

英語教師のための コーパス入門

第2回

すぐに利用できる オンラインコーパス



明海大学 助教授
投野 由紀夫

前回は英語コーパスの利用によって何が
変わるか、そして利用環境を整えるために
必要な最低限の情報を紹介した。個人で
コーパスを入手するには時間が多少かか
る。そこで今回はその間にすぐにも利用
してみたい、という皆さんのために、イン
ターネット上で簡単に利用できるコーパスを中
心に紹介してみよう。

British National Corpus Simple Search

<http://thetis.bl.uk/lookup.html>

1億語のイギリス英語のコーパスBNCのお試
しサイトは全コーパスからランダムに50例が検索
できる。

図1は textbook という単語を検索した結果
だ。このように全体が400件でその中からラン
ダムに50件が表示される。それぞれの用例は
完全文になっているので、途中で文が切れてし
まって内容が不明ということはない。ただし、文
脈は表示されないで、前後関係がないと理解
しづらいものもある。文の先頭にあるアルファ
ベット記号はこの文が含まれているファイル名

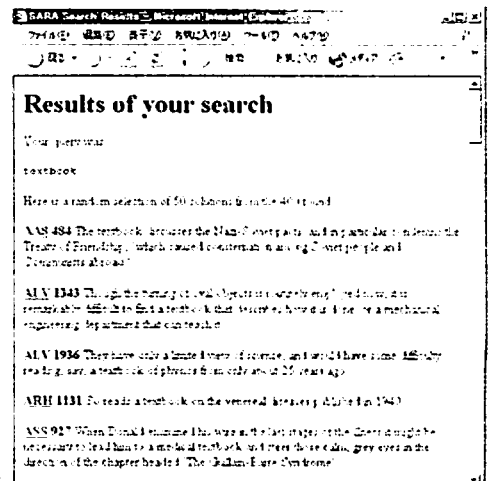


図1：BNC Simple Search の検索結果画面

で、そこをクリックすると出典の簡単な情報を得られるようになっている。

検索のコツ

BNCの例文は全体にむずかしいものが多いが、辞書に例文が載っていないような単語でも相当な確率で例文がヒットするので、用例データベースのようにつもりで使うとよい。

たとえば、目の前にある『ジーニアス英和』のpaperboundからpapyrusまでの20語は辞書中に用例がないが、BNCで検索すると2語を除いて全部用例が読める。

表1: BNCの用例ヒット件数

paperbound	3(paper-bound)	papillote	3
paperboy	10	papist	18
papergirl	3	papoose	4
paperhanger	1(paper hanger)	pappus	4
paperhanging	1	pappy	16
paperweight	44	paprika	46
paperwork	514	Pap test	0
papery	27	Papua	196
papier-mâché	0	papule	1
papilla	88	papyrus	61

これだけ用例がヒットすると語彙指導などには案外威力を発揮する。

たとえば、辞書などで意味をとらせても、まだ使い方が今ひとつわからないような場合に、BNCからの用例で適当な長さのものを10個くらい示せば、用例そのものが単語の意味や使い方を語ってくれる。むずかしい単語は、その場でオンライン辞書(Longman Web DictionaryとかBabylonなど)で検索すればよい。

COBUILD Concordance Sampler

<http://titania.cobuild.collins.co.uk/form.html>

コーパスの老舗COBUILDのサイトも必見だ。COBUILDは現在4億語以上の大規模なコーパスデータを持っているが、実際のサービスは5600万語に限定される。従来はtelnetとってUNIXのコマンドを使った原始的な検索サービスをやっていたCOBUILDも、現在はJAVAベースのGUIに刷新している。このページはサンプルのwebインタフェースなので、中身はいたってシンプルだ(図2参照)。

