



英語読解教材のCEFRレベルの推定 : CVLAの妥当性評価

内田, 諭
根岸, 雅史

(Citation)

Journal of Corpus-based Lexicology Studies, 3:1-14

(Issue Date)

2021-02-28

(Resource Type)

departmental bulletin paper

(Version)

Version of Record

(JaLCD0I)

<https://doi.org/10.24546/81012537>

(URL)

<https://hdl.handle.net/20.500.14094/81012537>



英語読解教材の CEFR レベルの推定

—CVLA の妥当性評価—

内田 諭(九州大学)・根岸雅史(東京外国語大学)
uchida@flc.kyushu-u.ac.jp, negishi@tufs.ac.jp

Estimating the CEFR Levels of English Reading Materials

—Evaluation of CVLA—

UCHIDA, Satoru (Kyushu University),
NEGISHI, Masashi (Tokyo University of Foreign Studies)

概要

本論文は、読解教材等の CEFR レベルを判定するアプリケーションである CVLA(CEFR-based Vocabulary Level Analyzer)の妥当性について検証する。CEFR レベルが明示されている 2 種類のデータ(Council of Europe のテストサンプルとケンブリッジ英検の過去問)を用いて検証したところ、6 値分類(A1, A2, B1, B2, C1, C2)で約 53%の一致率であった。また、推定値が大きく外れていないことを確かめるために隣接レベルも含めて一致率を計算したところ、約 81%という結果となった。この結果は、CVLA による推定値が比較的安定していることを示しており、検証データの平均値である 400 語程度の入力があれば、ある程度の精度で CEFR レベルが判定できることが明らかとなった。

キーワード

CEFR, レベル推定, 読解教材, CVLA

1. はじめに

外国語の習得において自分のレベルにあった入力を得ることは効率的に学習を進める上で非常に重要である。Krashen (1985)のインプット仮説では、「理解可能なインプット」(comprehensible input)の重要性を説いており、学習者にとって最適な入力は自身の現在の学習水準よりもわずかに高いものが最適であるとしている。学習者の現在の水準が i とすると、入力は「 $i+1$ 」のものが適切であると比喩的に述べている。学習のために重要な「理解可能なインプット」を得るためには、学習者のレベルを適切に把握することに加えて、入力となる素材のレベルを知ることが重要である。教師の立場から見ると、自分が受け持っている学習者のレベルを知り、そのレベルよりも少し高いレベルの教材を用いることが肝要であると言える。無論、外国語の習熟度を上げるためにはインプットのみならず、適切な分量のアウトプットやインタラクションも必要となるが(白井 2012), 教材のレベルを把握するということは、学習者にとっても教師にとっても欠かすことはでき

ない重要な側面である。

本論文では、英語の読解教材等を対象として CEFR (Common European Framework of Reference for Languages) レベルを推定するプログラムである CVLA (CEFR-based Vocabulary Level Analyzer)¹ について、その推定値の妥当性を検証する。CVLA は、入力された英文から算出した 4 つの統計量を、CEFR レベル別に分類された英語教材のコーパスで測定した基準値と比較して、入力文の CEFR-J レベル(日本の英語教育の文脈に合わせて CEFR レベルを細分化したもの)を推定するものである(Uchida & Negishi, 2018)。以下ではレベル推定の先行研究として Lexile Measure と Text Inspector について概観した後、CVLA の概要について示し、CEFR レベルが明示されている 2 種類の英語の文章データを用いて CVLA の推定値の妥当性を測定する。

2. 先行研究

テキストレベルの推定には伝統的にリーダビリティが使用されてきた。例えば、中條・長谷川(2004)では、大学入試センター試験および 40 の大学の 2 次試験について Flesch-Kincaid Formula, FORCAST Formula など複数のリーダビリティ指標を用いて検証し、センター試験は高校生にとって適切なレベル設定がなされている一方、大学個別の入試問題については大学によって差があり、一部の大学の英文は受験生にとって難易度が高すぎる傾向にあることを指摘した。しかしながら、リーダビリティは、指標の計算がシラブル数、文長など比較的単純な指標を用いており、真に英語学習者に適切な数値となっているかどうかは判断が難しい。本節では、リーダビリティよりも一歩進んだ指標を提供する 2 つのアプリケーション (Lexile Measure と Text Inspector) について概観し、それぞれのアプリケーションを使った研究事例を紹介する。

