

## はじめに

**前**回は、学習者コーパスとは何か、またそれによってどのようなことが可能になるかを、研究・教育の両面で具体例を示して考えてみた。今回はこの分野の研究の国内外の進捗状況を概観し、その現状を主要なコーパスの Web サイトの紹介や研究論文をもとに解説してみたい。

## 学習者コーパスの分類基準

世界の学習者コーパスを総覧する際に、いくつかの分類基準を設けてみよう。そうすることによって、既存の学習者コーパスの位置づけがより明解になるはずだ。

### (1) コーパスの規模

学習者コーパスはまだ緒に付いたばかりであるので、それほど大規模なものは開発されていないが、あまり規模が小さくては一般性に欠ける。学習者データの採取はかなり骨が折れる作業であるが、目標としては各レベルごとでも最低 20 万語、総量としては最低 100 万語くらいを目指したいものだ。

### (2) 学習者のレベル

コーパス・データを採取する際に、学習者のレベル分けは重要な要素である。学年などで分類することも考えられるが、世界の他の学習者コーパスと比較する際には不十分である。その学習者の客観的な英語力の指標が必要だ。そのために短時間である程度のレベル判定の出来る英語力小テストのようなものがあって、それによって英語力のレベルが統一されたデータであれば価値が高い。

### (3) 母語の相違

母語の背景が異なる学習者のデータを同列に扱うことはできないので、良質な学習者データは L1 background を情報として明記すべきである。母語が異なる学習者データは比較して興味深いデータが得られる場合もあれば、データをはっきり分けて考えないときちんとした結果が導かれてこない場合もあるので要注意だ。

### (4) タスクの種類

どのようなタスクによって発話(作文)データを得たかも重要な情報である。タスクが明解に記述されていれば、そのコーパスと同等のものを新たに作ることが出来、それをもとのコーパスに付加して規模を拡大したり比較したりすることが可能だ。このタスクの情報は、課題の内容、指示の仕方、作業時間、実際に使ったプリントなどさまざまな情報を含むもので、詳しくれば詳しいほどよいと言える。

## 世界の主要な学習者コーパス

p. 40 の表 1 は、筆者が知る限りでの学習者コーパスのプロジェクトをまとめたものである。このうちで特に注目を集めている学習者コーパスの動きを拾って解説してみよう。

■ **International Corpus of Learner English (ICLE)** (URL: <http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/abs.html>)

このコーパスは 1990 年から始まった International Corpus of English (ICE) という世界 18 地域の英語変種コーパス構築プロジェクトの傘下で、変種の 1 つとして特に EFL learner のコーパスを作ろうというものである。現在世界の学習者コーパスの中でも、サンプリング方法の科学的なこと、規模と整備状況から見ても最も本格的なものだと言える。

Louvan 大学(ベルギー)の Sylviane Granger が中心 (Granger 1996)。英語を主として外国語として用いる 11 の異なる母語の上級学習者データ(課題は意見を述べるような自由英作文)を 20 万語ずつ集め、品詞タグ、エラータグを施す作業を進めている (Meunier 1995)。

Granger によれば、中心的な研究テーマは異なる L1 background を持つ学習者がおかすエラーを比較して、母語関連エラー (L1-related error) と普遍のエラー (universal error) を特定することである。現在データは ICE project のメンバー以外には公開されていない。日本でも ICLE にデータを提供している研究グループがあり、Granger によ

# 学習者コーパスと

ると AILA'99 (来夏、東京で開かれる国際応用言語学会の世界大会) で ICLE 独自のシンポジウムを開くそうである。

### ■ HKUST Learner Corpus

香港科学技術大学の John Milton が中心になって集めている中国人大学受験生、大学生の英作文コーパス。電子メールで作文を書かせて集めたものと、大学入試の共通英作文問題の transcription が中心で、現在 1200 万語ほどある。その一部分は品詞タグ付きでエラータグの処理も施している (Milton 1994)。Milton はこの学習者コーパスのエラーデータをもとに独自の「学習者弱点克服メニュー」のようなものをコンコーダンサーに組み込み、AUTOWORD と題して 97 年 3 月のアジア辞書学会でデモを行なっている。

