

英語教科書コーパスの構築と利用—先行研究の概観

中條 清美

1. はじめに

教科書は数ある言語教材の中で、学校での教育・学習の核となるもので、教科書コーパスに基づく研究から教科書改善の手がかりが得られると言われる（石川, 2008: 165）。Johansson（2009: 40）は、教科書コーパスの研究は、言語教育における最も重要なインプットソースの1つに焦点をあてるものであると指摘しており、Römer（2004: 195）は、教科書コーパスデータと一般コーパスデータとの食い違いから教材改善の提案が可能であるとして、ドイツで多く使われている英語教科書を調査している。

実際に、中学および高校（以下、中高）教科書の英文を入力、校正して教科書コーパスを構築し、分析ツールを使用して出現語彙、特徴語、コロケーションなどを抽出してみると、英語教科書間に質的・量的な差があることが見えてくる。また、「時代の異なる教科書を同じ基準で比較」（小篠・江利川, 2004: 107）したり、外国の英語教科書や様々なコーパスなどの言語資料と比較することによって、日本の英語教科書の特徴が明確になってくる。

本稿では、最初に、「教科書コーパス」と「教材コーパス」の定義に触れ、続いて、英語教科書のコーパス化に際し留意すべき点を中村（2008）などの先行研究にもとづいて述べる。その後、我が国における教科書コーパスを利用した研究例、海外における研究例について報告する。

2. 「教科書コーパス」と「教材コーパス」の定義

日本における先行研究を見ていくと、例えば、「教科書コーパスから何が見えるか—方法論と中学校英語教科書の場合—」（中村, 2008）のように、教科書に特化して英文を集積した「教科書コーパス」を分析した研究が多い。一方、海外の研究例を見ていくと、「教科書コーパス」に当たる textbook corpus（Johansson,

2009: 40; Aijmer, 2009: 7) という言い方はあるものの、教科書の英文だけでなく補助教材などの英文も含めたコーパスを使用した研究が多い。Römer (2005: 175) はこのようなコーパスは“pedagogic corpus”に分類できるとコメントし、Hunston (2002: 16) の“pedagogic corpus”の定義「学習者が使用したもののすべて、すなわち、コースブック、リーダー、それに、テープなども含む」を引用した。¹ 本稿では、教科書と教科書以外の教材も含めたコーパスを「教材コーパス」とよぶ。一例をあげると、Bora and Fujiwara (2012-) の「小学校英語ウェブコンコーダンサー」の収録データは小学校外国語活動に関連する教材に加えて、指導書、教室英語の資料集、児童英語検定の書籍も収集した「教材コーパス」である。

3. 教科書コーパス作成の留意点

英語教科書をコーパス化する方法論について述べた文献には、鷹家・須賀 (1998)、中村・堀田 (2008)、赤野・堀・投野 (2014) などがある。これらに共通して指摘されている教科書コーパス作成の留意点は、1) 教科書をどう入手するか、2) どの部分をコーパス化するか、3) 編集のルールをどう決めるか、以上の3点である。これらについて、1980年代から2000年代の教科書コーパスを作成してきた筆者の経験もまじえて以下に述べる。

まず、1) 教科書の入手方法について、小篠・江利川 (2004: 169) は「教科書研究の出発点は第一に資料の入手である。資料なしには何も進まない。これが簡単なようでなかなか大変な作業である。…研究者同士が互いに協力し合って助け合うことが、もっとも実現の可能性が高い。」と、研究者の協力の必要性を強調した。

可能ならば研究チームに現場教員がいることが望ましい。例えば、英語教科書は約10年ごとに新しい学習指導要領に沿った内容に改訂されるが、学習指導要領の告示年と新教科書の使用開始年にずれがあり、さらに実施スケジュールが学校段階によって異なるため、教科書改訂の情報に精通した人材が必要である。また、小中高の現場教員の場合、出版社への教科書見本の請求やテキストファイルの入手要請など教科書を入手できる可能性は大学教員よりも高いと思われる。なお、発行されている教科書名は教科書協会のホームページ、² 教科書の採択部数は時事通信社の『内外教育』を参照するとよい。

