

のが常によいとは限りません。高頻度の語や一般的な構文を調べる場合にはそれほど大きなコーパスは必要ありませんが、頻度はそれほど高くはないが、使用パターンやコロケーションを知りたい語や周辺の構文を調べるときには大規模コーパスが必要です。調べたいことに対して正しいサイズのコーパスを使用することがコーパス検索成功の鍵となるでしょう。

第4章 汎用コーパス COCA の使用法

4.1 はじめに

この章では Corpus of Contemporary American English (COCA) の操作方法とその活用法を説明します。COCA は Brigham Young 大学の Mark Davies によって提供されているサイト, CORPUS.BYU.EDU で公開されている7つのコーパスの1つです。操作方法是共通ですので, COCA の検索方法を学べば他の6つのコーパスも同じように扱うことができます。詳細な説明や検索例は, 初期画面の [Help/information/contact] のメニューから見ることができます。特に [GENERAL] の Brief tour (general) は簡潔にまとめられているので, COCA を使って何ができるのかを具体例を見ながら手早く学ぶことができます。

表1 CORPUS.BYU.EDU で利用可能な英語コーパス

コーパス名	語数	国・地域	時期
Global Web-Based English (GloWbE)	1.9 billion	20 countries	2012-13
Corpus of Contemporary American English (COCA)	450 million	American	1990-2012
Corpus of Historical American English (COHA)	400 million	American	1810-2009
TIME Magazine Corpus	100 million	American	1923-2006
Corpus of American Soap Operas	100 million	American	2001-2012
British National Corpus (BYU-BNC)	100 million	British	1980s-1993
Strathy Corpus (Canada)	50 million	Canadian	1970s-2000s

4.2 利用者登録

CORPUS.BYU.EDU (<http://corpus.byu.edu>) にウェブブラウザでアクセスし、COCA を選んですぐに使用を開始することができますが、10-15 回程度検索を続けると、登録を促す画面が現れます (あるいは画面右上の [LOG IN] の横の [REGISTER] を選べば登録ページを表示できます)。ユーザー登録をすることにより、コーパスの使用を続けることができますようになります。登録者はその身分によって 1 (非研究者), 2 (準研究者), 3 (研究者) にランク分けされます。このランクによりアクセスレベル (1 日に実行できる検索回数, 1 ページに表示できる KWIC コンコーダンスの件数, 1 日に得ることができる KWIC コンコーダンスの件数の上限, 1 日に保存できる語数) が決められます。

ユーザー登録するには、登録画面に氏名、メールアドレス、パスワードを入力し、身分 (Category) を選びます。2 (準研究者) 以上のランクの Category を選択すると、新たに 3 つの記入事項 (Organization, Web page, Profile) が追加されます。2 以上のランクに分類されるためには、所属機関名と自分の氏名が掲載されている所属機関のウェブサイトの URL, 簡単な紹介文を入力する必要があります。登録したメールアドレスに確認のメールが届くので、当該部分をクリックすれば、登録は完了します。まずはランク 1 からスタートしますが、所属機関の情報が承認されれば、ランクが引き上げられ、より多くの検索・閲覧ができるようになります。

4.3 基本的な検索方法

4.3.1 インターフェース各部の働き

COCA のインターフェースは 3 つのフレームで構成されています (次ページ図 1)。画面左側 (次ページ図 2) は検索語の入力や条件の設定などを行う操作フレーム、右の 2 つのセクションは各種結果を表示するフレームです。右上 (⑤) は頻度情報, ジャンルや年代毎のグラフ, コロケーションなどを表示する統計値フレーム, 右下 (⑥) は LIST や KWIC 形式で用

例を表示する用例フレームです。またこのフレームは右端の [Help/information/contact] タブを選ぶことにより、ヘルプページの表示にも使われます。

