

CEFR 読解指標（CDS）に基づく作題例の妥当性に関わる特性抽出の試み
谷誠司（常葉大学）、宮崎佳典（静岡大学）、安志英（群山大学校）、元裕璟（高麗大学校
大学院博士課程）

**An attempt to extract characteristics regarding the validity of sample
problems based on the CEFR Reading Comprehension Index (CDS)**

Seiji TANI, Tokoha University Yoshinori MIYAZAKI, Shizuoka University Jiyoung AN,
Kunsan National University Yukyoung WON, Korea University

要旨：本研究チームで開発された「日本語 CEFR 読解指標（CDS）推定支援 Web アプリケーション」によって選定された適合度の高い例文をもとに日本語教育経験者がテスト問題を作題し、予備テストとして実施した。予備テスト実施後に作題者による事後コメントを分析した結果、Web アプリケーションで得られた高適合度の例文が必ずしもテスト問題として適切とは限らない事実が示唆された。そこで、CDS と例文の一致度や作問可能性について調査を実施し、その結果を報告する。

キーワード：CEFR、読解、例文適合度調査

1. はじめに

CEFR（Common European Framework of Reference for Languages）は言語能力を A1～C2 レベルの 6 段階に設定し、各レベルには何ができるかを示す複数の能力記述文（Can-Do Statements、以下 CDS）で記述されている。CEFR の CDS を読むと、抽象的な表現を含むことが多く、実利用とのイメージを結びつけるのが容易でない場合が見受けられる。故に、現場では CDS に基づいて読解のテスト問題や教材を準備する際、CDS の抽象的な表現を主観的に解釈して例文を選択／作成するのが一般的であると思われる。

本研究チームでは（DIALANG の）CEFR 読解 CDS に対応した日本語例文（約 1000 例文）を収集し、機械学習の手法を用いることで、専門性、文長、文書タイプなどの特徴から入力例文に対応する CDS を推測・分類する研究ならびに実験を行った（宮崎・平川・高田・谷、2019）。また、同分類手法を実装した「日本語 CEFR 読解指標（CDS）推定支援 Web アプリケーション」（以下、Web アプリケーション）も開発されている（宮崎・Vuong・谷・安・元、2020）（図 1）。なお、DIALANG とは欧州 14 言語を対象とした、学習者の言語運用力を CEFR のレベルに基づいて診断する無料オンラインシステムである。このシステムにある自己評価用の CEFR-DIALANG self-assessment statements という CDS は、CEFR の CDS より個別的・具体的であるため、本研究チームではこの CDS を使用することにした。

江崎経済危機：北方担当が27日辞任したことを受け、政府は沖縄振興を推進する態勢を立て直しを求めている。米軍普天間飛行場（沖縄県宜野湾市）の名護市辺野古への移設は、事前調査で明らかになった。米軍側は「沖縄・辺野古問題」を推進している。沖縄・北方担当をめぐり争い、上野と対峙した」と語った。福井氏は「沖縄・北方担当をめぐり争い、上野と対峙した」と語った。福井氏は「沖縄・北方担当をめぐり争い、上野と対峙した」と語った。

予測CDS（推定アルゴリズム:SVMを使用）					
Array					
CDS番号	CEFRレベル	CDS	CDS番号	CEFRレベル	CDS
1	B2	例文 さらに詳細に読む必要があるかどうかを決定するために、広範囲にわたる専門的な話題についてのニュース、記事、レポートの内容と関連性をすばやく確認することができる。	18	A1	
2	B1	例文 長いテキストや複数の短いテキストをざっと目を通して、課題を遂行するために必要な情報を探すことができる。	19	B1	
3	B1	例文 手紙、パンフレット、短い公的な文書といった、日常的な文章において必要とする一般的な情報を見つけて、理解することができる。	20	A2	例文 公共場
4	A2	例文 通り、レストラン、駅のような公共の場所や職場にある標識や指示を理解することができる。	21	A1	
5	A2	例文 広告、パンフレット、メニュー、時刻表などの簡単な言語資料の中の特定の情報を見つけることができる。	22	B2	例文 読む目
6	A1	例文 もっとも一般的で日常的な状況でよく出くわす、簡単な指示にでているような、なじみのある名前、単語、または非常に簡単な句を確認することができる。	23	B1	例文 自分
7	B2	例文 筆者が特別な立場や現況から取り上げた、現代の問題に関する記事やレポートを理解できる。	24	A2	例文 自分
8	B2	例文 専門用語を確認するために辞書が使えるのであれば、自分の専門以外の専門的な記事を理解することができる。	25	A2	
9	B1	例文 文章における議論の大まかな流れを確認することができるが、必ずしも詳細に認識できるわけではない。	26	A2	例文 もっとも
10	B1	例文 はっきりと主張が書かれたテキストの主要な結論を把握できる。	27	A1	例文 身近な
11	B1	例文 なじみのある話題に関する簡単な新聞記事において重要な点を確認することができる。	32	B2	例文 「短

