

多読学習用リーダー・コーパス構築と分析：

学習者が感じる「難しさ」の解明へ向けて

加 野 ま き み

要 旨

多読学習は大学の英語教育の場で、その効果が広く認められている課題の一つである。多読学習には通常、外国人学習者のために語彙・文法などが調整された Graded Reader (以下 GR) が使用されるが、本学では GR に加えて Youth Reader (以下 YR) と呼ばれるネイティブスピーカー用児童書も合わせて読むことを勧めている。しかし、YR に対して「難しい」と感想を述べる学習者がいるということをしっかきに、筆者は「難しさ」となり得る言語的特徴は何なのか解明しようと試みた。本稿では GR と YR のコーパスを構築し、YR の特徴を探った。その結果、YR は GR と比較して、基本 1000 語の語彙の割合が少なく、それ以上のレベルの語の割合が大きい、レベル毎に語彙レベルが着実に上昇している、受動態の割合が高くしかも動詞句の構造が複雑である、基本語に大きな頻度の差がある、1 単語が複数の語法で用いられる、複雑な文構造が見られる、叙事的な表現が多いなどの特徴があることが分かった。これらは学習者が「難しい」と感じる要素となり得ると考える。これまで多読学習の現場で使用されてきたものの、その言語学的性質について十分には分析されてこなかった YR をコーパス言語学的に GR と比較することによって、その特徴を明らかにした。

キーワード：Graded Reader, Youth Reader, 語彙レベル, 受動態, 文構造の複雑さ

1. はじめに

学生に高度な英語運用能力を身につけさせるという大学に課された社会的ニーズに応えるため、本学文化学部では、学生一人一人が授業時間外にたくさんの英文を読むという学習を行うことで、授業時間で学んだ語彙・文法に繰り返し触れ定着を図る「多読学習プログラム」を実施している。言語習得の為に多読学習が効果的であることは広く認められており、英語カリキュラムの一部として実施する教育機関も国内外で年々増加している。

多読学習では Graded Reader と Youth Reader という 2 種類のリーダーが使用されることが多いが、Youth Reader に対して「難しい」と感想を述べる学習者がいるということをしっかきに、筆者は「難しさ」となり得る言語的特徴は何なのかに興味を持った。本稿では、そのような「難しさ」の解明に向けて、まずはこの 2 種類のリーダーにはどのような違いがあるのか明らかにすることを目的とする。多読学習用に使用されているリーダーを収録したコーパスを構築し、コーパス間の様々な比較・分析により、2 種類のリーダーが持つ、異なる言語的特徴を導き出す。

本稿では、まず、第2節で多読学習、リーダーとそのレベル分けについての説明と、関連する先行研究の紹介を行った後、第3節でコーパスの作成・分析の方法、使用したツールを概説し、第4節で分析の結果を論じ、第5節でまとめと今後の課題を述べる。

2. 先行研究

2.1 多読学習について

言語習得の為に多読学習が効果的であることは広く認められている。多読学習の目的は英語により多く触れることで、既習語彙・文法の「反復練習」を行い、定着へと導くことである。多読学習では、個人の実際の英語力よりもやや低いレベルに設定された英語の Graded Reader (以下 GR) と呼ばれる、語彙レベル別に編集された小説等の簡略版やオリジナルの物語を、正確さより読む「量」に重きを置きながら読む。主に授業時間外に行うので、授業時間外の学習時間の確保にも繋がり、多数ある本の中から学生が自分で何を読むか選ぶので、楽しみながら学習することができる (ロブ & 加野, 2010)。

2.2 多読学習用リーダーについて

Claridge (2005) によると、学習者が使用する GR はオリジナルのストーリーを保ちつつ、低頻度語やコロケーションの置き換え、文の長さの調整、文体の均質化などのリライト作業が行われているが、英語としての真正性は保持しているとしている。しかし、GR は「生の英語」とは言えない (Swaffar, J., 1985, Honeyfield, J., 1977), 最終的に GR の限られた語彙と文法から脱却し、ネイティブスピーカー用に書かれた一般書を読めるようになるには GR と一般書の間には架け橋が必要である (Clafin, 2012) など、GR に依存しすぎることを疑問視する研究もある。そのため、本学では GR に加えて Youth Reader (以下 YR) と呼ばれるネイティブスピーカー用児童書も合わせて読むことを勧めている。YR も段階的に難易度が増すよう、語数、低頻度語、文の長さやストーリーの複雑さ、テーマ・ジャンルなどが調整され、レベル分けされている (LEVELING CRITERIA より)。YR は読み聞かせや、音読練習に使われることもあるネイティブスピーカーの子供用の読み物である。しかし、YR などのネイティブスピーカー用のリーダーは英語学習者の多読学習には適さないという研究報告もされており (Webb, S., & Macalister, J., 2013), YR の有効性の議論は分かれている。

2.3 リーダーのレベル分けについて

本学文化学部の「多読学習プログラム」は、主に1年次英語必修科目で英語力向上のため実施している。リーダーには出版社毎にレベル分けがされているが、それらはすべて Kyoto Scale と呼ばれる本学が設定したレベルに当てはめられている (図1)。各リーダーの Kyoto

