

英語長文読解用素材文の客観的難易度分析の方法について

——リーダビリティ指標と語彙レベル分析ツールを用いて——

秦野 進一（東北大学）

大学入試の英語試験などで一般的な長文読解問題に使用する素材文の難易度を客観的に分析・数値化する方法を紹介する。文章の内容に対して感じる質的な難しさの感覚は個人差があるので数値化できないが、英文の形状的な難しさや使用語彙の難しさは客観的に数値化可能である。本稿では主として文章の長さや単語の難しさを分析対象として英文の難易度を数値化するリーダビリティ指標と、語彙レベルリストに基づいた語彙レベル分析ツールを利用して英文で使用されている単語の難易度を分析する方法を紹介する。事前に素材文の候補を客観的な指標で分析することが、受験生の学力を考慮した、適切な難易度の作題を行う一助となることが期待される。

1 はじめに

英語の読解問題用の素材文は担当者の主観で選ばれることが多いように思われる。それぞれの担当者が自分のこれまでの経験を活かして適切な難易度の英文を選ぶことができれば問題ないが、経験の乏しい担当者の場合、自分の判断に自信が持てなかったり、あるいは受験生の学力を測るという目的に合っていない難易度の英文を選んでしまったりということは十分に起こりうることである。そのような際に英文の難易度を客観的に測定するスケールがあれば、安心して適切な難易度の英文を選ぶことができる。すでに英語の入試問題研究や語彙分析などの研究で用いられている指標や語彙リストが存在しているので、これらを素材文選択の際に活用すれば客観的な判断指標を元に素材文を選ぶことができるようになる。そこで本稿では適切な難易度の英文を選ぶための資料を提供し、かつ作題者の負担軽減にも寄与する難易度分析の方法を提案したい。

2 リーダビリティ指標

英文の難しさを数値化するためにすでに様々なリーダビリティ指標が利用されている。本稿では英語圏で一般的に利用されている、文章の長さや単語の音節の多さを変数にして分析を行う Flesch Reading Ease と Flesch-Kincaid Grade Level, それに文章の長さや使用語彙の意味的な難しさを変数に加えて分析を行う Lexile measures を紹介する。

2.1 Flesch Reading Ease と Flesch-Kincaid Grade Level

一番簡単に計測できるリーダビリティ指数はこの 2 つであろう。多くの人が使っているワープロソフトの

Word の校閲機能を利用することでインターネットに接続することもなく 2 種類のリーダビリティ指標の数値が得られる。どちらも文章の平均の長さ（文の数で割った単語数）と 1 単語あたりの平均音節数を元に公式にあてはめて算出する。1 文の長さが長く、単語の平均音節数の多い文章は難易度が高く、1 文が短く、単語の平均音節数が短い文章は難易度が低いという考えに基づく指標である。各指標には以下のような特色がある。

2.1.1 Flesch Reading Ease

この指標は以下の計算式によって得ることができる。¹⁾（日本語訳は筆者）

$$206.835 - (1.015 \times 1 \text{ 文あたりの平均単語数}) - (84.6 \times 1 \text{ 語あたりの平均音節数})$$

スコアは 0 から 100 までの数値で表され、60～70 が標準的な難しさで、数が小さいほど英文が難しく、数が大きいほど英文が簡単であることを表している。

2.1.2 Flesch-Kincaid Grade Level

上記の指標の公式に改良を加えた以下の計算式で指標が得られる。²⁾

$$(0.39 \times 1 \text{ 文あたりの平均単語数}) + (11.8 \times 1 \text{ 語あたりの平均音節数}) - 15.59$$

この指標は数値がアメリカの学校の児童・生徒の読解レベルを表す学年で表されるので難易度が理解しやすいという利点がある。この指標以外にも The

SMOG Index という指標は数値を学年で表している。リーダビリティ指標には様々なものがあるが、それぞれ特有の数値で表されるものが多いため 58 とか 980 といった数値がいったいどの程度難しいものなのかは慣れないと理解しにくい。その点この指標は「6」ならば小学校 6 年生レベル、「9」なら中学 3 年生レベルの英文であることを表しているのでイメージしやすい。

