

英語初級者向けコーパスデータとしての 教科書テキストの適性に関する研究

中條清美*, 西垣知佳子**, 山保太力***, 天野孝太郎****

Identifying the Suitability of Textbook English for Beginner-level Corpus Data

*Kiyomi CHUJO**, *Chikako NISHIGAKI***, *Motoyoshi YAMAHO**** and *Kotaro AMANO*****

Few attempts have been made to use corpora directly in the classroom, especially by foreign language teachers and learners, because of the difficulty learners have in understanding the concordance examples retrieved. Furthermore, very little research has examined how difficult corpora texts are for foreign language learners. This study will examine the reading grade level and word familiarity grade level of text samples from English high school textbooks used in Japan, China, Korea, and Taiwan, and various English e-texts available on the Internet. It aims to provide a collection of texts with rated difficulty level scores that may be particularly useful for beginner level English learners. It is hoped that using corpora that are more relevant may take corpus usage one step closer to its ideal application.

Keywords: English Textbooks, Difficulty Level, Readability, Grade Level, Corpus

1. はじめに

コーパスは、従来、言語学分野において研究ツールとして活用されてきたが、近年では、英語教育に活用して活用しようという試みが見られるようになっている (Braun, 2005; Breyer, 2006; Granath, 2009)^{1),2),3)}。コーパスを活用した英語学習はデータ駆動型学習 (DDL: Data-driven Learning) と呼ばれる。DDLでは、コンコーダンスと呼ばれるコーパス検索用ソフトを使って、ターゲットとする語や句を検索し、テキスト中での、その語の使用例を一覧にして多数観察することが可能である。DDLでは、学習者自らが検索結果を見て、文法規則や語彙の意味・用法等を発見し学習するという帰納的

な学習方法が用いられる (Johns, 1991)⁴⁾。DDLは学習者中心の指導法であり、言語に関する学習者の気づきを導き、自立的な学習活動を促すことによって動機付けと学習効果を増すと考えられている (Stevens, 1991; Johansson, 2009)^{5),6)}。

中條・内堀・西垣・宮崎 (2009)⁷⁾は日英パラレルコーパスを用いた実践研究において、DDLを活用した帰納的学習によるルールの取りだし (仮説形成) から始まり、教師が与える明示的説明、練習問題を活用した仮説検証を経てプロダクションに至る4ステップDDL指導法を提案し、その教育効果を実証した。DDLによる帰納的学習を通年の授業で経験した学習者から、「自分で調べているので記憶に残りやすい」「自分で検索するという自主性がある」という感想が寄せられ、能動的な学習法が学習

* 日本大学生産工学部教養・基礎科学系准教授

** 千葉大学教育学部教授

*** 日本大学大学院生産工学研究科博士前期課程数理情報工学専攻1年

**** 千葉大学大学院教育学研究科修士課程英語教育専攻2年

効果に結びつくことを確認した。

このように、DDLは学習者中心の帰納的学習を可能にする外国語学習法として期待されるものの、検索用に使われるコーパスの多くは、研究目的で構築され、成人母語話者の書きことばに基づいている。そのため、研究用に作成されたコーパスをそのまま教育現場で利用しようとすると、「検索結果の英文テキストが難しすぎる」ため、教材として機能しない。

こうしたコーパスの英文と学習者の英語レベルのギャップについて調査した研究に、中條・白井・内山・西垣・長谷川(2004)⁸⁾、Chujo, Utiyama, Nishigaki(2007)⁹⁾がある。彼らは、散文と新聞記事より成る日英パラレルコーパスの英語テキストの難易度を、「文の長さ」、「単文・複文数」、「受動態の割合」などの複数の指標を用いて調査した。その結果、指標として、「単語の長さ」と「リーダビリティ」が難易度の判断に適していることが報告された。また、初級学習者に適する英文テキストの数は少ないことも明らかにされた。さらに、Chujo and Utiyama(2006)¹⁰⁾などでは、上述の2指標の他に、テキストの難易度を推定する指標として、英語母語話者の「語彙習得学年」資料を用いる方法が有効であることを検証している。

教育効果の期待されるDDLを教育現場に導入するためには、学習者の英語の習熟度に適した英文テキストデータを提示する必要がある。しかしながら、これまでの調査で判明しているように、ウェブ上で公開されている入手可能なコーパスには、日本人の入門期学習者や初級学習者に適したレベルのコーパステキストを見つけることは容易でない。

そこで、我々は、学習者のレベルを考慮して作成されたテキストデータの候補として、学校英語教科書の英文テキストに着目した。学校教科書は、語彙や文構造を綿密にコントロールし、学習順序に配慮して執筆・編集され、学習者レベルに配慮が払われた英文テキストである。実際の言語使用に基づいた英文ではないため、本来コーパスが強みとする「真正性(authenticity)」¹¹⁾という点では劣るが、学習者の興味・関心、認知レベルに配慮されていることから、入門レベル、初級レベル学習者のための有力な言語データの候補であると考えられる。

筆者らは、大学生や大学院生の指導でDDLの指導効果を確認した(中條他, 2009; 西垣・中條・木島, 2010)^{12),13)}。その成果を踏まえて、中・高生の入門レベル、その上の初級レベル学習者の指導に応用したいと考え、教材開発と実践研究をスタートした(西垣・天野・吉森・中條, 2011)¹⁴⁾。入門レベル、初級レベルは、英語学習者の対象者数が多く、必要性の高いレベルと考える。

しかし、入門レベル、初級レベルDDL教材の開発には、上述のように使用する検索用の英文テキストの難易

度の問題が関わってくる。そこで将来的には、入門レベル、初級レベル学習者用の独自コーパスを作成し、より良い教材の開発に利用したいと考えている。

本研究では、入門レベル、初級レベルのコーパス構築の際に課題となる難易度の問題を解決するための参考データを得るという目的で、英語教科書の出現語彙を分析、調査する。さらにウェブ上で公開され、入手可能な小説や新聞記事などの英語散文データを併せて分析、調査し、より広範な基礎資料を得るようにした。