図1 Webアプリケーションの「推定結果」のページ（部分的に拡大表示）

谷・宮崎・安・元（2020）では（DIALANG の）CEFR 読解 CDS の中から A1～B2 レベルに属し、「必要な情報を探し出す」と「手紙やメールを読む」の下位尺度に該当する CDS（それぞれ 5 個と 6 個、計 11 個）に限定して Web アプリケーションから CDS との適合度が最も高い例文を 1 つ選択した。これらの例文に対して設問を加える形で日本語読解テストの作題を日本語教育経験者に依頼し、予備テストとして実施した。予備テスト後、作題者から CDS と例文レベルの不一致や、CDS で記述された読み方ができない例文があったなどの事後コメントが確認され、Web アプリケーションで得られた高適合度の例文がテスト問題として適切とは限らない事実が示唆された。

本研究チームでは Web アプリケーションの開発に加え、CEFR 読解 CDS からの問題作成支援の 1 つとして、CDS ごとのサンプル問題の提供を考えている。サンプル問題に相応しい例文を見つけ出し、またその特徴を明らかにするために Web アプリケーションが出力する適合度の高い例文に対して調査を実施した。

2. 調査の概要

調査協力者は韓国における日本語教育経験者 4 名であり、調査は 2020 年 10 月～12 月に行った。調査方法はまず谷他（2020）で使用した CDS に限定して Web アプリケーションから CDS との適合度の上位 5 例文を選択した。（11 個の CDS×5 例文 = 55 例文）、次に調査協力者に 55 例文を見せ、a) 例文と対応する CDS の一致度（以下「一致度」、b) CDS で記述されている「(読解) 行動」を問える問題の作問可能度（以下「作問可能度」、の 2 点から 5 段階評定と評定理由のコメントを依頼した。

3. 調査の結果

3.1 全体的な調査結果

本節では、調査結果の全体的な傾向を報告する。まず本調査では 4 名の評定者がいるの

で、評定者間信頼性係数（クロンバック α ）を調べた。結果は「一致度」においては「必要な情報を探し出す（30 例文）」では $\alpha = .27$ 、「手紙やメールを読む（25 例文）」では $\alpha = .75$ となり、「作問可能度」では「必要な情報を探し出す（30 例文）」では $\alpha = .76$ 、「手紙やメールを読む（25 例文）」では $\alpha = .70$ となった。

次に調査対象の 55 例文は CDS の適合度が高い上位 5 つの例文であり、適合度を推定する値（0%～100%で表示。アプリケーション上では「信頼度」と表示。以下「Web アプリケーションによる適合度」）が算出されている。この「Web アプリケーションによる適合度」と評定者 4 名の「一致度」の評定点の合計点（5 点満点×4 名=20 点満点）との相関係数（スピアマンの順位相関係数）を確認した。その結果、「必要な情報を探し出す（30 例文）」では $r = .55$ ($p < .01$)、「手紙やメールを読む（25 例文）」では $r = .23$ (n. s.) となった。更に「Web アプリケーションによる適合度」と評定者 4 名の「作問可能度」の評定点の合計点（5 点満点×4 名=20 点満点）との相関係数（スピアマンの順位相関係数）を確認したところ、「必要な情報を探し出す（30 例文）」では $r = .57$ ($p < .01$)、「手紙やメールを読む（25 例文）」では $r = .09$ (n. s.) となった。

また、CDS ごとの特徴を見るために、1 つの CDS に対応している 5 つの例文に関する 3 つの数値（「Web アプリケーションによる適合度」、「一致度」の判定得点の合計点、「作問可能度」の判定得点の合計点）の平均値を出した。（表 1 & 表 2）

