

L2 日本語学習者によるエッセイ評価：語彙的多様性と文法的複雑性に焦点を置いて

小畑文佳¹ 田川拓海¹ 小野雄一²

¹筑波大学大学院 ²人文社会ビジネス科学学術院人文社会科学研究群人文学学位プログラム

³ s2320036@u.tsukuba.ac.jp tagawa.takumi.kp@u.tsukuba.ac.jp

ono.yuichi.ga@u.tsukuba.ac.jp

概要

本研究は、日本語学習者のエッセイ評価において、言語的特徴量に基づく新しい指標の有効性について検討を試みたものである。従来、日本語のエッセイにおける言語的特徴量の研究は、主に語彙的多様性に関するものが多かったが、日本語学習者を対象とした研究は十分ではなかった。本研究では、日本語解析器 KWJA を使用して、用言及び体言の句単位の指標を分析し、これらが構造的複雑性の一側面を示すことを示した。特に、1文あたりの用言の割合 (Yougen_rate) が日本語学習者の言語習熟度と強く関連していることが、線形回帰モデル、決定木分析、ランダムフォレストモデルによって示された。この結果は、用言の使用頻度が日本語学習者の言語能力、特に複雑な文構造の生成能力と密接に関連していることを明らかにしている。

1 はじめに

本研究は、日本語学習者のエッセイ評価応用を念頭に置いた、言語的特徴量に基づく新しい指標の提案・検討を目的としている。従来の研究では、日本語エッセイにおける言語的特徴量については語彙的多様性が主に考察されてきたことが分かっている。[1]は、この語彙的多様性を、内容語と機能語の観点から分析し、それが ChatGPT による採点スコアよりもエッセイの習熟度をより適切に予測しているという結論を得た。

しかしながら、言語的特徴量を扱う研究は主に英語の文脈で行われており日本語学習者を対象とした研究はまだ十分ではない。本研究では日本語学習者のエッセイ評価に応用可能な言語的特徴量に焦点を当て、日本語解析器 KWJA を用いて、用言及び体言の句単位の指標を分析する。この指標は、構造的複雑

性の一側面を表すものと考えられるため、文中での出現割合や用言及び体言の句の数など文構造に関与する指標に基づいて、言語習熟度への寄与度を分析する。

2 先行研究

従来外国語学習の研究分野では、主に英語学習において言語的特徴量が、一方、日本語学習については、エッセイ評価を念頭に置いた上で言語的特徴量を取り扱ったものがあるものの、英語学習の文脈の研究ほど先行研究の数はおおくない。しかし、中には語彙多様性を言語習熟度の重要な指標として位置づけた研究がいくつかある。[2]は指標の記述量と語彙的多様性が日本語エッセイの習熟度と正の相関を持つことを明らかにしている。

また、英語学習の文脈では、CAF (Complexity, Accuracy, Fluency) に基づく分析が一般的に用いられている[3][4]。[5]は、ChatGPT によるプロンプト採点がある程度人間の評価に準拠していること、しかし CAF を示す指標となりうる言語的特徴量 (Lexical measures, Syntactic complexity measures, Fine-grained syntactic dependency, Verb argument construction, Cohesion measures) を評価項目に加えると性能が上がることを主張している。すなわち、これらの特徴量は習熟度の予測において重要なファクターであるといえる。

しかし、日本語学習者を対象とした研究は、これらの指標を十分に活用していない。具体的には、日本語学習者の言語的特徴量に基づく習熟度予測に関する研究はまだ十分ではない。そこで、本研究では日本語解析器 KWJA (Kyoto-Waseda Japanese Analyzer) を活用し、用言及び体言の句単位の指標を用いることで、語彙的多様性の他に構造的複雑性に寄与する言語指標がどの程度学習者の習熟度を予測できるの