Reader series written for language learners			
Starter	Foundation Series-Levels 1,2,3 BlackCat EasyReads Level 1	Building Blocks Library Level 5	Oxford Classic Tales B1,B2
Level 1	Foundation Series-Levels 4,5 Building Blocks Library Level 6	Oxford Bookworms Starter Page Turners Level 1	Penguin Readers Easystart Macmillan Starter
Level 2	BlackCat EasyReads Level 2 Foundation Series-Levels 6,7	Scholastic Popcorn 1 Page Turners Level 2	Oxford Classic Tales E1 Penguin Readers Level 1
Level 3	BlackCat EasyReads Level 3 Oxford Dominoes Starter	Building Blocks Library Level 7 Scholastic Starter	Macmillan Beginner Scholastic Popcorn 2 Oxford Classic Tales E2
Level 4	Cambridge Starter Helbling Level 1	Page Turners Level 3 Building Blocks Library Level 8	Penguin Readers Level 2 Oxford Classic Tales E3
Level 5	BlackCat EasyReads Levels 4,5 Helbling Readers 1	Scholastic Level 1	Scholastic Popcorn 2
Level 6	Cambridge Level 1 Helbling Level 2	Oxford Bookworms Stage 1 Page Turners Level 4,5	Cengage Footprints 800 Building Blocks Library Level 9
Level 7	BlackCat GreenApple Starter Helbling Readers 2	BlackCat R&T Stage 1 Scholastic Level 2	Compass Classics 1 Oxford Dominoes 1
Level 8	Cambridge Level 2 Helbling Level 3	Oxford Bookworms Stage 2 Page Turners Level 6	Penguin Readers Level 3 Macmillan Elementary
Level 9	BlackCat GreenApple Level 1 Helbling Readers 3	BlackCat R&T Stage 2 Oxford Dominoes 2	Cengage Footprints 1000 Compass Classics 2
Native speaker young reader series and youth literature			
Starter	Oxford Reading Tree Stage 5		
Level 1	Oxford Reading Tree Stages 6,7	Henry and Mudge	I Can Read Level 1
Level 2	I Can Read Level 2 Oxford Reading Tree Stages 8,9	All Aboard Reading Level 2	Step Into Reading 3
Level 3		Nate the Great	Step Into Reading 4
Level 4	Marvin Redpost Magic Tree House	Rainbow Magic Cam Jensen	Step Into Reading 5 Tashi
Level 5	A to Z Mysteries Secret Seven	Zack Junie B. Jones	Terry Deary's Historical Fiction The Secrets of Droon

図1 Kyoto Scale 別リーダー・シリーズ (mreader.org より)

Scale のレベル (以下リーダーレベル) は、まず出版社が公表している見出し語数レベルから大まかにレベル分けしたものを、一冊ずつネイティブスピーカーが点検し、レベルを調節して、Starter から Level 9 までの 10 段階に分け、決定される。最終的には「読みやすさレベル」¹⁾ という日本多読学会と英語多読研究会が共同で作成した指標を参照して、Kyoto Scale のレベル分けの信頼性を保っている。学生は入学時のプレースメントテストにより英語必修科目のクラス分けが行われ、そのクラスごとに読むリーダーレベルと目標語数が設定される。

2.4 リーダーのターゲットレベルと、実際の学生のレベル

Kyoto Scale の各リーダーレベルがどれくらいレベルの読者を想定しているかは表1の通りである。本学文化学部の場合、学生のがほとんどがCEFRレベルでA1からB1に当てはまる。

実際に2013年度入学の学生のプレースメントテスト (TOEIC Bridge) のスコアと設定された多読のレベルは表2に示す。2013年度の場合、Level 1 以外はターゲットレベルより実際の学生のスコアのほうがやや高くなっているが、やや易しいと思うリーダーを読むことを推奨していることや、またリーダーが簡単すぎると感じる場合は教員が個々の学生のレベルを調節することもできることを考えるとプレースメントに問題はなかったと考えられる。

表1 リーダーレベルと読者の英語レベル (mreader.org より)

Kyoto Scale	TOEIC	TOEFL	CEFR	読みやすさ レベル
Starter	200	300		0.4-0.7
Level 1	250	350	A1	0.7-1.0
Level 2	280	365		1.0-1.2
Level 3	320	380		1.2-1.6
Level 4	380	400	A2	1.6-2.6
Level 5	450	430		2.6-3.2
Level 6	500	450	B1	3.2-3.6

表2 2013年入学学生が多読レベル

KSU Level	T-Bridge	TOEIC	TOEFL
Level 1	88-120	230-310	376-404
Level 2	122-128	310-345	404-449
Level 3	130-138	345-395	449-432
Level 4	140-146	395-470	432-460
Level 5	148-162	470-570	460-494

学生はGR, YR どちらのリーダー・シリーズからでも読む本を選択できるが、読書後のこの2種類のリーダーに対する学生の反応は大きく異なる。多くの学生がYRを読むことに「難しさ」を感じると報告するのである。Post-Quiz Questionnaireの分析 (Robb, Clafin, & Gillis-Furutaka, 2014) の結果でも、特にYRのLevel 1とLevel 3のリーダーでは高いDifficultyの回答が得られた。リーダーのレベル分けは、GR, YRともに2.3節で述べた方法で行っており、レベルの指標にずれがあるとは考えられない。では、この「難しさ」の認識とはどこから来るのか、GRとYRにはどのような違いがあるのか、解明しようというのが本研究の出発点である。

学生が感じる「難しさ」の要因には、Claridge (2005) がGRの書き換えの対象としてあげている低頻度語や文の長さ・複雑さ、文法的な難しさなどがYRには含まれると推測される。また、文化的背景知識の不足や、YRに含まれるスラングや比喻表現の多さなども要因となり得るであろう。中條他 (2012) では、GRからは「多様な文法事項が含まれ、平易な語彙レベルで、かつ、文の長さやトピックも初級レベル英語学習者向けに適切な例文が多量に得られる」ことを確認しているが、GRとYRにおけるこれらの違いを客観的に示した研究はない。各出版社はリーダーの見出し語数などでレベル表示をしているが、その基準は出版社毎に異なり、各社で独自に開発されたレベル別語彙表は公表されておらず、書き換えの基準も明確に示され

ない。実際に多読学習において学習者がリーダーを難しいと感じるのはテキストのどのような特徴によるものかを解明するためには実証的な調査が必要であるが、本研究は、その第一歩として、コーパス言語学的アプローチにより「難しさ」の要因になり得る文構造、語法、意味などの要素を抽出することを目指す。