2. 研究の目的

本研究の目的は、教育的視点に基づいた客観的な指標を用いて、中・高英語検定教科書テキストと英語散文テキストの難易度を計測し、さらに、これらの英文テキストが英語初級者用の検索用コーパスとして使用可能かどうかを確認することであった。本研究の成果は、コーパスを活用した入門、および初級学習用DDL教材を開発する際の基礎的研究資料となるものである。

具体的には、日本・中国・韓国・台湾の中・高英語検定教科書テキストと、ウェブ上で公開されている散文テキストを収集して、1)英語リーダビリティ公式による読書学年、2)英語母語話者の語彙習得学年、という教育的視点に基づいた難易度判定として実績のある2つの指標を用いてテキストの難易度を調査した。

3. 研究の方法

3.1 使用したテキスト

3.1.1 日本・中国・韓国・台湾の学校英語教科書

調査には、日本・中国・韓国・台湾で使用された中・高英語検定教科書を計34冊収集し、教科書の英文をスキャナーで取り込み、人手で校正を行った。これらの教科書資料は、中條・西垣・長谷川・内山(2008)¹⁵⁾と西垣他(2011)¹⁶⁾で収集されたものを、新たにデータの入力基準を統一して再校正したものである。教科書は、一般的な傾向が得られるように、各国において採択率の高いものを選択した。選択の際には各国の有識者に問い合わせるとともに、赤井田(2001)¹⁷⁾、相川・赤井田(2002)¹⁸⁾、アダチ・楊(2003)¹⁹⁾、小沼(2003)²⁰⁾、大井・石川・田畑(2005)²¹⁾、相川・尾関・緑川・投野(2008)²²⁾、村上(2009)²³⁾、三菱総合研究所(2010)²⁴⁾等を参考にした。

使用した教科書は以下のとおりである。著者、出版社、出版年等の詳細はAppendixに記した。

① 日本

中学: *New Horizon English Course 1, 2, 3*

高校: *Unicorn English Course I, II, Reading*

② 中国

中学：英語（新標準）初中一年級，二年級，三年級，各上・下冊

高校：普通高中課程標準實驗教科書第一冊～第五冊

③ 韓国

中学：Middle School English 1, 2, 3

高校：High School English I, II

④ 台湾

中学：國民中學英語課本 English 1, 2, 3, 各上・下

高校：遠東新高中英文（一）～（六）

3.1.2 散文

難易度が調整されている教科書の英文テキストデータと比較して、自然な英文データの場合、難易度レベルはどのように分布しているかを教科書データと同じ指標を用いて計測して比較した。散文は、一般に公開されているものからランダムに抽出されたものとして、Chujo, et al (2007)²⁵⁾で使用された、①「日英対訳文対応付けデータ」(内山・高橋, 2003)²⁶⁾より32編の英語テキストと②「日英新聞記事対応付けデータ」(内山・井佐原, 2003)²⁷⁾から抽出した3サンプルのテキストであった。①は「Project Gutenberg」などで配布されている著作権が消滅した作品など再配布等が許可されている文書で、日英対訳文対応が付けられた古典的な小説や最近のコンピュータ関連のエッセイなどから構成されている。②は「読売新聞記事データ」の1989年9月から2001年12月までの「読売新聞」とThe Daily Yomiuri記事を対応付けたものであった。

3.2 調査方法

テキストの難易度を客観的に推定できる指標として、1) 英語リーダビリティ公式による「読書学年」、2) 英語母語話者の「語彙習得学年」の2指標を用いて調査した^{※1)}。2つの指標によるスコアを正確に算出するため、すべての英文テキストから人手で固有名詞、数字、略語等を取り除いた^{※2)}。

3.2.1 読書学年（リーダビリティ）

リーダビリティとは、「文章を読みやすくする要因、すなわち単語の難易、単語の長さ、センテンスの長さなどの要因を組み合わせ、公式に代入して計算し、その数字を読書学年レベルとするものである」と定義されており^{※3)}、算出結果は、通常、米国における生徒の読書能力を学年で表示した「読書学年レベル (reading grade level)」を予測する数値で表される。たとえば、リーダビリティ・スコアの「8.0」は米国の8年生の生徒がその文書・読み物を理解できることを意味する^{※3)}。米国では、最初のリーダビリティ公式が1928年に発表されて以来、リーダビリティの指標が教育界に広く普及しており、リーディングやESL (English as a Second Language) の担当教

師に教材の難易度判定のための指標として頻繁に利用されている²⁹⁾。現在は数種類の公式が組み込まれたソフトウェアが利用可能であり、リーダビリティ指標の実用性は高い。

本研究では、そのようなソフトウェアの1つであるReadability Calculations³⁰⁾を利用した。中條他(2004)³¹⁾では、このソフトで算出される9種類の公式について、リーダビリティ・スコアの算出に用いた要因、各公式に推奨されている最適対象学年、および、9種類の公式を適用した読書学年レベルの算出結果を照合して検討し、Flesch-Kincaid Formula³²⁾、SMOG Formula³³⁾、Fry Graph³⁴⁾の3指標がリーダビリティの予測に適切であることを報告した。本研究では、これらの3指標を試用した結果にもとづいて、読書学年を安定して算出したFlesch-Kincaid Formulaを用いることにした^{※4)}。Flesch-Kincaid Formulaは「語数、シラブル数、文数」の要因を組み合わせることで読書学年を算出する公式であり、国内外の多くの研究で用いられている指標である。なお、Fry Graphは一部の学校教科書の読書学年を表示しなかったため、また、SMOG Formulaは学校英語教科書のスコアを過度に高く表示したため本研究では採用しなかった。

3.2.2 語彙習得学年

英語母語話者の語彙習得学年資料であるThe Living Word Vocabularyは、Dale and O'Rourke (1981)³⁵⁾が、米国の4～16年生が基本的な語の意味を理解する学年を40,400項目にわたって調査したものである。この資料には米国の半数以上の生徒が当該単語の意味を理解できる学年が示されている。ただし、最低学年レベルは4年生になるため、本研究では4年生以下を1年から4年の学年に細分する資料としてHarris and Jacobson (1972)³⁶⁾のBasic Elementary Reading Vocabularyesを使用した。両資料に含まれない語は17年生として計算した。これらの資料は調査年代が古く、2000年代の英語テキストの難易度を測定するには理想的な資料とは言えないかもしれない。しかしながら、Hiebert (2005)³⁷⁾も指摘するように、包括的に米国の生徒の語彙習得学年を調査した信頼性における資料は他にない。そこで、本研究では、3.2.1のリーダビリティによる読書学年の計測結果とあわせて、2つの教育的視点から難易度レベルの観察を行うこととした。

