

Sketch Engine 操作マニュアル

東京外国語大学 投野研究室
(2018 年度改訂版)

<目次>

1. CONCORDANCE 画面を使いこなす	3
1.1. 検索の基礎	3
1.2. CQL 検索	5
1.2.1. 表層形検索	5
1.2.2. レマ検索	5
1.2.3. 品詞指定の仕方	5
1.2.4. より複雑な検索	6
1.2.5. 文境界を指定した検索	7
1.2.6. CQL attribute の指定	7
1.3. CQL BUILDER	7
2. さまざまなオプション機能	8
2.1. WORDLIST	8
2.2. THESAURUS / SKETCH DIFF	10
3. コーパスの作成	13
3.1. ファイルのアップロード	13
3.2. WEBBOOTCAT	16
4. 練習問題	18
4.1. 基本検索 1	18
4.2. 基本検索 2	19
4.3. CQL 検索 1 -基本検索と関連付けて-	20
4.4. CQL 検索 2	21
4.5. 頻度表	22

1. Concordance 画面を使いこなす

1.1. 検索の基礎

(1) Query Type を選ぶ

The screenshot shows the 'Query type' section of the Concordance search interface. It includes a 'Simple query' input field, a 'Make Concordance' button, and navigation links for 'Query types', 'Context', and 'Text types'. The 'Query type' section has radio buttons for 'simple', 'lemma', 'phrase', 'word', 'character', and 'CQL'. Below these are input fields for 'Lemma:', 'Phrase:', 'Word form:', and 'Character:', each with a 'PoS:' dropdown menu. There is also a 'match case' checkbox and a 'Default attribute:' dropdown menu. At the bottom, there are 'Make Concordance' and 'Clear All' buttons.

- Simple query...単語、フレーズ検索（各単語の活用形も含む＝lemma 検索）
- Lemma...単語のみ。活用形も含む。品詞指定可。
- Phrase...表層形のためのフレーズ検索（＝活用形は含まれない）。
- Word...表層形のための単語検索。品詞指定可。大文字・小文字指定可。
- Character...特定のアルファベットの並びで検索（＝接頭・接尾語の検索可）。

	単語	フレーズ	活用形	品詞指定	大文字・小文字指定
Simple	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>		
Lemma	<input type="radio"/>		<input type="radio"/>	<input type="radio"/>	
Phrase		<input type="radio"/>			
Word	<input type="radio"/>			<input type="radio"/>	<input type="radio"/>

(2) Context

The screenshot shows the 'Context' section of the Concordance search interface. It includes a 'Simple query' input field, a 'Make Concordance' button, and navigation links for 'Query types', 'Context', and 'Text types'. The 'Query type' section has radio buttons for 'simple', 'lemma', 'phrase', 'word', 'character', and 'CQL'. Below these are input fields for 'Lemma:', 'Phrase:', 'Word form:', and 'Character:', each with a 'PoS:' dropdown menu. There is also a 'match case' checkbox and a 'Default attribute:' dropdown menu. At the bottom, there are 'Make Concordance' and 'Clear All' buttons. The 'Context' section has two filters: 'Lemma filter' and 'PoS filter'. Each filter has a 'Window:' dropdown menu (set to 'both') and a 'tokens.' dropdown menu (set to '5'). The 'Lemma filter' has a 'Lemma(s):' input field and a dropdown menu (set to 'all'). The 'PoS filter' has a 'PoS:' dropdown menu (set to 'adjective') and a dropdown menu (set to 'all').

- Lemma Filter...前後○語以内の指定の単語（lemma）の有無によって絞り込み。
 - PoS Filter...前後○語以内の指定の品詞の有無によって絞り込み。
- ※いずれも複数指定可で、[all] にすると指定した単語/品詞の全てが含まれたもの、[any] にすると指定したもののうち少なくとも 1 つが含まれたもの、[none] にするとそのうちどれ

も含まれないものがコンコーダンス上に現れる。

(3) Text types

...サブコーパスによる絞り込みができる。サブコーパスを編集することもできる。

(4) Query 画面の活用

• Sort (並べ替え)

Left... 中心語の左隣 (L1) の語のアルファベット順で並べ替え

Right... 中心語の右隣 (R1) の語のアルファベット順で並べ替え

Node... 中心語のアルファベット順で並べ替え

References... ファイル情報で並べ替え

Query **write** 9,875 > Sort **Left** 9,875 > Sort **bncdoc.id/ 0>0** 9,875 > Shuffle 9,875 > Sort **Left** 9,875 (88.03 per million)

