

SW1 のテキストは、行番号を削除した後、Chasen によって形態素解析し、以後の分析の基礎データとする。RQ1（文法的正確性）については、I-JAS に含まれている誤用修正情報のほか、全例の目視検証により、文法的に逸脱が疑われる言語使用の有無を確認する。

RQ2（言語特性）については、作文ごとに、語数1（句読点を含まない）、語数2（句読点を含む）、読点数、句点数、句読点数、読点／句点率、100語あたりの読点数、100語あたりの句点数、100語あたりの句読点数を計量し、最小値と最大値の比率を確認する。また、それぞれの作文において粗頻度5以上となる上位形態素（句読点・記号は除く）を取り出し、15作文中での重なり度の合いを確認する。RQ3（内部的一体性）では、15作文中、6種以上で共通して使用されている全41語（句読点・記号は除く）を資料として、ケースクラスター分析（距離は $(2 \cdot 2r)$ の平方根で定義し、合併後の距離計算はWard法を使用）とコレスポンデンス分析を行い、15種の母語話者作文が内部的にどの程度一体的であるか、仮に内部でいくつか分割される場合は、どのような語の使用がそれに影響しているのかを確認する。

3.3 結果と考察

3.3.1 RQ1 文法的正確性

中間言語対照分析において、母語話者のL1産出からの逸脱を学習者の問題ととらえるのは、L1産出が文法的に無謬であるという前提に基づく。しかし、今回のデータを検証したところ、I-JASの開発者の側で誤用修正がなされた例が2点（JJJ03, JJJ26）、他にも誤用と考えられる例が1点（JJJ35）見つかった。

- (1) 目的地に到着し、お昼ご飯を食べるためにバスケットを開けると、犬が飛び出しして（→飛び出して）きました。（JJJ03）
- (2) そんな仲睦ましい（→仲睦まじい）二人を部屋の隅から、飼い犬が見ています。（JJJ026）
- (3) 行く場所を地図で確認してる[→している?] 隙に愛犬がバスケットの中に入ってしまった。（JJJ35）

誤用はいずれも軽微なものだが、母語話者作文15例中の3例において、日本語として問題となりうる言語使用が見つかったことになる。このことは、母語話者データが必ずしも言語の正確性のサンプルとなりえない場合があることを示唆している。

併せて注目すべきは、上記のようなはっきりした誤用以外にも、表記や語彙使用の点で問題を含む例が散見されたことである。下記はJJJ01の作文の一部である。

- (4) …出かける前に二人が地図を見ている間に、サンドイッチを入れたバスケットに犬が入ってしまいました…やがて突然犬がバスケットから飛び出し、二人は驚きました。

引用冒頭部の「出かける前に二人が地図を見ている間に」は、時を表す2種の副詞句が不適切に並列されたもので、たとえば「出かける前、二人が地図を見ている間に」や「出かけようとして二人が地図を見ている間に」などとするのが日本語としてより適切であろう。また、後半の「やがて突然犬がバスケットから飛び出し」の部分も、「やがて」と「突然」という意味の異なる副詞が同時に「飛び出し」という動詞を修飾する形になっており、「やがて、犬が突然バスケットから飛び出し」などとするのがより自然と言えよう。

上記は書き手の作文技術の不足やケアレスミスに起因するものとも言えるかもしれないが、興味深いことに、それらに起因しない不自然な日本語表現も見つかる。

- (5) …行く場所も決定し歩きだす二人。目的地に着いて「さあ食べよう」とバスケットを開けるとそこからは犬が・・・驚く二人。（JJJ09）（※全角省略符合は原文）
- (6) …「じゃあ、ここでお昼にしましょう。」とマリが言ったので、ケンもバスケットを地面に下ろすと、まあびっくり。…「おい、サンドイッチも林檎も全部食べられちゃたよ。」とケン。「お昼どうしましょう。」と困った顔のマリ。（JJJ10）

- (7) …余程楽しかったのでしょうか。飼い犬がその隙にサンドイッチの入ったバスケットに忍び込んだのも気づかないのですから。…さあ歩き疲れた二人はサンドイッチを食べようと丘の上でバスケットを開けました。するとあろう事か、飼い犬が飛び出したではありませんか。… (JJJ26)