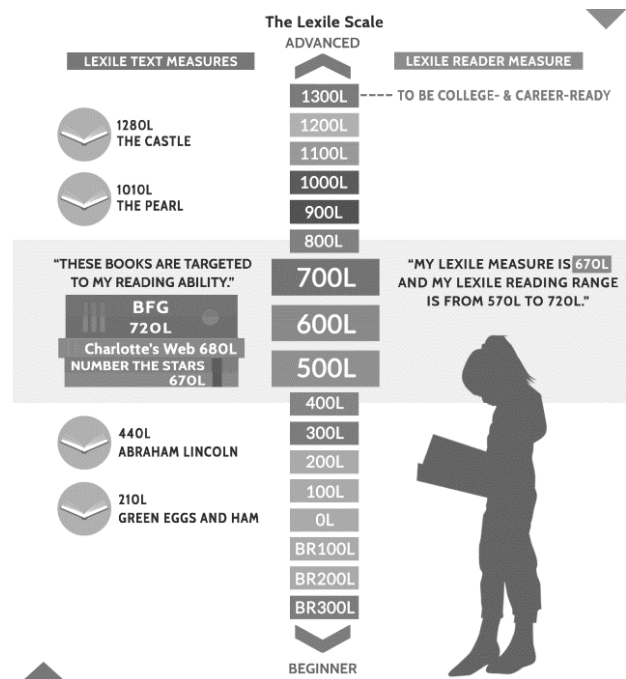
2.1 Lexile Measure

Lexile Measure は MetaMetrics 社が開発した指標で、主に母語話者を対象として「自分のレベルにあった書籍を見つける」という目的で開発されたものである。この指標は、Semantic Difficulty の指標である Mean Log Word Frequency (対数化した単語頻度の平均) と Syntactic Complexity の指標である Mean Sentence Length (文の長さの平均) の 2 つの値から算出される(Lennon & Burdick, 2004/2014)。詳細な数式は明示されていないが、Lexile Analyzer を用いることで手持ちのテキストのレベルを算出することができる。また、この指標には Lexile Reader Measure (LRM) と Lexile Text Measure (LTM) があり(図 1)、利用者は自身の指標 (LRM) に合わせて書籍を探す (LTM) ことが想定されており、LTM と LRM の差が -100L~50L あたりが最適なテキストであるとされている。

根岸(2015)は、Lexile Measure を使って高校英語教科書および入試問題の分析を実施した。その結果、学年進行とともに Lexile Measure の値が高くなることが明らかとなり、さらに中学 3 年と高校 1 年では他の学年と比較して大きなギャップ (約 300L) があることを

指摘している。大田(2016)も Lexile Measure を使って中高の英語教科書を比較し、同様の結論を得ている。Smith and Turner (2018)は、CEFR レベルが付与された Graded Reader を検証し、C2 を除くそれぞれのレベルについておおよその Lexile Measure を推定し、A1 (230L-340L), A2 (425L-715L), B1 (588L-860L), B2 (598L-993L), C1 (760L-1200L)であったことを報告している。

図 1
Lexile Scale のイメージ (Lexile ウェブサイトより)