表 1：「必要な情報を探し出す」にある CDS（6 個）の結果

CDS 番号 (CEFR レベル)	CDS	Web アプリケーション	判定者（4 名）	
		適合度の平均値	＜一致度＞の合計点の平均値	＜作問可能度＞の合計点の平均値
1 (B2)	さらに詳細に読む必要があるかどうかを決定するために、広範囲にわたる専門的な話題についてのニュース、記事、レポートの内容と関連性をすばやく確認することができる。	46.12%	13.00	10.80
2 (B1)	長いテキストや複数の短いテキストをざっと目を通して、課題を遂行するために必要な情報を探することができる。	55.47%	17.00	16.80
3 (B1)	手紙、パンフレット、短い公的な文書といった、日常的な文章において必要とする一般的な情報を見つけて、理解することができる。	57.95%	18.00	18.60
4 (A2)	通り、レストラン、駅のような公共の場所や職場にある標識や掲示を理解することができる。	47.05%	16.20	17.20
5 (A2)	広告、パンフレット、メニュー、時刻表などの簡単な言語資料の中の特定の情報を見つけることができる。	71.30%	17.20	18.20
6 (A1)	もっとも一般的で日常的な状況でよく出くわす、簡単な掲示にでている	63.43%	17.20	17.80

	ような、なじみのある名前、単語、または非常に簡単な句を認識することができる。			
平均		56.89%	16.43	16.57

表 1 は「必要な情報を探し出す」にある 6 個の CDS についての結果である。1(B2)はすべてにおいて平均値より数値が低い。4(A2)は「Web アプリケーションによる適合度」と判定者による「一致度」が低く、「作問可能度」が高い。一方、2(B1)・3(B1)・5(A2)・6(A1)は、2(B1)の「Web アプリケーションによる適合度」以外は平均値より数値が高い。

表 2：「手紙やメールを読む」にある CDS（5 個）の結果

CDS 番号 (CEFR レベル)	CDS	Web アプリケーション	判定者（4 名）	
		適合度の平均値	＜一致度＞の合計点の平均値	＜作問可能度＞の合計点の平均値
14(B2)	自分の専門分野に関連する通信文（手紙・メールなど）を読んで、楽に必要な意味が把握できる。	80.83%	15.20	15.80
15(B1)	個人の手紙を読んで、出来事、感情、希望の表現を理解することができる。友達や知り合いと文通できる。	90.48%	17.40	18.00
16(A2)	身近な話題について日常の定型型の手紙やファックスを理解することができる。	60.80%	15.00	17.20
17(A2)	短く、簡単な個人的な手紙を理解することができる。	71.62%	17.20	18.80
18(A1)	葉書などに書かれた、短く簡単なメッセージを理解することができる。	85.93%	14.20	16.60
平均		77.93%	15.80	17.28

表 2 は「手紙やメールを読む」にある 5 個の CDS 対しての結果である。14(B2)と 18(A1)は「Web アプリケーションによる適合度」が高いが、判定者による「一致度」と「作問可能度」が低い。16(A2)はすべてにおいて数値が低い、「一致度」と「作問可能度」は平均値に近い。一方、15(B1)と 17(A2)は、17(A2)の「Web アプリケーションによる適合度」が低いものの、「一致度」と「作問可能度」の数値はともに高い。

3.2 個別的な調査結果

本節では、前節で判定者による「一致度」と「作問可能度」の数値が低い CDS（1(B2)、4(A2)、14(B2)、18(A1)）を中心に個々の例文を見ていく。図 2 と図 3 は、「必要な情報を探し出す」にある 30 例文と「手紙やメールを読む」にある 25 例文の「一致度」と「作問可能度」の各判定得点（5 点満点）を平均値にし、散布図にしたものである。（紙面の関係上、縦軸及び横軸ともに 0 から始まっていない。）1 つの CDS に 5 例文あるので、

