

本文書では、2025 年 4 月時点での Sketch Engine の機能・インターフェイスに基づき、Sketch Engine の主要な機能と利用方法を説明しています。

## Sketch Engine の特徴

Sketch Engine は多くの大規模コーパスを擁する多言語対応コーパス分析システムです。機能は非常に多機能です。自分で作成したコーパスデータをアップロードして、Sketch Engine のシステムで分析したり、キーワードを指定することで独自のウェブコーパスを作成したりすることもできます。

## Sketch Engine を利用する方法

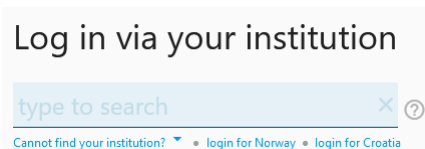
次の URL にアクセスします。

<https://auth.sketchengine.eu/#login>

ログイン画面が表示されたら、[Institutional login]を選択します。



Log in via your institution という画面が表示されたら、ボックスに[Seijo University]または[成城大学]と入力して[Enter]キーを押すか、途中まで入力することで表示される選択肢をクリックします。



大学のログイン画面が表示されるので、大学のアカウントでログインします。

初めて学認経由で Sketch Engine を利用する場合は、学認での認証後に Sketch Engine のログインページで Returning user or new user? というウィンドウが表示されるので、[NEW ACCOUNT]ボタンをクリックし、Sign up というウィンドウで必要な情報を入力し、使用条件等を適宜確認・同意し、[SIGN UP]ボタンをクリックします。“Verify your e-mail address for Sketch Engine”というタイトルの確認のメールが届くので、確認のリンクをクリックすると Sketch Engine が利用可能になります。

次回以降は大学選択と大学アカウントでログインするだけで利用可能です。

ログインページで右上に氏名と家のアイコンが表示され、“You are already logged in. You can go to Sketch Engine.”と表示されている場合は、ログイン済であるため、右上の家のアイコンをクリックすれば Sketch Engine のページが表示されます。

正しく認証されると、DASHBOARD という初期画面または SELECT CORPUS という検索対象言語選択画面が表示されます。SELECT CORPUS 以外の画面が表示された場合は、左のメニュー部分にカーソルを持っていくと、各モードへのボタンが表示されるので、[Select Corpus]をクリックします。



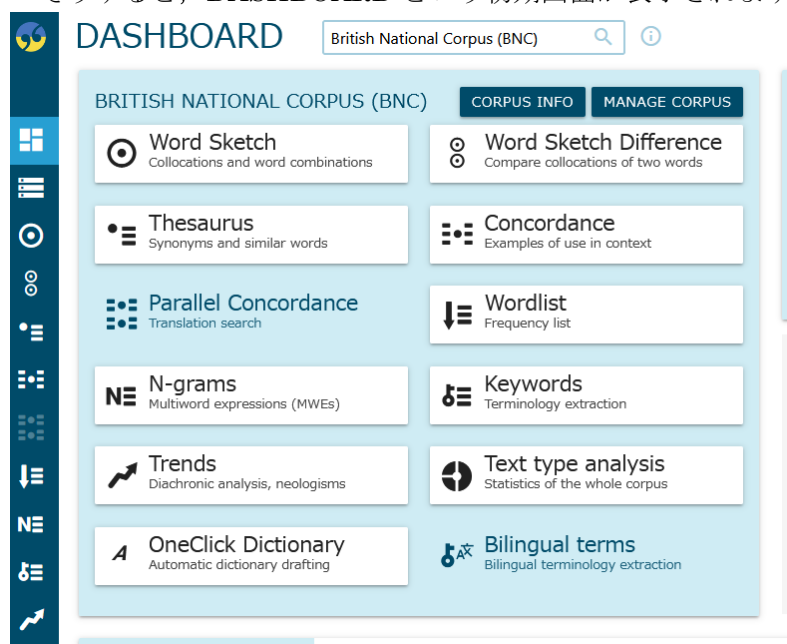
本資料では BNC (British National Corpus ; 1990 年前後を中心とする時期に収集されたイギリス英語約 1 億語を収集した汎用目的コーパス ; 書き言葉約 9,000 万語と話し言葉約 1,000 万語のデータから成る) を利用するので、ここでは[ADVANCED]タブをクリックします。([BASIC]タブの[ENGLISH]を選択すると、English Web 2021 (enTenTen21)が選択されます。)



利用可能な全てのコーパスがコーパス名のアルファベット順で表示されるので、English の British National Corpus (BNC)を選択します。「tagged by CLAWS」が付いていない方の British National Corpus (BNC)を選択してください。

English	British Law Report Corpus	8,515,749	...
English	British National Corpus (BNC)	96,132,981	...
English	British National Corpus (BNC), tagged by CLAWS	96,052,598	...

そうすると、DASHBOARD という初期画面が表示されます。



## Sketch Engine を利用したコーパス検索の例

### (A) 単語を検索して使い方をみる

ここでは単語“description”を例にします。ダッシュボードで[Concordance]をクリックし、コンコーダンスのモードにします。(または、左のメニュー部分にカーソルを持っていくと、各モードへのボタンが表示されるので、[Concordance]をクリックします。)[ADVANCED]タブをクリックし<sup>1</sup>、Query type で[word]を選択し、Word に[description]を入れて[GO]ボタンをクリックします。(標準でチェックボックスがチェックされている A=a は大文字と小文字を区別しないという設定です。)

British National Corpus (BNC)

BASIC **ADVANCED** LEARN

Simple search

Query type  
simple  
lemma  
phrase  
**word**  
character  
CQL

Part of speech  
any  
adjective  
adverb  
conjunction  
determiner  
noun  
numeral  
particle

Word  
**description**  
✓ A = a ?

Subcorpus  
none (the whole corpus)

Macro  
none

Filter context  
Text types

GO

そうすると、20 例が KWIC 形式 (keyword in context の略で検索語が中央に揃って表示される形式) で表示されます。ページ最下部の右にあるボタン・数字入力欄で他のページに移動することができます。

word

description

5,083

45.25 per million tokens

0.0045%

Details

Left context

KWIC

Right context

1

Written books a... be less convincing.

</s><s>4.4</s><s>

A general description of the two Leicester pavements is giv

2

Written books a... vas Plays Anytime Anywhere, and she gave a vivid description of playing in air-raid shelters, in the s

3

Written books a... showing chaos in that a fully quantum mechanical description of the phenomena is conceivable: the

4

Written books a... landscape.