4. 結果と考察

本稿では、日本人英語学習者に使用可能な難易度レベルのコーパステキストにどのようなものがあるのか、教科書コーパスと散文コーパスから実例を求めるという目標に鑑み、次の4段階のレベルを設定した。4段階のレ

ベルとは、読書学年および語彙習得学年の結果から得られたデータに基づいて、レベル1：日本の中学校教科書修了レベル、レベル2：日本の高校教科書「英語I・II」修了レベル、レベル3：日本の高校教科書「英語Reading」修了レベル、レベル4：それ以上の英語レベル、であった。

高校教科書を「英語I・II」と「英語Reading」に分ける理由として、まず、日本の高校英語の場合、同一の名前を持つ英語教科書シリーズ35シリーズ中11種が「英語I・II」から成り、他の24種は「英語I・II・Reading」から成る(中條・長谷川・西垣, 2008)³⁸⁾。大学進学者数の多くない高校では「英語I・II」で修了する場合もあり、「英語I・II」が一つの区切りと考えられる。一方、進学者数の多い高校では「英語I・II・Reading」を使用することが多い。本研究で使用しているUnicornシリーズは後者に属するものである。Unicornシリーズにおける「英語I・II」と「英語Reading」との間には、大幅な語数、レベル差があり、Unicornシリーズの「英語Reading」教科書は35シリーズ中で、異語数、延べ語数、学年レベルにおいて非常に高い(中條他, 2008)³⁹⁾。そのため、「英語I・II」と「英語Reading」を分けて、「英語Reading」修了時を一つの区分として取り扱うことは妥当と考えられた。

「読書学年」と「語彙習得学年」はともに単位は「(米国の) 学年」であるが、前述したように別個の基準に基づく指標である。各々の指標内でのスコアの整合性は取れているが、互いのスコアの表示範囲は異なる。そのため、同じサンプルに対する両指標のスコアは、同じ数値を表示するとは限らない。

Table 1には、日本の学校教科書を基準として、レベル1をホワイト、レベル2を薄いグレー、レベル3を濃いグレー、レベル4をブラックの4段階に分けて、読書学年と語彙習得学年を分類して示した。以下ではTable 1のように、色分けしたスコア表示を行う。

Table 2には、日本・中国・韓国・台湾の中・高英語検定教科書の各英語テキストについて、読書学年と語彙習得学年のスコアを示した。どちらも単位は「(米国) 学年」である。表には、参考データとして各教科書の異語数、延べ語数も示した。**Table 3**には散文テキストについて、読書学年および語彙習得学年のスコアを示した。表には、

各テキストのタイトル名、著者名、テキストのタイプ(N: narrative, E: expository)、テキストサンプルの異語数、延べ語数を示した。タイトルは読書学年のスコアにしたがって低い学年から高い学年の順に示した。No. 31~No. 33は英字新聞記事である。Table 2とTable 3は、読書学年と語彙習得学年の列をTable 1の段階区分に従い、レベル1, 2, 3, 4に分類して、色別に表示した。

Table 2のレベル分けを見ると、読書学年と語彙習得学年は大よそ一致しているように見える。Table 3のレベル分けを見ると、読書学年の方が語彙習得学年よりも高い学年レベルに評価される傾向が見られる。両指標のスコアの相関係数を求めると、Table 2の教科書テキストの場合は $r=0.91$ 、Table 3の散文テキストの場合は $r=0.78$ であり、両指標には、高い相関が確認できた。

Table 2を見ると、教科書テキストの読書学年は0.1~6.6、語彙習得学年は2.1~5.0の間に分布しており、評価結果は薄いグレーで表示されたレベル2に集まっている。また、Table 3から、散文テキストの読書学年は1.4~12.8、語彙習得学年は2.2~6.3の範囲に分布しており、教科書テキストよりも難易度レベルの幅が広いことがわかる。散文テキストでは読書学年のレベル1に評価されたものはなく、評価結果は薄いグレーのレベル2とブラックのレベル4が多い。語彙習得学年の結果は濃いグレーのレベル3と判定されたものが多い。

読書学年と語彙習得学年のレベル分布を明確にするため、**Table 4**には、学校教科書と散文テキストに分けて、各レベル別のテキストの割合とその実数を示した。

Fig. 1は2つの指標間の分布の関係を見やすいように図示したものである。学校教科書テキストでは、読書学年と語彙習得学年の両指標においてレベル2(各々68%, 64%)に属するものが多かった。散文テキストでは、読書学年ではレベル4(46%)、語彙習得学年ではレベル3(49%)がもっとも多く属するレベルであることが明確に示された。

教科書テキストではレベル1とレベル2を合わせると、読書学年で89%、語彙習得学年で78%を占めており、教科書テキストは、英語I, IIを学習した高校2年生を対象とするコーパステキストとしてほぼ適切なレベルと考えられる。散文テキストでは、レベル1は、読書学年

Table 1 Overview of Level Definitions

レベル	日本の教科書レベル	読書学年	語彙習得学年
レベル1	日本の中学校教科書修了レベル	US reading grade 1.3 以下	US grade 2.5 以下
レベル2	日本の高校「英語I・II」修了レベル	US reading grade 1.4~5.5	US grade 2.6~3.9
レベル3	日本の高校「英語Reading」修了レベル	US reading grade 5.6~6.5	US grade 4.0~5.0
レベル4	日本の高校「英語Reading」修了レベル以上	US reading grade 6.6 以上	US grade 5.1 以上

Table 2 The Reading Grade Level and Word Familiarity Grade Level of Each Textbook

