



絵描写作文課題におけるL2日本語学習者の動詞使用 と習熟度の関係 : I-JASのSW1課題データの計量的概 観

石川, 慎一郎

(Citation)

統計数理研究所共同研究リポート, 444:1-22

(Issue Date)

2021-03-15

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81012571>

(URL)

<https://hdl.handle.net/20.500.14094/81012571>



絵描写作文課題における L2 日本語学習者の動詞使用と 習熟度の関係

—I-JAS の SW1 課題データの計量的概観—

石川慎一郎
神戸大学

概要

本研究では、I-JAS の絵描写作文課題 (SW1) における産出データを用い、日本語学習者 850 人を J-CAT の 10 点刻みで群化した上で、習熟度の発達と動詞使用の関係を量的に概観した。まず、RQ1 (動詞使用量) については、学習者の使用する動詞のトークン数およびタイプ数の平均は母語話者より 22~28%程度少ないこと、また、習熟度別に見た場合、トークン数・タイプ数とも初級から中級にかけて増加し、母語話者を上回る水準に達した後、再び減少する逆 U 字型の分布を見せることが確認された。RQ2 (マーカー動詞) については、母語話者マーカー動詞 9 種、学習者マーカー動詞 15 種、初級マーカー動詞 3 種、上級マーカー動詞 19 種を特定した。RQ3 (動詞頻度による分類) については、対応分析により、学習者の動詞発達過程がおよそ 4 段階に分かれることが確認された。最後に、RQ4 (習熟度推定) については、J-CAT 総合スコア = $34.42 \times \text{「居る」頻度} + 35.97 \times \text{「見る」頻度} + 52.91 \times \text{「為る」頻度} + 107.86 \times \text{「作る」頻度} - 150.57 \times \text{「有る」頻度} + \text{定数 } 9.05$ という回帰式が得られた。

キーワード

I-JAS, 学習者コーパス, 作文, 動詞獲得過程, 習熟度推定

1. はじめに

各種の品詞の中で、動詞は「述語や修飾語を担う文法上きわめて重要な役割を果たす語群で、叙述を形成するその中核となる、いわば日本語の表現機能を支える担い手」である (森田, 2008, p. 21)。L2 学習者の日本語産出に関しても、学習者の全般的な習熟度が上昇すれば、動詞使用は段階的に高度なものになり、やがては母語話者に近接していくものと推定される。

しかし、習熟度の上昇に伴って本当に使用される動詞の語数が増加するのか、習熟度ごとにどのような動詞が選択的に使用されるのか、動詞使用の点で学習者はどのような発達段階をたどるのか、動詞を手掛かりとした習熟度の推定は可能か、といった点についてははっきりしないことが多い。

こうした点を実証的に検証するには、(1)統制的なタスクデザインによって、(2)大量の学習者の産出を集め、あわせて、(3)学習者の習熟度情報を直接的に調査した学習者コーパスが必須となる。2020 年に全面公開された「多言語母語の日本語学習者横断コーパス」(I-JAS) (迫田・石川・李(編), 2020)は、(1')言語タスクの内容と手順を明確に規定し、(2')1,000 名の学習者のデータを集め、(3')その全員に 2 種の日本語習熟度テストを受験させているという点で、要件を満たす初のデータベースであると言える。

本研究では、I-JAS に含まれる各種タスクの中で、参加者が複数枚の絵を見てその内容を作文する絵描写作文課題のデータを分析することで、学習者の習熟度と使用動詞の関係を計量的に調査する。本研究は、I-JAS における学習者の副詞使用を概観した石川(2020a)、および、同じく形容詞使用を概観した石川(2020b)を引き継ぐ研究である。

2. リサーチデザイン

2.1 目的と RQ

本研究の目的は、I-JAS の絵描写作文(SW1)課題における動詞使用を量的に分析することで、L2 日本語学習者の習熟度と動詞使用量の関係、習熟度別の典型的な使用動詞、動詞習得の段階性を解明し、動詞を手掛かりとした習熟度の推定の可能性を探る点にある。これらの目的に沿って以下の研究設問(RQ)を置いた。

RQ1 母語話者と学習者間で、また、異なる習熟度の学習者間で、動詞使用量はどのように変化するか？(動詞使用量)

RQ2 母語話者と学習者、また、異なる習熟度の学習者が典型的に使用するマーカー動詞にはどのようなものがあるか？(マーカー動詞)

RQ3 主要動詞の使用傾向において、母語話者と異なる習熟度の学習者はどのように分類され、学習者の動詞習得の発達段階はどのように整理できるか？(分類)

RQ4 動詞の使用量で学習者の習熟度を推定することは可能か？(習熟度推定)

2.2 データ

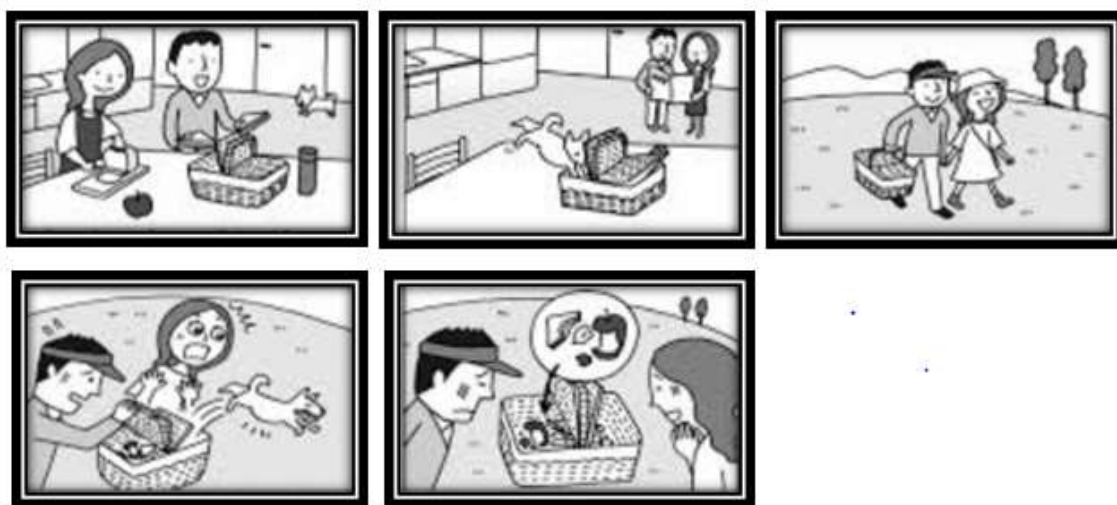
I-JAS では、2 種の絵描写作文課題(ストーリーライティング課題)を実施しているが、本研究では、そのうち、第 1 課題(SW1)のデータを使用する。これは、男女が昼食用のサンドイッチを持ってピクニックに出かけたところ、サンドイッチを入れたバスケットの中に知らぬ間に飼い犬が潜り込んでおり、ピクニック先でバスケットを開けると犬が飛び出してきて、サンドイッチは食べられていた、と

いう趣旨の 5 枚の絵を見て、その内容を作文(コンピュータ入力)するものである。なお、参加者は、同じ絵を見て内容を口頭で説明するストーリーテリング課題を先に済ませている。