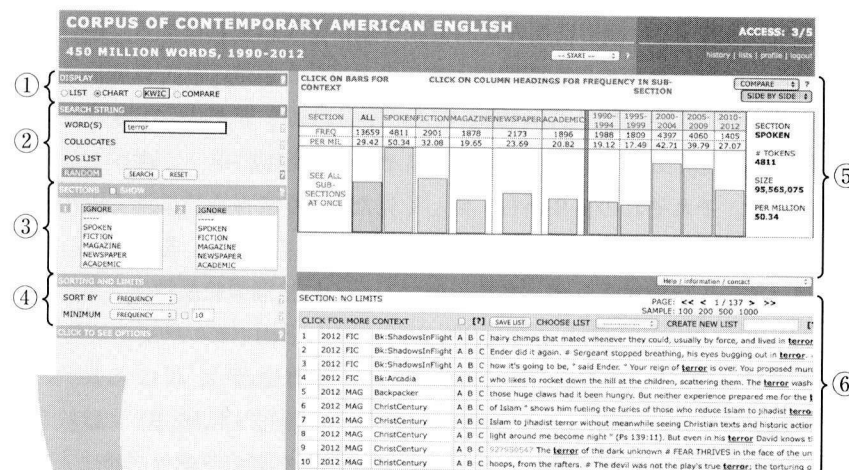


図 1 COCA のインターフェースと terror を検索した結果

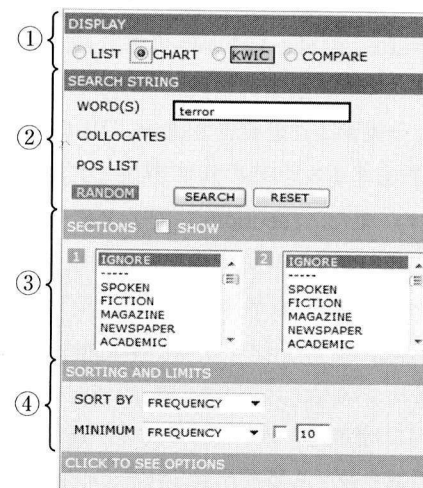


図 2 操作フレーム (拡大図)

操作フレームには様々なボタン、ボックス、タブがあります。図1の番号順にそれぞれの働きの概略を説明します。基本的な検索方法は本節で、活用法については4.4で説明します。さらにCOCAを使った用例作成については第5章を参照して下さい。

①【DISPLAY】検索結果の表示方法を選びます。それにより、統計値フレームにはそれぞれ異なる情報が表示されます。[LIST]では頻度情報や共起語リストが、[CHART]ではジャンル・年代別のグラフ、[COMPARE]では検索した2語のそれぞれの共起語リストが並べて表示されます。それぞれの結果から、詳しく見たい語をクリックすると、用例フレームに当該の用例が表示されます。ただし、[KWIC]を選ぶと、統計値フレームは使用されず、直接用例フレームに検索の結果抽出された例文がKWIC形式で表示されます。

②【SEARCH STRING】検索文字列の入力をするところです。[WORD(S)]([COMPARE]選択時にはボックスが2つになる)には検索する語・句をそのまま、あるいは文字列を制御する様々な記号と共に入力します。[COLLOCATES]は検索文字列と特定の語句や文字列が共起する例を探したいときに使います。[POS LIST]からは検索語の品詞を指定できます。また、[RANDOM]をクリックすれば、コーパス内の語から無作為に抽出した語に関する検索例が表示されるので、新しい検索方法が学べ、新たな発見につながることもあります。

③【SECTIONS】検索する文字列の用例をジャンル・年代で絞り込みたいときに使用します。また、[sections]で1,2にそれぞれ別のジャンルを選べば、ある文字列のジャンル間での比較ができます。セクションはCtrl+クリックで複数個選択できます。

④【SORTING AND LIMITS】ソートの基準(Frequency/Relevance/Alphabetical)や、検出する際の最低値(頻度, MI値:詳しくは16.4.2, 17.2.2)を設定します。[LIST]で[KWIC]を選ぶと、【DISPLAY/SORT】に変わり、用例を並べる順番を設定することができます。