次に、2) 教科書のどの部分をコーパス化するか、そして、3) テキストの

編集のルールをどう決めるかという問題である。これら2つの基準を策定するには、スタート地点として中村(2008)の「テキスト選択の基準」の7項目及び「編集のためのルール」の7項目を参考にした上で、利用目的やコーパス化する教科書の構成を考慮して独自の基準を設定するとよい。なお、テキストの整形、タグ付けなどについては赤野・堀・投野(2014)に詳しい。

4. 日本における利用例

本節では、我が国で行われた教科書コーパスの分析例の成果の一部を報告する。我が国では、1970年代以降、垣田他(1977)、小川(1977)、金田(1981)、三浦(1987)、中條・長谷川・竹蓋(1993)などによる、主に大型コンピュータを利用した中高英語教科書全シリーズのコーパス構築とその分析が行われた。2000年代にはパソコンを利用した研究が中心となり、塩見(2002)、杉浦(2002)、中條・西垣・長谷川・内山(2008)、中村(2008)、山添(2008)、石川(2008)などの英語教科書コーパスに関する研究成果が続々と公開された。

「日本の英語教科書を通時的に概観すれば、文法と語彙の統制が教科書編集の最大の関心事である」(小篠・江利川, 2004: 15)と指摘されるように、教科書コーパスの研究の中心は出現語彙の統計的調査であった。小篠・江利川(2004)の英語教科書の歴史的調査では、明治、大正、昭和、平成にわたり、各時代を代表する中高英語教科書各1シリーズの教科書を分析した結果、長らくにわたって教科書語彙の「異語数の減少」が続いていることが報告された。長谷川・中條(2004)では、1980年代、1990年代、2000年代という指導要領の変わった10年ごとの教科書語彙の年代比較を行った結果、2000年代にかけて、延べ語数、異語数、1語あたりの繰り返し回数は減少していることが明らかにされた。語彙学習における繰り返しの重要性を考慮すると、「ゆとり教育」を脱却後の2010年代にどう変化しているか、さらにグローバル化に対応した英語教育を掲げる学習指導要領³が実施される2020年代にどう変化するのかに関する継続調査が必要とされる。

小中高の各学校段階のテキストを縦断的に分析して語彙の学年比較をしてみるとわかることがある。中條・西垣・吉森・西岡(2007)では、文部科学省から外国語活動のテキストが発刊される以前にいち早く、⁴表1に示したように、5つの出版社から出された小学校外国語活動用テキストのコーパスを作成してシミュレーションを行った。1人の学習者が使用するテキストは各学校段階で1シリー

ズずつであるので、小学校テキスト5種のうち1種を、中学校教科書は *New Horizon* を、高校英語教科書は *Vivid* あるいは *Unicorn* の2種のうち1種を学習した場合を仮定して合計異語数を調査した。

表1 小中高を通して1人の学習者が出会うテキストの語彙数（異語数）

小学校 (A)	中学校 (B)	高校 (C)	小中高単純合計 (A) + (B) + (C)	重複を除いた 小中高累計
<i>One World Kids</i> (498語)	<i>New Horizon</i> (728語)	<i>Vivid</i> (1,917語)	3,143語	2,184語
		<i>Unicorn</i> (3,161語)	4,387語	3,348語
<i>Junior Columbus 21</i> (954語)		<i>Vivid</i> (1,917語)	3,599語	2,412語
		<i>Unicorn</i> (3,161語)	4,843語	3,514語
<i>Let's Have Fun!</i> (993語)		<i>Vivid</i> (1,917語)	3,638語	2,438語
		<i>Unicorn</i> (3,161語)	4,882語	3,554語
<i>Junior Horizon Hi, English!</i> (1,149語)		<i>Vivid</i> (1,917語)	3,794語	2,520語
		<i>Unicorn</i> (3,161語)	5,038語	3,626語
<i>Kids Crown</i> (2,084語)	<i>Vivid</i> (1,917語)	4,729語	3,144語	
	<i>Unicorn</i> (3,161語)	5,973語	4,155語	
平均			4,403語	3,090語

結果は、小学校の英語テキストと中高の英語教科書を通して出会う語彙数は、重複部分を除くと最終的に約3,000語程度にとどまることが判明した。小池(2006: 297)によると、中国、韓国、台湾では小中高で学習する基本語彙数は「6,000語か7,000語前後」が多いということなので3,000語程度の日本と大きな差がある。実用の英語力には7,000語程度の語彙が必要と報告されているので(中條・竹蓋, 1994), 「英語が使える日本人」の育成をめざすならば、大幅に語彙を増強する必要がある。