かについて分析を行う。

3 分析

分析対象としては、I-JAS (International Corpus of Japanese as a Second Language) に収録されているエッセイを使用した。I-JAS は国立国語研究所がデータ収集を行ったコーパスであり、日本語学習者 1,000 人による日本語のエッセイと発話を調査したものである。本稿では中国語母語話者によるデータを扱った。

習熟度のデータとしては、J-CAT (Japanese Computerized Adaptive Test) の (聴解分野を除く) スコアの合計を参照した。J-CAT は、一般社団法人日本語教育支援協会が運営する、非日本語母語話者に向けた日本語能力試験である。

学習者のデータは線形回帰モデル、決定木分析、およびランダムフォレストモデルを使用して線形回帰モデリング及び変数の寄与 (重要度) を分析した。この章では、それぞれのモデルによる結果を詳述する。

3.1 言語指標の評価

本研究で使用した指標は、[1]で利用した Proficiency (学習者の習熟度), GuiraudFunction (機能語多様性), GuiraudContent (内容語多様性), KanjiRate(漢字使用)に加え, Yougen(用言数), Taigen(体言数), Yougen_rate(1 文中の用言数), Taigen_rate(1 文中の体言数), Sentence (文数)も用いた。それぞれの指標と習熟度の散布図を図 1 に示す。

次に、これらの指標を独立変数、習熟度を従属変数として、それぞれの指標の評価を決定木分析を用いて分類を行い、ランダムフォレストを用いて変数の重要度分析を行った。それぞれの結果を図 2, 図 3 に示す。図 2 の決定木分析では、Yougen_rate が上位のノードを占め、大きな分類役割を果たしている点認められる。また、図 3 のランダムフォレストから得られる平均ジニ係数の結果からは、語彙的多様性に加え、用言、体言に関わる指標、特に Yougen_rate の高い重要度が確認できる。

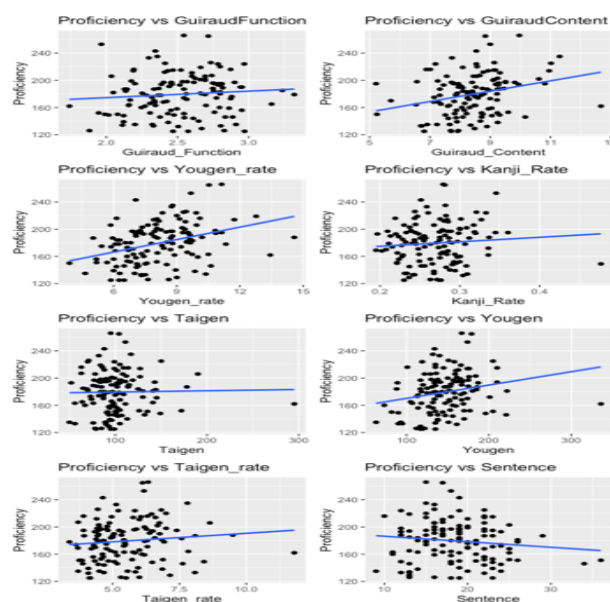


図 1 各指標と習熟度の相関

変数ごとの累積増加ノード純度の結果は以下の通りである: Taigen (10074.40), Yougen (13347.72), Sentence (10711.21), Taigen_rate (12765.05), Yougen_rate (19124.67), KanjiRate (12414.85), GuiraudContent (16487.11), GuiraudFunction (12126.76)。ここから、Yougen_rate がランダムフォレストモデルにおいて最も重要な予測変数であることが示唆される。

