



日本語学習者産出データにおけるオノマトペ使用の 規定要因 : 機械学習を用いた検討の試み

廉, 沢奇

(Citation)

統計数理研究所共同研究レポート, 484:73-91

(Issue Date)

2026-03-25

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/0100502564>

(URL)

<https://hdl.handle.net/20.500.14094/0100502564>



日本語学習者産出データにおけるオノマトペ使用の規定要因

—機械学習を用いた検討の試み—

廉 沢奇(神戸大学・大学院生)

Lzq980628@outlook.com

Factors Predicting Onomatopoeia Use in Japanese Learners' Production Data

—An Exploratory Study Using Machine Learning—
LIAN, Zeqi (Kobe University, Graduate Student)

概要

本研究は、廉(2024)で抽出した現代日本語で実際に使われているオノマトペ 200 語を対象に、日本語学習者が異なるタスクでオノマトペを「使用するか／どれくらい使用するか」を明らかにすることを旨としたものである。手法として、『多言語母語の日本語学習者横断コーパス』(I-JAS)に収録された 4 タスク(対話・ストーリーテリング・ストーリーライティング・作文)の産出データを用い、(i) 習熟度別の使用者率・使用量を調査したうえで、(ii) 使用有無と使用量を分けて捉える二段階の混合効果モデルにより習熟度とタスクの影響を推定し、さらに(iii) 機械学習により使用有無の予測と重要要因の検討を行った。これより、オノマトペはタスクによって出現の仕方が大きく異なり、いずれのタスクでも習熟度の上昇に伴って使用者率が大きく高まる一方、使用者に限った使用量の増加は比較的緩やかであることがわかった。加えて、機械学習は使用有無を一定の精度で予測できた(ROC-AUC \approx 0.82)が、学習歴・訪日経験などの背景要因を加えても性能向上は小さく、産出量・習熟度・タスクが主要な規定因であることが示された。

キーワード

オノマトペ、日本語学習者、機械学習、規定要因

1. はじめに

雨がしとしと降る。胸がどきどきする。人がぞろぞろ集まる。——日本語では、出来事の音だけでなく、主体の動き方や状態、感情の「手触り」までを一語で描き出すオノマトペが日常的に用いられる。会話や小説、漫画といった典型的な場面に限らず、広告や SNS の短い文でも、オノマトペは情報を圧縮しつつ臨場感を付け足す便利な資源である。辞典には 4500 語を超える項目が収録されるとも言われ(小野、2007)、その豊富さ自体が日本語の特徴の一つになっている。

しかし、この「便利さ」は学習者にとって必ずしも扱いやすいとは限らない。語数が多いだけでな

く、似た表現のニュアンス差、文体(話し言葉／書き言葉)との相性、共起しやすい語や場面の制約など、辞書の説明だけでは判断しにくい点が多いからである。結果として、学習者はオノマトペを避けて一般語彙で言い換えることもあれば、使いたい気持ちはあっても「どこで使えば自然か」が分からず不自然な使用につながることもある。にもかかわらず、日本語教育ではオノマトペが独立した指導項目として体系的に位置づけられているとは言い難く、現場の指導は教材・教師の経験に委ねられやすい。

より効果的な指導を設計するためには、まず学習者が実際の産出でオノマトペをどの程度用いているのか、どのような条件で用いられやすいのかを把握する必要がある。しかし、学習者研究全体として見れば、対象語彙やタスクが限定されることが多く、「タスクの違い」と「学習者の個人差」を同時に扱いつつ、使用の広がりや定量的に捉えた研究はまだ十分とは言えない。とりわけ、習熟度や母語だけでなく、学習歴や日本語との接触経験といった背景要因まで含めて、使用傾向を広く見渡す試みは限られている。

ここで、学習者の産出がいかに関「個人の条件」に左右されうるかを、具体例で確認したい。以下は、I-JAS の対話タスクにおけるやりとりの一部であり、いずれも「忙しさ／時間的余裕」を話題としながら、オノマトペの使用が分かれた例である。

(1) 調査者:昨日はどんな一日でしたか?

協力者:昨日も授業がなくて、お昼は何もしていません。ただ、寮でアニメを見たりして、映画を見たりして、のんびり過ごしていました。

(協力者番号:CCH09、母語＝中国語、年齢＝21、習熟度＝J-CAT(294/400)、教育機関で日本語学習歴あり、訪日経験あり)

(2) 調査者:昨日一日忙しかったですか?

協力者:最近そんなに忙しくないです、勉強、終わってます、そして、そんなに、忙しくないですよ。

(協力者番号:SES25、母語＝スペイン語、年齢＝21、習熟度＝J-CAT(212/400)、教育機関で日本語学習歴なし、訪日経験なし)

上記の例が示しているように、(1)ではオノマトペの「のんびり」が用いられ、時間に追われない過ごし方が一語で具体的に提示されているのに対し、(2)では同様の内容が「忙しくない」で表現され、オノマトペは用いられていない。もちろん、この2例だけで原因を断定することはできない。とはいえ、少なくとも「習熟度が高い学習者ほど使いやすいのか」「学習環境(機関学習や訪日経験など)が後押しするのか」「単に産出の長さが違うだけなのか」といった問いが自然に立ち上がる。個別の直観に頼らず、こうした問いをまとめて検討するには、多数の学習者を含み、タスクと背景情報が揃ったデータが必要になる。しかし、そのようなデータを扱って調査する研究は、筆者の管見の限り、極めて少ない。

そこで本研究では、I-JAS に収録された学習者の産出データを用い、廉(2024)で抽出した現

代日本語で実際に使われているオノマトペ 200 語を対象として分析を行いたい。

2. 先行研究

オノマトペの習得研究は、(a)意味理解・知識、(b)産出(課題実験)、(c)学習者コーパスに基づく産出分析に大別できる。本研究は産出データにおける使用実態(使用するか/どれくらい使うか)を扱うが、当該領域では依然として研究蓄積が多いとは言えない。ここでは、前田(2009)、グエン(2018)、張(2018)の3研究を取り上げて整理したい。

前田(2009)は、ドイツ人日本語学習者 27 人(10~30 歳代)と日本語母語話者 22 人(20~50 歳代)を対象に、豊語型オノマトペ 24 種(がらがら、どんどんなど)について意味理解を調査した。具体的には、オノマトペを音声で提示し、さらにそれらを含む文を読ませたうえで意味を推測させ、意味が対立する形容詞ペアを用いた 5 段階尺度(-2~+2)で回答させている(いわゆる SD 法)。加えて、日本語学習経験のないドイツ人 22 人にも同様の推測課題を実施した。その結果、学習者はおおよその意味を捉えられる一方、正確さは母語話者に及ばないこと、また意味推測の成績が日本語能力や学習経験と必ずしも対応しないことなどが示された。

グエン(2018)は、中国人(20名)とベトナム人(19名)の日本語専攻大学生(いずれも N2 合格)を対象に、オノマトペ産出を引き出す課題によって知識差を検討した。調査では、21 種の短いアニメーションを提示し、「風車が風で()回っている」のような穴埋め形式で適切なオノマトペを記入させている。その結果、正答率は中国人 29%、ベトナム人 55%であり、同程度の日本語能力であっても母語によってオノマトペ知識に差が生じることが示された。また誤答の傾向として、連想に基づき既知語の語基を反復させて豊語型を作る(例:まるまる、まわまわ)といった方略が観察され、学習者の産出メカニズムを考える上でも示唆的である。

