

# コーパスと英語教育

投野由紀夫

## 一 はじめに

コーパスの言語教育への応用で最も進んでいるのは英語教育であろう。個々の応用例に関しては、個別言語でより優れた実例があるかもしれないが、英語教育は少なくとももつともその応用範囲が広く、実践例も豊富である。本稿では、コーパスの英語教育への応用として、まずその基本的な分類に関して整理し、続いて個別の分野の主な応用例とその成果を概観してみよう。

## 二 コーパスと英語教育…その方法論と応用分野

コーパスは実際に使用されたことばの集積である。母語話

者コーパスからは、以下のようなさまざまな「言語使用」の統計が得られる…

- (i) 単語・句の頻度・分布
- (ii) 単語・句の用例
- (iii) 当該単語を含む構造（構文）のパターン・頻度
- (iv) 当該単語の語義頻度
- (v) 当該単語を含むチャンク・連鎖などの頻度
- (vi) 当該単語のコロケーション情報  
など

これらを詳細に調査することによって、今まで十分考慮に入れられてこなかった、言語使用パターンの頻度や分布を勘案した、新しい言語教育のデザインや方法論を創出すること

が可能になりつつある。

その根本的な資料はおそらく学習すべき基礎語彙の用法基準の設定であろう。すなわち、言語教育の根幹となる「目標言語 (target language)」のより深い理解を促し、学習すべき言語材料の精選に資する資料の整備である。その資料を活用して新たなシラバスの構築や教材の作成などが期待されている。

また上記の発想は、母語話者の自然な言語使用をコーパスから観察する、という発想であったが、視点を変えてこれらの発想を母語話者データのみならず、学習者データに適用することにより、学習者の使用することばの特徴分析も行うことができる。この分野を「学習者コーパス」と呼ぶ。そして学習者コーパスと母語話者コーパスとの言語使用頻度の違い(過剰使用・過少使用・誤用)を研究することによって、中間言語(学習者が習得中の目標言語の中間的段階の文法システム)の特徴を探り、かつ段階を経てどのような言語獲得のメカニズムが観察されるのかも記述可能になる。これらの中間段階の言語使用を特徴づける習得指標が数多く抽出できるようになれば、それらを教育に生かそうという視点も出てくる。すなわち、目標言語の状態だけでなく、習得の順序や道筋もコーパ

スで明らかになる、という、方法論的にも新しい可能性を秘めた重要な分野なのである。

コーパスの教育利用を考える場合、主として次の三分野が重要である (Leach 1997):

- ① コーパスの直接利用
- ② コーパスの間接利用
- ③ 教育用コーパスの構築

①の直接利用は、対象となる学習者をコーパス・データに直接触れさせることで教育効果を狙おうとするアプローチであり、②の間接利用は、コーパスを資料としてさまざまな授資料や学習教材を作成して教育に活かすことである。前述のシラバス構築・教材作成、等は、基本的にはこの間接利用の分野になる。さらに③の教育用コーパスの構築は始まったばかりであるが、一般的なコーパスとは異なる、教育利用に特化した明確な目的をもった教育コーパスの構築という分野である。英語教育においては現状では圧倒的に②の間接利用が多い。以下の節ではまずこの間接利用について概観し、次に直接利用を解説、最後に教育用コーパスの構築に関して述べたい。

### 三 コーパスの間接利用

コーパスは大別すると以下のような領域で積極的に活用されている。

- (a) 英語辞典
- (b) 学習語彙表
- (c) 英文法書
- (d) 英会話教材
- (e) 英語テスト

以下、簡潔に説明しよう。

#### 三ー一 英語辞典

英語辞典はコーパスの最初の教材への応用例である。一九八七年に英国バーミンガム大学とコリンズ社が共同で出版した COBUILD English Dictionary がその先駆的存在である。全用例を英米の書き言葉・話し言葉の資料から構築した Birmingham Corpus (後の Bank of English) の実例から引用、コーパスの用法パターンの頻度分析結果に基づいた語義配列を行った。それまでの辞書とは一線を画するきわめて独自性の高い辞書となった。この COBUILD の後、英国の

主要辞典出版社はすべて自社内にコーパス資源を持つようになり、一九九五年には COBUILD, LDOCE, OALD, CIDE という四つの英英学習辞典がみな「コーパス準拠」を謳うようになった。二一世紀に入って、マクミランが参入し、現在はこの五つの辞典を Big5 といって互いにしのぎを削っている。特に今世紀に入ってから、各出版社のコーパス規模は平均三、四億語規模になっており、またケンブリッジやロングマンは自社独自の学習者コーパスを構築して、学習者の誤りデータなどを組み込む工夫をしている。

