



日本語縦断学習者コーパス「B-JAS」のストーリーライティングタスクに見る中国人日本語学習者の語彙習得過程：語彙量・語彙多様性・語彙難度・品詞・特徴語の計量的分析

陳, 玲

(Citation)

統計数理研究所共同研究リポート, 456:1-20

(Issue Date)

2022-03-03

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81013065>

(URL)

<https://hdl.handle.net/20.500.14094/81013065>



日本語縦断学習者コーパス「B-JAS」のストーリーライティングタスクに

見る中国人日本語学習者の語彙習得過程

—語彙量・語彙多様性・語彙難度・品詞・特徴語の計量的分析—

陳玲(神戸大学大学院特別研究生)

1047997379@qq.com

Acquisition Process of L2 Japanese Vocabulary for Chinese Learners:
An Analysis of the Story Writing Task of B-JAS Corpus with Focus on the
Vocabulary Quantity, Diversity, Difficulty, POS, and Keywords
CHEN, Ling (Graduate School of Kobe University, Research Student)

概要

本研究では、縦断型日本語学習者コーパスである B-JAS のストーリーライティングタスク(SW1)で使用された語彙を調査する。中国人日本語学習者 16 人の語彙使用の発達過程について、語彙の産出量、多様性、難度、品詞使用率の観点から調査を行った。その結果、(1)学習者の使用する語彙の量(異なり語数・延べ語数)および語彙多様性(R 値)は段階的に上昇して母語話者に近づいていくが、2 年生以後は天井効果が存在すること、(2)語彙難度は、4 年間で 16%増加するものの、母語話者との間にはなお差が残ること、(3)動詞が増え、形容詞・形状詞・連体詞が減ること、(4)母語話者全般・学習者全般を特徴づける特定の名詞・動詞が存在すること、などが示された。

キーワード

B-JAS, 学習者コーパス, 作文教育, 語彙習得

1. はじめに

日本語を学習する上で、語彙の体系的習得はきわめて重要である。国立国語研究所(2005)の調査によると、雑誌で使われる語彙は約 45,000 語とされる。また、代表的な国語辞典の 1 つである『新明解国語辞典』の場合、収録語数は約 75,000 語に及ぶ。金田一(1988)は、1,000 個の語彙を知っていた場合、フランス語だと日常会話の 83.5%を理解できるが、日本語だと 60%しか理解できないと述べている。このように、日本語では、意味理解に必要な語の量が他の言語より多いといえる。

こうした背景もあり、日本語学習者は、ある程度のレベルに達しても、なお、語彙の理解や運用

に課題を抱えている場合が少なくない。以下は、日本語母語話者と日本語学習者に対して同じ課題を実施してデータを集めた「多言語母語の日本語学習者横断コーパス(I-JAS)」, および北京外国語大学日本学研究センター・北京師範大学外国語言語文学学院・国立国語研究所によって編纂された「北京日本語学習者縦断コーパス(B-JAS)」に含まれるストーリーライティング(SW)タスクの産出実例である。母語話者・学習者とも、同じイラスト(鍵を忘れた男性が自分の家に入れず、梯子で 2 階の窓から入ろうとしていて、それを警官が見ているシーン)を描写したものである。(1)は母語話者の用例で、(2)は中国人学習者の用例である。

(1) ケンは、しかたなく梯子をかけて 2 階から家に入ろうとしました。すると、パトロール中の警官が、ケンのことを怪しんで呼びとめました。(JJJ20)

(2) ケンは梯子を使いました、梯子を登りました。警官は見えました。(CCB07)

(1)と(2)を比較すると、学習者の日本語語彙使用には、母語話者と大きな違いが存在することがわかる。ここでは 5 点を指摘したい。

1 点目は産出する語彙の絶対量が圧倒的に少ないことである。上記の場合、(1)は 33 語(形態素)であるのに対し、(2)はその 52%にあたる 17 語に過ぎない。2 点目は、「使いました」「登りました」「見えました」のように学習者は同じ語を何度も反復するため、語彙の多様性が著しく低くなっていることである。タイプ数(異なり語数)をトークン数(延べ語数)で割った TTR 値を試算すると、(1)が 76%であるのに対し、(2)は 59%にとどまっている。3 点目は学習者の使用する語彙の難度が低いことである。動詞に限っても、母語話者が「(梯子を)架ける」「怪しむ」といった難度の高い語を使っているのに対し、学習者は「使う」「登る」「見える」のような基本動詞しか使えていない。4 点目は学習者の使用する語彙の品詞が偏っていることである。(1)と(2)を比較すると、学習者は、母語話者が使用している形容詞(「しかたなく」)や接尾辞(「(パトロール)中」)を一度も使用していない。5 点目は、(2)における「～ました」のように、間違いではないものの、いかにも学習者的な語彙使用が見られることである。これは、たとえば、(1)における「しかたなく」や「すると」といった、母語話者的な語彙と相補的な関係にあり、母語話者と区別できる学習者特有の語彙使用の傾向を示している。

これらの点をふまえると、学習者の語彙使用を見ていく際には、使用量・多様性・難度・品詞・特徴語の各点に注目する必要があると言える。では、これらの点に関して、中国人学習者はどのような段階的な変化を見せ、どの程度母語話者に近づいていくのであろうか。こうした問題を扱うには、一定数の学習者の産出を縦断的に収集したデータが必要になるが、これまで、そうしたデータはほとんど存在せず、研究は手薄であった。そこで本研究は、B-JAS の公開前データを許諾を得て使用し、上記の 5 点について、中国人の日本語専攻大学生の 4 年間の語彙運用力の変化を計量的に調査することを目指す。

2. 先行研究

ここでは、L2 日本語学習者の語彙発達研究を概観する。まず、先行研究で使われてきたコーパスについて触れ、その後それらを用いた研究内容について述べる。計量的な研究は、コーパスを使用することとなるが、過去の研究では、横断コーパスとしては、書き言葉・話し言葉の両方を収集した前述の I-JAS、書き言葉を収集した「YNU 書き言葉コーパス」、インタビュー発話を収集した「KY コーパス」などが使用されてきた。また、縦断コーパスとしては、台湾の大学生の書き言葉と話し言葉を収集した「LARP at SCU コーパス」や、「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)」などが使用されてきた。また、現在、I-JAS のデザインを踏襲した B-JAS の開発が進められており、開発中のデータを用いた研究がすでにいくつか出ている。

以下では、コーパス別に行われた研究と得られた知見を整理しておきたい。

I-JAS を用いた研究として、Ishikawa (2017) は作文分析を行い、テキスト長と品詞別の語彙使用量を調査した。その結果、学習者の習熟度レベルの上昇に伴い、語彙産出量が増加し、母語話者より多くなる場合もあること、助詞・副詞を過少使用する一方、名詞・助動詞・接続詞・連体詞・形容詞を過剰使用することなどが明らかになった。李 (2020) は I-JAS の学習者 1,000 名分と日本語母語話者 50 名分のストーリーライティングタスク (SW1・2) の作文データを分析した。これにより、学習者の日本語習熟度レベルの上昇に伴い、異なり語数・TTR 値・漢字比率・内容語使用率とともに上昇し、文長は長くなることが確認された。島崎 (2020) は I-JAS の「ストーリーテリング 1・2」、「対話」、「ロールプレイ 1・2」、「絵描写」タスクにおける中国人学習者 141 名 (中級、中級後半、上級前半) と日本語母語話者 50 名の産出データを用いて副詞の使用量を調査した。その結果、中国人学習者の副詞の使用量は習熟度の上昇につれて増加するものの、母語話者の半分程度であること、母語話者と中国人学習者に高頻度に使用される副詞として「そう」や「ちょっと」があるなどを明らかにした。石川 (2021) は I-JAS の日本語学習者 850 人のストーリーライティング (SW1) タスクにおける産出データを調査した。その結果、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、学習者の習熟度の上昇に伴い、トークン数とタイプ数とも逆 U 字型の分布を見せること、学習者の動詞発達過程がおおよそ 4 段階に分かれることなどが明らかにされた。