張(2020)は、学習者インタビューや作文等を収録した大規模な学習者コーパス『多言語母語の日本語学習者横断コーパス』(I-JAS)を用い、中国語・韓国語・英語母語話者および日本語母語話者を対象に、使用量・多様性・使用語・誤用の観点からオノマトペの使用実態を分析した。その結果、使用量・多様性・使用語の面で中国語母語話者は韓国語・英語母語話者より高いものの母語話者とは大きく異なること、誤用の割合は中国語母語話者で高いことなどが示された。また、レベル間の差が明確ではない点も報告されている。

以上より、先行研究はオノマトペの理解・知識、実際の産出の各面で重要な知見を蓄積してきた一方で、(i)複数タスク(話す/書く等)の条件差を明示的に組み込み、(ii)「使用有無」と「使用量」を分けて捉え、(iii)習熟度・母語に加えて学習歴や接触経験などの背景要因も含めて、大規模データで同時に検討する枠組みは十分に整っていないといった課題が存在する。そこで本研究は、学習者コーパスのタスクと学習者の背景データを調査資料とし、学習者の使用を調査することで、新しい観点から、学習者のオノマトペ使用の現状を把握することを目指す。

3. リサーチデザインと手法

3.1 研究目的と RQ

すでに述べたように、本研究の目的は、学習者の産出におけるオノマトペ使用を、タスクの違いと、学習者の個人差(習熟度、母語、年齢、性別、学習・接触経験等)の両面から捉え、学習者が「オノマトペを使用するか」と「使用する場合にどれくらい使用するか」が、どのような条件で変化するのかを明らかにすることである。あわせて、統計モデルによる説明に加え、機械学習による予測という観点から、どの要因が使用の有無と強く結びつくのかを検討する。この目的に沿って、以下の3つの研究課題を設定する。

- RQ1 タスクと習熟度によって、オノマトペの使用者率および使用量はどのように異なるか。(タスクと習熟度による使用傾向の調査)
- RQ2 タスクと習熟度は、(a)学習者の使用有無、(b)使用者における使用量にそれぞれどのような影響を与えるか。(混合効果モデルに基づく使用有無と使用量の推定)
- RQ3 学習者属性からオノマトペ使用有無はどの程度予測できるか。また、学習環境・接触経験などの背景要因を加えると予測性能は向上するか。(機械学習によるオノマトペ使用の予測と重要要因)

3.2 データ

本研究で用いるデータは、『多言語母語の日本語学習者横断コーパス』(I-JAS)である。I-JASは、学習者および日本語母語話者を対象とした対面調査と非対面調査により、発話データと作文データを複数タスクで収集した大規模コーパスであり、発話・作文に加えて言語テストの得点や詳細な背景情報を包含している点に特徴がある。

調査対象とするタスクは4種類である。具体的には、対話(I)、ストーリーテリング(ST)、ストーリーライティング(SW)、作文(E)であり、分析対象語は、廉(2024)で抽出した現代日本語で実際に使われているオノマトペ200語である。

調査協力者について、I-JASに収録された学習者データには、学習環境の観点から見ると、海外環境で学習する学習者と日本国内環境で学習する学習者という2種類がある。しかし、日本国内環境の日本語学習者は滞在形態や接触量などの変数が多様で、要因が複雑化しやすいため、本研究は日本国内環境を調査せず、海外環境の全850名の学習者を対象とする。以下は母語別参加者の詳細である。

表 1

本研究で使う I-JAS の参加者の詳細

母語	人数
インドネシア語、スペイン語、タイ語、トルコ語、ドイツ語、ハンガリー語、フランス語、ベトナム語、ロシア語	450 (各 50)
英語	100
韓国語	100
中国語	100
中国語 (台湾)	100
合計	850

参加者の背景について、I-JAS では、日本語能力を客観的に把握するため、J-CAT(聴解・語彙・文法・読解)および SPOT 等の言語テスト結果が収録されている。また、学習者の背景情報を収集するために、事前のアンケート(フェイスシート: FS)が整備されている。FS は学習者の属性、現在の言語環境、日本語学習の経緯などに関する複数項目からなり、回答負担とプライバシーに配慮して「答えたくない／答えられない」といった選択肢を設けるとともに、ウェブ入力を可能にする仕組みが採用されている。さらに、学習者が母語で回答できるよう複数言語版が用意され、入力漏れや不備があった場合には対面調査で確認が行われている。

本研究では、この背景情報のうち、欠損が比較的少なく、学習・接触経験を代表すると考えられる 5 項目(日本語環境、日本語母語話者の友人、他言語運用、教育機関での学習経験、訪日経験)を取り上げ、産出におけるオノマトペ使用との関係を検討する。

3.3 手法

3.3.1 データの事前処理

本研究では、分析を安定して行うために、I-JAS のデータから「分析用の一枚表」を作成した。分析単位は 1 行=1 学習者×1 タスク とし、同一学習者が複数タスクを実施しているという I-JAS の設計を、そのまま統計モデルで扱える形に整えた。

まず、対象タスクは 4 種類(対話 I、ストーリーテリング ST、ストーリーライティング SW、作文 E)とした。I-JAS にはストーリーテリングとストーリーライティングがそれぞれ 2 課題ずつ(ST1 と ST2、SW1 と SW2)含まれるため、本研究では ST1+ST2 を ST に、SW1+SW2 を SW に合算した。

次に、200 語のオノマトペを対象語彙とし、タスク別に各学習者におけるオノマトペの出現を集計した。具体的には、当該タスクの産出全体について、記号を除いた総トークン数を語数(words)とし、異なるタスクの語数差が大きいため、語数の影響を正しく確認できるよう、自然対数で圧縮された $\ln(\text{words})$ を使うことにする。さらに、対象 200 語の延べ出現回数を token として算出し、対象語彙が 1 回以上出現したかどうかを use (0/1)として定義した(token>0 を use=1)。記述統計で

は、タスク間で産出量が異なることを踏まえ、補助指標として 100 語あたりの token (token_per100) も算出した。

習熟度については、I-JAS に収録された J-CAT (聴解・語彙・文法・読解 4 下位得点) と SPOT の計 5 指標を用いた。これらは測っている側面や得点尺度が異なるため、単純平均ではなく、相関のある複数指標を一つの総合指標にまとめる目的で主成分分析 (PCA) を行い、主成分得点を総合的な習熟度指標 (prof) として採用した。なお、RQ1 では全体像を見やすくするため prof を ±1SD で 3 水準 (初級 / 中級 / 上級) に区分したが、RQ2・RQ3 では情報損失を避けるため連続変数のまま用いた。

