

コーパスを英語教育に生かす

投野 由紀夫

英語コーパス研究

第10号

2003

英語コーパス学会

コーパスを英語教育に生かす

投野 由紀夫

1. はじめに

コーパスの言語教育への活用は現在コーパス言語学の領域の中でも非常に熱い議論が交わされている分野の一つである。それはコーパス言語学の領域としては応用分野であり、本質的なコーパス構築・分析の技術的な事柄とは性質を異にする。しかしながら、言語教育、特に英語教育は欧州においてもアジアにおいても大きな市場を形成しているので、出版社や教育産業界におけるコーパスの英語教育への応用は大きな注目を集めている。

本論は英語コーパス言語学の見地からコーパスの英語教育への応用を考える際に、どのような視点や研究課題が必要かを筆者なりに整理し今後の問題点を明らかにしようとするものである。

2. コーパス言語学と言語教育に関する世界の潮流：歴史的概観と現在

コーパスと言語教育との接点はコーパスをどう定義するかによって異なってくる。ここでは Kennedy (1998) を参考に電子化される以前のコーパス (pre-electronic corpora) と、電子化コーパス (electronic corpora) に分けて歴史的流れと現状を概観してみたい。

2.1 電子化以前

電子化以前のコーパス分析はほとんどが印刷媒体のテキストを人手でカウントするという大事業であった。言語教育の分野における主要な貢献は Thorndike (1921), Lorge (1949), West (1953) などに代表される教育語彙表の分野がまず挙げられよう。特に、Carnegie Project (Faucett et al. 1936) はデータから得られた頻度の高い語彙を学習項目として優先するという明確な方向付けを示している点で特筆に値する。この頃にすでに Thorndike は規模にして数百万語のコーパスを分野別にサンプリングして処理しており、当時は周期的だった頻度の分布 (range credit と呼んだ) に関する統計も算出している。また Lorge は基本語 570 語に関して語義頻度 (semantic count) を数えるなど、

現在でも資料的に貴重な語彙統計情報を人手で行っている。この成果が West の General Service List (1953) に結実しているのは周知のとおりである。

日本ではこの時期 Harold E. Palmer が残した業績を忘れてはならない。1913年から1936年まで彼は英語教授研究所（現在の語学教育研究所）の責任者として日本の英語教育の発展に多大な貢献をした。彼は学習語彙表をその用法や機能を考慮した主観的方法で選定し、後の Carnegie Project で Thorndike らの機械的に頻度を求めて重要度を決定する語彙表作成の方法論だけでは不十分であると強く改善を求め、語彙表作成に教育的観点を組み入れることに大きな役割を果たした (Cowie 1999: 17)。それだけでなく、*A Grammar of English Words* (1938) ではコーパス・データから機械的に抽出したものではないが、現在のコーパス言語学で頻繁に用いられるコロケーション (collocation) の概念や、最近の理論言語学でも改めて重要度が認識されている「単語の文法」の視点をすでにこの頃の著作に明確に持っていたのは特筆に値する。これら先達の偉業を経て学習英語辞典という新しいジャンルが日本で生まれ英国で真剣に市場的に開拓されたわけであるが、コーパス・データから何を読み取るか、という部分のかなりの重要な要素がこの黎明期の著作に見られることは瞠目すべきことと言えよう。

2.2 電子化以後

電子化以後のコーパスに焦点を置くとするならば、やはりコーパス言語学の方法論を本格的に言語教育に応用したのは John Sinclair 率いる Birmingham 大学の COBUILD Project (Sinclair 1987) がその最初であると言ってよいだろう。¹ 実際、COBUILD Project の結果、言語教育の分野でコーパスに準拠した教材一式が完備された。COBUILD Series として大小さまざまなレベルの一般および特殊辞典が発表されただけでなく、文法書 (*Collins COBUILD English Grammar*; *Collins COBUILD Grammar Patterns*)、語法書 (*Collins COBUILD English Usage*)、単語集 (*COBUILD Vocabulary Builders*) などが次々と出版され、我々はコーパス準拠の教材作成の見本を COBUILD の一連の産物を通して見ることができるようになる。

この流れとともに Birmingham 大学では、Tim Jones を中心としてコーパス・データを言語教育に用いる実践的な試みが行われた。Classroom concordancing あるいは Data Driven Learning (DDL) という用語は彼らの実践から生まれたといえよう。² 特に Tim Jones の website³ では語法・文法の規則性をコンコーダンス・ラインから読み取らせるような練習問題や、パラレル・コーパスを利用した翻訳の練習など、教室で実際に行った活動の教授資料や授業実践の論文が数多く報告されている。