3. 方法論

3.1 リーダー・コーパスの概要

本研究では、GRとYRの比較のため、まずGraded Reader CorpusとYouth Reader Corpusの2つのコーパスを構築した。両コーパスの基本情報を表3に示す。Graded Reader Corpusの構築には、図書館の貸し出し情報や学生の読書記録を元に、本学学生によく読まれたGR79冊を選び、ストーリーの部分のみを集め収録した。総語数約30万語のコーパスで、Kyoto ScaleのLevel 1～4のリーダーが含まれ、品詞タグは付けられていない。現状では、YRの方が図書館の蔵書数が少なく、貸し出し頻度もGRより低いものが多いので、Youth Reader Corpusの構築に当たっては、GR Corpusとレベル毎の語数のバランスを保ちつつ、貸し出し頻度の高いものをKyoto ScaleのLevel 1～4のリーダーから選んだ。また、YRは、特に低いレベルのリーダーはGRより短いものが多いので、ほぼ同じ規模の語数のコーパスを構築するにはより多くのリーダーを使用しなければならなかったため、結果的に143冊を収録した。品詞タグは付けられていない。両コーパスに収録したリーダーのリストは付録に付した。

表3 YRコーパスとGRコーパスの基本情報

	リーダー数		1冊の平均語数		総語数		異なり語数	
	YR	GR	YR	GR	YR	GR	YR	GR
L1	11	9	853.45	1039.00	10155	10553	820	994
L2	58	29	1147.62	2192.34	67924	68421	4037	2592
L3	51	20	1957.39	4840.05	101656	103184	5896	2295
L4	23	21	4843.00	5413.05	115848	115123	5520	2517
全体	143	79	2008.15	3587.39	295583	297281	9178	4490

3.2 分析に使用するツール

分析には現在利用可能な様々なサイトやツール、語彙表を利用して行った。主なものは以下の通りである。

ツール：AntConc (Mac版 Build 3.2.4), AntWordProfiler (Mac版 Build 1.3.1), Microsoft Word for Mac 2011 「読みやすさの評価」

語彙表：BNC 語彙頻度表, The BNC/COCA word family lists, JACET8000

AntConc²⁾はAnthony Laurence氏により開発され、無料で提供されているコンコーダンス・プログラムであるが、KWICコンコーダンスを表示するだけでなく、様々なコーパスデータを分析する機能が備えられている。本研究では、語彙頻度表を作成するためにWord List機能、対象のコーパスに特徴的に見られる語彙を抽出するためにKeyword List機能、頻出する言い回しやフレーズを抽出するためにClusters/N-grams機能を使用した。Keyword List機能では、Adam Kilgarriff氏のサイトで無料で提供されているレマ化されたBNCの語彙頻度表³⁾を使用した。

AntWordProfiler⁴⁾はAntConcと同じくAnthony Laurence氏により開発され、無料で提供されている語彙分析ツールで、コーパス中に含まれる語彙を、汎用の語彙表の語彙レベルの分類に照らしてその分布を表すことができる。本研究では、Paul Nation氏のサイト⁵⁾から無料でダウンロードできる、BNCとCOCAに含まれる25000語を1000語毎にレベル分けしたBNC/COCA word family listsを汎用の語彙表として使用した⁶⁾。また、日本人学習者にとっての「難しさ」ということの解明を目指しているため、日本人にとって重要度の高い語を含めて作成されている語彙リストJACET8000もあわせて使用した。

リーダーの読みやすさを客観的に示すことができるか明らかにするためには、Microsoft Wordの文章校正機能の一つ機能である「読みやすさの評価」⁷⁾を使用した(図2)。これは文字数や単語数、単語の長さや文中に含まれる語数や受動態文の割合などを計算したり、Flesch

読みやすさの評価	
Counts	
Words	2072
Characters	9169
Paragraphs	87
Sentences	289
Averages	
Sentences per Paragraph	3.8
Words per Sentence	7.1
Characters per Word	4.3
Readability	
Passive Sentences	11%
Flesch Reading Ease	84.9
Flesch-Kincaid Grade Level	3.1

OK

図2 Microsoft Word「読みやすさの評価」の結果表示の例

Reading Ease (以下 FRE) や Flesch-Kincaid Grade Level (以下 FKG) などの読みやすさの指標を使ったりして、文章を評価するものである。FRE は、1 センテンス当たりの平均単語数と 1 単語当たりの平均シラブル数を元に計算され、0～100 の数字で表される。数字が多いほど読みやすいとされる。FKG は 1 センテンス当たりの平均単語数と 1 単語当たりの平均シラブル数を元に計算され、米国の学年でレベルが表示される。例えば 3.1 という値はおおよそ小学 3 年生レベルということになり、値が大きいほど難しいことになる。本研究ではリーダー 1 冊を 1 ファイルとして、1 ファイル毎に読みやすさの評価を行い、全体とレベル毎の平均値を出した。

4. 分析

以下、上記 2 つのコーパスに見られる YR と GR の相違点を分析した結果を述べる。

4.1 異なり語数

まず、単純な語数の比較から、大きな違いを見出すことができる。AntConc の Word List 機能を使用して、各コーパスの語彙頻度表を作成し、異なり語数を計算した。表 3 を見ると分かるとおり、各コーパスの異なり語数には全体で 2 倍近くの差がある。レベル毎に見てみると、Level 1 では YR、GR いずれも 1000 語以下であるが、GR は Level 2～4 まで 2000 語台半ばあたりで語数が顕著に大きくなることはないが、YR では Level 2 では既に約 4000 語、Level 3、4 では 5000 語台へと急激な増加をしていることがわかる。異なり語数の大きな差は「難しさ」の一因となり得るであろう。ただし、YR の一冊あたりに含まれる語数が少ないこと、そのため 2 つのコーパスサイズを揃えた際、YR の冊数が増えて GR より多くなっていることが異なり語数の差に影響していることも考えられる。

4.2 語彙レベル

次に、実際に使われている語彙のレベルに違いはあるか調査した。両コーパスの語彙を BNC/COCA や JACET8000 の語彙レベルと照らし合わせて、各レベルの Token (総語数) と Type (異なり語数) による語彙レベルを計測した (表 4-6)。

表 4 BNC/COCA による YR、GR の Token のレベルごとの分布 (%)

	YR	GR	YR-L1	YR-L2	YR-L3	YR-L4	GR-L1	GR-L2	GR-L3	GR-L4
1st 1000	82.46	89.10	88.12	84.04	83.75	79.91	86.92	86.97	90.50	89.32
2nd 1000	5.16	2.28	3.12	5.24	4.87	5.55	3.21	2.58	2.30	2.00
3rd 1000	0.88	0.41	0.43	0.62	0.87	1.09	0.65	0.52	0.30	0.42

