

COCA の使用方法について

小橋川 祥吾

コーパスの説明

COCA (CORPUS OF CONTEMPORARY AMERICAN ENGLISH)とは？

- Brigham Young University の Mark Davies 氏によって作られたインターフェイス
- 1990~2009 年までの各年 2000 万語、計 4 億語のアメリカ英語を収録
- 現在も更新中であるため、最近の言語変化を研究するのに最適
- BNC の 4 倍の大きさがある
- spoken , fiction , popular magazines , newspapers , academic journals の 5 ジャンルから各年、同じ分量を収録。1 ジャンル当たり約 4000 万語。
- 詳細リストをダウンロードできる
 - ◇ http://www.americancorpus.org/help/COCASources_2009_08_21.zip
- 各ジャンルは更に詳細に分かれており、特定分野に絞って検索することも可能。
例：spoken ジャンルの中から ABC と CNN と FOX のみを採用
例：fiction ジャンルの中から MOVIE のみを採用
例：magazine ジャンルの中から MEDICINE のみを採用
- **複雑な分析が容易に出来る**。例えば、以下のようなこと。詳しくは（注3）

COCA は以下のサイト上で、誰でも使用することができる。（無料）

◇ <http://www.americancorpus.org/>

10 数回検索すると登録が求められるので、すること。

メールアドレスとパスワードで使用すれば、時間制限はなくなる。

※このコーパスに関する詳しい情報は、全てここに記載されている。

◇ <http://davies-linguistics.byu.edu/personal/>

※また、他の巨大コーパスについてはこちらを参照。

◇ <http://corpus.byu.edu/>

（同様の方法で、1923~2006 の TIME 誌を集めた約 1 億語のコーパスなどがある）

使用方法

（※以下で説明することは、COCA にアクセスした後、“Where Should I Start?”を利用すると、詳しいヘルプも示されるのでわかりやすい。）

◆ はじめに、画面上の各種ツールの説明をする。

次ページは、名詞の laugh に前後 5 語以内で共起する形容詞の種類と頻度を、SPOKEN と NEWSPAPER を 1 つのコーパス、FICTION と MAGAZINE をもう 1 つのコーパスとして比較しようという複雑な検索の実行例であるが、ここではツールの説明を中心に行う。**入力方法や簡単な検索の説明は以降のページで行うので、基本を身につけたい場合は、まずそちらを先に参照してもよい。**

ツールの説明

The screenshot shows a search tool interface with the following sections:

- DISPLAY:** Radio buttons for LIST (selected), CHART, KWIC, and COMPARE.
- SEARCH STRING:** WORD(S) field contains "laugh.[n*]", CONTEXT field contains "[j*]", POS LIST dropdown is set to "adj.ALL".
- SECTIONS:** Two columns of genre lists. Column 1 includes SPOKEN, FICTION, MAGAZINE, NEWSPAPER, ACADEMIC. Column 2 includes SPOKEN, FICTION, MAGAZINE, NEWSPAPER, ACADEMIC.
- SORTING AND LIMITS:** SORTING dropdown is set to FREQUENCY. MINIMUM dropdown is set to FREQUENCY with input fields for 5 and 0.
- HIDE OPTIONS:** GROUP BY dropdown is set to LEMMAS. DISPLAY dropdown is set to RAW FREQ. SAVE LISTS dropdown is set to NO. # HITS is set to 100.

- ・ CHART : 各ジャンルでの頻度及び5年ごとの頻度がグラフ化される
- ・ LIST : 各語彙の頻度順のリストが作成される。また、リスト上で語彙をクリックすると KWIC が表示される。

・ここに、調べたい語彙を入力する。入力方法は下記参照。

- ・ WORD の前後に出現する語彙について調べるときは、ここに入力する。
- ・ CONTEXT (下参照) と書かれたところをクリックすることで、この項目を出現させたり省いたりする。
- ・ 数字は、前後の何語までを検索するかを決定する。

- ・ POS LIST と書かれたところをクリックすることで、この項目を出現させたり省いたりする。
 - ・ 特定の品詞について調べたいとき、ここを使用する。
 - ・ WORD や CONTEXT にカーソルを合わせ、この中から品詞を選択すると自動的にその品詞のタグが入力される。
- (例)CONTEXT にカーソルを合わせておき、POS LIST から adj.ALL を選択すると、画面のように[j*]が入力される。