90年代になると、コーパス言語学の学会も徐々に盛んになってくる。特にヨーロッパでコーパス言語学の先端的役割を果たしていた ICAME に集って

いた研究者らの間で、コーパスの調査結果を教育に応用し、コーパスそのものを言語教育や専門の言語学の教育に利用するというニーズが議論されるようになってきた。これを機会に、1994年に第1回の Teaching and Language Corpora (TALC) という国際会議が Lancaster 大学で開催された。この会議の主要な論文は Wichman et al. (1998) で読むことができる。その後、TALC は2年おきに開催され、2002年には第5回目の会議がイタリアの Bertinoro でもたれた。TALC に集まる研究者は最初コーパス言語学の専門家が多く、議論も新しいコーパスやツール類の紹介が多く一般の言語教育の研究者には敷居が高かったが、徐々に学際的な性格を帯びるようになり、現在は Michael McCarthy や Henry Widdowson といった応用言語学、言語教育分野の専門家でコーパスに興味のある研究者を交えた形で行われており、議論の中身も単純なコーパス利用の盲目的礼賛から、コーパスが提供するデータの役割に関する本質的な議論へと移行してきている。今後の動向が注目される国際会議である。

同様に米国においては、TALC98に参加したミシガン大学の Rita Simpson が呼びかけて北米応用コーパス言語学会 (North American Association for Applied Corpus Linguistics) が1999年に発足し今年で4回目の大会を行っている。この学会に出るとわかるのだが、アメリカの代表的な流れは Douglas Biber の Northern Arizona University (NAU) が作っていると言っただけであろう。Biber の register variation の研究方法は広くいろいろな分野に応用されており、毎年大会では関連する発表が必ず行われる。また同じ NAU の Randi Reppen を中心に進められている American National Corpus (ANC) の構築も大型プロジェクトとして注目を集めつつある。特に ANC はアメリカ英語の1億語規模のコーパスということで日本からも代表的な辞典出版社がコンソーシアムに参加しており、⁴ 今後そのデータを用いたさまざまな言語教材が日本でも現れることが予想される。またアメリカでは Michigan 大学の MICASE に代表されるような ESP (English for Specific Purposes) 関連のコーパスへの関心が高く、2002年に発足した Professional English Research Consortium (PERC) の動き⁵とともに、大規模専門分野英語のコーパスが今後 ESP あるいは EAP (English for Academic Purposes) の領域で広範囲に利用されることが期待されている。

3. コーパスの言語教育への応用：理想と現実

現在、上記のような国際学会などで議論の中心になっていることはどのようなことであろうか。本節では、これらの大会に参加した過去の経験から、コーパスの言語教育への応用に関してどのような考え方があるかをまとめ、関連する問題点を解説してみたい。

コーパスを言語教育に応用する際に大別すると1つの大きなアプローチがある。1つは間接的アプローチ (indirect approach), もう1つは直接的なアプローチ (direct approach) である。

3.1 間接的なアプローチ

間接的なアプローチはコーパスから得られた情報を言語教育シラバスや言語教材 (特に辞書や文法書) の作成や改善に生かすことが主眼になる。応用範囲としては、以下のような分野が考えられる：

- (1) 学習辞典：COBUILDを筆頭にLongman, Cambridge, Oxford, Macmillanなどが相次いでコーパス準拠の辞典を出版している。最も間接的アプローチで応用の盛んな分野だといってよいであろう。
- (2) 学習文法書：COBUILDの文法・語法書はコーパス準拠の先駆的なものだし、学習文法の範疇とは言えないがBiber et al. (1999) はコーパス準拠の最も新しい英文法書である。その他にもMindt (1995)などの既存文法のコーパスによる記述の見直し、などもこの分野の業績といえよう。
- (3) 教科書：Ljung (1900, 1991) の一連の教科書分析はコーパス言語学の手法を教科書の分析に利用したものとして特筆に値する。コーパス準拠の教科書というのは筆者の知る限りまだ現れていない。
- (4) シラバス：WillisのLexical Syllabus (Willis 1990)の考え方はCOBUILD Projectに長く関わった経験から生まれたものである。ただし、シラバスの中身自体はそれほどコーパス分析を入念に行ったものではなく、今後の研究が待たれる分野であろう。
- (5) 学習語彙表：コーパス分析による学習語彙表はCOBUILDの一連の辞典に載せてあるダイヤモンド印、LDOCEなどの話し言葉・書き言葉別の3000語、また『ジーニアス大英和』の重要語表示など辞典情報として頻繁に利用されている。最先端のコーパス処理の技術を用いたものとしてはJACET基本語改訂委員会で現在作成中の通称「JACET8000」がある。⁶
- (6) 言語テスト：コーパスを利用した言語テストはAldersonがTALCの第1回会議で発表した論文 (Alderson 1996) 以外はめぼしい業績はない。もっともテスト開発とまではいかないが、John Miltonの開発したWord Pilotというソフトでは、自分の作成したコーパスからコンコーダンス・ラインをもとにした穴埋めテストを作る機能がある。⁷ コーパスからの語彙頻度や統語構造の複雑度、コロケーション統計などを利用した言語テストは大きな可能性を秘めていると思われるが、言語テストの技術とコーパス言語学の研究領域の融合を考える研究者はまだ非常に少数であり、今後の発展が期待される。
- (7) 教員研修：コーパスを教員研修に利用しようとする考え方は特に学習者