I-JAS の絵描写課題は、絵を「描写」するのかそこから「創作」するのか、客観的に説明するのか昔話風に物語るのか、といった点で参加者ごとの受け止め方に若干の違いが見られ、完全な統制タスクとは言えない。しかし、対話やロールプレイといったほかのタスクに比べると、全員が同じ絵を見て産出を行っているために内容的な均質性が高く、幅広い学習者の L2 産出の言語特性を調査するという目的に適したデータである。

図 1

I-JAS の ST1/SW1 のプロンプト



2.3 対象者

本研究は、海外の日本語学習者 850 名を分析対象とし、50 名の日本語母語話者との比較を行う。学習者の母語別の内訳は、中国語母語 200 名、韓国語・英語母語各 100 名、タイ語・ベトナム語・インドネシア語・ドイツ語・フランス語・スペイン語・ロシア語・ハンガリー語・トルコ語母語各 50 名となる。

学習者コーパス研究では、母語別の分析・比較が重視される傾向にあるが、今回は、できるだけ大量の学習者の産出データを根拠として、習熟度と動詞使用の関係を探るという目的の元、分析にあたってあえて母語を区別しない方針を取る。これは、仮に母語別に分けた場合、習熟度が極端に偏ってしまうためである。なお、I-JAS に含まれる国内教室環境学習者および国内自然習得学習者は、一般的な学習者とは区別して議論すべきであると考えられるため、今回の分析対象からは除外する。

2.4 対象者の習熟度

I-JAS に参加した学習者は、全員が、SPOT (Simple Performance-Oriented Test: 音声テ-

プを聞いて文章の空欄 1 文字を埋める)と J-CAT (Japanese Computerized Adaptive Test) という 2 種類の日本語習熟度テストを受験している。2 つのテストはともに内外で広く使用されており、それぞれに利点を持つが、ここでは、J-CAT の総合スコアを日本語習熟度の目安とする。

J-CAT は、一般社団法人日本語教育支援協会が実施するテストで、聴解、文字・語彙、文法、読解の 4 セクションから構成され、試験時間は 45～90 分、配点はセクションごとに 100 点、合計で 400 点満点である。J-CAT は項目応答理論をふまえたコンピュータ適応型テストで、学習者の回答状況をふまえ、各段階で推定される習熟度に近い問題が出題される。これにより、短時間で信頼性の高い能力推定が行えるようデザインされている。

J-CAT の結果の解釈については、初級前半～超級の 9 段階別の対応スコアが示されている。また、日本語能力試験 (JLPT) およびヨーロッパ言語共通参照枠 (CEFR) との対応付けの目安も示されている (表 1)。

表 1

J-CAT のスコアの目安 (出典: <https://www.j-cat2.org/html/ja/pages/interpret.html>)

J-CAT Level	JLPT 対応付け	CEFR 対応付け
初級前半(～100)		
初級(100～)		
初級後半(150～)	N4(150～)	A2(175～)
中級前半(200～)	N3(200～)	B1(225～)
中級(250～)	N2(250～)	
中級後半(275～)		B2(275～)
上級前半(300～)	N1(300～)	C1(325～)
上級(325～)		
超級(350～)		

850 人の学習者の J-CAT 得点は 50 点台から 350 点台まで非常に広く分かれていたため、今回は、L50 (50 点台) から、L350 (350 点台) まで、10 点ごとに学習者をまとめて分析することとした。また、母語話者 (NS) 50 名も、別途、1 つの群にまとめる。

表 2 は、群別の人数を示したものである。

表 2

習熟度レベル別の人数

Lev	人数	Lev	人数	Lev	人数	Lev	人数
L50	1	L130	37	L210	57	L290	18
L60	6	L140	31	L220	55	L300	11
L70	3	L150	46	L230	47	L310	7
L80	9	L160	45	L240	39	L320	10
L90	11	L170	63	L250	46	L330	8
L100	13	L180	58	L260	33	L340	4
L110	16	L190	56	L270	26	L350	3
L120	20	L200	50	L280	21	NS	50

このうち、L50 は 1 名のみであり、結果が不安定になると思われるため、分析から除外することとした。これにより、学習者 30 群、母語話者 1 群、合計 31 群が本研究の分析対象となる。

2.5 使用動詞の抽出

国立国語研究所コーパス検索サイト「中納言」上で、上記対象者の SW1 作文の中から、品詞大区分が動詞となる用例を網羅的に抽出したところ、13,813 件を得た(検索日:2020 年 8 月 4 日)。得られたデータの全体を表計算ソフトに転記し(図 2)、J-CAT 得点で全体をソートした後、10 点ごとに、使用された全動詞語彙素を抽出し、1 つのテキストファイルにまとめた。

図 2

表計算ソフト上でのデータ

サンプル	協力者	タスク	データ	連番	発話	前文脈	キー	後	語彙	母語	J-CAT	
KKR43-SV	KKR43	SW1	第五次テ	70	10 K	<K>ケン	見	て	見る	韓国語	355	
KKR43-SV	KKR43	SW1	第五次テ	90	10 K	<K>ケン	いる	間	居る	韓国語	355	
KKR43-SV	KKR43	SW1	第五次テ	180	10 K	<K>ケン	入り	まし	入る	韓国語	355	
KKR43-SV	KKR43	SW1	第五次テ	270	20 K	<K>見	て	出	まし	出る	韓国語	355

学習者 30 群、母語話者 1 群、合計 31 種のテキストファイルから、それぞれ、動詞頻度表を作成し、それらを新しい表計算ソフトのシートにまとめ、以後の作業ベースとした。

上記ファイルを調査したところ、対象者によって 1 度以上使用された動詞は全 254 種であった。ただし、母語話者(JJJ07)の使用動詞とされる「驚き困る」は、文脈を確認したところ、「(二人は)驚き、困って…」のように 2 語と読むべきものと判断されたため、手作業で当該語を削除し、その頻度を「驚く」と「困る」の頻度に付け替えている。これにより、調査対象動詞は 253 種となった。

その後、群ごとに、使用された動詞のトークン数(総語数)とタイプ数(種別数)を調べた。各群の人数が異なるため、トークン数については 1 人あたりの調整頻度に換算した。一方、タイプ数については、人数に比例して線形増加するわけではないので、得られた値をそのまま分析対象とした。

2.6 手法

RQ1(動詞使用量)については、1 人あたりの調整トークン数と、タイプ数について、学習者 30 群の平均値と母語話者の値を比較した。その後、母語話者の動詞使用量を 100 として、学習者 30 群の動詞使用量を比率値に変換し、L60～L350 間の変化をグラフとして示した。

RQ2(マーカー動詞)については、2 つの観点から分析を行った。まず、全 253 種の動詞を対象に、学習者 30 群の総体と母語話者を比較し、母語話者のみを使用する母語話者マーカー動詞と、学習者のみを使用する学習者マーカー動詞を特定した。次に、母語話者・学習者全 31 群中のレンジが 20 以上になる基本動詞 35 種(表 3)を対象に、学習者群ごとに個々の動詞の使用頻度と J-CAT スコア(10 点刻みの値)の相関値を求めた。その後、正負ともに、中程度の相関($r \geq |0.4|$)を示す語を抽出し、習熟度と負相関する(習熟度が上がると使用頻度が下がる)初級マーカー動詞と、習熟度と正相関する(習熟度が上がると使用頻度も上がる)上級マーカー動詞を特定した。