日本国内では、全般的にまだ英語辞典へのコーパス利用はあまり進んでいないが、その中で三省堂は日本人学習者向けに独自の社内用コーパスを構築、『ウィズダム英和辞典』で初めて本格的にコーパスを使用した語法記述を行った。また小学館は British National Corpus (BNC) WordBanks Online (COBUILD 作成の元となったコーパス)、PERC などの大型コーパスを社内用に利用するインタフェースを教育用に公開、「小学館コーパス・ネットワーク」として商用サービスを開始している。このデータを用いてコロケーション統計情報を利用した類語辞典『コーパス英語類語使い分け200』(小学館)などが出版されている。他社も徐々にコーパスを編

集に使用したことを何らかの形でアピールするようになってきているが、まだ使用の実態は中途半端なものが多い。

### 三二 学習語彙表

コーパスの頻度情報をもとに、必要な重要語彙に関する基本データを作成し、学習語彙の選定を行うという発想は海外よりも日本国内において特に盛んである。『Standard Vocabulary List(SVL)12000』(アルク)、『JACET8000』(大英英語教育学会基本語選定委員会)といった語彙表がコーパスをベースにして新しい手法で作成されている。どちらもベースとして、一般公開されている一億語の均衡コーパスBNCの見出し語リストを用いているが、SVL12000ではそれに従来利用されていた三〇以上の学習語彙表のデータをからめるという作業をしており、JACET8000では日本人が接する可能性の高い英文テキストのサブコーパスを大量に作成して、その語彙表とBNCとの語彙ランキング比較を行い、より日本人に適した語彙表に調整する、という手法を用いている。これらの語彙表からは、アルクが数多くの単語学習教材を出版したり、JACET8000は『ロングマン英和辞典』のリンク表示や、桐原書店からの単語集などに利用されている。

ただし、コーパスの学習語彙表作成への応用には、まだまだ研究の余地が残されている。特に、現在利用可能な語彙表は単純な「単語のリスト」でしかない。Michael WestのGeneral Service Listのような、単語の意味別頻度情報などをコーパスからきちんと調べた語彙表も必要だし、各単語の使い方(重要な構文やコロケーションなど)をコーパスから調査した「活用語彙」表なども今後に期待がかかる。

### 三三 英文法書

コーパスを利用した文法書の刊行も英語分野では相次いでおり、主要なものとしてCollins COBUILD English Grammar(1996), Longman Grammar of Written and Spoken English(1999), Cambridge Grammar of English(2006)などがある。コーパスを用いることで、より実際の言語使用の状況を反映した文法記述を特徴としている。たとえば、ロングマンの文法書では、新聞、小説、学術書、会話という四領域をカバーする四、〇〇〇万語のコーパスを使用して文法の使用頻度を比較、その分析結果を三五〇以上のグラフや表を駆使して、文体や書き言葉・話し言葉のモードの違いによる用法の相違などを詳しく分析している。ケンブ

リッジの文法書は特に話し言葉に特徴的な文法記述をしており、会話には特有の文法がある、という主張を前面に押し出している。こういったコース準拠の英文法書の記述は学校教育にはあまりまだ反映されていない。しかし徐々にではあるが、現場での教員の言語観も、極度に規範的なものの方からより記述的・確率的な見方に変わってきている。

### 三一四 英会話教材

コースを英会話教材に活用したもつとも顕著な例は、筆者がNHKで二〇〇三―二〇〇五年度に担当した『100語でスタート!英会話』であろう。一〇〇ユニット(1ユニット10分)のプログラムで一〇〇のキーワードをとりあげ、そのキーワードの単語の使い方集中して学習する語彙シラバスを特徴としている。「コース君」というCGのキャラクターが登場し、そのキーワードが使用されるコロケーション・ラッキングを紹介する。日本の一般家庭に「コース」という言葉を浸透させた、という意味でも画期的な番組であったといえよう。その後、関連する英会話教材(例:『コース練習帳』NHK出版)などが続々と発売され、日本はコース関連の英会話教材では他国に先駆けて、最も商業的に成功し

ている国であるといえよう。

海外ではロングマン、ケンブリッジが英会話教材の一部にコースを利用し始めている。特にロングマンは学習者コースを、ケンブリッジは会話コースを利用した特別コラムなどを会話教材の中にちりばめている。しかしながら、上記『100語』のような完全な語彙ベースのシラバスで作られた総合会話教材はまだ海外では存在していない。

### 三一五 英語テスト

テスト開発にコースを利用する発想がこの一〇年ほどでかなり具体的になってきている。米国のETSではテスト受験者が書いた大量の英作文データをコース分析した結果をもとにCriterionという作文自動評価システムを作成している。特定のテーマを設定して書かせた自由英作文を、コンピューターが内容、文法、語彙などの側面で評価判定を下し、自動で添削して返却する、というシステムである。ETSはこれ以外に、米国の大学レベルで学生が触れると想定される学術英語コースを構築し、それに基づくテスト開発も進めている。