データをコーパス分析する研究者の間で顕著である。学習者コーパスを用いることによって、どのレベルの学習者がどのような誤りを犯すかに関する系統的な資料を得ることができ、教員研修における言語学習過程に関する新たな認識を与えてくれるという点で注目される。

- (8) 言語習得研究：特に学習者コーパスを用いた第2言語習得の分野の研究は海外ではICLEプロジェクトが、国内ではSST (Tono et al. 2002), JEFLL (Tono 2002) といったコーパスの作成⁸が行われ、この5年間ほどで非常に盛り上がりを見せている。前節の国際学会などでも4分の1程度の発表は学習者コーパスに関するものになっている。

以上、列挙した分野からもわかるようにコーパス言語学の教育への間接的な応用は辞典や文法書などの教材部分を除くとまだほとんど手付かずか、ようやくその緒に就いたと言ってもよい。むしろ、コーパスからの情報が一体どれほど役に立つのか、といった基本的な概念自体が一般の利用者側にわかっていないというのが現状である。

3.2 直接的なアプローチ

直接的なアプローチとしては前述のTim JonesのData Driven Learningに代表されるような一連のコーパスの教室利用が挙げられる。これらの代表的な研究者としては、Chris Tribble, Mike Scott, Guy Aston, Sylvia Bernardiniらが挙げられる。彼らの基本的な発想は、「学習者自身が探求者となる (learner as researcher)」である。次のKettemannの言葉がその本質をよく表している：

"As an extremely powerful hypothesis testing device on vast amounts of data, the computer allows controlled speculation, makes hidden structures visible, enhances at the same time imagination and checks it by inductivity, thus making higher degrees of objectivity possible. The students' work becomes more exploratory, self-directed and thus motivating and highly experiential. "

--- Bernhard Kettemann (1996 : 5)

このコーパス利用の言語教育へのアプローチは大別すると分散型 (divergent) と収斂型 (convergent) に分かれる。分散型はコンコーダンスーを用いて言語事実を観察した結果行き着く結論がばらばらでもかまわない、というタイプの活動で、たとえばBrown Corpusを使って何か調べてみる、トピックは何でもかまわない、というようなものはこれにあたる。一方、収斂型は教師の側で学習者が観察の末到達する結論をあらかじめ用意しておく。た

たとえば, some と any の用法の規則性を発見させるような場合には, ほぼ授業の最後に結論として全員が共通理解を得ることが目標になる。

コーパスを用いることにより, 教師が規則を学習者に押し付けるのではなく, 学習者自らに生きた言語資料から規則性を発見させ, 自分で言語に対する新しい認識を持てるように環境を整えてやることが教師の役割だという考えが DDL の特徴だ。そしてそこで学習者が発見するものは教師が意図したものと必ずしも一致せずともかまわない, という考え方も一部の主唱者には見られる。これを彼らは 'serendipity' (幸運な発見) と呼ぶ (Bernardini 2000)。コーパスを自ら検索していろいろな課題をこなしているうちに偶然発見するような言語に対する新たな知見が, 教育的に非常に意味があるという考え方なのだ。

これは言われてみるとなるほどと思うような説明なのだが, 我々は注意深くこの直接的なアプローチの効果に関する主張を吟味しなければならない。以下にその理由を詳しく述べてみたい。

3.3 コーパス・データの難易度

コーパスを言語教育に利用する際には, 「コーパスの中身」に関する議論を慎重にしなければならない。一例として, 以下のコンコードダンス・ラインを試してみよう。1分間この文脈を見ながら, 括弧にはいる2語の単語を推測してみたい。

in the problem than to a lack () () or diligence on the part of the asses
dog, and therefore is "devoid () () ". Piepsam calls the cyclist "cur" and
question: Has the long upswing () () rates during the past 15 years just a
ll produce a prolonged decline () () rates? My answer is in the negative b
hat road rose from 62 per cent () () requirements in calendar 1957 to 86 p
greater detail in their areas () () . Since the writer had not noticed thi
e to pan gold, etcetera may be () () . Some areas may provide archeological
ther delayed Completion. It is () () that her menarche was somewhat later
the public. The immense amount () () that the new jazz had for the younger
e assembling of selected cases () () to military medicine and of establish
new books, and other features () () to officers of the medical services.
m courts and fairs, which were () () to Othon, the builder of castles in W
t attention but Helva had been () () to several department heads in Centra
al reports, and other subjects () () to small business. The Office of Fo
e development is certain to be () () to the readers, for the idea has so o
es and clippings, or any items () () to you. It's getting interested in so
ntity. Feelings of a community () () will have to be recreated- in some of