体現止め・倒置・感嘆文などを含むこうした表現は、誤りではないものの、日本語教育でモデルとするような標準的な日本語とは言い難い。こうした表現の混在は、一部の母語話者がイラストからストーリーを作るという課題を狭義の「ストーリー」、つまりは、昔話や物語文、あるいはト書きのような作文を行う課題と理解し、あえてそうしたジャンルに典型的に見られる修辞法を使って作文したことを示している。筆者の検証では、15人中8人が課題を文字通りに受け取って出来事を中立的に報告しているのに対し、残りの7人（上記の3例に加えて JJJ30, 35, 37, 50）は少なくとも部分的に物語文的な特殊な修辞法を使用している。I-JAS のストーリーライティング課題は十分に練られたものであるが、にもかかわらず、母語話者の間ですら、その受け止めに差があり、予期せぬ言語的多様性が生じていることに注意が必要である。仮に、母語話者データの中身に十分な注意を払わず、機械的に対照分析を行った場合、こうした表現を日本語の「規準」としてしまう危険性もあるだろう。

3.3.2 RQ2 基本的言語特性

まず、語数および句読点の使用状況を検証したところ、表1の結果が得られた。表中の語数1および語数2は句読点を除いた語数と含めた語数を、読・句・句読・読／句は読点数・句点数・句読点数・読点／句読点比率を、読%・句%・句読%は100語あたりの読点・句点・句読点数を示す。なお、変動係数は標準偏差を平均で割った値で、平均の異なりを補正した上でばらつきの大きさを比較するための指標である。

表1 文長・句読点使用率

	語数1	語数2	読	句	句読	読／句	読%	句%	句読%
JJJ01	89	99	5	5	10	1.0	5.1	5.1	10.1
JJJ03	105	117	7	5	12	1.4	6.0	4.3	10.3
JJJ09	107	115	1	7	8	0.1	0.9	6.1	7.0
JJJ10	151	171	6	14	20	0.4	3.5	8.2	11.7
JJJ11	81	90	4	5	9	0.8	4.4	5.6	10.0
JJJ12	87	97	5	5	10	1.0	5.2	5.2	10.3
JJJ14	92	99	2	5	7	0.4	2.0	5.1	7.1
JJJ15	104	115	5	6	11	0.8	4.3	5.2	9.6
JJJ17	84	93	5	4	9	1.3	5.4	4.3	9.7
JJJ26	180	199	6	13	19	0.5	3.0	6.5	9.5
JJJ30	115	129	7	7	14	1.0	5.4	5.4	10.9
JJJ35	82	92	6	4	10	1.5	6.5	4.3	10.9
JJJ37	81	90	2	7	9	0.3	2.0	7.1	9.2
JJJ50	153	160	6	1	7	6.0	3.8	0.6	4.4
JJJ57	110	123	4	9	13	0.4	3.3	7.3	10.6
平均	108.1	119.3	4.7	6.5	11.2	1.1	4.1	5.4	9.4
標準偏差	30.4	33.1	1.8	3.4	3.9	1.4	1.6	1.8	1.9
変動係数	0.3	0.3	0.4	0.5	0.3	1.2	0.4	0.3	0.2
最大値	180.0	199.0	7.0	14.0	20.0	6.0	6.5	8.2	11.7
最小値	81.0	90.0	1.0	1.0	7.0	0.1	0.9	0.6	4.4
最大／最小比率	2.2	2.2	7.0	14.0	2.9	42.0	7.5	13.1	2.7

語数 1, 語数 2 に注目すると, 最小値と最大値の差が 2.2 倍に及ぶことが明らかになった。I-JAS では, イラストを与えているため, 産出の量はほぼ一定になると予想されたが, 実際には 2 倍を超える差が存在する。学習者の L2 産出の分析では, 一般に, 文章の長さは習熟度の高さを示すとされるが, 習熟度スケールの終値にあるはずの母語話者間にすらこうした差が生じていることは注目に値する。