Page 1 of 494 Go Next | Last Concordance is sorted. Jump to: [dropdown]

HR9	this Quigley gets too much,' he said, ' write here. Give me a bit of time to get settled
BNA	Madam'. </p><p> NB If an advertisement says, ' write for application form' then keep the letter
HH7	and had consequently been treated as a ' write off' by its insurer. </p><p> According to
EUS	the control unit of the computer sends a ' write ' signal to the store. After some delay
G00	original Canon laser engine is called ' write black' because it charges up those areas
FT0	longstanding tradition that Mozart could ' write down whole compositions, previously composed
HWF	times that the file has been opened for ' write ', the File Protection is updated to include
HAC	the Source disk or at least that it is ' write ' protected. If you leave it unmarked and
H7X	read only' memory. More strictly it is ' write once, read many times' memory. The pattern
J25	they find it. A bestseller in its own ' write ' - and no work of fiction either - Guinness

↑ファイル情報

↑L1 ↑node ↑R1

• Sample (標本抽出) ...ランダムに指定した数のコンコーダンスを出すことができる。

• Filter (絞り込み)

指定した語が中心語の前後〇語に現れる (positive) もしくは現れない (negative) ものでコンコーダンスをさらに絞り込むことができる。

• Frequency (頻度集計)

Frequency... 中心語のみだけでなく、前後〇番目にある単語の指定した形での頻度集計ができる。

Node tags... 中心語の時制ごとに頻度集計ができる。

Node forms... 中心語の表層形ごとに頻度集計ができる。

Doc IDs... ファイルの種類ごとに頻度集計ができる。

Text Types... テキストの種類ごとに頻度集計ができる。

→さらに P/N (Positive/Negative) でコンコーダンスを絞り込むことができる。

• Collocation (共起語)

Attribute... 共起語を word/tag/lempos/lemma... のどれによって分類するか決定。

Range... 共起するのが中心語の前後〇語以内にするか決定。

Minimum Frequency in corpus... 共起語 のコーパス内の総頻度の下限を決定。

Minimum Frequency in given range... 中心語 + 共起語の頻度の下限を決定。

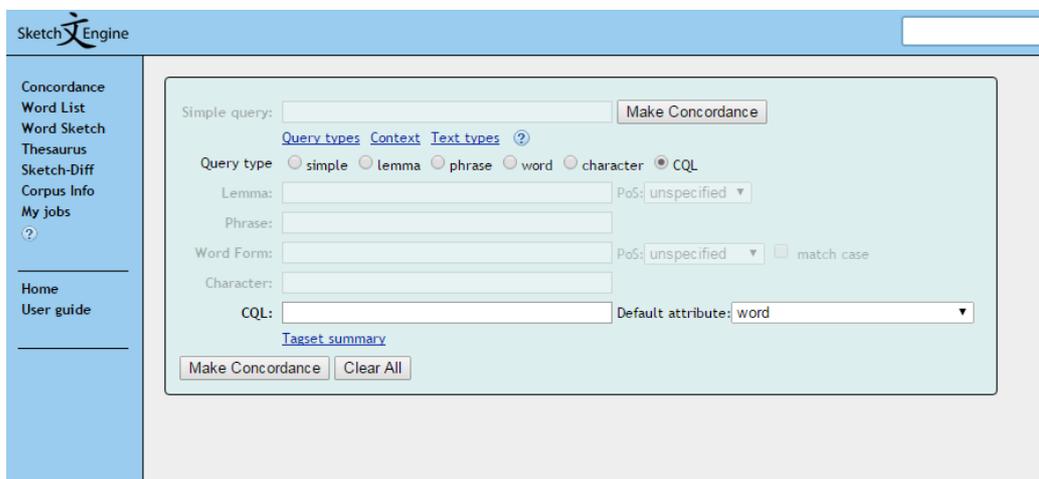
→さらに P/N (Positive/Negative) でコンコーダンスを絞り込むことができる。

• Visualize (グラフ化)

...全コーパス内における中心語の分布をグラフ化できる。全コーパスをどれくらいの束に分けるかは、グラフ下の数値で調節。

1.2. CQL 検索

機能：CQL(= Corpus Query Language)を使った検索式により、柔軟な検索ができる
CQL：1990年代初期に独 University of Stuttgart の IMS が開発