最後に、学習者背景要因を扱うため、I-JAS のフェイスシート (FS) から、自由記述項目を避けつつ欠損が比較的少ない二値項目 5 つを抽出した (設問 10「住環境で日常的に日本語が話されるか」、設問 11「親しい友人に日本語母語話者がいるか」、設問 13「母語以外に日常的に使える言語があるか」、設問 18「教育機関での日本語学習経験があるか」、設問 20「訪日経験があるか」)。回答は「はい / いいえ」に加え、未回答を分析から落とさないため unknown を設け、no / yes / unknown の 3 水準に統一した。背景情報は「1 人 1 行」の表として整形し、学習者 ID で主表に結合した (同一学習者の全タスク行に複製される)。また、年齢はモデルの解釈をわかりやすくするため中心化し、全体平均との差に変換した。

以上の処理により、以降の分析で共通して用いる、学習者 × タスク単位の分析テーブルを構築した。

3.3.2 分析の手順

RQ1 (タスクと習熟度による使用傾向の調査) では、タスク (I・ST・SW・E) と習熟度 (初級・中級・上級) ごとに、(a) 使用者率 (use) と (b) 使用量 (token_per100) を集計し、全体像を示す。また、差の有意性を確認するために、使用者率に関しては、0/1 という二値で、対応のあるデータを比較するための Cochran の Q 検定、使用量に関しては、分布が偏っている可能性が高いため、値そのものではなく順位に基づいてタスク差を検討する Friedman 検定を用いた上で、多重比較 (Holm 法) を行う。

RQ2 (混合効果モデルに基づく使用有無と使用量の推定) では、同一学習者が複数タスクを行っている点 (= データが独立でない点) を踏まえ、学習者ごとの個人差を統計的に扱える混合効果モデル (「同じ学習者は似た傾向をもつ」という点を織り込み、学習者 ID をランダム効果として導入する回帰モデル) を用いる。さらに、ゼロが多いというデータの性質に合わせ、次の二段階で分析する。

(a) 使用有無 (use)

まず「そのタスクでオノマトペを 1 回以上使ったか」を目的変数とし、ロジスティック混合効果回帰で、タスク・習熟度 (連続) ・その交互作用の影響を推定する。係数はオッズ比 (OR) として報告し、「他条件が同じとき、どの要因が“使いやすさ”をどれだけ押し上げるか」を読み取れる形にする。統制変数として年齢、性別、母語も加える。

(b)使用量(token)

次に、 $use=1$ ($token>0$)のデータに限定し、「使う学習者の中で、どれくらいの回数使うか」を推定する。回数データはばらつきが大きくなりやすいため、過剰分散に適している負の二項分布を用い、かつこの段階では0が存在しないため、「1以上の値しか取らない」ように修正したゼロ切断のモデルとする。

さらに重要なのが $offset$ である。タスクによって発話・文章の長さ(語数)が違う以上、単純に回数だけ比べると「長い人ほど多い」という当たり前の影響が混ざってしまう。そこで $offset(\ln(words))$ を入れることで、「長さが同じだと仮定したときの出現しやすさ」を比較できるようにする。直感的には「tokenをwordsで割って率を見る」ことに近いが、実際には回数データとしての性質(分散の大きさ等)を保ったまま、長さの差だけをモデル内で補正できる点に利点がある。係数は発生率比(IRR)として報告する。

加えて、本研究では「説明できた」だけでなく「どれくらい当てられるか」も示すため、学習者ID単位で訓練用/評価用に分割し、未知の学習者への予測性能も算出した。この点において、ID単位で分けるのは、同一学習者の別タスクが学習と評価の両方に入ると、実力以上に当たって見えてしまうためである。このとき、学習者固有のランダム効果に依存しない形で予測し、過大評価(リーク)を避ける。

RQ3(機械学習によるオノマトペ使用の予測と重要要因)では、統計モデル(RQ2)が主に「効果の推定(説明)」を目的とするのに対し、機械学習を用いて「どこまで当てられる情報か(予測)」という別の角度から検討する。目的変数は use (yes/no)とし、手法はランダムフォレストを用いる。

特徴量は、(A)baseline(タスク、語数、習熟度、年齢、性別、母語)と、(B)baselineに背景5変数(日本語環境、母語話者の友人、他言語運用、機関学習経験、訪日経験)を加えたモデルを用意し、背景要因の追加が性能改善につながるかを比較する。評価では、同一学習者のデータが学習と評価に同時に入らないよう、学習者IDを単位とする分割と検証を行う。

さらに、「何が効いているか」を説明するため、順列特徴重要度(Permutation Importance)(その変数の値を入れ替えて予測がどれだけ悪化するか)により重要度を算出し、部分依存(PDP)により、他条件を平均的に保ったまま当該変数だけを変化させたときの予測確率の動きから、影響の方向性(上がる/下がる)を整理する。

4. 結果と考察

4.1 RQ1 タスクと習熟度による使用傾向の調査

すでに述べたように、オノマトペの指導や学習支援を考えるうえでは、「学習者がどのタスクで(場面・文体の違い)、どの程度の習熟度でオノマトペを産出しているのか」という全体像をまず把握することが重要である。そこで RQ1 では、I-JAS の 4 タスク——対話(I)、ストーリーテリング(ST)、ストーリーライティング(SW)、作文(E)——について、(a)対象 200 語のオノマトペを 1 回以上使用した学習者の割合(使用者率)と、(b)使用者における使用密度(100 語あたり token)を記述的に概観したい。まず使用率について、以下の結果を得た。

図 1

タスク×習熟度別のオノマトペ使用者率

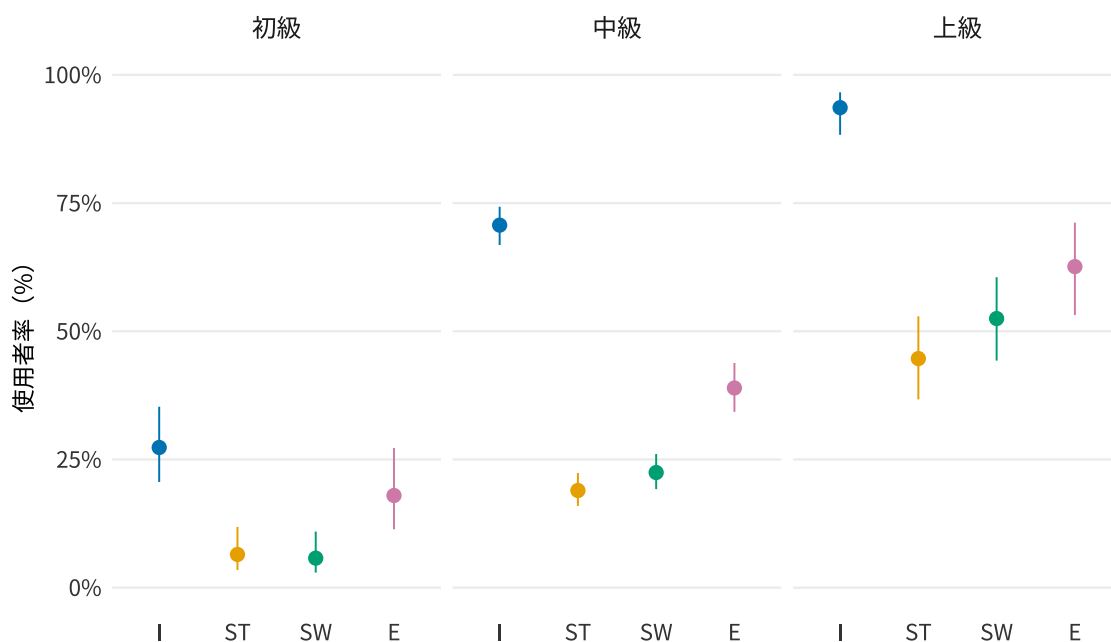


表 2

タスク×習熟度別の使用者率および使用量(記述統計)