また, 100 語あたりの句点, 読点, 句読点の数に注目すると, 最小値と最大値の差は 2.7 倍～13.1 倍に及ぶことがわかった。読点／句点率ではその差は 40 倍以上に達する。これらは, 内容の差の影響を受けにくいはずの句読点の使用についても, 母語話者間に大きなばらつきが存在することを示唆している。なお, 読点／句点率における著しい個体差は変動係数からも裏付けられる。

次に, 高頻度語の使用状況に注目したい。上位形態素の一覧は表 2 の通りである。表中, サンは「サンドイッチ」, バスは「バスケット」を略記したものである。

表 2 高頻度形態素

JJJ01		JJJ03		JJJ09		JJJ10		JJJ11		JJJ12		JJJ14		JJJ15	
に	7	に	8	た	6	た	8	た	5	た	8	た	6	た	7
た	6	が	6	に	5	て	6	て	5	が	6	て	6	て	5
て	6	た	6			と	6	と	5	と	6	を	6	と	5
は	5	と	6			は	6	に	5	に	6	は	5	に	5
まし	5	を	6			が	5	まし	5	バス	5	まし	5	まし	5
		て	5			の	5			まし	5				
						を	5								
JJJ17		JJJ26		JJJ30		JJJ35		JJJ37		JJJ50		JJJ57			
て	6	た	10	に	9	を	6	て	6	て	12	た	7		
を	6	の	10	た	7	に	5	た	5	た	9	を	7		
た	5	を	9	と	7			と	5	に	7	は	6		
		は	7	を	7			は	5	を	7	まし	6		
		に	6	て	6			まし	5	が	6	て	5		
		まし	6	まし	6					まし	6	と	5		
		サン	5	ケン	5					サン	5	バス	5		
		と	5	は	5										
				バス	5										
				マリ	5										

コーパス言語学では, 高頻度語の使用状況は高度に安定的であると言われるが, 今回のデータについて言えば, 頻度 1 位の形態素が 1 語に決まる 10 名 (JJJ01, 03, 09, 10, 12, 15, 30, 35, 37, 50) に限ってみても, 「た」が 4 名, 「に」が 3 名, 「て」が 2 名, 「を」が 1 名とばらついており, 完全な一致は確認できなかった。また, 基本助詞間の頻度の多少に関して, 一例として, 「を」と「に」のいずれの頻度が多いかを見たところ, 「を」を多用する者が 6 名 (JJJ10, 14, 17, 26, 35, 57), 「に」を多用する者が 7 名 (JJJ01, 03, 09, 11, 12, 15, 30), 同数が 2 名 (JJJ37, 50) となり, やはり, 完全な一致とはならなかった。

以上の結果は, 語数・句読点・高頻度語といったごく基本的な言語特性に限っても, 母語話者による L1 産出が想像以上に大きな多様性を持つことを例証する。こうしたばらつきの可能性を考慮せずに, 母語話者と学習者を単純に比較して対照研究を行った場合, 研究者の想定を超えて, 極めて不適切な結果を導いてしまう危険性も否定できない。

3.3.3 RQ3 内部的統一性

母語話者の L1 産出が安定的で、がある種の「一枚岩」のようなものであるとすれば、その内部にはっきりした下部構造は存在しないはずである。そこで、15 人中 6 人以上が共通して使用している 41 語を資料としてクラスター分析を行ったところ、図 3 の結果を得た。