表 3

分析に使用した基本動詞 35 種(レンジ 20 以上)

分析対象動詞(レンジ数)
居る／食べる／行く／入る／見る／出る(31)／為る／作る／有る(30)／開ける(29)／仕舞う／来る／入れる／成る(28)／知る(27)／歩く(26)／持つ／飛び出す／驚く／出来る(25)／気付く／着く(24)／思う／言う／終わる(23)／決める／分かる／出す(22)／飼う／飛ぶ／出掛ける／開く／置く／空く(21)／探す(20)

RQ3(分類)では、前述の基本動詞 35 種について、動詞を第 1 アイテム、参加者群を第 2 アイテムとするクロス表を作成し、対応分析を実施した。対応分析とは、クロス表の行列の相関を最大化する次元を取り出し、通例、寄与率が高い第 1 次元と第 2 次元を 2 つの軸として散布図を作成する多変量解析手法である。散布図上では、2 種のアイテムカテゴリデータが同時に布置され、関係が近いものは近傍に布置される。今回のデータから得られた散布図を根拠として、L60～L350 の学習者と母語話者がどのように分類されるか、また、それぞれを特徴付ける動詞にはどのようなものがあるかを検証した。

RQ4(習熟度推定)では、学習者全 30 群のうち、レンジ数が 29 以上になる動詞 9 種を対象に、1 人あたり使用量(トークン数)を説明変数、J-CAT スコア(10 点台ごとの中央値である 5 点に変換。たとえば、L80 は 85 点、L200 は 205 点とする)を目的変数として重回帰分析を実施した。

表 4

重回帰分析の説明変数候補とした動詞 9 種(レンジ 29 以上)

分析対象動詞(レンジ数)
居る／見る／入る／出る／食べる／行く(30)／為る／作る／有る(29)

重回帰分析では、手元のデータすべてを使ってモデルを作ること多いが、その場合、人数の少ない学習者群を含めることで回帰式の妥当性が下がったり、得られたモデルを別のデータで検証できなくなったりする危険性がある。そこで、今回の分析では、学習者群 30 種のうち、群内に 10 人以上の学習者を含む 23 群(L90～L320)のみを分析対象とすることにした。また、後者の問題点への対処として、23 群を 2 セット(セット A, セット B)に分割し、片方のみで回帰式を作り、それを残りのセットに適用してモデルの妥当性を検証することとした。セット A は L90, L110, L130, L150, L180, L190, L210, L230, L250, L270, L290, L320 を、セット B は L100, L120, L140, L160, L180, L200, L220, L240, L260, L280, L300 を含む。

重回帰分析には、用意した変数をすべて使用する全変数投入法のほか、変数を選択するステップワイズ投入法がある。今回は、9 種の変数に対して、ステップワイズ増減法で選択を行うこととした。なお、変数の投入・除去基準は $p=.2$ とした。

3. 結果と考察

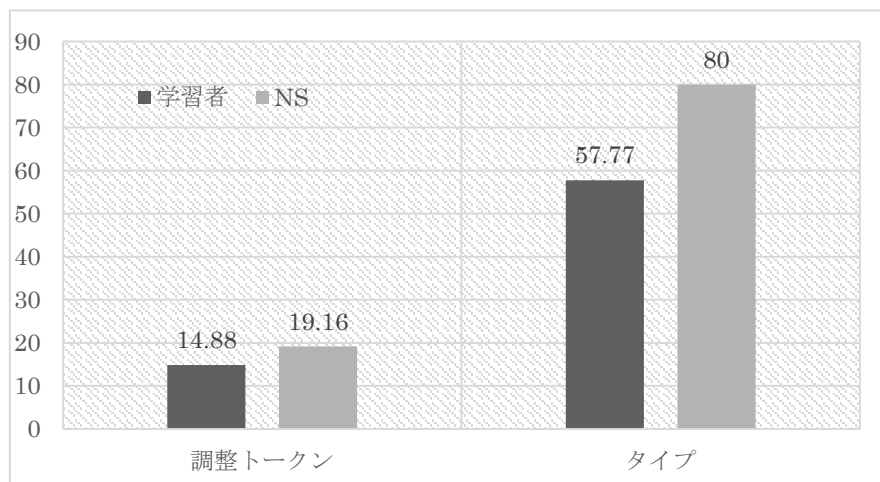
3.1 RQ1 動詞使用量

3.1.1 学習者・母語話者間比較

学習者(30 群平均)と母語話者の 1 人あたり調整トークン数とタイプ数を調査したところ、図 3 の結果を得た。

図 3

学習者と母語話者(NS)の動詞使用量



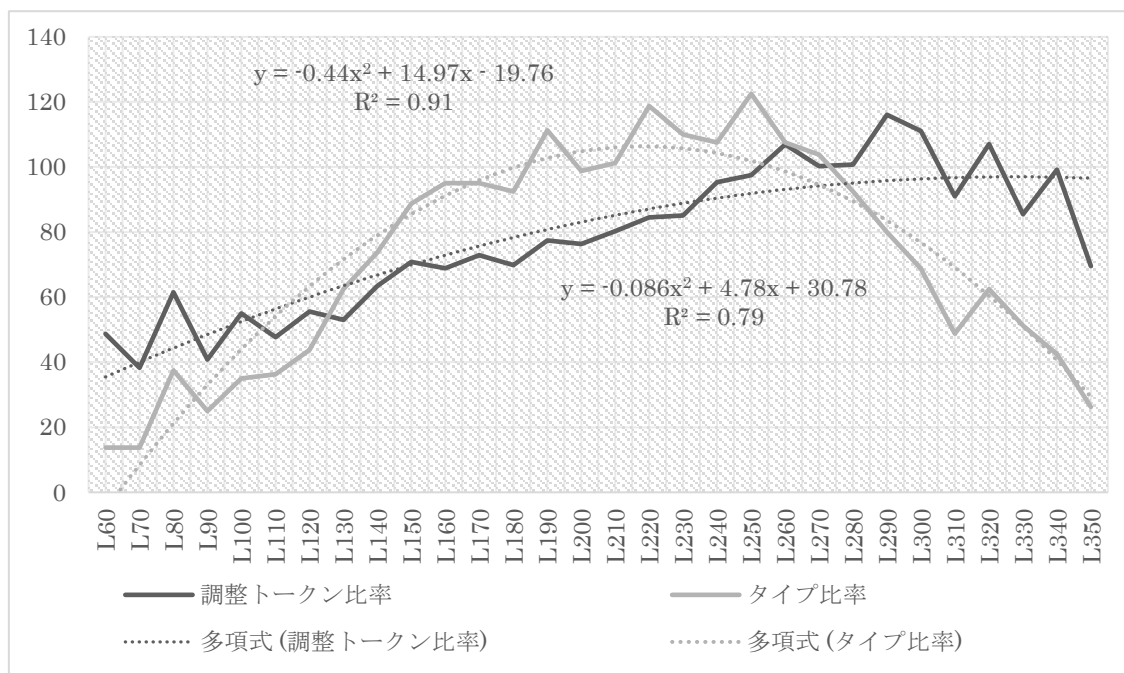
同じ絵を描写したわけであるが、母語話者に比べ、学習者が使用した動詞のトークン数は約 22% (4~5 語)、タイプ数は約 28% (約 22 種) 少ないことがわかった。この差は、学習者と母語話者の動詞知識の乖離を反映していると考えられる。