CQL の一般式：[\[attribute="value"\]](#)

- Attribute: word, lemma, tag, lempos がはいる
- [\[word="value"\]](#)
- [\[lemma="value"\]](#)
- [\[tag="value"\]](#)
- [\[lempos="value"\]](#)
- Value: 検索したい正規表現を含む文字列が入る。文字列は "" (ダブルクォーテーションマーク) で挟む。

1.2.1. 表層形検索

turn (表層形) を検索する場合の式

[\[word="turn"\]](#)

["\[tT\]hank"](#)

1.2.2. レマ検索

turn (レマ) を検索する場合の式

[\[lemma="turn"\]](#)

1.2.3. 品詞指定の仕方

turn の名詞を検索する場合の式

[\[lemma="turn"& tag="N.*"\]](#) または [\[lempos="turn-n"\]](#)

※ .(ピリオド)：任意の1文字

※ *：0以上n個 / +：1以上n個 / ?：0個または1個 / !：～以外

※ ※タグの記号に関しては [Tagset summary](#) を参照

Common tags:

POS	tag	lempos
noun	N.*	-n
verb	V.*	-v

adjective	J.*	-j
adverb	RB.?	-a
conjunction	CC	-c
preposition	IN	-p
pronoun	PP.?	-d

1.2.4. より複雑な検索

tag=""を使った他の複雑な例

"confuse.*" [tag="IN" | tag="PP"]

"confuse.*" ([tag="IN" | [tag="PP"])

"confuse.*" [tag="IN|PP"]

turn + 名詞 + 前置詞 の検索式

[lemma="turn"&tag="V.*"] [tag="N.*"] {1,2} [tag="PRP"]

※ [] と [] の間にはスペースがあってもなくても ok

※ {} で検索スパンを指定 ({1,2} は 1 語から 2 語 という意味)

この式で CQL 検索をしてみると・・・

Sketch Engine

Query turn, V.*, N.*, PRP 620 (5.50 per million)

Page 1 of 31 Go Next Last

J2W	, said" the Americans and Europeans have	turned whales into	a sacred animal, like the Hindu cow... If
J2R	</p> Waste and Recycling Old incinerator to	turn rubbish into	energy and money <p> Plans have been unveiled
J0P	landed property into individual property and	turning land into	a freely saleable commodity like anything
J0W	Sex appeal? In a way, the fat lady doctor	turned sex on	to its head and in those prepermissive
J5J	in politics </p><p> Victorian values have	turned Britain into	a more divided country. Homelessness and
JNF	services it helps youngsters help themselves by	turning moans into	action. Really we're trying to give young
J5F	way, the intention of these orders er to	turn auditors into	er snoopers or narks er and to do so I
J5G	responsibilities and in some way, as I say, to	turn auditors into	snoopers and narks er er and make more
J3B	Agriculture Minister Sotiris Chatzigakis. "	Turning regions like	Angistri into wildlife refuges automatically
J32	. They criticize them as a blueprint for	turning Britain into	Europe's toxic dump. </p><p> Guardian 13
J32	</p> Waste and Recycling Australian project	turns bottles into	pipes <p> A company in South Australia, Rib
J32	, Rib Loc, has developed a technique for	turning plastic bottles into	pipes. The pipe-making process was first
J18	which travel in groups: hummingbirds would	turn trees into	individual feeding territories and thus
HRJ	television interviewers, et hoc genus omne . It	turned Blackpool into	a sort of electoral Convention a l'Americaine
HRF	Sheppey. At the age of eighteen, Doris was	turning heads on	the island and was selected as Sheerness
HRC	follow him. </p><p> But then he could have	turned north to	the Tay in safety. With a tired army. But
HRC	, and especially Scone. Then they should	turn south past	Forteviot and march against us. By that
HRD	Disc Interactive <p> Whatever may be done to	turn CD-ROM into	a vehicle for multimedia, it is never likely
HRD	option making it possible, if desired, to	turn CDTV into	an overt computer system. </p><p> Although
HTP	the self in this characteristic way was to	turn egoism into	altruism, and <corr> aggression </corr> into

Page 1 of 31 Go Next Last

◇複雑な検索式◇

"help to do" vs. "help do" のコンテキスト差を調べる為、文を抽出する式

[lemma="help"&tag="V.*"] [word="to"] ? [tag="V.I"]