2.2 Text Inspector

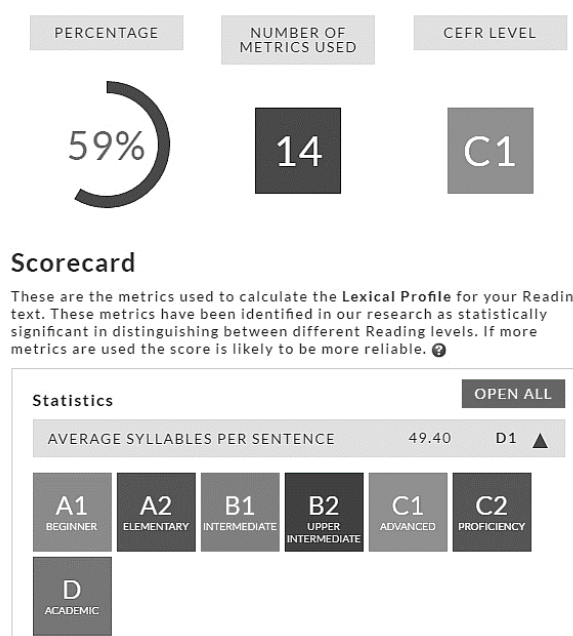
Text Inspector はオンラインで利用できるテキスト分析ツールがある(Bax, 2012)。このツールでは、入力された英文に対して、文数、語数、文長、Type/Token Ratio、音節数、Flesch Reading Ease、Flesch-Kincaid Grade などの様々な指標に加えて、British National Corpus や Corpus of Contemporary American English における単語の頻度ランクごとのグラフなどが表示される。また、Scorecard というセクション(図 2)では推定される CEFR レベルも表示される(ただし有料版のみの機能)。Text Inspector の CEFR スコアの推定方法の詳細は示されていないが、ウェブサイトの Help ページによると、Lexical Diversity、English Vocabulary Profile での語彙レベル、Type/token Ratio (TTR)、Average Sentence Length などが関連する指標として示されており、これらを使って推定されていると考えられる。また、独自の D レベル (Academic) も設けているという点も特徴的である。さらに、入力された英文で用いられている metadiscourse marker の分析が可能で、attitude marker (admittedly, amazingly など)、hedge (almost, apparently など)、booster (always, certainly

など)の頻度を検証することができる。

Bax, Nakatsuhara, and Waller (2019)は *Text Inspector* を使って比較的習熟度の高い学習者(B2, C1, C2 レベル)のライティングの分析を実施している。その結果, 上位レベルの学習者のほうが *metadiscourse marker* を使う頻度が下がるが, より幅広い表現を使う傾向があるということなどが指摘されている。このように *Text Inspector* では学習者の産出した英文も分析の対象とできることが特徴の1つであると言える。

図 2

Text Inspector の Scorecard の例 (Text Inspector ウェブサイトより)



2.3 まとめ

Lexile Measure, *Text Inspector* は, 旧来テキストレベルの推定に多く利用されてきたリーダビリティよりも詳細な分析を提供し, 教材や学習者のライティングの分析にも応用可能である。しかしながら, Lexile 値の算出方法や *Text Inspector* の CEFR レベル判定に関するアルゴリズムは公開されておらず, 学術的な検証が十分に実施できないという問題点がある。また, 日本の教材分析等を実施する場合は, CEFR よりもより詳細なレベル設定が設けられている CEFR-J レベルとのリンクがある方が望ましい。以下では, 本論文のリサーチクエスチョン(RQ)を述べた後, これらの問題点を有さない CVLA について概要を説明し, その妥当性について検証を行う。

3. リサーチデザインと手法

3.1 研究目的と RQ

本論文の目的は Uchida and Negishi (2018)による CVLA の妥当性を評価することである。CVLA は後述するように明示的な 4 つの指標により入力文章の CEFR-J レベルを推定するものである。本稿では CVLA のリーディングモードを対象として、CVLA の推定レベルの妥当性について CEFR レベルが明記されている文章を用いて検証する。具体的には、以下の 2 つのリサーチクエスチョンを設定する。

- (1) CVLA は Council of Europe が提示する CEFR レベルを示すサンプルテキストのレベルを正しく推定できるか
- (2) CVLA は CEFR レベルが明示されている Cambridge English Exams (ケンブリッジ英検) のリーディング素材についてレベルを正しく推定できるか

3.2 データ

本論文で使用する 1 つ目のデータは、Council of Europe (COE) が reading illustrative tasks として示すものを利用する²。Aptis (British Council が実施する英語能力試験)、Cambridge English Language Assessment の素材など、英語については全部で 13 の文章がある(A1:1 件, A2:3 件, B1:3 件, B2:5 件, C1:1 件)。本データは、97 語～915 語のテキストで、平均は 466 語である。これらのサンプルは図 3 のような形(PDF ファイル)で公開されており、その本文の部分のみを検証の対象とする。テストの難易度はタスクによっても変わらうが、ここで示されている文章はそのレベルの学習者が読めることを想定しているものであるということから、当該レベルを代表すると考えることができる。

図 3

Aptis (B2)のサンプル例

COUNCIL OF EUROPE
CEFR Illustrative Tasks: Reading
COUNCIL OF EUROPE

Level: B2

Providing Institution: BRITISH COUNCIL Aptis

READING

Read the passage below quickly. Match a heading (A-H) to each paragraph (1-7). There are more headings than you need.