正解は "of interest" である。この簡単なタスクを Lancaster 大学言語学科の Corpus Research Group という研究会での発表の際にその場にいた参加者12人に実施したことが会った (Tono 1999)。参加者のうち英語の母語話者が5名, 外国人が7名いて, タスクの正解者は母語話者が5名全員だったのに対

し、外国人は7名中たった1名しか正解できなかった。

言語教育ではCOBUILDの画期的な辞典の登場依頼、authenticな生きた英語を与えるということが非常にもてはやされる時期があった。例えば、Essex大学の言語習得や教育の分野で業績のあるVivian Cookは以下のように述べている：

"Designers of coursebooks and syllabuses may miss some of the aspects of language used in real-life situations. This lack can be filled most easily by giving students the appropriate real-life language."

--- V. Cook (1991 : 94)

このようなreal Englishという合言葉を追い風にコーパスは実際に使われた本物の英語を学習者に与える、ということがクローズアップされていく。しかし、このような潮流に釘を刺す研究者もいた：

"But the thing to avoid, if we can, is treating the use of corpora in teaching as a bandwagon. Teaching bandwagons, if driven too far and too fast, can do much harm to those on the receiving end."

--- G. Leech (1997 : 4)

"the corpus itself has no more than the facilitative 'pedagogue' role. It enables the learner/ student to explore, to investigate, to generalize, to test hypotheses; but it does not itself initiate or direct the path of learning."

--- *ibid.*: 5

先ほどのof interestの例だけでは一般化はできないが、コンコーダンス・ラインからの推測作業は外国人にとっては決して容易な作業ではないことがわかる。つまり手がかりとなる文脈の難易度が推測作業を大きく左右する、ということが容易に想像できるのに、コーパス言語学の研究者の立場からはあまり深刻に受け止められていないというのが現状であった。

またLeech (1997)の2番目の引用にあるように、コーパスを使えば自動的に言語学習が進むというような短絡的な発想があるように思われる。これはプログラム学習やCALLのような新しいテクノロジーを教育に応用しようとする際の1つの典型的な現象である。しっかりとした中身の定義や効果の吟味が行われずに、流行として一斉に皆が飛びつくようなことは戒められなければならない。ではこのような現実に対して我々ほどのような研究を行っていけばよいのだろうか？次章でその点を考察してみよう。

4. 言語教育におけるコーパス利用の効果検証

言語教育におけるコーパス利用を考える際には、先ほどの直接的・間接的なアプローチという視点以外に、もっと教育現象の変数定義という点からの整理が必要になる。すなわち、(i) 教材・教具としてのコーパスの内容に関する変数、(ii) 提示するモードに関する変数、そして (iii) 学習者の中間言語 (interlanguage) の発達過程において変容を期待する部分に関する変数、の3つの視点が最低必要だ。⁹ それらの変数を図示したのが、図1である。

コーパスから取り出す情報としては大まかに言って単語リスト、コロケーション、構文パターン、語法などに関する頻度 (frequency) とその分布 (distribution) が考えられる。それらの情報を学習者に提供する際には、学習者内の要因 (年齢、認知スキル、学習ニーズ、4技能のレベル)、および学習環境要因 (教員の指導能力、ITスキル、学校のITインフラ、コーパスに接することのできる物理的な時間、など) を考え合わせなければならない。また、3つ目の要素として、ただコーパスを使えばよいというのではなく、中間言語 (英語学習者の英語能力) のどのような部分に焦点を当ててコーパス・データを提示するのか、それによってどのような学習者の側の言語知識・能力の変容を期待するのか、といった具体的な視点が必要になる。最低でもこれらの3要素が相互に関係しあってコーパスの教育利用に関する変数を複雑に構成していると考えられる。