KY コーパスと YNU 書き言葉コーパスを用いた研究として、李 (2019) は語彙の難度や語種と習熟度の関係を調べた。その結果、KY コーパスでは中級語彙使用率と習熟度が強く関係していること、YNU 書き言葉コーパスでは、下位レベルでは中級後半語、中位レベルでは中級語彙、上位レベルでは上級前半語彙の使用率が高くなっていることが示された。語種に関しては、習熟度が上昇しても漢語使用率にはっきりした変化は見られず、むしろ和語使用率が増加し、とくに KY コーパスにおいて上昇が顕著であることが観察された。

LARP at SCU コーパスを用いた研究として、石川 (2018a) は台湾の日本語専攻大学生の 3 年半の作文産出におけるタイプ・トークン使用率、品詞使用率、高頻度語彙使用率などを調査した。その結果、3 年半でトークン数は 27%、タイプ数は 65%、タイプ／トークン使用率は 30% 増加し、学習開始 272 カ月あたりで増加が止まること、代名詞が大幅に減り、助動詞と一部の名詞・副詞が

減少するのに対して、名詞、動詞、形容詞、接続詞、助詞の大半は増加することなどが示された。

C-JASを用いた研究として、石川(2018b)は3名の中国語母語の日本語学習者の3年間の発話データを調査し、トークン数とタイプ数はともに増加するものの、語彙多様性指標は明確に上昇しないこと、初期には機械的な応答標識が多く、中期・後期には、テンス標識(「た」)や文要素の接合に関連する機能語(「か」「の」「て」)などのように文法的に複雑な文を構成する上で必要となる語が多く使用されること、3名の学習者の語彙習得段階がおおよそ3つの時期に区分されることを示した。

B-JASを用いた研究として、費(2019)はトピック作文タスクにおける5回分の作文データを使い、漢語の使用率の変化を調査した。その結果、初級段階では漢語使用率が顕著に高いが、中級と上級になると段階的に減少した後に安定すること、中級と上級の間には有意差がないこと、学習者の習熟度の上昇に伴い、日中同形同義語・異形語・同形異義語の使用率が順に減少することなどが示された。董(2020)はB-JASのトピック作文タスクにおける3回分の産出データを用いて並列・継起の接続詞の使用実態を調査した。その結果、1年次後半から2年次前半にかけて、並列・継起の接続詞の使用率が低下し、接続助詞の使用率は上昇すること、そして並列と継起の形態上の区別が可能になることが確認された。また、2年次前半から2年次後半にかけて、段落の冒頭で接続詞を使えるようになり、使用位置に合わせて接続表現を選択できるようになることが明らかにされた。

以上で概観したように、先行研究は、それぞれ多くの価値ある知見を示し、とくに中国人日本語学習者の語彙発達について多くの事実を明らかにしたが、いくつかの点には課題も見られる。とくに、(1)語彙を見る視点・観点の制約、(2)タスクの未統制、(3)具体的な学習レベル別の語彙特定が少ないといった点は問題となりうる。そこで本研究は、(1')使用語彙の量・多様性・難度・品詞・特徴語の5点を同時に見る多元的アプローチを採用した上で、(2')統制度の高いB-JASの絵描写タスクを分析対象とし、(3')学習者の語彙使用の段階的变化の概観に加え、初級者・中級者・上級者、また学習者と母語話者の差を決定する特徴語を抽出することを目指す。

3. リサーチデザインと手法

3.1 研究目的とRQ

本研究の目的は、内容が固定されたストーリーライティング(SW1)タスクにおける、中国人大学生の産出データを分析し、4年間の学習過程に見られる語彙使用の変化のパターンを解明し、各段階の学習者を区分する特徴語を発見することにある。この目的のもと、以下の研究設問を設定する。

RQ1 語彙量(延べ語数・異なり語数)および語彙多様性は増加するか？

RQ2 語彙難度は上昇するか？

RQ3 どの品詞において使用量が増加するか？

RQ4 学習者・母語話者を特徴づけるマーカー語にはどのようなものがあるか？

4 つの RQ に対する事前仮説は以下のとおりである。

- H1 内容が固定された絵描写課題であるから、延べ語数・異なり語数・語彙多様性は増加しないのではないかな？
- H2 使用語彙の難度レベルは学習時間の増加に連れて上昇するのではないかな？
- H3 全ての品詞は一貫して増加・減少するのではないかな？また、石川(2018a)によると、名詞、動詞、形容詞、接続詞、助詞は増加するのではないかな？
- H4 石川(2021)によると、母語話者マーカーには複合動詞、慣用表現中の語などが含まれ、学習者マーカーには意思性に関わる使い分けを誤った語、文脈と合致しない語などが含まれるのではないかな？

以下、具体的なデータ調査を通して RQ ごとに仮説の妥当性を検証していく。

3.2 データ

すでに述べたように、B-JAS は北京外国語大学日本学研究センター・北京師範大学外国語言語文学学院・国立国語研究所によって編纂中の大型の日本語学習者縦断コーパスである。北京師範大学の 17 名の日本語専攻学習者の産出データが経時的に収集されている。特筆すべきは、すでに公開されている I-JAS のタスクデザインを踏襲することで、横断コーパスである I-JAS を補完する縦断コーパスとして使用できるように設計されていることである。B-JAS で収集したデータの概要は以下の表 1 の通りである。

表 1 B-JAS のタスクデザイン

テスト項目		所要時間	調査サイクル	備考
対面調査	ストーリーテリング(ST)	10 分	年 22 回(4 月, 10 月)計 8 回 ※初回は 2016 年 1 月に実施	・4 月分は B-JAS 版イラスト(2 組)を使用・計 4 回 ・10 月分は I-JAS 版イラスト(2 組)と同じものを使用・計 4 回
	ストーリーライティング(SW)	20 分(5 文以上書く条件を付ける)	上記と同様	・ST タスクと同一課題を書く
	対話(I)	30 分	上記と同様	・4 月分は B-JAS 版の話題を採用・計 4 回 ・10 月分は I-JAS 版と同じ話題を採用・計 4 回
	ロールプレイ(RP)	10 分	上記と同様	・4 月分は B-JAS 版の状況を採用・計 4 回 ・10 月分は I-JAS 版同じ状況を採用・計 4 回

	絵を見て話す タスク(D)	5 分	上記と同様	・I-JAS と同じイラストを使用
PC テスト	SPOT	30 分	年 1 回(10 月) 計 4 回 ※初回 は 2016 年 1 月 に実施	
	J-CAT (Can-do)	60 分	年 1 回(10 月) 計 4 回 ※初回 は 2016 年 1 月 に実施	
オンライン 調査	トピック作文	指定なし	年 3 回(1 月, 5 月, 10 月)	

上表に示すように、B-JAS では 4 種のストーリーライティングタスクを実施しているが、本研究では、ピクニックを話題とした計 4 回の作文データを使用する。プロンプトは以下のイラストである。

図 1 B-JAS の ST1/SW1 のプロンプト



なお、17 名の参加者の初期(入学から 4 カ月後)における J-CAT 得点を確認したところ、参加者 CCB02 は学習開始の段階ですでに中級レベルに達していることが確認された。本研究は初学者の語彙習得過程を探るという目的を持つため、この学習者のデータは分析から削除し、残り 16 名のデータのみを扱う。また、母語話者データは、I-JAS に参加した 50 名の日本語母語話者データを使用する。