3.1.2 習熟度間比較

次に、母語話者の動詞使用量(トークン数・タイプ数)を 100 として、習熟度別の学習者群の動詞使用量を比率値に換算し、習熟度による比率の変化を調査したところ、図 4 の結果を得た。なお、図 4 には、それぞれの比率値のほか、比率値の変動をスムージングした多項(二次)近似曲線と、曲線の式を示している。

図 4

習熟度別の動詞使用量の変化



調整トークン数については、L60 から比率値が増加し、L250 で母語話者と並んだ後、L290 で母語話者より 16%ほど増え、それ以降、減少する傾向が認められた。また、タイプ数については、L60 から比率値が急速に増加し、L190 で母語話者と並んだ後、L250 では母語話者より 20%以上増え、それ以降は大きく減少する傾向が認められた。

今回のデータは、初級と上級レベルの人数が中級レベルに比べて少ないため、そのことで、グラフの両端付近の値が不安定になっている可能性はあるものの、データに基づく限り、トークン数・タイプ数ともに、初級者より中級者にかけて大きく増加し、母語話者を一時的に上回った後、中上級レベルになると、ふたたび減少するという逆 U 字型の変化パターンを見せていることになる。とくに、タイプ数については、二項近似でもそのことがはっきり示されている。

動詞使用量に関するこうした逆 U 字型の変化は、第 2 言語習得で広く見られる U 字型発達曲線 (U-shaped behavioral development) の裏返しとしてとらえることができるだろう。たとえば、言語産出の正確性は、初級者は高く、中級者は低く、上級者は再び高くなることが知られている (Kellerman, 1985)。初級者は確実に使えるものだけを用心深く使用するため、誤りは少なく、正確性は高い。一方、中級者は誤用を恐れず多種多様な表現を試すため、誤りが増え、正確性は低下する。しかし、そうした試行錯誤の体験を通して正しい言語規則を内在化した上級者になると、誤用は減り、正確性は再び上昇する。

動詞使用量は、U 字型発達の鏡像になっている可能性がある。つまり、初級者はそもそも知っている動詞が少ないため、また、発話量そのものが少ないため、動詞の使用量は少ない。しかし、新しい動詞を多く学んだ中級者は、覚えてたての多種・多量の動詞を試行的に使用するため、動詞の使用量は場合によって母語話者を上回る。しかし、試行錯誤の過程を経て、動詞の運用ルールを自分なりに確立した上級者になると、文脈に応じた適切な動詞を適量使用するようになり、結果的に、動詞の使用量はむしろ低下していくものと考えられる。

3.2 RQ2 マーカー動詞

まず、253 種の動詞すべてを対象に、学習者の全群と母語話者を比較することで、(a)母語話者マーカー動詞(母語話者のみが使用して学習者は使用しない)9 種、(b)学習者マーカー動詞(学習者のみが使用して母語話者は使用しない)173 種が特定された。ただし、(b)の多くは、特定の習熟度群の学習者のみが使用しているもので、学習者全体の傾向を表すものとは言えない。そこで、学習者 30 群中のレンジが 10 以上のものに限ったところ、15 種となった。

その後、母語話者を含む全 31 群中のレンジが 20 以上になる基本動詞 35 種を対象に、習熟度スコア(10 点刻み)と動詞頻度の相関を調べ、中程度あるいはそれ以上の相関を示す動詞を抽出した。これにより、(c)習熟度と負相関となる初級マーカー動詞(初級者が多く使用し、中上級になると使用頻度が下がる)3 種と、(d)習熟度と正相関となる上級マーカー動詞(初級者はあまり使わないが、中上級者になると使用頻度が上がる)19 種がそれぞれ特定された。

表 5 は、これら 4 種の動詞群を一覧にしたものである。母語話者・学習者マーカーは 1 人あたりの平均頻度 (fa)、初級・上級者マーカーは相関強度 (r) によって 2 つのグループに整理して示している。

表 5

タイプ別マーカー動詞

マーカー種別	該当動詞
母語話者マーカー (9 種)	$fa \geq 0.1$: 詰める $fa < 0.1$: 漁る／歩き疲れる／荒らす／参る／抱える／掛ける／紛れ込む／詰め込む

学習者マーカー (15 種)	$fa \geq 0.1$: 終わる／飛ぶ $fa < 0.1$: 座る／読む／見付ける／選ぶ／返る／残る／調べる／遊ぶ／見える／走る／隠れる／連れる／見付かる
初級者マーカー (3 種)	$r \geq 0.4$ (相関中): 有る／出す／入れる
上級者マーカー (19 種)	$r \geq 0.7$ (相関強): 飛び出す／知る／気付く／飼う／居る／来る $r \geq 0.4$ (相関中): 着く／終わる／空く／仕舞う／為る／持つ／決める／出掛ける／開ける／驚く／思う／成る／開く

以下、母語話者マーカーと、学習者系のマーカーに分けて文脈を概観していく。

3.2.1 母語話者マーカーの使用文脈

850 名の学習者が一度も使用していない母語話者マーカーは 9 種ある。これらは、形式・用法・意味の点で一定の傾向性を有する。

形式面について言えば、複合動詞が多く、「(二人は) 歩き疲れる」「(犬がバスケットに) 紛れ込む」「(サンドイッチをバスケットに) 詰め込む」などがその例である。辞書の見出し語調査では、現代日本語における複合動詞の数は 1,817 語におよび、単純動詞の 2,083 語に近接する数があるとされる(森田, 2008, p. 40)。複合動詞は日本語の動詞体系の中核を占めるものであるが、学習者には運用が難しいとされ、陳(2008)によれば、学習者の複合動詞使用頻度は母語話者の三分の一程度であるとされている。この意味で、母語話者マーカーの三分の一が複合動詞であったことは注目に値する。

次に、用法面で言えば、慣用表現中での使用がいくつか見つかる。「(サンドイッチをバスケットに) 詰める／詰め込む」や、「(腕によりを) 掛け(てサンドイッチをつくる)」「参っ(たなあ)[嘆息のせりふとして]」などがその例である。これらの動詞は、決まった共起語や文脈とセットになって使用され、通例、類義動詞での言い換えができない。また、教科書に掲載される典型的な語義からはかなり逸脱している。こうした慣用表現を使いこなすことは上級学習者にとっても困難であると思われる。

最後に、意味の面で言えば、評価的ニュアンスを含む動詞が含まれており、「(犬がバスケットの中の食べ物を) 漁る」や「(犬がサンドイッチを) 食い荒らす」などがこの例である。これらは、話者の視点から見た被害を強調し、犬への貶めのスタンスを含意するものとなっている。一般に、敬意を表出するポライトネス表現は日本語においてとくに習得困難な学習項目の 1 つとされるが、敬意のみならず、反感や不快のスタンスの表出もまた習得しにくい項目と言えよう。

以上をまとめると、母語話者のみが使用する動詞は、(1)複合動詞、(2)慣用表現中で派生的語義で使用される動詞、(3)評価ニュアンスを含む動詞に区分され、これらを使用できるかどうか、上級者を含む学習者と母語話者を分かつ垣根になっていると言える。