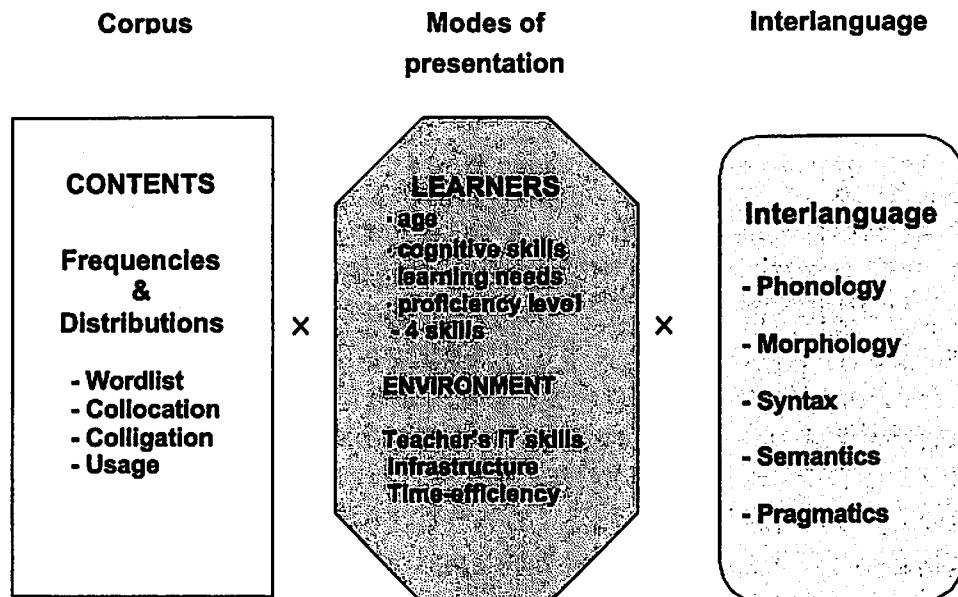


図1：コーパスの教育利用における変数

実際の教育研究ではこのように教材・学習者・環境といった要因を変数的に制御して実験を行うのであるが、このような実験的な手法でコーパス利用の効果を調査したものは筆者の知る限りでも極めて少ない。この辺の事情は80年代初頭のCALLの効果研究に似ていて新しい教育機器を「使った場合」「使わない場合」という大まかな区分けはあっても、その中身やどのような中間言語の側面に変化を期待するか、などに関する厳密な定義がないから効果の特定が難しいという皮肉な結果になっているのだ。

5. コーパスの中身に関する視点

前節で取り上げたさまざまな変数と関連して、コーパスの中身に関する問題をもう少し掘り下げてみたい。従来のコーパスの教育利用の議論では、大部分がBNCやBank of Englishなどの大型コーパスからの一般的な語彙統計情報を学習者に示すという発想だった。しかし、前述のKWICラインを外国人が読み取る能力の例からもわかるように、英語学習者に与える素材としてBNCなどの英語がふさわしいかという点必ずしもそうとは言えない。ではどのような中身が望ましいのだろうか。この点に関しては3点ほど具体的な視点を挙げる事が出来る。

5.1 コーパスに含む英語そのものを簡易化する

1つの方向性としては、コーパス・データの中身自体を簡易化するというものだ。我々は教育利用のために制限語彙で定義された英英辞典や、語彙レベルを制御したgraded readersなどを使っている。そのような英語の難易度レベルを調整することで、接する英語の内容理解を深め、文脈知識などが増えれば新しい単語や文法事項の類推や帰納的な学習がより効果的になる、という考え方は妥当であると思われる。

この路線でいけば、我々英語学習者が使うための特別に語彙制限されたコーパス・データというものがあってもよい。¹⁰ COBUILDなどのアプローチでは「コーパスは実際に使用された"real English"でなければいけない」という発想があるが、このような考えは間接的なアプローチでならまだしも、直接的アプローチではもう古いと言わねばならない。Widdowsonが2002年のTALCで頻繁にコメントしたように、教室は擬似的に英語を学習するように設定された環境であり、我々は学習材を單元ごとに区切り、時間配当を決め、毎回提示・練習する内容を制御して学習者に与えている。その理由はフリーに英語に触れて身につくためには膨大な接触量が必要だがそれが日本の状況では不可能なのでより効率的なシラバスの工夫が必要だからである。このような観点からいけば、コーパスだけが神聖視され絶対に生の英語でなければいけない、というのは現実を無視した理想論であると言えよう。もちろん目標としては

「生きた英語」のサンプルを入れたコーパスが参照できることは望ましいに違いない。そういうコーパスの存在は否定しない。しかし、その中間過程でより accessible なコーパス・データの存在を認めないとすれば、それはコーパス言語学の教育利用を大きく阻む原因になるだろう。

具体的には、graded readers のコーパス化、海外の小学校レベルの教科書や百科事典などをコーパス・データとして用いることは大いに期待できる。このような中身を学習者レベルに適應させたコーパスと一般のコーパスとで、コンコーダンス・ラインの理解度などを実証的に検証してみることが出来れば、非常に興味深い結果が得られるに違いない。