3.3 手法

3.3.1 データの事前処理

はじめに、I-JAS より日本語母語話者 50 名のデータを集め、1 ファイルにまとめた。次に、B-JAS に含まれる 1 年生から 4 年生まで学習者各 16 名のデータを学年別の 4 ファイルにまとめた。

以後の分析では、これらの 5 ファイルを検索対象とする。なお、目視で形態素解析の妥当性を検証し、問題が見つかった個所(例:I-JAS-JJJ07:「驚き困る」→正しくは「驚き」+「困る」、B-JAS-CCB14:「飛ぶ」「込む」→「飛び込む」、B-JAS-CCB02:「犬」(接尾辞)→「犬」(名詞)など)については修正を行っている。また、ストーリーライティングの冒頭に与えられた「朝、ケンとマリはサンドイッチを作りました。」という書き出し指示文については一律に削除している。

3.3.2 分析の手順

本研究では、学習段階を学年とみなし、1 年時、2 年時、3 年時、4 年時のデータを比較することで、4 年間の語彙使用の変遷を調査する。

まず、RQ1 については、延べ語数と異なり語数を調べる。また、語彙多様性指標として、Guiraud's R 値(異なり語数÷[延べ語数の平方根])を計算する。一般に、語彙多様性の目安として、異なり語数を延べ語数で割った TTR (type / token ratio) が使用されることが多いが、総語数が増えると TTR 値は自然に減少する。今回は、異なる総語数のデータ間で語彙多様性を比較するため、補正指標の 1 つである R 値を使用することとしたい。なお、語数の計算にあたり、形態素解析で「補助記号」や「解析困難箇所」と判定されたものは除いている。

RQ2 については、学年別および母語話者の全 5 ファイルに含まれるすべての語について、「J-readability」(李, 2017)を用いて難度レベルを付与する。このレベルは、日本語教育語彙表のデータに基づくものである。日本語教育語彙表は、「現代日本語書き言葉均衡コーパス(BCCWJ)」や日本語教科書コーパスの頻度調査によって選定された 17, 929 語の見出し語からなっている。各見出し語には、経験豊富な日本語教師の主観判定に基づく 6 段階の難易度の他、品詞・語種・表記・アクセントなどの情報や、語義と用例が搭載されている(砂川, 2012)。今回は、初級前半から上級後半までの 6 段階の語彙レベルを 1~6 のスコアに変換し、ファイルごとに平均値を算出する。なお、日本語教育語彙表に含まれていない語は分析から除外する。計算の実例として、仮に、あるテキストに含まれる異なり語数が 3 種のみで、それらの難度レベルが 3, 3, 6, 頻度が 10, 20, 30(合計で延べ 60 語)であったとすると、当該テキスト内の語彙難度の平均値は $(3 \times 10 + 3 \times 20 + 6 \times 30) \div 60 = 4.5$ となる。

RQ3 について、品詞の分類には精粗の別があるが、今回は、UniDic 辞書における「大分類」を品詞の認定基準とする。5 つのファイルごとに、12 種の品詞(大分類)別に総頻度を求め、ファイルの総語数で割って百分率に変換する。次に、12 種の品詞ごとに、4 つの学年ごとの平均頻度を算出し、増減の傾向を線型回帰分析する。一般に、相関係数は $r = 0.7$ 以上であれば「強い相関」と判断される。このとき相関係数の二乗で求められる説明力は 49%となる。そこで、得られた回帰式の説明力が 49%を上回り、また、回帰係数が 10 を超える場合に注目すべき線型的な増加または減少が起こっていると判断する。また、対数尤度比検定を使って 4 年生の学習者と母語話者の品詞使用率を比較し、差の有意性を確認する。

RQ4 については、学習者と母語話者の使用語彙を同等の使用場面で比較するため、筆者の判断に基づき、プロンプトに含まれる 5 枚のイラストのいずれかと関連付け、すべての語彙に「絵 1」,

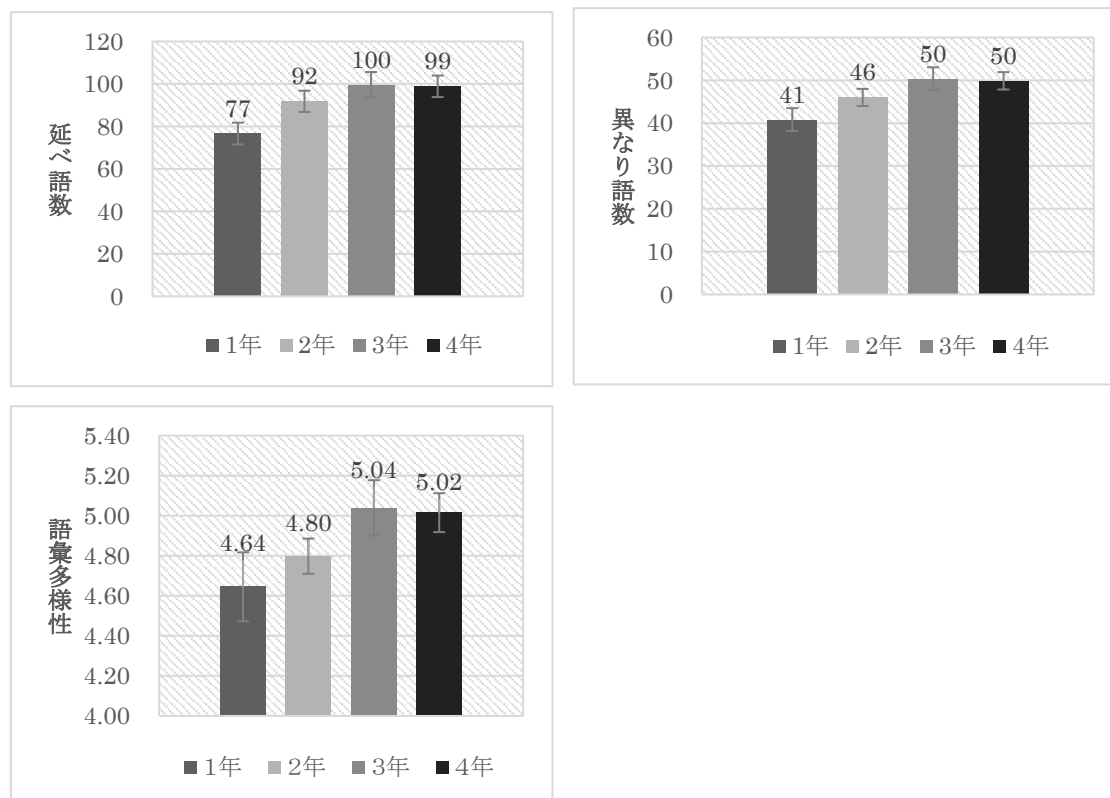
「絵 3」などのコーディングを行う。たとえば、「今日は二人でピクニックに行く予定です。ケンとマリは昼に食べるサンドイッチをバスケットに詰め、今日の行き先を確認するために地図を広げていると、二人が飼っている犬のリサちゃんがバスケットの中に入り込んでしまいました。」という例文について言うと、前半部(今日は二人でピクニックに行く予定です。ケンとマリは昼に食べるサンドイッチをバスケットに詰め)はプロンプトの 1 枚目に該当する内容なので、この箇所のすべての語に「絵 1」のコードを付与する。また、後半部(今日の行き先を確認するために地図を広げていると、二人が飼っている犬のリサちゃんがバスケットの中に入り込んでしまいました)はプロンプトの 2 枚目に該当する内容なので、この箇所のすべての語に「絵 2」のコードを付与する。次に、コードごとに母語話者と学習者の語彙を比較し、母語話者のみが使用する語彙を母語話者マーカー語、学習者のみが使用する語彙を学習者マーカー語として特定する。なお、ここでは主要品詞である名詞と動詞に調査対象を限定する。