</s><s>

Since motorways fit this same description they would seem to be the perfect loc

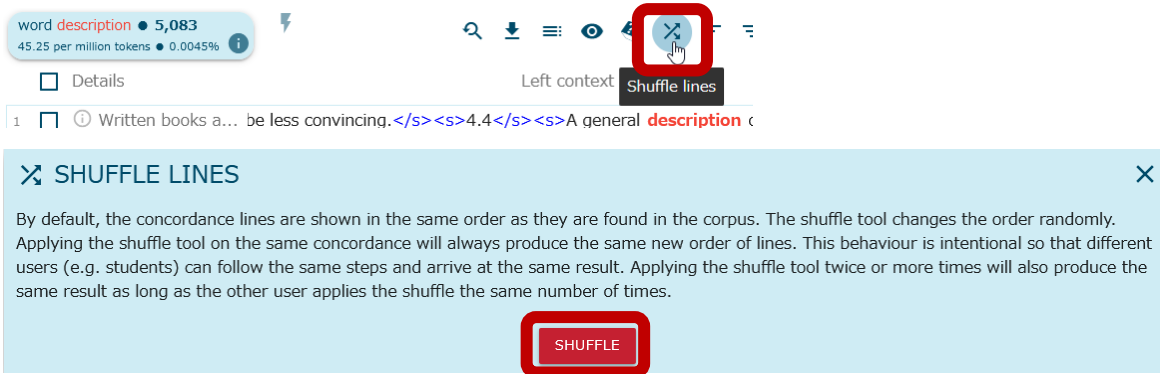
5

Written books a... gical point of view, the most obvious omission is a description of the various stable periodic orbits ol

例文の表示順は BNC のファイルの ID 順です。左上に表示される数値は、検索条件（この場合は“description”という単語）に合致する件数 (5,083) と、それを 100 万語あたりの相対頻度に換算した値 (45.25) です。検索に時間がかかる場合は、件数のところの数値に「…」などが表示されて順次書き換わるので、件数を調べたい場合には、最終的な数値が表示されるまで待つ必要があります。

<sup>1</sup> 初期状態で選択されている BASIC モードでは入力した語句がレマ（原形）として解釈され、全ての変化形も含めて検索されます。(下記(I)を参照。)

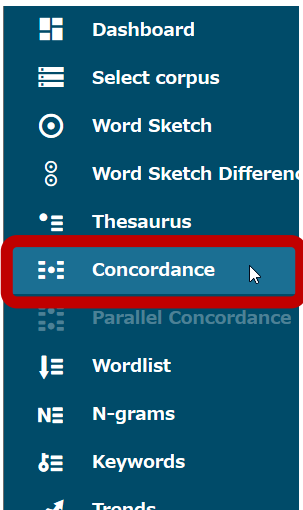
次に、上部にある[Shuffle lines]のボタンをクリックし、[SHUFFLE]ボタンをクリックします。



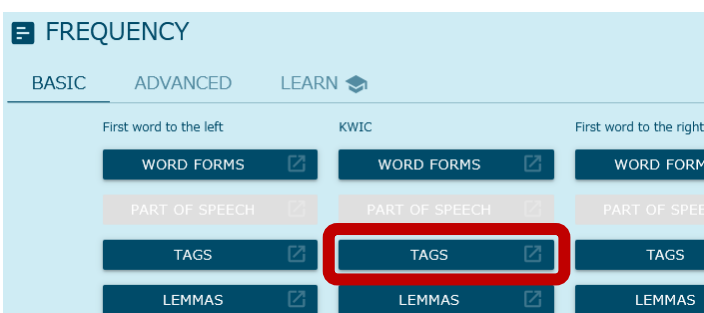
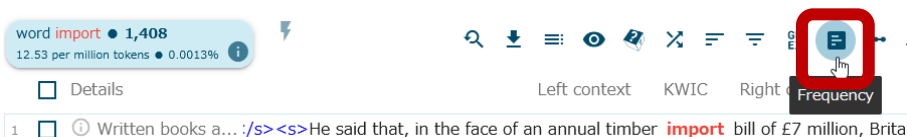
そうすると、表示順がランダムになります。標準ではファイルの ID 順で例文が表示されるため、同一のテキストでの似たような使用例が集中して表示されることがあります。そのような場合には、検索した表現の特殊な使い方が目立ってしまい、本来の多様な使い方が見えにくくなってしまうことがあります。そのような事態を避けるための一つの有効な方法が、このランダム順表示です。

## (B) 検索条件に合致する件数の内訳を調べる

(A)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニュー部分にカーソルを持っていくと、各モードへのボタンが表示されるので、[Concordance]をクリックします。



そうすると、コンコーダンスモードの検索の初期画面に戻ります。BASIC モードになっている場合には[ADVANCED]モードに変更し、Query type で[word]を選択し、Word に import を入れて[GO]ボタンをクリックすると、“import”の検索結果が表示されます。その状態で、上部にある[Frequency]のボタンをクリックし、KWIC の列の[TAGS]ボタンをクリックします。



そうすると、品詞別の集計結果が表示されます。**Tag** は品詞情報を表していて、**NN** は単数または不可算の名詞、**VV** は一般動詞（**be** と **have** を除く）の不定形（原形）、**VVP** は一般動詞（**be** と **have** を除く）の 3 人称単数を除く現在形を意味します。

(3 items, 1,408 total frequency)

	Tag	Frequency	Relative <sup>2</sup>
1 <input type="checkbox"/>	NN	1,066	9.49 ***
2 <input type="checkbox"/>	VV	281	2.50 ***
3 <input type="checkbox"/>	VVP	61	0.54 ***

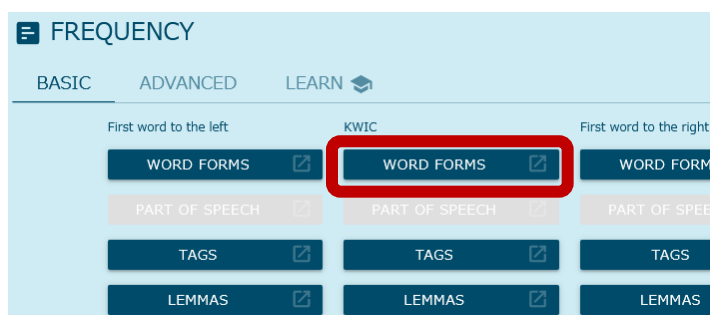
各行の右にある[...]をクリックし、[Only this]をクリックすると、そのタグでの使用に該当する用例が表示されます<sup>2</sup>。（それ以外の用例は[Whole concordance but this]で表示されます。）

次に、上部にある[CHANGE CRITERIA]をクリックし、KWIC の列の[WORD FORMS]ボタンをクリックします。

Frequency

CHANGE CRITERIA

BACK TO CONCORDANCE



そうすると、大文字と小文字も区別した、表層形（実際の語形）別の集計結果が表示されます。

(3 items, 1,408 total frequency)