Burning Man

1. On the 21st June 1986 – the date of the summer solstice – Larry Harvey and Jerry James invited a few friends to Baker Beach in San Francisco to celebrate the longest day of the year. They marked the event by setting fire to a nine foot wooden model of a man and a dog. As they watched the flames consume the crudely made figures, none of them could have known how their small gathering would grow over the next two decades and develop into an annual seven day experience that has become known simply as Burning Man. In 2011, so many people wanted to be part of it that the number of participants had to be limited to 50,000.

2. Even the organisers of the Burning Man project find it difficult to describe the experience. 'Trying to explain what Burning Man is to someone who has never been to the event,' says their official website, 'is a bit like trying to explain what a particular colour looks like to someone who is blind.' Part celebration, part act of rebellion, part community theatre, the week long event – now held in the Black Rock Desert of Nevada – attracts tens of thousands of people. They still come to watch the ritual destruction of a wooden figure (now more than forty feet high), but most of all they

1 A. Birth of a tradition.

2

3

4

5

6

7

COE のデータに加えて、本稿では CVLA の検証にケンブリッジ英検の過去問データを利用する。ケンブリッジ英検は KET (Key English Test; A2), PET (Preliminary English Test; B1), FCE (First Certificate in English; B2), CAE³ (Certificate of Advanced English; C1), CPE (Certificate of Proficiency in English; C2) の 5 種類があり、それぞれ CEFR レベルが指定されている。これらの試験問題からリーディングパッセージを取り出し、合計 68 件のデータを作成した (A2: 7 件, B1: 14 件, B2: 8 件, C1: 20 件, C2: 19 件)。なお、本データは、最小値 161 語～856 語で、平均は 392 語である。

3.3 手法

本研究では上記の 2 種類のデータについて、(1) 文書個別に CVLA で分析したもの、および(2) レベルごとにテキストをまとめて CVLA で分析したものを用いて妥当性を評価する。以下では CVLA の概要とその分析例を示す。

3.3.1 CVLA の概要

CVLA は入力テキストの CEFR-J レベルを推定するオンラインアプリケーションである。CEFR-J は PreA1, A1.1, A1.2, A1.3, A2.1, A2.2, B1.1, B1.2, B2.1, B2.2, C1, C2 の 12 段階に区切られており、CEFR よりも詳細な評価が可能な枠組みである (投野 (編) 2013)。リーディングモードとリスニングモードが用意されているが、本稿ではリーディングモードのみを検証の対象とする。

CVLA の CEFR-J レベルの推定は、ARI, VperSent, AvrDiff, BperA の 4 つの指標を用いて行う。ARI とは Automated Readability Index の略で文字数や 1 文の単語数などから回帰的に推定されるリーダビリティ指標である (Senter & Smith 1967)⁴。VperSent (Verb per Sentence) は 1 文中に含まれる動詞の数の平均である。この指標は文の複雑さを間接的に表していると考えられる。例えば、The window was broken. のような受動態の文では was と broken という 2 つの動詞を含むため、受動態の多い文ではこの値は上昇する。したがって、文法的に複雑な構造を持つ文が多い場合、この値が高くなる傾向がある。AvrDiff (Average Difficulty) は単語の難易度の平均である。CEFR-J Wordlist⁵ を参考に、A1 の場合 1, A2 の場合 2, B1 の場合 3, B2 の場合 4 として文章中に含まれる内容語のレベルを平均したものである (CEFR-J Wordlist には C レベルがないため A1~B2 の単語のみで計算をしている)。BperA は A レベルの内容語に対する B レベルの内容語の割合を表す。AvrDiff と合わせてこれらは語彙の難易度指標として機能する。

これら 4 つの指標について、内田(2015)で示した CEFR レベル別の教材コーパス (リーディングパート) を用いてそれぞれのレベルの平均値を示したものが表 1 である。