4. 結果と考察

4.1 RQ1 語彙量・語彙多様性の変化

語彙量(延べ語数・異なり語数)および語彙多様性について、4 年間の変化を調査したところ、以下の結果を得た。

図 2-4 語彙量(延べ語数・異なり語数)および語彙多様性(R 値)の変化



注:エラーバーは標準誤差を示す。

グラフからは、何れの観点についても、およそ、学年進行に応じて上昇する傾向が認められた。そこで、まず、一元配置の分散分析で学習段階(学年)の主効果を確認したところ、延べ語数($F(3, 45) = 7.785, p < .001, \eta_p^2 = .342$)、異なり語数($F(3, 45) = 9.033, p < .001, \eta_p^2 = .376$)、語彙多様性($F(3, 45) = 5.224, p = .005, \eta_p^2 = .258$)ともに有意であった。学年が変われば、延べ語数・異なり語数・語彙多様性ともに何らかの変化が生じることが証明されたとと言える。

次に、事後検定(Holm)によって隣接学年間の差の有意性を検証したところ、以下の結果を得た。なお、上段は平均値、下段は効果量 d の値と小中大の目安を示す。

表 2 語彙量(延べ語数・異なり語数)および語彙多様性(R 値)の隣接学年間変化

観点	隣接学年間変化
延べ語数	1 年(77)<2 年(92)≒3 年(100)≒4 年(99) 1 年--(.73:大)--2 年--(.35:小)--3 年--(.03:無)--4 年
異なり語数	1 年(41)<2 年(46)≒3 年(50)≒4 年(50) 1 年--(.54:中)--2 年--(.45:中)--3 年--(.05:無)--4 年
語彙多様性	1 年(4.64)≒2 年(4.80)<3 年(5.04)≒4 年(5.02) 1 年--(.27:小)--2 年--(.51:中)--3 年--(.05:無)--4 年

以上より、当初の仮説(H1)とは異なり、統制的な絵描写課題であっても、語彙使用量・語彙多様性が 2～3 割増加することが証明された。この間の学習者の語彙力の増大は大きく、同じ内容を描写する場合であっても、より幅広い語彙で、より多くの文を産出できていると言える。

なお、縦断的な作文データを分析した石川(2018a)の結果と比較すると、今回のデータにおける学年間の増加率は、延べ語数ではほぼ同様(27%/29%)、異なり語数では小さく(65%/38%)、多様性でも小さく(30%/22%)になった。異なり語数と多様性において、今回のデータにおいて伸び幅が小さかったのは、統制的課題を使用したためであると考えられる。

ここで、4 年間の語彙量および語彙多様性の隣接学年間ならびに 4 年間の増減幅を調べたところ、以下の結果を得た。

表 3 語彙量(延べ語数・異なり語数)および語彙多様性(R 値)の学年間増減幅

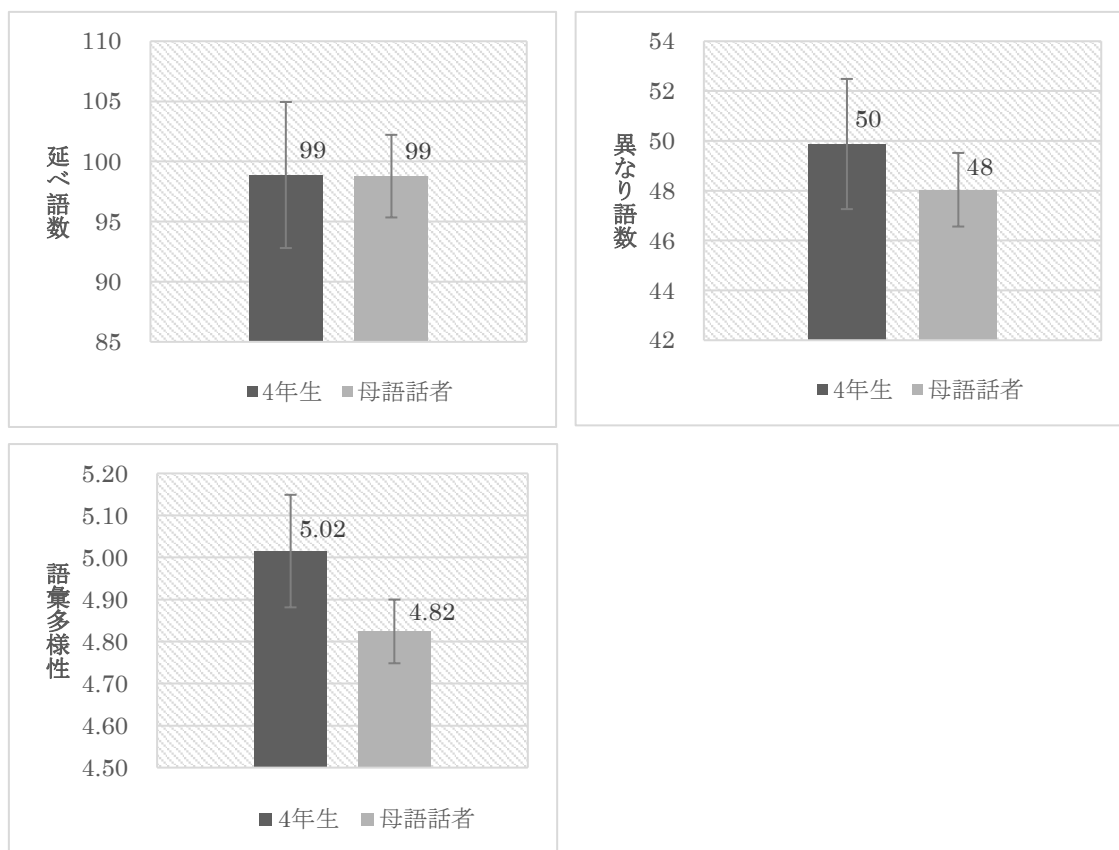
観点	隣接学年間の変化			4 年間の変化
	1-2 年	2-3 年	3-4 年	1-4 年
延べ語数	20%	9%	-1%	29%
異なり語数	20%	13%	2%	38%
語彙多様性	10%	9%	2%	22%

まず、4 年間の変化に着目すると、延べ語数・異なり語数・語彙多様性とも 2～3 割程度増加す

ることが示された。ただし、隣接学年間の増加幅は、学年があがるにつれて縮小しており、学習者の語彙力の伸長には「天井効果」(ceiling effect)が存在することが確認された。

では、4年間の学習を経ると、中国人学習者の語彙使用は母語話者に並ぶと言えるのであろうか。この点を確認するため、4年生と母語話者の語彙量および語彙多様性を比較する。

図 5-7 4 年生と母語話者の語彙量(延べ語数・異なり語数)および語彙多様性(R 値)



注: エラーバーは標準誤差を示す。

対応なし t 検定を行ったところ、4年生と母語話者間では、異なり語数($t(34, 335) = 0.611$, $p = .479$, $d = .174$), 延べ語数($t(31.469) = 0.014$, $p = .988$, $d = .004$), 語彙多様性($t(37.618) = 1.243$, $p = .139$, $d = .353$)とも、差は有意でなかった。つまり、4年生の段階では、学習者の語彙量と語彙多様性は母語話者と同じレベルに到達していることが証明されたと言える。