表5 BNC/COCAによるYR, GRのTypeのレベルごとの分布(%)

BNC+COCA	YR	GR	YR-L1	YR-L2	YR-L3	YR-L4	GR-L1	GR-L2	GR-L3	GR-L4
1st 1000	29.31	43.77	73.54	45.69	37.83	38.09	75.13	58.6	59.92	58.01
2nd 1000	18.61	13.62	10.73	19.17	20.13	19.38	9.43	13.16	10.07	11.03
3rd 1000	5.74	3.31	0.98	3.47	5.15	5.53	1.3	3.23	1.95	1.65
4th 1000	7.90	3.58	3.17	6.04	6.83	8.17	1.3	2.81	1.91	2.99
固有名詞	4.22	15.91	1.59	4.51	5.34	3.37	5.52	10.73	12.46	11.46

BNC/COCAによる結果(表4)を見てみると、全体では基本1000語の割合がGRの方が高く、YRのほうが2000語レベルの語を多く含んでいることが分かる。リーダーレベル別で見ると、Level 1ではYRの方が基本1000語をやや多く含んでいるが、その後のレベルではGRの方が多くなり、その差はレベルがあがるにつれて大きくなっている。GRはレベルが上がるにつれてむしろ基本1000語の割合が高くなる傾向にある。逆に2000語レベルの語の割合は、YRのレベルがあがるにつれて高くなる傾向があるのに対して、GRでは下がっている。3000語レベルの語を含む割合はどちらのコーパスでも1%前後と少ないが、Level 1以外のすべてのレベルでYRの方が高い割合を示している。

Typeによる結果をみると、同様の傾向をより顕著に見ることができる(表5)。全体的な基本1000語の割合の差は大きく、YRのレベル上昇に伴う基本1000語の減少は急激で、特に高いレベルでGRと大きく異なる。2000語、3000語、4000語レベルの語の割合はYRがGRを大きく上回っている。このことから、YRのほうがレベルがあがることに基本1000語の使用が減り、語彙レベルのあがり方が急であると言える。

日本人にとって重要度の高い語を含めて作成されている語彙リスト、JACET8000を用いて、同様の分析結果を行ったが、ここでもほぼ同じ傾向を見ることができる(表6)。GRの語彙レベルはリーダーレベルがあがることで大きな変化がない一方、YRはむしろGRより多くの基本語彙でスタートするが、リーダーレベルがあがるにつれ、より高いレベルの語の使用が増え、語彙レベルが着実に上がっていくと言えるであろう。

表6 JACET8000によるYR, GRのTokenのレベルごとの分布(%)

	YR	GR	YR-L1	YR-L2	YR-L3	YR-L4	GR-L1	GR-L2	GR-L3	GR-L4
1st 1000	77.30	79.73	82.11	79.11	78.40	74.62	75.26	78.82	79.84	80.58
2nd 1000	5.53	3.38	5.32	5.18	5.46	5.85	4.68	3.66	3.33	3.13
3rd 1000	3.22	1.50	3.13	3.62	2.84	3.28	2.65	1.64	1.43	1.37

4.3 読みやすさの評価

4.1-4.2節では、語彙レベルでの検討を行ったが、ここでは文章全体としての読みやすさの評価を比較する。1文に含まれる語数、1語中の文字数（単語の長さ）、受動態文の占める割合、FREとFKGのスコアから、文章評価を行い、レベル毎に平均値を出した結果を表7に示す。

表7 YR, GRのリーダーレベル別「読みやすさの評価」

	W/S		C/W		P.S.		FRE		FKG	
	YR	GR	YR	GR	YR	GR	YR	GR	YR	GR
L1	6.35	5.60	3.89	3.98	0.18	0.00	97.39	92.71	1.20	1.67
L2	7.14	6.42	4.03	4.05	1.26	0.41	92.60	90.13	2.04	2.21
L3	7.75	6.86	4.10	3.99	1.92	0.00	89.83	92.96	2.59	1.94
L4	7.13	7.90	4.24	3.93	1.09	0.00	90.36	90.49	2.38	3.44
平均	7.09	6.69	4.06	3.99	1.11	0.10	92.55	91.57	2.05	2.31

*W/S: Words per sentence, C/W: Characters per word, P.S.: Passive sentences, FRE: Flesch Reading Ease, FKG: Flesch-Kincaid Grade Level

1文中に含まれる語数はYRで6.35～7.75語、GRで5.60～7.90語と多少の違いはあるものの、極端に文の長さが異なるわけではない⁸⁾。YR、GRそれぞれの1語中の文字数の相違も1文字以下である。FREやFKGの数値を見ても、大きな差はなく、むしろYRのほうが読みやすいと評価されているレベルもある。これらの数値の中には、必ずしもリーダーレベルの上昇に従って数値が変化している訳ではない部分もある。これは、特にLevel 2-3はGR、YRともにコーパスに含まれるリーダーの数も多く多様であるので、読みやすさレベルの幅が大きいためと思われる。これらの数値ではリーダーの読みやすさの違いを厳密に表すことができないと判明した。

しかし、受動態文の割合には違いが見られる。両コーパスとも受動態文の割合は低いが、GRでは受動態文が全体の文の数の1%以上あるのはL2の4冊のみと非常に少ないが、一方YRでは約半数が1%以上の受動態文を持つ(71冊)。しかも、実際にどのような受動態文が使われているかを見てみると、構造に違いがあることが分かる。GRに見られた受動態文は、最も単純なbe動詞+過去分詞の形であるが、YRに見られる受動態文にはL1の段階から複雑な構造が見られる。以下に各コーパスに見られる受動態文の例を挙げる。

- GRの受動態の例

Your arm is broken. (Dangerous Journey), They were trapped. (The Long Tunnel), The theater was destroyed by a fire in 1613. (This is London)

- YRの受動態の例

L1: *The children were fed up. I know what that one is called. (The Outing)*

L2: *His stall needs to be cleaned. Do you see how his ears are pointed? (Ponies)*