操作に慣れるために、例えばterrorという語が、どのジャンルでいつ

頃よく用いられたかを調べてみましょう。[CHART]を選び、[WORD(S)]のボックスにterrorと入力し、[SEARCH]ボタンをクリックするだけです(図1, 2)。話し言葉(SPOKEN)でもっともよく使用され、2000年以降使用頻度が上がっていることが見て取れます(⑤)。グラフのジャンル名をクリックすれば、そのジャンル内でのさらに細かい分布が示されます。グラフの棒の部分をクリックすれば、用例フレームに当該ジャンルの用例が表示されます(⑥)。表示された用例のヘッダ部分(年代やジャンルが表示されている部分)をクリックすると、その用例の出典情報やより広い文脈を見ることができます。

4.3.2 入力上の注意

操作フレームの【SEARCH STRING】下の[WORD(S)]ボックスに語句を入力し、[SEARCH]ボタンをクリックすれば、指定語句の検索ができますが、検索文字列に所有格('s), 縮約形('m, 'n't, 'll)および句読点(., !, ?)を含む場合には注意が必要です。それぞれが1つの単語として扱われるため、その前に半角スペースを入れます。例えば、haveの縮約形'veも単語として扱いますので、they'veはthey' ve('は半角スペースを表す)と入力します。付加疑問won't he?はwo' n't' he?と入力します。アスタリスク(*)はワイルドカードと呼ばれ、検索時に特別な意味を持ちます。詳しくは4.3.4や5.2.2を参照して下さい。また、新しい検索に移るときには必ず[RESET]ボタンをクリックして下さい。

4.3.3 活用形を一括で検索する

コーパスを検索するとき、その語の活用形を含めるかどうかで結果が異なります。例えば、experienceという単語を検索するとき、[WORD(S)]のボックスにexperienceをそのまま入力すると、名詞の単数形あるいは動詞の原形であるexperienceを含む文のみが検索の対象となります。動詞の活用形のexperiences, experiencing, experienced(および名詞の複数形のexperiences)も同時に検索の対象にしたいときには、検索の対象

の文字列を半角の角括弧 ([]) で囲むことでレマ化することができ、検索の対象に活用形を含むことができます。つまり、experience の場合、[experience] として検索することで進行形や過去形なども同時に表示されるようになります。また、「|」で区切ることで OR 検索が可能になります。例えば、experiences|experiencing などとすることで特定の活用形だけを指定して検索することもできます。なお、have など短縮形 ('ve) のあるものは、[have] としてレマで指定することでそれらも一括して検索されます。

4.3.4 品詞を指定する

experience には動詞の用法と名詞の用法があります。[experience] で検索すると、experience と experiences では動詞と名詞ともにヒットしてしましますが、COCA では品詞を指定して検索する方法が用意されています。品詞は、単語の直後にピリオドを打って品詞タグを [] に入れて、[品詞タグ] のように書きます。表2は代表的な品詞タグ(英国ランカスター大学が開発した自動品詞タグ付与プログラム CLAWS7 に基づい

表2 COCA で使用されている品詞タグ(抜粋)

タグ	意味
ND1	方向を表す名詞の単数形(例: north, southeast)
NN	数の変化のない一般名詞(例: sheep, headquarters)
NN1	一般名詞の単数形(例: book, girl)
NN2	一般名詞の複数形(例: books, girls)
NP	数の変化のない固有名詞(例: IBM, Andes)
NP1	固有名詞の単数形(例: London, Jane, Frederick)
NP2	固有名詞の複数形(例: Browns, Reagans, Koreans)
VV0	一般動詞の原形(例: give, work)
VVD	一般動詞の過去形(例: gave, worked)
VVN	一般動詞の過去分詞形(例: given, worked)
JJ	一般的な形容詞