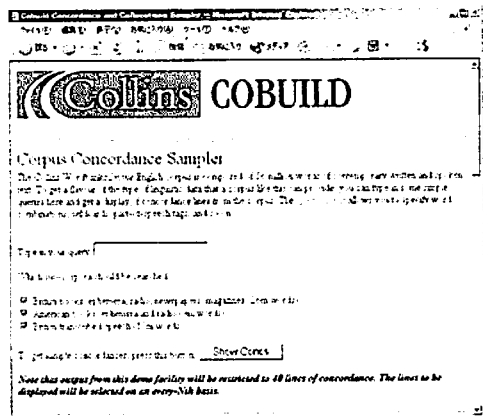


図2: COBUILD Corpus Concordance Sampler メイン画

Type in your query と書かれているフィールドに検索したい単語を入力して [Show Concs] というボタンを押すだけ。サブコーパスとしてはアメリカ英語(900万語)、イギリス英語(2600万語)、イギリス英語の話し言葉(1000万語)という3種類のコーパスを選択できる。

検索結果は図3のようになる。COBUILDの画面はBNCに比べるとちょっと使いづらい。結果は40例しか出てこないし、KWIC (Keyword in Context) という表示形式は一般的なものだが、検索語の両側の文脈が短いので、用例データベース的に使うには少々不便である。

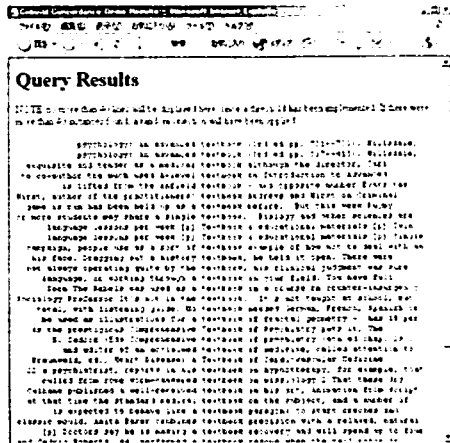


図3: COBUILD sampler の検索結果

検索のコツ

COBUILD の sampler が威力を発揮するのは、Query syntax という特別な検索式を使ったときだ。たとえば、@ という記号を用いて“give@”のようにすると動詞の活用形がすべて一度に検索できる。また単語の品詞を指定することも可能で、“spring/VERB”とすると「春;泉」といった名詞の意味の spring ははじかれる。

こういった高度な検索は元のコーパス・データに品詞情報が埋め込まれているから初めてできるのであって、普通我々がインターネットでダウンロードするようなデータではこうはいかない。

またサブコーパスの選択も一般の利用者にはありがたい。前述の3つのサブコーパス(英米の書き言葉、英の話し言葉)の選択ができれば、ちょっとしたジャンルやレジスターの違いをざっと調べることができる。また話し言葉コーパスから用例を取ったほうが一般に簡単な例文がとれる。

もっとも先ほど述べたように KWIC ラインが拡大できないのが最大の欠点で、用例データベースとして用いるには少々情報不足という感は否めない。

**Virtual Language Centre
Web Concordancer**
<http://vlc.polyu.edu.hk/default.htm>

香港工科大学の Chris Greaves のページは web-based の語学教材サイトとして非常に有名なものだ。そのメイン・メニューの1つがこの Web Concordancer である(図4)。

この Web Concordancer のリンクをクリックすると、さらに詳細な検索ページに移動し、ここでは英語、フランス語、日本語、また英語・中国語の対訳テキストも検索できる。Web 上で提供されている無料の検索サービスとしては非常にコンテンツが豊富である。