表 1

教材コーパス (リーディングパート) の各指標のレベル別平均値

Reading	ARI	VperSent	AvrDiff	BperA
A1	5.73	1.49	1.31	0.08
A2	7.03	1.82	1.41	0.12
B1	10.00	2.37	1.57	0.18
B2	12.33	2.88	1.71	0.26

この表からわかる通り、これら 4 つの指標はすべて CEFR レベルと比例関係にある。つまり、CEFR レベルが高くなれば、ARI, VperSent, AvrDiff, BperA も高くなるということである。CVLA では A1=1, A2=2, B1=3, B2=4 としてこれらの値から次のような回帰式を作成し、入力テキストの統計値からレベルの推定を行う。

$$\begin{aligned} \text{CEFR_level_ARI} &= 0.4298 * \text{ARI} - 1.27085 \\ \text{CEFR_level_VperSent} &= 2.1075 * \text{VperSent} - 2.01 \\ \text{CEFR_level_AvrDiff} &= 7.2961 * \text{AvrDiff} - 8.4442 \\ \text{CEFR_level_BperA} &= 16.3043 * \text{BperA} - 0.1087 \end{aligned}$$

この結果、それぞれの指標から CEFR レベルが推定されるが、入力テキストの最終的なレベル判定はこれらの平均値から算出する。また、CEFR-J への対応付けに関しては回帰によるレベル推定値を Pre-A1 (<0.5), A1.1 (<0.84), A1.2 (<1.17), A1.3 (<1.5), A2.1 (<2), A2.2 (<2.5), B1.1 (<3), B1.2 (<3.5), B2.1 (<4), B2.2 (<4.5), C1 (<5), C2 (<6) のように分割して推定する (cf. Uchida and Negishi, 2018)。これらの指標は (特に ARI と VperSent, AvrDiff と BperA) 相関関係が強いことが想定されるが、別角度から文および語彙を指標化して平均することで結果の安定化を図っている。

CVLA の推定は CEFR-J レベル (12 段階) で出力されるが、本稿で用いるデータは CEFR レベル (6 段階) が付与されているため、A1.1, A1.2, A1.3 を A1, A2.1, A2.2 を A2, B1.1, B1.2 を B1, B2.1, B2.2 を B2 と読み替えて評価する (Pre A1 は範囲外とする)。

3.3.2 CVLA による分析例

図 4 は CVLA に COE の A1 のテキスト (Aptis) を入力して分析した結果である。この分析では、ARI が 1.83 (PreA1 相当), VperSent が 2.08 (A2.2 相当), AvrDiff が 1.16 (PreA1 相当), BperA が 0.05 (A1.1 相当) とそれぞれ判定され、これらの平均値から算出したテキスト全体の推定レベルは A1.1 という結果になった。このように CVLA を利用することで明確な基準で入力テキストの CEFR-J レベルを推定することが可能である。

図 4

CVLA の分析サンプル

Dear₈₄₅₃ Sarah, How are you? I'm fine₇₂₀. I'm having a wonderful₁₅₀₀ time₅₁ in France. My course₆₂₄ is great₁₄₇ and I'm enjoying₉₁₅ learning₃₅₇ French₁₄₅₃. Paris is a very₉₈ old₁₆₂ and interesting₁₀₈₈ city₁₈₆. There are lots₂₂₀ of beautiful₁₀₁₃ old₁₆₂ buildings₄₂₅ to visit₁₀₀₁. I am staying₄₀₇ with₃₅₈ a French₁₄₅₃ family₁₅₁. They are very₉₈ nice₈₆₂ and I feel₁₃₅ very₉₈ welcome₁₅₅₀ in their house₁₅₂. I usually₇₅₅ have dinner₁₁₀₃ with₃₅₈ the family₁₅₁ in the evening₁₀₀₉. They do n't speak₃₁₉ English₁₈₄₇ so I can practice₂₉₃₅ my French₁₄₅₃. The other₇₇ students₁₃₈ in my class₄₃₂ are friendly₂₈₅₄. They are from all around the world₁₂₂. At the weekend₁₁₉₅ we are planning₇₇₂ to go₃₅ out₆₃ together₃₃₁ to watch₃₀₀ a film₆₀₉. See₅₆ you soon₆₃₁, Mark₂₂₈₉