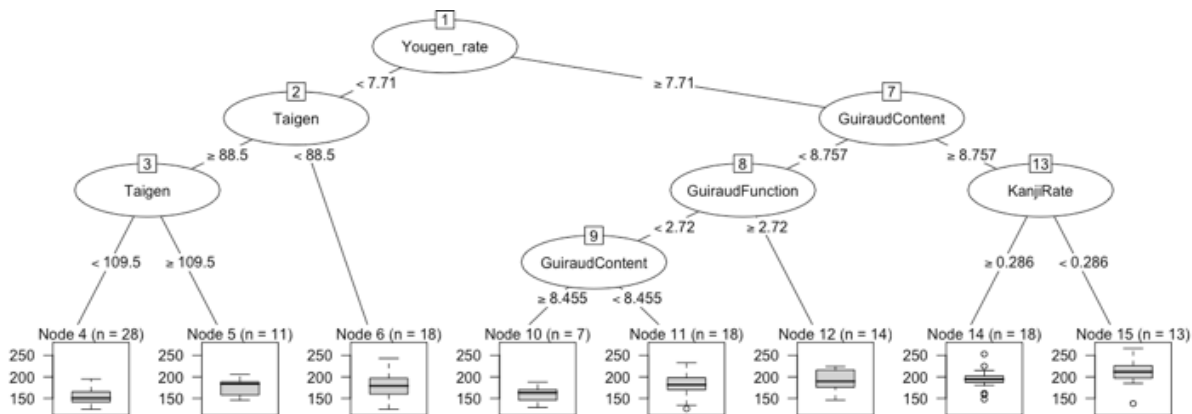


図 2 決定木分析の結果

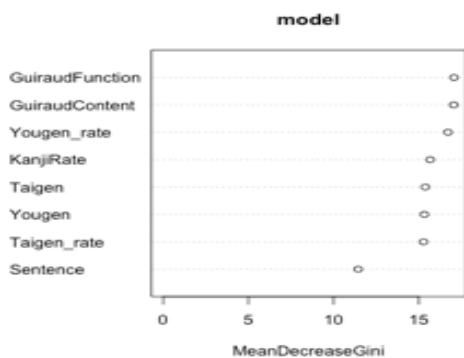


図 3 ランダムフォレストによるジニ係数

モデル 6 : $\text{lm}(\text{Proficiency} \sim \text{GuiraudFunction} + \text{GuiraudContent} + \text{Yougen_rate} + \text{KanjiRate} + \text{Taigen} + \text{Yougen})$

モデル 7 : $\text{lm}(\text{Proficiency} \sim \text{GuiraudFunction} + \text{GuiraudContent} + \text{Yougen_rate} + \text{KanjiRate} + \text{Taigen} + \text{Yougen} + \text{Taigen_rate})$

モデル 8 : $\text{lm}(\text{Proficiency} \sim \text{GuiraudFunction} + \text{GuiraudContent} + \text{Yougen_rate} + \text{KanjiRate} + \text{Taigen} + \text{Yougen} + \text{Taigen_rate} + \text{Sentence})$

3.2 線形回帰モデル

次に、線形回帰モデルによる分析から、言語達成度の予測因子を包括的に分析した。習熟度

(Proficiency) を従属変数とし、語彙的多様性や文中の用言・体言特徴量などの統語的指標を予測変数とした。F 値の比較により、習熟度の主要な予測因子を特定するため、モデル 1 からモデル 8 までの複数のモデルを構築した。

モデル 1 : $\text{lm}(\text{Proficiency} \sim \text{GuiraudFunction})$

モデル 2 : $\text{lm}(\text{Proficiency} \sim \text{GuiraudFunction} + \text{GuiraudContent})$

モデル 3 : $\text{lm}(\text{Proficiency} \sim \text{GuiraudFunction} + \text{GuiraudContent} + \text{Yougen_rate})$

モデル 4 : $\text{lm}(\text{Proficiency} \sim \text{GuiraudFunction} + \text{GuiraudContent} + \text{Yougen_rate} + \text{KanjiRate})$

モデル 5 : $\text{lm}(\text{Proficiency} \sim \text{GuiraudFunction} + \text{GuiraudContent} + \text{Yougen_rate} + \text{KanjiRate} + \text{Taigen})$

表 1 各モデルの F 値

モデル	F 値
モデル 3	1.228143
モデル 7	1.226504
モデル 4	1.223245
モデル 6	1.222564
モデル 5	1.219351
モデル 8	1.218937
モデル 2	1.146421
モデル 1	1.068799