2.2 Word を使った分析

Word 2016 を利用してリーダビリティの分析を行うための事前設定方法は Microsoft サポート³⁾ に図 1 のように説明されている。

1	[ファイル>オプション]に移動します。
2	[文章校正]を選択します。
3	[Word のスペルチェックと文章校正]で [文章校正とスペルチェックを一緒に行う] チェックボックスがオンになっていることを確認します。
4	[読みやすさの統計情報を表示]を選択します。

図1. リーダビリティ指標の事前設定方法

この設定を行った後、分析対象のファイルを開き、F7 キーを押してスペルチェックを実行するか、「校閲」→「スペルチェックと文章校正」をクリックすると表記上の修正候補などが提示され、文章校正が終了すると読みやすさの評価スコアが出てくる。一番下の Readability の欄に Flesch Reading Ease と Flesch-Kincaid Grade Level のスコアが表示される。

2.3 Lexile measures

Lexile measures は米国 MetaMetrics 社が開発した指標で、英文の難しさを表す指標であると同時に読み手の読解力を表す指標でもある。Lexile reading measure と Lexile text measure の 2 種類あり、前者はテストの結果などによって計測された受験者の読解力を数値で示し、後者は「1 文あたりの長さ」と「単語の出現頻度」から分析した単語の難易度によって文章の難易度を数値で示す。そのため教育現場では生徒の読解力に合わせたテキストを選択する際になどに使われている。Flesch Reading Ease と Flesch-Kincaid Grade Level が単語の難しさを単語の音節の長さで計測しているのに対し、Lexile measures は 500 万語に及ぶコーパス（言語資料の集合体・用例集）に多く出てくる単語（出現頻度の高い単語）は難易度の低い単語

であり、コーパスにあまり出てこない単語（出現頻度の低い単語）は難易度の高い単語であるという考えに基づいて分析を行う。1 文の長さが長く、また出現頻度が低い単語が使用されている文章は難易度が高く、1 文が短く、使用されている単語の出現頻度が高い文章は難易度が低いという考えに基づく指標である。この指標は 900L, 1000L のように数値に L の記号を付けて表され、数値の大きい方が難易度が高い。

分析手順としては Lexile Analyzer⁴⁾ というウェブサイト上の分析ツールのテキストボックスに分析対象の英文を入れて分析ボタンを押すことでスコアが得られる。なおこの分析は 250 語未満の語数なら登録なしで利用できるが、250 語を超える英文を分析するには簡単な登録をしてアカウントを作成することで 60 日間は無料で利用できる。その期間以降も利用する場合には年間 2000 円程度の料金がかかる。

2.4 実際の分析例

同じテキストを使って上記 3 つのリーダビリティの分析を行ってみた結果が以下の表 1 である。それぞれの指標が難易度の異なる 2 つのテキストをどのように数値化しているかをみるために令和 3 年度共通テスト第 2 日程第 6 問 (B) と TIME 誌の記事”We May Never Eliminate COVID-19. But We Can Learn to Live With It” (Jamie Ducharme, 2021) の 2 つの英文を使用した。第 6 問 B の語数 (628 語) とほぼ同じ語数で分析をするために TIME の記事は前半部分 (621 語) のみを使用した。いずれの指標も第 2 日程第 6 問 (B) より TIME の記事の方が難易度が高いことを示している。

	第2日程6問(B)	TIME記事
Flesch Reading Ease	52.9	38.3
Flesch-Kincaid Grade Level	9.4	13.1
Lexile text measure	1010L-1200L	1210L-1400L

2.5 その他のリーダビリティ指標

素材のおおよその難易度を知るには上記のいずれかの指標を用いれば十分だと思われるが、もっと多くの指標を参考にしたいという方は、Readability Formulas⁵⁾ というウェブサイトにある Automatic Readability Checker⁶⁾ を利用して分析したい英文を送れば同時に 7 種類のリーダビリティ指標の数値を返

してくれる。研究目的でなければここまで多くの指標は不要なので、自分に合った使いやすいものを選べばよいであろう。

リーダビリティ指標は主として英文の形状的な難易度を利用して計測しているため、受験生にとって本当にその指数が示す難易度通りに難しく感じるかどうかは実際に人間の目で英文を読んで確かめる必要がある。あくまで参考資料として利用することをお勧めする。