頻度順かアルファベット順かを選択する。

・何番目までリスト化するかを決める。

・ GROUP BY は WORD なら WORD に入力した語彙の頻度のみ、LEMMA なら LEMMA 化された語彙の頻度が、NONE ならタグの違いごとに頻度がリスト上に示される。

・ 但し、LEMMA 化は WORD への入力の仕方によって可能であり、そちらの方がよい。

(理由)clean と入力し LEMMA を指定すると、[clean]のレマ化された頻度のみが示されるが、[clean]を入力し WORD を指定すると[clean][cleaned][cleaner][cleanest]など分かれて頻度が調べられる。

さらに、clean には動詞、形容詞、副詞があるが形容詞の clean のレマ化されたもののみを調べることも可能。

(※[clean].[j*]を WORD に入力)

- ・ 2つのコーパスを比較するとき使用する。
- ・ 4億語全てから検索する場合は、どちらも IGNORE を選択。

(例)画面上は、FICTION、NEWSPAPER のコーパスと FICTION、MAGAZINE のコーパスの比較を指示している。

※MAGAZINE 中の FINANCIALsection などに限定することも可能

- ・ CONTEXT とは、対象となる語彙の前後に出現する共起語を調べたいときに用いるツールである。
- ・ 使用しない場合は、この項目を消しておく

?を押せば詳しくわかりやすい説明が出てくる。具体例も豊富。

◆ **上の設定で、SEARCH をクリックすると次のような画面が現れる。**

(ここでは、2つのコーパスの比較なので下のように SEC1 と SEC 2が表示されているが、単一のコーパスで調べたときは SEC 1 のみが表示される。また、TOKEN 2, PM2,RATIO は表示されない)

左表で TOKENS1 は SPOKEN,NEWSPAPER での合計の頻度

TOKEN2 は FICTION,MAGAZINE での合計の頻度

PM1 は SPOKEN,NEWSPAPER での 100 万語当たりの頻度

PM2 は FICTION,MAGAZINE での 100 万語当たりの頻度

RATIO は PM1/PM2 の割合。

例えば、下の場合 RATIO0.50 は great が laugh の 5 つ前以内に出現する割合が、spoken と newspaper の合計が,fiction と magazine の合計の半分であることを示す。

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION)

SEC 1: 161,058,671 WORDS

SEC 2: 162,026,972 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1
1	[GOOD]	82	188	0.51	1.16	0.44	1	[GOOD]	188	82	1.16	0.51
2	[BIG]	59	122	0.37	0.75	0.49	2	[LITTLE]	168	14	1.04	0.09
3	[HEARTY]	32	58	0.20	0.36	0.56	3	[BIG]	122	59	0.75	0.37
4	[GREAT]	16	32	0.10	0.20	0.50	4	[SHORT]	120	1	0.74	0.01
5	[EASY]	15	31	0.09	0.19	0.49	5	[SMALL]	77	3	0.48	0.02

◆ **次に、各語彙をクリックすると下表のように KWIC が表示される。**

(単一のコーパスで調べたときも同じ)

KEYWORD IN CONTEXT (KWIC) [FREE] FREQUENCY LISTS More information...

CLICK ON TITLE FOR MORE CONTEXT SECTION: NO LIMITS PAGE: << < 1 / 3 > > SAMPLE: 100 200

1	2009	SPOK	Fox_Susteren	is always perfect to needle. You know, she is always good for a laugh . I think it' s because -- I know
2	2009	SPOK	NPR_FreshAir	was any particular interest to me or to him. It was just a good laugh gimmick and a good romantic g
3	2009	FIC	Bk:Missing	of war were the stories, and this one was good only for a derisive laugh . # Robicheaux poked him in
4	2009	FIC	Bk:EverythingHurts	sell this book in two months? " # The book packager packaged a good laugh . " No, pal, " he said, " g
5	2009	FIC	Analog	Miss Frozen Face, is no help. At least Harry was good for a laugh once in a while. " " There's Verme
6	2009	FIC	Analog	infections; this was well known. Many of the station women had a good laugh when one of them sedu
7	2009	MAG	GolfMag	, tinted blue from being dropped in the toilet. We all have a good laugh until Fulton takes a swing anc
8	2009	MAG	NatlParks	's leader, Nelson is clearly second-incommand. When he isn't enjoying a good laugh or a plug of toba