国名	学校	教科書名	異語数	延べ語数	読書学年	語彙習得学年
日本	中学	NEW HORIZON ENGLISH COURSE 1	278	1,195	0.2	2.1
		NEW HORIZON ENGLISH COURSE 2	466	2,507	0.8	2.2
		NEW HORIZON ENGLISH COURSE 3	463	2,446	1.3	2.5
	高校	UNICORN ENGLISH COURSE I	1,064	6,439	5.2	3.4
		UNICORN ENGLISH COURSE II	1,505	10,443	5.5	3.9
		UNICORN ENGLISH COURSE READING	2,461	15,894	6.5	5.0
中国	中学	英語(新標準)初中一年級上・下冊	735	13,542	1.6	2.9
		英語(新標準)初中二年級上・下冊	1,270	26,455	2.1	3.1
		英語(新標準)初中三年級上・下冊	1,654	31,239	2.5	3.6
	高校	普通高中課程標準實驗教科書第一冊(必修1)	945	7,271	3.2	3.5
		普通高中課程標準實驗教科書第二冊(必修2)	940	6,978	3.2	3.5
		普通高中課程標準實驗教科書第三冊(必修3)	852	6,571	3.4	3.4
		普通高中課程標準實驗教科書第四冊(必修4)	1,208	8,079	3.1	3.6
		普通高中課程標準實驗教科書第五冊(必修5)	1,349	10,278	3.3	3.8
韓国	中学	Middle School English 1	580	4,432	1.3	2.6
		Middle School English 2	757	6,067	1.7	2.7
		Middle School English 3	1,098	13,085	2.5	3.4
	高校	High School English I	1,586	9,851	5.7	4.2
		High School English II	1,727	10,158	6.6	4.5
台湾	中学	國民中學英語課本 English 1 上・下	384	3,041	0.1	2.3
		國民中學英語課本 English 2 上・下	706	6,010	1.0	2.7
		國民中學英語課本 English 3 上・下	918	6,326	1.9	2.8
	高校	遠東新高中英文(一)	918	4,844	3.3	3.1
		遠東新高中英文(二)	1,072	5,701	3.9	3.8
		遠東新高中英文(三)	1,277	5,959	3.8	3.9
		遠東新高中英文(四)	1,454	7,078	4.4	4.0
		遠東新高中英文(五)	1,568	7,900	4.5	4.5
		遠東新高中英文(六)	1,670	8,634	4.3	4.7



レベル1：日本の中学校教科書修了レベル
 レベル2：日本の高校「英語I・II」修了レベル
 レベル3：日本の高校「英語Reading」修了レベル
 レベル4：日本の高校「英語Reading」修了レベル以上

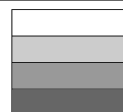
の指標で0%，語彙習得学年で6%であった。レベル2は、前者が43%，後者が26%であり、高校生を学習者対象にしたコーパステキストとして使用可能と考えられるテキストは多くはないものの、使えるテキストも存在することがわかる。

宮浦(2002)⁴⁰⁾によると、英文読解の研究では、散文テ

キストは物語文(N: narrative)と説明文(E: expository)に二分され、一般に物語文の読解は説明文に比べて易しい。その理由は、物語文で描かれる対象は日常生活での社会的関係であり、語彙もなじみのあるもので、物語の構造は、比較的平易なので、読み手は物語スキーマの知識を活かすことができるためであるとしている。本稿で

Table 3 The Reading Grade Level and Word Familiarity Grade Level of Each Text Collection

	タイトル	著者	テキスト タイプ	異語数	延べ語数	読書学年	語彙習得 学年
1	Jack and the Beanstalk	Joseph Jacobs	N	366	1,884	1.4	2.8
2	The Story of the Three Little Pigs	Joseph Jacobs	N	189	943	1.5	2.2
3	The Happy Prince	Oscar Wilde	N	476	1,731	2.4	3.1
4	What The Tortoise Said To Achilles	Lewis Carroll	N	304	1,146	3.1	4.1
5	The Last Leaf	O. Henry	N	602	1,894	3.3	4.1
6	The Little Match Girl	Hans Christian Andersen	N	339	997	3.4	3.0
7	The Gift of the Magi	O. Henry	N	587	1,849	3.5	3.8
8	The Tell-Tale Heart	Edgar Allan Poe	N	520	1,981	3.5	4.0
9	The Tempest	Mary Lamb	N	541	1,888	3.6	3.6
10	The Selfish Giant	Oscar Wilde	N	355	1,553	4.2	2.5
11	The Adventuer of Charles Augustus Milverton	Arthur Conan Doyle	N	641	2,189	4.4	4.0
12	To Build a Fire	Jack London	N	635	2,531	4.6	4.1
13	The Fad of the Fisherman	Gilbert K. Chesterton	N	602	2,106	4.6	3.7
14	In Midsummer Days	August Strindberg	N	642	2,485	4.9	3.6
15	The Shadow and the Flash	Jack London	N	694	2,169	5.3	4.5
16	A Tale about a Queer Client	Charles Dickens	N	806	2,704	5.8	4.5
17	A Horseman in the Sky	Ambrose Bierce	N	748	2,448	6.1	4.4
18	An Imperfect Conflagration	Ambrose Bierce	N	439	1,187	6.1	4.0
19	The Manager FAQ	Peter Seebach	E	503	2,054	6.1	4.3
20	The Man and the Snake	Ambrose Bierce	N	828	2,445	6.7	4.8
21	The Manager FAQ	Peter Seebach	E	522	2,266	6.8	4.4
22	Waiting for the Knock	Richard Stallman	E	368	1,078	6.8	4.1
23	RMS lecture at KTH (Sweden), 30. October 1986.	Richard M. Stallman	E	598	3,551	7.7	3.8
24	Interview with Karl Marx	R. Landor	E	737	2,495	7.8	4.7
25	Twenty Rules for Writing Detective Stories	S.S. Van Dine	E	454	1,289	8.2	5.2
26	A Modest Proposal for Preventing the Children...	Jonathan Swift	E	865	3,195	8.4	5.0
27	The Pragmatist of Free Software...	Hiroo Yamagata	E	392	1,611	8.7	4.4
28	Why Free Software is better than Open Source	Richard Stallman	E	461	1,987	8.7	4.3
29	The Fall of the House of Usher	Edgar Allan Poe	N	1,075	3,502	9.4	5.9
30	Why you should use the GNU FDL	Richard Stallman	E	94	160	9.8	3.2
31	The Daily Yomiuri Article Sample 1		E	822	2,180	12.0	5.2
32	The Daily Yomiuri Article Sample 2		E	792	2,084	2.4	5.2
33	The Daily Yomiuri Article Sample 3		E	878	2,378	2.5	5.2
34	Discourse on the Method of Rightly Conducting the Reason, ...	Rene Descartes	E	839	4,144	12.7	5.3
35	The Instinct of Workmanship and the Iirksomeness of Labor	Thorstein Veblen	E	808	3,551	12.8	6.3


