

日本語学習者作文評価システム「jWriter」の自動採点の精度

李在鎬（早稲田大学）、村田裕美子（ミュンヘン大学）、長谷部陽一郎（同志社大学）

Accuracy of automatic scoring of Japanese language learner's writing evaluation system "jWriter"

Jaeho LEE(Waseda University), Yumiko MURATA(LMU Munich), Yoichiro HASEBE(Doshisha University)

要旨：本発表では、日本語学習者の作文執筆を支援するための自動採点システム「jWriter」の測定精度を検証するために行ったデータ分析の結果を報告する。分析データは「住みやすい国コーパス」に収録されている120名の意見文である。分析では、120名の日本語学習者が受けた客観テスト「SPOT」の得点と「jWriter」の自動採点のスコアの相関を確認したところ、は「 $r=.688$ 」と高い相関が見られた。そして、コレスポンデンス分析においても SPOT のレベルと jWriter のレベルが明確に対応していることが明らかになり、「jWriter」は学習者の作文を十分な精度で測定していると結論づけられる。

キーワード：自動採点、作文評価、相関分析、コレスポンデンス分析、住みやすい国コーパス

1. 研究目的

私たちは科研費による補助のもと、2016年度より日本語学習者の文章執筆を支援するウェブシステム「jWriter」を開発している（図1）。

図1のテキストボックスに文章を貼り付け、「実行」ボタンをクリックするだけで、自動評価を行うと同時に、文章が持つ様々な情報を可視化する機能を持っている。とりわけ重要な機能として、習熟度（Proficiency）を自動採点する機能を持っているが、採点の精度については、十分な研究がなされていない状況であった（李ほか2019、小森ほか2022）。この問題を解決するため、本研究では、村田(2021)が構築した「住みやすい国コーパス」(<https://sumiyasui.jpn.org/>)のデータを利用し、「jWriter」の採点精度に関する統計的な分析を行った。



図1 jWriter の初期画面

2. 研究方法（データと方法）

本研究で使用した調査データは、「住みやすい国コーパス」である。このデータは、村田（2021）の方法で構築されたデータで、ヨーロッパの日本語学習者の意見文を収録した小規模コーパスである。データは、<https://sumiyasui.jpn.org/>からダウンロードできる。

2022年11月にバージョン1.0が公開されてから、徐々に規模が大きくなっているデータであるが、本研究ではドイツ話者（60名）、セルビア話者（20名）、クロアチア話者（20名）、日本語話者（20名）のデータを収録した「住みやすい国コーパス」バージョン1.0を使用した。このコーパスの特徴として、データ提供者の日本語能力を判断するための客観的な情報として、小林（2015）の「SPOT」の点数を提供している点が挙げられる。「SPOT」のスコアをもとに、日本語能力の内訳をみると、セルビア話者20名とクロアチア話者20名は全員が中級、ドイツ話者に関しては初級20名、中級20名、上級20名、日本語話者は未受験ということになっている。