★また、CHART を選択して SEARCH をクリックすると WORD の各ジャンルでの頻度及び5年ごとの頻度がグラフ化される。(下のグラフは WORD に[[LAUGH.[N*]]を入力したもの)

CLICK ON BARS FOR [LAUGH.[N*]] IN CONTEXT

(HELP)

SECTION	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC	1990-1994	1995-1999	2000-2004	2005-2009
SEE SUB-SECTIONS									
PER MIL	6.40	48.37	13.62	15.57	1.93	15.91	16.69	17.59	18.29
SIZE (MW)	81.7	78.8	83.3	79.4	79.3	103.3	102.9	102.6	93.6
FREQ	523	3809	1134	1236	153	1644	1718	1804	1711

◆ それでは、WORD や CONTEXT を使った、検索の基本を説明する。

1. COCA を使用して特定の語彙の頻度を調べたいとき。

(例)COCA 全体に”dog”は何回出現するか？

- ✓ DISPLAY は LIST を選択。
- ✓ SEARCH STRING の WORD に dog と入力。
- ✓ SECTION は 1, 2 とも IGNORE のままでよい。
- ✓ SORTING AND LIMITS の SORTING は FREQUENCY のままでよい。
- ✓ OPTION の GROUP BY は WORD のままでよい。
- ✓ SEARCH をクリックすると以下の画面が表示される。

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT]

	<input type="checkbox"/>	CONTEXT	TOT <input type="checkbox"/>
1	<input type="checkbox"/>	DOG	34444

3,922

※“dog”を含む文脈を表示したいとき (KWIC 機能) は、DOG をクリックすると次のように表示される。

5	2009	MAG	SportsIll	asked if Savard was out of his doghouse. The Zen-like reply: " A dog has more than one paw. "
6	2009	MAG	SportsIll	bigger than life, was his wife, Danielle, chomping down on a hot dog . After the game Cason gav
7	2009	MAG	SportsIll	was James at a hotel in Los Angeles last summer, busting out some downward dog by the pool i
8	2009	MAG	SatEvenPost	babysitter that no matter what he tells you, little Johnny can not teach the dog to do a handstan
9	2009	MAG	SatEvenPost	much of the time, which makes it fun. As a pet sitter and dog walker for Fetch! Pet Care, she ca

左から番号、年度、ジャンル名、更に細かなジャンル、文脈である。

2. COCA を使用して特定の語彙の共起語を調べたいとき。

(例)COCA 全体に dog の前 4 語以内に white を含む表現は何回出現するか？

- ✓ DISPLAY は LIST を選択。
- ✓ SEARCH STRING の WORD に dog と入力。
- ✓ " の CONTEXT に white と入力。その右の変数バーを、左 4、右 0 に設定。
※CONTEXT 入力ツールが表示されていない場合は CONTEXT の上をクリック。
- ✓ SECTION は 1, 2 とも IGNORE のままでよい。
- ✓ SORTING AND LIMITS の SORTING は FREQUENCY のままでよい。
- ✓ OPTION の GROUP BY を BOTH WORD に変更。
- ✓ SEARCH をクリックすると以下の画面が表示される。

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT]

	<input type="checkbox"/>	CONTEXT	TOT <input type="checkbox"/>
1	<input type="checkbox"/>	DOG WHITE	171

4,520

※“white..dog”を含む文脈を表示したいときは、DOG | WHITE をクリックすると次のように表示される。

1	2009	SPOK	Fox_Baier	family. STEWART: Introducing Bo-bama, America' s first openly gay White House dog . LAUC
2	2009	FIC	Bk:WaterStoneHeart	figure, a woman, striding along the cliffs, a large brown and white dog running circles aroun
3	2009	FIC	Bk:CoventryWedding	was being decided by a bogus coin toss. Except for being white , this dog was nothing like th
4	2008	SPOK	NBC_Matthews	delivered this elegant bit when he came to his search for a new White House dog . Senator C
5	2007	FIC	Analog	have real ailments? How can you tell? " We sat in the White Dog , a self-consciously political

3. COCA 内で 2 つのコーパスを作成し、特定の語彙の頻度を比較したいとき。

(例)COCA に収録されている ABC TV と FOX TV の中で"obama"の出現回数の違いは？

- ✓ DISPLAY は LIST を選択。
- ✓ SEARCH STRING の WORD に obama と入力。
- ✓ SECTION 1 で SPOK:ABC を選択、SECTION 2 で SPOK:FOX を選択。
- ✓ SORTING AND LIMITS の SORTING は FREQUENCY のままでよい。
- ✓ OPTION の GROUP BY は WORD のままでよい。
- ✓ SEARCH をクリックすると以下の画面が表示される。

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION)

SEC 1: 13,132,214 WORDS

SEC 2: 3,883,546 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM
1	OBAMA	1448	2915	110.26	750.60	0.15	1	OBAMA	2915	1448	750.60	110.