[116 words, mean rank: 839.87, min: 35, max: 8453]

Lexical Index	Raw Frequency	Explanation		
Mode: R				
CEFR	ARI	VperSent	AvrDiff	BperA
A1	5.73	1.49	1.31	0.08
A2	7.03	1.82	1.41	0.12
B1	10.00	2.37	1.57	0.18
B2	12.33	2.88	1.71	0.26
Input	1.83	2.08	1.16	0.05
Estimated level	PreA1	A2.2	PreA1	A1.1
Estimated Text Level:A1.1				

4. 結果と考察

4.1 COE サンプルによる CVLA の検証(RQ1)

COE のサンプルパッセージによる分析を表 2 および表 3 に示す。以下、CEFR-J の A1.1, A1.2, A1.3 と分析されたものは A1, A2.1, A2.2 は A2, B1.1, B1.2 は B1, B2.1, B2.2 は B2 とみなして一致率を算出し、個別分析の表中では一致している箇所を塗りつぶして示す。

表 2 は個別の文章ごとに分析した結果をまとめたものである。13 文章中 7 文章(約 54%)が一致という結果になった。この結果は、ランダム分類(6 値分類のため約 16.7%)をベースラインとして比較するとある程度の精度であるといえるが、語数が十分に多くない文章も含むため(最小 97 語, 最大 915 語, 平均 466 語), 結果が安定しないことが考えられる。つまり, 語数が少ない場合, 極端に短いまたは長い文が 1 つでもあれば ARI への影響が大きくなり, 判定結果がぶれることになる。また, 語彙レベルに関しても文章の語数が少なければ 1 つ 1 つの語が AvrDiff に与える影響が過剰に大きくなると考えられる。そこで文章をレベルごとに結合し, 1 つのテキストとして分析した。例えば, A2 レベルには 3 つの文章があるが, これらすべてを 1 つのファイルにまとめて CVLA で分析した。結果を表 3 に示す。分析の結果, C2 以外のレベルでは CVLA の推定値と COE のサンプルのレベルが一致した(一致率 80%)。また, 不一致となった B2 は C1 と判定されているが, 最終スコア(4.68)は B1(2.95)と C1(5.24)の間に位置しており, レベルの連続性を捉えた結果となった。

表 2

COE のサンプルの個別分析

COE	PreA1	A1.1	A1.2	A1.3	A2.1	A2.2	B1.1	B1.2	B2.1	B2.2	C1	C2	Total
A1		1											1
A2					2		1						3
B1					1		1					1	3
B2									2		3		5
C1											1		1
Total	0	1	0	0	3	0	2	0	2	0	4	1	13

表 3

COE のサンプルの全体分析

CEFR	ARI	VperSent	AvrDiff	BperA	Final
COE_A1 (116 語)	1.83	2.08	1.16	0.05	0.65
	PreA1	A2.2	PreA1	A1.1	A1.1
COE_A2 (462 語)	6.88	2.15	1.52	0.12	2.18
	A2.1	B1.1	B1.1	A2.1	A2.2
COE_B1 (1479 語)	7.32	2.79	1.58	0.19	2.95
	A2.1	B2.1	B1.2	B1.1	B1.1
COE_B2 (3091 語)	10.54	3.67	1.8	0.3	4.68
	B1.2	C2	C1	C1	C1
COE_C1 (915 語)	11.19	4.72	1.9	0.39	5.24
	B2.1	C2	C1	C2	C1

4.2 ケンブリッジ英検のデータによる CVLA の検証(RQ2)

次にケンブリッジ英検のデータでの検証を行った。文章個別の分析では 68 文章中 36 件(約 53%)が一致する結果となった(表 4)。前節での検証同様、文章の長さによって結果が不安定になると考えられるため(最小値 161 語, 最大値 856 語, 平均 392 語), レベルごとにすべての文章を結合したものでの検証も実施した(表 5)。その結果, A2 レベルは A1.3, C2 レベルは C1 と判定されたが, それ以外のレベルでは一致した(一致率 60%)。前節と同じく, 推定レベルの判定値は目標 CEFR レベルと比例関係にあり, レベルの上下関係を捉えている。具体的には, A1.3 と判定された A2 の文章は 1.47 であり, B1 の 3.13 よりも低い。また, ともに C1 と判定された CAE_C1 (4.76), CPE_C2 (4.98)では, 相対的にレベルが下である前者の判定値のほうが低く算出されている。

