

英語教師のための コーパス入門

第4回

コンコーダンスを 活用する



明海大学 助教授
投野 由紀夫

さて、今回はコーパス・データから自分の知りたい単語の用例を抽出するコンコーダンス機能 (concordancing) について解説しよう。コーパスを利用する強みの1つは、我々が教材研究、言語タスクやテストの作成などで英文の用例を見たいときに、大量のデータから一瞬にして用例を抽出でき、それを自分の知りたいことに合わせて分析できるという点にあるだろう。今回は基本的な用例検索の方法、そこから発展した用例のいろいろな分析方法を簡単にまとめて紹介したい。

単語の用例を調べる

まずは手始めに単語の用例を検索してみよう。今回も、標準ツールとしては WordSmith を利用する。検索には 1990 年代のアメリカ英語を 100 万語集めた Frown Corpus を利用している。このコーパスは連載の初回で紹介した ICAME Corpus Collection という CD に含まれている。

例として、run という動詞の使い方を調べてみる。

- 1) WordSmith を起動する。
- 2) Tools Controller の画面で [File] - [Choose Texts] を選択 → Choose Texts のウィンドウが開くので、Frown のデータを選択して OK を押す。
- 3) Tools Controller の [Settings] - [Adjust settings] を選択し、[Tags] のタブを開く。
- 4) [Only part of file] をクリック → 一番上の [Cut start of each line...] というところに "9" という数字を入れて OK を押す → これで Frown の各行先頭のファイル属性情報をカットできる。
- 5) Tools Controller から [Tools] - [Concord] を選択 → Concord の画面が開く。

- 6) [File]-[Start]を選択 ⇒ [specify search-word]をクリック。
- 7) 図1のように調べたい単語を入力する。
- 8) [Go Now!]をクリック

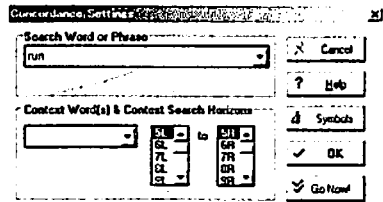


図1: WordSmith Concordの検索画面

以上の手順でコンコーダンス・ラインが現れる。結果は図2のようになる。

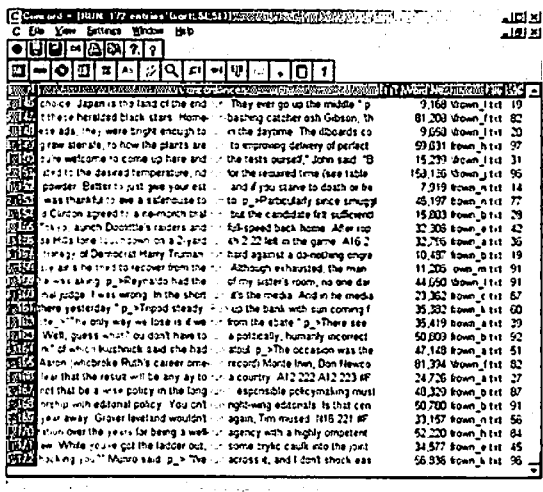


図2: runの検索結果

用例を見やすく整理する

172例が run の用例で出てきた。これを1件ずつ見ていくのもよいが、ここではいくつか整理する方法を覚えておこう。

(A) 用例もとの文脈を見たい

コンコーダンス・ラインだけではその使い方が十分理解できないこともある。Concordの画面で、メニュー一番左の2つのピンクのボタンがあり、これがコンコーダンス行の幅を拡げたり (grow)

縮めたり (shrink) する機能がある。試してみたい。Grow ボタンを押すごとにラインが1行から2行、3—4行と増えていき、文脈を読み込むことが可能だ。

(B) コンコーダンス・ラインをソートする

runと共起する単語を知りたいときにはソート (sort) 機能を用いる。これはどんな単語がよく一緒に使われるかを視覚的に知るのに有効だ。たとえば、runの右側、つまりrunの直後にくる要素でソートをかけてみよう。

1) コンコーダンス結果画面のメニューボタンを押すと図3のような設定画面が現れる。

- 2) そこに runの右1語目 (1R) を第1ソート、次に2語目 (2R) で第2ソートと指定する。
- 3) これで再度、OK を押せばコンコーダンス画面が図4のように右側で整列されて表示される。