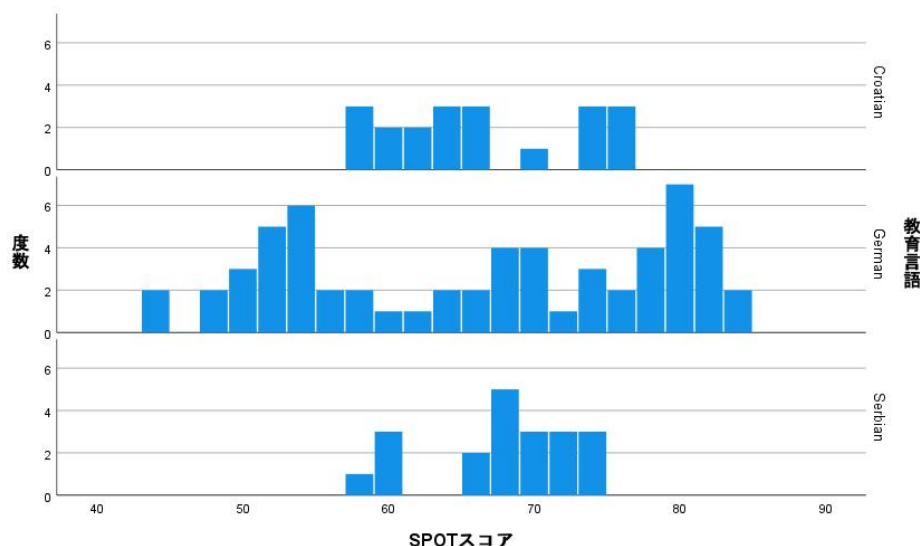


図2: 「住みやすい国コーパス」の話者別の SPOT のスコアのヒストグラム

本研究では、120名が書いた「住みやすい国の条件と理由」というタイトルの意見文を分析する。分析の手順は以下のとおりである。

- Step1: 120名の意見文を「jWriter」で解析し、作文評価値を算出
- Step2: 「jWriter」の作文評価値と「SPOT」のスコアを相関分析する
- Step3: 「jWriter」の作文評価値にもとに算出した習熟度と、「SPOT」のスコアをもとに算出した習熟度をコレスポンデンス分析する

なお、Step1の作文評価値は、1から4の間で連続量として表示される数値で、計算式は、Lee&hasebe(2020)に従って行っている。また、同じ計算式は <https://jreadability.net/jwriter/>にも実装されており、誰でも確認や利用ができる。

3 分析結果

Step1 の分析として「jWriter」の作文評価値を計算した。話者別の分布を確認したところ、図3の結果となった。

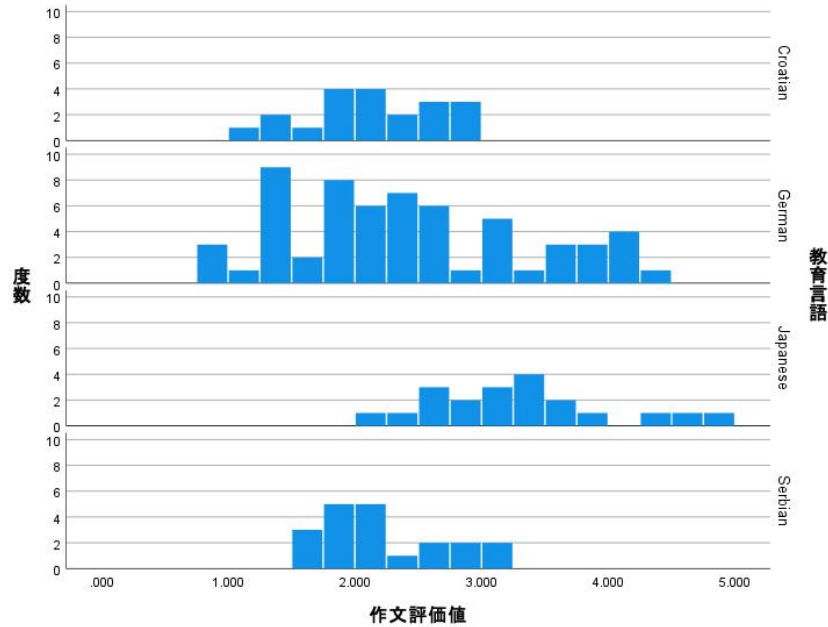


図3：話者別の作文評価値のヒストグラム

「jWriter」の作文評価値は、数値が高ければ高いほど、習熟度が高いことを意味する。1はもっともレベルが低く、5がもっとも高い。日本語教育領域の慣習に従えば、1は初級レベルに相当し、2は中級レベル、3は上級レベル、4は超級レベルに相当する。