	Word	Frequency	Relative <sup>2</sup>
1 <input type="checkbox"/>	import	1,345	11.97 ***
2 <input type="checkbox"/>	Import	57	0.51 ***
3 <input type="checkbox"/>	IMPORT	6	0.05 ***

レマ検索（下記(I)参照）やワイルドカード検索（下記(J)参照）をした場合には、この機能で実際の語形の分布を知ることができます。

### (C) 1 画面での表示件数を変更する

(B)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[word]を選択し、Word に **statistics** を入れて[GO] ボタンをクリックします。検索結果が表示された状態で下部にある Rows per page を[100]に変更すると、1 画面に 100 例が表示されるようになります。



表示件数の設定は、そのまま検索を続ける間は保存され、他の条件で検索をしても、100 例が表示されます。

<sup>2</sup> 複数のタグの用例を表示したい場合は、各行の左にあるチェックボックスにチェックを入れて、下部に表示される[SELECTED]をクリックします。

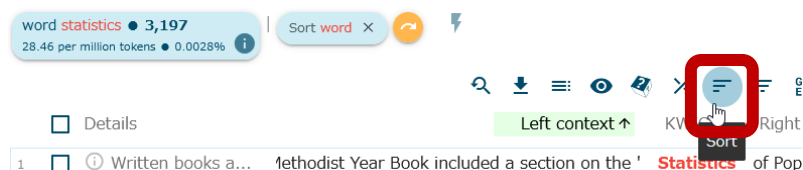
#### (D) 例文をソートする

(C)から続けて試している場合はそのまま続けます。(D)から読み始めた場合は、コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[word]を選択し、Word に **statistics** を入れて[GO]ボタンをクリックします。検索結果が表示された状態で上部にある[Right context]をクリックすると、検索した語句（この場合は“statistics”）の直後の位置でソートされ、表現の実際の使い方が分かりやすくなります。（ただし、最初の方には記号類が並びます。）

Left context    KWIC    **Right context**

次に、上部にある[KWIC]をクリックすると、キーワードを対象にソートすることができます。今回の例の場合は検索対象のキーワードは **statistics** という形のためのためソートの意味はありませんが、レマ検索（下記(I)参照）やワイルドカード検索（下記(J)参照）をした際には、実際の語形によってソートすることができます。

次に、上部にある[Left context]をクリックします。そうすると、左の単語でソートされた結果が表示されます。詳細なソート条件を指定するには、上部にある[Sort]のボタンをクリックします。



Sort first by の図は、[Left context]で左の単語でソートされた状態を表しています。[ADD NEXT CRITERION]をクリックします。



左側の[2]をクリックし、[GO]ボタンをクリックします。



そうすると、左 1 語の位置・左 2 語の位置の優先順でソートされます。

## (E) 使用分野などの内訳と特定分野での使用例を表示する

(D)から続けて試している場合はそのまま続けます。(E)から読み始めた場合は、コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[word]を選択し、Word に **statistics** を入れて[GO]ボタンをクリックします。検索結果が表示された状態で上部にある[Frequency]のボタンをクリックし、[TEXT TYPES]をクリックします。

The screenshot shows the top of the Frequency interface. At the top, there's a search bar with 'word statistics' and '3,197' results. Below it, there are tabs for 'Details', 'Left context', 'KWIC', 'Right context', and 'Frequency'. The 'Frequency' tab is selected and highlighted with a red box. Below the tabs, there's a section titled 'FREQUENCY' with sub-tabs 'BASIC', 'ADVANCED', and 'LEARN'. Under 'BASIC', there are three columns: 'First word to the left', 'KWIC', and 'First word to the right'. Each column has buttons for 'WORD FORMS', 'PART OF SPEECH', 'TAGS', and 'LEMMAS'. To the right of these columns, there's a 'More presets' section with buttons for 'TEXT TYPES' (highlighted with a red box) and 'LINE DETAILS'.

そうすると、詳細なサブコーパス別の頻度が **Frequency** の数値で表現されます。[Show relative in text types] をチェックすると、各サブコーパスにおける 100 万語当たりの相対頻度 (Relative in text type) が表示され、[Show relative density] をチェックすると、そのサブコーパスでの使用頻度がコーパス全体での使用頻度と比較してどの程度の割合であるかという情報 (Relative density) がグラフと共に表示されます。

The screenshot shows the 'Frequency' section of the interface. At the top, there are two buttons: 'CHANGE CRITERIA' and 'BACK TO CONCORDANCE'. Below them, there's a section with two checkboxes: 'Show relative in text types' (checked) and 'Show relative density' (unchecked). Below this, there are two tables. The first table is titled '(5 items, 3,197 total frequency)' and has columns 'Text type' and 'Frequency'. The second table is titled '(4 items, 3,197 total frequency)' and has columns 'Publication date' and 'Frequency'.

Text type	Frequency
1 <input type="checkbox"/> Written books and periodicals	2,383 ...
2 <input type="checkbox"/> Written miscellaneous	610 ...
3 <input type="checkbox"/> Spoken context-governed	172 ...
4 <input type="checkbox"/> Written-to-be-spoken	20 ...
5 <input type="checkbox"/> Spoken demographic	12 ...

Publication date	Frequency
1 <input type="checkbox"/> 1985-1993	2,714 ...
2 <input type="checkbox"/> Unknown	274 ...
3 <input type="checkbox"/> 1975-1984	179 ...

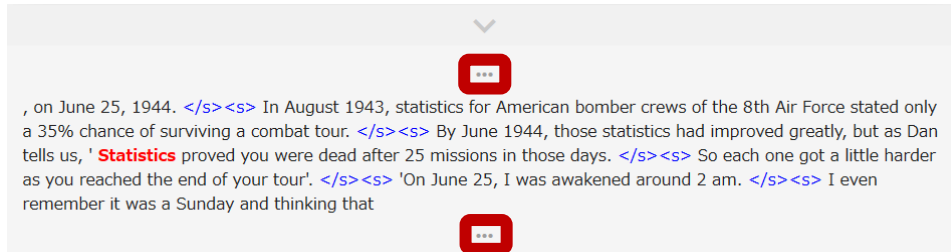
この中で、Written books and periodicals の右にある[...]をクリックし、[Only this]をクリックすると、書籍と定期行物での検索結果 2,383 件が表示されます。