 レベル1：日本の中学校教科書修了レベル
 レベル2：日本の高校「英語 I・II」修了レベル
 レベル3：日本の高校「英語 Reading」修了レベル
 レベル4：日本の高校「英語 Reading」修了レベル以上

レベル2に分類されるテキストの多くは、小説や童話など物語文(N)が多い。物語文は初級レベル学習者に使用可能なコーパステキストを供給してくれるということが期待できる。

しかしながら、今回指標に用いた読書学年と語彙習得学年の指標は英語母語話者を基準としたものである。母語話者にとっては、身近であるために容易な語であっても、英語を外国語として学ぶ日本の学習者にはそうした語彙を学ぶ機会は少なく、未知語となることもあるので、

物語文をDDL用のコーパスとして活用するには、教科書テキストデータとは異なる注意を払う必要もあろう。

これに対し、Table 3において説明文(E)に分類される新聞記事、学術論文、解説文、投稿文、演説原稿、エッセイ等は、読書学年の指標における評価レベルは高く、レベル4の上級者向けテキストレベルである。英文読解の研究においても、説明文の読解は比較的難しいとされている。その理由としては、題材が読み手の知識範囲を超えていることが多く、用語の意味のアクセスが困難を

Table 4 Comparing English Text Difficulty by Two Indices

	読書学年				語彙習得学年			
	レベル 1	レベル 2	レベル 3	レベル 4	レベル 1	レベル 2	レベル 3	レベル 4
学校教科書	21% (6冊)	68% (19冊)	7% (2冊)	4% (1冊)	14% (4冊)	64% (18冊)	21% (6冊)	0% (0冊)
散文	0% (0編)	43% (15編)	11% (4編)	46% (16編)	6% (2編)	26% (9編)	49% (17編)	20% (7編)

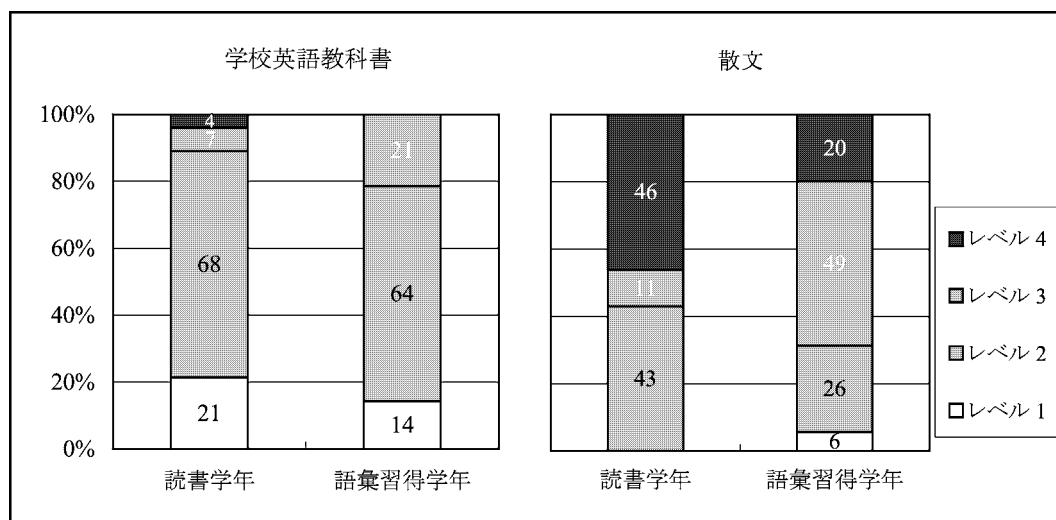


Fig. 1 English Text Difficulty by Two Indices

伴い、利用できる関連知識も少ない等さまざまな要因が考えられている(宮浦, 2002)⁴¹⁾。語彙習得学年においても、説明文は物語文より学年レベルの高い語の割合が多いことが読み取れる。外国語としての英語教育においてはとりわけ語彙の影響が大きいので、テキストを選択するには注意が必要である(堀場, 2002)⁴²⁾。

以上の結果に基づいて、1)レベル1：日本の中学校教科書修了レベル、2)レベル2：日本の高校教科書「英語II」修了レベル、3)レベル3：日本の高校教科書「英語 Reading」修了レベル、4)レベル4：レベル3以上のレベル、の各レベルの学習に使用可能なコーパステキストにはどのようなテキスト例があるか以下に具体的に考察する。

1)レベル1の日本の中学校教科書修了レベルの学習者に使用可能なテキスト例は、読書学年の指標では、学校英語教科書の6冊、内訳は日本の中1から中3の教科書と韓国の中1、台湾の中1と中2の教科書である。散文は圏外であり、使用できるものがない。語彙習得学年から見ると、教科書テキストでは日本の3冊と台湾の中1、散文では *The Story of the Three Little Pigs* と *The Selfish Giant* の2編である。このように、中学校英語教科書修了レベルの学習者に適したコーパステキストは非常に少ない。DDL 用教材開発の際には、多読指導に用いられる graded readers や幼児物語等を調査して、レベル

1のテキストを探す必要があると思われる⁴³⁾。

2)レベル2の日本の高校教科書「英語II」修了レベルの学習者に使用可能なテキストには、教科書テキストの78%が含まれる。散文の中で使用可能なテキストはNo. 1からNo. 15までの15編 (*The Happy Prince* や *The Last Leaf* などの物語文のテキスト)であり、語彙習得学年で見てもほぼ同様の結果であった。