L3: *She said that since they are called sandwiches, they should be made of sand. (Nate the Great and the Boring Beach Bag)*

L4: *Two other men in the expedition are chosen to head for the top. (To The Top), Stones covered with blue ice are scattered all over. (Ruby the Red Fairy)*

このように、関係詞節や名詞節の中で使われているものや、不定詞や助動詞と組み合わされて使われているものなど、YRで使用されていた受動態文には学習者には難しいと思われる構造が多く見られた。

4.4 YR に特徴的な語彙

YR に特徴的な語彙を抽出するために、比較コーパスと語彙リストと比較して特徴的である度合いを示す *Keyness* を算出した。本研究では *Keyness* の値の算出には *Log-likelihood* を用いた。YR, GR を BNC の語彙リストと比較した場合、比較的類似した結果となる (図3)。本コーパスでは *it's* や *don't* のようにアポストロフィで短縮されているものはそれぞれ、*it* と *s*, *don* と *t* のように別々の語として数える。また大文字と小文字は区別されない。両コーパスに共通して特徴的な語彙として表れたものを太字で示した。固有名詞の差を除けば、人称代名詞、接続詞 *and/but*, 前置詞など会話が多く含まれる小説や話し言葉の特徴と思われるような語がどちらのコーパスの特徴としても表れる。

<p>YR/BNC: a, and, annie, at, but, for, had, he, her, his, i, in, is, it, jack, of, on, out, s, said, she, t, that, they, to, up, was, we, with, you</p> <p>GR/BNC: a, and, at, but, can, for, he, her, him, his, i, in, is, it, me, my, of, on, s, said, she, t, that, there, they, to, very, was, with, you</p>

図3 YR, GR の BNC 語彙頻度表との比較

一方で YR コーパスを GR コーパスから作成した語彙頻度表と比較して見ると、異なる特徴が浮かび上がってくる。図4には *Keyness* の高い順に上位 50 語を示した。そのうち半数以上がリーダーの登場人物の固有名詞⁹⁾ (網掛けで表示) である。

YR/GR: annie, jack, rachel, kirsty, bear, as, biff, frog, toad, rosamond, said, if, sludge, arthur, marvin, around, casey, even, bedelia, fang, nate, amelia, tree, just, fairy, kipper, had, air, over, moon, little, up, wilma, wilf, cam, book, chip, amber, toward, rainbow, out, would, been, nadim, owl, still, off, grabbed, dad, teddy

図4 YR に特徴的な語 (GR との比較)

その上位 50 語から固有名詞を除いた 24 語を、機能語と内容語に分けて GR と頻度の割合 (YR/GR) が大きく異なるものの順に並べ、YR・GR コーパスでの頻度数を割合と共に示したのが表 8 である。

表 8 YR で特徴的に使われている語

機能語	YR	GR	YR/GR	内容語	YR	GR	YR/GR
toward	158	0	-	rainbow	157	0	-
even	288	9	32.00	grabbed	140	0	-
been	205	12	17.08	moon	215	7	30.71
if	472	49	9.63	fairy	238	8	29.75
around	448	55	8.15	air	226	9	25.11
as	903	137	6.59	dad	239	29	8.24
still	345	69	5.00	tree	536	119	4.50
would	389	86	4.52	book	477	122	3.91
just	680	194	3.51	little	848	324	2.62
off	489	143	3.42	said	6492	4409	1.47
over	635	195	3.26				
had	1160	506	2.29				
up	1481	756	1.96				
out	1395	726	1.92				

この 24 語のうち、rainbow, grabbed, moon, fairy の 4 語以外はすべて BNC/COCA 語彙リストの基本 1000 語である¹⁰⁾。それに関わらず、この表からは GR では全く使われていない語や、YR と比べて非常に頻度が低い語があることが分かる。以下に、表 8 から YR の割合が高い語を取り上げ、実際の用法を確認する。

• toward (YR : 158 回, GR : 0 回)

まず、toward については GR では一度も使用されていない。YR で toward が使われている例を見ると、大きく意味を変えることなく、よりシンプルな前置詞 to や for での代用が可能な場

合がほとんどであるので (*walked toward the door, working with his teammates toward a common goal* など), 書き換えの対象となったものと思われる。

• even (YR : 288 回, GR : 9 回)

GR では, even は主語, 動詞, 比較級を強調 (例 : *Even my mother ..., he even loved ..., even better ...*) する用法でのみ使用されているが, YR ではそれ以外に, *don't even know/notice* という定形表現が多く使われ, 接続詞や前置詞と繋がって *even if/though, even after/as ..., even at/in/with/without ...* のように多くの異なる表現が使われ, 学習者には解釈が難しいかもしれない *even than* や複雑な構造の *have never even been named* なども見られた。

• been (YR : 205 回, GR : 12 回)

been は GR では主に *have been to, there has been* という現在完了の決まった形で使用され, 過去完了での使用は *had been* は 1 回のみであったが, YR では過去完了で 61 回使用され, それ以外にも *have been Ving, have been Vp.p., must have been* などのように進行形, 受動態, 助動詞などと組み合わされた複雑な構造のものも多く見られた。

• if (YR : 472 回, GR : 49 回)

GR でも if は使用されているが, ほとんどが条件文での使用で, 名詞節の例はわずかであった (*ask* (1 回), *see* (5 回), *know* (1 回))。YR では, 条件文やこれらの名詞節に加えて, *if only, even if, as if, don't care if, what if, wonder if* などで使用されており, 語法の幅は GR よりもずっと広いことが分かった。

• around (YR : 448 回, GR : 55 回)

around の GR での用法は, *put one's arm(s) around* (13 回), *look around* (21 回) の二つに集中している。YR では *all around* (28 回), *around and around* (8 回) がもっとも頻度の高いフレーズだが, それ以外に特徴的なのは動詞句として使用される例で, GR には *look around* 以外ほとんど見られない (*drive/lying/walked around* 各 1 例) が, YR にはその動詞のパリエーションが多い (*blow, buzz, circle, dance, fly, glance, gather, go, run, sit, spin, swirl, wrap, whirl* など)。ほとんどの場合 around なしでも場面のおおよその状況が伝わらないわけではないが, あると動作をより精緻に描き出すことができ, また, 読み聞かせる際にはリズムがよく, 表現が生き生きする効果もあると思われる。そうした豊かな叙述表現が GR ではそぎ落とされる傾向があるようだ。