ある表現がどのような分野で使われるのかという分布を知りたい場合や、特定の分野に限定してある表現の使われ方を見たい場合に、この機能が役立ちます。



## (F) 出典・原文を表示する

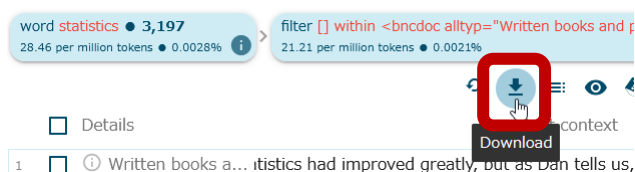
(E)から続けて試している場合はそのまま続けます。(F)から読み始めた場合は、コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[word]を選択し、Word に **statistics** を入れて[GO]ボタンをクリックします。検索結果が表示された状態で、各行頭の ⓘ またはジャンル情報が表示されている部分（灰色）をクリックすると当該データの出典が表示されます。各行中央のマッチ部分（赤字）をクリックすると当該箇所付近の原文が表示されます。原文の上と下にある[...]をクリックすると、前後の文脈を広げて、より大きな文脈を確認できます。



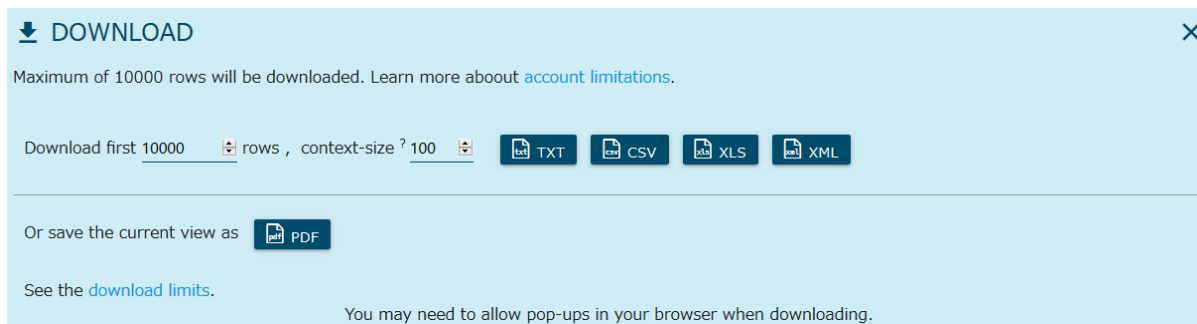
上部にある ▼ をクリックすると、原文表示を消すことができます。

## (G) 検索結果を保存する

(F)から続けて試している場合はそのまま続けます。(G)から読み始めた場合は、コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[word]を選択し、Word に **statistics** を入れて[GO]ボタンをクリックします。検索結果が表示された状態で上部にある[Download]のボタンをクリックします。



適宜オプションを変更して保存したいファイル形式のボタンをクリックすると、検索結果を保存することができます。Sketch Engine で提供されているコーパスの検索結果をダウンロードする際にダウンロードできる最大の行数は 10,000 行です。



ダウンロードされるファイルは、行頭に表示されるテキストタイプの情報を含んでいます。



## (H) フレーズを検索する

(G)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[phrase]を選択し<sup>3</sup>、Phrase に **variety of** を入れて[GO]ボタンをクリックすると、“variety of”というフレーズを含む例が表示されます。

The screenshot shows the search interface with the 'ADVANCED' tab selected. The 'Query type' dropdown menu is open, and 'phrase' is selected. The search term 'variety of' is entered in the 'Phrase' field. The 'GO' button is highlighted with a red box.

## (I) 変化形も含めて検索する

(H)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[simple]を選択し<sup>4</sup>、Simple に **study show** を入れて[GO]ボタンをクリックします。

The screenshot shows the search interface with the 'ADVANCED' tab selected. The 'Query type' dropdown menu is open, and 'simple' is selected. The search term 'study show' is entered in the 'Simple' field. The 'GO' button is highlighted with a red box.

<sup>3</sup> 初期状態で選択されている[simple]モードで variety of を検索すると、varieties of も検索結果に含まれますが、phrase モードだと variety of のみが検索されます。(下記(I)を参照。)

<sup>4</sup> フレーズではなく単語をレマ（原形）で検索する場合には lemma モードも利用できます。

そうすると、study shows/showed や studies show/showed などがまとめて検索されます。

1	<input type="checkbox"/> ⓘ Written books a... ial in London, would be extended, after initial <b>studies showed</b> that it had
2	<input type="checkbox"/> ⓘ Written books a... id 37 per cent of hydrocarbons. </s><s> The <b>study showed</b> that childr
3	<input type="checkbox"/> ⓘ Written books a... st in a growing list of official and quasi-official <b>studies showing</b> how the pr
4	<input type="checkbox"/> ⓘ Spoken context-... ig or anything like that and er certainly all the <b>studies show</b> that the nc
5	<input type="checkbox"/> ⓘ Spoken context-... our slots? </s><s> No you're not but in fact <b>studies show</b> after about
6	<input type="checkbox"/> ⓘ Spoken context-... orbing much anyway </s><s> It's true, most <b>studies show</b> I mean if, I
7	<input type="checkbox"/> ⓘ Written books a... lear if this would succeed </c><c> A recent <b>study showed</b> that nollut

simple モードでは、入力した語句の中で単語のレマ（原形）に相当する語については全ての変化形が検索対象となり、変化形で入力した語についてはその形だけが検索対象になります。混在している場合でも同様に、例えば study showed を検索すると、study showed と studies showed が検索されます<sup>5</sup>。

## (J) ワイルドカードを使って検索する

(I)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコードダンスモードの検索の初期画面に戻ります。コンコードダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[simple]を選択し、Simple に **test \* hypothesis** (test の後と hypothesis の前にそれぞれ半角スペースを 1 つ入れます) を入れて[GO]ボタンをクリックします。

そうすると、test (またはその変化形) と hypothesis (またはその変化形) の間に 1 語が入った検索結果が表示されます。「\*」は前後にスペースがある場合には任意の 1 語を表すワイルドカードとして機能するためです。「\*」を複数使うことで、任意の語数を指定でき、例えば、任意の 2 語は「\* \*」(2 つの「\*」の間に半角スペースを 1 つ入れます) で表すことができます。

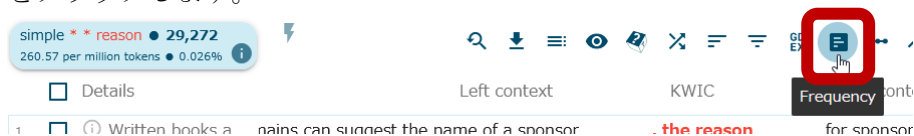
左のメニューにある[Concordance]をクリックしてコンコードダンスモードの検索の初期画面に戻ります。コンコードダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[simple]を選択し、**c?t**を検索します。そうすると、cat や cut とこれらの変化形などが検索されます。「?」は語中に置かれた場合には任意の 1 文字を表し、さらに simple モードではレマ検索が行われるためです。