※各々の"obama"を含む文脈を表示したいときは、表示したい方の **OBAMA** をクリックする。

4. COCA を使用してレマ化された特定の語彙を調べたいとき。

(例)COCA の SPOKEN&FICTION ジャンルにレマ化された clear は何種類、各何回出現するか？

- ✓ DISPLAY は LIST を選択。
- ✓ SEARCH STRING の WORD に[clear]と入力。
※ここでの、[]がレマ化を指示する。詳しくは後述。
- ✓ SECTION 1 で SPOKEN と FICTION の両方を選択、SECTION 2 は IGNORE のままでよい。
- ✓ SORTING AND LIMITS の SORTING は FREQUENCY のままでよい。
- ✓ OPTION の GROUP BY は WORD のままでよい。
※ここで、LEMMA を選択すると、[clear]の頻度のみになってしまう。
- ✓ SEARCH をクリックすると以下の画面が表示される。

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT]

	<input type="checkbox"/>	CONTEXT	TOT <input type="checkbox"/>
1	<input type="checkbox"/>	CLEAR	31862
2	<input type="checkbox"/>	CLEARED	4373
3	<input type="checkbox"/>	CLEARING	1036
4	<input type="checkbox"/>	CLEARs	819

※ここでの CLEAR は動詞、形容詞、副詞の合計で表示されてしまう。それを解消するための方法は後述。

5. COCA を使用して特定の品詞の種類と頻度を調べたいとき。

(例)COCA の SPOKEN ジャンルと NEWSPAPER ジャンルの動詞 (レマ化済み) の種類と頻度は？

- ✓ DISPLAY は LIST を選択。
- ✓ [方法 1] SEARCH STRING の WORD に[v*]と入力。
[方法 2] " の WORD にカーソルを合わせ、POS LIST をクリック。
表示されたものの中から verb.ALL を選択、クリック。
WORD に[v*]が入力される。
- ✓ SECTION 1 で SPOKEN を、SECTION 2 で NEWSPAPER を選択。
- ✓ SORTING AND LIMITS の SORTING は FREQUENCY のままでよい。
- ✓ OPTION の GROUP BY を LEMMA に変更する。
※先ほどと違い、[[v*]]とした上で、WORD を選択してもレマ化はされない。
- ✓ SEARCH をクリックすると以下の画面が表示される。

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION) [↑]

SEC 1: 81,690,404 WORDS

SEC 2: 79,368,267 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2
1	IS	1023314	691357	12,526.73	8,710.75	1.44	1	IS	691357	1023314	8,710.75
2	'S	908532	293247	11,121.65	3,694.76	3.01	2	WAS	472994	642075	5,959.49
3	WAS	642075	472994	7,859.86	5,959.49	1.32	3	ARE	366535	470560	4,618.16
4	HAVE	578986	339538	7,087.56	4,278.01	1.66	4	SAID	359802	145086	4,533.32
5	DO	526992	177833	6,451.09	2,240.61	2.88	5	HAVE	339538	578986	4,278.01

6. COCA を使用して特定の語彙の共起語を調べたいとき。(発展編)

(例)COCA 全体に[人称代名詞主語]+[think(レマ化)]の後ろ 5 語以内に certainly を含む表現 (I think ...certainly, you thought...certainly etc) は何回出現するか？

- ✓ DISPLAY は LIST を選択。
- ✓ SEARCH STRING の WORD に[pp*][think]と入力。
" の CONTEXT に certainly と入力。その右の変数バーを、左 0、右 5 に設定。
※ここで、[*pp]は 5. の例同様に、カーソルを WORD にあわせ、POS LIST から選ぶとよい。
- ✓ SECTION は 1, 2 とも IGNORE のままでよい。
- ✓ SORTING AND LIMITS の SORTING は FREQUENCY のままでよい。
- ✓ OPTION の GROUP BY を BOTH WORD に変更。
※think を[]でくり、BOTH WORD の方を選択しないと、think や thought や thinks の頻度はわからない。BOTH LEMMA はだめ。
- ✓ SEARCH をクリックすると以下の画面が表示される。