* <http://ucrel.lancs.ac.uk/claws7tags.html> を参考に作成

たもの)を抜粋して示したものです(動詞タグについては5.2.1 参照)。

この表から、N で始まるものは名詞を、V で始まるものは動詞を、J で始まるものは形容詞を表していることが分かります。また、NN で始まるものが一般名詞、NP で始まるものが固有名詞です。複数形 (NN2) や現在分詞 (VVG) などの細かい区分を指定することもできますが、任意の文字列を表すワイルドカード (*) を用いて NN* や V* (大文字・小文字は区別されませんので nn* や v* でも可能です。4.4 では小文字で統一します) とすることで「一般名詞」や「動詞」といった一般的なカテゴリーを指定することが可能になります。なお、このワイルドカードはその部分に文字列がない場合も含みますので、NN* には NN も含まれる点に注意して下さい。

具体例として experience の品詞指定の例を見てみましょう(表3)。例えば、3人称単数の動詞として使用される experiences を検索したい場合は、experiences.[v*] とします。一方、experiences.[n*] とすると、複数形の名詞として使用される場合が抽出されます。また、レマ化と組み合わせで使用することも可能で、[experience].[n*] は名詞として使用される experience, experiences, experiencing を抜き出すことができます。後ろに.[VVN] を添えれば動詞 experience の過去分詞のみを抜き出すことができます。構文検索に品詞指定を利用する方法については5.2.1を参照して下さい。

表3 検索式の例

検索対象	検索式
【動詞】3人称単数の experiences	experiences.[v*]
【名詞】複数形の experiences	experiences.[n*]
【名詞】活用形を含む名詞の experience	[experience].[n*]

4.4 COCA で単語のふるまいを調べる

前節で紹介した COCA の検索機能を使うことで、単語のふるまいや類

義語の違いなどを明らかにすることができます。ここではその方法を紹介しします。

4.4.1 語と語の相性(コロケーション)を調べる

まず特定の語がどのような語と結びつきやすいかその相性を調べてみましょう。この場合、[LIST] 表示を使用します。[WORD(S)] に指定した語を中心に、それと共起する単語を集計することができます。画面上の[COLLOCATES] の文字列をクリックする入力ボックスが現れ、ここで共起する要素を指定します(ボックスを消すには再度クリックするか[RESET] ボタンを押します)。**[COLLOCATES]** に何も入力しなければ検索を実行したときに*が自動的に入力されます。これは「任意の単語」という意味で、すべての品詞の共起語を対象に集計します。また、**[COLLOCATES]** は品詞を指定して共起語を集計することもできます。品詞の指定は **[POS LIST]** の文字列をクリックすると入力欄と品詞の一覧を含んだドロップダウンリストが出てきますので(図3)、ここから選択すれば、品詞のコードが自動的に検索ボックスに入力されます。**[COLLOCATES]** の直後にある数字は(図4)、**[WORD(S)]** に指定した語の前後何語までを集計に含むかを指定するものです。