日本の教科書コーパスを外国の教科書コーパスや他の大規模コーパスと比較するとわかることがある。例えば、石川(2008)は、日本と韓国的高校英語教科書を FLOB コーパス、FROWN コーパス、映画コーパスと比較した。結果、日本の高校教科書は韓国と比較して、語彙量は少なく、語彙の多様性は低く、語彙レベルも低く、品詞の使用状況も意味領域分布も標準から逸脱していることが明確に示された。中條・内山・西垣(2005)は、日本の中高英語教科書コーパスと米

国の英語（国語）教科書コーパスを用いて、日米の教科書語彙で18種の実用英語（日常生活用語、医者との対話、ビジネストーク、経済ニュース、現代小説、サイエンスニュースなど）の英文テキストをどの程度カバーできるかを比較した。結果、日本の中高英語教科書は日常生活用語のカバー率が極端に低いことが判明した。このような欠点を補うために期待されるのが小学校英語である。中條・西垣（2010）は文部科学省から共通教材として配布された『英語ノート』の語彙の意味領域の分布率を調査した。結果、『英語ノート』では中高英語教科書で分布の低い領域となっている「生物」、「娯楽・スポーツ・ゲーム」、「飲食物」が高くなっていることから、小学校英語のテキストは中高英語教科書を補完するような英文で構成されていることが明らかに示された。

5. 海外における利用例

海外における教科書を扱った研究論文には、Römer（2004, 2005）がよく引用される。彼女はドイツの中等学校（grades 5-10）で一番広く使われている英語（外国語）教科書シリーズ及び付属の文法冊子から The German English as a Foreign Language Textbook Corpus (GEFL TC)（延べ語数108,424語）を構築した。Römer（2004）では、学校英語教科書における9種の法助動詞の使用状況を British National Corpus の話し言葉と比較した。その結果、例えば、ドイツの英語教科書では、will, can, must が過剰使用され、would, could, should, might が過少使用されていることが明白になり、教科書での法助動詞の提示順序の改善が提案された。Römer（2005）では同様の手法を用いて進行形の使用状況を調査し詳細な比較結果をモノグラフとして公開した。

Meunier and Gouverneur（2007）は大学の中級・上級学習者向け英語（外国語）テキストとワークブック（テープスクリプト、語彙練習問題、指示文）から724,174語の TeMa コーパスを構築し、⁵ 英語教科書における phraseology の扱いを調査した。彼らによると、第二言語習得、コーパス言語学、英語教育の3つの分野がそれぞれ別個に発展してきたため、その3分野を統合して教科書の構成や内容を評価する「教科書研究」は比較的新しいトピックであるとされる。確かに、海外の教科書研究を見ていくと、小規模な調査（e.g., Simpson and Mendis, 2003; Yuen, 2011）が中心で件数も多くないように思われた。⁶

6. まとめと今後の課題

本稿では、最初に、教科書コーパス及び教材コーパスの定義に触れ、続いて、英語教科書をコーパス化する際の留意点に言及した。その後、日本および海外における英語教科書コーパスの構築と利用例に関する成果の一部を報告した。第4節に述べたように日本ではこれまでに教科書コーパスに関する研究のデータが積み重ねられてきた。今後も継続して、2010年代、2020年代の教科書調査の実施が望まれる。それに加えて、「教科書コーパスから得られる情報は、語彙リストとそれから得られる基本的な統計が主なもので、コンコーダンスやコロケーションに関する情報を扱ったものはあまり見られない。」(中村, 2008: 121) という見解や, “It is only recently that such textbook studies have been carried out. The main focus however has been on grammar teaching.” (Meunier and Gouverneur, 2007: 123) と指摘されるように、語彙、文法、コロケーションに加えて、さらに談話分析、語用論、異文化理解など多様な方面の研究が待たれる。⁷ また、教科書だけでなく、教育効果を高めるための補助教材にも対象を広げた教材コーパスに関する研究を進めていくことも今後の課題と考える。