※動詞の不定詞形タグは VBI (be)、VDI (do)、VHI (have)、VVI (lexical verbs)

名詞 + be + -ed 形で終わる動詞 の検索式

[tag="N.*"] [lemma="be"] [tag="V.*&word="*.ed"]

look/bring + up/down の検索式

[lemma="look|bring"&tag="V.*"][tag!="V.*"]{0,5}"up|down"

OR の意味のバー (|) の例

[tag="JJ.*"] [tag="N.*"] "and|or" [tag="N.*"]

1.2.5. 文境界を指定した検索

✧within を使った検索式✧

文境界を指定した式

[word="confus.*"][tag!="V.*"]*[word="by"]within<s/>

動詞で始まり、動詞で終わる連鎖の中にあるすべての名詞句を抽出する式

[tag="N.*"]+within[tag="VB.*"][*][tag="VB.*"]

(動詞句を一緒にとってこないための式)

※ Query within Query という式も可能である

※ **Containing** もある (cf. **Sketch Engine**)

1.2.6. CQL attribute の指定

CQL attribute を指定することで、少し検索式が短くなります。が、簡単になるのかは不明。エラーが起きたときにどこが間違っているのか特定しにくくなりそうな印象

default attribute	通常の CQL の検索式	attribute を指定した場合の検索式の例
word	[word= "refil"]	"refill"
tag		"work"
lempos		
<u>lempos-lc</u>		
<u>lemma</u>		
<u>lc=word</u>		
<u>lemma-lc</u>		

1.3. CQL builder

CQL builder という機能が追加されていますが、CQL を知っていないと活用できないので基礎的なことはしっかり抑えること！

1.4. 新しいインターフェイス

2018 年より新しいインターフェイスが導入されているが、右下に表示される[Back to the original interface]をクリックすることで従来のインターフェイスに戻ることができる。

2. さまざまなオプション機能

2.1. Word list

コーパス（サブコーパス）内から語彙表を抽出できる。

上から順に

- Subcorpus：サブコーパスを選択、新たに作成できます。

- Search attribute: word, lemma, tag (POS)などが選べます。

Use n-grams では n 語の連鎖の語彙表を作成できます。

ここまでで検索してリストを作成できます。また以下のオプションを使うこともできます。

ここで、コーパスを BNC、サブコーパスを Written_Medium_Book、Search attribute を lemma にして word list を作成すると以下ようになります。

Sketch Engine British National Corpus (BNC) / Written_Medium_Book

Sketch Engine homepage

- Home
- Concordance
- Word list
- Word sketch
- Thesaurus
- Sketch diff
- Trends
- Corpus info
- My jobs
- User guide ↗

- Save
- Change options

Word list

Corpus: British National Corpus (BNC)
Subcorpus: Written_Medium_Book

Page [Next >](#)

<u>lemma</u>	<u>Freq</u>
the	3,246,582
be	2,120,850
of	1,701,293
and	1,386,210
to	1,352,316
a	1,121,858
in	1,026,932
have	660,482
that	553,152
it	521,370
for	420,060
he	409,936
not	402,571
I	398,018
as	385,700
with	343,420
on	338,677
you	290,158
his	280,464
she	265,410
by	263,321
at	262,123
do	253,091

(頻度が高い順に lemma を並べた語彙表)

【Filter options】

- Regular expressions: 正規表現で検索できます。 `.*` がワイルドカード (何が何文字入っても OK) を表すので、「`th.*`」で検索すると `the, that, this` 等の語彙表が作成されます。(その他、`+`、`?`、`!` などがあります)
- Minimum frequency: 最小頻度を指定できます。
- Maximum frequency: 最大頻度を指定できます。
- Whitelist: 語彙表に含めたい特定の単語リストがある場合、アップロードできます。
- Blacklist: 語彙表に含めたくない特定の単語リストがある場合、アップロードできます。
- Include non-words: 句読点や記号などを含めたいときに使います。

【Output options】

- Frequency figures: Hit counts →粗頻度 (= raw frequency)

Document counts →語彙表中の単語を含むドキュメントの数
ARF (Average Reduced Frequency) →ひとつの単語が近距離
(e.g. 同じドキュメント内) で複数回現れるときに頻度を調節する
機能です。