• would (YR : 389 回, GR : 86 回)

would は GR でも決して頻度が低いわけではないが, その用法は非常に限られている。主に *would like (to)* (32 回) *would you (like) ...?* (43 回) という 2 種類である。YR では, この 2 つの表現は少なく (*would like* (13 回), *would you?* (11 回)), それ以外の様々な用法 (推量, 時制の一致, 過去の習慣, 拒絶 (*would not/never*), 仮定法過去・過去完了, 意外 (*Why would you ...*)) などで多く使用されている。

このように GR と YR の頻度に大きな異なりがある語を調べてみると、頻度以上の差があることが分かる。YR ではそれら語が使用される語法の数・種類が多く、しかも、他の構造と組み合わせあってより複雑な文を形成しているということが明らかになった。

さらに、GR コーパスを YR コーパスから作成した語彙頻度表と比較し、GR で特徴的な語彙を算出した。図 5 には Keyness の高い順に上位 50 語を示した。

GR/YR: mr, i, says, very, man, me, you, money, want, christine, anna, but, raoul, him, angry, joe, woman, m, about, paul, holmes, bowen, sara, jenny, ok, yes, david, t, afraid, aladdin, room, young, mary, einon, inspector, police, nicky, old, john, going, alex, mike, sarah, tom, and, happy, opera, laurie, boy, earl

図 5 GR に特徴的な語彙 (YR との比較)

そのうち 21 語が特定のリーダー・シリーズの登場人物の固有名詞 (網掛けで表示) で、opera と earl も主に固有名詞として用いられている。固有名詞を除いたこれらの語はすべて BNC/COCA 語彙リストの基本 1000 語に含まれており、難解な語はない。GR では強調語は very に、感情は angry, afraid, happy にというように、多様な表現が基本語に書き換えられていることが推測される。

4.5 N-gram

では、YR にはどのような語の連鎖が特徴的に見られるのであろうか。本節では頻出する n-gram を観察することによってその特徴を探る。先述の通り本コーパスでは I'm や don't のようにアポストロフィで短縮されているものはそれぞれ、I と m, don と t のように別々の語として数える¹¹⁾。また、1冊のリーダーにおける特定の語や表現の頻度の偏りを避けるため、コーパス全体での生起頻度ではなく、その連鎖が生起するリーダー数を指標として用いた¹²⁾。2-gram のような短い連鎖を比較すると、前置詞・冠詞・接続詞・代名詞などの組み合わせで、両コーパスに共通する連鎖が多く見られ、特徴的なものを見つけるのは難しいので、本稿では YR と GR の相違が十分に観察可能であった 4-gram を分析した結果を述べる。出現したリーダー数の多いものから表 9 に列挙した。ここでは最も特徴がよく表れる上位 20 件のみを表示した。

GR には *I don't want/know/like/understand*, *I'm (not) going to* のように I を含む 4-gram が半分を占める。代名詞 you や we を含むものや、その直前にほとんどの場合 I が来ていると思われる *don't want to* などを含めると、1・2 人称を使った文が圧倒的に多いことが分かる (表 9 の 20 件中 16 件)。you はいずれも疑問文で使用されている (*do you want to, what are you doing, are you going to*)。これらのことから、GR では、会話で I を主体に語ったり、相手に問いかけたり

表9 YR,GR コーパスにおける高頻度 4-gram

	YR	GR		YR	GR
1	i don t know	i don t know	11	the edge of the	didn t want to
2	the top of the	i m going to	12	magic key began to	we re going to
3	the end of the	i don t want	13	the bottom of the	i don t understand
4	i don t think	don t want to	14	the magic key began	a lot of people
5	what are you doing	for a long time	15	at the end of	are you going to
6	key began to glow	do you want to	16	i don t like	and looked at the
7	a picture of a	what are you doing	17	in the middle of	i m not going
8	don t know said	i don t like	18	didn t want to	i want to see
9	i m going to	a lot of money	19	at the bottom of	out of the window
10	on the other side	but i don t	20	it was time for	m not going to

している文が多いことが分かる。一方 YR では、場所や状況を説明するような前置詞句 (*at the end of*, *in the middle of*, *on the other side* など) や、前置詞句の一部と思われる (*the top of the*, *the end of the*, *the edge of the* など) が多く見られ (表9中8件)、会話であることを強く思わせる 1・2 人称を含んだものは GR ほど多くない (直前に I が来ていると思われるものを含め6件)。このことから、YR の方が場面を説明する叙述的な表現が多いことが分かる。

5. まとめと今後の課題

本稿では GR と YR のコーパスを構築し、YR の特徴を探った。その結果、YR は GR と比較して、以下のような相違点があることが明らかになった。

- (1) 基本 1000 語の語彙の割合が少なくそれ以上のレベルの語の割合が大きい
- (2) レベル毎に語彙レベルが着実に上昇している
- (3) 受動態の割合が高く、しかも動詞句の構造が複雑である
- (4) 基本語に大きな頻度の差がある
- (5) 1 単語が複数の語法で用いられる
- (6) 複雑な文構造が見られる
- (7) 叙述的な表現が多い

これらは学習者が「難しい」と感じる要素となり得ると考える。今回はこれまで多読学習の現場で使用されてきたものの、その言語学的性質について十分には分析されてこなかった YR をコーパス言語学的に GR と比較することによって、その特徴を明らかにした。

今後の課題としては、まずタグ付けの必要性が挙げられる。今回は主に語彙表などと比較することで、特徴的な語彙を見つけ出し、その語について個別に調査をするという方策を採った

が、両コーパスにタグ付けを行えば、文の構造を指定した検索を行うことができ、より定量的に文の複雑性などを測ることができるようになる。今回はコーパス言語学的アプローチで「難しさ」の要因になり得る要素を抽出することができたので、各要素がどれくらい実際に学生が「難しさ」の感覚に影響しているのか、実証的に調査し解明していきたい。文構造の複雑さや、複数の語法があることなどは「難しさ」に直接繋がる可能性があるかと予測される。