3)レベル3の *Unicorn Reading* 修了レベル学習者には、韓国の *English II* を除く96%の教科書が使用可能なテキスト例となる。散文テキストは読書学年から見るとNo. 19までの散文の54%が使用可能、語彙習得学年から見ると散文の81%がコーパステキストの候補として考えられる。

4)レベル4の英語上級レベル学習者は、教科書、散文のすべてのテキストがコーパステキストとして使用可能な候補となる。とりわけ、散文テキストでは、読書学年から見て散文全体の46%にあたるテキストがこのレベル4の学習者のみが使用可能なテキストに該当し、それらの大部分は説明文(E)である。

5. まとめ

コーパスの教育利用を困難にしている原因の1つに、コーパスを構成するテキストの英文が難し過ぎることが

ある。学習者がコンコーダンスの出力結果である英文を理解してはじめて、文法の規則や語彙の意味・用法の発見学習が可能となり、DDL用の言語データとして機能する。そのため、DDL教材用のコーパスには、学習者の習熟度レベルに合致したテキストを使用する必要がある。そこで、本研究では、DDL導入の障碍となっている英語コーパステキストの難易度に関する基礎的研究として、日本・中国・韓国・台湾の中・高英語検定教科書テキストと、小説などの公開散文テキストの難易度レベルを測定し、比較することによって英語入門レベル、初級レベルのコーパステキストとして適切なテキストの具体例を把握しようと試みた。

英語教科書34冊(上巻と下巻で構成されるものをまとめると28冊になる)と英語散文35編を構成する英文テキストについて、読書学年と語彙習得学年という2つの指標を用いてテキストの難易度を調査した。その結果、以下のことが導かれた。

- ・2つのいずれの指標を用いてもそれぞれの基準にもとづいてテキストの難易度別の分類が可能である。ただし、本稿で用いた指標では、読書学年の方が語彙習得学年よりも高い学年にテキストレベルを評価する傾向がある。
- ・中学校教科書修了レベルの学習者に使用可能なテキストは非常に少ない。
- ・高校「英語I・II」を修了したレベルの学習者には、本稿で調査したアジアの英語教科書テキストのほぼ8割が使用可能圏内であり、散文のうち説明文は圏外であるが、物語文はコーパステキストとして使用可能である。
- ・高校「英語Reading」(*Unicorn Reading*)を修了したレベルの学習者は、アジア英語教科書の96%、散文の5割強が使用可能となる。
- ・散文の5割を説明文が占める。それらは高校「英語Reading」修了(高校3年生)以上のレベルである。

本研究の結果、日本の高校教科書を修了した学習者には、アジア各国の英語教科書テキストは学習者の習熟度レベルに合致したテキストと言えることが判明した。本研究の結果は、コーパスを外国語学習に利用する際、学習者のレベルに適した難易度のテキストを収集し活用するための基礎資料のひとつとなると考える。実際の活用例として、たとえば、コンコーダンスの出力結果の表示に際して、コーパステキストレベルI、II、III、IVをラジオボタンで選択できれば、よりユーザビリティの高いコーパス利用につながると考える。

また、外国語学習者のレベルを考慮した英文テキストデータには、教科書テキストのほかに、graded readersや母語話者用の幼児・児童向け教材も考えられる。そうした教育用テキストを広く収集し、入門レベル、初級レ

ベル学習者向けの教育用コーパスを構築する可能性を追求していきたい。そしてコーパスを活用した外国語学習DDL教材開発のための基礎資料としたいと考える。

謝辞：本研究は平成21-24年度科学研究費補助金基盤研究(B)(課題番号21320107)を受けて行われました。

注

注1) 筆者らの先行研究では、上記のリーダビリティ以外に、「単語の長さ」の指標も実用的で適切であることが報告されている。本研究では、当該先行研究で「単語の長さ」の算出に用いたプログラムを利用できなかったため、「単語の長さ」を調査の指標として使用しなかった。

注2) “Terms like *USA*, abbreviations like *lbs*, numerics like *123*, symbols like *+*, and monetary amounts like *\$3.87*—all are treated as words. With this in mind, for readability evaluation purposes, it is normally a good idea to either select sample text that is absent such entries, or text that has had such edited out before using it in Readability.” (*Readability Calculations*, p.6)

注3) “a score of 8.0 means that an eighth grader would understand the document” (“Readability and its Implications for Web Content Accessibility,” <http://wats.ca/resources/determiningreadability/1>)

注4) “By using and comparing their results a number of times—with one another, with comprehension test scores, and with your own judgment—you may soon determine which of the formulas is most reliable with your materials.” (*Readability Calculations*, p.2)

注5) 投野(2003)⁴³⁾には graded readers のコーパス化、海外の小学校レベルの教科書等をコーパスデータとして用いることが提案されている。

参考文献

- 1) Braun, S., From Pedagogically Relevant Corpora to Authentic Language Learning Contents, *Recall*, 17 (1), 2005, 47-64.
- 2) Breyer, Y., My Concordancer: Tailor-made Software for Language Learners and Teachers, in Braun, S., Kohn, K., and Mukherjee, J. (eds.), *Corpus Technology and Language Pedagogy* (New