表 4

ケンブリッジ英検の個別分析

Cam.	PreA1	A1.1	A1.2	A1.3	A2.1	A2.2	B1.1	B1.2	B2.1	B2.2	C1	C2	Total
A2			1	1	5								7
B1						3	2	7		2			14
B2								3	3	2			8
C1							1	2	2	2	13		20
C2								1	2	2	10	4	19
Total	0	0	1	1	5	3	3	13	7	8	23	4	68

表 5

ケンブリッジ英検の全体分析

CEFR	ARI	VperSent	AvrDiff	BperA	Final
KET_A2 (1411 語)	5.59	2.59	1.26	0.04	1.47
	A1.2	B1.2	A1.1	A1.1	A1.3
PET_B1 (4880 語)	7.84	3.02	1.58	0.19	3.13
	A2.2	B2.2	B1.2	B1.1	B1.2
FCE_B2 (5356 語)	8.13	3.67	1.64	0.22	3.81
	A2.2	C2	B2.1	B1.2	B2.1
CAE_C1 (8561 語)	9.53	3.44	1.84	0.35	4.76
	B1.1	C1	C1	C2	C1
CPE_C2 (6437 語)	11.14	3.73	1.84	0.34	4.98
	B2.1	C2	C1	C1	C1

4.3 考察

2つのデータを用いた検証の結果、個別データの分析では全体で81件中43件が一致となり、約53%程度の精度で推定できることが明らかになった。ランダム分類と比較すると、ある程度の精度であるといえるが、不一致となった場合について、どの程度外れているかという点がCVLAの妥当性検証において重要な側面である。レベルの一つ違いまでを正答とみなし(例えばB1が正答の場合でCVLAの結果がA2.2と推定した場合など)、隣接レベルを含めて検証すると、COEでは10/13(約77%)、ケンブリッジ英検では55/68(約81%)がその範囲に含まれる。これらを合わせると全体で65/81(約80%)となる。このことは、ほとんどの場合において推定結果が大きく外れていないことを示している。また、それぞれのデータについてレベルごとに文章をまとめた全体分析の結果を見ても、不一致であった部分のCEFR-Jレベルを判定する数値はレベルの上下関係を

正しく捉えており、推定結果が大きくずれないことを示唆する。

一方、レベル別の正答率をみると、A1: 100%(1/1), A2: 70%(7/10), B1: 約 59%(10/17), B2: 約 54%(7/13), C1: 約 67%(14/21), C2: 約 21%(4/19)となっており、C2 レベルの正答率が最も低い。このことは CVLA が依拠している教材コーパスのデータが B2 までであり、C1, C2 レベルについては回帰直線を伸ばして擬似的に推定していることが影響していると考えられる。この点は CVLA の将来的な改善課題の1つである。

今回検証に用いたデータは、81 件あり、その平均語数は約 404 語(最小値 97, 最大値 915, 標準偏差 235.3)であった。一致の場合の平均語数は 410.2 語, 不一致の場合の平均語数は約 396.5 語(N=37)であり、一致のほうがやや語数が多かった。標準偏差からはややばらつきのあるデータ群であることが読み取れるが、読解教材の場合、実験の結果から、およそ 400 語程度あれば約 80%の場合で CEFR レベル 1 つのずれ幅内にとどまることが明らかとなり、ある程度の精度で CEFR レベルを判定できることが示唆される。

5. まとめ

本稿では、Uchida and Negishi (2018)で提案された CVLA の妥当性について 2 つのデータソースを使って検証した。その結果、一致率は約 53%, 隣接レベルも含めた一致率は約 80%となり、CVLA の推定値は概ね妥当なものであり、400 語程度の入力があれば、大きくずれたレベルを判定することは少ないことが明らかとなった。CVLA の特徴の一つとして、Lexile Measure や Text Inspector とは違って指標を完全にオープンにしているという点が挙げられる。この性質を利用することで入力テキストをリライトする指針を得ることができる。例えば、ある入力テキストの VperSent が他の指標に比べて高い場合、受動態や関係節などを減らし、よりシンプルな文法構造にすることでテキストのレベルを下げる可以降低ことができる。一方、AvrDiff や BperA が対象となる学習者や受験者よりも低い場合、単語をより上級なものに置き換えることで文章のレベルを上げることができる。CVLA のすべての指標は CEFR レベルが上げれば高くなるため、このように直感的に文章レベルの調整ができるのである。