5.2 パラレル・コーパスの利用

2つ目の大きな中身に関する議論としてはパラレル・コーパスを利用することが考えられる。これはTim Jones のページなどでも以前から挙げられていることで、別に目新しいことではないのだが、日本ではまだ言語教育の具体的な実践としてパラレル・データを利用している研究はほとんど皆無に近い。¹¹ パラレル・データは英語のコーパス・データと対応する日本語の対訳データを併せ持つ。表示も英語のコーパス・データと対応する形で日本語が表示される。そこで前述のような英語のコンコーダンス・ラインを理解する負荷が日本語のサポートによって大きく軽減されるという効果がある。パラレル・データを扱えるソフトとしてはMichael Barlow のParaConc (Barlow 1996)、Tim Jones らの開発したMulti-Concord¹² などがあり、日本の英語教育にも即応用できるようなツールが利用可能だ。また自然言語処理の分野では対訳データの自動対応付け (automatic alignment) の技術が日本でも非常に進んでおり、これらの技術をテキストの前処理で利用することが出来れば、ParaConcなどのUnicodeを扱えるソフトでは日本語表示も問題なく可能になるだろう。¹³ このようにパラレル・コーパス利用の環境はかなり整ってきつつあるので、この方面での活用のノウハウの蓄積が望まれる。

5.3 学習者コーパスの利用

コーパスの中身に関する第三の可能性として学習者コーパスを挙げておこう。学習者データをコーパス化することは、最近第2言語習得などの分野で急速に関心を集めているが、学習者自身によるコーパスの直接利用としてのデータとしても意味がある。1つは、学習者の発表したい内容がより身近な形でデータとして読んだり検索したりできる点だ。例えば、筆者が中心に構築しているJEFLL Corpus (Tono 2002) は6種類の共通トピックを作文課題として課している。将来的に研究用に公開されれば、そのトピックと同じテーマで自分の生徒に作文を書かせてみて、自分と同じまたは異なるレベルの他の学習者がどのような文章を書いているか、どのような表現や語彙を使っているのかをコーパ

スを用いて閲覧・検索できる。コーパスそのものが直接発信用の表現辞典のようにも使える。

また学習者データにエラー情報が組み込まれていれば、それをもとに日本人のよくやる英語の誤りに関して系統的な情報を得ることが出来る。Data Driven Learning に英語学習者のエラーパターンを視点として組み込めば、より学習を深化させることも可能になろう。日本では明治大学のKevin Markがこのような発想に近い日英対訳エラー・データベースを作成しており (Mark 1998)、英作文をいろいろなレベルでより自然な表現へと修正する過程をデータベース化している。このような資料が教材作成や教員研修、実際の学習者の参考資料として活用されることが強く望まれる。

6. 待たれる実証研究

コーパスの中身を前節のように英語学習者のニーズに合わせて調節したとしても、そのようなコーパスを用いてどのような学習効果を挙げるかは実際にやってみないとわからない。そういった意味で、現状ではまだ実践報告の類も少ないし、ましてや実証的な効果研究と言うのは極めて稀である。過去のTALCにおける発表論文は90余りになるが、そのうちで質問肢などにより実際の効果に関して評価を行ったものは3,4本しかない。ほとんどがこんなやり方で授業中にコーパスを使ってみた、という類のものが多く、図1のような複雑なコーパス利用の変数を制御した実験研究はまだ皆無に近い。今後、実証研究が積み重ねられていくために必要な観点としては以下のような項目を挙げることができる：

1. コンコーダンサーを使うと何がよいのか？使わない場合との具体的な違いは？
2. 異なるコーパス（例：一般コーパス/学習者コーパス/ESPコーパス）を利用する利点は？
3. コーパス使用により言語能力のどのような側面に効果を期待するのか？
4. DDL（収斂型・分散型）の具体的な効果は？
5. コーパス・データとその他のデータ（辞書・文法書・教科書など）との違いは？

これらの研究課題が具体的な変数操作を伴って実証的に仮説検証されることでコーパス利用の言語教育の真の意義が解明されてくると思われる。

7. 今後の課題

本稿を締めくくるにあたり、コーパスの言語教育利用に関する今後の課題を総括しておきたい。

7.1 言語教育用コーパスと利用環境の整備

Leech (1991: 20) はコーパス言語学の技術的な課題に関して以下の3点を挙げている。これが言語教育用のコーパスにも同様に当てはまる：

- A) 基礎的なコーパス構築・整備 (Basic corpus development)
- B) 検索性ツールの開発・整備 (Corpus tools development)
- C) コーパス注釈付けの開発・整備 (Development of corpus annotations)

言語教育用には書き言葉と同時にレベル調整した会話データのコーパスが必須だ。またESPなどの専門英語コーパス、学習者コーパス、教科書コーパス、日英対訳コーパスなど、一般の英語コーパスを補強するさまざまなコーパスの開発が強く望まれている。またそれらのコーパスを検索するためのツール類も重要だ。検索ツールには学生が利用することを想定したユーザー・フレンドリーなインタフェースが期待される。また、注釈付け (annotation) の整備も重要だ。学習者データならばエラー情報、教科書コーパスであれば新出単語・構文、本文、練習問題などの文章構造を別々に検索するための標識など、いろいろな工夫が必要になってくるだろう。