- Resources, New Tools, New Methods*), Frankfurt am Main, Germany, Peter Lang, 2006, 157-276.
- 3) Granath, S., Who Benefits from Learning How to Use Corpora, in Aijmer, K. (ed.), *Corpora and Language Teaching*, Amsterdam, John Benjamins Publishing Co., 2009, 47-65.
 - 4) Johns, T., Should You Be Persuaded—Two Samples of Data-driven Learning Materials, *ELR Journal*, 4, 1991, 1-16.
 - 5) Stevens, V., Classroom Concordancing: Vocabulary Materials Derived From Relevant, Authentic Text, *English for Specific Purposes*, 10, 1991, 35-46.
 - 6) Johansson, S., Some Thoughts on Corpora and Second-language Acquisition, in Aijmer, K. (ed.), *Corpora and Language Teaching*, Amsterdam, John Benjamins Publishing Co., 2009, 33-44.
 - 7) 中條清美, 内堀朝子, 西垣知佳子, 宮崎海理, 「コーパスを利用した基礎文法指導とその評価」, 『日本大学生産工学部研究報告 B(文系)』, 42, 2009, 53-65.
 - 8) 中條清美, 白井篤義, 内山将夫, 西垣知佳子, 長谷川修治, 「日英パラレルコーパスを構成するテキストの難易度分類に関する研究」, 『日本大学生産工学部研究報告 B(文系)』, 37, 2004, 57-68.
 - 9) Chujo, K., Utiyama, M. and Nishigaki, C., Towards Building a Usable Corpus Collection for the ELT Classroom, in Hidalgo, E., Quereda, L. and Santana J. (eds.), *Corpora in the Foreign Language Classroom*, Amsterdam, Rodopi, 2007, 47-69.
 - 10) Chujo, K. and Utiyama, M., Selecting Level-specific Specialized Vocabulary Using Statistical Measures, *System*, 34 (2), 2006, 255-269.
 - 11) 米山朝二, 『英語教育指導法事典』, 東京, 研究社出版, 2003.
 - 12) 中條清美, 内堀朝子, 西垣知佳子, 宮崎海理(2009), 前掲論文.
 - 13) 西垣知佳子, 中條清美, 木島綾子, 「パラレルコーパスを利用した英語上級者用データ駆動型英語学習の実践の試み」, 『千葉大学教育学部研究紀要』, 58, 2010, 279-286.
 - 14) 西垣知佳子, 天野孝太郎, 吉森智大, 中條清美, 「中・高生のためのコンコーダンス・ラインを利用したデータ駆動型英語学習教材の開発の試み」, 『千葉大学教育学部研究紀要』, 59, 2011, 235-240.
 - 15) 中條清美, 西垣知佳子, 長谷川修治, 内山将夫, 「「ゆとり教育」時代の高校教科書語彙を考える—1980年代と2000年代の高校英語教科書語彙の比較分析からの考察—」, 『英語コーパス研究』, 15, 2008, 57-79.
 - 16) 西垣知佳子, 天野孝太郎, 吉森智大, 中條清美(2011), 前掲論文.
 - 17) 赤井田拓弥, 「アジアの英語教育 第1回 韓国の英語教育事情」, 1, *Kirihara Kyouiku Net, Web Peripatos*, Kirihara Shoten, Ltd., 2001, <http://www.kirihara-kyoiku.net/peripatos/01/01.html>
 - 18) 相川真佐夫, 赤井田拓弥, 「アジアの英語教育 第2回 台湾の英語教育事情」, *Kirihara Kyouiku Net, Web Peripatos*, 2, 2002, Kirihara Shoten, Ltd., http://www.kirihara-kyoiku.net/peripatos/02/01_3.html
 - 19) アグチ徹子, 楊肖力, 「中国と日本の英語教科書の内容の比較研究」, 『宮崎大学教育文化学部紀要 教育科学』, 9, 2003, 43-51.
 - 20) 小沼忠雄, 「アジアの英語教育 第4回 中国の英語教育事情」, *Kirihara Kyouiku Net, Web Peripatos*, Kirihara Shoten, Ltd., 4, 2003, <http://www.kirihara-kyoiku.net/peripatos/04/01.html>
 - 21) 大井恭子, 石川直美, 田畑光義, 「日本と韓国の中学校英語教科書の比較—論理的思考を育てるという観点から—」, 『千葉大学教育学部研究紀要』, 53, 2005, 249-258.
 - 22) 相川真佐夫, 尾関直子, 緑川日出子, 投野由紀夫, 「韓国, 中国, 台湾における英語教科書に関する調査研究」, 小池生夫(編)『第二言語習得研究を基盤とする小, 中, 高, 大の連携をはかる英語教育の先導的基礎研究』(平成16~平成19年度科学研究費補助金基盤研究(A)研究成果報告書, 2008, 71-109.
 - 23) 村上明, 「多次元分析による日本とアジア諸国の英語教科書研究」, 英語コーパス学会第33回大会, 2009年4月25日.
 - 24) 三菱総合研究所, 『平成21年度「教科書の質・量改善推進事業」—英語教科書改善・充実のための調査研究—』, 文部科学省委託事業報告書, 2010.
 - 25) Chujo, K., Utiyama, M. and Nishigaki, C. (2007), 前掲論文.
 - 26) 内山将夫, 高橋真弓, 「日英対訳文対応付けデータ」, 2003. <http://www2.nict.go.jp/x/x161/members/mutiyama/align/index.html>
 - 27) 内山将夫, 井佐原均, 「日英新聞の記事および文を対応付けるための高信頼性尺度」, 『自然言語処理』, 10(4), 2003, 201-220. <http://www2.nict.go.jp/jt/a132/members/mutiyama/jea/>
 - 28) 高梨庸雄, 卯城祐司, 『英語リーディング事典』, 東京, 研究社出版, 2000.
 - 29) Fry, E.B., Kress J.E., and Fountoukidids, D.L., *The Reading Teacher's Book of Lists*, West

- Nyack, New York, The Center for Applied Research in Education, 1993.
- 30) Micro Power and Light Co., *Readability Calculations*, 2003.
- 31) 中條清美, 白井篤義, 内山将夫, 西垣知佳子, 長谷川修治 (2004), 前掲論文.
- 32) Flesch R., *The Art of Readable Writing*. New York, Harper and Row, 1974. (as cited in Smith, C.R. and Smith, C.A. Patient Education Information: Readability of Prosthetic Publications, *Journal of Prosthetics & Orthotics*, 6 (4), 1994, 113-118.)
- 33) McLaughlin, G. SMOG Grading: A New Readability Formula, *Journal of Reading*, 12 (8), 1969, 639-646.
- 34) Fry, E.B., A Readability Formula That Saves Time, *Journal of Reading*, 11 (7), 1968, 265-271.
- 35) Dale, E. and O'Rourke, J., *The Living Word Vocabulary*, Chicago, World Book-Childcraft International, Inc., 1981.
- 36) Harris, A.J. and Jacobson, M.D., *Basic Elementary Reading Vocabularies*, New York, The Macmillan Company, 1972.
- 37) Hiebert, E.H., and Kamil, M.L., *Teaching and Learning Vocabulary*, Mahwah, NJ, Lawrence Erlbaum Associates, Publishers, 2005.
- 38) 中條清美, 長谷川修治, 西垣知佳子, 「1980年代と2000年代の高等学校英語教科書語彙」, 『日本大学生産工学部研究報告 B (文系)』, 41, 2008, 49-56.
- 39) 中條清美, 長谷川修治, 西垣知佳子 (2008), 前掲論文.
- 40) 宮浦国江, 「テキスト・タイプ」, 『英文読解のプロセスと指導』(津田塾大学言語文化研究所読解研究グループ編), 東京, 大修館書店, 2002, 118-136.
- 41) 宮浦国江 (2002), 前掲論文.
- 42) 堀場裕紀江, 「アセスメント」, 『英文読解のプロセスと指導』(津田塾大学言語文化研究所読解研究グループ編), 東京, 大修館書店, 2002, 243-265.
- 43) 投野由紀夫, 「コーパスを英語教育に生かす」, 『英語コーパス研究』, 10, 2003, 249-264.