左のメニューにある[Concordance]をクリックしてコンコードダンスモードの検索の初期画面に戻ります。コンコードダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[simple]を選択し、**c\*t**を検索します。そうすると、cat や cut とこれらの変化形に加えて、craft や client とこれらの変化形なども検索されます。「\*」は語中に置かれた場合には 1 つ以上の文字を表し、さらに simple モードではレマ検索が行われるためです。

lemma・phrase・word の各モードでは上記のワイルドカードは使えません<sup>6</sup>。

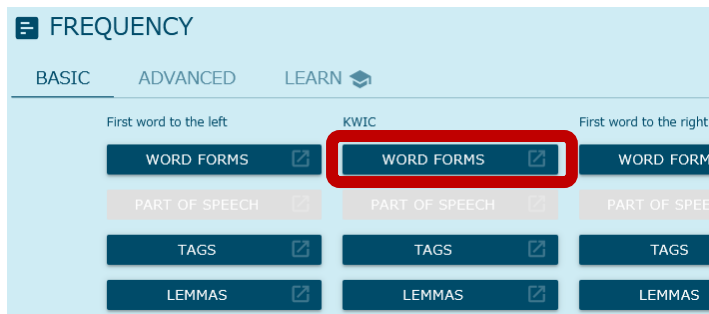
## (K) 使用頻度の高い表現を調べる

(J)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコードダンスモードの検索の初期画面に戻ります。コンコードダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[simple]を選択し、Simple に **\* \* reason** (2 つの「\*」の間と 2 つ目の「\*」の後にそれぞれ半角スペースを 1 つ入れます) を入れて[GO]ボタンをクリックします。検索結果が表示された状態で上部にある[Frequency]のボタンをクリックし、KWIC の列の[WORD FORMS]をクリックします。



<sup>5</sup> この検索方法では study は名詞に限定されないため、studied（後置修飾の過去分子や関係節の最後に位置する過去形など）の直後に showed が続くパターンも検索されます。

<sup>6</sup> ただし、正規表現という特別な記法を使ってより詳細なパターンを検索することはできます。例えば語中に inter を含み ly で終わる語は、word モードで正規表現を使って[.+inter.\*ly]で調べることができます。また、character モードでは指定した文字列が含まれる語を全て検索することができます。



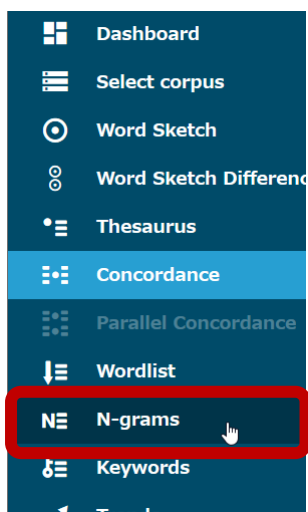
そうすると、2 単語+reason(s)で構成される表現のリストが表示されます。(カンマやピリオドなどの句読点もそれぞれ 1 語扱いとなります。) [Show percentage of concordance lines]をチェックすると、コンコーダンスの総件数に占める割合がグラフと共に表示されます。



(7,461 items, 29,272 total frequency)

	Word	Frequency	Relative ?
1	<input type="checkbox"/> . The reason	1,134	10.09 ***
2	<input type="checkbox"/> of the reasons	686	6.11 ***
3	<input type="checkbox"/> is no reason	632	5.63 ***
4	<input type="checkbox"/> For this reason	574	5.11 ***
5	<input type="checkbox"/> for some reason	569	5.07 ***
6	<input type="checkbox"/> for this reason	406	3.61 ***

次に、左のメニュー部分にカーソルを持っていくと、各モードへのボタンが表示されるので、[N-grams]をクリックします。



[ADVANCED]タブをクリックし、N-gram length で[3]をクリックし、A=a をチェックし、Frequency min に[10]を指定し、[GO]ボタンをクリックします。（この処理には数秒の時間がかかります。）

N-GRAMS

BASIC **ADVANCED** LEARN

N-gram length ? 2 **3** 4 5 6

Attribute ? word

Frequency min ? 10 Frequency max ? 0 Subcorpus ? none (the whole corpus)

☐ Nest n-grams ? ☐ Include nonwords ? ☐ Exclude these words: ?

☒ A = a ?

☐ Key n-grams ?

Additional criteria ?

- all**
- starting with letters
- ending with letters
- containing letters
- starting with word
- containing word
- ending with word
- matching regular expression

Text types ?

**GO**

そうすると、頻度の高い3語連鎖が表示されます。

### 3-grams, word (lowercase) (items: 688,068 , total frequency: 29,526,279 )

N-gram	Frequency ?	N-gram	Frequency ?	
1 i do n't	37,229 ...	18 to be a	11,796 ...	35
2 one of the	35,580 ...	19 there is no	11,675 ...	36
3 the end of	21,058 ...	20 you do n't	11,668 ...	37
4 as well as	18,253 ...	21 per cent of	11,344 ...	38
5 part of the	17,335 ...	22 it would be	10,628 ...	39
6 do n't know	16,009 ...	23 at the end	10,608 ...	40
7 out of the	15,644 ...	24 i did n't	10,374 ...	41
8 there is a	15,310 ...	25 it is not	10,364 ...	42
9 some of the	15,191 ...	26 in terms of	10,072 ...	43
10 a number of	15,161 ...	27 i ca n't	10,032 ...	44
11 a lot of	14,647 ...	28 it is a	9,776 ...	45

特定のキーワードを指定することなく、使用頻度の高いフレーズを調べる際には、この方法を使います。

#### (L) 句読点を検索する

(K)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[simple]を選択し、Simple に finally, (句読点も単語扱いなので「,」の前に半角スペースを1つ入れます) を入れて[GO]ボタンをクリックすると、“finally,”を含む例が表示されます。

### **(M)** 縮約形や所有の“...’s”を検索する

(L)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[simple]を選択し、Simple に **that 's** (「that」の後に半角スペースを 1 つ入れます) を入れて[GO]ボタンをクリックすると、“that’s”を含む例が表示されます。

縮約形を検索するには、2 つの部分に分けて入力する必要があります。例えば、“I’m”を検索するには **I 'm** で検索する必要があります。また、縮約に伴って形が変わる場合は、検索する際に注意が必要です。“will not”の縮約形である“won’t”は **wo n't** で検索する必要があります<sup>7</sup>。