3.2.2 学習者系マーカースの使用文脈

3.2.2.1 学習者マーカース

学習者系マーカースには、学習者マーカース・初級者マーカース・上級者マーカースの 3 つがあるが、ここではまず、50 名の母語話者が一度も使用していない学習者マーカース 15 種を概観する。この中には誤用も一定数含まれている。15 種の動詞は 4 つのタイプに分類できる。

1 つ目は、意思の有無によって使い分けがなされる動詞群で、(終える／)「終わる」、(見る／)「見える」、「見付ける」／「見付かる」などが該当する。このうち、とくに頻度が高いものは「終わる」である。

(1) ケンとマリはサンドイッチ作りを終わって… (RRS36:L160)

(2) せっかく作ったサンドイッチを犬が全部食べ終わったのでした… (KKR29:L290)

学習者は、サンドイッチを作ったり食べたりする行為の完了を含意する文脈で自動詞の「終わる」を使用しているが、母語話者は、こうした場合、通例、他動詞の「終える」を用いる。一般に、意思や意図を前提とする行為の完了は他動詞で、意思を前提としない行為の完了は自動詞で表現するわけだが、学習者はこの点の使い分けが曖昧で、結果として、意思を前提とした行為(サンドイッチを作る／食べる)の完了に自動詞の「終わる」を多用している。

このほか、いわゆる「見る」系動詞に関して、「(犬／サンドイッチを)見えない」、「(食べ物／食べられたことを)見つける」、「(犬／サンドイッチ／食べる場所／食べたことを)見付かる」などの用例も認められる。「見る」系動詞については、意図せず自然と目に入ってくる(自動詞の「見える」)→意図をもって対象を凝視する(他動詞の「見る」)→意図をもって探していた物を発見する(他動詞の「見付ける」)／意図をもって探していた物が発見される(自動詞の「見付かる」)のように、自他性の違いや、意思性・意図性の強弱によって異なる動詞を選択する必要があるが、学習者はしばしばこれらを混同し、文脈に合わない不自然な動詞を選択している。

2 つ目は、上記とも関連するが、文脈に合致しない動詞群で、「読む」「調べる」「連れる」などが相当する。このうち、前の 2 語は「(地図を)読む／調べる」という形で多く用いられている。このように言えば、ふつう、机の上に地図を広げ、場合によってはルーペなども使いながら地図を詳細に分析している姿が思い浮かぶが、実際には、二人は立ったまま手に持った地図をざっと眺めているだけで、はっきりした意思や意図をもって地図を吟味しているわけではない。また、「連れる」は、「(犬を)連れ(ていく／てくる)」という形で用いられている。このように言えば、たとえば動物病院で犬を診察してもらう場合など、話者の側に犬を連れだす明確な意思や意図があると感じられるが、実際には犬が知らぬ前に入っていたというだけで、話者の側ではそうした意図はない。1 点目と合わせ、意思や意図の有無ないし強弱による動詞の選択が、学習者の苦手項目であることが示唆される。

3 つ目は、複合動詞の構成成分となる動詞群で、学習者は、しばしば、「飛ぶ」を「来る」につな

げて使用している。

(3) 犬はバスケットの中に飛んできた…(IID25:L150)

(4) 犬がバスケットの中から飛んできて…(CCT24:L 320)

「飛んでくる」という表現自体は間違っただけのものではないが、上記の文脈では、それぞれ、バスケットに「飛び込む」、またはバスケットから「飛び出す」などの複合動詞を使用すべきであろう。すでに述べたように、学習者の多くは複合動詞を苦手としており、ここでも同様の問題が確認できる。また、上記には視点の問題も認められる。母語話者が絵の世界の外側に視点を置いて全体を客観的に描写しがちであるのに対し、学習者はしばしば絵の中の登場人物に視点を置いている。

最後に、4 つ目は、絵を描写する上で、必ずしも言う必要のない行為を言語化したことで使われた動詞群である。これらは、「(芝生に)座る」「(地図／公園／座る場所を) 見つける」「(ピクニックの場所／公園で座る場所を) 選ぶ」「(家に) かえる」「(犬が) 隠れる」「(公園で) 遊ぶ／遊び(にいく)」 「(犬が) 走っ(てバスケットに入る)」 「(サンドイッチが) 残っ(ていない)／(残飯が) 残る」のような形で使用されている。母語話者は絵に直接描かれていない内容を言語化することはほとんどなく、結果的に学習者のみがこうした表現を使用している。

以上をまとめると、学習者のみが使用する動詞は、(1)意思性や意図性に関わる使い分けを誤った動詞、(2)とくに意思性や意図性の点で文脈と合致していない動詞、(3)複合動詞構成要素、(4)絵に描かれていない行為を表す動詞に区分される。なお、(4)については、すでに述べたように、学習者の一部が、与えられたタスクを、絵の描写ではなく、絵をもとに創作を行うことだと受け取っていたためとも考えられる。

3.2.2.2 初級者マーカー

頻度と習熟度が負相関を示し、初級者は多く使うが、中上級者になると使わなくなる初級者マーカーは「有る」「出す」「入れる」の 3 つであった。これらについて、用例を見ておこう。

(5) バスケットを開けた時は食べ物がありません…(IID54:L100)

(6) 犬はバスケットの中にある食べ物を食べました…(TTR16:L100)

(7) そこで犬はバスケットの中から出しました…(TTR55:L70)

(8) ケンがバスケットを開きました、でも犬が出しました。(SES58:L80)

(9) 犬をピクニックのバスケットの中に入れました。(TTR36:L90)

(10) 犬はバスケットの中に入れます。(SES47:L100)

「有る」の用例は、文法的誤用とまでは言い切れないものの、若干の不自然さが感じられる。そもそも日本語において「～がある」という言い方は必ずしも一般的ではなく、上級者や母語話者であ

れば「有る」を使う代わりに、(5)については「食べ物が(残って)いなかった」のような、(6)については「バスケットの中の食べ物…」のような表現を使うようになる。一方、初級の学習者は、物が存在する場合は機械的に「有る」を選択しがちである。

「出す」については、「犬が(バスケットから)出る」を誤用した「犬が出す」という形が初級者に散見される。また、「入れる」については、「犬が(バスケットに)入る」を誤用した「犬が(を)入れる」という誤用が散見される。すでに述べたように、意思や意図を前提とする他動詞と、それらを前提としない自動詞の区別は日本語学習者の間違いやすい項目の 1 つであり、そうした誤用は初級者に多く見られるが、習熟度の上昇につれて減少し、上級者ではほとんど見られなくなる。

以上より、初級者が多く使う動詞は、(1)存在を表す「有る」と、(2)自他の使い分けエラーによる動詞であることが確認された。

3.2.2.3 上級者マーカー

一方、頻度と習熟度が正相関を示し、初級者はほとんど使わないが、中上級者になると広く使うようになる上級者マーカーは 19 種であった。これらは大きく 4 種に区分できる。

1 つ目は、複合動詞(飛び出す／出掛ける)である。初級者は「(犬がバスケットから)出る」や、「(二人がピクニックに)行く」と言うことが多いが、上級者になると、文脈により適切な複合動詞が使えるようになる。

(11) 中から犬が飛び出してきて…(CCM51:L350)

(12) 二人は気づかないまま外に出かけましたが…(TTH09:L330)