3 単語の難易度

我々が英文を読むときに、その英文を「難しい」と思う大きな要因の1つは単語の難しさである。知らない単語が多く出てくる英文を読むことはつらく、辞書に頼ることなく内容を理解することが難しいことも多いであろう。作題者は受験生にある程度の語彙力を持っていることを期待して素材文を選ぶであろうが、受験生の英語力に比して難しすぎる英文では英語力を適切に測ることはできない。受験生の英語力を測るのに適した難易度の英文を選ぶためには使用されている単語の難易度を理解した上で素材文を選ぶことが大切である。単語の難易度は、しかるべき語彙レベルリストに基づいて使用されている単語について分析をし、難易度の高い語がどの程度含まれているかをみることで判断できる。以下に日本人英語学習者を対象とした単語の分析に適していると思われる語彙レベルリストである新JACET8000とCEFR-J Wordlist, それに単語レベルチェッカー⁷⁾(語彙レベル分析ソフト)とその具体的な利用方法について紹介する。

3.1 新JACET8000

新JACET8000(大学英語教育学会基本語改定特別委員会, 2016)は2005年に大学英語教育学会により日本人英語学習者のための教育語彙集として出版された8000語の英単語リスト「JACET8000」(相澤ほか, 2005)の2016年改訂版である。旧版はBNC(The British National Corpus)という1990年前後に作られた大規模なイギリス英語コーパスを母体に編集されたが、新版ではBNCに加えて、アメリカ英語のコーパスであるCOCA(The Corpus of Contemporary American English)を母体としてベースとなるリストを作成し、さらに日本人英語学習者の英語学習の実態および英語学習の目標に合わせたものに補正するための資料を基に補正された。資料として使われたのは、中学・高校の検定教科書、大学入試センター試験問題、47都道府県の公立高校入試問題、英検・TOEFL・TOEICの問題、日本の英字新聞、英

語による学術入門書などである。

JACET8000は1000語単位でレベル分けがされており、2005年版の目次では以下の表2のようにおおよそのレベルについての説明が掲載されている。

レベル	順位	説明
Level 1	1-1000	中学校の教科書に頻出する基本単語
Level 2	1001-2000	高校初級レベルの単語。英検準2級レベル。
Level 3	2001-3000	高校英語教科書レベルの単語。大学入試センター試
Level 4	3001-4000	大学受験、及び大学一般教養の初級レベルに相当。
Level 5	4001-5000	難関大学受験、及び大学一般教養に相当。
Level 6	5001-6000	英語を専門としない大学生やビジネスマンが目指すレベル。英検準1級レベル。
Level 7	6001-7000	英語専攻の大学生や仕事で英語を使うビジネスマンの到達目標とするレベル。英検1級、TOEICの95%以上の単語をカバーしている。
Level 8	7001-8000	日本人の英語学習者一般的な単語学習の最終到達目標です。

また新JACET8000には本体としてのリスト以外に「中学・高校コミュニケーション支援語彙リスト」、「共通学術語彙リスト」、「発信語彙リスト」の3つの付加リストも付いている。

2020年度から小学校で全面实施されている新しい小学校学習指導要領(文部科学省, 2017)では小学校でも英語の教科化が始まり、600語から700語の単語学習が行われている。したがって小学校・中学校・高校を通した学習単語数は4000語から5000語へと大幅に増えることになり、現在では上記リストのLevel 3が「高校英語教科書レベルの単語」であるが、新学習指導要領が完全実施された場合にはLevel 4もしくはLevel 5までの単語が「高校英語教科書レベルの単語」となる。

仮にこの分類でLevel 4レベルまでの単語力(項目

数累計 4000 語) を受験生に求めるならば, Level 5 レベル以上の語について注を付けるかどうかを検討すればよい。また選抜性の高い大学であれば Level 5 までの 5000 語を受験生に求めることとして Level 6 以上の語について注を付けるかどうか検討することになる。前後関係から意味を推測できる語や, すでに日本語として使われている語などには注は不要であるので, 設定したレベル以上の語だからといって機械的に注を付ける必要はない。