また、YR以外にも、英語ネイティブスピーカー対象に書かれた書籍の中には、日本人英語学習者の多読学習に役立つ可能性があるものも多数存在する。日本のマンガの英訳や、本を読まない成人対象に書かれたシリーズ、難読症などの障害がある人のために書かれたリーダー、そして有名作家による子ども向けのストーリーなどである。これらについても、コーパスの構築を行い、それぞれの言語的特徴を見出し、GR、YR以外の英語学習に効果的なリーダーや、GRから一般書への架け橋になるようなシリーズの発掘を目指していきたい。

※本研究は2013～2015年度科学研究費補助金・基盤研究(C)課題番号25370664の援助を受けて行った。本稿の内容の一部は「英語コーパス学会第40回大会（於 熊本学園大学、2014年10月5日）の口頭発表に於いて報告したものである。

注

- 1) 「読みやすさレベル」についての詳細は<https://www.seg.co.jp/sss/YL/>を参照のこと。
- 2) http://www.laurenceanthony.net/antconc_index.html
- 3) <http://www.kilgarriff.co.uk/bnc-readme.html>
- 4) http://www.laurenceanthony.net/antconc_index.html
- 5) <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- 6) 実際にはGeneral Service List（基本1000語と2000語レベル）+ Academic Word List（570語）とコーパスに含まれる語彙レベルの分析も行った。これらの語彙表は作成された年代が異なるため結果が異なる可能性があったが、実際には大きな差が見られなかったため、本稿ではBNC/COCA word family lists, JACET8000と比較した結果のみを提示する。
- 7) 「読みやすさの評価」は必ずしも専門的な分析のために使われるのではなく、一般的な評価手段として使用されている指標である。本研究ではFREやFKGなどの一般的指標では大きな相違が見られない2種類のテキスト間でも、詳細に分析すれば、読みやすさに影響を及ぼす可能性のある大きな相違が見出されるということを示すため、この評価方法を分析に含めた。
- 8) Words per Sentenceの計算は、文の途中で改行記号が入っていると、ピリオドがなくても一文と計算されるので注意が必要であるが、GR、YRコーパス共に文の途中で改行が入らないよう後処理を行っているため、本稿のデータについてはその問題はない。
- 9) frogやowlなどの普通名詞のように見える語も固有名詞として使用されている。
- 10) rainbowは5000語レベル、fairy、grabbed、moonは2000語レベル。
- 11) ただし、COCAのn-gram頻度表では語句切りが異なり、doとn'tと分けられる。
- 12) それでも、シリーズを通して繰り返し使われる*The magic key began to glow.*などの決まり文句からのn-gramは排除できなかった。

参考文献

- Clafin, M. (2012) "Bridging the Gap Between Readers and Native Speaker Literature." *Extensive Reading World Congress Proceedings* 1: 156–159.
- Claridge, G. (2005) "Simplification in graded readers: Measuring the authenticity of graded texts." *Reading in a Foreign Language* 17(2) [online]. URL: <http://nflrc.hawaii.edu/rfl/October2005/claridge/claridge.html>.
- Honeyfield, J. (1977) "Simplification." *TESOL Quarterly* 11(4): 431–440.
- LEVELING CRITERIA. Reading A-Z [online]. URL: <http://www.readinga-z.com/readinga-z-levels/leveling-criteria/>
- Robb T., Clafin M., and Gillis-Furutaka A. (2014). "Evaluating Post-Quiz Responses to MoodleReader and MReader" The 7th Annual Extensive Reading Seminar [口頭発表].
- Swaffar, J. (1985) "Reading authentic texts in a foreign language." *The Modern Language Journal* 69: 115–134.
- Webb, S., and Macalister, J. (2013) "Is text written for children useful for L2 extensive reading?" *TESOL Quarterly* 47(2): 300–322.
- 大学英語教育学会基本語改訂委員会編 (2003) 『大学英語教育学会基本語リスト JACET List of 8000 Basic Words』 大学英語教育学会.
- 中條清美, 赤瀬川史朗, 西垣知佳子, 横田賢司, 長谷川修治 (2012) 「LagoWordProfiler による英語 Graded Reader Corpus の Collocation/Colligation 頻度分析」『日本大学生産工学部研究報告 B, 文系』 45: 55–71.
- ロブ・トーマス, 加野まきみ (2010) 「授業時間外の学習時間の増大による英語力の向上」『ICT 活用教育方法研究』 13: 6–10.

〈付録〉

1. Youth Reader リスト (下線はシリーズ名)

Level 1 I Can Read: *Father Bear Comes Home, Little Bear, Little Bear's Friend, Little Bear's Visit, Oxford Reading Tree: In the Garden, Pirate Adventure, Red Planet, The Magic Key, The Outing, The Willow Pattern Plot, Land of the Dinosaurs*

Level 2 All Aboard Reading: *Gorillas, Knights, Martin Luther King Jr. and the March on Washington, Pink Snow and Other Weird Weather, Ponies, Red White and Blue The Story of the American Flag, Wagon Train, I Can Read: Amazing Dolphins, Amelia Bedelia, Big Max, Come Back Amelia Bedelia, Frog and Toad All Year, Frog and Toad are Friends, Frog and Toad Together, Here Comes the Strikeout, Mouse Soup, Owl at Home, Play Ball Amelia Bedelia, Small Pig, Zacks Alligator, Oxford Reading Tree: Dutch Adventure, Flood, Green Island, Key Trouble, Rescue, Storm Castle, Superdog, Survival Adventure, The Blue Eye, The Finest in the Land, The Flying Carpet, The Flying Machine, The Litter Queen, The Quest, Step into Reading: Abe Lincoln's Hat, Arthur Tricks the Tooth Fairy, Arthur's Fire Drill, Arthur's Reading Race, Barack Obama Out of Many One, Baseball Ballerina, Dinosaur Days, Eat My Dust, George Washington and the General's Dog, Johnny Appleseed, Little Witch Goes to School, Little Witch Learns to Read, Little Witch's Big Night, Pirate Mom, Samantha the Snob, The Bravest Dog Ever, Oxford Reading Tree: A Day in London, The Rainbow Machine, Viking Adventure, The Kidnappers, Pocket Money, The Evil Genie, Save Floppy, What Was It Like?*