続いて、4年間の変化を具体的に確認するため、同一学生(CCB18)の1年次の用例(3)・4年次の用例(4)・母語話者用例(5)を質的に検討したい。なお、語彙多様性を示すため、1回しか使われていない語にはアンダーラインを引いている。

- (3) ケンとマリはパンを作りました。その時、犬がありました、パンを見えました。作ることが終わったら、マリとケンと一緒に地図を見ました。注意を持ってなかつた、犬はバスケットに入り

ました。その後、公園で一緒に散歩をしました。犬は出ました。食べ物はすべて食べられなかったから、ケンとマリはとても悲しかったです。(CCB18:延べ語数:41 語 異なり語数:82 語 語彙多様性:4.53)

- (4) その時、サンドイッチの匂いに沿って、犬がやってきました。サンドイッチをバスケットに入れておいてから、ケンとマリはピクニックの場所を確認するために、地図を開いて見ていました。しかし犬がバスケットに跳んできて、二人ともそれに気づきませんでした。ケンとマリはバスケットを持ってピクニックをしに行きました。その時、犬が急にバスケットから飛び出しました。中に入っている食べ物がほとんど犬に噛まれて、もう食べなくなりました。ケンとマリはとても失望していました。(CCB18:延べ語数:57 語 異なり語数:119 語 語彙多様性:5.23)

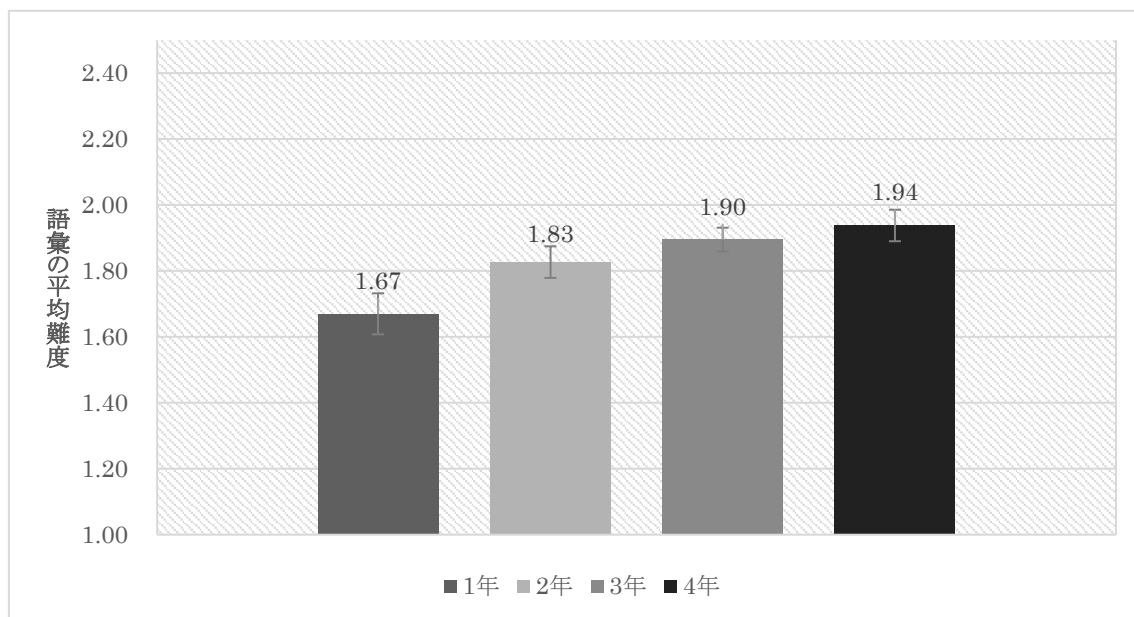
- (5) 二人が作っているところを飼っている犬がじっと見ていました。その後、ケン達はどこにピクニックに行こうかと地図を見て考えていました。その隙に犬がテーブルの上のサンドイッチが入っているバスケットの中にすりを入れてしまいました。二人はそれに気付かずバスケットを持ってピクニックに出かけました。そして、お昼になりサンドイッチを食べようとバスケットを開けると、なんと飼っている犬が飛び出してきました。二人はびっくりして中を見たところ、サンドイッチが全部なくなっていました！(JJJ16:延べ語数:54 語 異なり語数:122 語 語彙多様性:4.89)

用例でも明らかなように、1年次と比較して、3年間の学習を重ねた4年次の学習者の産出量は飛躍的な上昇を見せている。延べ語数は16語(41→57)、異なり語数は37語(87→119)、語彙多様性は0.7(4.53→5.23)増加しており、母語話者とはほぼ同じレベルに到達している。

4.2 RQ2 語彙難度の変化

語彙難度について、4年間の変化を調査したところ、以下の結果を得た。

図 8 語彙の平均難度の変化



注: エラーバーは標準誤差を示す。

グラフからは、語彙の平均難度は学年進行に応じて上昇する傾向が認められた。一元配置の分散分析で学習段階(学年)の主効果を確認したところ、主効果は有意で($F(3, 45) = 9.430$, $p < .001$, $\eta_p^2 = .386$), 事後検定(Holm)により、以下の結果を得た。表内の上段は平均値、下段は効果量 d と効果量の小中大の目安を示す。

表 4 語彙の平均難度の隣接学年間変化

観点	隣接学年間変化
語彙難度	1 年(1.67) < 2 年(1.83) $\div$ 3 年(1.90) $\div$ 4 年(1.94) 1 年--(.69: 大)--2 年--(.39: 中)--3 年--(.25: 小)--4 年

また、4 年間の語彙量および語彙多様性の隣接学年間ならびに 4 年間の増減幅を調べたところ、以下の結果を得た。

表 5 語彙難度の学年間増減幅

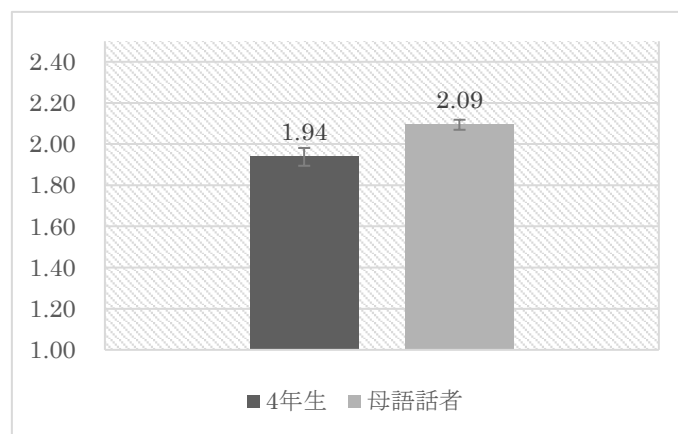
観点	隣接学年間の変化			4 年間の変化
	1-2 年	2-3 年	3-4 年	1-4 年
平均難度増加率	11%	2%	3%	16%

仮説(H2)に関して、4 年間で語彙難度は 16%増加することがわかった。学習者の語彙力は増加し、同じ内容を描写する時にもより難しい語彙で表現できるようになったと考えられる。伸び幅は、

1～2 年(11%)>2～3 年(2%)>3～4 年(3%)となり、学年があがるにつれて縮小する傾向が確認された。語彙産出量や語彙多様性の場合と同じく、学習者の語彙力の発達には「天井効果」があると思われる。

次に、4 年生と母語話者の語彙難度を比較する。

図 9 4 年生と母語話者の語彙難度



注:エラーバーは標準誤差を示す。

語彙難度は母語話者のほうが高くなっており、対応なし t 検定でも差は有意であった($t(22, 808) = 3.146, p = .008, d = .893$)。つまり、1 年生に比べると 4 年生が使える語彙難度は若干上昇しているものの、母語話者にははるかに及ばないことがわかった。