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT]

	<input type="checkbox"/>	CONTEXT	TOT <input type="checkbox"/>
1	<input type="checkbox"/>	THINK CERTAINLY	787.5
2	<input type="checkbox"/>	THOUGHT CERTAINLY	38
3	<input type="checkbox"/>	THINKS CERTAINLY	4

◆ 更に詳細な分析は、以下のような方法でタグを入力し、検索を実行する。

- 語彙をレマ化するとき、[word]とする。

(例) [tall] は tall,taller,tallest / [sing] は sing,sang,singing etc

- シノニムを調べるときは、[=word]で検索すれば、代表的なものは調べられる。

(例)[=strong] は formidable,muscular,fervent etc / [=big] は great,high,large,important etc

※シノニムを更にレマ化したいときは、[[=word]]とする

(例)[[=big]] は great,large の他に greater,larger,largest も出現する。

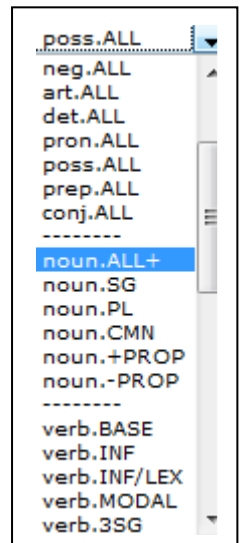
- 品詞を特定したいときは、[word].[品詞タグ]とする。

(例)[clean].[j*] は clean の形容詞

- その品詞全てを調べたいときは、[品詞タグ*]とする。

(例)[*j] は全ての形容詞 / [v?z*] は三人称単数現在形

なお、カーソルを WORD や CONTEXT に合わせ、POS LIST より必要な品詞を選べばタグは自動的に入力される (タグ付けは <http://ucrel.lancs.ac.uk/claws7tags.html> を参照)



- 数語を同時に調べたいときは word|word|word|word のように | で区切る。以下参照

- 特定の語彙を省きたいときは、-word とする。

(例)全ての名詞を除外したい場合は、-[nn*]

- sing,sang,sung のように S#NG を調べたいとき、necessarily,usually のように # # # #LY を調べたいときは、#が一字なら ? を、#が複数なら * を入力する。

(例)*ly は necessarily,usually,wally,ugly etc / s?ng は sing,sang,sung

上の例で、wally,ugly などを除いておきたければ、*ly.[j*] にすればよい。

上の方法をさまざまに組み合わせることで、複雑な検索が容易にできる。

strong のシノニムに名詞の共起語リストは以下のように作成する。

SEARCH STRING

WORD(S)

CONTEXT

POS LIST

SEC 1: 81,690,404 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM
1	[GREAT] [DEAL]	2222	729	27.2
2	[GREAT] [THING]	983	174	12.0
3	[GREAT] [JOB]	893	139	10.9
4	[TOUGH] [TIME]	632	77	7.74
5	[GREAT] [TIME]	596	185	7.30
6	[GREAT] [IDEA]	530	255	6.45

stunning,charminf,gorgeous と名詞の共起。離れた位置での共起を調べる場合は CONTEXT を使用

SEARCH STRING

WORD(S)

CONTEXT

POS LIST

RANDOM

SEC 2: 78,752,154 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2
1	CHARMING SMILE	41	0	0.52
2	STUNNING BEAUTY	12	0	0.15
3	CHARMING CHILD	11	0	0.14
4	GORGEOUS CREATURE	11	1	0.14
5	GORGEOUS EYES	11	2	0.14
6	STUNNING WOMAN	12	3	0.15

8	<input type="checkbox"/>	EXPENSIVE HATS [S]
9	<input type="checkbox"/>	EXPENSIVE SHOE [S]
10	<input type="checkbox"/>	EXPENSIVE SKIRT [S]
11	<input type="checkbox"/>	COSTLY SHOES [S]
12	<input type="checkbox"/>	PRICEY SHIRT [S]
13	<input type="checkbox"/>	EXPENSIVE BLOUSE [S]
14	<input type="checkbox"/>	EXPENSIVE BELTS [S]
15	<input type="checkbox"/>	DEAR SHIRT [S]