タスク	LV	人数	使用人数	使用率	使用量 (全体 平均)	使用量 (全体 中央 値)	使用量 (使用 者平 均)	使用量 (使用 者中 央 値)
I	初級	139	38	0.27	0.03	0.00	0.12	0.09
	中級	570	403	0.71	0.10	0.06	0.14	0.11
	上級	141	132	0.94	0.18	0.16	0.19	0.16
ST	初級	139	9	0.06	0.03	0.00	0.46	0.42
	中級	570	108	0.19	0.10	0.00	0.51	0.42
	上級	141	63	0.45	0.23	0.00	0.52	0.42
SW	初級	139	8	0.06	0.03	0.00	0.54	0.52
	中級	570	128	0.22	0.15	0.00	0.65	0.58
	上級	141	74	0.52	0.31	0.31	0.59	0.50
E	初級	89	16	0.18	0.10	0.00	0.58	0.44
	中級	398	155	0.39	0.17	0.00	0.44	0.32
	上級	107	67	0.63	0.29	0.28	0.47	0.36

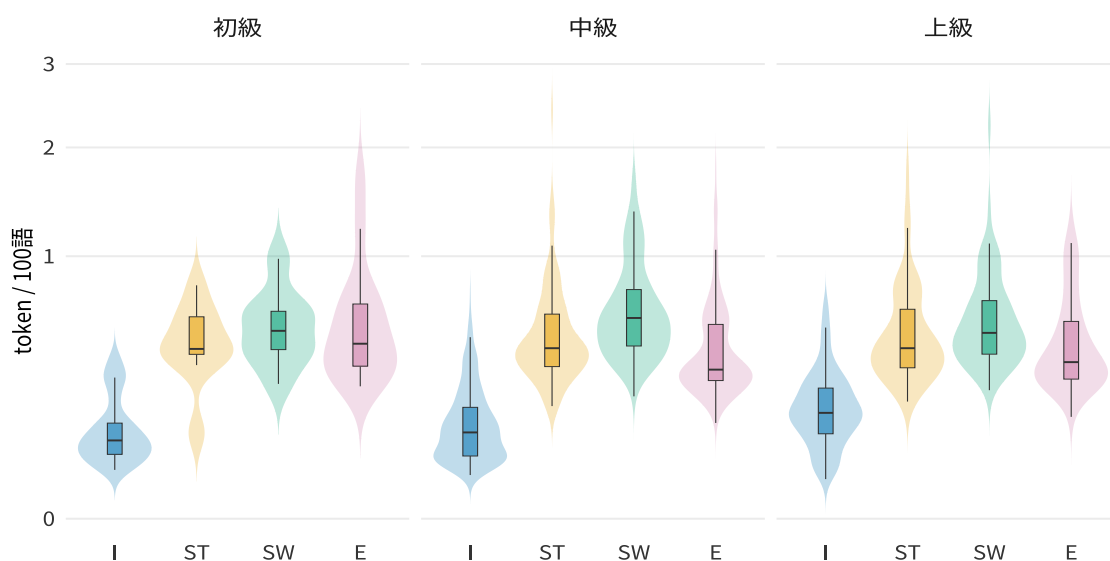
上記の図表からわかったことを2点まとめたい。まず、どのタスクでも習熟度が上がるほど使用者率が上昇することが分かる。例えば対話(I)では、初級 27.3%→中級 70.7%→上級 93.6%と大きく上昇している。同様に、ST(6.5%→18.9%→44.7%)、SW(5.8%→22.5%→52.5%)、E(18.0%→38.9%→62.6%)でも、上級に向けて「使う学習者」が増える。つまり、オノマトペは「一部の上級者だけが使う特殊表現」ではなく、習熟の進行に伴って使用が徐々に「当たり前の選択肢」になっていく語彙であることが、少なくとも使用有無の段階では確認できる。

そして、タスク差が非常に大きいことも明瞭である。全体として、対話(I)は常に使用者率が高く、上級ではほぼ全員が何らかのオノマトペを用いる。一方でSTとSWは、初級では使用者が1割未満であり、上級になっても約半数にとどまる。Eはその中間で、初級でもST・SWよりは使われやすく、上級では6割を超える。

このタスク差を、同一学習者内の比較として検定するため、4タスクすべてに参加した学習者について Cochran の Q 検定を行ったところ、タスク差は有意であった ($Q(3) = 437.14, p < .001$ 、Kendall's W = 0.245)。事後の検定(Holm 補正)では、Iは他のすべてのタスクより使用者率が高く、またEはST・SWより高いことが確認された。一方で、STとSWの差は有意ではなかった ($p = .33$)。この結果は、「話す／書く」という二分法だけでは説明しきれないタスク特性(相互行為の有無、描写要求の強さ、産出の自由度など)が、オノマトペの出やすさを左右している可能性を示唆する。

次に、「使うかどうか」だけでなく、「使う学習者が、どれくらいの頻度で使うか」に目を向けたい。非使用者が多いデータでは、100語あたり token を全体平均で見ると0が大量に混ざり、実態が見えにくくなる。そこで、使用者(token>0)に限定した使用頻度の分布を調査したところ、以下の結果を得た。

図2
使用者におけるオノマトペ使用量の分布



注:淡色の左右対称の形(バイオリンプロット)は値の分布を表し、幅が広いほどその付近に観測値が多いことを示す。中央の色付きの箱は第1四分位数から第3四分位数までの範囲、箱の中の横線は中央値、上下の黒線は分布の広がりを示す。

ここで最も重要な点は、図1では習熟度が上がるにつれて「使う人」の割合が劇的に増えていたのに対し、図2の分布を見ると、習熟度が上がっても使用量は初級から上級までそれほど大きく変化していないことである。一方、「習熟度差」よりも「タスク差」のほうがはっきり現れることである。使用者に限ってみると、対話(I)は中央値が概ね0.1~0.16程度であるのに対し、STは0.4前後、SWは0.5前後、Eは0.3~0.4程度に位置している。言い換えると、対話では「使う人は多いが、一人ひとりの使用は薄く広がる」一方で、物語産出や作文では「使う人は限られるが、使うときは比較的まとまって出やすい」。この対照は、次節のRQ2で二段階(使用有無/使用量)に分けて扱う必要性を、RQ1の時点で既に裏づけている。

さらに、タスク差を含めた全体の使用量(非使用者を含む)についてFriedman検定を行うと、タスク差は有意であった($\chi^2(3) = 114.69, p < .001$)。ただし効果量(Kendall's W = 0.064)は小さく、これは多くの条件で中央値が0になる(=非使用者が多い)ことに由来する。つまり、この段階で重要なのは「平均値が少し違う」ことよりも、「0か1か」の境界(使う/使わない)そのものが大きく動くという事実である。