「出る」でなく「飛び出す」を使うことで、犬が勢いよくバスケットから出たことが伝わる。また、「行く」でなく「出掛ける」を使うことで、二人にとってのピクニックが、意図的・計画的・本格的なものではなく、すぐに「帰ることを前提として(ちょっとの間だけ)外出」(『明鏡国語辞典』第 2 版)したものであったことが描写できる。

2 つ目は、知覚に関わる動詞(知る／気付く)である。上級学習者は、これらを否定形にして、副詞的に使用している。

(13) 二人はそれを知らずに…(KKD28:L350)

(14) 二人はそれに気付かず…(CCM51:L350)

3 つ目は、主動詞に添えてアスペクトを追加したり(居る／来る／仕舞う)や、漢語や外来語名詞に添えてそれらを動詞化したり(為る)、「～に」に後続して状態変化を示したりする(為る／成る)補助的・機能的な動詞群である。

- (15) ケンとマリが地図を見ている間に… (KKR43, 355)
- (16) 犬がバスケットから出てきました… (CCS39, 347)
- (17) 食べ物は全部食べられてしまいました (CCM50, 310)
- (18) どこにピクニックしに行こうかと、地図を見ながら相談しています。(HHG46:L300)
- (19) すごく楽しみにしていたピクニックなので台無しにならないように… (KKR10:L300)

森田(2008)は、「日本語では、対象の上に生ずる行為や現象をそのまま生の動詞でもって言い表すと、どうしても抽象的になりやすい」として、たとえば、水が「流れる」と言うのではなく、水が「流れている／来る」あるいは「流れて行く／しまう」など、「補助的な動詞」を添えることで、「具体的な『場』における話者の視点から」状況をとらえなおし、抽象的状況が「具体的な状況」になると述べている(p.125)。こうした補助動詞をうまく使えることが上級学習者の特徴になっている。

4 つ目は、慣用表現の一部となる動詞である。例えば、「お腹が空いて」(CCM35:L320)などがこれに当たる。

5 つ目は、登場人物や犬の行為を詳しく説明する動詞群で、「(犬を)飼う」「(〜と)思う」「(公園に)着く」「(犬が食べ)終わる」「(バスケットを)持つ(てくる)」「(行き先の場所を)決める」「(バスケットを)開く・開ける」「(飛び出した犬に)驚く」などの形で使用されている。このうち、「開く」と「開ける」は自他の識別が難しいペア動詞だが、初級者と異なり、中上級者はこうした語も正しく使えるようになっている。

以上より、上級者が多く使う動詞には、(1)複合動詞、(2)副詞句内で使用される否定形知覚動詞、(3)アスペクト等を添加する補助動詞、(4)慣用表現中の動詞、(5)登場人物の動作を詳しく説明する動詞群、などが含まれることが確認された。

本節では、母語話者マーカー動詞を見た後、学習者マーカー動詞、初級者マーカー動詞、上級者マーカー動詞の使用環境を具体的に概観してきた。全く同じ絵を描写する課題でありながら、学習者と母語話者間で、さらには習熟度の異なる学習者間で、それぞれ典型的なマーカー動詞が存在するというのは、学習者の動詞獲得プロセスを考える上で興味深い発見であると言える。

3.3 RQ3 動詞頻度による分類

RQ1 および RQ2 の検討を通して、学習者と母語話者間、また、異なる習熟度の学習者間で、動詞の使用量についても、また、典型的に使用される動詞についても一定の差異があることが確認された。では、母語話者を含む全 31 群中のレンジ数が 20 以上になる 35 種の基本動詞の使用頻度を手掛かりとした場合、学習者の各群および母語話者はどのように区分され、それらはどのような動詞で特徴づけられるのであろうか。学習者群がいくつかのかたまりに分かれるかを見ることで、学習者による動詞獲得過程がどのように進行しているかを考える手掛かりが得られる。

対応分析で得られた第 1 軸の寄与率は 33.8%、第 2 軸の寄与率は 13.8%であり、両軸で全体の分散の約半分が説明できることが確認された。ゆえに、第 1 次元を横軸、第 2 次元を縦軸として

散布図を描いたところ、図 5 の結果が得られた。寄与率は第 1 軸のほうが大きく、分析対象のデータは、まずもって第 1 軸で左右に分割され、次いで、第 2 軸で上下に分割されることになる。

ここで、学習者の布置状況に注目しながら 2 つの軸の意味を考えてみたい。まず、第 1 軸について、左側には L60～L180 の学習者が、右側には L190～L350 の学習者と母語話者 (NS) がかたまっている。また、左側に注目すると、左端半分 (座標値が $-2 \sim -0.8$ あたり) に入門段階に相当する L60～L100 の学習者が、その右側半分 (座標値が $-0.8 \sim 0$ あたり) にそれに続く L110～L180 の学習者がかたまっている。これらの事実から、第 1 軸は、入門者・初級者を左方に、それ以外の中上級者および母語話者を右方に分割する軸と解釈できる。

次に、第 2 軸について、左側では上下の差がはっきりしないものの (L60 と L70 はそれぞれ上下に分かれているが、これは各々が「有る」または「出す」という異なる動詞を典型的に使用していることによる)、右側を見ると、上側には L190～L270 の学習者が、下側には L260～L350 の学習者と母語話者がかたまっている。このことから、第 2 軸は、中上級者のうち、中級者を上方に、上級者と母語話者を下方に分割する軸と解釈できる。

2 つの軸は、いずれも習熟度に関係するものであるが、第 1 軸は入門者・初級者とそれ以外を、第 2 軸は中級者とそれ以外を分かちものになっており、学習者の動詞獲得過程が L60～L180 と、K190 以降とで質的に変容している可能性が示唆される。以上をあわせて考えると、学習者の動詞獲得段階 (Verb Acquisition Stage) は、第 1 軸左側の左端部 (VAS1)、その右隣部 (VAS2)、右側上部 (VAS3)、右側下部 (VAS4) の 4 つに区分でき、動詞の習得は VAS1～VAS4 の順に進んでいくものと考えられる。

表 6 は、従前の習熟度段階と、今回の分析から新たに推定された動詞獲得段階 (VAS) を対応させて表示したものである。VAS の各段階には関係の深い動詞項目を併せて示している。

表に示されるように、およそ入門～初級の入り口に相当する第 1 段階の学習者は、ごく限られた種類の動詞しか使用できない。その中には、「(犬が)出す」や「(食べ物)有る」のように、自他の誤用や、日本語として不適切な動詞の使用も含まれる。その後、初級に相当する第 2 段階に入ると、学習者の動詞知識は大きく拡張し、広義の移動動詞 (入る／歩く／出る／行く／入れる) や基本的な行為動詞 (探す／置く／見る／食べる／作る) を含め、幅広い事象を言語化できるようになる。

