

シンポジウム

「英語コーパス活用最前線 —現状と今後の展開を考える—」

奥 田 隆 一
(和歌山大学)

1. 目的と反響

このシンポジウムでは、英語のコーパスを使ってどのようなことが出来るのかを3つの違った分野の研究者の方に説明してもらうことにより、コーパス活用の現状と、今後の可能性と問題点を探ろうとした。ただ、残念であったのは、永田昌明講師の体調不良による不参加のため、代わりに司会者が永田講師の原稿を代読することになり、その結果、直接いろいろな質問に答えていただけなかったことと、3人の講師に対し、同一のトピックについて意見をお伺いしようとしたが、賛成を得られず、3人がバラバラの内容でプレゼンテーションをされたことである。しかしながら、将来的に、コーパスを活用して、まだまだいろいろなことが出来るということや、個人レベルでも研究にコーパスを活用することができるということを学べたことで非常に役に立ったという感想が複数の参加者からあったので、それだけでも、このシンポジウムを行った価値があったと考えられる。

2. 基本用語・概念

コーパスを活用した研究をしようとする、コーパス言語学の文献が役に立つのだが、その場合に、知っておくべき基本用語や概念があるので、簡単に触れておくことにする。

- (1) 「コーパス／コーポラ corpus (corpora)」：コンピュータによる検索が可能になっている大量の言語データのこと。
- (2) 「コンコーダンス concordance」：用例索引のことで、文書中で使われているキーワードの全出現箇所が、文脈の一部とともに記載されているもの。
- (3) 「クイック（形式）KWIC = key word in context」：検索対象であるキーワードを中心に、その左右の文脈を合わせて表示するコンコーダンスの形式を言う。
- (4) 「コンコーダンスライン concordance line」はコンコーダンスに示されたそれぞれの1行のこと。
- (5) 「タイプ／トークン type / token」：タイプとは、資料中の相互に異なる語形のこと。例えば、単数、複数で異なる語は別のタイプと数えられます。トークンは、単数・複数で違った形でもそれを一つと考える数え方です。
- (6) 「タグ付き tagged」：コーパスに、特殊なタグ（品詞や構文の情報など）が付いているもの。
- (7) 「ラーナーズ・コーパス learner's corpus」：学習者コーパスとも呼ばれ、母国語話者ではなく、学習者が話したり、書いたりした言語データのデータベースのこと。
- (8) 「パラレル・コーパス parallel corpus」：まったく同じ内容の2カ国語コーパスをパラレル・コーパスという。

3. コーパス、オンライン検索可能コーパス

それでは実際にコーパスを活用しようとする、どうすればいいのかということになるが、大規模コーパスを購入（または使用料を払い）してソフトを使い検索するという方法がコーパス言語学では主流だが、インターネットで独自のコーパスを公開し検索可能なサイトもあるので、以下にはこのようなすぐに使える主な無料のサイトを列挙しておく。

- (1) British National Corpusの一部分のコーパスを無料で検索できるサイト：Simple Search of BNC-World (<http://sara.natcorp.ox.ac.uk/lookup.html>)
- (2) ビジネスレターのコーパスを検索できるサイト：Business Letter Corpus Online KWIC Concordancer (<http://ysomeya.hp.infoseek.co.jp/>)
- (3) ミシガン大学で行われた講演をテキスト化し検索できるようにしたサイト：Michigan Corpus of Academic Spoken English (<http://www.hti.umich.edu/m/micase/>)
- (4) Brown Corpus, Lob Corpus, The Times 1995年1~3月等の検索が出来る：The Web Concordances (<http://www.edict.com.hk/concordance/WWWConcappE.htm>)
- (5) コーパス言語学専門のMark Davies教授のサイトだが、現代アメリカ英語のコーパスや*Time*やBNCなどが検索できる：[\(http://davies-linguistics.byu.edu/personal/\)](http://davies-linguistics.byu.edu/personal/)
- (6) 映画の英語を中心に扱っているスクリーンプレイ社のサイトだが、映画の台詞を検索できる：映語犬サク (<http://www.screenplay.co.jp/>)
- (7) アメリカ、イギリス、オーストラリア新聞英語を比較することが出来る：Sudachi KWIC Search (<http://uzu.ias.tokushima-u.ac.jp/staff/nakasima/skwics.html>)

4. 将来的に活用度の増す可能性のあるコーパス

これから、活用度が増すだろうと思われるコーパスは、上で触れたパラレル・コーパスとラーナーズ・コーパスであろう。特に、パラレル・コーパスは日英語の比較研究や翻訳などに、ラーナーズ・コーパスは英語教育に活用できると思われる。以下に、簡単にアクセスできるもののみを挙げる。

- (1) 日英語パラレルコーパス
 - a. 立命館大朝尾幸次郎氏のサイト
(<http://www.eng.ritsumei.ac.jp/asao/corpus/ej.html>)
 - b. 読売新聞とThe Daily Yomiuriの日英対応付けコーパス (JENAAD)
(<http://www2.nict.go.jp/x/x161/members/mutiyama/jea/index-ja.html>)
 - c. 日英対訳文対応付けデータ (Project Gutenbergや青空文庫やプロジェクト杉田玄白などの作品に、日本語文と英語文との対訳文対応を付けたもの)
(<http://www2.nict.go.jp/x/x161/members/mutiyama/align/index.html>)
- (2) 日本人英語学習者の書いた英語の学習者コーパス：
 - a. 名古屋大学の杉浦正利氏が中心になって作成された学習者コーパスNICE：
(<http://sugiura5.gsid.nagoya-u.ac.jp/~sakaue/nice/index.html>)
 - b. 立命館大学の朝尾幸次郎氏のサイト (<http://www.eng.ritsumei.ac.jp/lcorpus/>)