具体例として、動詞の experience の後にくる名詞で目的語として使われていると考えられるものを集計する場合(次ページ図5)について見てみ

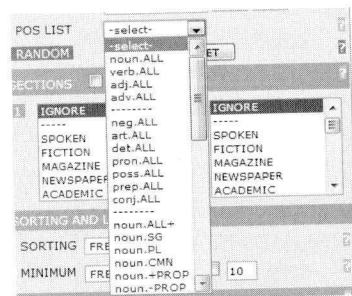


図3 COLLOCATESのPOS LIST

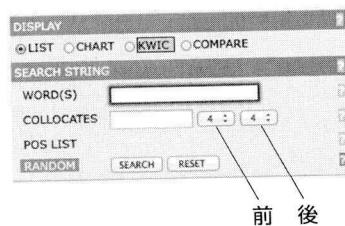


図4 コロケーションの検索画面

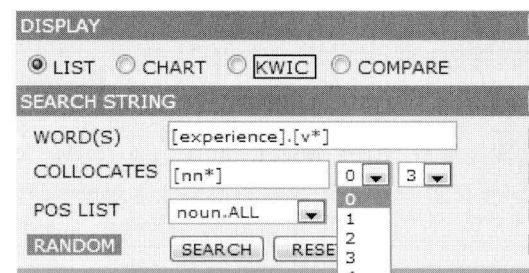


図5 experience (動詞) のコロケーションの検索

みましょう。まず、[LIST] 表示を選択し、**[WORD(S)]** に [experience][v*] を指定します。これがコロケーションの中心となる単語になります。次に **[COLLOCATES]** に名詞を指定するため、直下にある **[POS LIST]** から noun.ALL を選択すると、入力欄にタグが自動的に挿入されます。あるいは、[nn*] を直接入力してもかまいません。ここでは後ろにくる名詞を集計することが目的ですので、前の数字は0、後ろの数字は3を指定します。「直後」のみを集計するためには1となりますが、名詞の場合、冠詞や修飾語が入ることがありますので、少し余裕をもって3としておきます。

例えば、後ろのスパンを3とすると、experience problems (後1の名詞), experience a problem (後2の名詞), experience a serious problem (後3の名詞) までを含み、それらを合わせて集計します。

この条件で検索すると、動詞 experience の目的語として共起しやすい名詞のリストを得ることができます。

problems, life, pain, growth, difficulties, symptoms, levels, loss, stress, difficulty (頻度上位10語)

このように [LIST] 表示のコロケーション集計を使うことで特定の動詞と共起しやすい名詞、あるいは特定の動詞と共起しやすい副詞、特定の名詞と共起しやすい形容詞など、品詞ごとに傾向をみることができます。

さらなる例として, experience と共起する副詞と形容詞の検索について見てみましょう。動詞の experience と共起する副詞を検索する場合は, [WORD(S)] に [experience].[v*], 前後 3 のスパンで [COLLOCATES] に [r*] (副詞) を指定すると, ever, never, more, also などが上位にきます。名詞の experience と共起する形容詞を検索する場合, [WORD(S)] に [experience].[n*], 前 1・後 0 のスパンで [COLLOCATES] に [j*] (形容詞) を指定すると personal, human, sexual, past などがヒットします。

この検索方法を使えば, 生徒の英作文中の表現をチェックすることができます。例えば「充実した 1 日でした」の訳だと思われる It was a massive day. という作文があったとします。確かに和英辞典で「充実」を引くと massive という語が出てきますが, 果たして a massive day というのは自然な表現なのか確かめてみましょう。

[LIST] 表示を選択し, [WORD(S)] に [massive] を, [COLLOCATES] に [nn*] を指定し, 後ろのスパンを 1 とすると, massive の共起語リストが得られますが, day は含まれていません。また, [COLLOCATES] に day を入れて検索してみても, massive の直後に day が使われている例はわずか 4 例しか検出されず, その 4 例を見ても「充実した」という意味で massive が使われている例はありません。例えば, 用例 1 “Sept. 11, all the people who died that day. It was just a massive day, a black scar on the history of the United States...” は 9.11 の連続テロ事件の日を「アメリカ合衆国の歴史に深い傷跡を残した重大な日」だとして a massive day を使っています。これらのことから, a massive day は「充実した 1 日」という意味で使われる表現ではないと判断できます。代わりに表現を調べるには, [WORD(S)] に [day] を, [COLLOCATES] に [j*] を指定し, 前のスパンを 1 とします。a full day, a great day, a busy day などを使うように指導できます。

4.4.2 類義語の使い分けを調べる

互いに意味が類似した単語でも, 共起語は異なります。共起語の違いが

The screenshot shows the COCA interface with the following data:

CORPUS OF CONTEMPORARY AMERICAN ENGLISH									
450 MILLION WORDS, 1990-2012									
ACCESS: 3/5									
CHOOSE CONTEXT: CLICK ON NUMBERS (WORD 1 OR 2)									
WORD 1 (W1): BIG (1.59)					WORD 2 (W2): LARGE (0.53)				
WORD	W1	W2	W1/W2	SCORE	WORD	W1	W2	W1/W2	SCORE
1 DEAL	5486	12	457.3	288.5	1 NUMBER	4213	333	12.7	20.1
2 DIFFERENCE	2237	85	26.3	16.6	2 BOWL	3436	207	16.6	26.3
3 PROBLEM	2139	116	18.4	11.6	3 PART	3383	1807	1.9	3.0
4 MAN	2059	574	3.6	2.3	4 NUMBERS	3311	345	9.6	15.2
5 QUESTION	2053	56	36.7	23.1	5 SKILLET	1513	9	168.1	266.5
6 BUSINESS	1884	164	11.5	7.2	6 AMOUNTS	1486	13	114.3	181.2
7 PART	1807	3383	0.5	0.3	7 PEOPLE	1255	887	1.4	2.2
8 PICTURE	1738	120	14.5	9.1	8 HEAT	1094	28	39.1	61.9
9 THING	1622	38	42.7	26.9	9 COMPANIES	1080	1483	0.7	1.2
10 TIME	1575	249	6.3	4.0	10 EGGS	1051	21	50.0	79.3

図 6 big と large の共起語比較

分かれば, 両者の単語を使い分けることができます。検索モード [COMPARE] で big と large の相違を調べてみましょう (図 6)。

[COMPARE] を選ぶと, [WORD(S)] にボックスがもう 1 つ追加され, 2 語間の共起語の比較を行うことができます。一方に big を, もう一方に large を入れ, 前 0・後 4 のスパンで [COLLOCATES] に [nn*] (一般名詞) を指定します。

一方の語にのみ強い共起傾向を示す語は緑で, 両方に共通して共起する語はピンクでハイライトされます。big には deal, difference, problem などが, large には number, bowl, amount などが強い共起傾向を示していることが分かります。日本語では同じ「大きい」という語で表される 2 語ですが, その語の後ろにはどんな語がくるのかを示すことで, 「問題」, 「差」が大きいときには big, 「数」や「量」が大きいときは large というように, 類義語の使い分けを教えることができます。

4.4.3 ある単語と同じように使える別の語を探す

CORPUS.BYU.EDU には 6 万近くの語の類義語情報が組み込まれていますので, 類義語検索を行うことができます。[LIST] 表示で [WORD(S)] のボックスに [=hard] のように [= と] で検索語を囲んで入力すると,

その語の類義語のリストが頻度順に表示されます。ここまでは単に類語辞典(シソーラス)を引いているのと同じことですが、この類義語検索を応用すれば、生徒の表現を豊かにすることができます。例えば「きつい仕事」を生徒に英語で言わせると、ほぼ全員が hard work と答えるでしょう。hard 以外にもっとびったりの形容詞はないのでしょうか。そのような場合、[hard] work と入力すれば、difficult, challenging, tough, demanding などの語がたちどころに見つかります。同じように、[good] idea, [new] idea, [bad] idea でも調べてみて下さい。

少し高度ですが、[[=clean]][v*] the [n*] と入力すれば、動詞 clean ([v*]) のすべての活用形 ([]) の類義語 ([=clean]) で、そのあとに the の付く名詞 (the [n*]) を目的語にとる表現を調べることができます。その上、clean の類義語の動詞がどんな目的語を取るのか、すなわちどこをきれいにするのかによって、どんな動詞を使うことができるのかがよく分かります。wipe the sweat/tears, clean the house/air, mop the floor などがある例です。

4.4.4 コーパス間の比較で地域差を調べる

統計値フレームの右上端に [COMPARE] と [SIDE-BY-SIDE] というタブがあります。これらは同じ検索を CORPUS.BYU.EDU のサイト内の COCA とは別のコーパスと比較するためのツールです。[COMPARE] タブからコーパスを選べば、別のコーパスに切り替わり、同じ文字列・条件で検索した結果が表示されます。[SIDE-BY-SIDE] という機能を使えば、2つのコーパス間の結果が比較できます。図7(次ページ)は good の直前に来る副詞を検索し、[SIDE-BY-SIDE] で BNC と比較した結果を示しています。RELEVANCE でソートしてあるので、それぞれのコーパスにより強い共起関係を示す語が上位に並んでいます。COCA では real good や mighty good が、BNC では okay good, outstandingly good, bloody good, jolly good などが多く使われているのが分かります。これはアメリカ英語とイギリス英語の特徴的な差を表していると言えるでしょう。

Figure 7 shows the BYU-BNC: BRITISH NATIONAL CORPUS * interface. The search results for 'good' are displayed in a table comparing COCA (450,000,000 words) and BNC (100,000,000 words). The table lists various adjectives and their frequencies in both corpora.