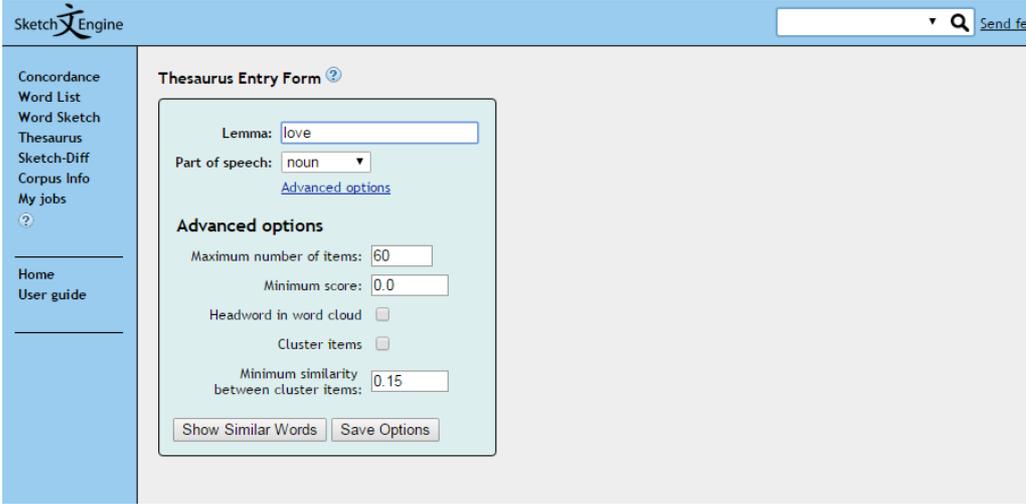
- Output type: Simple Keywords →他の (サブ) コーパスを参照して比較する場合にキーワードを抽出できます。
- Reference (sub)corpus →比較するサブコーパスを選択できます。
- Prefer: rare/common words →頻度が高い・低い単語に高いスコアが当てられるように調節できます。
- Change output attribute(s):

2.2. Thesaurus / Sketch diff

機能：同義語、あるいは異なる2語間の共起関係を調べることができる

◇ 実際に見てみましょう — 検索対象コーパス：BNC ◇

検索画面 (名詞の"love"で検索、POS を指定)