例文は CDS 番号(CEFR レベル)例文識別記号(a～e)で表示されている。

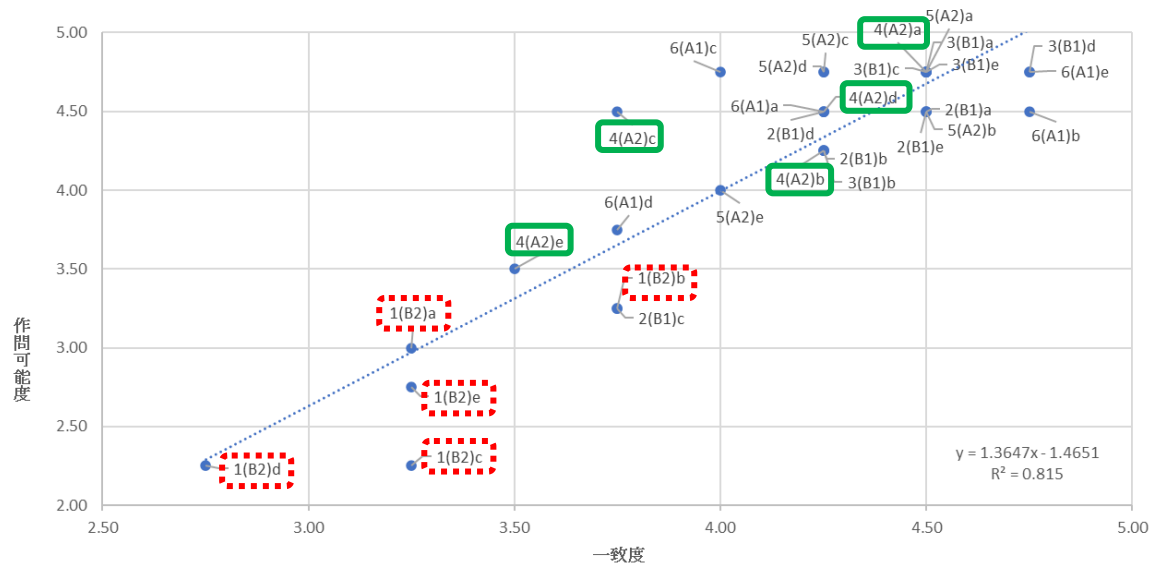


図 2 : 「必要な情報を探し出す」にある 30 例文の一致度と作問可能性の散布図

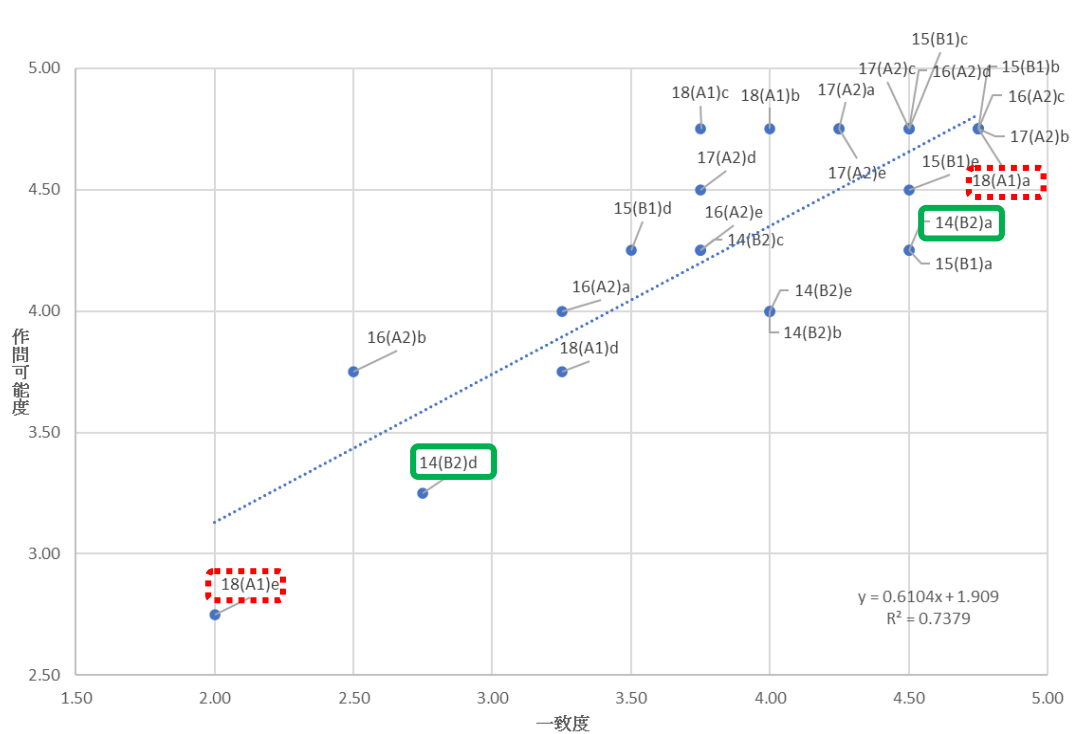


図 3 : 「手紙やメールを読む」にある 25 例文の一致度と作問可能性の散布図

図 2 を見ると、まず 1(B2)の例文（赤点線の枠）が他の CDS と比較して、「一致度」と「作問可能性」の数値が低いことが分かる。作題者の事後コメントでは b の例文は長すぎ、c、d、e の例文は短すぎるとあった。a～e の例文の文字数を数えると、a は 1061、b は

7883、c は 186、d は 369、e は 204 であり、文字数にかなりの違いがある。a は毎日新聞の社説であり沖縄県・米軍普天間飛行場の名護市辺野古への移設計画に関するもので専門性の高い話題であるとともに文章量や漢字が多い点、また b は大人の引きこもりに関するネット上の文章であり文章量がとびぬけて多い点から、CDS にある「すばやく確認することは難しい」という事後コメントもあった。c と d は文章の長さによる作題のしにくさが指摘された。4(A2) の例文（緑の枠）においては e の例文（駅の自動改札機の手前の床に貼られた注意書き「のりかえはみどり色の改札機へ」）はテキストが短すぎる（e の文字数は 14 であり、a は 124、b は 43、c は 56、d は 94）、自動改札機の写真と一緒に提示する必要があると指摘された。

図 3 を見ると、18(A1) では a と e の例文間で評定が大きく異なっている点が目立つ（赤点線の枠）。a は「あけましておめでとうございます」、e は「あなたの心がいつも晴れていますように」であり、事後コメントでは e は「慣用的な表現」「比喩的表現」「A1 レベルには難しい」といった意見が出され、「一致度」が低く評定がされ、「作問可能性」についても A1 レベルには難しいとの指摘がされた。