表 6

習熟度段階と動詞獲得段階

J-CAT Level	JLPT 対応付け	CEFR 対応付け	動詞獲得段 階(VAS)	VAS 段階別特徴動詞
初 級 前 半 (～100)			VAS1 (～100)	出す／有る
初級 (100～)			VAS2 (110～)	入る／探す／置く／歩く／見る／出 る／食べる／行く／出来る／入れる ／作る
初 級 後 半 (150～)	N4 (150～)	A2 (175～)		
中 級 前 半 (200～)	N3 (200～)	B1 (225～)	VAS3 (190～)	着く／仕舞う／思う／開く／言う／為 る／飛ぶ／開ける／分かる
中級 (250～)	N2 (250～)			
中 級 後 半 (275～)			VAS4 (260～)	出掛ける／気付く／飼う／飛び出す ／決める／空く／持つ／終わる／驚 く／知る／来る／成る／居る
上 級 前 半 (300～)	N1 (300～)	B2 (275～)		
上級 (325～)				
超級 (350～)		C1 (325～)		

続いて、第 3 段階に入ると、これまでの動詞獲得過程が次第に質的に変容し、単に知っている動詞の数が増えるというだけでなく、より高度な動詞使用ができるようになってくる。たとえば、思考・発話動詞(思う／分かる／言う)や、基本語の自他対(開く／開ける)、また、否定的なスタンスを表出する補助動詞(仕舞う)や、名詞に付加されて動詞表現の幅を広げるスル動詞(為る)などである。

そして、第 4 段階に入ると、動詞の獲得は一層深化し、母語話者同様、複合動詞(出掛ける・飛び出す)、知覚判断動詞(気付く・決める・知る・驚く)、アスペクトの意味を添加する各種の補助動詞(～に成る／～して居る／～しに来る)などが使えるようになってくる。

3.4 RQ4 習熟度推定

最後に、動詞頻度を手掛かりとした習熟度推定の可能性について考えたい。一定の人数を有する L90～L320 の学習者のうち、セット A に含まれる 12 の学習者群について、9 種の動詞の 1 人あたり使用頻度を説明変数、J-CAT スコア(10 点台ごとの中央値)を目的変数としてステップワイズ

による重回帰分析を実施したところ、5 種の動詞がモデルに採用され、以下の結果が得られた。

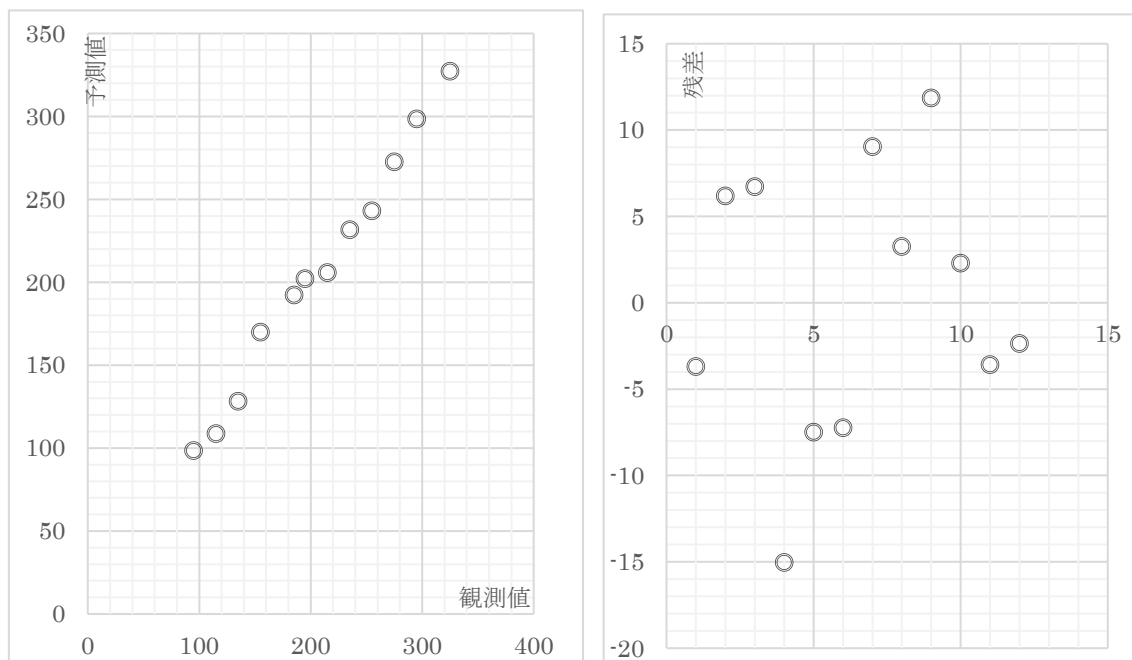
表 7

重回帰モデルに選ばれた変数

変数	偏回帰係数	標準誤差	標準偏回帰係数	F	t	p	VIF
居る	34.42	11.54	0.31	8.89	2.98	.02	5.41
見る	35.97	24.77	0.10	2.11	1.45	.20	2.38
為る	52.91	11.85	0.51	19.93	4.46	<.01	6.70
作る	107.86	20.14	0.29	28.67	5.35	<.01	1.50
有る	-150.57	36.74	-0.19	16.80	-4.10	.01	1.14
定数項	9.05	35.74	---	0.06	0.25	.81	---

図 6-7

観測値・予測値対応(左)／残差プロット(右)



注:図 7(右)の X 軸は項目番号で、セット A の L90, L110…L320 に対応する。

上記より, J-CAT スコアは下記の式で説明できることになる。

$$y = 34.42 * \text{「居る」頻度} + 35.97 * \text{「見る」頻度} + 52.91 * \text{「為る」頻度} + 107.86 * \text{「作る」頻度} \\ - 150.57 * \text{「有る」頻度} + \text{定数 } 9.05$$

L90～L320 の範囲に限った推定ではあるが、図 5 の予測値・観測値グラフに示唆されるように、モデルの精度は高く、決定係数(修正 r^2)は.9786 となっている。係数は、「見る」と定数項以外、有意である。残差 VIF において基準値 10 を超えるものはなく、多重共線性の問題はないと判断できる。

得られたモデルの妥当性を検証するため、モデルづくりに使用しなかった 11 群のデータ(セット B)を使い、上記の式に頻度を代入して推定値を算出し、実際の値と比較したところ、以下の結果を得た。誤差は推定値－実測値の値、誤差率は実測値に対する誤差の絶対値の比率(%)である。

表 8

回帰モデルに基づく 5 種動詞頻度からの習熟度推定(セット B)

	説明変数					目的変数		誤差	
	居る	見る	為る	作る	有る	推定値	実測値	誤差	誤差率
L100	1.000	1.231	1.077	0.462	0.769	78.8	105	-26.2	25.0
L120	0.850	1.400	1.100	0.500	0.350	148.1	125	23.1	18.5
L140	0.968	1.129	1.548	0.548	0.226	190.0	145	45	31.0
L160	1.022	1.467	1.711	0.333	0.244	186.7	165	21.7	13.2
L180	0.828	1.328	1.397	0.345	0.172	170.5	185	-14.5	7.8
L200	1.120	1.200	1.56	0.540	0.160	207.5	205	2.5	1.2
L220	1.236	1.364	2.109	0.418	0.218	224.5	225	-0.5	0.2
L240	1.667	1.436	2.000	0.513	0.256	240.7	245	-4.3	1.8
L260	2.424	1.455	2.091	0.455	0.424	240.7	265	-24.3	9.2
L280	2.571	1.381	2.429	0.333	0.143	290.1	285	5.1	1.8
L300	2.545	1.000	3.364	0.091	0.000	320.4	305	15.4	5.0