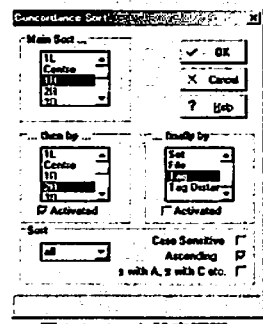


図3: ソート設定画面

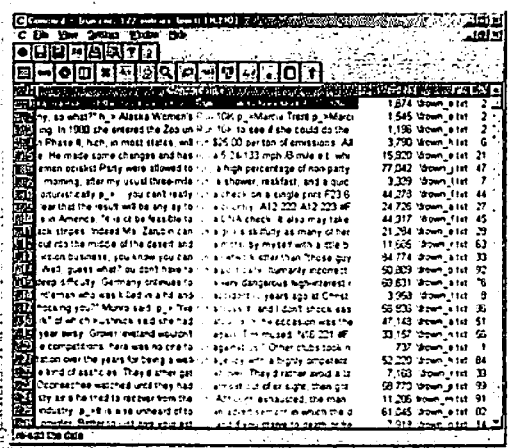


図4: ソート後のコンコーダンス画面

のクラスターは in the long で2番目が the long run で、これらがつながられれば in the long run だとわかるが、このようなものが散見される場合もある。

不要なクラスターはカーソルをその位置にもって行き、削除してしまえばよい。

データ駆動型の学習

以上のようなコンコーダンスの機能ははたして中高の英語教師にどのように役立つのだろうか？ その1つの事例として、検索語をブランクにする機能を紹介してみよう。

WordSmith やその他いくつかのコンコーダンスには検索語をブランクにする機能がある。これを用いると、文法のポイントを単に教師のほうで一方的に説明するよりも効果的に提示することが可能だ。図7、8はWordSmithでanyとsomeを検索して、それぞれの検索語の右側でソートし、かつほぼ同じような連結語の部分で検索

語をブランクにした画面である。

このような画面を生徒に見せて、some と any のどちらが入るかを考えさせる。単純に次の単語が単数か複数かというような基準だけでは正解は出せない。Some の場合には came のように次が動詞であったりして、決定詞だけでなく代名詞の使い方も頭に入れておかねばならない。Any の場合も単数・複数両方ありその意味の違いにも注意しなければならない、など多面的な理解が必要になる。

このような導入の方法をデータ駆動型学習 (Data-driven Learning) という。決して万能の学習方法ではないが、この some/any の対比のように、一定の英語力レベルの生徒に、語の結びつきや規則性を帰納的に学習させる際には威力を発揮する。中学レベルの学習者の場合は、コンコーダンスを読み込むのは限界があるので、一般の英語コーパスを素材に使うのは難しい。しかし、同様の方法で自作コーパスを使って、たとえば中学の教科書のデータを電子化して持っていれば、用例がやさしいので十分初級レベルの学習者にも活用できるはずだ。

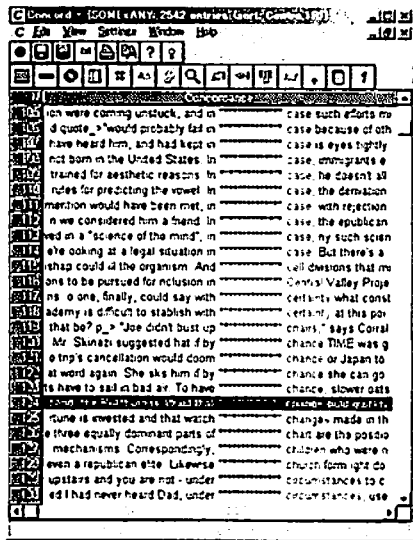


図7: any のブランク画面

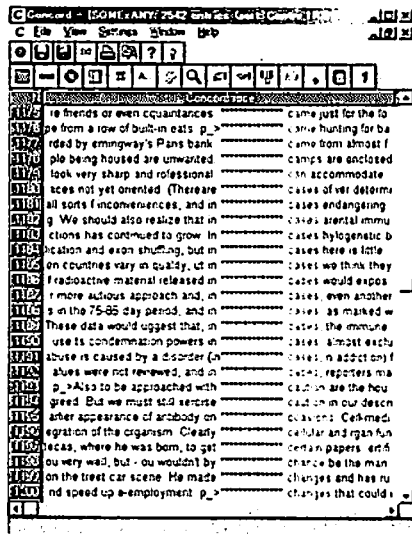


図8: some のブランク画面

★ ★ ★

今回はコンコーダンスの活用の初歩的な解説をした。次回は一般の英語コーパスを学習者の作文データなどのコーパスと連動させることで、より現場に密着した指導のヒントを考えてみよう。