14(B2)では a と d の評定が異なる（緑の枠）。評定が低い d はロードセルという荷重（力）を電気信号に変換する荷重変換器の販売方法の変更に関するビジネス文書である。精密機器に関する文書のため専門用語も多く、レベルが高いとのコメントが見られた。一方、a は研究会の案内メールであり、評定も最も高い。14(B2)の CDS は「自分の専門分野に関連する通信文（手紙・メールなど）を読んで、楽に必要な意味が把握できる。」とあり、専門分野と合わせての出題は困難ではないかとのコメントもあった。

本調査では、1)「Web アプリケーションによる適合度」と評定者 4 名による評定点の相関関係は、「必要な情報を探し出す（30 例文）」はやや相関あり、「手紙やメールを読む（25 例文）」は弱い相関あるいはほとんど相関なし、2) B2 レベルの CDS の例文が評定者による評定点が低い傾向、3)「Web アプリケーションによる適合度」が同程度に高い例文であっても、評定者の評価が低かったり評定者間の評定が割れたりする例文あり、4) 例文ごとの文字数の違いや例文の専門性の扱いにも注意が必要、が明らかになった。

参考文献

- 谷誠司, 宮崎佳典, 安志英, 元裕璟. 2020. 「CEFR 読解指標に基づく日本語能力テスト開発の試み」, 『常葉大学大学院国際言語文化研究科研究紀要』創刊号 43-54.
- 宮崎佳典, 平川遼汰, 高田宏輝, 谷誠司. 2019. 「CEFR 読解指標に基づく日本語例文分類手法」, 當作靖彦(監修), 李在鎬(編)『ICT×日本語教育 ―情報通信技術を利用した日本語教育の理論と実践』54-67. 東京: ひつじ書房.
- 宮崎佳典, Vuong Hong Duc, 谷誠司, 安志英, 元裕璟. 2020. 「CEFR Companion Volume に対応した日本語例文自動分類手法」, 『日本學報』第 125 輯 153-175. 韓国日本学会.