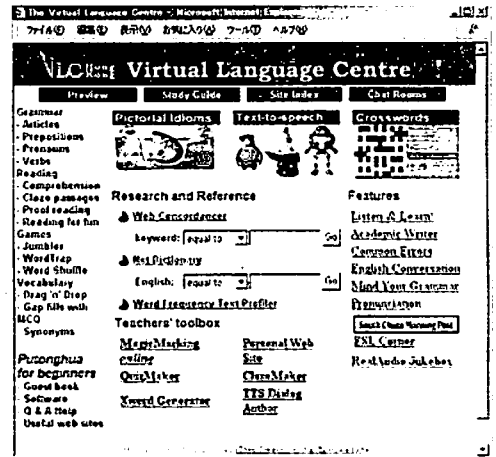


図4: Virtual Language Centre のメイン画面

ただし、コーパスの中身を見ると、Brown Corpus や TIMES というライセンスや著作権取得が必要なテキストが一般向けに提供されており、クレームがないからまだよいものの、実際は public domain にこういった類のテキストを置くのは問題があると言わざるをえない。我々も便利だからといってインターネットでダウンロードしたテキストなどをもとに自作コーパスを作って配布したりしたくなるだろうが、あくまでも個人利用の範囲内でのみ許されていることで、第三者に配布す

際には著作権に抵触することを覚えておく必要がある。

さて、とりあえず実際の検索を簡単にやってみなければ、メインページの中央にある[keyword]に単語を入力して、[Go]ボタンを押せば勝手に検索してくれる。

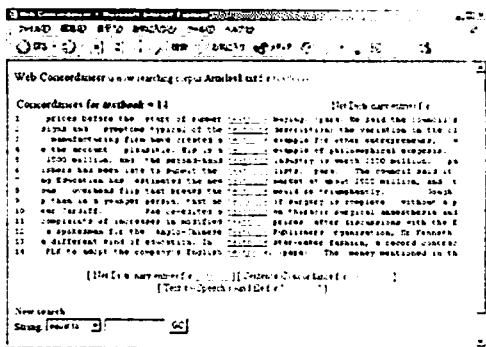


図5：Web Concordancer 検索結果画面

図5がその検索結果である。画面はKWIC (Keyword in Context) という形式で出力され、検索した単語が中央にそろって並ぶので見やすい。この画面から、検索語のオンライン辞書に飛べたり、発音が聞けたりと機能も豊富だ。

検索のコツ

VLC Web ConcordancerはBrown, LOBなど研究用にしか利用できないコーパスがオンラインで検索できる数少ないサイトだ。最初はメイン画面からのキーワード検索を使っているいろいろ試してみるとよい。そのうち、キーワードとソート(検索語の前の単語と後ろの単語のどちらでソートするか選べる)をうまく行えば、コロケーションの簡単なパターンを見たりすることも可能になる。

検索結果画面で、コンテキストが見たい場合には、中央の検索語をクリックすると出典のファイルの文脈を読むことができる。また、KWICではなく、文単位の用例が欲しいという場合には、

検索結果画面の下に表示される [Sentence Concordancer for ...] というリンクをクリックすると、同じ検索語で文単位の用例が表示される。

CALL (computer-assisted language learning) 環境で英作文などをさせている先生は、このVLCのリソースをいろいろ活用されるといだろう。特に少しまとまった自由英作文などをさせるときには、やはり英和・和英辞典の例文だけでは物足りない。作文の中にトピック語彙としていくつかターゲットの単語を設定し、それらを辞書で引いて得られる情報と、辞書だけでは不十分な情報に分け、後者の場合にはオンラインのコーパスで用例を読ませる、といった工夫が欲しい。意味を知っていても、使い方が案外わからなかった単語も10~20の用例を一度に読ませれば、使い方のイメージが膨らみ、定着度も高いはずだ。

まとめ

今回は実際のコーパス・データが到着するまでの間、インターネット上のリソースを使いたいという人のために3つのwebベースのコーパス検索サイトを紹介した。

この他にもMICASEのようにESP/EAP用途のコンコーダンサーなども含めれば、まだまだ紹介できるサイトがあるが、今回は汎用的なサイトを紹介するに留めよう。

次回からはいよいよ自分の手元にコーパス・データを置いて検索する際の具体的なヒントや方法を伝授したい。ぜひ、ご自分でいろいろ触ってみて、教材研究や授業への応用を試みていただきたい。