L100～L160 の学習者群では最大 31%の誤差が生じているが、L180 以降、推定誤差は安定して小さくなる。誤差率の平均値は全体で 10.43%となり、CEFR の A2 に相当する L180 以降に限れば、3.86%となる。わずか 5 動詞の頻度による推定としては、実用的なレベルの推定モデルになっていると結論してよいだろう。

4. まとめ

4.1 得られた知見

本研究では、I-JAS の SW1 タスクにおける産出データを用い、海外の日本語学習者 850 人を J-CAT の 10 点刻みで群化した上で、習熟度の発達と動詞使用の関係を量的に概観した。その結果、以下の知見が得られた。

まず、RQ1(動詞使用量)については、学習者の使用する動詞のトークン数およびタイプ数の平均は母語話者より22~28%程度少ないこと、また、習熟度別に見た場合、タイプ数はL190前後、トークン数はL250前後で、それぞれ母語話者の使用量に到達し、その後、使用量はさらに増加するが、タイプ数はL250以降、トークン数はL290以降、減少に転じ、両者ともに逆U字型の分布を見せることが確認された。

RQ2(マーカー動詞)については、母語話者マーカー動詞(母語話者のみが使用して学習者は使用しない)9種、学習者マーカー動詞(学習者のみが使用して母語話者は使用しない)173種(うち、レンジ10以上が15種)、初級マーカー動詞(初級者が多く使用し、中上級になると使用頻度が下がる)3種、上級マーカー動詞(初級者はあまり使わないが、中上級者になると使用頻度が上がる)19種を特定した。母語話者マーカー動詞には、複合動詞、慣用表現中の動詞、評価ニュアンスの含まれる動詞群などが含まれ、学習者マーカー動詞には、意思性に関わる使い分けを誤った動詞、意思性の点で文脈と合致していない動詞、複合動詞構成要素、プロンプトとしての絵に直接描かれていない行為を表す動詞などが含まれる。

RQ3(動詞頻度による分類)については、対応分析により、動詞発達過程が、およそ、VAS1(〜100:出す／有る)、VAS2(110〜:入る／探す、など)、VAS3(190〜:着く／仕舞う、など)、VAS4(260〜:出掛ける／気付く、など)の4段階に区分できること、L190に相当するVAS2とVAS3の間に最も重要な境界が存在することが確認された。

最後に、RQ4(習熟度推定)については、L90~L320の学習者の習熟度推定に関して、J-CAT 総合スコア = $34.42 * 「居る」頻度 + 35.97 * 「見る」頻度 + 52.91 * 「為る」頻度 + 107.86 * 「作る」頻度 - 150.57 * 「有る」頻度 + 定数 9.05$ という回帰式が得られた。回帰式の妥当性について、モデルの作成に使用していない別データで検証したところ、約10%(A2以上であれば約4%)の誤差でスコアを推定できることが確認された。

4.2 教育的応用の可能性

では、これらの結果をどのように日本語教育に応用していくことができるのだろうか。たとえば、動詞指導においては、今回の研究で習熟度と相関が強いとされたものについて各段階で適切に指導していくことができるだろう。とくに複合動詞や補助動詞については、インプットからの自然習得だけに任せず、明示的に焦点を当てた指導も必要になるとと思われる。

また、SW1の絵を利用し、回帰式に含まれる動詞の使用状況を確認することで、学習者の習熟度の簡易判定が可能になるかもしれない。初級者マーカーである「有る」を使っていない場合は、「(サンドイッチを)作る」、「(確認・専念・到着・失望など／〜しようと) + 為る」、「(地図などを)見る」、「(〜して) + 居る」の4語の頻度を数えるだけで、高い精度でJ-CATスコアが予測できることになる。もっとも、こうした習熟度推定を行う場合は、統計的推定の鍵となるマーカー動詞を学習者から秘匿しておくことが必須である。

4.3 今後の課題

さて、以上で知見の整理と教育的応用の可能性をまとめたが、本研究は試行的なもので、さらになすべき作業は多く残っている。ここでは 3 点に絞って述べる。

1 点目は、分析するデータの拡張と、分析対象の拡張である。前者については、今回使わなかった SW2, さらには I-JAS 内のその他のタスク産出データを用いて同様の分析を行う必要がある。後者については、語彙素レベルだけでなく、活用形を含めた動詞の分析や、動詞以外の品詞を組み込んだ分析が必要だろう。

2 点目は、母語別の比較である。今回の分析では学習者の総数を増やすことを優先し、あえて母語の差を見なかったわけであるが、母語の影響についても検証の必要があるだろう。ただし、中国語と韓国語話者を除くと、上位層の数は著しく少なくなり、母語を組み込んだ分析では、幅広い習熟度の学習者のモデル化が困難になる恐れもある。

3 点目は、分析結果の再現性の検証である。およそこうした研究は、データが変われば結果が変わる危険性と常に隣り合わせである。学習者の習熟度と語彙レベルの関係を調査した李 (2019) によれば、たとえば「初級前半語彙」の使用率は、インタビュー発話を書き起こした KY コーパスでは習熟度上昇によって線形的に減少するのに対し、YNU 書き言葉コーパスでは一度減ってまた増える V 字カーブになることを報告している。書き言葉と話し言葉の違いや、習熟度の測定法の違い (KY コーパスは総合的なパフォーマンス評価、YNU コーパスは作文中での事前に設定したタスクの遂行度による評価) があるにせよ、語彙レベルといった基本的観点であってもコーパスにより、違う結果が得られたことになる。さらに、迫田 (2019) は、本研究でも使用した絵描写作文課題と、同一の絵をプロンプトに用いた発話課題を比較して、プランニングや産出モードの違いにより、受け身・テシマウ形・有対自他動詞 (入る／入れるなど)・助詞の正確さが変化する可能性を示した。これらの報告をふまえると、本研究で得られた知見についても、データを代えて、結果の再現性を確認することが不可欠であろう。

引用文献

Kellerman, E. (1985). If at first you do succeed... In S. M. Gass & C. Madden (Eds.), *Input in second language acquisition*. (pp. 345-353). Newbury House.

石川慎一郎 (2020a) 「第 14 章: 発話における副詞の使用」 迫田久美子・石川慎一郎・李在鎬 (編) (2020) 『日本語学習者コーパス I-JAS 入門』 (pp. 167-184). くろしお出版.

石川慎一郎 (2020b) 「日本語・中国語・韓国語・英語母語話者の日本語発話における形容詞使用実態 —I-JAS に基づく調査—」 『日本語学習者コーパス「I-JAS」完成記念シンポジウム予稿集』, 27-34.

迫田久美子 (2019) 「話すタスクと書くタスクに見る日本語のバリエーション: 日本語学習者コーパス I-JAS の分析に基づいて」 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』(pp. 151-168). くろしお出版.

迫田久美子・石川慎一郎・李在鎬(編)(2020)『日本語学習者コーパス I-JAS 入門』くろしお出版.

陳曦(2008)「日本語学習者と母語話者における日本語複合動詞使用状況の比較一作文データベースを用いて一」『小出記念日本語教育研究会論文集』16, 83-96.

森田良行(2008)『動詞・形容詞・副詞の事典』東京堂出版.

李在鎬(2019)「学習者の語彙使用は習熟度を反映しているか」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』(pp. 87-106). くろしお出版.