また英国のCambridge ESOLは自社の開発しているさまざまな英語運用能力試験のテスト結果のうち英作文データ

を Cambridge University Press の辞典編集部と共同で大量にコーパス化しており、Cambridge Learner Corpusとして現在三、〇〇〇万語以上の規模を誇る。さらにこのコーパスを補強すべく、The English Profileという大規模プログラムを計画中で、世界中の英語学習者のコーパスを作つて、教育・研究・評価に活かすという構想を持っている。

### 三一六 間接利用…その他の可能性

コーパスの間接利用では、上記のような分野のほかに、英語教科書への応用が期待されているが、まだきちんとした活用例は見られない。しかし、可能性としては、三一二で述べた語彙表をより精密化し、語彙と構造・表現したい意味が習得レベルごとに導入時期を明示した語彙シラバスを設計し、それに基づいた教科書が作成できれば素晴らしい。それらの提示順序の研究に学習者コーパス・データなどが有効利用される可能性が大いにある。

また、効果が期待されるのは教員養成・研修などのトレーニングにおける利用である。英語教師にコーパスの分析結果をもとにした英語の基礎的な骨格になる語彙やそれらの用法頻度に関する客観的データを提示することで、英語教師とし

て重視すべき項目や指導の指針を明確に得ることができる。英語教師の言語観を変えることは言語教育観を変えることにもつながる。

### 四 コーパスの直接利用

コーパスの「直接利用」というのは、コーパスそのものを学習者に触れさせることによる効果をねらうものである。コーパスは大量のテキストの集積であるから、それをもとに当該単語の用例を大量に抽出したり、その単語とよく用いられる共起語のパターンを一覧できる。そのような「言葉の使い方」を調べる体験はおのずと「発見型」の学習になる。これを称して「データ駆動型学習 (Data Driven Learning: DDL)」と呼ぶ。DDLでは、従来の言語教育が「3P (提示: Presentation-練習: Practice-産出: Production)」という流れを取っていたのに対し、「観察 (Observation) - 分類 (Classification) - 一般化 (Generalization)」という学習者の主体性を活かしたことは特徴や規則性の発見を重視した教育法になる。さらにこのDDLでも、教師がある程度学習者の到達する答えを予測して行う収斂型学習 (convergent learning) と、教師の与えた課題から学習者各自が自由

に結論に達すればよい分岐型学習 (divergent learning) とに分けられる (Leach 1997)。

コーパスを実際に教育に直接利用する試みでは、すでに退職したバーミンガム大学の Tim Johns が DDL を提唱して、コンコーダンスを利用したさまざまな学習教材を web で公開している (<http://www.eisu2.bham.ac.uk/johnstf/index.html>)。最近では Tom Cobb の語彙学習ページ (The Compleat Lexical Tutor, <http://www.lexutor.ca/>) が、DDL の発送を受け継いでいろいろなコーパス資料や語彙学習のアイデアを web 教材として提供している。日本においては中條らの日英パラレル・コーパスを用いたライティング指導の例 (中條他、2005)、太田・日臺 (2006) の学習者コーパスを用いた語彙指導活用例などが注目される。

## 五 教育用コーパスの構築

コーパスと言語教育第三の接点は「教育用コーパスの構築」である。一般的に研究用の汎用コーパスは、母語話者の言語使用の分析には役立つが、コーパス本体を直接教育利用しようとする、その内容が専門的すぎたり難しすぎたりすることが多い。これは日本語の汎用コーパスを日本語学習者

が使った場合を考えてもその困難さは容易に想像できるであろう。コロケーション統計をパターンとして取り出す程度であればよいが、実際の例文をそのまま使おうとするとはほど上級レベルの学習者でないと敷居が高過ぎる。

そのため、「教育用コーパス」という発想が英語教育でも議論されている。これにはいくつかの可能性がある。それらを簡単に紹介する。

### 五―一 語彙制限コーパス

1つは語彙レベルを制限した英文コーパスである。これは英語教育の出版社が大量に作っている語彙制限リーダー (graded readers) のような語彙制限されたテキストをコーパス化するというアイデアである。それを活用した DDL の実践などは、まだ体系的には誰もやっておらず面白い可能性を秘めている。

また別の発想で、BNC のような汎用コーパスを語彙制限でフィルタリングする、という試みも行われている。東京外国語大学の佐野洋研究室の作成した BNC 用例検索では、コンコーダンスラインを小学館『プログレッシブ英和中辞典』の見出し語リストによってフィルタをかけ、基本語彙 1 万 5