ただし、所有等を表す“...’s”は検索時には分割してはいけません。例えば **Jack's** (スペースを 入れません) で検索する必要があります。

---

<sup>7</sup> “gonna”と“wanna”は Sketch Engine ではこれらの形で 1 語という扱いのため分割しません。

**(N)** サブコーパスを指定して検索する

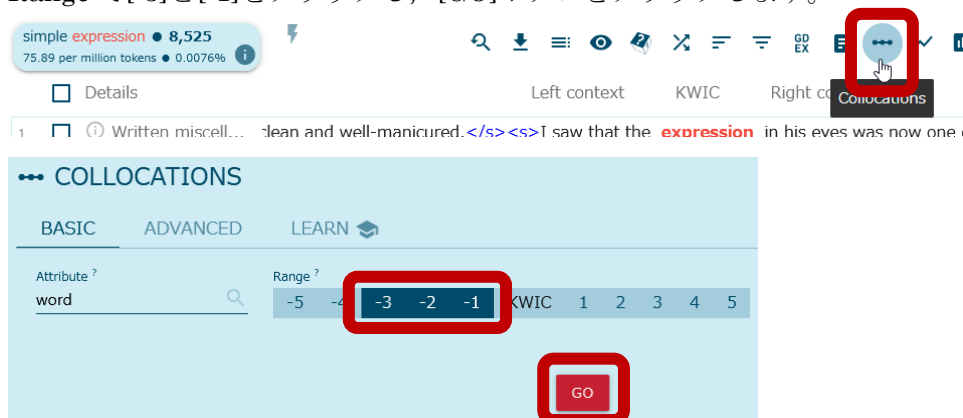
(M)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[simple]を選択します。[Text types]をクリックして、新たに表示される中で左上の[Text type]をクリックして[Spoken context-governed]と[Spoken demographic]を選択し、Simple に you mind を入れて[GO]ボタンをクリックします。

The screenshot shows the 'ADVANCED' search mode. The 'Query type' dropdown is set to 'simple'. The search term 'you mind' is entered in the search box. The 'Text types' section shows 'Spoken context-governed' and 'Spoken demographic' selected. The 'GO' button is highlighted at the bottom.

そうすると、話し言葉データでの“you mind”の用例が表示されます。テキストタイプを指定しないと 423 件が表示されますが、話し言葉に限定すると 137 件が表示されます。

## (Q) コロケーションを調べる

(N)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[simple]を選択し、Simple に **expression** を入れて[GO]ボタンをクリックします。検索結果が表示された状態で上部にある[Collocations]のボタンをクリックし、Range で[-3]と[-1]をクリックし、[GO]ボタンをクリックします。



そうすると、**expression** とその変化形の左 3～1 語の位置に現れる語のリストが表示されます。標準では、コロケーションの結びつきの指標のひとつである LogDice の値でソートされています。

	Word	Cooccurrences ?	Candidates ?	T-score	MI	LogDice ↓
1	<input type="checkbox"/> facial	137	509	11.70	11.79	8.96 ...
2	<input type="checkbox"/> freedom	170	5,364	13.01	8.71	8.65 ...
3	<input type="checkbox"/> gene	74	2,022	8.58	8.91	7.84 ...
4	<input type="checkbox"/> finds	45	3,070	6.67	7.59	6.99 ...
5	<input type="checkbox"/> linguistic	40	2,390	6.30	7.78	6.91 ...

続いて、第 2 位の **freedom** の行の右にある[...]をクリックし、[Only this]をクリックすると、**freedom+0～2** 語の任意の単語+**expression** の用例が表示されます。(この場合は実際には間に入る単語数が 0 であるパターン、つまり **freedom expression** という表現はありません。)

なお、オプションの設定時に Attribute を lemma に変えるとレマ（原形）単位で集計することができます。



## (P) 品詞を指定して検索する

(O)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[word]を選択し、Part of speech に[noun]を選択し、Word に **study** を入れて[GO]ボタンをクリックします。

The screenshot shows the 'ADVANCED' search mode. The 'Query type' dropdown menu is open, and 'word' is selected. The 'Part of speech' dropdown menu is also open, and 'noun' is selected. The 'Word' input field contains the text 'study'. The 'GO' button at the bottom right is highlighted with a red box.

そうすると、名詞の **study** の検索結果が表示されます。ただし、変化形である studies は含まれません。

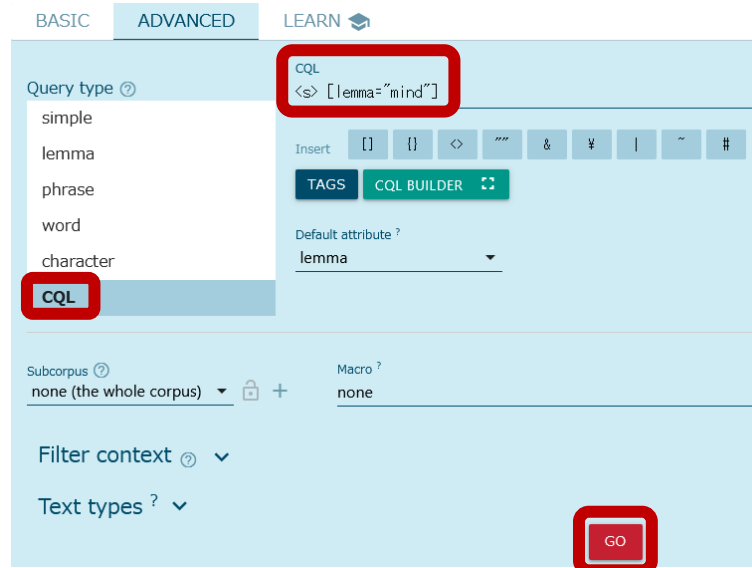
左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[lemma]を選択し、Part of speech に[noun]を選択し、Lemma に **study** を入れて[GO]ボタンをクリックします。

The screenshot shows the 'ADVANCED' search mode. The 'Query type' dropdown menu is open, and 'lemma' is selected. The 'Part of speech' dropdown menu is also open, and 'noun' is selected. The 'Lemma' input field contains the text 'study'. The 'GO' button at the bottom right is highlighted with a red box.