表 1 に示すように、最も高い F 値を示したモデルはモデル 3 (Proficiency ~ GuiraudFunction + GuiraudContent + Yougen_rate) であった。

表2 モデル3の結果

	Estimate	Std. Error	t-value	p-value
(Intercept)	89.668	28.909	3.102	0.002**
GuiraudFunction	5.717	8.706	0.657	0.512
GuiraudContent	4.197	2.54	1.652	0.101
Yougen_rate	4.928	1.571	3.137	0.002**
Residual standard error	28.56			
Adjusted R-squared	0.1302			

表2の結果から1文あたりの用言の割合と習熟度との間に有意な関連性があることがわかった

(Yougen_rate, 推定値=4.928, $p < 0.01^{**}$) ている。調整済みR二乗値は0.1302で、モデル3は言語達成度の約13.02%の変動が説明できるとしている。以上の結果は、機能語の多様性、内容語の多様性、および1文あたりの用言使用の組み合わせがモデルの説明力に有意な寄与をしていることを示唆しており、特にYougen_rateは習熟度との正の相関を示している。

4 おわりに

本研究では、日本語学習者の習熟度予測において、言語的特徴量がどのように関連しているかを、線形回帰モデル、決定木分析、ランダムフォレストモデルを用いて分析した。

まず、今回の調査における最も重要な結論は、1文あたりの用言の割合 (Yougen_rate) が日本語学習者の言語習熟度と強く関連していることである。これは、線形回帰モデル、決定木分析、ランダムフォレストモデルの三つの異なる統計的手法によって一貫して示された。これにより、用言の使用頻度が日本語学習者の言語能力、特に複雑な文構造の生成能力と密接に関連していることが明らかになった。

しかし、本研究にはいくつかの課題も存在する。まず、研究のサンプルサイズが限定的であったことが挙げられる。また、研究に使用したデータは中国語母語話者の日本語学習者集団に限定されているため、結果の一般化には慎重である必要がある。将来的には、より広範なデータセットを用いて、さらに詳細な分析が求められる。

総括として、本研究は、日本語学習者の言語習熟度予測において、用言の使用頻度が重要な指標であ

ることを示した。この知見は、日本語教育や言語処理技術の発展に貢献するものと期待される。今後の研究においては、今回見られた課題を克服し、より包括的な理解を目指す。

謝辞

本研究はJSPS 科研費 21H00905 の助成を受けたものです。

参考文献

1. **Ayaka Obata, Takumi Tagawa, & Yuichi Ono.** : Assessment of ChatGPT’s Validity in Scoring Essays by Foreign Language Learners of Japanese and English. Proceedings of 15th International Congress on Advanced Applied Informatics, 2023 December 11-13, Bali, Indonesia.
2. **Hiroaki Hatano,** “An investigation of the linguistic features of B1 level of the JF Standard for Japanese Language Education: An analysis based on the writing data by Japanese language learners,” in Learner Corpus Stud. Asia World, vol. 3, pp. 189-206, Mar. 12, 2018.
3. **Alex Housen, Folkert Kuiken, & Ineke Vedder:** Complexity, accuracy, and fluency: Definitions, measurement, and research. In A. Housen, F. Kuiken, & I. Vedder (Eds.), Dimensions of L2 performance and proficiency: Complexity, accuracy, and fluency in SLA, pp.1-20, Amsterdam, The Netherlands: John Benjamins, 2012
4. **Kato Takeshi.** (2019). Constructing measurement models of L2 linguistic complexity: A structural equation modeling approach. JLTA journal, 22, p23-43., T. (2019). Constructing measurement models of L2 linguistic complexity: A structural equation modeling approach. JLTA journal, 22, pp.23-43, 2019
5. **Atsushi Mizumoto & Masaki Eguchi,**: Exploring the potential of using an AI language model for automated essay scoring, Res. Methods Appl. Linguist., vol. 2, no. 2, pp. 100050, 2023.