7.2 コーパス使用の厳密な定義と実証的研究の必要性

コーパスを英語教育に真に生かすためには、間接的なアプローチと直接的なアプローチの両面で吟味が必要になる。間接的なアプローチの恩恵は学習英語辞典をはじめ、いろいろな分野で活用が期待されており、その貢献度が高いことは衆目の一致するところであろう。問題は、直接的アプローチにおける効果である。これには4. で指摘したようなコーパスから得られるデータの性質・提示モード・学習の側面といった多様な変数の整理が必要であり、これらを厳密に定義した上での実証研究の積み重ねが重要になってくる。

7.3 学習環境全体の変容を考慮に入れた提案の必要性

実証研究の積み重ねとは別に、コーパスを利用する環境や言語学習へのインフラ全体を変えるような提案がなされていく必要がある。というのは、コーパス利用はその性質上、教師主導よりも学習者主導の形態で効果を発揮する。学習者が探索的に仮説検証を行えるためにはコーパスに触れる設備や時間が十分に確保されていなければならない。現在の限られた教室空間や授業時間数の中

だけで、コーパスの利用を限定的に評価したのでは、本当の真価が発揮されないという可能性がある。そういった意味で、現状の学習環境変数を統一した実験的手法を駆使した研究とは別個に、新しいインフラ整備や学習環境の提案を行いながらコーパスの位置づけを考えるという大胆な発想も不可欠となるであろう。

8. まとめ

本稿では、コーパスの英語教育への応用というテーマでその歴史的な背景と現状を分析し、問題のありかや今後の研究の方向性などに関して論じてみた。本稿で述べてきたようなコーパスの教育効果を検証するための実証的研究の積み重ねが可能かどうかは、コーパス言語学の専門家と一般の英語教員との研究協力や情報交換に懸かっている。英語コーパス学会のような専門家集団は、高度なコーパス分析の技術や知識を追求するだけでなく、英語教育界へ理解しやすい形でコーパス利用の効用を伝えていく必要がある。そのうえでコーパスを利用する英語教師の裾野を広げ、実際の教育方法としての効果を検証するような問題意識を現場に育てていくことが大切だ。過去のCALLの分野などに顕著なのだが、新しい教育機器・教材は一部の利用者にもみその効果を「信奉」され、一般の英語教育界からはあまり関心を寄せられずに次の新たな技術が導入されると廃れてしまうことが多い。コーパスは機器に依存しない、いわば言語資源である点、またself-access centerなど大きく変化する学習環境の総合的な核となるデータとして必ずやその本質が理解され重視される時代が来るだろう。その際に、我々が最先端の情報と専門的な知見をもとにさまざまな提案を行えるかどうかは、本稿で詳述したような研究課題に今後我々一人一人がどのように取り組んでいくか次第であろう。

注

- 1 もっとも米国ではAmerican Heritageのチームが新しい発想で教育語彙に取り組み、Carroll et al. (1971) など高度な統計処理を行った頻度表の出現などがあったことを忘れてはならない。
- 2 Classroom Concordancingの実践についてはTribble and Jones (1997)を参照。
- 3 <http://sun7.bham.ac.uk/johnstf/homepage.htm>
- 4 ANCコンソーシアムの正式メンバーとして、研究社、大修館書店、三省堂書店、小学館、ベネッセ・コーポレーション、アルク、旺文社が辞典出版社として参加している。
- 5 <http://www.perc21.org>

- 6 JACET8000 に関しては, <http://tonolab.meikai.ac.jp/~jacetvoc/> を参照のこと。
- 7 Word Pilot に関しては,
http://home.ust.hk/~autolang/download_WP.htm#C を参照のこと。
- 8 学習者コーパスの総合的な情報は筆者のwebpage (<http://leo.meikai.ac.jp/~tono/>) を参照されたい。
- 9 これらの変数以外に教師や教室環境, 個人差などのさまざまな学習・環境要因が関わってくるわけだが, これらはどのような教育効果の測定でもいえることなのでここでは議論から除外しておく。
- 10 同様の主張は Seidlhofer (2001), Osborne (2002) などでもなされている。
- 11 パラレル・コーパスの言語研究の利用に関しては JAECS の大会でも何度か取り上げられていることを付記しておく。
- 12 http://web.bham.ac.uk/johnstf/l_text.htm
- 13 この論を執筆中に, 読売新聞と Daily Yomiuri を対応付けしたデータが通信総合研究所から研究用に利用可能になった。徐々にパラレル・データを利用する環境が整ってきつつある。