WORD に [=expensive] を、CONTEXT に [[devies:clothes]]を入力すると、expensive の意味を表す形容詞に、衣類の意味を表す名詞のレマ化されたものが共起する頻度が検索できる。この場合、両方の語彙を画面上に表したい場合は、OPTION の GROUP BY で BOTH WORD を選択しておく。

ちなみに devies は user customized されたもの。詳しくは http://www.americancorpus.org/help/custom_e.asp

例題 1 (基本編)

1. COCA 全体に”manhattan”は何回出現するか？
2. COCA 全体に girl の前後 4 語以内に lovely を含む表現は何回出現するか？
3. COCA に収録されている 1990~1999 のコーパスと 2000~2009 のコーパスで”terrorism”の出現回数の違いは？
4. COCA の SPOKEN&FICTION&MAGAZINE ジャンルをまとめて1つのコーパスにすると、レマ化された”pretty”は何種類出現するか？また、出現数の最も多いものから、最も少ないものを引いた数はいくらか？
5. COCA の SPOKEN ジャンルと NEWSPAPER ジャンルの三人称単数現在形動詞（レマ化済み）のリストを作成するとどのようになるか？
6. COCA 全体に feel(レマ化)+like の後ろ 5 語以内に動名詞の ing 型を含む表現 (feels like going etc) のリストを作成するとどのようになるか？<ヒント：動名詞のタグは[*vvg*]である>

例題 2 (応用編)

1. 基本編の 4. で形容詞の pretty のみを選び出し、リストを作成するとどのようになるか？
2. [過去形の動詞+over]の熟語の種類と頻度を SPOKEN と NEWSPAPER の場合で比較した結果のリストを作成するとどのようになるか？
3. 否定の接頭辞 dis を含む動詞の過去形のリストを作成するとどうなるか？
4. SPOKEN と NEWSPAPER を1つのコーパスとし FICTION と MAGAZINE をもう1つのコーパスとして、beat と同様の意味を表す動詞（レマ化）のリストを作成するとどのようになるか？

例題 1 (基本編) :解答

1. 方法は<検索の基本 1.>と同じ。 11347
2. 方法は<検索の基本 2.>と同じ。 172
3. 方法は<検索の基本 3.>と同じ。 3012 と 108848
4. 方法は<検索の基本 4.>と同じ。 5 と 51733
5. WORD にカーソルを合わせ、POST LIST から verb.3SG を選択しクリック。

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION)

[HELP...]

SEC 1: 81,690,404 WORDS

SEC 2: 79,368,267 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO		WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1	RATIO
1	[BE]	1890855	979644	23,146.60	12,343.02	1.88	1	[BE]	979644	1890855	12,343.02	23,146.60	0.53
2	[HAVE]	248238	287638	3,038.77	3,624.09	0.84	2	[HAVE]	287638	248238	3,624.09	3,038.77	1.19
3	[DO]	100164	54588	1,226.14	687.78	1.78	3	[SAY]	166811	77194	2,101.73	944.96	2.22
4	[SAY]	77194	166811	944.96	2,101.73	0.45	4	[DO]	54588	100164	687.78	1,226.14	0.56
5	[COME]	21803	15394	266.90	193.96	1.38	5	[MAKE]	19854	17507	250.15	214.31	1.17
6	[GO]	20128	11698	246.39	147.39	1.67	6	[COME]	15394	21803	193.96	266.90	0.73
7	[SEEM]	19833	14307	242.78	180.26	1.35	7	[SEEM]	14307	19833	180.26	242.78	0.74
8	[WANT]	18997	13185	232.55	166.12	1.40	8	[WANT]	13185	18997	166.12	232.55	0.71

このとき、OPTION で LEMMA でなく WORD にしてしまうと'S が表示されてしまうが、is か has かの判別がつかなくなるので LEMMA にする方がよい。但し、その他の動詞は本来、この場合 LEMMA でも WORD でも頻度は等しくなるはずにもかかわらず、ほんのわずかに異なってしまう。なぜ？

6. WORD に[feel] like を、CONTEXT に[*vvg*]を入力。下は OPTION が WORD

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT]