謝 辞

本稿は、英語コーパス学会第40回大会シンポジウム「英語教育・研究のための教材コーパスの構築と利用：実践例と課題」での発表を加筆修正したものである。本稿をまとめるにあたり、千葉大学教育学部の西垣知佳子氏に、貴重なご意見をいただきました。また、シンポジウムの参加者、司会者、および編集委員長、査読委員の方々からの貴重なご指摘をいただきました。ここに皆様に感謝いたします。本研究の一部は、平成25-28年度科学研究費補助金基盤研究(B)「多言語パラレルコーパスに基づくDDL オープンプラットフォームの構築と教育への応用」(25284108, 代表: 中條清美) の助成を受けて行われました。

注

1. Hunston (2002: 16) によると, “pedagogic corpus” は D. Willis (1993) の用語であるという。
2. 教科別発行教科書: <http://www.textbook.or.jp/about-us/publishing.html>
3. グローバル化に対応した英語教育改革実施計画: http://www.mext.go.jp/a_menu/kokusai/gaikokugo/_icsFiles/afieldfile/2014/01/31/1343704_01.pdf

4. 小学校外国語活動で使用する外国語活動教材は、2009年から2011年に『英語ノート』が配布され、2012年からは新たに *Hi, friends!* が作成された。
http://www.mext.go.jp/b_menu/houdou/24/01/1314922.htm
5. TeMa corpus: <http://www.uclouvain.be/en-cecl-tema.html>
6. Carroll, et al. (1971) は500万語の “published materials for grades 3 through 9” を収集して AHI コーパスを構築した。Harris and Jacobson (1972) は450万語の “textbooks for grades 1 through 6” を収集した。Zeno, et al. (1971) は1,700万語の “texts from kindergarten through college” を収集した。これらは英語（国語）教科書に加えて、それ以外の科目の教科書や様々な読み物も含んでいるため、主に英語教科書のテキストを収めたコーパスを対象とする本稿に含めなかった。
7. 文化の観点から英語教科書を分析した Yamanaka (2004), 相づちや繰り返しの機能を分析した山本 (2014), 教科書トピックを分析した Siegel (2014), コロケーション割合を考察した清水 (2014) などの研究がある。

参考文献

- Aijmer, K. (ed.) (2009) *Corpora and Language Teaching*. Amsterdam: John Benjamins Publishing Co.
- Bora, S. and Y. Fujiwara (2012) The Web Concordancer for Elementary School English. <http://corpus.gaikoku.aichi-edu.ac.jp/view/disclaimer-in-japanese/>
- Carroll, J. B., P. Davies and B. Richman (1971) *The American Heritage Word Frequency Book*. Boston: Houghton Mifflin Company and New York: American Heritage Publishing Company.
- Harris, A. and M. Jacobson (1972) *Basic Elementary Reading Vocabulary*. New York: The Macmillan Company.
- Hunston, S. (2002) *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Johansson, S. (2009) “Some Thoughts on Corpora and Second-language Acquisition.” In Aijmer, K. (ed.), *Corpora and Language Teaching*. Amsterdam: John Benjamins Publishing Co., pp. 33-44.
- Meunier, F. and C. Gouverneur (2007) “The Treatment of Phraseology in ELT Textbooks.” In Hidalgo, E., L. Quereda and J. Santana (eds.), *Corpora in the Foreign Language Classroom*. Amsterdam: Rodopi, pp. 119-139.
- Römer, U. (2004) “A Corpus-driven Approach to Modal Auxiliaries and Their Didactics.” In Sinclair, J. (ed.), *How to Use Corpora in Language Teaching*. Amsterdam: John Benjamins Publishing Co., pp. 185-199.
- Römer, U. (2005) *Progressives, Patterns, Pedagogy: A Corpus-driven Approach to English Progressive Forms, Functions, Contexts and Didactics*. Amsterdam: John Benjamins Publishing Co.
- Siegel, A. (2014) “What Should We Talk about? The Authenticity of Textbook Topics.” *ELTJ* 68, 4: 363-375.
- Simpson, R. and D. Mendis (2003) “A Corpus-Based Study of Idioms in Academic Speech.” *TESOL Quarterly* 37, 3: 419-441.
- Yamanaka, N. (2004) “An Evaluation of English Textbooks from the Viewpoint of Culture Based