次に Step2 の分析として、Pearson の積率相関係数を計算してみたところ、 $r=.688$ という高い値が確認できた。なお、母語別に比較するにはデータ量が少ないという問題があるが、クロアチア語話者の 20 名に関しては、 $r=.442$ 、セルビア語話者の 20 名に関しては $r=.250$ 、ドイツ語話者の 60 名に関しては、 $r=.704$ となった。

最後に Step3 として「SPOT」の習熟度レベル（初級～上級）と「jWriter」の習熟度レ

表 1: 「SPOT」の習熟度レベルと「jWriter」の習熟度レベルのクロス表

		jWriter の習熟度レベル			
		初級	中級	上級	超級
SPOT の 習熟度 レベル	初級	11	8	1	0
	中級	5	35	15	5
	上級	0	5	9	6
	未受験	0	2	12	6

ベル（初級～超級）の対応を確認するためクロス表分析（表1）とコレスポンデンス分析（図4）を行った。

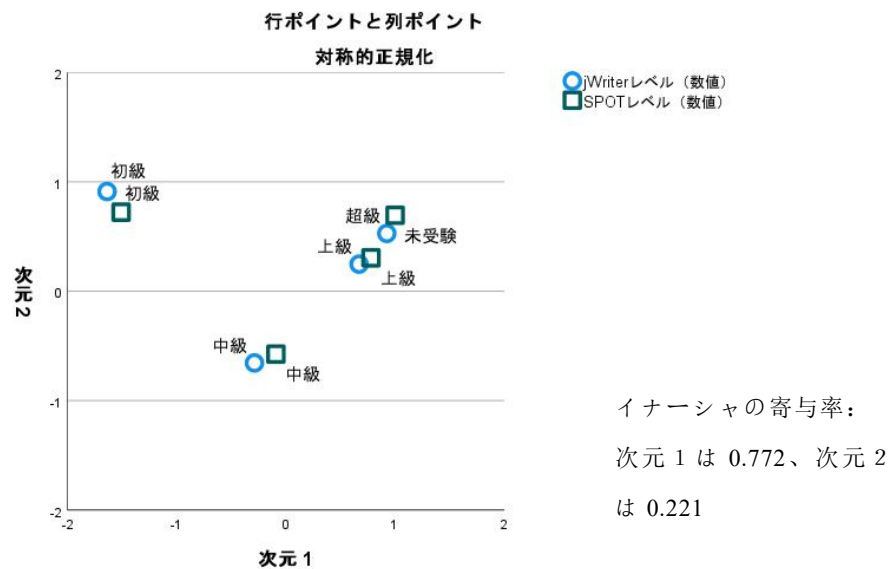


図4：コレスポンデンス分析の結果

「SPOT」の習熟度レベルで未受験となっているのは、日本語話者のことである。以上の表1および図4の結果から「SPOT」の習熟度レベルと「jWriter」の習熟度レベルに対応関係が明らかになった。

参考文献

- 小林典子（2015）「SPOT」李在鎬（編）『日本語教育のための言語テストガイドブック』110-126. 東京：くろしお出版.
- 小森和子・伊集院郁子・李在鎬（2022）「日本語学習者の作文における自動評価と教師評価の比較」『明治大学国際日本学研究』14(1)、41-68.
- 村田裕美子（2021）「小規模コーパスの構築方法」李在鎬（編）『データ科学×日本語教育』34-53. 東京：ひつじ書房.
- Lee, J., and Hasebe, Y. 2020. Quantitative Analysis of JFL Learners' Writing Abilities and the Development of a Computational System to Estimate Writing Proficiency, *Learner Corpus Studies in Asia and the World* 5, 105-120.
- 李在鎬・長谷部陽一郎・村田裕美子（2019）「学習者作文の習熟度に関する自動判定とWebシステムの開発について」『ICT×日本語教育』38-53. 東京：ひつじ書房.

コーパス

「住みやすい国コーパス」<https://sumiyasui.jpn.org/>