Sketch Engine

Concordance
Word List
Word Sketch
Thesaurus
Sketch-Diff
Corpus Info
My jobs

Home
User guide

Thesaurus Entry Form

Lemma: love

Part of speech: noun

Advanced options

Maximum number of items: 60

Minimum score: 0.0

Headword in word cloud

Cluster items

Minimum similarity between cluster items: 0.15

Show Similar Words Save Options

図 1

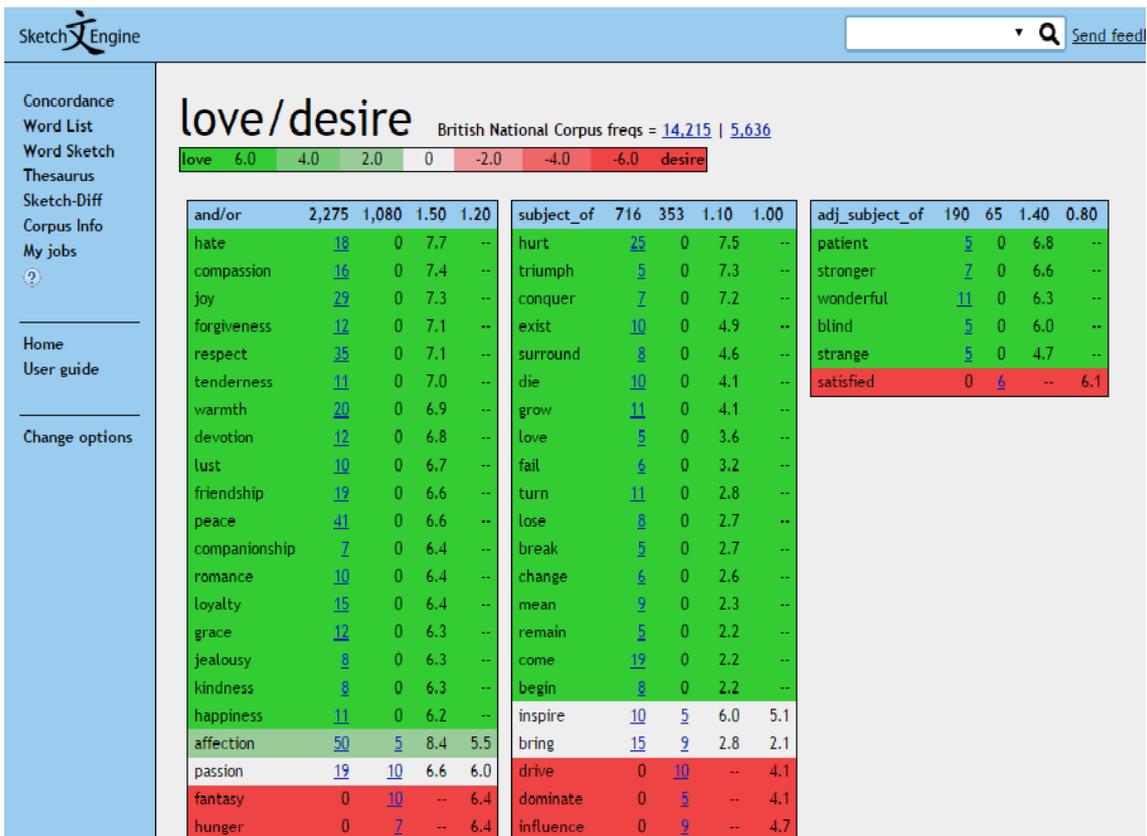


図3 (下に結果が続く)

Word sketch 機能の2単語比較バージョンで、Thesaurus からだけでなく、メニュー左の Sketch-Diff からこの機能を使うことができる。

この図3では、名詞 love と desire の違いをコロケーションの観点から学び取られる。

- ☆ 左から、and/or で並列共起する名詞、主語のときに共起する動詞、そして修飾を受けて共起する形容詞の3つの観点から love と desire の違いが分かる (他にもたくさんの違いを示す表が下に続く)

- ☆ 緑であればあるほど love との共起が多く、赤であるほど desire との共起が多い (ここで love には satisfied は修飾されていない・・・ということは・・・)

3. コーパスの作成

3.1. ファイルのアップロード

☆自分がコーパス化したいパソコン上のテキストファイルをスケッチエンジンにアップロードすると、既存のコーパスと同じようにスケッチエンジン上で検索できるようになります。

〈テキストの整形〉 アップロードしたいファイルに xml タグをつける (適宜)。
xml → 情報を記述するためのタグ

```
<bookstore>
  <book category="COOKING">
    <title lang="en">Everyday Italian</title>
    <author>Giada De Laurentiis</author>
    <year>2005</year>
    <price>30.00</price>
  </book>
  <book category="CHILDREN">
    <title lang="en">Harry Potter</title>
    <author>J K. Rowling</author>
    <year>2005</year>
    <price>29.99</price>
  </book>
  <book category="WEB">
    <title lang="en">Learning XML</title>
    <author>Erik T. Ray</author>
    <year>2003</year>
    <price>39.95</price>
  </book>
</bookstore>
```

< w3schools.com(http://www.w3schools.com/xml/xml_tree.asp)より >

【タグ付与例】

(i)ヘッダー (開始タグ)

```
<xml>
<doc Level="〇〇" Title="テキスト名">
<text Types="〇〇">
```

(ii) フッター (終了タグ)

```
</text>
</doc>
</xml>
```

このようにテキスト内にコンピューターが理解できるような形で情報を付与すると、アップロードした後に自動的にサブコーパスに分けてくれたり、検索結果画面でテキストの細かい情報が表示できたりするようになります！

〈スケッチエンジンへのアップロード〉

(i) 新規のコーパスを作成する

The screenshot shows the 'Corpora' management interface. On the left, a sidebar lists actions: '+ Create corpus', '+ WebBootCaT', '+ Upload TMX', 'Parallel corpora', 'Compare corpora', and 'My jobs'. The main area has tabs for 'Recent', 'My own', 'Featured', 'Parallel', and 'All'. Below the tabs, a table shows 'No corpora found' with columns for 'Language', 'Name', and 'Words'. At the bottom of the table, there are three links: '+ Create new corpus', '+ WebBootCaT', and '+ Upload TMX'. A blue box highlights the sidebar, and an arrow points to the 'WebBootCaT' link with the text 'いずれかをクリック' (Click either one).