### ■ Longman Learner's Corpus

Longman が所有する商用の学習者コーパス。約 1000 万語の規模は世界最大級。ただしデータの採取方法はかなりばらばらで、文法問題あり、自社

の courseware のモニター資料あり、Cambridge 英検の essay データありと雑多である。これらを世界 70 カ国以上から集めたデータベースを Michael Rundell が corpus manager として全体の統括をしている。品詞・エラータグなどは現在のところついていない。このデータをもとにして Longman Dictionary of Common Errors や Longman Essential Activator などのコメント欄が書かれている。商用で本格的に出版物に学習者コーパスを利用したのはロングマン社が最初である。

## 日本の学習者コーパス

### ■ Corpus of Japanese Learners of English (http://www.lb.u-tokai.ac.jp/lcorpus/)

JACET'96 で呼びかけがあり、ハイパーメディア研究会のメンバーが中心となって発足した日本人英語学習者コーパス作成のプロジェクト。中・高・大と連携したデータの広範囲な採取、WWW 上での共有化、エラータグ開発、音声コーパスへの取り組みなど全国規模で展開する予定である。97 年度からは東海大学の朝尾幸次郎氏が代表となって科研のプロジェクトが 3 年計画で進行中。

日本初の学習者コーパス関連のメーリングリストもある(下記 Web site を参照)。

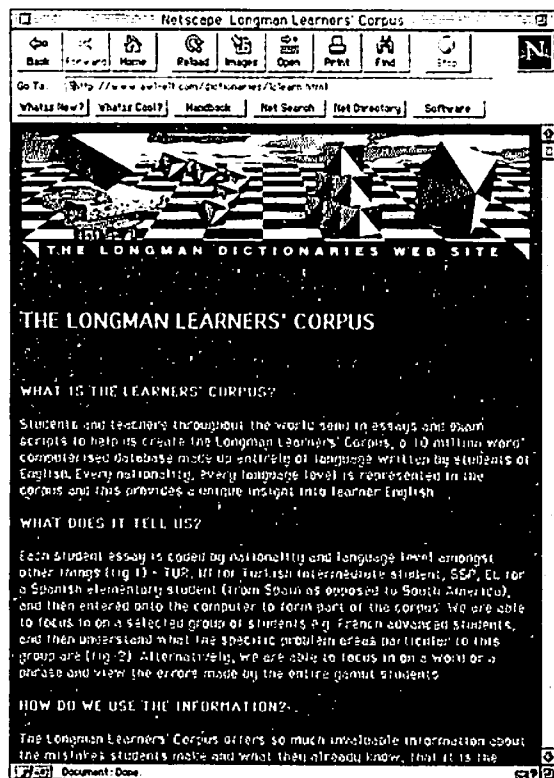


図 1 Longman Learners' Corpus (http://www.awl-elt.com/dictionaries/lclearn.html)

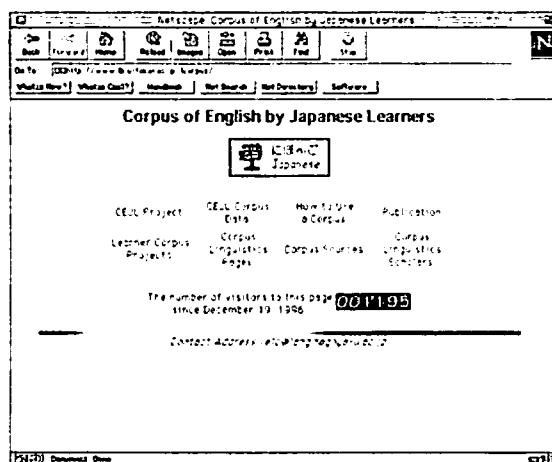


図 2 学習者コーパス構築のページ(東海大学朝尾研究室) (http://www.lb.u-tokai.ac.jp/lcorpus/)