続いて、4 年間の変化を具体的に確認するため、同一学生(CCB14)の 1 年次の用例(6)・4 年次の用例(7)・母語話者用例(8)を質的に検討したい。なお、語彙難度の違いを示すため、中級後半以上の語にはアンダーラインを引いている。

- (6) あの時、ケンの犬は部屋にいますが、二人も注意しませんでした。二人は地図を見ています時、この犬はバスケットに行きました。そして、ケンとマリと一緒に公園へ行きました。ランチを食べたいですから、バスケットに犬が発見したです。犬はバスケットに出したが、食べ物は全然食べられませんでした。ケンとマリはとても悲しいです。(CCB14:語彙の平均難度:1.49)
- (7) 今日はいい天気ですので、二人はピクニックに行こうと思っていました。ケンとマリは地図を見ているうちに、ワンちゃんが勝手にバスケットに飛び込んでしまいましたが、二人は何も気づいていなかったです。うれしい気持ちで、二人は出発しました。しかし、バスケットのふたを開けたとき、ワンちゃんが飛び出しました。びっくりした二人は慌ててバスケットの中に見てみると、食べ物は全てめちゃくちゃになりました。ケンとマリは落ち込みました。(CCB14:語彙の平均難度:2.12)
- (8) 今日は二人でピクニックに出かけるのです。その様子を、お留守番のポチがうらやましそう

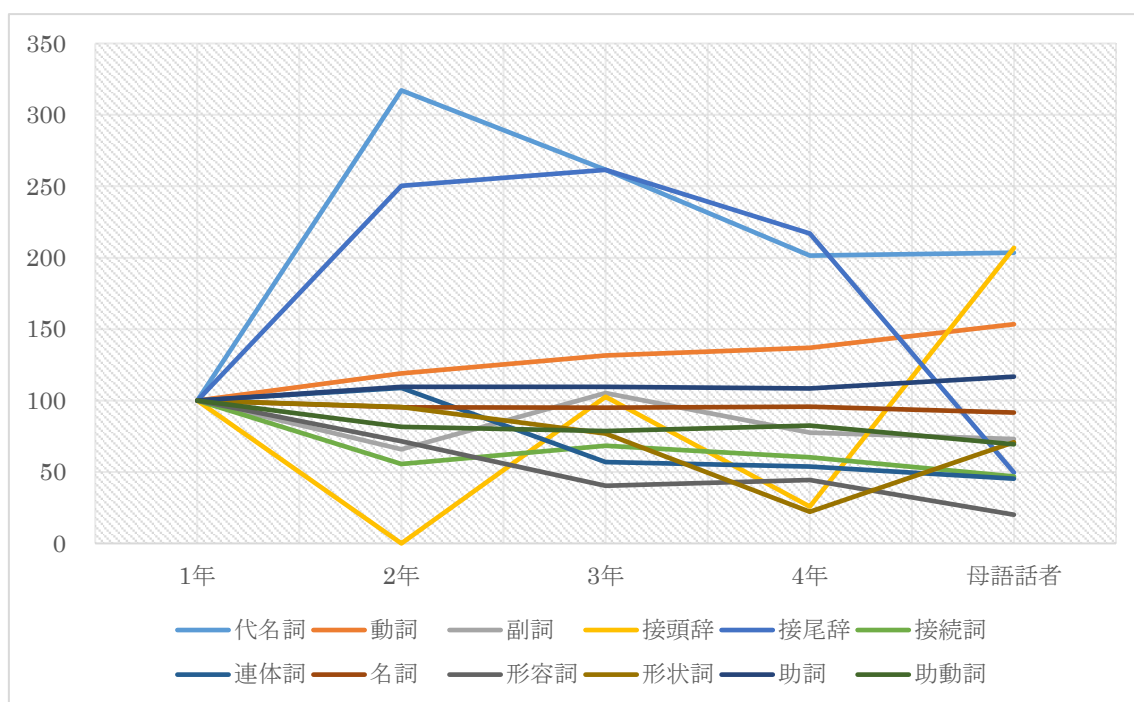
に見ていました。二人が今日の地図を眺めている隙に、ポチはバスケットに忍び込みます。そうとは知らず、ケンとマリは草原までやってきました。絶好のピクニック日和です。見晴らしのよいところでお昼を食べようとしたそのときでした。バスケットの中から、ポチがとびだしてきたのです。驚くマリの横で、がっかりするケン。二人のお昼ご飯は、すっかりポチに食べられていたのです。(JJJ08:語彙の平均難度:2.33)

中級後半以上の語の使用状況に着目すると、1年生では「ランチ」のみであるが、4年生になると、「勝手」「飛び込む」「仕舞う」「飛び出す」「慌てる」「目茶苦茶」「落ち込む」の7語が使われている。さらに母語話者になると、「様子」「うらやましい」「眺める」「隙」「忍び込む」「草原」「絶好」「日和」「見晴らし」「飛び出す」「驚く」「がっかり」の12語が使われている。上級の語の使用という点で見ると、4年生であっても母語話者のレベルには到達していないと思われる。

4.3 RQ3 品詞使用率の変化

品詞別の使用量の変化を調査したところ、図10の結果を得た。なお、下図においては、縦軸は、1年生の使用量を100%とした場合の使用率を示す。

図10 4年間の品詞使用率の変化



この中には、一貫して増加するもの(動詞)や一貫して減少するもの(形容詞など)だけでなく、接頭辞のように、増減を繰り返し、はっきりした増減のパターンが見られないものもある。そこで、12種の品詞ごとに、学年を説明変数、使用率を目的変数とする単回帰分析を行い、得られた回帰式につ

いて、回帰係数(増減の傾きの大きさ)と説明力(全体の分散のうち回帰式で説明できる分散の比率)を調査した。ここでは、回帰係数が 10 以上で、かつ、説明力が 49%より大きい場合に、4 年間に、一定の増加または減少が生じていると判断する。該当するのは 12 品詞中、4 品詞のみであった。

表 6 学習歴と品詞の使用率の関係($r^2 > 49\%$ の項目)

品詞	回帰係数	増減パターン	説明力 r^2
動詞	12.35	増加	94%
形容詞	-19.84	減少	85%
形状詞	-25.20	減少	83%
連体詞	-19.03	減少	74%

学習時間に伴って増加しているのは動詞のみであった。動詞は「日本語の表現機能を支える担い手」であり(森田, 2008), 使用率の上昇は学習者の表現力の増加を示していると考えられる。一方、形容詞や形状詞の使用率の減少は、初級学習者が人物の内面の感情を主観的に描写するのに対し、中上級になると、人物の外見や関係性を客観的に描写するようになっていることを示唆する。連体詞の使用率の変化は、学習の最初の段階から「コソアド」の使い方を重点的に教えられるため、初級学習者が「コソアド」を過剰使用するが、中上級になると、「コソアド」を適宜に使えるようになっていくことを示唆する。

次に、4 年生と母語話者の品詞使用率を比較する。

表 7 4 年生と母語話者の品詞使用率の比較($p < 5\%$ の項目)

品詞	4 年生使用率	母語話者使用率	対数尤度比	p 値	効果量 ϕ
助動詞	17.76%	14.94%	20.02	0.000	0.06
名詞	27.50%	26.30%	16.18	0.000	0.05
接尾辞	0.88%	0.20%	12.67	0.000	0.05
接頭辞	0.06%	0.51%	8.27	0.004	0.03
感動詞	0.00%	0.18%	5.01	0.025	0.02
形容詞	0.76%	0.34%	4.12	0.042	0.03