そうすると、名詞の **study** の検索結果が表示されますが、今度は変化形である studies も含まれます。

## (Q) 文頭・文末の位置指定

(P)から続けて試している場合など、何らかの検索結果が表示されている場合は、左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[CQL]を選択し、CQL に `<s> [lemma="mind"]` (「<s>」の後に半角スペースを1つ入れます)を入れて、[GO]ボタンをクリックすると、文(あるいはそれに相当する単位)の最初の位置にある“mind”またはその変化形の使用例が表示されます。

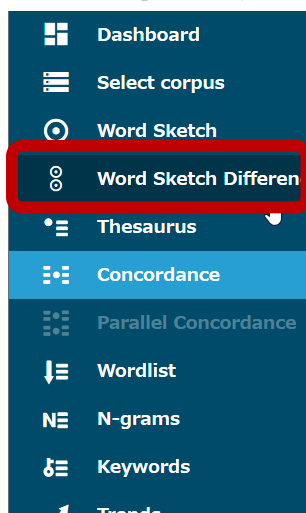


左のメニューにある[Concordance]をクリックしてコンコーダンスモードの検索の初期画面に戻ります。コンコーダンスモードの検索の初期画面で[ADVANCED]モードを選択し、Query type で[CQL]を選択し、CQL に `[lemma="mind"] [word="."] </s>` (「[lemma="mind"]」の後に「[word="."]」の後にそれぞれ半角スペースを1つ入れます)を入れて、[GO]ボタンをクリックすると、文(あるいはそれに相当する単位)の末尾の位置にある“mind”またはその変化形と1文字から成る語(実質句読点)<sup>9</sup>の使用例が表示されます。

BNCでは<s>と</s>がそれぞれ文頭・文末を表しますが、どのようなタグが使われているかはコーパスによって異なります。

## (R) 2語のコロケーションの違いを示す

左のメニュー部分にカーソルを持っていくと、各モードへのボタンが表示されるので、[Word Sketch Difference]をクリックします。



<sup>8</sup> CQLはCorpus Query Languageの略で、コーパスを検索するための記法を表します。CQLの仕様はシステムによって異なり、ここで紹介しているCQLはあくまでSketch Engineで利用できるCQLの記法です。

<sup>9</sup> CQLのモードではピリオドは任意の1文字を表す正規表現として解釈されます。ピリオドそのものを表すためには、前に半角のバックスラッシュをつけて、“\.”とする必要があります。

First lemma に **large** を入れ、Second lemma に **big** を入れて、[GO]ボタンをクリックします。

BASIC ADVANCED LEARN

First lemma ?  
large

Second lemma ?  
big

GO

そうすると、2つの語のコロケーションの違いが構造パターン別に表示されます。緑色が濃いほど **First lemma** で指定した語 (**large**) と、赤色が濃いほど **Second lemma** で指定した語 (**big**) との共起が強いということが示されます。

large 47,370x    big 31,432x

"large/big" and/or ...				subjects of "be large/big"			
large	106	0	...	pore	47	0	...
small	404	73	...	shield	31	0	...
several	92	24	...	writ	28	0	...
open	70	22	...	area	27	0	...
white	101	64	...	room	29	12	...
enough	68	75	...	time	51	28	...
black	98	127	...	size	37	31	...
good	24	112	...	house	16	18	...
strong	18	91	...	hole	11	11	...
fat	11	84	...	something	18	73	...
big	0	254	...	bit	6	82	...
great	0	390	...	lot	0	30	...

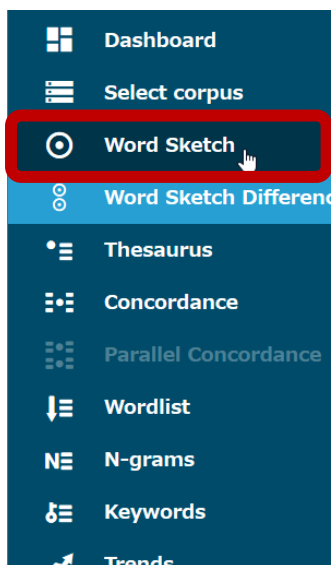
  

nouns modified by "large/big"				infinitive objects of "large/big"			
quantity	498	0	...	accommodate	23	0	...
amount	798	9	...	cover	12	0	...
proportion	569	7	...	allow	15	0	...
number	3,461	51	...	carry	10	0	...
part	758	122	...	hold	22	14	...
company	745	321	...	support	9	6	...
city	260	211	...	take	23	21	...
house	336	371	...	fit	6	9	...
difference	63	210	...	go	0	9	...
business	74	288	...	get	0	12	...
problem	47	352	...	put	0	7	...
bang	0	154	...	handle	0	6	...

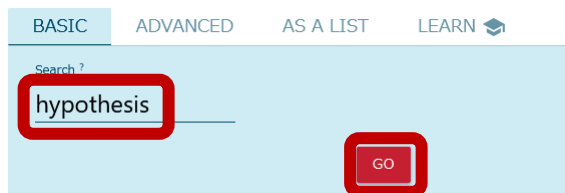
例えば nouns modified by "large/big" は large/big が修飾する名詞を表していて、large は量や程度の大きさを、big は把握される質的な大きさを表すことが多いということが分かります。このように、コロケーションの違いから意味の違いが読み取れることは少なくありません。