(ii) コーパスの名前と扱う言語の設定

Sketch Engine

About Home Settings Change password Log out

Search in Help

user: Prof. Yukio Tono used words: 20% / 10,000,000 days left: 267 Search in Kaken RLD

Create new corpus

Corpus name

Language

Cancel Create

(iii) ファイルの選択

English: Add new file: Step 1

Supported file types: .doc, .docx, .htm, .html, .pdf, .ps, .tar.bz2, .tar.gz, .tgz, .tmx, .txt, .vert, .xml, .zip.

It is possible to upload multiple documents in an archive file. Supported archive types include .zip, .tar, .tar.gz, and .tar.bz2. By default, the contents of the archived documents are extracted into a single text file. If you wish to expand the archive into individual files instead, please select this option at the following page after uploading the archive. Expanding the archive is also necessary if it contains vertical files. Files with an unknown extension will be ignored. Vertical files included in the archive need to contain the following tab-separated values: word, tag, lempos (in this order).

You are responsible for the copyright and other intellectual property issues of the uploaded content.

Upload from disk ファイルを選択 選択されていません

Download from location

Use file or directory on the server

show files in subdirectories

FTP to the.sketchengine.co.uk at port 10021 to upload files. Use the same user name and password as for logging into this web interface.

Paste text

→ライブラリからファイルを選択し、Next をクリックしてアップロード

(iv) ファイルの種類と文字コードを選択

English: Add new file:
 001_A1_International_English_for_Speakers_of_Other_Languages_Book_1_P

File added successfully:
 001_A1_International_English_for_Speakers_of_Other_Languages_Book_1_Preliminary_Listening.xml.txt
 Detected character encoding: utf_8

File type: Plain text

Character encoding: UTF-8 (all languages)

If some characters are not displayed correctly in the preview, changing this may help.

クリック

File preview (plain text)

Showing bytes 1..1500 / 54928

```
<xml>
<file CEFR="A1" Title="International English for Speakers of Other Languages">
<text Skills="Reading">
How do we say the letters A to Z in English? Look at the letters. Do you know
how we say them? Do you know them when you hear them? Put a circle around the
```

(v) すべてのファイルをアップロードし終わったらコーパスを compile する

Sketch Engine

About Home Settings Change password

Search

user: Prof. Yukio Tono used words: 34% / 10,000,000 days left: 252

Search in ELT-CourseBook-02

ELT-CourseBook-01
 kaken_rld

[Add new file](#) /
 [Add data from web using WebBootCaT](#) /
 [Compile corpus](#) /
 [Search corpus](#)

#	Original file	Plain text	Vertical	Tokens	Owner
1	001_A1_Internati...istening.xml.txt	✓	✓	12,255	Prof. Yukio Tono
2	002_A2_Internati...2_Access.xml.txt	✓	✓	18,559	Prof. Yukio Tono
3	003_B1_Internati...istening.xml.txt	✓	✓	23,818	Prof. Yukio Tono
4	004_B2_Internati...or_Listening.txt	✓	✓	36,756	Prof. Yukio Tono
5	005_C1_Internati...5_Expert.xml.txt	✓	✓	47,526	Prof. Yukio Tono
6	006_C2_Internati..._Mastery.xml.txt	✓	✓	34,859	Prof. Yukio Tono
7	007_B1_Close-Up.xml.txt	✓	✓	16,776	Prof. Yukio Tono
8	008_A1_English_Explorer_1.xml.txt	✓	✓	8,699	Prof. Yukio Tono
9	009_B1_English_Explorer_4.xml.txt	✓	✓	17,214	Prof. Yukio Tono
10	010_A1_Holiday_Explorer_1.xml.txt	✓	✓	3,660	Prof. Yukio Tono
11	011_B1_Just_Righ...rmediate.xml.txt	✓	✓	10,587	Prof. Yukio Tono
12	012_B1_Just_Righ...o_Script.xml.txt	✓	✓	4,642	Prof. Yukio Tono

クリック

3.2. WebBootCaT

◎WebBootCaT とは？

インターネットをクロールしてテキストを自動収集し、コーパスを作成する機能。

◎WebBootCaT の基本操作 (図1 参照)

1. Home 画面の左側のメニューから WebBootCaT をクリック。
2. コーパス名と言語を設定。
3. Input type で Seed words/URLs のどちらかを選択。(後述)

WebBootCaT: Create corpus ?

[Get seed words from Wikipedia](#)

Corpus name

Language

WebBootCaT is unavailable for languages which cannot be automatically tokenised.