この結果は、オノマトペ習得の最大の壁は「0から1への壁」であることを強く示唆している。学習者は、上級になったからといって文章中にオノマトペをむやみに使用するわけではない。使えるようになった学習者は、長い発話や文章のなかで自然で使いやすい場面に絞って、1・2回だけ効果的にオノマトペを配置しているのである。

以上、RQ1では、(i)どのタスクでも習熟度の上昇に伴って使用者率が上がること、(ii)使用者率は対話(I)が高いこと、(iii)一方で使用者に限った使用量はST・SW・Eで高く、Iでは相対的に低いことが確認された。次のRQ2では、この傾向が「年齢・性別・母語の違い」や「同一学習者の反復(複数タスク)」を考慮しても成立するのかを、混合効果モデルにより推定したい。

4.2 RQ2 混合効果モデルに基づく使用有無と使用量の推定

上記RQ1は使用有無と使用量の全体像をつかむうえで有効であった一方、(a)同じ学習者が複数タスクを行っていること(データの非独立性)、(b)学習者の年齢・性別・母語といった要因が同時に絡むこと、(c)産出の長さの違いが回数に直結しうることを、十分に切り分けてはいない。そこでRQ2では、学習者IDをランダム効果として入れた混合効果モデルを用い、「使うか」と「使う場合にどれくらい使うか」を分けて推定する。

4.2.1 使用有無

まず、「使用有無」について、ロジスティック混合効果回帰で、年齢や母語などの条件を平均的な状態に揃えたとき、タスク、習熟度とその交互作用の影響を調査したところ、以下の結果を得た。

図 3

タスク×習熟度に基づく使用確率の予測曲線

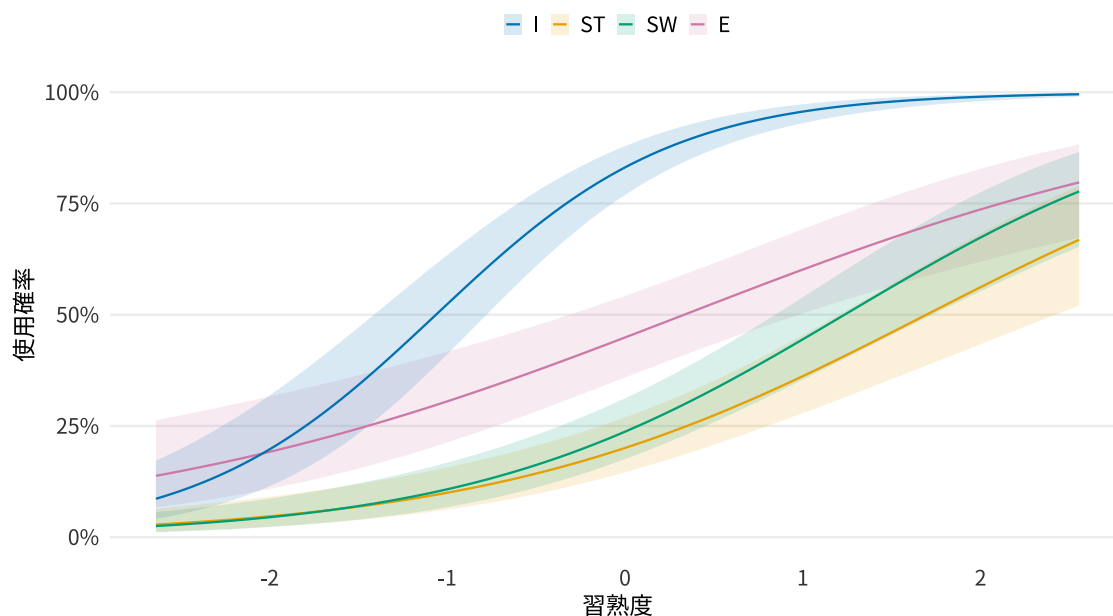


表 3

使用確率モデルにおける各変数の効果(オッズ比、基準=対話(I))

変数	オッズ比	p 値
taskST	0.051	<.001
taskSW	0.063	<.001
taskE	0.165	<.001
prof	4.476	<.001
taskST:prof	0.505	<.001
taskSW:prof	0.576	<.001
taskE:prof	0.414	<.001

図が示しているように、使用確率はすべてのタスクで習熟度とともに上昇する。しかし上昇の仕方はタスクによって異なり、とりわけ対話(I)では、低習熟で低いところから急激に立ち上がり、高習熟ではほぼ 100%に達する。対話が最も使用されやすいという傾向は RQ1 と一致するが、混合効果モデルにより、年齢・性別・母語を同時に統制してもこの差が残る点は重要である。

また、表を見ると、習熟度(prof)は対話(I)を基準としたとき、1SD 上昇で使用のオッズが約 4.48 倍となる(OR=4.48)。また、同じ習熟度でも ST・SW・E では I より使用されにくく、例えば ST は I の約 0.05 倍(OR=0.051)、SW は約 0.06 倍(OR=0.063)、E でも約 0.17 倍(OR=0.165)である。

さらに、交互作用 ($\text{task} \times \text{prof}$) がすべて 1 より小さいことは、「習熟度が上がったときの伸び幅」が I で最も大きく、ST・SW・E では相対的に緩やかになることを意味する。言い換えれば、オノマトペは「習熟度が上がればどのタスクでも使えるようになる」が、とくに対話では、習熟度が「使う／使わない」を分ける決定的な要因になりやすい。

また、このモデルでは学習者間の個人差(同じ人がどのタスクでも使いやすい／使いにくい)が一定程度存在し、ICC は約 0.20 であった。つまり、タスクや習熟度だけでは説明しきれない「その人らしさ」が、使用有無の段階でも確かに残っている。

補助的に、未知の学習者に対する予測性能を確認すると、ROC-AUC=0.818、PR-AUC=0.736、F1=0.673 であった。ここで ROC-AUC は「使う人を使わない人より上位に並べられる度合い」、PR-AUC と F1 は「使う(yes)をどれくらい取りこぼさずに、かつ誤って yes と言いつぎないか」を表す。完全ではないが、タスク・習熟度・基本属性だけでも、使用有無は一定程度見通せることが分かる。

4.2.2 使用量(token)

上記では、タスク、習熟度とその交互作用が学習者の使用有無に及ぼす影響を確認した。では、「使う」側に回った学習者は、産出の長さを揃えた場合、どのタスクでより多くオノマトペを産出しているのだろうか。次に、使用者 ($\text{token} > 0$) に限定し、語数を統制したうえで使用量を調査したところ、以下の結果を得た。