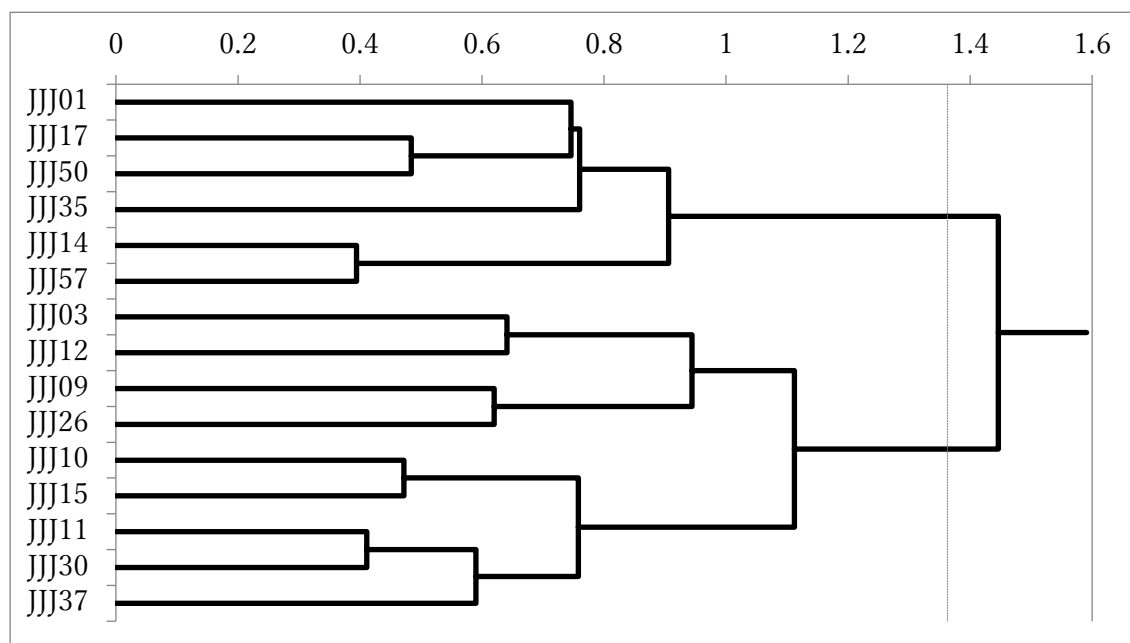


図 3 クラスター分析に基づく樹形図

定常状態が最も長く継続する場所（距離 1.2～1.4）にカッティングポイントを置くと、15 人の母語話者は大きく 2 グループ（クラスター A : JJJ01, 14, 17, 35, 50, 57 ; クラスター B : JJJ03, 09, 10, 11, 12, 15, 26, 30, 37）に分割されることとなった。このことは、高頻度語の使用状況においても母語話者の L1 産出が均質でなく、はっきりした下部構造が存在することを示している。

次に、コレスポンデンス分析の結果を概観する（図 4）。散布図に見られるように、15 名の母語話者は、まず、第 1 次元（横軸）上で、左右の 2 グループに分割される。この時、左側には先ほどの分析で得られたクラスター B が、右側にはクラスター A が位置する。それぞれの側に布置された形態素に注目すると、左側には、動作主体を表す名詞（「マリ」「ケン」「犬」）、行為動詞（「飛び出す」「開ける」「入っ」「出かけ」「食べ」）、否定辞（「ない」）などがあり、右側には、意味の希薄化した動詞（「い」「し」）、丁寧性を表す文末表現（「ます」）、照応指示語（「その」）、意思助動詞（「う」）、助詞（「の」「から」「で」「は」）などがある。母語話者の L1 産出は、内容語中心型産出と機能語中心型産出に二分されると言えよう。

また、第 2 次元（縦軸）を合わせて分析すると、全体は第 1 象限（Z1+ / Z2+）、第 2 象限（Z1- / Z2+）、第 3 象限（Z1- / Z2-）、第 4 象限（Z1+ / Z2-）に 4 区分され、第 1 象限は「その」や「し・ます」などの機能的語群、第 2 象限は「マリ」や「ケン」などの固有名詞、第 3 象限は「出かけ」「入っ」「食べ」などの動詞群、第 4 象限は「二・人」という人称代名詞的語群によってそれぞれ特徴づけられることがわかった。コレスポンデンス分析の結果もまた、母語話者の L1 産出が必ずしも一枚岩的な均質性を持っていないことを例証している。