3.2 CEFR-J Wordlist

CEFR-J のウェブサイト⁸⁾ によれば, 「CEFR-J は欧州共通言語参照枠 (CEFR) をベースに, 日本の英語教育での利用を目的に構築された, 新しい英語能力の到達度指標」である。指標自体は「聞くこと」, 「読むこと」, 「書くこと」, 「話すこと (やりとり)」, 「話すこと (発表)」の 5 技能それぞれについて 12 レベルの「言葉を使って何が出来るか」を示す CAN-DO リストの形式を取っている。例えば B1.1 レベルの「話すこと (やりとり)」では, 「個人的に関心のある具体的なトピックについて, 簡単な英語を多様に用いて, 社交的な会話を続けることができる。」, B1.2 レベルの「聞くこと」では, 「はっきりとなじみのある発音で話されれば, 身近なトピックの短いラジオニュースなどを聞いて, 要点を理解することができる。」などと表記されている。この CEFR-J に付属する資料として CEFR-J Wordlist がある。約 7000 語が A1 から B2 までの 4 レベルに分類されている。おおよその相当レベルを表 3 に示す。仮にこの分類で B2 レベルまでの単語力 (項目数累計 6868 語) を受験生に求めるならば, (B2 までのリストに) 該当なし (NA : Not Applicable) と表示された単語について注を付けるかどうかを検討すればよい。

レベル	単語数	日本の学校教育での相当レベル
A1	1068	小学校～中学1年程度
A2	1352	中学2年～高校1年程度
B1	2353	高校2年～大学受験レベル
B2	2692	大学受験～大学教養レベル
合計	6868	

3.3 単語レベルチェッカー

単語レベルチェッカーはイー・キャスト社製の市販ソフトで, 中学校・高校の各教科書の持つシェアのパーセンテージに基づいて生徒の何%がその単語を知っているかという「認知率」を学年別に数値化し, さら

に過去 10 年分のセンター試験の本試験の全単語を分析して, 6 段階のレベルに分けた 8517 語のリストに基づいて語彙レベルのチェックを行う。自分のパソコンにインストールして使用するのでインターネットにつながることなく分析ができる。毎年新しいセンター試験 (今後は共通テスト) のデータを元に改訂版を作成しているため最新の単語などもリストに反映されている。分析後にはレベル外単語数, レベル外単語率などが表示され, 難語を平易語への置き換えをサポートするソーラス (類義語辞典) 機能も搭載されている。レベルのおおよその目安を表 4 に示す。

レベル	単語数	基準
中1	482	中1の認知率50%以上
中2	419	中2の認知率45%以上
中3	445	中3の認知率35%以上
高1	965	高1の認知率25%以上
高2	1588	高2の認知率15%以上
高3 / センター試験	4618	高2の認知率14.9%以下, または高3新出語, センター試験
合計	8517	

仮にこの分類で「高3 / センター試験」レベルまでの単語力 (項目数累計 8517 語) を受験生に求めるならば, レベル外と示された単語について注を付けるかどうかを検討すればよい。

3.4 実際の語彙分析

語彙リストの特徴を知っただけでは実際にどのように語彙分析を進め, どのように結果を利用するのかイメージがつかみにくいと思うので, 同じテキストを使って実際に分析を行ってみることとする。使用するテキストは 1.4 で使用した令和3年度共通テスト第2日程第6問 (B) と TIME 誌の記事 (前半部分) である。

3.4.1 New Word Level Checker⁹⁾ 利用の語彙分析

オンラインで利用できる英文語彙難易度解析プログラムである New Word Level Checker を利用すれば上記の新 JACET8000 と CEFR-J の語彙リストに基づいた難易度分析ができる。このサイトでは他にも SVL12000¹⁰⁾ などの 5 種類の語彙リストに基づく語彙分析が可能である。利用方法はいたって簡単で, ウェブサイトのテキストボックスに分析したい英文を入れ,