図 4

タスク×習熟度に基づく使用量の予測曲線

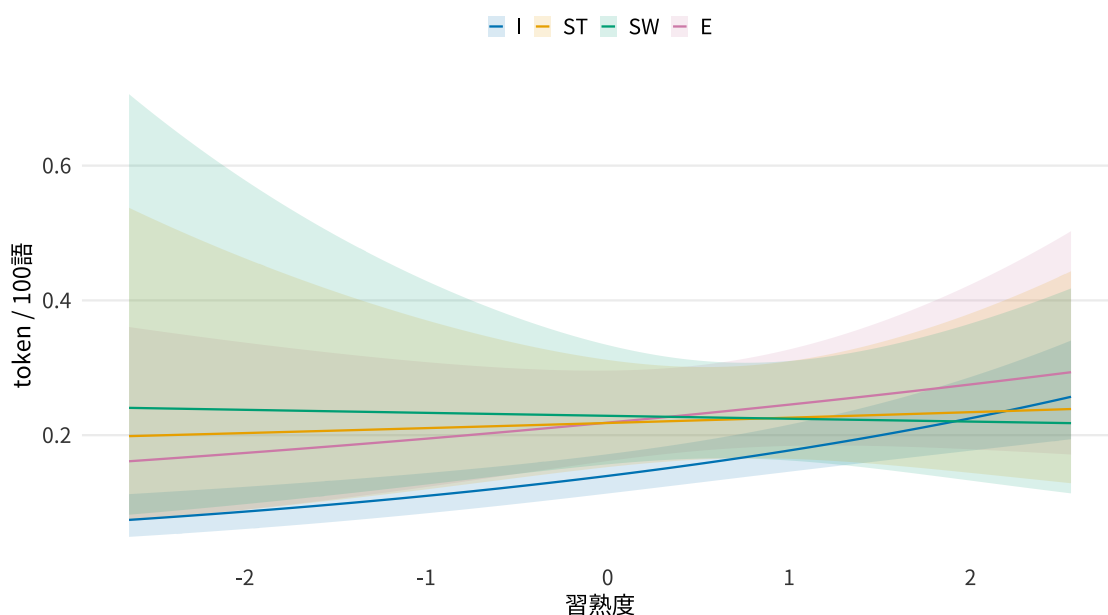


表 4

使用確率モデルにおける各変数の効果(発生率比 IRR、基準=対話(I))

変数	発生率比	p 値
taskST	1.562	0.005
taskSW	1.639	0.005
taskE	1.566	0.001
prof	1.270	<0.001

図が示しているように、使用者に限ると、ST・SW・E は I より高い水準にあり、タスク差が残る。上表で変数の効果を確認しても、I を基準として ST は約 1.56 倍、SW は約 1.64 倍、E は約 1.57 倍の使用量となり (IRR=1.56~1.64)、「使うなら物語・作文のほうが濃く出やすい」ことが確認された。これは RQ1 の分布とも整合的である。

一方、習熟度の効果は有意ではあるものの (IRR=1.27)、使用有無ほど劇的ではない。直感的に言い換えると、習熟度が 1SD 上がると、同じ長さの産出でオノマトペ回数は約 1.3 倍になるが、それ以上に「そもそも使うかどうか」の壁が大きい、という構図である。

ここで注目すべきは、モデルの説明力 (marginal $R^2=0.047$) が低い点である。これは「モデルの性能が悪い」というより、「使う人の中で、どれだけ使うか」は、タスク・習熟度・基本属性だけでは決まりにくいことを示していると考えられる。物語のどの場面に焦点を当てたか、どの語を好むか、書き言葉の文体をどう意識したかといった、より局所的・内容依存的な要因が強く働く可能性が高い。

この意味で、RQ1 で見えた「I は使用者率が高いが密度は低い」「ST・SW は使用者率が低い密度は高い」という二面性は、RQ2 によって平均的傾向としては確かに存在する一方、使用量にはなお大きな個人差・内容差が残ることが明確になった。

以上、RQ2 では混合効果モデルにより、(i)使用有無は習熟度とタスクの影響を強く受け、とくに対話で習熟度差が大きいこと、(ii)使用者における使用量は習熟度・タスク差に有意に影響されるが、モデルの説明力が低く、むしろ「使う／使わない」の段階が主要な分岐点であることが示された。次の RQ3 では、この知見を別角度から確かめるため、ランダムフォレストによって使用有無をどの程度予測できるか、また背景 5 変数を追加したときに性能が向上するかを検討する。

4.3 RQ3 機械学習によるオノマトペ使用の予測と重要要因

以上、RQ1 は、習熟度・タスク別のオノマトペ使用の全体像を掴み、RQ2 は習熟度・タスクが使用に与える影響を統計的に解明し、オノマトペの「使う／使わない」が学習者の主要な分岐点であることがわかった。次に重要なのは、その使用有無に関係する要因を探ることである。つまり、「学習者属性から使用有無がどれくらい当てられるのか」、また「背景要因(接触経験など)を加えると、本当に情報が増えるのか」という問いが重要である。そこで RQ3 では、ランダムフォレストで使用有無に影響する要因を探る。以下順に、モデルの予測性能、重要変数と変数の方向性の 3 点に絞っ

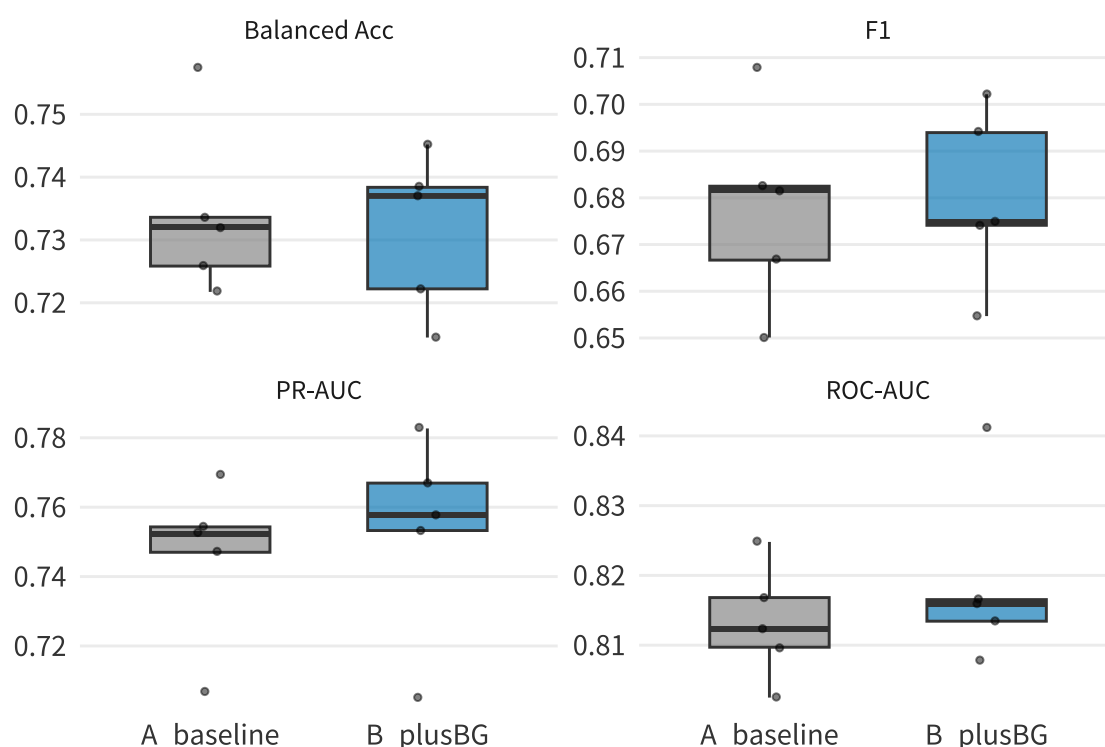
で紹介したい。

4.3.1 予測性能

まず、ベースラインとなるモデル A(タスク、語数、習熟度、年齢、性別、母語)と、ベースラインに背景 5 変数(日本語環境、母語話者の友人、他言語運用、機関学習経験、訪日経験)を加えたモデル B の予測性能を比較したところ、以下の結果を得た。