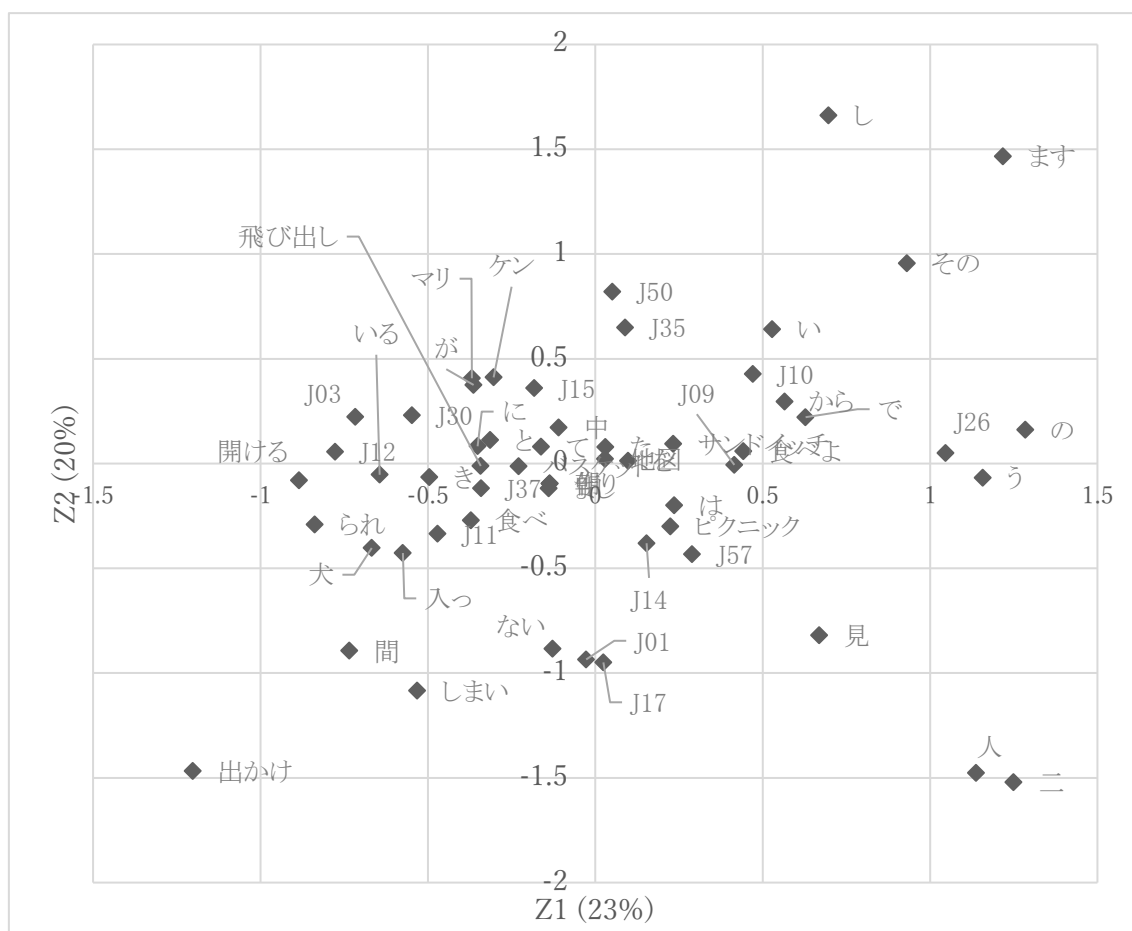


図4 コレスポネンス分析に基づく散布図（国籍コードをJで略記）

4. まとめ

本研究では、日本語学習者コーパスの教育応用における問題点として、比較の規準となる L1 産出の安定性、比較に使用する産出データの統制性、比較結果の教育応用の是非の 3 点を概観し、その後、I-JAS を用い、母語話者による L1 産出の安定性について実証的観点から検証を行った。その結果、3 つのリサーチクエスチョンに即して以下の事実が明らかになった。

まず、RQ1（文法的正確性）については、母語話者であっても明示的な文法的エラーがおおよそ 3 分の 1 の割合で見つかり、また、明らかな誤用とまでは言えないにせよ、日本語として不自然表現も随所に散見されることが確認された。さらには、統制的なタスクを与えたにもかかわらず、書き手によってタスクの受け取り方に差があり、そのことが非標準的な日本語表現の使用につながっていることがわかった。