上記より、4 年生になっても、学習者は助動詞・名詞・接尾辞・形容詞を過剰使用し、接頭辞・感動詞を過少使用する傾向が確認された。これらの違いは、4 年生であっても、品詞使用の点で、学習者と母語話者の差が完全には埋められていないことを示唆する。

4.4 RQ4 マーカー語の分析

学習者全体と母語話者を特徴づけるマーカー語を調査したところ、以下の結果を得た。

表 8 学習者マーカ―と母語話者マーカ―

イラスト	学習者マーカ―	母語話者マーカ―
1 コマ目 	【名詞】スライス／ハイキング／料理／一緒／注意／間／今／内／全部／時／カップル／フード／ワン／愛犬／表情／側／茶／果物／積み／隣り／名前／目／妻／食べ物／食物／水／同時／外／匂い／周り／作り／回り 【動詞】来る／遣る／有る／終わる／待つ／見える／買う／入る／思う／洗う／沿う／遊ぶ	【名詞】ポチ／留守／予定／支度／日／今日／遠く／中／昼／途中／ランチ／後ろ／飼い犬／隅／仲良し／子犬／番／具 【動詞】出来る／持つ／出来上がる／見詰める／詰める／詰め込む／飼う／眺める／挟む
2 コマ目 	【名詞】チェック／参考／発見／経由／旅行／完了／研究／原因／注意／日／明日／全部／途中／勝手／真剣／スケジュール／ルート／ワン／弁当／側／道順／夫婦／果物／机／具体／隣り／路線／食べ物／匂い／周り 【動詞】出す／出来る／付ける／有る／置く／終わる／つく／持ち出す／持つ／読む／飛ぶ／見える／開く／取り出す／入れ込む／上る／食べる／思う／選ぶ／言う／隠れる／遊ぶ／知る	【名詞】ジョン／ポチ／リサ／出発／計画／決定／位置／今日／前／最中／子／ランチ／段落／飼い犬／万端／隙／子犬／度 【動詞】遣る／始める／広げる／話し合う／話す／就く／決まる／決める／考える／立てる／忍び込む／眺める／向かう／漁る
3 コマ目 	【名詞】出発／一緒／予定／注意／日／頃／時／途中／午後／翌日／カップル／スポーツ／道／夫婦／家／郊外／景色／外／走り／通り／一／十 【動詞】見る／有る／辿り着く／出る／分かる／話し合う／話す／決める／連れる／喜ぶ／選ぶ／隠れる／遊ぶ／重ねる／走る	【名詞】先／中／草原／風景／感じ／高原／見晴らし／絶好／麓／気／気分／日和／天辺／行き先／野原／原っぱ 【動詞】出す／付く／遣る／仕舞う／抱える／出掛ける／紛れ込む／入る

4 コマ目 	【名詞】発見／料理／休憩／用意／前／全て／瞬間／途中／急／子／ランチ／ワン／夫婦／果物／積み／口／生塵／食べ物／外／晩御飯／御腹／一 【動詞】出す／始める／仕舞う／行く／残る／飛ぶ／見付ける／考える／空く／思う／逃げる／逃げ出す／喜ぶ／現われる／止まる／走る／選ぶ	【名詞】ジョン／リサ／ポチ／満悦／食事／物／仰天／朝／上／為／ランチタイム／愛犬／弁当／林檎／気／丘／事／飼い犬／顔／野原／中身／子犬 【動詞】歩き疲れる／飛び出る／広げる／入る／飼う／楽しむ／作る
5 コマ目 	【名詞】チェック／びっくり／発見／話／料理／失望／物／半分／結局／今／複雑／急／目茶苦茶／パン／ペット／ワン／本当／部分／菓子／裏／気持ち／食べ掛け／食物／仕様／屑／終わり／状況／最後 【動詞】切る／終わる／残す／残る／持つ／壊れる／急ぐ／近づく／開く／落ち込む／噛む／食べ残す／食べ掛ける／言う／知る	【名詞】ポチ／丹精／確認／用意／さつき／朝／間／所／一生懸命／昼／ショック／デザート／テンション／ランチ／愛犬／弁当／表情／縋り／底／横／台無し／腕／様子／中身／昼食／状態／子犬 【動詞】参る／出す／掛ける／来る／置く／荒らす／驚く／齧る／気付く／取る／入れる／散らかす／受ける／無くなる／下がる／込める／漁る／作る

I-JAS のデータを対象として同様の分析を行った石川(2021)では、母語話者を特徴付けるのは(1)複合動詞、(2)慣用表現中での使用、(3)評価ニュアンスを含む動詞の 3 種の語彙であり、学習者を特徴づけるのは(4)意思性や意図性に関わる使い分けを誤った動詞、(5)文脈と合致しない語、(6)複合動詞構成要素、(7)絵に描かれていない行為を表す語の 4 種の語彙であるという結果が報告されている。石川(2021)が指摘した点の大部分は、B-JAS のデータを対象とする今回の分析でも同じように確認されたが、一部で異なる傾向も認められた。

以下、まず、石川(2021)の結果と重なる点について確認していく。母語話者の特徴の 1 点目は、複合動詞の使用である。「出来上がったサンドイッチ」「(サンドイッチをバスケットに)詰め込む」「(犬がバスケットに)紛れ込む／忍び込む」「(犬がバスケットから)飛び出る」などがその例である。学習者は複合動詞の代わりに「作ったサンドイッチ」「(サンドイッチをバスケットに)入れる」「(犬がバスケットに)入る」「(犬がバスケットから)出る」のような簡単な動詞を使用する。このことは、学習者にとって、複合動詞の運用が困難であることを示す。

2 点目は、慣用的な表現の使用である。「(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(たサンドイッチ)」などがその例である。母語話者が慣用的に使用しているが教科書には掲載されていない語義、また、決まったコロケーションの中でしか使用されない語は、学習者が接触する機会が少なく、結果的に、習得が困難な項目になっていると考えられる。

3 点目は、犬に対する否定的感情を含意する語の使用である。「(犬がバスケットを)漁る」や「(食べ物が犬に)荒ら(された)」などがこの例である。母語話者は、被害者の立場から犬の行為を

貶めるため、ネガティブな感情を含む動詞を自在に使用しているが、学習者にとって、スタンスの表出は困難であり、こうした表現は十分に習得されていない。

次に、学習者の特徴の 1 点目は、意思性や意図性に関わる使い分けを誤った動詞の使用である。「(サンドイッチを作り)終わる」「(りんごをバスケットに)入る」などがその例である。一般に、母語話者が意図を持つ行為の完了を表現するために他動詞の「(サンドイッチを作り)終える」、そして意図的にものを狭いところに置くときに「(りんごをバスケットに)入れる」を使用するが、学習者は意図性の有無に準じて動詞を選択する能力を身につけていないと言えるだろう。

2 点目は、文脈に完全には合致しない、とくに意味合いがずれた動詞や漢語を多用することである。たとえば、地図を見る・眺めると言うべき個所で、「読む」「研究」という語を使ったり、犬が食べ物の匂いに誘われてやってきたと言いたい場合に、「(匂いを)経由」してやってきたと書いたりしている。

3 点目は、複合動詞構成要素の使用である。学習者が「(犬がバスケットに)跳んできて…」 「(犬がバスケットから)飛んできた」のように「飛ぶ」と「来る」をつなげて使用し、「飛び込む」や「飛び出す」などの複合動詞の使用を回避している。

4 点目は、文法的な誤りではないものの、絵に描かれていない行為やものを表す語の使用である。たとえば、1 コマ目で言えば、水筒の絵しか描かれていないのに「水」や「茶」などの内容物を指す語を使ったり、単に犬が描かれているだけであるのに、犬が「匂い」に「沿って」やってきたり、犬が「回り」で「遊んで」いたとしたりするなど、内容を補填している例である。また、「リンゴ」と言うべき個所で総称的な「果物」という語を使ったりする例もみられる。