図 5

両モデルの予測性能比較(A: ベースライン、B: +背景 5 変数)



図が示しているように、ベースラインモデルだけでも、Balanced Acc(偏りに補正した全体の正解率)は約 0.73、F1(予測の「当てた確かさ」と「取りこぼしの少なさ」を総合的に評価する指標)は約 0.68、PR-AUC(正解予測の正確さ)は約 0.75、ROC-AUC(正解と誤解を上手く分類する能力)は約0.81であり、使用有無は一定程度予測できる。一方で、背景 5 変数を追加したモデル(+BG)は、ROC-AUC や PR-AUC がわずかに上がるものの、差は小さく、ばらつきの範囲内に収まっている。すなわち、本データにおいては、背景要因は「まったく無関係」ではない可能性があるが、少なくとも予測性能を大きく押し上げるほどの追加情報にはなりにくい。

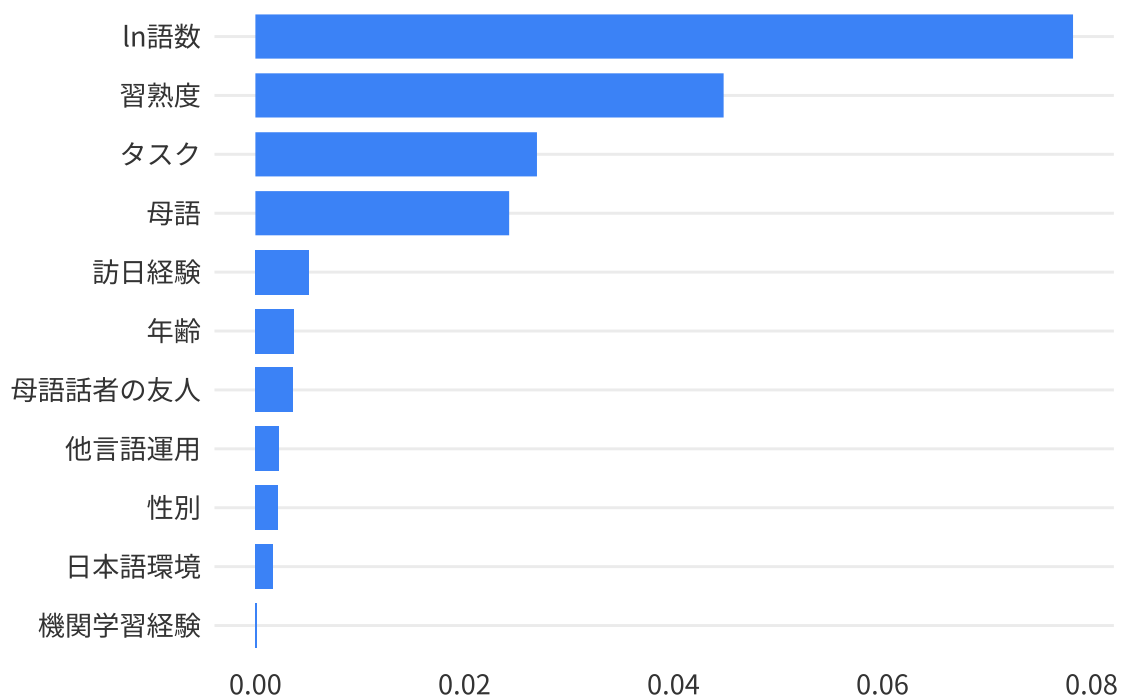
この結果は、背景要因が重要でないとは断定するものではない。むしろ、(i)背景項目が「はい/いいえ/未回答」という粗い情報であること、(ii)背景要因の効果が習熟度に一部吸収されること、

(iii)オノマトペ使用が内容・文体選好に強く依存すること等により、「予測」という観点では増分が小さく見える可能性がある、と解釈するのが妥当であろう。しかし、使用有無を影響する要因を探るには、モデル全体だけではなく、変数それぞれの重要度を確認する必要があるだろう。これに関しては次節で検討したい。

4.3.2 重要変数

次に、モデル内で「どの変数が予測に強く貢献したか」を示す順列特徴重要度 (Permutation Importance)を確認したところ、以下の結果を得た。

図 6
変数重要度



図で示されているように、重要度の上位が語数>習熟度>タスク>母語の順になっている。産出量が最上位であるのは、「1回でも出現する」確率が、産出量が多いほど上がるという極めて自然な性質を反映している。これは同時に、使用者率の比較 (RQ1) や使用有無のモデル (RQ2) を考察する際に、「タスクや学習者によって産出量が異なる」という前提を意識する必要があることを証明したものである。また、ここで重要なのは、背景 5 変数の中に「訪日経験」が TOP5 に入っていることである。これは日本という目標言語環境で滞在することが、オノマトペを実際に産出できるかを影響することを示唆している。

また、残りの 4 つの背景要因は上位要因に比べると重要度が小さいが、これは決して「他言語運

用や友人の存在が無意味である」ということではない。むしろ、そうした豊かな背景経験は、すでに「習熟度」や「産出量」という形に吸収・凝縮されている可能性があると思われる。

では、上位変数は実際どのように使用率に影響しているだろうか。次節では変数の方向性を確認したい。

4.3.3 方向性 (PDP)

最後に、これら上位の重要変数が、使用確率を「プラスに押し上げるのか、マイナスに下げるのか」、その方向性を部分依存プロット (PDP) で確認したところ、以下の結果を得た。

図 7

上位 5 変数の方向性まとめ



まず、語数について、長いほど使用確率が上がり、上位 10%点と下位 10%点の差は約 39%である。これは単に「長いほど出る機会が増える」という面に加え、長い産出ほど描写や評価が増え、オノマトペを挿入する余地が増える可能性も示唆する。

習熟度について、高いほど使用確率は上昇する。上位 10%点と下位 10%点の差は約 31%であり、習熟度は「使う／使わない」を大きく分ける要因である。習熟度が高いほど使用確率が上がるという当然の事実を機械学習も独自に見つけ出している。

タスクに関しては、平均的には対話(I)が最も高く、STが最も低い。SWとEは中間に位置する。これは RQ1 の使用者率、RQ2 の推定とも整合的であり、「タスクによってオノマトペが出やすい／出にくい」という性質が、手法を変えても一貫して観察されることを示している。

訪日経験は「yes」が「no」よりわずかに高く、差は約+0.02(2 ポイント強)である。大きな効果とは言えないが、実地接触が「使ってみる」方向に働く可能性は残る。ただし因果は不明であり、動機づけや学習機会の豊富さといった交絡要因も考えられる。

特筆すべきは、重要度第 4 位の「母語(L1)」の影響である。最後に母語差を見ると、中国語を基準とした場合、中国語(台湾)はやや高く(約+3.5%高)、それ以外の多くの母語群は低い方向に位置する(最小で約-13.7%)。この差が「言語類型」だけで説明できるとは限らず、教育環境、教材、メディア接触、評価観など複数の要因が混ざっている可能性が高い。しかし少なくとも、習熟度やタスク、産出量を考慮しても母語差が残ることは、オノマトペの学習が「一般的な日本語能力」だけでは捉えきれない側面を含むことを示唆する。