Appendix Selected English Textbooks

国名	学校	教科書名	著者	出版地, 出版社
日本	中学	NEW HORIZON ENGLISH COURSE 1	笠島準一他 (2003)	東京: 東京書籍株式会社
		NEW HORIZON ENGLISH COURSE 2	笠島準一他 (2003)	東京: 東京書籍株式会社
		NEW HORIZON ENGLISH COURSE 3	笠島準一他 (2003)	東京: 東京書籍株式会社
	高校	UNICORN ENGLISH COURSE I	市川泰男他 (2006)	東京: 文英堂
		UNICORN ENGLISH COURSE II	市川泰男他 (2006)	東京: 文英堂
		UNICORN ENGLISH COURSE READING	市川泰男他 (2006)	東京: 文英堂
中国	中学	英语(新标准)初中一年級上册学生用	陈琳, Simon Greenall(英) (主编) (2008)	北京市: 外语教学与研究出版社
		英语(新标准)初中一年級下册学生用	陈琳, Simon Greenall(英) (主编) (2007)	北京市: 外语教学与研究出版社
		英语(新标准)初中二年級上册学生用	陈琳, Simon Greenall(英) (主编) (2008)	北京市: 外语教学与研究出版社
		英语(新标准)初中二年級下册学生用	陈琳, Simon Greenall(英) (主编) (2007)	北京市: 外语教学与研究出版社
		英语(新标准)初中三年級上册学生用	陈琳, Simon Greenall(英) (主编) (2008)	北京市: 外语教学与研究出版社
		英语(新标准)初中三年級下册学生用	陈琳, Simon Greenall(英) (主编) (2008)	北京市: 外语教学与研究出版社
	高校	普通高中課程標準實驗教科書第一冊(必修1)(供高中一年級上學期使用)	陈琳, Simon Greenall(英) (主编) (2008)	北京市: 外语教学与研究出版社
		普通高中課程標準實驗教科書第二冊(必修2)(供高中一年級下學期使用)	陈琳, Simon Greenall(英) (主编) (2008)	北京市: 外语教学与研究出版社
		普通高中課程標準實驗教科書第三冊(必修3)(供高中一年級上學期使用)	陈琳, Simon Greenall(英) (主编) (2008)	北京市: 外语教学与研究出版社
		普通高中課程標準實驗教科書第四冊(必修4)(供高中一年級下學期使用)	陈琳, Simon Greenall(英) (主编) (2008)	北京市: 外语教学与研究出版社
韩国	中学	Middle School English 1	김성곤 송미정 윤정미 김영 윤미정 (2007)	(주) 두산 http://textbook.doosandong.com
		Middle School English 2	김성곤 송미정 윤정미 김영 윤미정 (2007)	(주) 두산 http://textbook.doosandong.com
		Middle School English 3	김성곤 송미정 윤정미 김영 윤미정 (2007)	(주) 두산 http://textbook.doosandong.com
	高校	High School English I	김성곤 권익수 한정원 박용애 김재영, Phillip a.o' Neill (2005)	(주) 두산 http://textbook.doosandong.com
		High School English II	김성곤 권익수 한정원 박용애 김재영, Phillip a.o' Neill (2005)	(주) 두산 http://textbook.doosandong.com
台湾	中学	國民中學英語課本 English 1上	田超英 (主編) (2006)	臺北市: 佳音事業股份有限公司, 臺南市: 翰林出版事業股份有限公司
		國民中學英語課本 English 1下	田超英 (主編) (2007)	臺北市: 佳音事業股份有限公司, 臺南市: 翰林出版事業股份有限公司
		國民中學英語課本 English 2上	田超英 (主編) (2006)	臺北市: 佳音事業股份有限公司, 臺南市: 翰林出版事業股份有限公司
		國民中學英語課本 English 2下	田超英 (主編) (2007)	臺北市: 佳音事業股份有限公司, 臺南市: 翰林出版事業股份有限公司
		國民中學英語課本 English 3上	田超英 (主編) (2006)	臺北市: 佳音事業股份有限公司, 臺南市: 翰林出版事業股份有限公司
		國民中學英語課本 English 3下	田超英 (主編) (2007)	臺北市: 佳音事業股份有限公司, 臺南市: 翰林出版事業股份有限公司
	高校	遠東新高中英文(一)	施玉惠, 林茂松, 黃崇術, Sarah Brooks (編著) (2007)	臺北市: 遠東圖書股份有限公司
		遠東新高中英文(二)	施玉惠, 林茂松, 黃崇術, Sarah Brooks (編著) (2008)	臺北市: 遠東圖書股份有限公司
		遠東新高中英文(三)	施玉惠, 林茂松, 黃崇術, Sarah Brooks (編著) (2007)	臺北市: 遠東圖書股份有限公司
		遠東新高中英文(四)	施玉惠, 林茂松, 黃崇術, Sarah Brooks (編著) (2008)	臺北市: 遠東圖書股份有限公司
		遠東新高中英文(五)	施玉惠, 林茂松, 黃崇術, Sarah Brooks (編著) (2007)	臺北市: 遠東圖書股份有限公司
		遠東新高中英文(六)	施玉惠, 林茂松, 黃崇術, Sarah Brooks (編著) (2008)	臺北市: 遠東圖書股份有限公司

(H 23. 2. 9 受理)