CVLA の課題として、C レベル以上の文章の判定精度を向上するということが挙げられる。今後は C1, C2 レベルの教材コーパスを整備し、それぞれのレベルの統計値を算出したいと考えている。また、今回はリーディングのみの検証にとどまったため、リスニングスクリプトでの検証も必要である。さらに、テキストレベルとタスクレベルの関係については今後さらなる検討が必要であると考えている。

注

- ¹ <http://dd.kyushu-u.ac.jp/~uchida/cvla.html> にて利用できる。また, <https://cvla.langedu.jp/>にて文ごとに指標を表示する version 2 を公開している。
- ² 次の URL よりアクセス可能である (最終アクセス日: 2021 年 1 月 10 日)。 <https://www.coe.int/en/web/common-european-framework-reference-languages/reading-comprehension>
- ³ ただし, CAE が Grade A 評価の場合は C2 として判定される。
- ⁴ ARI は次の数式で算出される。

$$4.71 \left(\frac{\text{characters}}{\text{words}} \right) + 0.5 \left(\frac{\text{words}}{\text{sentences}} \right) - 21.43$$

- ⁵ 『CEFR-J Wordlist Version 1.3』 東京外国語大学投野由紀夫研究室. <http://www.cefr-j.org/download.html>

付記

本論文の一部は 2019 年 12 月に横浜で実施された British Council 主催の国際会議 New Directions 2019 で発表した Uchida, S. & Negishi, M. (2019) “Assigning CEFR and CEFR-J levels to Lexile measures: A corpus-based approach”に基づいている。

謝辞

本研究は JSPS 科研費 JP18H00693 および JP20H00095 の助成を受けたものである。

引用文献

- Bax, S. (2012). Text Inspector: Online text analysis tool. Available at: <https://textinspector.com/>.
- Bax, S., Nakatsuhara, F., & Waller, D. (2019). Researching L2 writers' use of metadiscourse markers at intermediate and advanced levels. *System*, 83, 79-95.
- 中條清美・長谷川修治(2004)「語彙のカバー率とリーダビリティから見た大学英語入試問題の難易度」『日本大学生産工学部研究報告 B』 37, 45-55.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Addison-Wesley Longman Ltd.
- Lennon, C., & Burdick, H. (2004/2014). The Lexile framework as an approach for reading measurement and success. Electronic publication (http://cdn.lexile.com/cms_page_media/135/The%20Lexile%20Framework%20for%20Reading.pdf; accessed January 10, 2021).
- 根岸雅史(2015)「Lexile Measure による中高大の英語教科書のテキスト難易度の研究」『ARCLE REVIEW』 9, 6-16.
- 大田悦子(2016)「Lexile Measure を用いた中高英語教科書の難易度比較」『白山英米文学』41, 1-20.
- Senter, R. J., & Smith, E. A. (1967). Automated readability index. AMRL-TR-6620. Aerospace Medical Division, Wright Patterson AFB.

- 白井恭弘(2012)『英語教師のための第二言語習得論入門』大修館書店.
- Smith, M., & Turner, J. (2018). The Common European Framework of Reference for Languages (CEFR) and the Lexile framework for reading bringing more precision to language learning. Electronic publication (https://metametricsinc.com/wp-content/uploads/2018/01/CEFR_1.pdf; accessed January 10, 2021).
- 内田諭(2015)「CEFR レベルに基づいた教材コーパス: レベル別基準特性の抽出に向けて」『英語コーパス研究』22, 87-99.
- Uchida, S., & Negishi, M. (2018). Assigning CEFR-J levels to English texts based on textual features. In Tono, Y. & Isahara, H. (eds.) *Proceedings of Asia Pacific Corpus Linguistics Conference, 4*, 463-467.
- 投野由紀夫(編)(2013)『英語到達度指標 CEFR-J ガイドブック』大修館書店.