使用したい語彙リストを選択して「Check」ボタンをクリックするだけで語彙レベルごとにテキストカバー率とそのグラフ、難易度ごとに色分けしたテキストを返してくれる。その後「Word List」をクリックすればすべての語のレベルの一覧表が返ってくる。このWord Listはボタン1つでExcelファイルとしてダウンロードできるので、レベル順に並び替えて難しい語の一覧を表示させるなどの使い方ができる。4~5色で色分けされたテキストの表示は色によってはやや見分けが付きにくいという難点がある。その場合にはダウンロードしたWord Listの一覧表も併用して注を付けるかどうかの判断の際に使うとよい。

New Word Level Checkerでは固有名詞や数字は既知語として分類される(PropNoun_Numと表示)ので、それ以外の語についてそれぞれの語彙レベルが表示される。

今回は受験生に求める語彙レベルをLevel 5(新JACET8000)とB2(CEFR-J)として分析し、リスト外(NA)に分類された語からアルファベット単体(AやXなど)や記号(%など)を除いた語を難語と定義し、総数を表2にまとめた。

	条件	第2日程6B	TIME記事
新JACET8000	Level 6以上 +リスト外の語*	7語	39語
CEFR-J	リスト(B2まで)外の語*	15語	43語
単語レベルチェッカー	「高3/センター」レベル外の語	5語	40語
*アルファベット単体と記号は除く			

3.4.2 単語レベルチェッカー利用の語彙分析

単語レベルチェッカーのテキストボックスに分析対象の英文をコピー&ペーストし、レベルを高3/センターに設定してレベル判定を行った結果も表2にまとめている。

設定したレベルが「高3/センター」なので、そのレベルを超える難易度の語(難語)が赤色、それ以外の語は黒字で表示されるので難語が一目でわかり見やすい。前述の2つのリストでは難語に分類されてしまったアルファベット単体や記号は単語レベルチェッカーでは難語には分類されていない。

またそれぞれのリストに基づいて難語と認定された語が3つのリストに共通なのかどうかを調べてみた結果を以下に記す。

第2日程第6問(B)について単語レベルチェッカー

で難語となった5語(floss, gel, harden, sealant, prevention)のうちprevention以外の4語は他の2つのリストでも難語となっている。preventionは新JACET8000ではL5, CEFR-JではB2にレベル分けされているためそれぞれのリストでは難語となっていない。新JACET8000とCEFR-Jでは難語となっているが、単語レベルチェッカーでは難語となっていない語は以下表6の通りである。

新JACET8000	coating, conversely, Finnish, well-maintained
CEFR-J	age(動詞), brushing, chewing, coating, digest, Finnish, impact, Japanese, low(名詞), Ministry, seal, well-maintained

TIMEの記事についてはそれぞれおよそ40語前後が難語と分類されており、3つのリストに共通して難語となっている語は24語であった。残りの語も多く、3つのリストのうち2つのリストでは共通して難語となっていた。リストが扱っている語数の違いもあるので多少の違いはあるものの、英文で使用されている単語のおおよその難易度を把握することは、ここにあげた3つのリストのいずれでも可能であると思われる。

いずれの語彙レベルリストを用いても、ときどきどうしてこの語が難語と分類されているのだろうか(あるいはその逆もある)ということがあるので必ず自分の目でリストを確認する必要がある。

3.5 難語の言い換えと著作権

難しい単語があるのなら、すべて平易な単語に置き換えれば素材文として使えるかというところ簡単にはいかない。著作権の問題が絡むからである。文化庁(n.d.)では入学試験問題として詩や論文を利用する場合の改変の許容程度について以下のように説明している。

「A 著作権者は著作権者人格権の一つとして、著作物の内容や題名を他人に無断で改変されない権利を持っています(第20条)。しかし、どのような場合においても、改変が認められないわけではなく、例えば、教育目的上必要な用字・用語の変更(例難しい漢字をひらがなに改める)、空白の個所に正しい用語を入れさせる穴埋め問題など、真に

やむを得ない改変は許されます(第20条第2項)。なお、(社)日本文芸家協会から、入試問題への文芸作品の使用について『出題に際しみにだりに作品を改変しないこと』などを内容とした要望書が、全国の大学などへ送付されています。」