	<input type="checkbox"/>	CONTEXT	TOT <input type="checkbox"/>
1	<input type="checkbox"/>	GOING	578
2	<input type="checkbox"/>	GETTING	228
3	<input type="checkbox"/>	TALKING	127
4	<input type="checkbox"/>	CRYING	126
5	<input type="checkbox"/>	TAKING	87
6	<input type="checkbox"/>	TRYING	83
7	<input type="checkbox"/>	SAYING	82

(参考) OPTION を BOTH WORD にすると次のようになる。

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT]

	<input type="checkbox"/>	CONTEXT	TOT <input type="checkbox"/>
1	<input type="checkbox"/>	FEEL GOING	339.5
2	<input type="checkbox"/>	FELT GOING	173
3	<input type="checkbox"/>	FEEL GETTING	151.5
4	<input type="checkbox"/>	FEEL TALKING	100
5	<input type="checkbox"/>	FELT CRYING	67
6	<input type="checkbox"/>	FEELS GOING	51

例題 2 (応用編) 解答

1. WORD に[pretty].[j*]を入力。OPTION は WORD を選択。[j*]入力は POS LIST から adj.ALL を選択してもよい。

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT]

	<input type="checkbox"/>	CONTEXT	TOT <input type="checkbox"/>
1	<input type="checkbox"/>	PRETTY	10890
2	<input type="checkbox"/>	PRETTIEST	511
3	<input type="checkbox"/>	PRETTIER	482
TOTAL			11883

2. WORD に[vvd] overを入力。[vvd]は POS LIST から verb.ED を選択してもよい。タグは[v?d*]で入力されるが結果はたぶん、同じになる。

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION)

SEC 1: 81,690,404 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	TOOK OVER	806	1698	9.87	21.39	0.46
2	WENT OVER	726	213	8.89	2.68	3.31
3	CAME OVER	546	282	6.68	3.55	1.88
4	TURNED OVER	149	180	1.82	2.27	0.80
5	LOOKED OVER	121	88	1.48	1.11	1.34
6	FLEW OVER	111	113	1.36	1.42	0.95
7	GOT OVER	108	61	1.32	0.77	1.72
8	WALKED OVER	97	102	1.19	1.29	0.92

SEC 2: 79,368,267 WORDS

	WORD/PHRASE	TOKENS 2	TOKENS 1	PM 2	PM 1
1	TOOK OVER	1698	806	21.39	9.87
2	CAME OVER	282	546	3.55	6.68
3	WENT OVER	213	726	2.68	8.89
4	TURNED OVER	180	149	2.27	1.82
5	PRESIDED OVER	174	95	2.19	1.16
6	SAID OVER	118	74	1.49	0.91
7	FLEW OVER	113	111	1.42	1.36
8	HANDED OVER	106	61	1.34	0.75

3. WORD に dis*.[vvd]を入力。

SEE CONTEXT: CLICK ON WORD OR SELECT WORDS + [CONTEXT]

	<input type="checkbox"/>	CONTEXT	TOT <input type="checkbox"/>
1	<input type="checkbox"/>	DISCOVERED	16107
2	<input type="checkbox"/>	DISAPPEARED	8100
3	<input type="checkbox"/>	DISCUSSED	4901
4	<input type="checkbox"/>	DISMISSED	2663
5	<input type="checkbox"/>	DISPLAYED	2478
6	<input type="checkbox"/>	DISAGREED	2019
7	<input type="checkbox"/>	DISLIKED	841

4. WORD に[[=beat]].[v*]を入力。名詞の hit や strike を省くことができる。

SEE CONTEXT: CLICK ON WORD (ALL SECTIONS) OR NUMBER (SPECIFIED SECTION)

SEC 1: 161,058,671 WORDS

	WORD/PHRASE	TOKENS 1	TOKENS 2	PM 1	PM 2	RATIO
1	HIT	28284	26117	175.61	161.19	1.09
2	BEAT	12467	10860	77.41	67.03	1.15
3	BROKE	10541	12250	65.45	75.60	0.87
4	BROKE	9158	11024	56.86	68.04	0.84
5	STRUCK	6879	8164	42.71	50.39	0.85
6	BREAKING	5299	6070	32.90	37.46	0.88
7	HITTING	5027	4295	31.21	26.51	1.18
8	BROKEN	4446	6764	27.60	41.75	0.66

SEC 2: 162,026,400 WORDS

	WORD/PHRASE
1	HIT
2	BREAK
3	BROKE
4	BEAT
5	STRUCK
6	STIR
7	BROKEN
8	BREAKING