- on the 2003 Ministry of Education's Course of Study Guidelines." *JACET Bulletin* 39: 87-101.
- Yuen, Ka-Ming (2011) "The Representation of Foreign Cultures in English Textbooks." *ELT J* 65, 4: 458-466.
- Zeno, S., S. H. Ivens, R. T. Millard and R. Duvvuri (1995) *The Educator's Word Frequency Guide*. New York: Touchstone Applied Science Associates.
- 赤野一郎・投野由紀夫・堀正広 (編) (2014) 『英語教師のためのコーパス活用ガイド』大修館書店。
- 石川慎一郎 (2008) 『英語コーパスと言語教育—データとしてのテキスト』大修館書店。
- 小川清 (1977) 『昭和53年度用中学校英語教科書使用語の初出ページ総合索引 (コンピュータによる)』静岡県教育委員会学校教育課。
- 小篠敏明・江利川春雄 (2004) 『英語教科書の歴史的研究』辞游社。
- 垣田直巳・三浦省五・友枝謙二・河田孝義 (1977) 『電子計算機による英語教科書の使用語彙総覧：中学校編』溪水社。
- 金田正也 (1981) 「英語学習必要語い量の事例研究 (中間報告)」『名古屋学院大学外国語教育紀要』第4号: 25-37。
- 小池生夫 (2006) 『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年度～平成19年度科学研究費補助金 (基盤研究 (A)) 中間報告書。
- 塩見知之 (2002) 『文部省検定済中学校・高等学校教科書に現れた英語の語彙』北星堂。
- 清水真弓 (2014) 「高校検定英語教科書のテキストに占めるコロケーション割合の考察」英語コーパス学会東支部《講習会・研究発表会》, 日本大学文理学部。
- 杉浦千早 (2002) 「高校英語教科書語彙リストの作成と使用語彙の検討」*Language Education & Technology* 39: 117-136.
- 鷹家秀史・須賀廣 (1998) 『実践コーパス言語学：英語教師のインターネット活用』桐原ユニ.
- 中條清美・長谷川修治・竹蓋幸生 (1993) 「日米英語教科書の比較研究から」『現代英語教育』第29巻第12号: 14-16.
- 中條清美・竹蓋幸生 (1994) 「現代英語のキーワード『プラス α 2000』—定義と効果の検証—」『千葉大学教育実践研究』第1号: 253-267.
- 中條清美・内山将夫・西垣知佳子 (2005) 「絵辞書と子供話し言葉コーパスを利用した日常生活語彙の選定」, 大学英語教育学会第44回全国大会, 玉川大学。
- 中條清美・西垣知佳子・吉森智大・西岡菜穂子 (2007) 「小, 中, 高一貫型英語語彙シラバス開発のための基礎研究」*Language Education & Technology* 44: 23-42.
- 中條清美・西垣知佳子・長谷川修治・内山将夫 (2008) 「『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』第15号: 57-79.
- 中條清美・西垣知佳子 (2010) 「小学校『英語ノート』の語彙分析」『英語コーパス研究』第17号: 115-126.
- 中村純作 (2008) 「教科書コーパスから何が見えるか—方法論と中学校英語教科書の場合—」中村純作・堀田秀吾 (編) 『コーパスと英語教育の接点』松柏社, pp. 121-150.
- 中村純作・堀田秀吾 (編) (2008) 『コーパスと英語教育の接点』松柏社。

- 長谷川修治・中條清美 (2004) 「学習指導要領の改訂に伴う学校英語教科書語彙の時代的変化—1980年代から現在まで—」 *Language Education & Technology* 41: 141-155.
- 三浦省五 (1987) 『英語教科書使用語彙：文部省検定済高等学校用 英語 I・英語 II・英語 IIB』 溪水社.
- 山添孝夫 (2008) 「教科書コーパスから何が見えるか—高等学校英語教科書の場合—」 中村純作・堀田秀吾 (編) 『コーパスと英語教育の接点』 松柏社, pp. 151-179.
- 山本智史 (2014) 「中学校英語教科書に見られる『相づち』と『繰り返し』—それぞれの機能に焦点を当てて—」 『関東甲信越英語教育学会誌』 第28巻: 109-120.

(日本大学 E-mail: chuujou.kiyomi@nihon-u.ac.jp)