以上、RQ3 では、ランダムフォレストにより使用有無が ROC-AUC 約 0.82 程度で予測可能であること、ただし背景 5 変数の追加による性能向上は小さいことが示された。また重要度と PDP から、使用有無に強く関わるのは産出量・習熟度・タスク・母語であり、訪日経験は小さく正向であることが分かった。

5. まとめ

本研究は、日本語学習者にとって習得が困難とされるオノマトペが、実際の産出においてどのような条件で現れるのかを解明することを出発点とし、大規模な横断コーパス(I-JAS)のデータを用いて実証的な分析を行ったものである。具体的には、高頻度オノマトペ 200 語を対象に、記述的な集計にとどまらず、混合効果モデルによる統計的推論と、ランダムフォレストによる機械学習的な予測を組み合わせることで、タスクの違いと学習者の個人差がオノマトペ使用に与える影響を多角的に検討した。3 つの研究設問について、以下の知見が得られた。

RQ1(タスクと習熟度による使用傾向の調査)については、いずれのタスクにおいても、習熟度の上昇に伴いオノマトペを産出する学習者の割合(使用者率)が着実に増加することが確認された。同時に、使用者率には大きなタスク差があり、とりわけ対話(I)で使用者率が際立って高いことが分かった。一方、使用者に限って使用量を見ると、対話(I)よりも、ストーリーテリング(ST)・ストーリーライティング(SW)・作文(E)のほうが相対的に高い点である。すなわち、対話は「使い始める」ことが起きやすいが、物語産出や作文は「使うと決めた学習者が、文章の中で効果的に織り込む」場

になりやすいという、使用有無と使用量の二面的な振る舞いが浮き彫りとなった。

RQ2(混合効果モデルに基づく使用有無と使用量の推定)では、この二面的な現象に対し、「使うかどうか(0 か 1 か)」と「使う場合にどれくらい使うか」を分離するアプローチをとった。学習者の年齢・性別・母語、および産出の長さを統計的に厳密に統制した結果、使用有無は習熟度とタスクの影響を強く受け、タスクによって習熟度効果の現れ方も異なることが明確になった。とくに、対話(I)では習熟度上昇に伴う使用確率の伸びが大きく、学習者が一定の日本語力に達すると「会話の流れの中でオノマトペを出せる」状態に移行しやすいことが示された。一方、使用者における使用量については、習熟度・タスク差は統計的に確認されるものの、モデル全体の説明力は低く留まった。これは、使用量が「習熟度」だけで直線的に決まるというより、オノマトペ産出における最大の障壁が「いかにして一回目の使用に踏み切るか」という入口の段階にあり、一旦その壁を越えた後、学習者個人の表現意図や局所的な文脈に強く依存していることを示唆している。

RQ3(ランダムフォレストに基づく使用の予測と重要変数の特定)では、機械学習を用いた非線形な予測モデルにより、オノマトペの使用有無は高い精度で予測可能であることが実証された。一方で、学習・接触経験に関する背景 5 変数を加えても、性能向上は小さかった。予測に寄与する重要変数を解析した結果、使用に強く関わるのは「語数」、次いで「習熟度」「タスク」「母語」であり、背景情報である訪日経験の有無がオノマトペ使用に対して微小ながらも正の方向(使用を促す方向)に作用しているという、線形モデルでは捉えきれなかった実践的な兆しも確認された。今後のオノマトペ教育的には、学習者の背景を細かく分類する以前に、(i)どのタスクで、(ii)どれだけ発話・作文の量を確保し、(iii)どの習熟度に対して、(iv)どの母語圏でとくに支援が必要か、という見立てが重要になるだろう。

以上、本研究では、統計モデリングと機械学習を併用することで、これまで学習者の「個人差」や「タスクの違い」として漠然と片付けられがちであったオノマトペ使用のメカニズムに関して、一定の精度を持つモデルを得ることができた。しかし一方で、本研究には改善すべき課題も残されている。ここでは主に 3 点に絞って言及したい。1 点目は、本研究の対象語彙は中核的な 200 語に限定されており、低頻度語や創作的用法まで含めたときに同様の傾向が維持されるかは追加検証の必要があることである。2 点目は、本研究では、機械学習モデルの安定性を担保するため、欠損の少ない「はい/いいえ」の二値的なアンケート項目を採用したことである。これにより、予測性能の増分が小さく留まったが、日本人との「接触の量(時間)」や「質の深さ」を測る連続的なデータがあれば、接触経験がオノマトペ産出に与える真の影響をより鮮明に描き出せる可能性がある。3 点目は、本研究は「使用したか/どれくらい使ったか」を中心に扱ったため、用法の適切さ(自然さ、文体との整合、誤用の種類)までは踏み込めていないことである。今後は、量的分析で得られた「出やすい条件/出にくい条件」を足がかりに、質的な用例分析を組み合わせることで、教育現場により直結する具体的提案に結び付けることを考えている。

謝辞

本研究は、JST 次世代研究者挑戦的研究プログラム JPMJSP2148 の支援を受けたものである。

研究インテグリティ宣誓

本稿著者は、本論文が著者自身によるオリジナルの研究成果であって、他者の著作物の剽窃や盗用、データの捏造や改竄、二重投稿、貢献度を反映しない著者名表記といった一切の不適切行為を行っていないことを宣誓する。なお、生成 AI に関しては一切使用していない。

引用文献

- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer.
- Molnar, C. (2025). *Interpretable machine learning: A guide for making black box models explainable* (3rd ed.). <https://christophm.github.io/interpretable-ml-book/>
- 小野正弘(2007) (編) 『擬音語・擬態語 4500: 日本語オノマトペ辞典』 小学館.
- 北原保雄(2003) (著) 『日本国語大辞典 第二版』 小学館.
- グエンティタイントゥイ(2018) 「中国人日本語学習者による日本語オノマトペの使用実態と産出傾向: ベトナム人日本語学習者との比較」 『一橋日本語教育研究』 6, 41–50.
- 張晶鑫(2020) 「現代日本語におけるオノマトペの用法解明と中国人日本語学習者のためのオノマトペ指導に対する提言 —コーパス言語学の教育的応用の可能性をめぐって—」 神戸大学国際文化学研究科博士論文.
- 前田未央(2009) 「日本語学習者による擬音語・擬態語の意味推測に影響する要因」 『ニダバ』 38, 128-137.
- 山田忠雄・柴田武・酒井憲二・倉持保男・山田明雄・上野善道・井島正博・笹原宏之(2011) (編) 『新明解国語辞典 第七版』 三省堂.
- 廉沢奇(2023) 「学習者コーパス調査をふまえた日本語学習者の ABAB 型基本オノマトペの使用実態の解明」 『統計数理研究所共同研究レポート』 465, 118–134.
- 廉沢奇(2024) 「パタンマッチングによるオノマトペ候補語抽出の試み—オノマトペ形態変換プログラムを用いて—」 『言語資源ワークショップ2024発表論文集』 255–267.