参考文献

- Alderson, C. (1996) "Do corpora have a role in language assessment?" In Thomas, J. and Short, M. (eds.), *Using Corpora for Language Research*. London: Longman, pp. 248-259.
- Barlow, M. 1996. "Parallel texts in language teaching". In Botley, S.P., Glass J., McEnery, T., and Wilson, A. (eds), *Proceedings of Teaching and Language Corpora 1996*. UCREL Technical Papers 9 (Special Issue), Lancaster University.
- Bernardini, S. (2000) "Systematising serendipity: proposals for concordancing large corpora with language learners." L. Burnard & T. McEnery (eds.), *Rethinking Language Pedagogy from a Corpus Perspective: Papers from the Third International Conference on Teaching and Language Corpora*. Frankfurt am Main: Peter Lang, pp.225 - 234.
- Biber, D., Johansson, S., Leech, G., and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. London: Longman.
- Carroll, J.B., Davies, P. and Richman, B. (eds.), (1971) *The American Heritage Word Frequency Book*. New York: American Heritage Publishing Co.
- Cook, V. (1991) *Second Language Learning and Language Teaching*. London: Edward Arnold

- Cowie, A.P. (1999) *English Dictionaries for Foreign Learners*. Oxford: Clarendon Press.
- Faucett, L., Palmer, H.E., West, M. and Thorndike, E.L. (1936) *Interim Report on Vocabulary Selection*. London: P.S. King.
- Kennedy, G. (1998) *An Introduction to Corpus Linguistics*. London: Logman.
- Kettemann, B. (1996) "Concordancing in English Language Teaching." In Botley, S. P., J. Glass, T. McEnery, A. Wilson (eds.), *Proceedings of Teaching and Language Corpora 1996*. UCREL Technical Papers 9, Lancaster University, 1996, pp. 4-16.
- Leech, G. (1991) "The state of the art in corpus linguistics." In Aijmer, K. & Altenberg, B. (eds.), *English Corpus Linguistics: Studies in Honour of Jan Svartvik*. London: Longman, pp.8-29.
- Leech, G. (1997) "Teaching and Language Corpora: a Convergence." In Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (eds.), *Teaching and Language Corpora*. London: Longman, pp.1-23.
- Ljung, M. (1990) *A Study of TEFL Vocabulary*. Stockholm Studies in English 78. Stockholm: Almqvist & Wiksell.
- Ljung, M. (1991) "Swedish TEFL meets reality." In Johansson, S. & Stenstrom, A.-B. (eds.), *English Computer Corpora*. Berlin: Mouton de Gruyter, pp. 245-256.
- Lorge, I. (1949) *Semantic Count of the 570 Commonest English Words*. New York: Columbia University Press.
- Mark, K. (1998) "A parallel learner corpus approach to English curriculum development at a Japanese university." In *Proceedings of First International Symposium on Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching*. (14-16 December, 1998), pp. 89-90.
- Mindt, D. (1995) *An Empirical Grammar of the English Verbs: Modal Verbs*. Berlin: Cornelsen.
- Osborne, J. (2002) "Top-down and bottom-up approaches to corpora in language teaching." Paper presented at North American Symposium on Corpus Linguistics. Indianapolis, 1 - 3 November, 2002.
- Seidlhofer, B. (2001) "The case for a corpus of English as a lingua franca." In Aston, G. & Burnard, L. (eds.), *Corpora in the description and teaching of English*. Bologna: CLUEB, pp. 70-85.
- Sinclair, J.M. (1987) *Looking Up: An account of the COBUILD Project in lexical computing*. London and Glasgow: Collins ELT.

- Thorndike, E.L. (1921) *Teacher's Wordbook*. New York: Columbia Teachers College.
- Tono, Y. (1999) "Application of Corpora to Classroom Language Teaching." A paper given at the Corpus Research Group meeting. 10 May, 1999, Lancaster University.
- Tono, Y. (2002) *The Role of Learner Corpora in SLA Research and Foreign Language Teaching: The Multiple Comparison Approach*. Unpublished Ph.D. thesis, Lancaster University.
- Tono, Y., Kaneko, T., Isahara, H., Saiga, T. and Izumi, E. (2002) "The Standard Speaking Test Corpus." *Studies in Lexicography* 11 (2), 7-18.
- Tribble, C. and Jones, G. (1997) *Concordances in the Classroom*. Athelstan.
- West, M. (1953) *A General Service List of English Words*. London: Longman.
- Wichmann, A., Fligelstone, S., McEnery, T. and Knowles, G. (eds.), (1997) *Teaching and Language Corpora*. London: Longman.
- Willis, D. (1990) *The Lexical Syllabus*. Collins ELT. (Available online at <http://www.cels.bham.ac.uk/resources/LexSyllabus/lexsch1.pdf>)