### ■ TGU Learner Corpus

東京学芸大学で行なわれた 10 年間の科研の英作

英語指導 .....(2)世界の学習者コーパス

words 投野由紀夫

表1 世界の学習者コーパス

学習者コーパスの名称	総語数(目標)	開発国	学習者国籍	学習者レベル	タスクの種類	タグ整備	公開性
International Corpus of Learner English (ICLE)	各グループ20万語ずつ	ベルギー(ルバン大学)	11の異なるL1を持つ学習者	上級者	作文(論理的な文章)	エラータグ 品詞タグ	非公開、将来公開の予定あり
HKUST Corpus	1200万語	香港(香港科技大)	中国	中・上級	作文	エラータグ 品詞タグ	非公開
Hungarian EFL Learner Corpora	12万語(現在)	ハンガリー	ハンガリー	中・上級	作文	なし	部分的に公開
Longman Learners' Corpus	1000万語	英国	70カ国以上	全レベル	作文 文法問題	なし	商用
CBACLE (Corpus-Based Analysis of Chinese Learner English)	100万語(目標)	中国	中国	中学生～大学院生	作文	予定あり	公開予定
Colman Bernath's EFL Corpus	12万語	台湾 Soochow University	台湾	大学生	作文	エラータグ	公開
Corpus of Japanese Learners of English	未定	日本	日本	中学生～大学生	作文 会話 その他	エラータグ 品詞タグ (予定)	公開
TGU Learner Corpus	100万語(現在60万語)	日本(東京学芸大学)	日本	中学生～大学生	作文 会話	エラータグ 品詞タグ (予定)	公開予定

文プロジェクトで蓄積されたデータをコーパス化したもの (Tono 1996)。中学2年から高校3年まで同一のトピックで自由英作文をさせたデータを電子化している。現在、ファイル管理、品詞タグ、エラータグなどに関して整備中。現在の規模は約60万語。今後、ホームページで公開を前提に100万語まで増殖する予定。

このように学習者コーパスの構築は現在始まったばかりだが、おそらくこの5年くらいで状況はかなり激変するだろう。世界中でかなり多くのプロジェクトが立ち上がっており、今後は、表1に記したような学習者の発達のデータのサンプリング方法、タグ整備、タスクの開発などでコーパスの有効性や質が問われてくるに違いない。

なお、この原稿を書いている最中に Granger の編著の *Learner English on Computer* (Addison Wesley Longman, London & New York ISBN 0-582-29883-0) が出版されるとメーリングリストで案内があった。これは学習者コーパスを特集した初の本格的な研究書である。また研究社出版から最近刊行された『英語コーパス言語学——基礎

と実践』も本邦初の概説書として是非一読を勧めたい。

今回はいよいよ学習者コーパスを用いた分析例を見ながら、英語教育にどのように光を当ててるのか、その具体例を見ることにしよう。

〔参考文献〕

Granger, S. (1996) "Learner English around the World." In S. Greenbaum (ed.) *Comparing English Worldwide: the International Corpus of English*. Clarendon Press: 13-24.

Meunier, F. (1995) "Tagging and Parsing Interlanguage." In L. Beheydt (ed.) *Linguistique appliquee dans les annees nonante*, ABLA papers no. 16: 21-29.

Milton, J. (1994) "Tagging the interlanguage of Chinese learners of English" In L. Flowerdew & A.K.K. Tong (eds.) *Entering Text*. Language Centre, Hong Kong University of Science and Technology: 127-143.

Tono, Y. (1996) *Using Learner Corpora for L2 Lexicography*. LEXICOS 6 (AFRILEX SERIES 6) Stellenbosch: Universiteit van Stellenbosch.

(とうの・ゆきお / 元東京学芸大学講師、  
ランカスター大学博士課程在籍)

