

# 汎用コーパス COCAの使用方法

6211666

外国語学部ドイツ語専攻 4年

大熊 桃子

# 目次

1. COCAのプロフィール(Corpus of Contemporary American English)
2. 利用者登録
3. 基本的な検索方法
  - 3.1 インターフェース各部
  - 3.2 入力上の注意
  - 3.3 活用形を一括検索
  - 3.4 品詞指定
4. COCAで単語のふるまいを調べる
  - 4.1 コロケーション
  - 4.2 類義語の使い分け
  - 4.3 類似表現の検索
  - 4.4 コーパス間比較
5. リストの保存
6. おわりに

# 1. COCAのプロフィール

(Corpus of Contemporary American English)

- 1990年～  
2000万語／年ずつ拡張
- 2013年8月現在、4億6000万語
- 無料(その中でも最大規模)
- アメリカ英語
- balanced Corpus
- 5つのジャンルから構成  
(spoken/fiction/大衆紙/新聞/学術誌)

# 1. COCAのプロフィール

(Corpus of Contemporary American English)

- CORPUS.BYU.EDUサイトで公開されている7つのコーパスのうちの一つ
- COCAとその他の6コーパスの使用法は同じ  
→COCAの使用法をマスターすれば、ほかのコーパスも操作できるようになる

## 2. 利用者登録

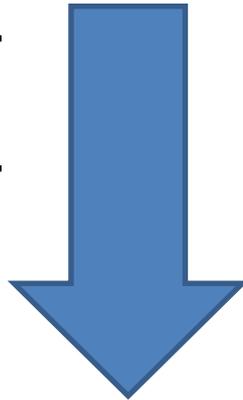
- ① CORPUS.BYU.EDUのWebサイト  
(<http://corpus.byu.edu>)にアクセス
- ② 画面右上 [REGISTER] から、  
Name/E-Mail adress/Password/Category(身分)を登録
- ③ 登録メールアドレスに完了メール→リンクをクリック

完了

## 2. 利用者登録

### ※ Category(身分)に関して

1. 非研究者
2. 準研究者
3. 研究者



- 1日に実行できる検索数
- 1ページに表示できるKWICコンコーダンス数
- 1日に得ることができるKWICコンコーダンス数
- 1日に保存できる語数

注: KWICコンコーダンス(テキストp.5)

...中央にKey Wordを置き、その前後に一定の長さのコンテクストを表示する形式  
(Key Word In Context)

## 2. 利用者登録

非研究者の場合

検索数... 50語／日 (研究者...上限400)

KWIC数... 5000／日 (研究者...上限20000)

準研究者へのランクUPには、

- ・所属機関名
- ・自分の氏名が記載された所属機関のサイトURL
- ・簡単な紹介文

以上を追加する必要がある。

# 3. 基本的な検索方法

## 3.1 インターフェース各部

### ・各部の名称

左： 操作フレーム（検索語入力／条件設定）

右上： 統計値フレーム

（頻度情報／グラフ／コロケーション）

右下： 用例フレーム

（リスト、KWIC形式で表示）

# 3. 基本的な検索方法

## 3.1 インターフェース各部

### ・操作フレーム

#### ①[DISPLAY] 検索結果の表示方法を切り替える

[DISPLAY]内のコマンド	統計値フレームに表示される情報
[LIST]	頻度情報／共起語
[CHART]	ジャンル・年代別のグラフ
[COMPARE]	検索した2語のそれぞれの共起リスト
[KWIC]	検索の結果、抽出された例文 (KWIC形式) ※用例フレームに表示される

# 3. 基本的な検索方法

## 3.1 インターフェース各部

### ②[SEARCH STRING] 検索文字列の入力

項目	操作
[WORD(S)]	検索語句(+記号)を入力
[COLLOCATES]	検索文字列⇔特定の語句の共起例を探す
[POS LIST]	品詞を指定
[RANDOM]	コーパス内から無作為に語を抽出して検索

# 3. 基本的な検索方法

## 3.1 インターフェース各部

### ③[SECTIONS]

検索文字列の用例をジャンル・年代別に絞り込む

- 1, 2にそれぞれ別ジャンルを選んで比較可能
- Ctrl+クリックで複数条件を選択可能

### ④[SORTING AND LIMITS]

- ・ソートの基準[FREQUENCY/RELEVANCE/ALPHABETIAL]を設定
- ・検出時の最低値を設定
- [LIST]で[KWIC]を選択した場合、[DISPLAY/SORT]に変わり、用例の並び替えが可能

# 3. 基本的な検索方法

## 3.2 入力上の注意

-縮約形、所有格、句読点の入力

... その前に半角スペースを入れる

ex.) 付加疑問文won't he? → wo┐n't┐he┐?

-新しい検索に移るとき

... [SEARCH STRING]下の、[RESET]ボタンをクリック

# 3. 基本的な検索方法

## 3.3 活用形を一括検索

- 検索対象の文字列を[半角/角括弧]で囲む  
→「レマ化」
- “|”で区切る→ OR検索
- レマの中には縮約形も含まれる  
ex) [have]→ ('ve)も検索結果に表示される

# 3. 基本的な検索方法

## 3.4 品詞指定

単語の直後に、“ .(ピリオド)[品詞タグ] ”

-代表的な品詞タグは、テキストp.40に記載

初めの文字	品詞
N	名詞
NN	一般名詞
NP	固有名詞
V	動詞
J	形容詞

# 3. 基本的な検索方法

## 3.4 品詞指定

### ・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*]

\* (ワイルドカード)...任意の文字を表す

# 4.COCAで単語のふるまいを調べる

## 4.1 コロケーション

- ・特定の語と結びつきやすい語を調べる

[LIST]を選択



[WORD(S)]に検索語を入力



[COLLOCATES]に共起する語の要素を入力

-品詞:[POS LIST] から選択

-前後の語数を指定

# 4.COCAで単語のふるまいを調べる

## 4.1 コロケーション

- ・品詞ごとにコロケーションの傾向を調べることが出来る
  - 動詞⇔名詞, 動詞⇔副詞, 名詞⇔形容詞など
- ・英作文中の表現の妥当性を調べることが出来る
  - より適切な表現を探し、提示できる

# 4.COCAで単語のふるまいを調べる

## 4.2 類義語の使い分け

- ・共起語の違い→類似した語の使い分けが出来る

[COMPARE]を選択



[WORD(S)]の2つのボックスに、比較する語を入力



[COLLOCATES]から要素を設定

結果： 一方のみに強い共起→緑  
両方に共起→ピンク

# 4.COCAで単語のふるまいを調べる

## 4.3 類似表現の検索

- ・類語辞典(シソーラス)のように使うことも出来る

[LIST]を選択



[WORD(S)]に、**[=検索語]** を入力



類義語のリストが表示される

# 4.COCAで単語のふるまいを調べる

## 4.4 コーパス間比較

統計値フレーム右上端

[COMPARE]

→別のコーパスに切り替え

同じ条件・文字列の検索結果を表示

[SIDE-BY-SIDE]

→2つのコーパスでの検索結果を並べて表示

ex.) COCA vs BNC で、地域差を調べる事が出来る

# 5. リストの保存

用例フレームから保存したい用例をクリック



[CREATE NEW LIST]のボックスにリスト名を入力



[SAVE LIST]をクリック

→保存したリストは、[CHOOSE LIST]から取り出し可能

# 6. おわりに

豊富な機能...慣れれば使いこなせるようになる  
高度な分析が可能に