次に、RQ2（基本的言語特性）については、最小値と最大値の比率が、語数では 2.2 倍、句点・読点・句読点数では 2.7 倍～13.1 倍に及び、読点／句点率では 40 倍以上に達することが明らかになった。また、高頻度形態素に限っても、頻度順位上の一致度は予想以上に低いことが示された。

最後に、RQ3（内部的一体性）については、クラスター分析およびコレスポネンス分析の結果、15 人の母語話者の L1 産出が内容語中心型と機能語中心型に大きく二分され、さらに、より細かく見れば、機能的語群、固有名詞、動詞、人称代名詞的語群に特徴づけられる 4 つのグループに区分されることが分かった。

以上の結果は、中間言語対照分析において、母語話者の L1 産出を絶対的で安定的な規準

とみなし、母語話者と学習者を比較して学習者の側の「逸脱」を論じることが、場合によってきわめて危うい行為になりかねないことを示すものである。学習者コーパス研究では、計量的なテキスト比較を行うことで異なるテキスト間の差異を容易に検出することができるわけだが、検出された差異が果たして何を意味しているのか、また、検出された差異を教育現場で扱うべきかどうかについては、慎重な判断が求められる。

謝 辞

本稿は、2016年12月3日に開催された第1回学習者コーパス・ワークショップー学習者コーパス(I-JAS)を利用するためにー(於:国立国語研究所)における招待講演「世界の英語学習者コーパス研究の潮流:How から Why へ」で口頭報告した内容の一部を大幅に加筆修正し、新規に原稿化したものである。同ワークショップは、石黒圭氏(国立国語研究所)がリーダーを務める「日本語学習者のコミュニケーションの多角的解明」プロジェクトおよび迫田久美子氏(国立国語研究所)による科研費(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」プロジェクトに基づく。本稿執筆の契機となる講演機会をお与え下さったことに対し、迫田教授・石黒教授に御礼申し上げる。また、当日のワークショップのパネリストであり、本稿初校に対して貴重な意見をいただいた小西円氏(国立国語研究所)および奥野由紀子氏(首都大学東京)にあわせて感謝申し上げます。

文 献

- 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.), *Language in contrast: Text-based cross-linguistic studies* (pp. 37-51). Lund, Sweden: Lund University Press.
- Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 3-18). Harlow, England: Addison Wesley Longman.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In K. Aijmer (Ed.), *Corpora and language teaching* (pp. 13-32). Amsterdam, The Netherlands: John Benjamins.
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (Eds.). (2003). *International corpus of learner English*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.
- Granger, S., Dagneaux, E., Meunier, F., & Paquot, M. (Eds.). (2009). *International corpus of learner English. Version 2*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.
- Granger, S., Hung, J., & Petch-Tyson, S. (Eds.). (2002). *Computer learner corpora, second language acquisition and foreign language learning*. Amsterdam, The Netherlands: John Benjamins.
- 石川慎一郎 (2012). 『ベーシックコーパス言語学』 東京: ひつじ書房.
- Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian Learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 1* (pp. 91-118). Kobe, Japan: Kobe University.
- Ishikawa, S. (2014). Design of the ICNALE-Spoken: A new database for multi-modal contrastive interlanguage analysis. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 2* (pp. 63-76). Kobe, Japan: Kobe University.

- Leech, G. (1998). Preface. In S. Granger (Ed.), *Learner English on computer* (pp. xiv-xx). Harlow, England: Addison Wesley Longman.
- 奥野由紀子・リスダ=ディアンニ (2015). 「『話す』課題と『書く』課題に見られる中間言語変異性：ストーリー描写課題における『食べられてしまっていた』部を対象に」『国立国語研究所論集』9, 121-134.
- 迫田久美子 (2016a). 「I-JAS 使用の手引き簡易版」東京：国立国語研究所.
- 迫田久美子 (2016b). 「学習者コーパスをどう使うか：I-JAS, C-JAS 検索法入門」『コーパスと日本語教育研究国際シンポジウム予稿集』（湖南，中国：湖南大学）38-45.
- 迫田久美子・小西円・佐々木藍子・須賀和香子・細井陽子 (2016). 「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」『国語研プロジェクトレビュー』6(3), 93-110.