文芸作品については「みにだりに作品を改変」せず、それ以外の文章についても「真にやむを得ない」「教育目的上必要な用字・用語の変更」についてのみ許されるのであるから、平易な語に書き換えることは真に必要な場合に限る、注として説明を加えるなどの方法と併用して対応すべきであろう。もし注語が多くなりすぎるから平易な語への改変を多くしなくてはならないとしたら、それはそもそも素材文として不適格であると考えべきである。

4 おわりに

本稿で紹介したリーダビリティ指標や語彙レベル分析ツールは慣れないと作業が煩雑に思えるかもしれない。また得られた数値もどう解釈しているのかよくわからないこともあるであろう。しかしどれも慣れれば5分程度で終わるものばかりであり、自分にとって理解しやすい数値で表記されたものを使っていればすぐに慣れると思われる。そしてぜひ経年で同じ指標を用いて分析していくとともに、試験結果の分析も合わせて行って英文を検討してみたい。そうすれば自分の大学の受験生の英語力を測定するのに適切と思える難易度の英文がどの程度のものか見えてくるであろう。初めは過去に出題した問題の中でうまくできていたと思われる英文を分析して、その難易度に近いものを選ぶということも可能である。勘や主観に基づいた素材文選択から客観的な指標に基づいた選択に移行するために今回紹介した分析ツールが活用されることを期待している。

注

- 1) <https://www.readabilityformulas.com/flesch-reading-ease-readability-formula.php> (2021.2.16)
- 2) <https://readabilityformulas.com/flesch-grade-level-readability-formula.php> (2021.2.16)
- 3) <https://support.microsoft.com/ja-jp/office/%E6%96%87%E6%9B%B8%E3%81%AE%E8%AA%AD%E3%81%BF%E3%82%84%E3%81%99%E3%81%95%E3%81%A8%E3%83%AC%E3%83%99%E3%83%AB%E3%81%AE%E7%B5%B1%E8%A8%88%E6%83%85%E5%A0%B1%E3%82%92%E5%8F%96%E5%BE%97%E3%81%99%E3%82%8B-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2> (2021.2.10)

- 4) <https://hub.lexile.com/analyzer> (2021.2.10)
- 5) <https://readabilityformulas.com/> (2021.3.8)
- 6) <https://readabilityformulas.com/free-readability-formula-tests.php> (2021.2.16)
- 7) 単語レベルチェッカー2020 イー・キャスト
- 8) <http://cefr-j.org/cefrj.html> (2021.2.10)
- 9) 関西大学大学院外国語教育研究大学院応用言語学部の水本敦教授によって開発されたプログラムで、前身のプログラムは青山学院大学文学部英米文学科の染谷泰正教授が開発したWord Level Checkerである。
<https://nwlc.pythonanywhere.com/> (2021.3.23)
- 10) 株式会社アルクが独自に開発した語彙リストで最上級のレベルの難語を含む12000語を1000語ずつの12レベルに分類している。

謝辞

本研究はJSPS 科研費JP20K20421の助成を受けたものである。

参考文献

- 相澤一美, 石川慎一郎, 村田年編集代表; デビッド・クルソン 英文校閲 (2005). 『「大学英語教育学会基本語リスト」に基づくJACET8000英単語』 桐原書店.
- 文化庁 (n.d.). 著作権なるほど質問箱, https://pf.bunka.go.jp/chosaku/chosakuken/naruhodo/answer.asp?Q_ID=0000369, (2021.3.23)
- 大学英語教育学会基本語改定特別委員会 (2016). 『大学英語教育学会基本語リスト 新JACET8000』 桐原書店.
- Jamie Ducharme, (2021.2.4). “We May Never Eliminate COVID-19. But We Can Learn to Live With It” *TIME*
- 文部科学省 (文部科学省, 2009). 「高等学校学習指導要領」
- 文部科学省 (文部科学省, 2017). 「小学校学習指導要領」
- 染谷泰正 (2009). オンライン版「英文語彙難易度解析プログラム」(Word Level Checker)の概要およびその教育研究分野での応用可能性, 『青山学院大学文学部紀要』 51, 99-122.