千語の範囲で例文を選択して表示できる (<http://scn02.corpora.jp/~n-cube/n-cube.html>)。

### 五二 教科書コーパス

語彙制限コーパスと類似した発想であるが、より学校現場に密着しているのが「教科書コーパス」である。中学・高校での教科書データをコーパス化して、それを例文データベースのように使う、という発想である。著作権などがあり、公開は難しいが、いくつかの大学でこの教科書コーパス構築の試みが既に始まっている。教科書コーパスに関する論考は山添 (2006)、小中高の主要語彙分析には中條他 (2007) などがある。また、投野 (2008) のアジア諸国の英語教科書コーパスの分析も外国語政策的な観点から参考になろう。

### 五三 学習者コーパス

学習者コーパスは外国語学習者の産出データをコーパス化する発想である。産出データは大別すると会話データと作文データに分けられる。これらをコーパス化することで、言語習得データをコーパス分析できる。それによって、冒頭に述べたような、学習者言語の特徴を理解し、それに必要な言語

材料の配列や指導法の改善などが期待できる。

海外ではロングマン、ケンブリッジなどの出版社が辞書作成用に、数千万語規模の巨大な学習者コーパスを構築している。またヨーロッパを中心に International Corpus of Learner English (ICLE) というプロジェクトがあり、世界二〇カ国前後の母語の異なる英語専攻の大学生による英作文コーパスがよく知られている。また授業における教師と生徒のインタラクションをコーパス化しているプロジェクトとしてはシンガポールの教育省が中心で行っている SCORE プロジェクトがある (Hong 2005)。

国内の主要な英語学習者コーパスとして、まず作文コーパスでは、投野が中心で中高生の英作文データ約1万件を JEFLL Corpus として一般公開している (投野 2007: <http://scn02.corpora.jp/~jefll03/jeflltop.html>)。会話コーパスでは情報通信機構 (NICT) が構築した NICT JLE Corpus という二二〇〇人余りの口頭英語テストデータのコーパスが有名である (和泉他 2005)。

以上の分類からもわかるように、言語学習環境のインプット (教科書・教材)、インタラクション (授業観察)、アウトプット (作文・会話などの産出データ) の三種類のデータが相互



に連携しながら、一定期間、学校教育を追いかけるようなコース作りができると、シラバスの効果検証などに大きく貢献をするであろう。

## 六 おわりに

本稿では、進歩が著しいコース言語学の成果を言語教育に応用する事例として、英語教育を中心に論じてみた。まだこの分野は本格的に議論され始めて一〇年程度である。その間にコース資源の爆発的な増加とインターネット環境の激変で、議論の方向性もこれら技術的な進歩によりさまざまに影響を受けている。しかし、本質的には、どのような質と量の学習が学習者の頭の中に実際に起きるか、という点が重要である。その学習者中心の視点を失わないよう、今後もこの分野の動向を見守っていききたい。

### 【参考文献】

- 和泉絵美・井佐原均・内元清貴 (2005) 『日本人1200人の英語スピーキングコース』東京：アルク。
- 中條清美・西垣知佳子・内山将夫・原田康也・山崎淳史 (2005) 『日英パラレルコースを活用した英語語彙指導の試み』『日本大学生産工学部研究報告B』38：17-37
- 中條清美・吉森智大・長谷川修治・西垣知佳子・山崎淳史「高等

学校英語教科書の語彙」『日本大学生産工学部研究報告B』40：71-92

投野由紀夫 (2007) (編著) 『日本人中高生1万人の英語コース 中高生が書く英文の実態とその分析』東京：小学館。

投野由紀夫 (2008) 『アジア各国と日本の英語教科書比較』第3回教育再生懇談会公届審議第2セッション資料

(二〇〇八年五月一六日)

<http://www.kantei.go.jp/jp/singi/kyouiku/kondan/kaisai/dai3/2s-3gijisidai.html>

太田洋・日暮滋之 (2006) 『新しい語彙指導のカタチー学習者コースを活用して』東京：明治図書。

山添孝夫 (2006) 『教科書コースから何が見えるか：高等学校英語教科書の場合』『立命館言語文化研究』17 (4)：167-187

Hong, Huqing. (2005) SCoRE: A multimodal corpus database of education discourse in Singapore schools. *Proceedings of the Corpus Linguistics Conference Series*, vol11, No.1. (ISSN1747-9389) Available online.

Leech, Geoffrey. (1997) "Teaching and language corpora: a convergence". In Anne Wichmann, Steven Fligelstone, Tony McEnery, & Gerry Knowles (eds.) *Teaching and Language Corpora*. London: Longman.

【このの・ゆきお 東京外国語大学准教授】