Input type

Seed words

URLs

Select "URLs" to download data from specified URLs rather than use seed words for finding the URLs.

Seed words

Random tuples will be selected from the seed words to query a search engine. Input 3 to 20 words or multiword expressions. Use space as separator. Enclose multiword expressions into quotes ("").

Compile corpus when finished Automatically compile corpus when WebBootCaT processing is finished.

[Show advanced options](#)

Cancel Next >

図 1 :

WebBootCaT の基本画面

◎WebBootCaT の具体的な使用例 (Seed words/URLs)

★ “Seed word”で Web corpus を作成する。

1. 上記の 1~2 を経て、Input type の Seed words にチェックを入れる。

2. Seed words の欄にキーワード（3～20個）を入力。
3. Seed words を3語ずつランダムに組み合わせたものをインターネット検索にかけた結果が表示される（図2参照）。
4. Next をクリック→自動的にテキストをダウンロード。
5. OK をクリック→コーパスの完成。

図2：URLの一覧

★“URLs”で Web corpus を作成する。

1. 上記の1~2を経て、Input type のURLsにチェックを入れる。
2. URLs の欄に指定するURLを入力（図3参照）。
3. Next→OK→コーパスの完成（図4参照）。

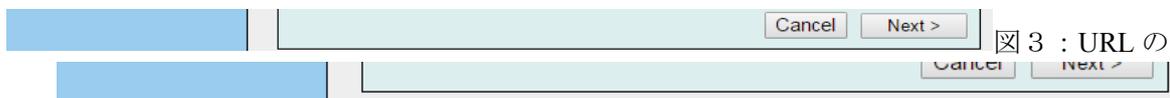


図3：URLの指定

4. 練習問題
4.1. 基本検索 1

練習問題
～基本検索画面編～

以下の 1～8 に挙げる用例を抽出したいときの基本検索を用いた検索方法を考えてみましょう！

1. talk (大文字・小文字・活用形を問わない)を含む用例を抽出したい。
>> []
2. 名詞としての talk の用例を抽出したい。
>> []
3. 名詞としての talk のうち、単数形の用例だけ抽出したい。
>> []
4. 動詞としての Talk (大文字 T で始まる talk)を抽出したい。
>> []
5. talk over (大文字・小文字・活用形を問わない) の用例を抽出したい。
>> []
6. talking over の用例を抽出したい。
>> []
7. trans- (接辞) を含む用例を抽出したい
>> []
8. talk (大文字・小文字・活用形を問わない)のあとに、with, talk, about のいずれも続かない用例を抽出したい。
>> []
9. 直前に形容詞で修飾されている名詞の talks (複数形に限る) の用例を抽出したい。
>> []

4.2. 基本検索 2

練習問題

～基本検索画面編～

以下の 1～11 に挙げる用例を抽出したいときの基本検索を用いた検索方法を考えてみましょう！

4.3. CQL 検索 1 -基本検索と関連付けて-

4.4. CQL 検索 2

4.5. 頻度表

Concordance 画面から frequency 機能を使って、検索結果の頻度内訳を見てみましょう！

疑問1) talk (活用形・大文字・小文字問わない) の検索結果のうち、最も用例数が多い表層形はなん
でしょうか？

[]

疑問2) talk の名詞用法のうち、単数形と複数形のどちらがより多く使われるのでしょうか？

[]

疑問3) talk は名詞用法、動詞用法のどちらがより多く使われるのでしょうか？

[]

疑問4) talk の直後に来る前置詞のうち、最も頻度の高いものはなんのでしょうか？

[]

1.5. ワードリスト

【実際に検索してみよう】

- ①BNC 全体で search attribute を pos、 minimum frequency を 0
- ②BNC の中のサブコーパス Written_Domain_Imaginative で search attribute を lemma、 regular expressions を wh.*、 minimum frequency を 1、 maximum frequency を 0
- ③BNC の中のサブコーパス Written_Domain_Informative で search attribute を word、 regular expression を .*ing、 Frequency figures を Document counts
- ④BNC 全体で search attribute を word、 use n-grams で n=4
- ⑤BNC の中のサブコーパス Written_Medium_Book で search attribute を word、 Output type を Keywords にして、 Reference subcorpus を BNC のサブコーパス Written_Medium_To-be-spoken
- ⑥BNC 全体で search attribute を word、 regular expression を .*ing、 Output type を Change output attributes にして、 lemma, pos, word を選択