WORD/PHRASE	1: COCA	2: BNC	PM 1	PM 2	RATIO
1 REAL	40	1	0.09	0.01	8.99
2 MIGHTY	33	1	0.07	0.01	7.33
3 MOSTLY	61	2	0.14	0.02	6.78
4 FUNDAMENTALLY	23	0	0.05	0.00	5.11
5 ENOUGH	45	2	0.10	0.02	5.00
6 ALMOST	20	0	0.04	0.00	4.44
7 OFF	19	1	0.04	0.01	4.22
8 PRETTY	8220	464	18.27	4.64	3.94
9 ABOUT	17	1	0.04	0.01	3.78
10 EXACTLY	17	1	0.04	0.01	3.78

図7 good を修飾する副詞のコーパス間比較

4.5 リストの保存

最後に、用例の保存について紹介します。検索の結果得られた用例フレームの中の用例から保存したい用例をクリックで選択(あるいは[SAVE LIST] ボタンのすぐ右のチェックボックスをクリックして全選択)し、[CREATE NEW LIST] のボックスにリストにつけたい名前を入力して、[SAVE LIST] をクリックすれば、用例のリストをBYUのサーバー内に保存することができます。ただし用例自体はダウンロードできません。保存したリストは、後から[CHOOSE LIST] から選んで再利用することができます。また、各用例のヘッダの後にある ABC というアルファベットは、クリックして用例を色分けすることができる機能です。リストを意味や用法などで分類したり、並べ替えたりしたい場合のグループ分けに使用できます。

4.6 おわりに

COCA を含む7つのコーパスと多様な検索機能の両方が備わっている CORPUS.BYU.EDU は、インターネット環境があればブラウザーだけで英語の様々な側面を調べることができます。しかしながら機能が豊富な

め、自由に使いこなすには時間がかかるかもしれません。英語の表現に関する疑問が生じたら辞書を引くのと同じように、毎日辞書代わりに使うことをお勧めします。そうしているうちにそれぞれの機能が理解でき、機能と機能を組み合わせることで高度な分析ができるようになります。

第5章 汎用コーパス COCA を使った例文作成

5.1 はじめに

新しい文法項目を導入するとき、定着を図るためには様々な例文を提示することが有効だと考えられます。例えば、未来を表す仮定法の *were to* がどのような動詞とともに使われているかを示せば、具体的なイメージを持たせることができるでしょう。しかしながら、辞書や参考書など手元の資料を見ると例文が似ていて、十分なバリエーションを確保できないなど、適切な例文が見つからないこともあります。

本章の目的は、コーパスから、感嘆文、不定詞、仮定法など、文法項目に沿った例文を効率的に抜き出す方法を提示することです。そのためには4章で紹介した検索方法から一歩進んで、文法形式を定式化し、検索式を考えるということが必要になります。以下では学校文法の枠組みに沿った形で、文法形式の定式化の方法と、COCAでの検索結果および例文の作成方法について示していきます。検索の方法が理解できたら実際に手を動かしてCOCAで検索してみてください。従来、文法は統語構造が中心に研究されてきましたが、コーパス言語学の発展により、語彙を中心とした文法研究に注目が集まるようになってきています。本節で提示する結果を見れば、語彙と文法が密接に関連していることが分かるでしょう。

5.2 文法項目を定式化する

コーパスから文法項目に沿った用例を抜き出すには、語句を入れて検索する場合と比較して一段階高度な処理が必要です。文型、時制、完了形、助動詞などの文法項目は、多くの場合「パターン」が存在しますが、これ