続いて、石川(2021)では触れられておらず、今回の分析で新たに見出された点について述べる。まず、母語話者の特徴としては、すでに述べた 3 点に加え、新たに以下の 3 つの特徴が検出された。

1 点目は、慣用的 1 語名詞の使用である。たとえば、「具」「隅」「隙」「度」などがこれに当たる。二文字の漢語と比べ、1 語の漢語の数が少ないため、学習者はこのような語に接触する機会が少なく、体系的に学習することもないと推測される。1 語の漢語は漢語学習の盲点と言えるだろう。

2 点目は、難度の高い語の使用である。「支度」「丹精」などがその例である。「支度」の難度レベルは中級前半であり、中上級学習者であれば当該語を知っている可能性が高いと思われるが、実際には、中国人学習者はより簡単な日中同形同義語の「準備」を選択していることが分かった。また、しばしば決まった共起語になって使用される「丹精(を込める)」のような高度な語も学習者にとって習得しにくい項目であると言えるだろう。

3 点目は、日中同形異義語の使用である。「(二人は一)段落し…」 「(二人はびっくり)仰天」などがその例である。中国語では「段落」は一般に「文章中の意味の上での大きな切れ目」という意味で使用され、動詞的な用法がないため、日本語での語義と用法を勉強したとしても、中国語と似る語義と用法しか把握していない可能性も考えられる。また、中国語では「仰天」が「天を仰ぐ」の意味であるが、日本語では「非常に驚くこと」の意味が含まれる。そのような日中同形異義語に対する理解と使用が学習者にとって必要な能力であると思われる。

一方、学習者の特徴としては、石川(2021)の4点に加え、以下の2点が新たに検出された。1点目は、説明上、必ずしも言及の必要性がない時間表現を多用することである。「(あの／その)時」「同時」「(バスケットをあけた)瞬間」「(歩いている)途中」「最後」のような時間表現の使用は必ずしも問題ではないが、5コマ漫画のように、時系列的な流れが明白な文脈では、時間をその都度強調することは不自然に感じられる場合もある。この点については、すでに述べたように、時間表現に限らず、総じて中国人学習者は絵に描いていないことを余分に描写する傾向が認められる。これが中国語の影響によるものであるのか、文化的な背景によるものであるのか、あるいは指示文の受け取り方の違いによるものであるのかについて、更に詳しく見ていく必要があるだろう。

2点目は、日中同形語を多用することである。たとえば、犬がバスケットに入ったことに気付かない状況を描写する際に、学習者は「気付く」ではなく、「注意」を使用する。また、ピクニック先での二人の休息を描写する際には、「休む」ではなく「休憩」を選択する。これらの背景としては、「注意」や「休憩」が日中同形語であることが考えられるだろう。なお、日中同形語の知識は、時に、負の転移をもたらす。たとえば、一部の学習者は、ピクニックで二人が通る道すじを「路線」と表現しているが、これは日本語として不適切である。

学習者にとっては、学習者マーカーの使用を再考するとともに、母語話者マーカーについて理解を深め、それらを適切に使用できるようにしていくことが1つの課題と言えるだろう。

5. まとめ

以上、本研究では、B-JASのストーリーライティングタスク(SW1)における産出データを用い、中国人日本語学習者16人の語彙使用の発達過程を母語話者と比較しながら量的・質的に概観した。その結果、以下の知見が得られた。

まず、RQ1(語彙量・語彙多様性)については、学習者の使用する語彙の量(異なり語数・延べ語数)および語彙多様性(R値)は段階的に上昇して母語話者に近づいていくが、2年生以後は天井効果が存在することが確認された。

RQ2(語彙難度)については、4年間で16%増加するものの、母語話者との間にはなお差が残ること、語彙産出量や語彙多様性の場合と同じく、学習後期において「天井効果」が存在することが確認された。

RQ3(品詞使用率)については、動詞だけが増加し、形容詞・形状詞・連体詞がともに減少すること、母語話者と比べ、4年生は助動詞・名詞・接尾辞・形容詞を過剰使用し、接頭辞・感動詞を過少使用することが確認された。

RQ4(マーカー語)については、5つのイラストごとに母語話者・学習者を特徴づける母語話者マーカー語と学習者マーカー語を特定した。母語話者マーカーには、(1)複合動詞、(2)慣用表現中での使用、(3)評価ニュアンスを含む動詞、(4)慣用的1語名詞、(5)難度の高い語、(6)日中同形異義語などが含まれ、学習者マーカーには、(7)意思性や意図性に関わる使い分けを誤った動詞、(8)文脈と合致しない語、(9)複合動詞構成要素、(10)絵に描かれていない行為やものを表す語、(11)時間表現、(12)日中同形語などが含まれる。

今後は、B-JAS に含まれるほかのデータを用いて学習者の語彙習得過程を全般的に考察するとともに、得られた知見を日本語語彙教育に応用する方法を探索したい。

注)本研究は北京外国語大学北京日本学研究センター、北京師範大学日本語教育教学研究所、国立国語研究所の連携による研究プロジェクト「北京日本語学習者縦断コーパス(B-JAS)」の研究成果の一部である。

引用文献

- Ishikawa, S. (2017). Learners' acquisition and use of L2 Japanese vocabulary: Influence of L1 backgrounds and L2 proficiency levels: A learner corpus-based analysis. 『第2言語としての日本語の習得研究』 20, 10-27.
- 石川慎一郎(2018a)「L2 日本語語彙の習得プロセスについて:LARP コーパスに見る台湾人学習者による日本語作文の縦断分析」『統計数理研究所共同研究レポート』 400, 1-13.
- 石川慎一郎(2018b)「中国語母語の日本語学習者の発話における使用語彙の変遷:発達段階の差と個体の差をめぐって」『第4回学習者コーパス・ワークショップ&シンポジウム「第2言語習得における語彙の役割」予稿集』(国立国語研究所) 62-76.
- 石川慎一郎(2021)「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」『統計数理研究所共同研究レポート』 444, 1-22.
- 金田一春彦(1988)『日本語』 岩波書店.
- 国立国語研究所(編)(2005)『現代雑誌の語彙調査:1994年発行 70誌』 国立国語研究所.
- 島崎英香(2020)「中国語を母語とする日本語学習者の話し言葉における副詞の使用傾向:I-JASを用いて習熟度別に」『言語資源活用ワークショップ発表論文集』 5, 20-32.
- 砂川有里子(2012)「学習辞書編集支援データベース作成について-『学習辞書科研』プロジェクトの紹介」『日本語教育連絡会議論文集』 24, 164-169.
- 董芸(2020)「日本語学習者の作文における並列・継起の接続表現の習得:中国語母語話者の縦断コーパスの分析を通じて」『国立国語研究所論集』 19, 127-138.
- 费晓东(2020)「基于历史语料库的中国学习者日语汉字词汇习得研究」『高等日语教育』 6, 30-42.
- 森田良行(2008)『動詞・形容詞・副詞の事典』 東京堂出版.
- 李在鎬(2017)(編著)『文章を科学する』 ひつじ書房.
- 李在鎬(2019)「学習者の語彙使用は習熟度を反映しているか:学習者コーパスの定量的分析」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. 87-104). くろしお出版.
- 李在鎬(2020)「作文における産出量と語彙特徴」迫田久美子・石川慎一郎・李在鎬(編)『日本語学習者コーパス I-JAS 入門』(pp. 144-156). くろしお出版.