(S) コロケーション・フレーズを一覧表示する

左のメニュー部分にカーソルを持っていくと、各モードへのボタンが表示されるので、[Word Sketch]をクリックします。



Search に **hypothesis** を入れ、[GO]ボタンをクリックします。

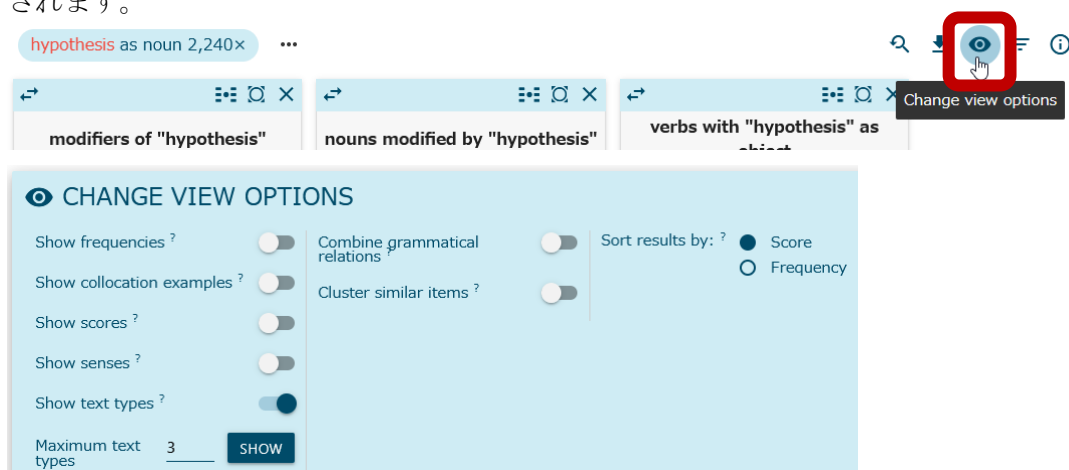


そうすると、hypothesis の共起語が、構造パターン別に、結びつきの強さを示すスコア順に表示されます。

hypothesis as noun 2,240x ... 🔍 ⬇

modifiers of "hypothesis"	nouns modified by "hypothesis"	verbs with "hypothesis" as object
<b>expectation</b> ... <ul style="list-style-type: none"><li>concentrated in: W_commerce ?</li><li>usually in: W_commerce ?</li></ul>	<b>testing</b> ...	<b>test</b> ... <ul style="list-style-type: none"><li>concentrated in: W_commerce ?</li><li>concentrated in: W_ac_soc_science ?</li></ul>
<b>null</b> ...		<b>support</b> ...
<b>rational</b> ... <ul style="list-style-type: none"><li>concentrated in: W_commerce ?</li><li>only in: W_commerce ?</li></ul>		<b>formulate</b> ...
<b>Samuelson</b> ...		<b>compete</b> ... <ul style="list-style-type: none"><li>usually in: W_ac_tech_engin ?</li><li>usually in: Written miscellaneous ?</li></ul>
<b>speculative</b> ...		<b>refute</b> ...
<b>accelerationist</b> ...		<b>reject</b> ...
<b>clausal</b> ...		<b>span</b> ...
<b>efficiency</b> ...		<b>disprove</b> ...

検索結果が表示された状態で上部にある[Change view options]のボタンをクリックすると、表示の設定が表示されます。



[Show frequencies]のトグルボタンをオンにすると各コロケーションの度数が表示され、[Show collocation examples]のトグルボタンをオンにすると用例が表示され、[Show scores]のトグルボタンをオンにすると結びつきの強さを示す値（LogDice）が表示されます。（[Cluster similar items]のトグルボタンをオンにすると、構造パターンが近いもの（多くの場合、類義語や反義語）がまとめて表示されます。）

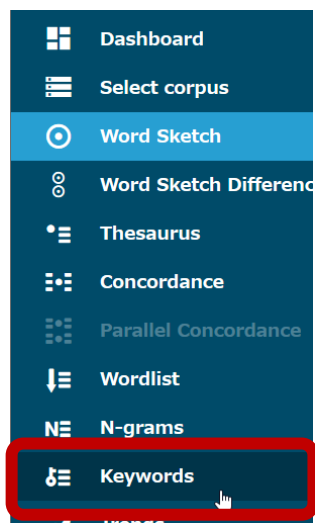
verbs with "hypothesis" as object の test の右にある[⋮]をクリックすると、4つの項目が表示されます。

verbs with "hypothesis" as object	verbs with "hypothesis"
<b>test</b> 120 10.0 test this hypothesis • concentrated in: W_commerce ? • concentrated in: W_ac_soc_science ?	hypothesis + test hypothesis + test
<b>support</b> 105 8.1 support the hypothesis	hypothesis + test test
<b>formulate</b> 15 8.1 formulating hypotheses	

1 番上の項目をクリックすると、検索語(hypothesis)とコロケーション(hypothesis を目的語とする動詞 test)のコンコーダンスが表示されます。2 番目の項目は、コンコーダンス表示後に行う操作をマクロとして保存しておいたものがあればそれを呼び出して実行することができます。3 番目の項目をクリックすると、test + hypothesis というパターン内での、それぞれの語のコロケーション等が表示されます。4 番目の項目をクリックするとコロケーション (test) の類義語が表示されます。類義語は、元のコロケーション画面での品詞（この場合は test は動詞）は考慮されていません（この場合は名詞の test の類義語が表示されます）が、左上のボタンで品詞を変更することができます。

(T) 他のコーパスと比較して特徴的な表現を調べる

左のメニュー部分にカーソルを持っていくと、各モードへのボタンが表示されるので、[Keywords]をクリックします。



[ADVANCED]タブをクリックし、Focus subcorpus に[Spoken transcripts]を選択し、Reference corpus に[British National Corpus (BNC)]を、Reference subcorpus に[Written Academic]を選択し、Minimum frequency に[10]を指定し、[GO]ボタンをクリックします。

The screenshot shows the 'ADVANCED' tab of a search interface. On the left, a red box highlights the 'Focus subcorpus' dropdown set to 'Spoken transcripts', the 'Reference corpus' dropdown set to 'British National Corpus (BNC)', and the 'Reference subcorpus' dropdown set to 'Written Academic'. To the right, the 'Minimum frequency' is set to '10' in a red box. Further right, a 'Focus on' slider is set to '1' (neutral). On the far right, several checkboxes are visible: 'A = a ?' (checked), 'At least one alphanumeric ?' (checked), 'Only alphanumeric ?' (checked), 'Include nonwords ?' (unchecked), 'Exclude these words: ?' (unchecked), and 'From list ?' (unchecked). Below these are three settings panels: 'Keywords settings' (Attribute: lemma, Matching regex: .\*), 'Terms settings' (Matching regex: .\*), and 'N-grams settings' (Attribute: word, Matching regex: .\*, N-gram length: 2, 3, 4, 5, 6). At the bottom, there is a 'GO' button in a red box. A warning message at the bottom left states 'Subcorpus and Text types cannot be combined'.

そうすると、BNC の Written Academic のサブコーパスと比較して、BNC の Spoken transcripts のサブコーパスで特徴的なフレーズが表示されます。

Spoken transcripts ▾ ×

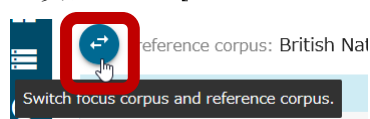
SINGLE-WORDS ✓ MULTI-WORD TERMS ✓

reference corpus: British National Corpus (BNC) subcorpus: Written Academic

Term	Term	Term	Term
1 er er ...	11 other day ...	21 cup of tea ...	31 fox repor
2 little bit ...	12 mr chairman ...	22 er er er ...	32 good boy
3 i dunno ...	13 new settlement ...	23 sort of thing ...	33 p p
4 erm er ...	14 lot of people ...	24 bye bye ...	34 u k
5 last night ...	15 last week ...	25 sort of things ...	35 good mo
6 oh aad ...	16 erm erm ...	26 a m h ...	36 not some

上部の[SINGLE-WORDS]タブをクリックすると単語が表示されます。

次に左上の[Switch focus corpus and reference corpus.]のボタンをクリックします。



そうすると、対象コーパスと参照コーパスを入れ替えた結果が表示されます。(Spoken transcripts のサブコーパスと比較して、Written Academic のサブコーパスで特徴的な単語・フレーズが表示されます。)

再度左上の[Switch focus corpus and reference corpus.]のボタンをクリックすると、検索対象コーパスが Spoken transcripts のサブコーパスに戻ります。