Level 3 All Aboard Reading: *Alice in Wonderland, I Can Read: Amelia Bedelia 4 Mayor, Arthur's Birthday Party, Arthur's Camp-Out, Arthur's Funny Money, Arthur's Great Big Valentine, Bread and Jam for Frances, Magic Secrets, Minnie and Moo The Night of the Living Bed, Minnie and Moo Wanted Dead or Alive, Mouse Tales, Marvin Redpost: A Magic Crystal, Nate the Great: Nate the Great, Nate the Great and the Boring Beach Bag, Nate the Great and the Crunchy Christmas, Nate the Great and the Fishy Prize, Nate the Great and the Halloween Hunt, Nate the Great and the Hungry Book Club, Nate the Great and the Lost List, Nate the Great and the Missing Key, Nate the Great and the Monster Mess, Nate the Great and the Snowy Trail, Nate the Great and the Sticky Case, Nate the Great Talks Turkey, Puffin Easy to Read: Get Ready for Second Grade Amber Brown, Its a Fair Day Amber Brown, Its*

Justin Time Amber Brown, What a Trip Amber Brown, Step into Reading: 20,000 Baseball Cards Under the Sea, Ben Franklin and the Magic Squares, Choppers, Escape North The Story of Harriet Tubman, Helen Keller Courage in the Dark, How Not to Babysit Your Brother, How Not to Start Third Grade, Hungry Plants, No Tooth, No Quarter!, Pompeii Buried Alive, Soccer Sam, The Fairy Berry BakeOff, The Fly on the Ceiling, The Great Houdini, The Little Mermaid, The Mystery of the Pirate Ghost, The Titanic Lost and Found, Thomas Jefferson's Feast, Tuts Mummy, Volcanoes, Wild Cats, Walker: It Moved, My Dad the Hero

Level 4 Cam Jansen: *Cam Jansen The Mystery of the U.F.O., Magic Tree House: Afternoon on the Amazon, Dinosaurs before Dark, Dolphins at Daybreak, Ghost Town at Sundown, Lions at Lunchtime, Midnight on the Moon, Mummies in the Morning, Night of the Ninjas, Pirates Past Noon, Polar Bears Past Bedtime, Sunset of the Sabertooth, The Knight at Dawn, Tigers at Twilight, Tonight on the Titanic, Rainbow Magic: Amber the Orange Fairy, Fern the Green Fairy, Ruby the Red Fairy, Saffron the Yellow Fairy, Sky the Blue Fairy, Step into Reading: Basketballs Greatest Players, Moonwalk The First Trip to the Moon, To the Top*

2. Graded Reader リスト

Level 1 Macmillan: *Lost Ship, Lucky Number, Sara Says No!, The Umbrella, The Well, Oxford Bookworms: Orca, Red Roses, Sally's Phone, Vampire Killer*

Level 2 Cengage Foundations: *A Helping Hand, Do It!, Does He Love Me?, Let's Party!, Love Online, My Mom, the Movie Star, No, You Can't!, The Lost Wallet, Macmillan: Anna and the Fighter, Dangerous Journey, L. A. Raid, Marco, Money for a Motorbike, Picture Puzzle, Rich Man, Poor Man, The House on the Hill, The Long Tunnel, The Truth Machine, This is London, Penguin: Ali and his Camera, Girl Meets Boy, Karen and the Artist, Lisa in London, Marcel and the Shakespeare Letters, Marcel Goes to Hollywood, Mike's Lucky Day, Surfer!, The Barcelona Games, The Battle of Newton Road*

Level 3 Cambridge: *A Death in Oxford, Dirty Money, Let Me Out!, The Black Pearls, The Girl at the Window, The Penang File, What A Lottery!, Why?, Penguin: Alice In Wonderland, Dragonheart, Five Famous Fairy Tales, Freckles, Ghost Of Genny Castle, Lost In New York, Lost Love and Other Stories, Project Omega, The Railway Children, The Wave, Three Short Stories of Sherlock Holmes, Walkabout*

Level 4 Cambridge: *Bad Love, Help!, Hotel Casanova, Just Like a Movie, Next Door to Love, The Big Picture, Macmillan: The Phantom of the Opera, Oxford Bookworms: A Little Princess, Aladdin and the Enchanted Lamp, Last Chance, Little Lord Fauntleroy, Love or Money, Mary, Queen of Scots, Pocahontas, Remember Miranda, The Adventures of Tom Sawyer, The Elephant Man, The Phantom of the Opera, The Witches of Pendle, The Wizard of Oz, White Death*

Compiling and Analyzing Reader Corpora:

Toward Revealing Factors Affecting Learners' Sense of “Difficulty” in Extensive Reading

Makimi KANO

Abstract

Extensive Reading (ER) is widely accepted as an effective assignment in many university English curricula worldwide. Graded Readers (GRs) specifically designed for learners of English as a foreign language are generally used, but Youth Readers (YRs), written for native speaker children, are also used in some programs. However, students often find YRs more difficult even when they are categorized as being at the same level as GRs. Though YRs have been put into levels and used in many ER programs, the differences in vocabulary and word usage between GRs and YRs have never been objectively analyzed. To find out the differences, two kinds of corpora have been compiled and compared. As a result, some differences were observed. The YR corpus contains a lower percentage of the basic 1000-word-level vocabulary and a higher percentage of more advanced levels, and unlike the GR corpus, it shows a steady increase of vocabulary level as the reader levels go up. In the YR corpus, there is also a higher percentage of passive sentences and complicated sentence structures. Some basic words, such as *even*, *if*, *been* and *around* have much higher frequencies and usage varieties in YRs than in GRs. YRs contain more descriptive expressions in contrast to the narrative expressions seen in GRs. These characteristics of YRs may be considered as factors that affect learners' comprehension.

Keywords: Graded Readers, Youth Readers, vocabulary levels, passive voice, complex structures